Journal of Network Intelligence ©2023 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **8**, Number **3**, August **2023**

Multibehavioral Pattern-based Graph Neural Network with Topology Awareness to Detect Healthcare Fraud

Jiang-Tao Lu

Department of Computer Science and Technology
Xiamen University of Technology
Xiamen, 361024, China
2022031417@s.xmut.edu.cn

Kai-Biao Lin*

Department of Computer Science and Technology
Xiamen University of Technology
Xiamen, 361024, China
2010110706@t.xmut.edu.cn

Xin Chen

Department of Computer Science and Technology
Xiamen University of Technology
Xiamen, 361024, China
2022031466@s.xmut.edu.cn

Jin-Po Chen

Department of Computer Science and Technology
Xiamen University of Technology
Xiamen, 361024, China
2022032015@s.xmut.edu.cn

*Corresponding author: Kai-Biao Lin

Received November 11, 2022, revised January 3, 2023, accepted March 23, 2023.

Abstract. *Healthcare fraud is a major problem that affects the quality of healthcare services. Traditional machine learning methods to address healthcare fraud typically focus on only characteristic patient attributes and ignore the underlying behavioral patterns implicit in a patient's treatment history. To mine behavioral information implied by healthcare fraudsters, we propose a multibehavioral pattern-based graph neural network (MBGNN) to detect healthcare fraud. The proposed MBGNN first mines behavioral attributes in healthcare heterogeneous graphs by multibehavior pattern decomposition. Here, three sampling strategies are employed to capture groups with the same behavioral patterns. Then, a two-step attention mechanism is employed to aggregate the compound semantic information of patients with different behavior patterns. Finally, the learned embeddings by combining semantic information with topological information are used to detect patients with abnormal behaviors. Extensive experiments were conducted on two real-world datasets, and the proposed MBGNN exhibited superior performance compared to state-of-the-art baseline methods. In the anomaly detection task, the proposed MBGNN scored high on F1 and precision metrics. We found that the proposed MBGNN outperformed existing methods in terms of accuracy. The multiple sampling strategies employed in the proposed method reveal the hidden behavior patterns of healthcare fraudsters. The results obtained by the proposed method are effective and explainable; thus, we believe the proposed can be used to detect healthcare fraud in practical applications.*
Keywords: Healthcare, Fraud detection, Graph neural network

1. Introduction.

Healthcare is an important part of the daily life, and due to the COVID-19 pandemic, many countries with universal coverage under a public healthcare system have suffered significant economic losses. In addition, the loss of healthcare due to fraud can increase the burden of an already fragile healthcare system. Healthcare fraud causes huge financial losses and puts patients at risk because it indirectly damages the healthcare system [1].

The National Health Care Anti-Fraud Association in Washington, D.C. states that "healthcare fraud refers to intentional deception or false statements by an individual or entity, knowing that false statements may lead to unauthorized benefit for the individual, entity, or other party" [2, 3]. According to data published by the World Health Organization, losses from healthcare fraud have been increasing steadily [4]. The Center for Medicare and Medicaid Services reported that national health expenditure reached 3.5 trillion USD in 2017, accounting for 17.9% of the US GDP [5], and the National Health Care Anti-Fraud Association estimates that approximately 3% of annual health care expenditures are the result of fraud [6]. To reduce the loss of health care funds, the Office of Inspector General created the healthcare fraud Strike Force [7] to detect and prosecute fraudulent physicians. As of 2017, the Strike Force has recovered $2.52 billion through 1,791 criminal prosecutions and 2,326 indictments. In China, the National Health Insurance Administration recovered $3.23 billion in health insurance funds in 2021 alone [8]. In 2022, 1 USD = 7.24 RMB [9]; therefore, guaranteeing health insurance is essential to protect public health and promote and protect economic development [10].

The high level of confidentiality in the healthcare industry and the uniqueness of each medical record make it difficult to detect healthcare fraud. Previous detection methods have relied on manual reviews. However, having experienced industry personnel review medical records manually is inefficient and labor intensive. Subsequent studies have employed a knowledge engineering approach to analyze the healthcare fraud problem and construct knowledge models. This method learns certain detection rules that can reduce labor and time costs significantly [11]. Inspired by data mining, some researchers have considered health insurance fraud to be a claim classification problem. Here, statistical methods are used to construct fraud identification models to detect fraud cases automatically [14]. This approach requires the selection of appropriate fraud assessment metrics; however, common healthcare fraud [12, 13] assessment metrics may pose personal privacy concerns and do not consider the potential patterns of behavior implicit in a patient's treatment history. Behavior is a critical aspect of healthcare fraud that is recorded as medical data. Many healthcare fraud cases tend to be gang fraud, where members of the same group behave fraudulently in a similar manner. For example, they will go to the same doctor at the same hospital to receive prescription medications. In such cases, the fraudster will typically have more medical data than the ordinary patient. Thus, the probability of finding the fraudster may be improved if such behavioral characteristics are considered.

Medical data are intricate [15], consisting of tables of personal information, medical information, prescription information, etc. Statistical-based methods make it difficult to mine behavioral characteristics from medical data, which is why few studies have investigated the behavioral characteristics of people who have committed fraud in the past. Emerging graph network technologies have achieved good results in various areas, e.g., citation network analysis [16, 17], social networks [18], and e-commerce [19]. Graph networks provide a new direction for mining behavioral attributes. Graph structures are also widely available in medical data. A medical dataset can be viewed as a heterogeneous graph comprising patients, hospital departments, medications, and visit dates. The connections between these entities constitute both the edges between the nodes in the heterogeneous graph and the behavioral trajectories of the patients. The patient's behavior patterns in medical data are transformed into topological information in the medical heterogeneous graph. The topological structure information of heterogeneous graphs can be mined easily using graph neural network (GNN)-based techniques. Then, the patient's characteristic attributes and graph-specific topological information are fused to detect anomalies.

However, there are some challenges to solve the healthcare fraud detection problem using GNN techniques. First, due to the complexity of healthcare data, there is a large number of noisy nodes in the network formed by healthcare data. Healthcare fraud detection can be considered a classification task, and a large number of noisy nodes can impact the performance of the learning graph network model, which increases the difficulty of the classification task. Second, distinguishing fraudulent and normal patients in the medical heterogeneous graph is a challenging process. Third, the detection results must be interpretable, and we must distinguish which behavioral features are more important to analyze healthcare fraud.

In this paper, we proposed the multibehavioral pattern-based GNN (MBGNN) to mine the behavioral information hidden by healthcare fraudsters. For the first two challenges, to realize fraud detection on heterogeneous graphs, where each patient node is marked as fraudulent or normal, the proposed MBGNN implements three unique strategies on sample patient nodes with the same behavioral trajectory in the medical heterogeneous graph. The sampling process can significantly reduce the number of noisy nodes in the graph. After sampling, some structural information in the graph may be lost; thus, the proposed MBGNN includes a topology-aware module to learn the structural semantic information of the graph. Finally, the proposed MBGNN employs an alternating two-step attention aggregation mechanism to learn multiple embeddings of behaviors for each target node. This attention mechanism provides some interpretability for the analysis of the importance of different behavioral patterns. An comparison of the proposed method with several baseline approaches on two real-world medicare datasets demonstrates the effectiveness of the proposed MBGNN.

