2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Addressing Noise and Stochasticity in Fraud Detection for Service Networks

Wenxin Zhang[1,+][0009-0000-8916-6944], Ding Xu[2,+], Xi Xuan[3], Lei Jiang[1], Guangzhen Yao[4], Renda Han[5], Xiangxiang Lang[5(✉)], and Cuicui Luo[1(✉)][0000-0002-4570-5990]

[1] University of Chinese Academy of Sciences, Beijing, China
[2] Harbin Institute of Technology, Harbin, China
[3] City University of Hong Kong, Hong Kong SAR, China
[4] National University of Defense Technology, Changsha, China
[5] Hainan University, Haikou, China
[+] Equal contributions
langxiangxiang@dlut.edu.cn,luocuicui@ucas.ac.cn

Abstract. Fraud detection is crucial in social service networks to maintain user trust and improve service network security. Existing spectral graph-based methods address this challenge by leveraging different graph filters to capture signals with different frequencies in service networks. However, most graph filter-based methods struggle with deriving clean and discriminative graph signals. On the one hand, they overlook the noise in the information propagation process, resulting in a degradation of filtering ability. On the other hand, they fail to discriminate the frequency-specific characteristics of graph signals, leading to the distortion of signal fusion. To address these issues, we develop a novel spectral graph network based on information bottleneck theory (SGNN-IB) for fraud detection in service networks. SGNN-IB splits the original graph into homophilic and heterophilic subgraphs to better capture the signals at different frequencies. For the first limitation, SGNN-IB applies information bottleneck theory to extract key characteristics of encoded representations. For the second limitation, SGNN-IB introduces prototype learning to implement signal fusion, preserving the frequency-specific characteristics of signals. Extensive experiments on three real-world datasets demonstrate that SGNN-IB outperforms state-of-the-art fraud detection methods.

Keywords: Fraud Detection, Graph Neural Network, Heterophily.

1 Introduction

The rapid growth of digital service networks has transformed how services are delivered across industries, enabling seamless interactions across platforms, from financial services to e-commerce. However, this transformation has introduced new risks, particularly from sophisticated fraud schemes that undermine service quality, erode customer trust, and threaten operational stability. In service-oriented industries, where transaction networks and customer relationships form graph-structured systems,

leveraging advanced analytics to address these risks is becoming a critical area for data-driven decision-making. This is particularly evident in financial platforms, where transaction records structured as graphs can reveal intricate patterns characteristic of fraudulent behavior. Developing effective fraud detection methods is essential, not only for enhancing system security but also for maintaining user trust and protecting the reputation of online platforms. As digital fraud schemes continue to grow in complexity, it is crucial to refine and advance graph-based detection methods to keep pace with emerging threats.

In this context, graph neural networks (GNNs) have emerged as a transformative technology for social service networks due to their exceptional ability to perceive interactive information, as demonstrated in various social service scenarios, such as fraud detection [1]. GNNs are particularly well-suited for identifying risky and fraudulent behaviors that may be hidden within dense, high-dimensional interactive information. By integrating both interaction data and user-specific attributes, GNNs can detect suspicious activities with high accuracy, significantly enhancing the security of digital service platforms and establishing a more trustworthy online environment.

However, GNN-based fraud detection faces two main challenges: (1) **Data imbalance**. In real-world service ecosystems, fraudulent entities (such as fake accounts, malicious transactions, or service abuse) are often a minority within the network. The dominance of legitimate service nodes and regular interactions makes it difficult for detection models to capture the subtle anomalies associated with fraudulent behavior. This imbalance reduces the model's sensitivity to minority-class samples and weakens its ability to differentiate between normal service patterns and sophisticated fraud tactics, ultimately lowering both detection accuracy and generalization performance.

(2) **Heterophily**. Traditional GNNs, designed around homophily (the assumption that connected nodes exhibit similar features and behaviors), are poorly suited for service fraud detection. A significant limitation of these models is the over-smoothing effect, which is especially problematic in service networks. These models assume that interconnected nodes in a network share similar features and behaviors, thereby diminishing the ability to distinguish between linked entities. Fraudsters exploit this design flaw by creating cross-service-cluster relationships, such as generating high-frequency interactions or embedding themselves within legitimate transaction pathways to hide their fraudulent actions. Through these heterophilic strategies, fraudulent nodes can contaminate their local neighborhoods, obscuring their anomalous behavior and evading detection by GNNs. As a result, GNNs fail to identify the abnormal patterns, operational irregularities, and behavioral deviations that distinguish malicious users from legitimate participants in service networks.

To address these challenges, existing methods primarily focus on spatial domain analysis, which includes strategies like attention mechanisms [2] and auxiliary loss functions [3]. For example, attention mechanisms can dynamically allocate the weights to the neighbors and manage to boost the contributions of nodes with high affinity; resampling techniques can adaptively determine which neighboring nodes to retain through feedback. However, these methods often face high computational costs and may alter the underlying structure of the service network. Recently, spectral domain analysis has been explored as a promising alternative [4]. By filtering high- and low-

frequency signals in the service network, spectral GNNs are better equipped to capture the distinct characteristics of anomalies, offering improved efficiency and accuracy over spatial approaches.

