Identification of Malicious E-commerce Users Based on User Rating Behavior and GNN

Chunyan Wu*, Zemei Liu, Zhaocui Li
Department of Senior Technician, Shandong Labor Vocational and Technical College, Jianan 250022, China
E-mail: yanzi691177@126.com
*Corresponding author

Keywords: user rating, GNN, E-commerce, malicious users, recognition

Received: October 21, 2024

With the rapid rise of e-commerce, authentic user evaluations are particularly important in purchasing decisions. The increase in fake evaluations has made effective identification a top priority. The study addresses this challenge by introducing a reputation strategy that combines scoring patterns and differences, GNR metrics, and adversarial data augmentation technology, to improve the effectiveness of fraud detection. The study conducted experiments using three datasets, Netflix, Movielens2, and Movielens_100, which record user ratings of movies at different scales. The main performance metrics include recall, F1 value and area under the curve (AUC). The experiment showed that when the proportion of fraudulent users was 0.035, the recall value of reputation ranking technology strategy based on user evaluation mode and bias was 0.79, with stability exceeding 0.990. After applying the GNR metrics, the deviation ranking method showed a significant reduction in peak user count and improved the overall performance by 9.40%. The accuracy of the iterative group-based ranking and the iterative balance ranking increased by 2.89% and 2.54%, respectively. After introducing the adversarial data augmentation technology, the fraud detector based on graphical neural networks improved recall and F1 by 1.20% and 1.34%, respectively compared with the disguised fraudster model in the case of data scarcity. It can be seen that the method combining multiple strategies and technologies demonstrates improved performance in e-commerce user evaluation fraud detection, far surpassing traditional methods. This study has brought significant significance and value to the e-commerce field.

Povzetek: Predstavljena je kombinacija analize uporabniških ocen in grafnih nevronskih mrež (GNN) za identifikacijo zlonamernih e-trgovinskih uporabnikov. Predlagana metoda izboljšuje zaznavanje prevar ter povečuje zanesljivost modela z uporabo metrik GNR in tehnike nasprotnega učenja podatkov.

1 Introduction

With the vigorous development of e-commerce, user review has become a key decision-making factor for purchasing goods on e-commerce platforms. However, given the increasing number of fraudulent user comments, their impact on buyers and businesses is becoming increasingly significant, leading to distorted consumer decisions and loss of commercial benefits [1-3]. Therefore, accurately identifying forged user reviews and ensuring the integrity and health of the e-commerce environment are important issues that urgently need to be addressed in the current e-commerce field. The main research challenge is how to accurately distinguish between genuine and fraudulent evaluations, while also considering the concealment and variability of fraudulent behavior. This makes traditional methods often ineffective to a certain extent and unable to meet the needs for efficient and accurate identification of counterfeit evaluations [4-6]. Traditional rule-based methods are often difficult to handle large amounts of complex evaluation data. Machine learning methods also have certain limitations in dealing with data sparsity and cold start problems. Graph Neural Networks (GNN) is an emerging deep learning model, which can effectively learn features on graph structured data. Compared with traditional machine learning methods, GNN has shown significant advantages in handling complex user behavior patterns and large-scale data. By modeling users and their interaction behaviors in e-commerce platforms as graphs, GNN can better capture the relationships and behavior patterns between users, thereby improving the accuracy and robustness of malicious user identification. The main goal of this research is to address the problem of how to accurately identify fake user reviews among many genuine reviews. To this end, this paper proposes an integrated approach that combines a user rating mechanism with a Reputation Ranking Method (RRM) based on Rating Patterns and Rating Bias (RPRD), metrics, and adversarial data augmentation technology, aiming to improve the accuracy of identifying fake user ratings. The innovation of this study is to combine user rating behavior analysis with GNN to improve the robustness and accuracy of the model through adversarial data augmentation technology. The study introduces GNR metrics to improve the accuracy of identifying fraudulent user behaviors and adds adversarial data augmentation technology to gain insights into the

behavioral patterns of malicious users to improve the stability of identifying user behaviors. The overall structure of the study includes four sections. Firstly, the research achievements and shortcomings of neural networks and user ratings both domestically and internationally are summarized. Secondly, a reputation technology that combines user rating mechanisms with evaluation biases is introduced, along with GNR metrics and adversarial data augmentation technology. Then, related experiments are conducted to identify and analyze malicious e-commerce users based on user rating behavior and GNN. Finally, the experimental results are summarized, and the shortcomings and future research directions are proposed.