Our primary contributions are summarized as follows.

(1) We propose the MBGNN for healthcare fraud. To the best of our knowledge, the proposed MBGNN is the first method to employ multiple sampling strategies to mine patients' behavioral patterns.

(2) We present a topology-aware module. This module learns the global topological information of medical heterogeneous graphs to guide the merging of information between different behavioral patterns.

(3) Extensive experiments on two real-world datasets demonstrate that the proposed MBGNN outperforms existing graph representation learning methods.

2. **Related Work.**

2.1. **Healthcare Fraud.**

Fraud has a significant impact on health care and daily life. Hall [20] employed fuzzy logic and neural network-based induction algorithms to detect the abnormal behavior of fraudsters. However, this approach required the assistance of medical experts to achieve good results and did not allow for effective automatic detection. Subsequent research has shifted toward automating fraud detection. Healthcare fraud detection models are based on supervised learning techniques, e.g., k-nearest neighbors, decision trees, and support vector machines. He et al. [21] proposed neural networks with an error backpropagation algorithm using 28 relevant features to analyze fraudsters. In addition, Yang and Hwang [3] used logistic regression, neural networks, and classification trees to detect fraud among health care providers in the Taiwanese health insurance system. Bauder and Khoshgoftaar [22] employed a random forest technique to classify the imbalanced data. Supervised methods require large amount of labeled data. However, the development of supervised learning for fraud detection is limited due to the diversity of healthcare data and the difficulty obtaining labeled data from related fields in the healthcare domain.

One of the most common unsupervised methods used to find fraudulent records is outlier detection. Anbarasi and Dhivya [23] employed an outlier detection-based approach to identify fraud in health care records. They identified outliers by evaluating health care provider information and claims management data. Yamanishi et al. [24] proposed the SmartSifter algorithm to process and detect medical data. This algorithm evaluates scores based on the similarity between variables in medical data and identifies medical records with high scores as outliers. However, healthcare-related businesses are becoming increasingly granular, and fraud patterns becoming increasingly complex. Gupta et al. [25] proposed an implementation of correlation and regression models to investigate new fraud patterns in response to the emerging fraud challenges facing healthcare systems under COVID-19.

Intelligence and information technologies have help realize significant improvements in terms of the efficiency and accuracy of healthcare fraud detection. However, automated fraud detection is not sufficient to guarantee healthy operation of a healthcare system. Most existing technologies rely on the statistical characteristics of patients, e.g., age, gender, and the amount of money involved. Limited research has been done on patient behavior; however, behavior is one of the most critical aspects of health insurance fraud. Thus, we must understand the behavioral characteristics of fraudsters, and the goal of this study is to identify fraudsters with abnormal behaviors by mining the behavioral patterns of patients.

2.2. **Graph-based Approach.**

Many GNN methods have been used in the anomaly detection field. For example, Belle et al. [26] applied inductive graph representation learning to credit card fraud detection. Here, they used two node embedding frameworks, i.e., GraphSage and FI-GRL, to mine anomalous transaction behavior, and they conducted extensive experiments on financial transaction datasets and achieved good results. Yoo et al. [27] transformed medical datasets into heterogeneous graphs and improved the GraphSage model in the medical industry context. They used the optimized GraphSage model to capture the graph structure to enhance the accuracy of detecting abnormal medical events. Chen et al. [28] proposed the StGNN algorithm for the abnormal behavior trajectory of healthcare fraudsters. This algorithm constructs a graph network comprising hospitals, drugs, and patients, and it employs spatial constraints to find groups with similar behavioral trajectories. In addition, the StGNN algorithm implements a time constraint to determine whether the patient demonstrates abnormal medical behavior. Zhang et al. [29] developed a longitudinal and multimodal data fraud detector based on GNN. They believed that the interval and frequency of medical records are beneficial in identifying fraudsters with abnormal behaviors. Therefore, the idea of this method is to model the electronic medical record in the form of dynamic graphs. They proposed a novel HMF-GNN algorithm that fuses the unique temporal information of dynamic graphs with medical multimodal information to identify abnormal medical behaviors.

A medical heterogeneous graph is a form of medical dataset mapping in which the nodes and edges in the graph are the connections of different entities in a medical scenario. It can be used to analyze the complex relationships between entities and effectively mine medical data for abnormal medical visit behaviors. Learning the embedding representation of nodes in heterogeneous graphs using GNN [30] techniques enables the use of feature information and the mining of patient behavior patterns, i.e., the topological information in the graph. The GNN techniques can also improve the robustness of downstream tasks compared to statistical-based approaches that use feature information.

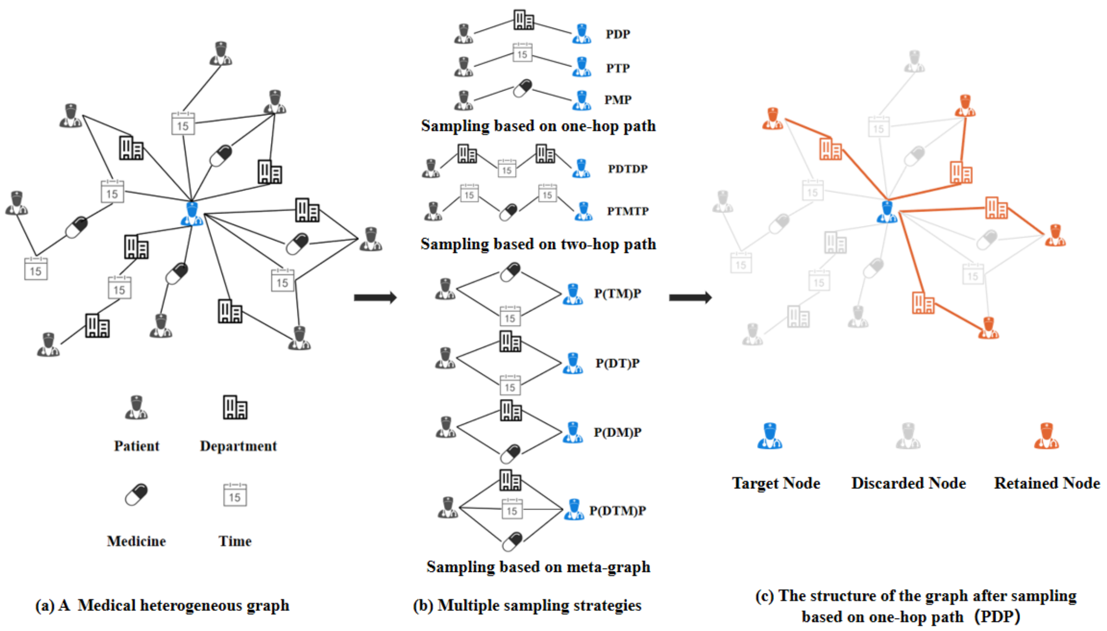

Figure 1. Medical heterogeneous graph (node types and sampling strategies) and the graph structure after applying a sampling strategy

3. Preliminaries.

3.1. Heterogeneous Graph.

A heterogeneous graph can be defined as $G = \{v_h, \varepsilon_h, \varphi, \delta\}$, where v_h is the set of vertices in the graph, and ε_h is the set of edges. $\delta : \varepsilon_h \rightarrow R$ are mapping functions that assign vertices and edges with type information. Here, A and R denote the set of predefined object types and link types, respectively, where $|A| + |R| > 2$.