Despite these advancements, spectral GNN-based fraud detection still have poor ability to capture clean and discriminative latent representations, which can be attributed to the following limitations: (1) Although graph filters can capture signals in different frequency domains, these filters still assume that information interaction between nodes in the network is effective behavior, ignoring noise variables introduced by malicious propagation and irrelevant behavior patterns. (2) A prevalent solution to heterophily is to leverage different graph filters to capture the signals at different frequencies. However, these signals from different graph filters lack the frequency-specific semantic discrimination, which makes the model hard to explicitly identify signal characteristics with different frequency domains, resulting in the distortion of the fused signals at the fusion node.

To address these issues, we propose a novel spectral graph network based on information bottleneck theory (SGNN-IB) for fraud detection. SGNN-IB first splits the original graph into homophilic and heterophilic subgraphs using a heterophily-aware classifier. It then applies multi-scale graph filters to capture both low- and high-frequency signals from the subgraphs and the original graph. For the first limitation, SGNN-IB incorporates information bottleneck theory [5] to enhance the encoding quality of graph filters with different frequencies, alleviating the noise interference in the encoded node embeddings. For the second limitation, SGNN-IB employs prototype learning to boost the semantic discrepancy between high- and low-frequency signals, thereby helping the model to identify diverse graph signals and fuse frequency-specific graph signals.

In summary, our contributions are as follows:

- We present a novel SGNN-IB model to derive clean and discriminative characteristics for fraud detection, which employs an edge classifier to split the original graph into homophilic and heterophilic subgraphs and then leverages Beta wavelet graph filters to capture critical characteristics of fraudsters.
- We introduce an IB-based loss function to decrease the noise in different signals and utilize prototype learning to capture the frequency-specific characteristics and improve the signals' integration.
- Extensive experiments on widely used datasets demonstrate that our method significantly outperforms baseline approaches. Additionally, our ablation study validates the effectiveness of each component in the SGNN-IB framework.

2 Related work

Graph-based methods for fraud detection in service networks leverage the inherent topological structure of service interactions to facilitate information propagation across individuals. A major challenge in fraud detection is data imbalance, as fraudsters often blend in with legitimate users, making their presence hard to detect. GNN-based fraud detection methods typically use various strategies to mitigate the impact of data

imbalance and improve detection accuracy. For instance, ASA-GNN [6] adopts adaptive sampling strategies to filter out noisy nodes and propagate more representative information. Although these methods can effectively mitigate the issue of outliers in service networks, the sampling strategies may disrupt the inherent structure of service interactions, leading to the loss of important information.

Another challenge is that fraudsters often hide by frequently interacting with benign users, leading to heterophily, where connected nodes exhibit different patterns. To tackle this, GAGA [7] introduces a group-based strategy to mitigate the impact of high heterophily. Although these methods are effective, they suffer from significant computational complexity.

Recent studies have used graph filters to capture both low- and high-frequency signals. For example, IDGL [8] applies dual-channel graph convolution filters to propagate multi-scale frequency information. Additionally, some research addresses the ``right-shift" phenomenon caused by heterophily, using Beta wavelet transformations as spectral filters to capture important information [4].

Many filter-based methods rely on sophisticated graph filters to update node features, achieving success in identifying fraudsters in service networks. These methods often use classical graph filters, such as polynomial and wavelet transformations, to capture both low- and high-frequency information. Given the complexity of graph structures, some approaches apply filters at different levels or perspectives, such as global vs. local views, homophilic vs. heterophilic views, and relation-based views, to enhance model representation. Despite these advances, such methods are still limited in obtaining representative characteristics of nodes and are vulnerable to noise interference across different frequency domains.

3 Preliminaries

3.1 Definitions

Definition 1 *(Graph)*: Let a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X}, \mathcal{A}, \mathcal{Y})$ denotes a service network. $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ represents node set of graph $\mathcal{G}$, where N is the number of nodes. $\mathcal{E}$ is the edge set of graph $\mathcal{G}$ and $e_{uv} \in \mathcal{E}$ denotes an edge from node u to node v. $\mathcal{X} \in \mathbb{R}^{N \times D}$ indicates the feature matrix of N nodes, where D is the feature dimension. $\mathcal{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix of $\mathcal{G}$. If $e_{uv} \in \mathcal{E}$, $a_{uv} \in \mathcal{A} = 1$, otherwise $a_{uv} = 0$. $\mathcal{Y} \in \mathbb{R}^{N \times 1}$ denotes the label of all nodes, where $y_v \in \mathcal{Y} = 0$ if node v is a benign sample and $y_v \in \mathcal{Y} = 1$ if node v is a fraudster.

Definition 2 *(Multi-relation graph)*: If there are different relations between nodes in the graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E}_r|_{r=1}^{R}, \mathcal{X}, \mathcal{A}_r|_{r=1}^{R}, \mathcal{Y})$ can be denoted as a multi-relation graph, where R is the number of relation categories. For simplicity, a multi-relation graph can be identified as $\mathcal{G} = (\mathcal{X}, \mathcal{A}_r|_{r=1}^{R}, \mathcal{Y})$.