2 Related works

The scoring behavior of e-commerce users has become a research focus in recent years. With the widespread application of neural networks in data analysis, their value in user rating analysis is increasingly significant. Rating not only reveals consumer experience, but may also involve malicious behavior [7]. The following introduces some research on neural networks and user ratings. Wang *et al.* proposed a model based on neural networks-Cross domain Explicit Implicit Hybrid Neural Network (CEICFNet). This model combined deep neural networks to learn potential factors from explicit scoring and implicit interaction, thereby achieving cross domain learning. It included a domain shared multi-layer perception network that learned the potential factors of user and project ratings, serving as a knowledge transfer bridge. The experimental data proved that the model had better performance [8]. Guo *et al.* proposed a deep GNN social recommendation framework suitable for future applications of the Internet of Things. This method first converted the user and item feature space into two graph networks, and encoded them through a GNN strategy. Next, these two encoding spaces were nested into the potential factors of matrix decomposition to fill in the missing ratings in the user product rating matrix. The efficiency and stability of the model were verified through experiments on three real datasets [9]. Liu *et al.* proposed a hybrid neural recommendation model that extracted deep representations of users and items from ratings and comments. This model included a comment-based encoder to model users and items, as well as a prediction module to make recommendations based on ratings and comments. To fully utilize the information in comments, the study introduced a comment level attention mechanism to select key comments. The tests on multiple datasets showed that this model outperformed existing methods in recommendation tasks [10]. Da *et al.* combined fine-grained user item semantic information and used neural attention techniques to learn representation. The model learned heterogeneous user/item representations through comments and specific interactions. The model further integrated scoring-based features with comment specific features. A factor decomposition machine was applied to make predictions on a shared hidden layer. The experimental results showed that this method outperformed the baseline method in rating prediction and ranking [11].

Tang *et al.* proposed a neural joint model based on comments, product categories, and user co-purchase information. The comment module was responsible for learning user and product information, while the heterogeneous information network module extracted associated features from the heterogeneous information network. These data were subsequently fed into feature interaction methods for rating prediction. Testing on three datasets on Amazon showed that the model performed better than other baseline methods [12]. Wang *et al.* proposed a multi-attention deep neural network recommendation model that combined embedding and matrix decomposition, aiming to address data sparsity and cold start issues. The GPU-based deep network ensured the scalability of the model. Compared with traditional matrix decomposition methods, this model exhibited better prediction performance on real datasets, further improving the quality and performance of recommendations [13]. Shi *et al.* proposed an emotion enhanced neural graph recommendation method that integrated text comments and bipartite graph information. Through a hierarchical attention mechanism and emotional assistance tasks, users could identify their multifaceted preferences for items from comments. Graph convolutional networks were used to simulate information diffusion in user item interaction graphs, thereby capturing user interactions and preferences. Finally, the decomposition machine model was used to implement recommendations. The experiment on two datasets showed that the prediction accuracy of this model surpassed other related methods [14]. Liang *et al.* combined rating and topic level comment information into a deep neural framework. User preferences and item attributes in comments were captured through topic alignment operations and attention mechanisms. The neural prediction layer of this model extended user and item representations, integrating potential scoring factors and text information. The experimental results showed that this method surpassed the current state-of-the-art recommendation techniques in rating prediction, which could classify user/project comments by topic [15]. Li *et al.* proposed a GNN based on local and global perceptual memory (LGM-GNN) to optimize the fraud detection task. The network fused and utilized local and global information through relationship-aware embedding and interactive aggregation of local and global memory networks. The results showed that the LGM-GNN outperformed other methods on a real-world fraud detection dataset [16]. To improve the performance of GNN in e-commerce review fraud detection, and solve the sample class imbalance and fraud camouflage, Li *et al.* proposed a fraud detection method based on self-paced graph contrast learning (SPCL-GNN). The graph structure was first optimized by label equalization and self-paced graph contrast learning. Then the attention mechanism was introduced for node embedding. The results showed that SPCL-GNN outperformed the baseline method on Amazon and

YelpChi datasets [17]. To deal with gang fraud in e-commerce platforms, Yu *et al.* propose a novel end-to-end semi-supervised Group-based Fraud Detection Network (GFDN). The model supported real-world fraud detection by analyzing the characteristics of group fraud behavior. The results showed that the GFDN exhibited better effectiveness and efficiency in group fraud detection on bi-directional graphs on Taobao and Bitcoin datasets [18].

In summary, many scholars have conducted in-depth research on user ratings and recommendation accuracy, and have successfully applied them in multiple fields. However, there is still relatively little comprehensive research on e-commerce user ratings and GNN e-commerce malicious user identification. Therefore, this study proposes a new solution strategy for e-commerce user rating and recognition. The study first introduces a reputation strategy based on user rating mechanism and evaluation bias. Then the reputation calculation method is enhanced by combining GNR metrics. Finally, the Free Large-scale Adversarial Augmentation on Graphs (FLAG) technology further strengthens the deception recognition model. This comprehensive model provides an efficient and reasonable new solution for user rating and recognition in the e-commerce field, which is of great significance for the development of this field.