Example 3.1. *As shown in Figure 1(a), we construct a heterogeneous graph to model a medical dataset. The dataset includes multiple types of objects (patients (P), hospital departments (D), time (T), and medications (M)) and relationships (visit relationship between patients and hospital departments, purchase relationship between patients and medications, etc.).*

3.2. Metapath [31].

A metapath is formed by transforms of a graph schema and can capture rich semantic information preserved in heterogeneous graphs. As an abstract sequence of node types connected by link types, a metapath is denoted as a path in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} A_3 \xrightarrow{R_3} \cdots \xrightarrow{R_l} A_{l+1}$(abbreviated as $A_1 A_2 A_3 \cdots A_{l+1}$) that describes a composite relation $R = R_1 {}^\circ R_2 {}^\circ \cdots {}^\circ R_l$ between node types A_1 and A_l, where $^\circ$ denotes the composition operator on relations.

3.3. Metagraph [32].

A metagraph can be defined as a special kind of path, i.e., $A_1 \xrightarrow{E_1} A_2 \xrightarrow{E_2} A_3 \xrightarrow{E_3} \cdots \xrightarrow{E_4} A_{l+1}$ (abbreviated as $A_1 A_2 A_3 \cdots A_{l+1}$). where E is the set of edges, represents the semantic relationship between two nodes, and $|E| > 2$.

3.4. Fraud Detection on Heterogeneous Graph.

Fraud detection on heterogeneous graphs, where each patient node is marked as fraud or normal node. The goal of the proposed method is to identify anomalous nodes demonstrating fraudulent behavior.

4. Methods.

To mine the implicit behavioral information of healthcare fraudsters, we propose a healthcare fraud detection method based on a MBGNN. In this section, we describe the architecture of the proposed MBGNN. Its basic components are multibehavior pattern decomposition and multipattern fusion. Multibehavior pattern decomposition is a preprocessing stage that finds nodes with similar behavioral patterns. Multipattern fusion comprises two aggregated modules, i.e., the intrapattern fusion and interpattern merging modules. The embedding of nodes with multiple behavioral patterns is learned via continuous iteration. In addition, global topological information is introduced to guide the aggregation process and maintain the quality of embeddings with multiple behavioral patterns. Figure 2 illustrates the overall flow of the proposed MBGNN.

4.1. Multibehavior Pattern Decomposition.

Our medical heterogeneous graph contains different types of nodes, e.g., patients, drugs, and hospital departments. Different types of nodes have different feature vector dimensions, and they are in different feature spaces. However, it is difficult to deal with feature vectors of different dimensions in a uniform framework. Thus, we consider a type-specific transformation matrix M_A to project different types of features into the same latent vector space. As shown in Equation (1), for nodes $u \in V_A$ of type $A \in \mathcal{A}$, we have:

$$W_u = M_A \cdot X_u \tag{1}$$

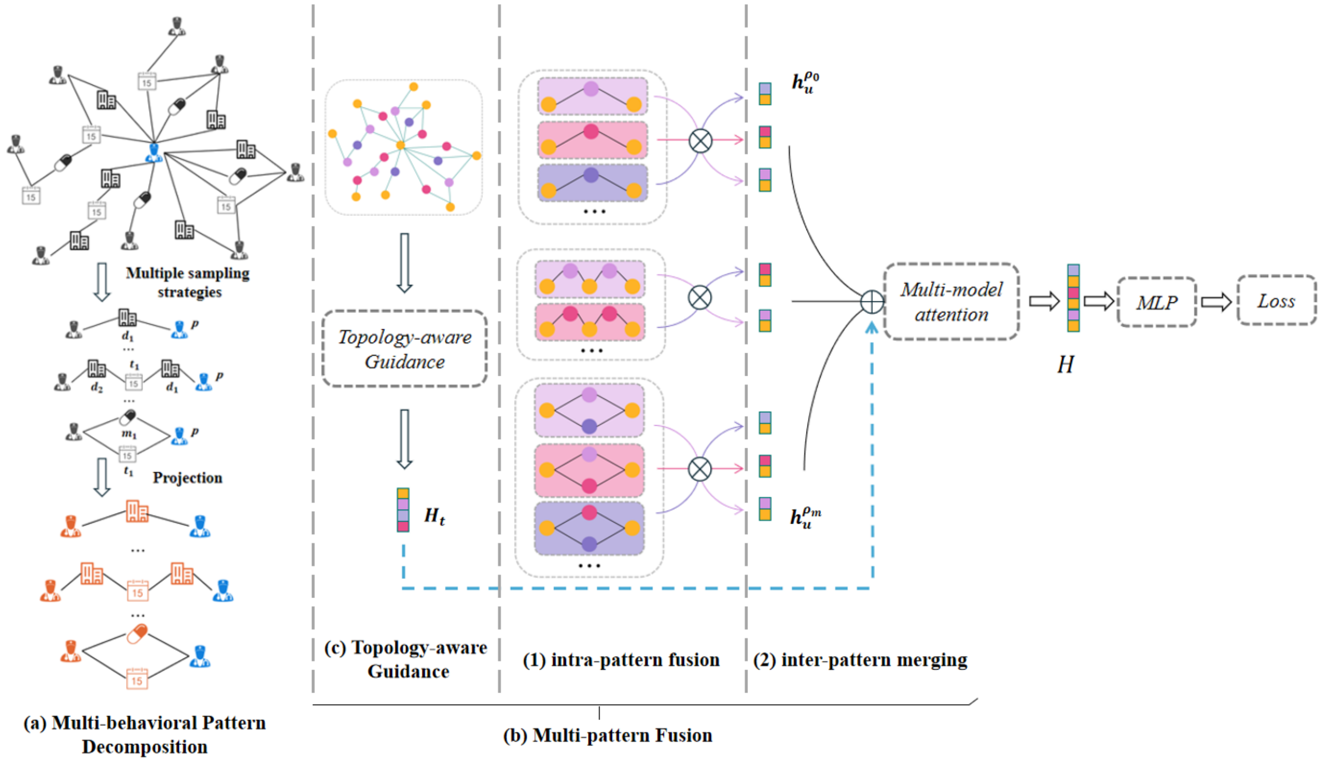

FIGURE 2. Overall framework of the proposed MBGNN

where $X_u \in \mathbb{R}^{d_A}$ is the original feature vector, and $W_u \in \mathbb{R}^{d'}$ is the projected feature vector of node u. Here, $M_A \in \mathbb{R}^{d' \times d_A}$ is the parameter weight matrix of nodes of type A.