4 Methodology

4.1 The SGNN-IB framework

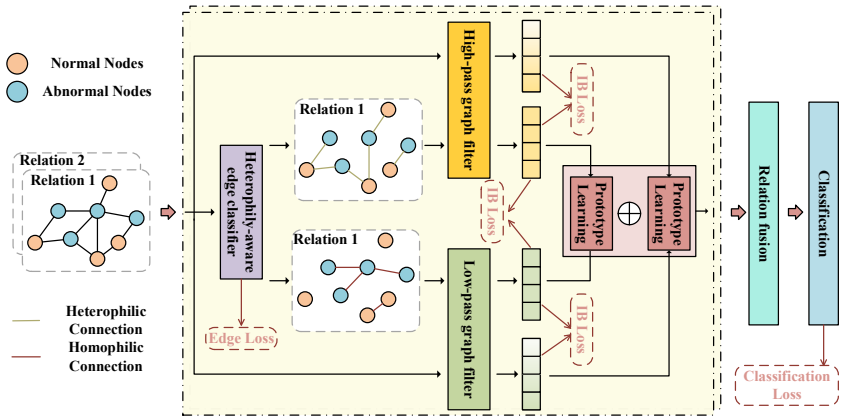

Fig. 1. The framework of SGNN-IB. First, SGNN-IB leverages an edge classifier to perceive heterophilic subgraphs. Then, SGNN-IB utilizes multi-scale graph filters to obtain the high- and low-frequency signals in the graph. Subsequently, SGNN-IB integrates the signals from different frequencies based on prototype learning. Finally, SGNN-IB is trained by the joint loss function, integrated with IB-loss.

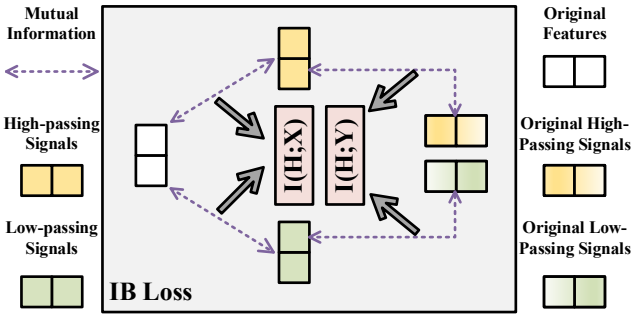

Fig. 2. The architecture of IB loss. To solve the noise issue, the model leverages classical IB theory, maximizing the mutual information between the latent features and the ground truths and minimizing the mutual information between the latent features and the original features. Here, latent features denote the high-pass and low-pass signals, and ground truths represent the band-pass signals. To solve the stochasticity issue, the model introduces the mutual information between high-pass and low-pass signals.

The framework of SGNN-IB is shown in Fig. 1. First, SGNN-IB employs an edge classifier to identify and extract heterophilic subgraphs within the graph structure. Subsequently, SGNN-IB applies diverse graph filters to encode the original graph and specific subgraphs. the graph signals into high-pass, low-pass, and band-pass components, capturing diverse frequency-specific information. To enhance frequency-specific semantic discrimination, SGNN-IB introduces prototype learning to obtain the affinity of

signals and performs information fusion. To conquer the noise problem, SGNN-IB introduces an IB loss to alleviate the interference of noise in the process of information propagation. Finally, SGNN-IB is trained with an objective function that comprises IB loss, classification loss, and edge loss, ensuring a balanced and comprehensive learning process. The architecture of IB loss is shown in Fig. 2.

4.2 Heterophily-aware edge classifier

Traditional GNNs are established on the assumption that the connections between nodes exhibit homophily, which means the connected nodes belong to the same category. In other words, traditional GNNs serve as a smoothing function for the graph signals. However, many connections show heterophily, indicating that the connected nodes have different labels. Simply deploying traditional GNNs may dilute the categorical characteristics of nodes, which hinders accurate node identification. Therefore, to avoid the loss of discriminative information in the graph, it is important to split homophilic and heterophilic connections.

To perceive the heterophily in graph topology, we design a heterophily-aware edge classifier, which aims to identify the edge type of each edge. In the context of training data containing labeled nodes, we meticulously establish homophilic and heterophilic edges based on the labels of source and target nodes in the training set. The edge classifier, designed as a binary classification model, leverages the feature representations of both the source node u and the target node v to predict the type of edge e_{uv}. This classifier is implemented using a multi-layer perceptron (MLP) architecture, thereby facilitating the discrimination between different edge types within the graph.

For an edge e_{uv} with the source node u and target node v, the computations are as follows:

$$\boldsymbol{h}_u = \sigma(\boldsymbol{W}_h \cdot \boldsymbol{x}_u + \boldsymbol{b}_h), \boldsymbol{h}_v = \sigma(\boldsymbol{W}_h \cdot \boldsymbol{x}_v + \boldsymbol{b}_h) \tag{1}$$

$$\phi_{uv} = Sigmoid(\boldsymbol{W}[\boldsymbol{h}_u || \boldsymbol{h}_v || (\boldsymbol{h}_u - \boldsymbol{h}_v)]), \pi_{uv} = 2 * \phi_{uv} - 1 \tag{2}$$

where $\sigma(\cdot)$ is a nonlinear activation function, $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ are respectively the original features of node u and v, $\boldsymbol{W}_h$, $\boldsymbol{b}_h$ and $\boldsymbol{W}$ are learnable parameters of the feature transformation $\boldsymbol{h}_u$ and $\boldsymbol{h}_v$ are respectively transformed features of node u and v, $Sigmoid(\cdot)$ is Sigmoid activation function, $[\cdot \,||\, \cdot]$ is concatenation function. π_{uv} is limited to [-1, 1] to discriminate the heterophilic connections.

To partition the original graph into a homophilic subgraph $\mathcal{G}_{homo}$ and a heterophilic subgraph $\mathcal{G}_{heter}$, we leverage the prediction outcomes of all edges within the graph. The homophilic subgraph exclusively comprises edges predicted to exhibit homophily, whereas the heterophilic subgraph merely encompasses edges anticipated to display heterophily.

The precise classification of edges is of paramount importance for subsequent procedures, as it directly influences the quality of the resultant partitioned subgraphs. To this end, we devise an auxiliary loss function tailored for training the edge classifier. This loss is derived from the constructed training edge set $\mathcal{E}_{tr}$ and the corresponding

prediction outcomes. The heterophily-aware edge classifier is optimized using the training edge set $\mathcal{E}_{tr}$ with the following loss function:

$$\mathcal{L}_H = -\sum_{e_{uv}\in\mathcal{E}_{tr}}\left[y_{e_{uv}}\log(\phi_{uv}) + \left(1-y_{e_{uv}}\right)\log(1-\phi_{uv})\right] \quad (3)$$

where $y_{e_{uv}}$ is the label of edge e_{uv}. If the edge exhibits homophily, the label $y_{e_{uv}}$ is 1, otherwise the label $y_{e_{uv}}$ is 0.

4.3 The design of graph filter and information fusion

Upon dividing the original graph, the resultant homophilic subgraph $\mathcal{G}_{homo}$ manifests an enrichment of low-frequency signals, whereas the heterophilic subgraph $\mathcal{G}_{heter}$ predominantly exhibits high-frequency signals. To capture signals within distinct frequency bands, diverse filters are applied to these partitioned graphs. Notably, since the splitting process yields frequency-specific signals from the original graph, subgraphs inevitably lose the holistic structural information contained within the original graph. To bolster the overall semantic richness and the fidelity to original information, it is imperative to also apply filters to the original graph.

Formally, consider the original graph $\mathcal{G}$, alongside the predicted homophilic subgraph $\mathcal{G}_{homo}$ characterized by its Laplacian $\boldsymbol{L}_{homo}$, and the predicted heterophilic subgraph $\mathcal{G}_{heter}$ with $\boldsymbol{L}_{heter}$. Given the model's need to discern signals of varying frequencies across these three graphs, a versatile band-pass filter becomes indispensable. Crafting an apt graph filter for the partitioned subgraphs presents a non-negligible challenge, as contemporary GNNs predominantly leverage low-pass filters [4]. Recently, research endeavors have introduced methodologies to learn arbitrary graph filters via polynomial approximation or Transformer architectures, exemplified by PolyFormer [9]. However, these methodologies fall short in the context of fraud detection tasks, where the minute proportion of fraudulent nodes within the graph exacerbates the issue of severe class imbalance. Consequently, high-frequency signals become relatively scant, leading the trained filter to potentially demonstrate a propensity for prioritizing low-frequency signals.

Consequently, we adopt design band-pass filters based on Beta wavelet [10] to capture distinct frequency bands. Based on the Beta distribution, Beta wavelet transformation is defined as follows:

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, x\in[0,1] \quad (4)$$

where $\Gamma(\cdot)$ is Gamma distribution, α and β are the parameters of Beta distribution. Given the eigenvalues of the normalized graph Laplacian $\boldsymbol{L}\in[0,2]$, we leverage $f^*(x;\alpha,\beta) = \frac{1}{2}f\left(\frac{1}{2}x;\alpha,\beta\right)$ to cover the whole spectral range of $\boldsymbol{L}$.

For simplicity, we constrain the $\alpha,\beta\in\mathbb{N}^+$ and only generate a low-pass filter $f^*_{low}(x;\alpha,\beta)$ and a high-pass filter $f^*_{high}(x;\alpha,\beta)$ to avoid computational complexity problems.

Then we apply the high f_{low}^{*} to $\boldsymbol{L}_{homo}$ to capture low-frequency information from $\mathcal{G}_{homo}$. Correspondingly, we can obtain high-frequency signals by deploying f_{high}^{*} on the normalized Laplacian $\boldsymbol{L}_{heter}$ of $\mathcal{G}_{heter}$. The formulations can be defined as follows:

$$\boldsymbol{H}_i = f_i^{*}(\boldsymbol{L}_i, \boldsymbol{H}) = f_i\left(\frac{1}{2}\boldsymbol{L}_i; \alpha, \beta\right)\boldsymbol{H}, i \in \{low, high\} \tag{5}$$

where $\boldsymbol{H}$ is the features matrix. Then, we integrate the obtained signals from different frequency domains:

$$\widehat{\boldsymbol{H}} = \Phi\left(\boldsymbol{H}_{low}, \boldsymbol{H}_{high}\right) \tag{6}$$

where $\Phi(\cdot,\cdot)$ is an adaptive frequency fusion function, which is illustrated in Section 4.4.

We have derived representations from both homophilic and heterophilic subgraphs utilizing low- and high-pass filters. Nevertheless, the structural integrity of these two subgraphs remains incomplete. To enhance the expressive power of our model, we employ the band-pass filters on the original graph and generate fused embeddings of band-pass filters:

$$\boldsymbol{H}_i^{o} = f_i^{*}(\boldsymbol{L}_0, \boldsymbol{H}) = f_i\left(\frac{1}{2}\boldsymbol{L}_o; \alpha, \beta\right)\boldsymbol{H} \tag{7}$$

$$\widehat{\boldsymbol{H}}^{o} = \Phi\left(\boldsymbol{H}_{low}^{o}, \boldsymbol{H}_{high}^{o}\right) \tag{8}$$

where $\boldsymbol{H}_i^{o}$ represents the transformed features by a single low- or high-pass filter. To protect the original semantic information of node features, the ultimate embedding of the node is constructed by concatenating the filtered representations and the linearly transformed residual representations from the original graph:

$$\overline{\boldsymbol{H}} = \sigma\left(\boldsymbol{W}_o\left[\widehat{\boldsymbol{H}}^{o}, \widehat{\boldsymbol{H}}\right]\right) \tag{9}$$

where $\boldsymbol{W}_o$ is learnable parameters.