Different behavioral patterns of healthcare fraudsters have different meanings; thus, to mine potential behavioral patterns in a heterogeneous graph, we perform multibehavioral pattern-based decomposition of the medical heterogeneous graph. Given nodes with multiple behavioral patterns in the medical heterogeneous graph, we must first sample the nodes in different ways. The proposed MBGNN employs three sampling strategies, i.e., the one-hop neighbor-based metapath, two-neighbor-based metapath, and metagraph-based sampling strategies, to focus on patient nodes that contain multiple behavior patterns, which typically provide more useful information to detect fraud. The processes of these sampling strategies are shown in Figure 1. Through the multibehavior pattern decomposition, nodes with similar trajectories under different behavior patterns can be obtained. In addition, fraudsters tend to have more medical records; thus, they are represented as nodes that are more closely connected in the heterogeneous graph. The feature vectors of the nodes are preprocessed according to this property, and Equation (2) shows the process. After preprocessing, the proposed MBGNN can focus on this type of nodes when learning node embeddings.

$$h_u = W_u \cdot \lambda \frac{d_u}{n} \tag{2}$$

Here, W_u is the projected feature vector of the node, d_u is the degree of node u, n is the number of nodes, and λ is a hyperparameter. After applying the preprocessing operation, the more densely connected nodes are able to get higher attention in the next model.

4.2. **Multipattern Fusion.**

The multipattern fusion module is the core component of the proposed MBGNN and follows a hierarchical fusion structure, i.e., intrapattern fusion, topology-aware module, and interpattern merging.

4.3. **Intrapattern Fusion.**

In a certain behavioral pattern, neighboring nodes have different degrees, which may be due to the different types of nodes or their different local topologies. This means that each neighbor node plays a different role in the intrapattern fusion process and has different importance. Previous studies that used an underlying GNN directly for aggregation could not capture the importance of different neighboring nodes. Thus, the proposed MBGNN employs an attention mechanism to perform aggregation in the intrapattern fusion process. Here, based on a node pair (u, v) under some behavior pattern ρ, the attention mechanism can learn the importance $e_{uv}^{\rho_i}$, which represents the contribution of node v to node u. The importance of node v is calculated as follows.

$$e_{uv}^{\rho_i} = att_{\text{intra}}\ \left(\left[h_u \| h_v\right], \rho_i\right) \tag{3}$$

where h_u and h_v are the projection feature vectors of nodes u and v after processing, respectively, and $||$ denotes the vector connectivity operator. Here, $att_{\text{intra}}\left(\cdot\right)$ denotes the deep neural network performing the attention mechanism. From the node pair importance calculation, the weights between node pairs are entirely dependent on their projection features. Thus, different neighboring nodes of node u can be assigned different contributions.

After obtaining the importance $e_{uv}^{\rho_i}$ between node pairs, we use the $Softmax$ function to obtain the normalized weighting coefficients $a_{uv}^{\rho_i}$, which represents the weight of neighbor node v among all neighbors of node u. $a_{uv}^{\rho_i}$ is calculated as follows.

$$a_{uv}^{\rho_{\mathrm{i}}} = \operatorname{softmax}_v\left(e_{uv}^{\rho_i}\right) = \frac{\exp\left(\operatorname{LeakyReLU}\left(e_{uv}^{\rho_i}\right)\right)}{\sum_{k \in N_u^\rho} \exp\left(\operatorname{LeakyReLU}\left(e_{uk}^{\rho_i}\right)\right)} \tag{4}$$

The proposed MBGNN can perform aggregation within the behavior pattern of node u based on the weight coefficients $a_{uv}^{\rho_i}$ obtained above. Then, the output is performed by the activation function $\sigma(\cdot)$ as follows.

$$h_u^{\rho_i} = \sigma\left(\sum_{v \in N_u^\rho} a_{uv}^{\rho_i} \cdot h_v\right) \tag{5}$$

Note that the weight coefficient $a_{uv}^{\rho_i}$ is produced in a certain behavioral pattern; thus, the specific semantic information of different behavioral patterns can be mined. Considering the high variance of the training process for heterogeneous graph data, we employ a multiheaded attention mechanism to solve this problem. Here, we execute k independent attention mechanisms that connect their outputs. This process is described in Equation (6), and this method can stabilize the entire training learning process effectively.

$$h_u^{\rho_i} = \mathop{\Big\|}_{k=1}^{K} \sigma\left(\sum_{v \in N_u^\rho} a_{uv}^{\rho_i} \cdot h_v\right) \tag{6}$$

Thus, $h_u^{\rho_i}$ can be interpreted as a summary of the target patient node u under behavioral model ρ_i. This shows one aspect of the semantic information contained in node u. Here, assume there are behavioral patterns $\rho_0, \rho_1, \ldots, \rho_m$. After performing the interpattern fusion process, we obtain a sequence $\{h_u^{\rho_0}, h_u^{\rho_1}, \ldots, h_u^{\rho_m}\}$ represented by m behavioral pattern-specific vectors.

4.4. **Topology-aware Module.**

A decomposition of multiple behavioral patterns has been performed previously to study the complex and different behavioral trajectories of healthcare fraudsters. However, after three sampling strategies, although groups with similar behaviors can be found, some topological structure information in the graph will be lost to some extent [33, 34]. These missing topologies are the links between different behavioral patterns. The nature of the multibehavior pattern decomposition step is similar to local clustering; thus, it is difficult to obtain global knowledge in the learning of behavior pattern merging. Therefore, the proposed MBGNN employs the topology-aware module to guide interpattern merging and mine the structure between behavioral patterns more deeply. As shown in process (c) of Figure 2, inspired by node sampling in graph pooling, this process employs a GNN to learn the global topological structure information. The global topological embedding H_t is calculated as follows.

$$H_t = GNN(A, x) \tag{7}$$

Note that the dimension of H_t is kept consistent with the dimension of the embedding $h_u^{\rho_i}$ learned under a single behavioral pattern. Then, the global topological embeddings and the embeddings under individual behavioral patterns are input together to the interpattern merging module for fusion learning.

4.5. **Interpattern Merging.**

Nodes learn different information in different behavior patterns; thus, the information must be merged in this module. Specifically, the proposed MBGNN uses a multimodal focus to assess the importance of different behavioral patterns in the final task. According to the importance, each behavioral pattern is assigned a different weight. The embeddings learned under each behavior pattern are then weighted and summed to obtain the final embedding, as shown in Figure 1(b). To learn the importance of different behavioral patterns, a weight matrix W_ρ is used to transform the specific representation under each behavioral pattern. We then average all transformed embedding representations to summarize the importance of each behavior pattern as follows.

$$S^{\rho_i} = \frac{1}{|G|} \sum_{i \in G} q^T \tanh\left(W_\rho \cdot h_u^{\rho_i} + b_\rho\right) \tag{8}$$

where b_ρ and q^T are the learnable bias vector and parameterized attention vector, respectively. The topology-aware module was implemented to guide the intermodal merging; thus, the importance of the global topology information in the aggregation process must also be evaluated. Equation (9) describes the evaluation process.

$$S^t = q^T \tanh\left(W_t \cdot H_t + b_t\right) \tag{9}$$

Here, W_t is the weight matrix, and b_t is the bias vector. Then, the contribution of each behavior pattern and the global topology are obtained by normalizing it with the $Softmax$ function using Equations (10) and (11).

$$\beta^{\rho_i} = \frac{\exp\left(S^{\rho_i}\right)}{\sum_{i=1}^{m} \exp\left(S^{\rho_i}\right) + \exp\left(S^t\right)} \tag{10}$$

$$\beta^t = \frac{\exp\left(S^t\right)}{\sum_{i=1}^{m} \exp\left(S^{\rho_i}\right) + \exp\left(S^t\right)} \tag{11}$$

This can be interpreted as the contribution of each behavior pattern ρ_i to the fraud detection, where a higher β^{ρ_i} value indicates that behavior pattern ρ_i has higher importance. Finally, all behavior patterns are merged, and the global topology information is used to guide the fusion to obtain the final embedding H, which is formulated as follows.

$$H = \sum_{i=1}^{m} \beta^{\rho_i} \cdot h_u^{\rho_i} + \beta^t \cdot H_t \tag{12}$$

Here, the proposed MBGNN employs cross-entropy as the loss function and optimizes the model via backpropagation. The cross-entropy is expressed as follows:

$$L = -\sum_{l \in y_L} Y^l \ln \left(M \cdot H^l \right) \tag{13}$$

where Y^l and H^l are the labels of the labeled patient nodes and learned embeddings, respectively and M represents the classifier's parameters.