In practical scenarios, the majority of fraud graphs encompass diverse relationships. After acquiring representations for each relation, we integrate the node representations stemming from these various relations, thereby constructing the definitive embedding for the nodes. For the sake of brevity, we have omitted the explicit representation of these relations in the aforementioned equations. The relation fusion formulation can be defined as follows:

$$\boldsymbol{H}_{all} = \boldsymbol{W}_r||_{r=1}^{R}\overline{\boldsymbol{H}}_r \tag{10}$$

where $\overline{\boldsymbol{H}}_r$ is the ultimate filtered embedding in homogeneous graph under relation r, R is the relation set of graph $\mathcal{G}$ and $\boldsymbol{W}_r$ is the learnable weights.

4.4 Frequency-specific feature fusion based on prototype learning

The high-frequency and low-frequency should reflect the behavior characteristics of nodes in different frequency domain modes. However, due to the interactive pattern of

nodes, these signals lack discernible frequency-specific semantic information, which loses significant discrimination after feature fusion. Therefore, we introduce an adaptive frequency fusion function $\Phi(\cdot,\cdot)$, a prototype learning mechanism, to enhance the semantic representations in each frequency domain.

Take high-frequency features as an example. Given the latent representations of frequency domain $\boldsymbol{H}_{high}$, we first calculate the prototype of high-frequency domain:

$$\boldsymbol{c}_{high} = Readout(\boldsymbol{H}_{high}) \tag{11}$$

where $Readout(\cdot)$ is average readout function. Then we can obtain the affinity score of the node features with prototype:

$$s_{high} = \frac{1}{\text{len}(\boldsymbol{H}_{high})} \sum_{i=1}^{\text{len}(\boldsymbol{H}_{high})} \cos(\boldsymbol{h}_{(i,high)}, \boldsymbol{c}_{high}) \tag{12}$$

where $\boldsymbol{h}_{(i,high)} = \boldsymbol{H}_{high}[i,:]$, $\cos(\cdot,\cdot)$ denotes the cosine distance, and $\text{len}(\cdot)$ denotes the sample size in $\boldsymbol{H}_{high}$. Similarly, we can obtain the affinity score s_{low} in low-frequency domain. A higher score indicates that the frequency-specific characteristics are more representative.

To enhance the frequency-specific semantic discrimination, the fused representations should approach to frequency domain signals with high affinity. Therefore, we integrate the signals from high-frequency and low-frequency domain based on the affinity score:

$$\Phi(\boldsymbol{H}_{high}, \boldsymbol{H}_{low}) = \frac{s_{high}}{s_{high}+s_{low}} \boldsymbol{H}_{high} + \frac{s_{low}}{s_{high}+s_{low}} \boldsymbol{H}_{low} \tag{13}$$

To capture signals within distinct frequency bands, diverse filters are applied to these partitioned graphs. Notably, since the splitting process yields frequency-specific signals from the original graph and there is interference in the propagation of information in interactive behavior, subgraphs inevitably lose the holistic structural information contained within the original graph. To bolster the overall semantic richness and the fidelity to original information, it is imperative to also apply filters to the original graph.

4.5 IB-based representation denoising

Even though high-pass and low-pass filters encapsulate distinct semantic information within graphical representations, as illustrated in Section 1, there is interference in the propagation of information in interactive behavior, which results in noise problems in the propagation of information. These issues leave the graph filtering capability constrained and hindering the generation of sufficiently discriminative representations across diverse frequency domains.

To this end, we introduce the IB theory to improve the quality of latent representations against noise. According to IB theory, the training objective is twofold: (1) to maximize mutual information between encoded embeddings $\boldsymbol{H}$ and labels $\boldsymbol{Y}$, and (2) to

minimize mutual information between the encoded embeddings $\boldsymbol{H}$ and the node features $\boldsymbol{X}$.

$$arg\max_{\boldsymbol{H}} -I(\boldsymbol{H};\boldsymbol{Y}) + \mu \cdot I(\boldsymbol{H};\boldsymbol{X}) \tag{14}$$

where μ is a balanced coefficient. The IB Theory can compress the information within input data to distill and preserve the most task-relevant knowledge, effectively reducing noise and redundant information while extracting the most predictive and useful features.

Based on this idea, we develop an IB-based information-enhancing module to improve the quality of graph filters and provide more optimization guidance for signals in different frequencies. First, our basic objective function is consistent with IB theory: (1) to maximize the mutual information between the latent embeddings $\boldsymbol{H}$ and the labels $\boldsymbol{Y}$, and (2) to minimize the mutual information between the latent embeddings and input features $\boldsymbol{X}$. However, due to the lack of prior knowledge of different frequency signals, it is impractical to calculate the mutual information directly using ground truth labels. To this end, we regard the latent embeddings from the encoded original graph using different graph filters as labels $\boldsymbol{Y}$, and the representations encoded from the heterophilic and homophilic using corresponding graph filters as the latent embeddings $\boldsymbol{H}$. Then, the IB-based loss function can be defined as follows:

$$I(\boldsymbol{H};\boldsymbol{Y}) = I\left(\boldsymbol{H}_{high};\boldsymbol{H}^{o}_{high}\right) + I(\boldsymbol{H}_{low};\boldsymbol{H}^{o}_{low}) \tag{15}$$

$$I(\boldsymbol{H};\boldsymbol{X}) = I\left(\boldsymbol{H}_{high};\boldsymbol{H}\right) + I(\boldsymbol{H}_{low};\boldsymbol{H}) \tag{16}$$

The overall IB-based loss function is defined by averaging each term of mutual information:

$$\mathcal{L}_{IB} = \frac{1}{2} \times [-I(\boldsymbol{H};\boldsymbol{Y}) + \mu \cdot I(\boldsymbol{H};\boldsymbol{X})] \tag{17}$$

Through the implementation of the information-enhancing module based on IB theory, the graph filters obtain explicit guidance to effectively counteract noise within features. This ensures that the encoded representations not only preserve the vital characteristics of the original features but also meticulously filter out redundant and irrelevant information. Additionally, the graph filters operate across different frequency channels, maintaining their specificity and ensuring that each channel remains relatively independent. This approach enables the generation of high-quality, fused features that are crucial for the accuracy of the model.

5 Experiments

5.1 Experimental setup

Datasets. We execute experiments on three public fraud detection datasets, YelpChi [11], Amazon [3], and FDCompCN [4]. In the YelpChi dataset, nodes represent reviews, and it includes three types of relations: 1) R-U-R represents the reviews posted by the

same user, 2) R-S-R denotes reviews related to the same product with the same star rating, and 3) R-T-R stands for the reviews related to the same product posted in the same period. In the Amazon dataset, nodes represent users, with three types of relations: 1) U-P-U denotes users reviewing at least one same product, 2) U-S-U represents users having at least one same star rating within a specific period, and 3) U-V-U indicates subscribers with the top 5 percent mutual review text similarities. In the FDCompCN dataset, nodes represent companies, and it includes three types of relations: 1) C-I-C represents companies that have investment relationships, 2) C-P-C indicates companies and their disclosed customers, and 3) C-S-C suggests companies and their disclosed suppliers. The dataset statistics are summarized in Table 1.

Table 1. Statistics of datasets

Dataset	Application	Node	Dimension	Fraud (%)	Relation	Edge
Yelp	review	45954	32	14.53%	R-U-R	98630
					R-T-R	1147232
					R-S-R	7693958
Amazon	review	11944	24	6.87%	U-P-U	351216
					U-S-U	7132958
					U-V-U	2073474
FDCompCN	financial	5317	57	10.50%	C-I-C	5686
					C-P-C	760
					C-S-C	1043

Baselines. We select ten baselines to validate the advancement of our model. We categorize the baselines into three groups: shallow methods, GNNs, and GNN-based fraud detection frameworks. Among these, MLP and XGBoost are typically shallow methods based on feature learning, which ignore graph topology. GCN [12], GAT [13], FAGCN [14], and GPR-GNN [15] are GNN-based methods. CARE [11], Fdetector [3], BWGNN [10], and SEFraud [1] are fraud detection frameworks based on GNNs.

Evaluation settings. Since the fraud detection problem exhibits data imbalance, we select four metrics to evaluate all models, including AUC, Recall (R), GMean (G), and F1-score (F).

Implementation details. The experiments utilize PyTorch in Python 3.9.12, deploying a single NVIDIA A40 GPU, 40GB of RAM, and a 2.60GHz Xeon (R) Gold 6240 CPU. All the baselines can be reproduced by public source codes and Python dependencies.

5.2 Overall performance

The experimental results of our study are summarized in Table 2. The best results are highlighted in bold, while the second-best results are underlined.

For shallow methods, we find that MLP performs better than XGBoost. This is because MLPs can adaptively learn and represent nonlinear relationships in the data. Among GNN-based models, GPRGNN and FAGCN demonstrate better performance

compared to traditional GCN and GAT models. GPRGNN excels by capturing both structural and feature information. FAGCN is particularly effective at identifying fraud-related features. Graph-based approaches outperform feature-based shallow methods, as they are better equipped to capture the complex relational and interactive information embedded within the graph structure. This advantage stems from the ability of GNNs to handle the data imbalance commonly seen in fraud detection tasks

Table 2. Performance of the proposed SGNN-IB model and comparative model on three datasets. All results are in %.

Dataset	Yelp				Amazon				FDCompCN			
Metric	R	F	AUC	G	R	F	AUC	G	R	F	AUC	G
XGBoost	19.15	61.72	59.01	43.51	69.09	72.68	79.54	78.87	61.25	61.17	50.64	58.04
MLP	69.37	61.48	77.43	70.73	78.18	72.95	87.78	82.93	57.08	54.80	43.06	58.48
GCN	77.53	36.67	59.33	49.46	80.00	56.43	84.61	73.72	52.92	51.01	40.89	43.95
GAT	62.15	42.77	56.13	53.13	80.00	71.46	88.03	83.04	52.55	51.36	38.20	42.94
GPRGNN	75.16	57.34	77.12	69.84	80.09	64.15	89.08	82.32	56.40	47.52	50.31	52.09
FAGCN	70.64	61.11	77.90	70.88	81.21	69.30	90.48	84.33	57.90	48.48	51.59	49.50
CARE	72.32	60.40	77.41	70.86	75.76	70.45	86.19	81.71	57.21	43.59	49.00	50.10
FDetector	84.61	70.78	88.90	81.64	82.12	71.36	89.84	84.29	55.96	48.33	47.89	49.10
BWGNN	82.56	72.32	89.72	81.92	83.94	69.43	91.91	84.67	58.01	47.91	49.79	52.33
SEFraud	78.64	72.51	86.77	82.44	88.67	71.28	91.50	85.13	57.49	50.31	50.41	53.74
SGNN-IB	**86.37**	**74.64**	**92.06**	**84.40**	**90.30**	**71.56**	**93.03**	**86.65**	**58.93**	**52.22**	**56.43**	**54.17**