5. Experimental and Result.

5.1. Datasets.

Healthcare-1 and Healthcare-2 are desensitized health insurance datasets from a health insurance management system from a city in China. The Healthcare-1 dataset contains 440 patients, 2,328 drugs, 708 hospital departments, and 351 timestamps after data preprocessing. The ratio of positive to negative samples is approximately 1:2. The Healthcare-2 dataset contains 10,647 patients, 4,718 drugs, 2,751 hospital departments, and 364 timestamps. Of these, 1.5% of the patients were fraudsters. Each dataset was divided randomly into a training set (60%), a validation set (20%), and a test set (20%).

5.2. Baselines.

Isomorphic graph embedding methods: GCN [35], and GAT [36]. We generate isomorphic graphs from medical datasets by leaving patient type nodes and show the best results.

Heterogeneous graph embedding methods: Metapath2Vec [37], HAN [38], and StGNN [28]. These methods were experimented on constructed medical heterogeneous graphs.

All methods were implemented in Pytorch 1.4.0 [39] and Python 3.6.2. The Metapath2Vec, GCN, and GAT methods were implemented based on DGL 0.6.1 [40]. The HAN and StGNN methods were implemented using official source code. All methods were set to run for 300 epochs on the training set. Optimization was performed using the Adam optimizer, and the cross-entropy loss function was employed. We then used the best model on the validation set.

5.3. Node Classification.

We employed an end-to-end training approach for all models; thus, all models were optimized by connecting an MLP layer at the end of each model. In addition, the cross-entropy loss function was employed. The embeddings of the labeled nodes generated by each model were input to a linear support vector machine classifier for node classification. The variance of graph structure data can be quite high. Thus, to avoid chance, the process was repeated 10 times. In this study, we considered two metrics, i.e., Micro-f1 and Macro-f1, to evaluate the node classification effectiveness of each model. The results are shown in Table 1.

As shown in Table 1, the proposed MBGNN obtained the best performance on different training ratios. The peak performance was obtained at training set ratios of 40% and 60%. For the GCN and GAT methods, we found that the homogeneous graph-based embedding approach cannot learn the semantic information between different entities in the medical heterogeneous graph well. This resulted in better quality embeddings learned by the heterogeneous graph-based methods and better classification results. The GCN method aggregates neighbor information by simple averaging, and the GAT and HAN methods employ an attention mechanism to aggregate neighbor information. The attention mechanism correctly weighted the importance of neighbor node information; thus, the GAT and HAN methods could learn better embeddings. Compared to GAT, the HAN method was superior due to its ability to learn semantic information in heterogeneous graphs using a metapath approach. For medical datasets, the StGNN method adds additional time constraints to HAN, which learns more complete timestamp information in medical heterogeneous graphs and demonstrated better classification results. In contrast, the proposed MBGNN found patients with the same behavioral trajectory using multiple strategies and adding global topological embedding to guide the learning of structural information from the medical heterogeneous graphs. The final results are all about a 5% improvement on average compared to StGNN, which has the best results in the baseline model.

These results demonstrate that learning information in the graph using only the metapath approach is insufficient and that the global topology is also important for learning the graph structure.

TABLE 1. Results (%) of node classification on the Healthcare-1 dataset

Dataset	Metrics	Training	Metapath2Vec	GCN	GAT	HAN	StGNN	**MBGNN**
Healthcare-1	Macro-f1	20%	63.81	77.67	79.79	81.11	83.15	**88.33**
		40%	65.91	77.43	80.03	83.27	84.74	**89.23**
		60%	68.97	78.24	81.15	83.92	85.27	**90.17**
		80%	68.25	78.96	81.21	84.86	85.94	**89.17**
	Micro-f1	20%	69.82	78.45	80.61	82.45	84.61	**89.43**
		40%	71.36	79.24	80.84	84.27	85.13	**90.12**
		60%	74.03	79.68	81.26	84.71	85.88	**91.03**
		80%	73.29	80.19	81.14	85.47	86.52	**90.01**

5.4. **Node Clustering.**

We also performed a clustering task on the Healthcare-1 dataset to evaluate the embeddings learned by the different models from the medical heterogeneous graph. Here, the embedding of the labeled nodes generated by each model was input to the K-Means algorithm for node clustering. Since the clustering results of the K-Means algorithm are influenced by the initial center of mass, this process was repeated 10 times, and the average results are reported in Table 2. We used normalized mutual information (NMI) and the adjusted Rand index (ARI) as evaluation metrics.

As shown in Table 2, the proposed MBGNN consistently performed better than other baselines. In addition, Metapath2Vec performed better in clustering compared to the traditional isomorphic graph embedding methods, i.e., GCN and GAT. Metapath2Vec is a random wandering method based on metapaths, and the sequence nodes obtained are close to the graph and projected to the embedding space. Thus, the embedding learned by Metapath2Vec fits better with the idea that K-means clusters nodes based on the Euclidean distance between embeddings. However, Metapath2Vec cannot distinguish the importance of metapaths; thus, it was outperformed significantly by the HAN method. StGNN also sampled nodes with the same trajectory through metapaths. However, the proposed MBGNN samples nodes with a more comprehensive strategy; thus, it obtained the best node clustering performance.

Table 2. Results (%) of node clustering on Healthcare-1 dataset

Dataset	Metrics	Metapath2Vec	GCN	GAT	HAN	StGNN	**MBGNN**
Healthcare-1	NMI	17.19	14.36	15.31	27.58	31.98	**38.22**
	ARI	15.42	13.21	14.47	20.45	30.24	**39.02**

5.5. **Anomaly Detection.**

In subsequent experiments, we validated the performance of different models on the node anomaly detection task on the Healthcare-2 dataset. Approximately 1.5% of the patients in the Healthcare-2 dataset are fraudsters, which accurately simulates real-world medical scenarios. As shown in Table 3, the effect of the isomorphic graph embedding-based method is significantly reduced on the unbalanced sample dataset due to the large proportion of ordinary patients and the noise of isomorphic graph embedding in the neighbor node aggregation process. In contrast, the HAN and StGNN methods selected nodes with similar action patterns based on metapaths. In addition, these methods employ an attention mechanism to distinguish the importance of paths and nodes, which greatly mitigated the impact of data imbalance. The proposed MBGNN captured more comprehensive structural information in the medical graph by mining the behavior patterns among fraudsters using multiple strategies. The embeddings of the proposed MBGNN integrate more comprehensive structural information and incorporate feature information. Compared to the StGNN method, which was the best performing heterogeneous graph embedding model, the proposed MBGNN method improves the average F1 score by 2%.

Table 3. Results (%) of anomaly detection on the Healthcare-2 dataset

Dataset	Training Size	Metrics	Metapath2Vec	GCN	GAT	HAN	StGNN	**MBGNN**
Healthcare-2	20%	F1	71.03	60.67	75.51	77.58	85.19	**87.21**
		Precision	79.5	62.96	71.15	80.41	88.73	**89.17**
	40%	F1	71.66	57.37	72.13	75.63	83.19	**85.29**
		Precision	80.03	58.94	69.84	79.81	89.1	**90.03**
	60%	F1	73.21	60.36	71.28	78.46	87.5	**89.14**
		Precision	80.96	61.21	75.86	81.16	89.98	**90.57**

5.6. **Ablation Study.**

To validate the effectiveness of each component of our model, we evaluated several variant models of the proposed MBGNN: $MBGNN_{\text{one}}$, $MBGNN_{\text{two}}$, $MBGNN_{\text{graph}}$, and $MBGNN_{\text{top}}$. Table 4 shows the results of these experiments for the clustering classification task on the Healthcare-1 dataset. The $MBGNN_{\text{one}}$, $MBGNN_{\text{two}}$, and $MBGNN_{\text{graph}}$ variants are models without the one-hop neighbor path sampling based strategy, without the two-hop neighbor path sampling based strategy, and without the metagraph sampling based strategy, respectively. The $MBGNN_{\text{top}}$ variant does not include the topology-aware module. To demonstrate the effect of the comparison more intuitively, the Micro-f1, Macro-f1, and Weight-f1 results are visualized using bar charts in Figure 3. The performance of the learned embedding representations of several variants of the proposed MBGNN under node classification is shown in Figure 3. As can be seen, the overall effect of $MBGNN_{\text{one}}$ decreased more significantly compared to $MBGNN_{\text{two}}$ and $MBGNN_{\text{graph}}$. This indicates that the one-hop neighbor sampling strategy is more important for the overall model to learn the complete result information and feature information. The $MBGNN_{\text{top}}$ variant without the topology-aware module also exhibited a gap in performance compared to the proposed MBGNN, which highlights the effectiveness of the topology-aware module.

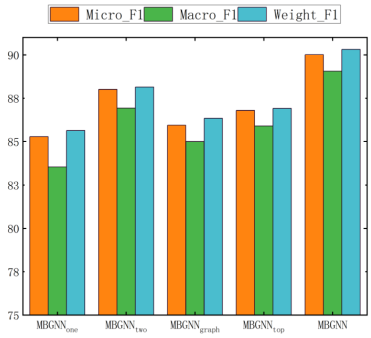

FIGURE 3. Results of MBGNN and its variant models for node classification on the Healthcare-1 dataset

TABLE 4. Quantitative results (%) for ablation study

Dataset	Metrics	Training Size	MBGNN$_{one}$	MBGNN$_{two}$	MBGNN$_{graph}$	MBGNN$_{top}$	**MBGNN**
Healthcare-1	Macro-f1	20%	83.06	86.51	83.39	85.13	**88.33**
		40%	83.39	86.61	84.59	85.17	**89.23**
		60%	83.54	86.93	85.01	85.91	**90.18**
		80%	82.91	85.94	84.64	85.01	**89.17**
	Micro-f1	20%	85.37	87.87	84.57	86.17	**89.43**
		40%	85.34	87.73	85.59	86.16	**90.12**
		60%	85.28	88.01	85.94	86.79	**91.03**
		80%	84.52	86.76	85.28	85.66	**90.01**
		NMI	31.21	35.11	34.81	33.61	**38.22**
		ARI	33.66	34.23	31.51	28.98	**39.02**

5.7. **Visualization.**

In this section, we visually compare the learned embedding representation capabilities of different GNN models. Here, we learned the node embeddings of the appeal methods (i.e., GCN, GAT, Metapath2Vec, HAN, StGNN, and MBGNN) on the Healthcare-2 dataset and projected the embeddings into two-dimensional space. We then used t-SNE to visualize the embeddings in Healthcare-2 (Figure 4). As can be seen, the GCN, GAT, and Metapath2Vec methods did not perform very well because positive and negative sample nodes are mixed together. In contrast, the HAN and StGNN methods separated positive and negative sample nodes effectively by selecting groups with the same action trajectory through metapaths; however, the boundary is still somewhat fuzzy. Compared to the HAN and StGNN methods, the proposed MBGNN found groups with the same behavioral pattern using multiple sampling strategies at a finer granularity. The topology-aware module guides the learning of node embedding representations to distinguish different classes in the visualization. The proposed MBGNN exhibits clearer boundaries and denser cluster distribution, which further highlights the effectiveness of the global topology awareness and multiple sampling strategies.

5.8. **Analysis of Model and Parameters.**

A remarkable feature of the proposed MBGNN is that the importance of different sampling strategies and global topologies is considered in interpath merging using an attention

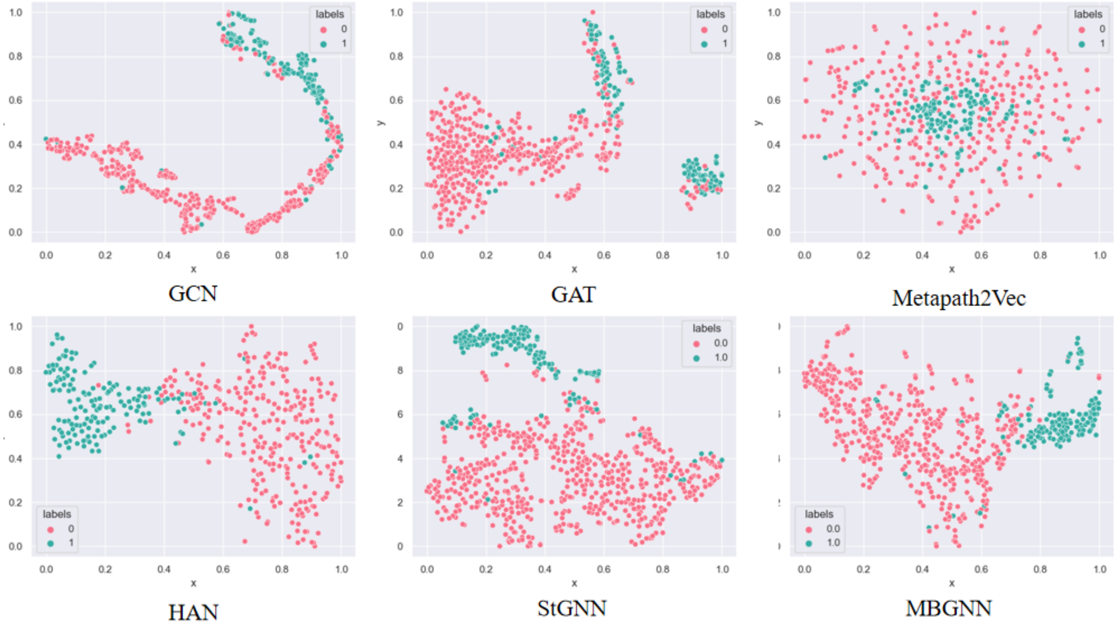

FIGURE 4. Visualization of the learned node embeddings on the Healthcare-2 dataset

mechanism. Thus, we analyzed the attention mechanism to better understand the importance of the different strategies. Table 5 shows the results of attention mechanism. As can be seen, the metagraph sampling strategy has the greatest impact during model training. This means that the characteristics of neighbor nodes selected by metagraph sampling strategy are similar to each other and highly correlated. The level of attention shown in Figure 5 also reflects the interpretability of the model to some degree. Each strategy contributes to the learning of the proposed model, and the global topological information plays a role in the fusion process.