The proposed SGNN-IB outperforms all these baseline models. In comparison to the best performance in baselines, SGNN-IB shows an absolute improvement of 1.76%, 2.13%, 2.34%, and 1.96% in Recall, F1-Macro, AUC, and GMean on the YelpChi dataset. On the Amazon dataset, SGNN-IB achieves absolute improvements of 1.63%, 0.20%, 1.12%, and 1.52%, respectively. For the FDCompCN dataset, SGNN-IB improves by 0.92%,1.91%, 6.02%, and 0.43% in Recall, F1-Macro, AUC, and GMean.

The success of SGNN-IB can be attributed to several key factors. First, SGNN-IB uses both low-pass and high-pass filters to selectively extract relevant information from homogeneous and heterogeneous structures, respectively. It also employs a prototype learning method to maintain the discriminative information of different frequency domain. In addition, to enhance the robustness of the filtering process against noise, SGNN-IB integrates an IB-based enhancement module. This module guides the graph filter, enabling it to generate high-quality, encoded features that improve fraud detection performance.

5.3 Ablation experiments

To evaluate the contribution of each component in the SGNN-IB framework, we conduct ablation studies by examining five variants. -edge denotes SGNN-IB without heterophily-aware edge classifier, -low denotes SGNN-IB without low-pass filter, -high represents SGNN-IB without high-pass filter, -rel denotes SGNN-IB without relation

fusion, and -IB represents SGNN-IB without IB-based information enhancer. The results of these ablation experiments are presented in Table 3, with the best results highlighted in bold and the second-best results underlined.

Table 3. Performance of the ablation experiments on three datasets. All results are in %.

Dataset	Yelp				Amazon				FDCompCN			
Metric	R	F	AUC	G	R	F	AUC	G	R	F	AUC	G
-edge	84.32	68.28	85.62	74.36	83.21	66.34	84.51	81.51	51.21	46.32	50.11	48.29
-low	85.31	67.48	85.43	76.73	82.37	64.74	85.97	76.94	52.14	45.80	49.39	48.22
-high	83.53	66.67	83.33	79.46	80.44	63.48	82.11	79.55	51.39	43.82	50.73	46.77
-rel	82.64	66.15	77.90	74.86	83.64	65.54	88.54	80.28	55.83	49.33	50.81	48.96
-IB	81.53	67.57	89.13	79.26	88.99	69.45	90.42	84.68	53.93	50.48	53.61	51.28
SGNN-IB	**86.37**	**74.64**	**92.06**	**84.40**	**90.30**	**71.56**	**93.03**	**86.65**	**58.93**	**52.22**	**56.43**	**54.17**

The results indicate that SGNN-IB outperforms all its variants, demonstrating the effectiveness of each component in the framework. Meanwhile, -low performs relatively close to SGNN-IB, while -high shows lower performance. This suggests that high-pass signals play a particularly important role in detecting fraudulent activities. Additionally, the performance of -IB reinforces the effectiveness of the IB-based information enhancement module, which contributes to noise reduction and improved model robustness.

5.4 Sensitivity experiments

We conduct sensitivity experiments by selecting three key model hyperparameters: μ, λ, and η. The parameter μ controls the contribution of mutual information between the input features and the filtered features, as well as between different filter channels. The parameter λ controls the influence of the heterophily-aware edge classifier, while η controls the contribution of the information enhancement loss based on the information bottleneck (IB) theory. The values of λ and η range from 0.1 to 1.5, with a step size of 0.1. The range for μ is from 0.000001 to 0.1, with an exponential step size. The results of these sensitivity experiments for the YelpChi and Amazon datasets are shown in Fig. 3 and 4.

From the sensitivity experiments across these three datasets, we observe that the parameters λ and μ have a significant impact on model performance, while η plays a relatively minor role. Specifically, take the YelpChi dataset as an example. As shown in Fig. 3(a), a small value of λ limits the effectiveness of the edge classifier, leading to incorrect identification of heterophilic edges. This misclassification hampers the capture of high-frequency signals, which are crucial for identifying fraudulent behavior, thus reducing the model's ability to detect fraudsters. On the other hand, increasing λ enhances the classifier's capacity, but its effect on performance is relatively small beyond a certain threshold. Fig. 3(b) and 3(c) further show that η and μ mainly affect the model's ability to filter noise and extract key features. However, when these values are too large, the loss function tends to converge rapidly to negative values during training,

resulting in a slight decline in performance. In particular, for μ, which regulates the data purification and compression between the input data and the filtered features, smaller exponential values are more effective. This allows SGNN-IB to focus on the most essential components of the original features, improving its ability to capture the key information related to fraud.