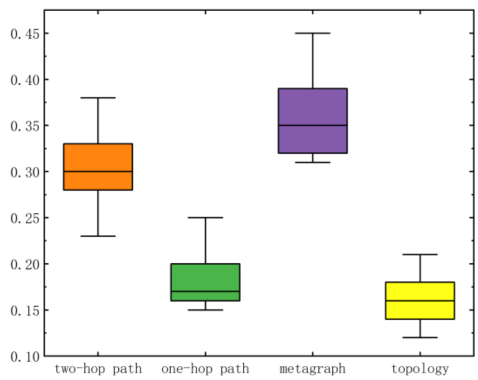

FIGURE 5. Attention contribution values for different strategies and global topology information

We also investigated the sensitivity of the parameters. Here, the node clustering scores (i.e., NMI and ARI) obtained on the Healthcare-1 dataset are reported using the dimensionality of the final embedding Z in the proposed MBGNN, the number of heads

Table 5. Attention result in Healthcare-1

Different strategies	one-hop	two-hop	metagraph	topology
Attention value for different strategies	0.162	0.334	0.372	0.132

k of attention, and the weighting factor λ of the preprocessing. Note that the reported scores are the average of the scores for different training scales (see Section 5.4 for an explanation). The results are shown as a line graph in Figure 6.

First, the effect of the dimensionality of the final embedding Z was tested experimentally, and the results are shown in Figure 6(a). As can be seen, with increasing dimension, the performance shows a trend of initially increasing rising and then slowly decreasing. The reason for this is that the proposed MBGNN selects patient group nodes with the same behavior trajectory using multiple sampling strategies, and it requires an appropriate dimension to encode behavior information. A smaller dimension cannot encode behavior information effectively, which may lead to missing semantic information. In addition, an overly dimension may cause additional redundancy, which leads to further degradation of performance.

We then investigate the performance of the model under different k values by varying the value of attention head number k. The results are shown in Figure 6(b), where the model clustering scores increase gradually when the k value increases gradually. This is due to the high variance of the heterogeneous map. When the value of k is 1, the multiheaded attention is removed, which makes training possible by chance and makes clustering less effective. In addition, when the value of k is increased, the multiheaded attention improves the stability of the training process.

Finally, to obtain the best model performance, we tested the effect of weighting factor λ on the model. The results are shown in Figure 6(c), where the performance shows a trend of increasing and then slowly decreasing as the weighting factor λ increases. It is reasonable that when the weighting factor is too small, the target node is not as sensitive to more similar neighbor nodes in the neighborhood and only evaluates the importance of the neighbor nodes by the initial feature vector. When the weighting factor is too large, the target node spends most of its attention on aggregating the information of such neighbor nodes. However, such nodes only account for a small fraction of the neighbor nodes and cannot learn the information in the complete neighborhood. Thus, a suitable weighting factor is more beneficial for the model's performance.

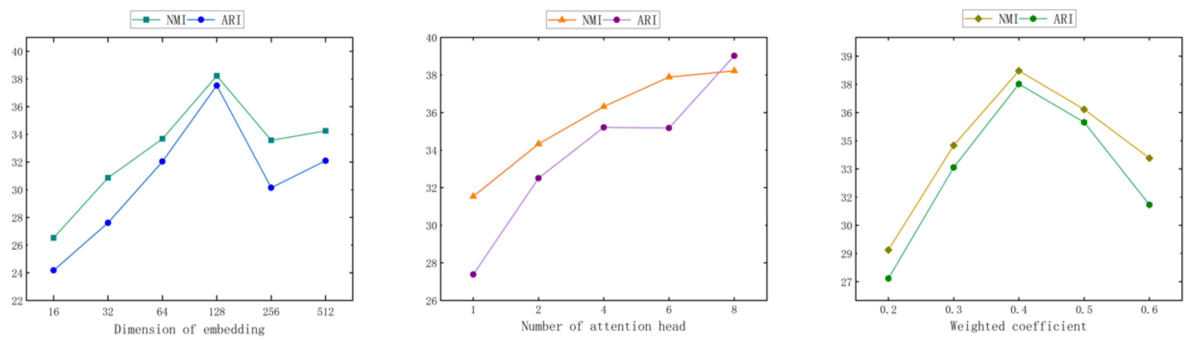

Figure 6. Parameter sensitivity of proposed MBGNN: (a) dimension of the final embedding Z; (b) number of attention heads k; (c) weighting factor λ

6. Conclusion.

In this paper, we have proposed the MBGNN model to mine behavioral attributes in medical heterogeneous graphs through multibehavior pattern decomposition. The proposed MBGNN employs three unique sampling strategies to capture groups with the same behavioral patterns. In addition, the semantic information of nodes in different behavior patterns is aggregated using an attention mechanism, and the semantic information contained in different behavior patterns is fused under the guidance of the proposed topology-aware module.

The learned final embeddings were employed in different downstream tasks to evaluate the performance of the model. Clustering, classification, and anomaly detection experiments were conducted on two real-world datasets to verify the performance of the proposed MBGNN compared to existing methods. Visualizations and analysis of the model's attention mechanism were found to be interpretable. The proposed MBGNN can effectively mine implicit behavioral patterns in historical healthcare data, which is beneficial relative to detecting healthcare fraudsters.

In future, we plan to address two primary limitations. (1) The topology-aware module. The proposed method employs the most basic GNN model to learn global topological information. However, there are better models to realize structure learning of graphs. (2) Time is very important factor, and we currently only use a timestamp as an entity in medical heterogeneous graphs. However, from the perspective of dynamic graphs, it possible to capture information about the structural changes of the graph.

Acknowledgment. This research was funded by the Science Foundation of Fujian Province (No. 2021J011188), the Open Fund of Engineering Research Center of Big Data Application in Private Health Medicine, Fujian Provincial University (KF2020003), and the XMUT Scientific Research Project (No.YKJCX2020103) The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] L. Settipalli and G. Gangadharan, "Healthcare fraud detection using primitive sub peer group analysis," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 23, e6275, 2021.
[2] National Health Care Anti-Fraud Association (NHCAA), "Guidelines to health care fraud," 2020. [Online]. Available: http//www.nhcaa.org
[3] W. S. Yang and S. Y. Hwang, "A process-mining framework for the detection of healthcare fraud and abuse," *Pergamon*, vol. 31, no. 1, pp. 56–68, 2006.
[4] W.-L. Shiau, K. Siau, Y. Yu, and J. Guo, "Research commentary on is/it role in emergency and pandemic management: Current and future research," *Journal of Database Management (JDM)*, vol. 32, no. 2, pp. 67–75, 2021.
[5] S. P. Keehan, G. A. Cuckler, J. A. Poisal, A. M. Sisko, S. D. Smith, A. J. Madison, K. E. Rennie, J. A. Fiore, and J. C. Hardesty, "National health expenditure projections, 2019–28: Expected rebound in prices drives rising spending growth: National health expenditure projections for the period 2019–2028," *Health Affairs*, vol. 39, no. 4, pp. 704–714, 2020.
[6] R. Müller and J. Hillegersberg, "Fraud detection within medicaid," MA thesis, University of Twente, 2012.
[7] Medicare Fraud Strike Force, "Office of inspector general," 2013. [Online]. Available: https://www.oig.hhs. gov/fraud/strike-force/
[8] China Medical Security Bureau, "Statistical bulletin on the development of china's medical security business in 2021," 2021. [Online]. Available: http://www.nhsa.gov.cn/art/2022/6/8/art_7_8276.html
[9] Bank Of China, "Bank of china foreign exchange rates," 2022. [Online]. Available: https://www.boc.cn/sourcedb/whpj/
[10] A. J. Mestres, G. L. Casasnovas, and J. V. Castelló, "The deadly effects of losing health insurance," *European Economic Review*, vol. 131, 103608, 2021.

[11] R. Paul and A. S. M. L. Hoque, "Mining irregular association rules based on action & non-action type data," in *2010 Fifth International Conference on Digital Information Management (ICDIM)*. IEEE, pp. 63–68, 2010.

[12] S. Kumari, P. Chaudhary, C. M. Chen, and M. K. Khan, "Questioning key compromise attack on ostad- sharif et al.'s authentication and session key generation scheme for healthcare applications," *IEEE Access*, vol. 7, pp. 39717–39720, 2019.

[13] C. M. Chen, S. Liu, S. A. Chaudhry, Y. C. Chen, and M. A. Khan, "A lightweight and robust user authentication protocol with user anonymity for iot-based healthcare," *CMES-Computer Modeling in Engineering & Sciences*, vol. 131, no. 1, pp. 307–329, 2022.

[14] P. A. Ortega, C. J. Figueroa, and G. A. Ruz, "A medical claim fraud/abuse detection system based on data mining: A case study in chile," *DMIN*, vol. 6, pp. 26–29, 2006.

[15] M. Pavithra and D. Parvathi, "A survey on data mining approaches for healthcare domain," *International Journal of Bio-Science and Bio-Technology*, vol. 5, no. 5, pp. 241–266, 2013.

[16] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," *Advances in Neural Information Processing Systems*. Curran Associates Inc, 2016, pp. 2001–2009.

[17] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in Neural Information Processing Systems*. Curran Associates Inc, 2017, pp. 1025–1035.

[18] D. Wang, P. Cui, and W. Zhu, "Structural deep network embedding," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2016, pp. 1225–1234.

[19] J. Chang, C. Gao, Y. Zheng, Y. Hui, Y. Niu, Y. Song, D. Jin, and Y. Li, "Sequential recommendation with graph neural networks," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, 2021, pp. 378–387.

[20] H. C, "Intelligent data mining at ibm: new products and applications," *Intelligent Software Strategies*, vol. 7, pp. 1–16, 1996.

[21] H. He, W. Graco, and X. Yao, "Application of genetic algorithm and k-nearest neighbour method in medical fraud detection," in *Asia-Pacific Conference on Simulated Evolution and Learning*. Springer, 1999, pp. 74–81.

[22] R. Bauder and T. Khoshgoftaar, "Medicare fraud detection using random forest with class imbalanced big data," in *IEEE International Conference on Information Reuse and Integration for Data Science*. IEEE, 2018, pp. 80–87.

[23] M. Anbarasi and S. Dhivya, "Fraud detection using outlier predictor in health insurance data," in *2017 International Conference on Information Communication and Embedded Systems (ICICES)*. IEEE, 2017, pp. 1–6.

[24] K. Yamanishi, J.-I. Takeuchi, G. Williams, and P. Milne, "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms," *Data Mining & Knowledge Discovery*, vol. 8, no. 3, pp. 275–300, 2000.

[25] R. Y. Gupta, S. S. Mudigonda, P. K. Baruah, and P. K. Kandala, "Implementation of correlation and regression models for health insurance fraud in covid-19 environment using actuarial and data science techniques," *International Journal of Recent Technology and Engineering* , vol. 9, no. 3, pp. 699–706, 2020.

[26] R. Van Belle, C. Van Damme, H. Tytgat, and J. De Weerdt, "Inductive graph representation learning for fraud detection," *Expert Systems with Applications*, vol. 193, 116463, 2022.

[27] Y. Yoo, D. Shin, D. Han, S. Kyeong, and J. Shin, "Medicare fraud detection using graph neural networks," in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE, 2022, pp. 1–5.

[28] J.-P. Chen, P. Lu, F. Yang, R. Chen, and K. Lin, "Medical insurance fraud detection using graph neural networks with spatio-temporal constraints," *Journal of Network Intelligence*, vol. 7, no. 2, pp. 480–498, 2022.

[29] J. Zhang, F. Yang, K. Lin, and Y. Lai, "Hierarchical multi-modal fusion on dynamic heterogeneous graph for health insurance fraud detection," in *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2022, pp. 1–6.

[30] Z. Wang, Z. Wei, Y. Li, W. Kuang, and B. Ding, "Graph neural networks with node-wise architecture," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2022, pp. 1949–1958.

[31] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks," *Proceedings of the Vldb Endowment*, vol. 4, no. 11, pp. 992–1003, 2011.

[32] Y. Fang, W. Lin, V. W. Zheng, M. Wu, J. Shi, K. C.-C. Chang, and X.-L. Li, "Metagraph-based learning on heterogeneous graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 1, pp. 154–168, 2019.

[33] X. Liu, J. Cheng, Y. Song, and X. Jiang, "Boosting graph structure learning with dummy nodes," in *International Conference on Machine Learning.* PMLR, 2022, pp. 13704–13716.

[34] S. Liu, R. Ying, H. Dong, L. Li, T. Xu, Y. Rong, P. Zhao, J. Huang, and D. Wu, "Local augmentation for graph neural networks," in *International Conference on Machine Learning.* PMLR, 2022, pp. 14054–14072.

[35] T. N. Kip F and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2017.

[36] P. Velikovi, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *International Conference on Learning Representations*, 2018.

[37] Y. Dong, N. V. Chawla, and A. Swami, "metapath2vec: Scalable representation learning for heterogeneous networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Association for Computing Machinery, 2017, pp. 135–144.

[38] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *The World Wide Web Conference.* Association for Computing Machinery, 2017, pp. 2022–2032.

[39] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems.* Curran Associates, Inc, 2019, pp. 1–12.

[40] M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, and Y. Gai, "Deep graph library: Towards efficient and scalable deep learning on graphs," in *ICLR workshop on representation learning on graphs and manifolds*, 2019, pp. 1–18.