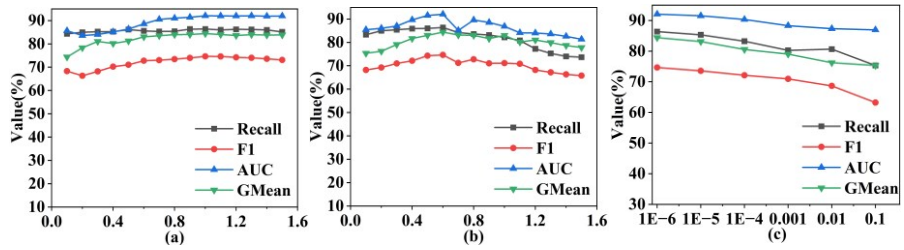

Fig. 3. Sensitivity experimental results on YelpChi dataset: (a) Sensitivity results for parameter λ; (b) Sensitivity results for parameter η; (c) Sensitivity results for parameter μ.

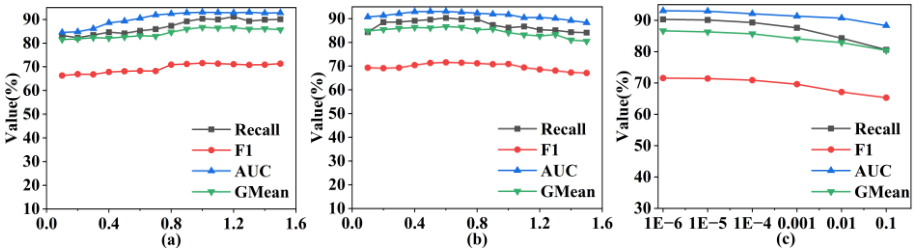

Fig. 4. Sensitivity experimental results on the Amazon dataset: (a) Sensitivity results for parameter λ; (b) Sensitivity results for parameter η; (c) Sensitivity results for parameter μ.

6 Conclusion

In this paper, we propose a novel spectral graph network based on information bottleneck (SGNN-IB) for fraud detection in service networks. SGNN-IB innovatively utilizes an edge classifier to dissect the original service network into heterophilic and homophilic sub-networks. It then applies band-pass graph filters to effectively extract high- and low-frequency service patterns from each subgraph. The framework integrates these signals from multiple relational dimensions to enhance the representation of fraudulent behavior. To improve the robustness and filtering capabilities of the spectral graph network, we introduce an information bottleneck-based learning module. To evaluate the effectiveness and improvements of SGNN-IB, we conduct comprehensive experiments on three publicly available datasets. The results show that our model outperforms existing state-of-the-art methods in terms of detection accuracy. Future research will focus on developing more efficient and scalable methods for large-scale fraud detection. Additionally, exploring the potential role of multi-modal information in fraud detection presents an exciting avenue for future work.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant 72210107001, the Beijing Natural Science Foundation under Grant IS23128, the Fundamental Research Funds for the Central

Disclosure of Interests. The authors have no competing interests.

References

1. Li K, Yang T, Zhou M, et al.: SEFraud: Graph-based Self-Explainable Fraud Detection via Interpretative Mask Learning. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 5329-5338 (2024).
2. Liu C, Sun L, Ao X, et al.: Intention-aware heterogeneous graph attention networks for fraud transactions detection. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining, 3280-3288 (2021).
3. Shi F, Cao Y, Shang Y, et al.: H2-fdetector: A gnn-based fraud detector with homophilic and heterophilic connections. In: Proceedings of the ACM web conference, 1486-1494 (2022).
4. Wu B, Yao X, Zhang B, et al.: Splitgnn: Spectral graph neural network for fraud detection against heterophily. In: Proceedings of the 32nd ACM international conference on information and knowledge management, 2737-2746 (2023).
5. Yan X, Mao Y, Ye Y, et al.: Cross-modal clustering with deep correlated information bottleneck method. IEEE Transactions on Neural Networks and Learning Systems (2023).
6. Tian Y, Liu G, Wang J, et al.: ASA-GNN: Adaptive sampling and aggregation-based graph neural network for transaction fraud detection. IEEE Transactions on Computational Social Systems , 11(3): 3536-3549 (2023).
7. Wang Y, Zhang J, Huang Z, et al.: Label information enhanced fraud detection against low homophily in graphs. In: Proceedings of the ACM Web Conference, 406-416 (2023).
8. Wu J, Hu R, Li D, et al.: A GNN-based fraud detector with dual resistance to graph disassortativity and imbalance. Information Sciences, 669: 120580 (2024).
9. Ma J, He M, Wei Z.: Polyformer: Scalable node-wise filters via polynomial graph transformer. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2118-2129 (2024).
10. Tang J, Li J, Gao Z, et al.: Rethinking graph neural networks for anomaly detection. In: International conference on machine learning, 21076-21089 (2022).
11. Dou Y, Liu Z, Sun L, et al.: Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In: Proceedings of the 29th ACM international conference on information & knowledge management, 315-324 (2020).
12. Kipf T N, Welling M.: Semi-Supervised Classification with Graph Convolutional Networks. In: International Conference on Learning Representations (2017).
13. Veličković P, Cucurull G, Casanova A, et al.: Graph Attention Networks. In: International Conference on Learning Representations (2018).
14. Bo D, Wang X, Shi C, et al.: Beyond low-frequency information in graph convolutional networks. In: Proceedings of the AAAI conference on artificial intelligence, 35(5): 3950-3957 (2021).
15. Chien E, Peng J, Li P, et al.: Adaptive Universal Generalized PageRank Graph Neural Network. In: International Conference on Learning Representations (2022).