Table 1 Different literature and research methodology gaps

Literatures	Research purpose	Results and gaps
Wang et al. [8]	CEICFNet	Can learn potential factors, but cannot achieve user fraud identification and analysis
Guo et al. [9]	Deep GNN Social Recommendation Framework	Able to fill in the potential factors of user coding, but under-applied for user fraud identification and correlation analysis
Liu et al. [10]	Hybrid Neural Recommendation Model	Capable of user rating and recommendation, not able to refine the user fraud identification process
Da et al. [11]	Neural attention techniques	Capable of integrating user rating features but unable to apply user fraud prediction
Tang et al. [12]	Neural Joint Model	Capable of user data interaction, but poor for user behavior recognition
Wang et al. [13]	Multi-attention deep neural network recommendation model	With good matrix decomposition ability, but cannot perfectly guarantee the predictive analysis of user fraudulent behavior
Shi et al. [14]	Sentiment-enhanced neural graph recommendation method	User comment recognition can be achieved, but the effect on user behavior recognition still needs to be improved
Liang et al. [15]	Deep Neural Networks	Able to recommend users based on topics does not build a more effective user behavior system
Li et al. [16]	LGM-GNN	Can combine global information, but the model performance is poor
Li et al. [17]	SPCL-GNN	Able to improve user fraud problem camouflage, but the method is more complicated to use
Yu et al. [18]	GFDN	Can analyze team fraud behavior, but cannot identify individual behavior

3 E-commerce malicious user identification method construction based on user rating behavior and GNN

This study focuses on the e-commerce user rating and recognition. Firstly, a reputation strategy based on user rating mechanism and evaluation bias is proposed. Subsequently, a GNR-based indicator is provided to enhance reputation calculation methods. Finally, the FLAG technology is used to enhance the deception recognition model.

3.1 Reputation model construction based on user rating mechanism and evaluation bias

Malicious users are becoming increasingly apparent on online e-commerce trading platforms. According to user rating behavior, these actors often have a clear purpose. To distinguish these consumers, an evaluation mechanism is designed. Consumers with high evaluation values are considered trustworthy, while low evaluation values may belong to negative actors. This study introduces raking habits and tendencies to identify these users. To ensure the authenticity of the data, the study uses a ternary model to improve computational efficiency, expressed as (i, j, r), where i represents consumers, j represents items, and r represents the corresponding raking. Based on the rating trend, the evaluation values are calculated and the classification strategy is optimized. The specific strategy is that most ratings should reflect the actual value of the project and follow a normal distribution [19-20]. The probability distribution of each participant is shown in equation (1).

$$g_i(q_a^{abs}) = N(f_i(O_a), \sigma_i^2) \quad (1)$$

In equation (1), μ and σ respectively represent the mean and difference. q_a^{abs} represents the absolute value of the identification item a. O_a represents the a -numbered item. $f_i(\cdot)$ is the i consumer's identification method. Due to the difficulty in knowing the true value of items in actual situations, this study replaces their true value with the average number of evaluations obtained from each item. The actual value of each item is shown in equation (2).

$$q_a^{avg} \equiv \frac{1}{|U_a|}\sum_{i \in U_a} b_{ia} \quad (2)$$

In equation (2), $|U_a|$ represents the number of users who rate the item a. $|U_a|$ represents the rating of consumer i to item a. If the evaluation scores of the item O_a are sufficiently concentrated, the average value q_a^{avg} of the item should be close to its intrinsic value q_a^{abs}. Therefore, equation (1) can be optimized, as shown in equation (3).

$$g_i(q_a^{avg}) = N(b_{ia}, \sigma_i^2) \quad (3)$$

In equation (3), σ_i^2 refers to the variation coefficient of consumer users i. Due to differences between consumer feedback and the actual value of the item, the deviation degree of each consumer identification method needs to be calculated in the future. The threshold ranges in equations (1)-(3) are determined by analyzing the distribution of user ratings and identifying the points where the behavior of malicious users deviates significantly from that of real users. It can be adjusted according to the standard deviation of user ratings. The average price of an item follows a normal model within a subset O_{is}. Therefore, the probability model for consumer rating is shown in equation (4).

$$g_i(q_{is}) = N(s, \frac{\sigma_i^2}{|O_{is}|}) \quad (4)$$

In equation (4), $q_{is} = (1/|O_{is}|)$ represents the evaluation quality of consumers i providing s level evaluation. $|O_{is}|$ represents the number of s level items provided by the user i. This study uses Z value to measure the reliability of evaluations submitted by consumers, as shown in equation (5).

$$Z_{is} = \frac{\sqrt{|O_{is}|}(q_{is} - s)}{\sigma_i} \quad (5)$$

In equation (5), σ_i represents the deviation of consumers i. The previous steps are further integrated to assign evaluation credibility values r_i to each consumer. In the construction stage of this method, the Z value of the s level proposed by consumers i is negatively correlated with their evaluation reputation. Therefore, the evaluation reputation method for each consumer should be the $|Z_{is}|$ descent method. To integrate each $|Z_{is}|$ into a reputation value r_i, it is calculated from equation (6).

$$r_i = -\frac{\sigma_i \sum_{S=1}^{L} |Z_{is}|}{\sqrt{|O_i|}} = -\frac{\sum_{S=1}^{L} \sqrt{|O_{is}|}\,|q_{is} - S|}{\sqrt{|O_i|}} \quad (6)$$

In equation (6), L represents the maximum rating limit for consumers. Z_{is} is the adjusted Z value of consumer i rating level. On this basis, the motivations, rating trends, and biases of malicious consumer users are further analyzed. To improve the accuracy of rating reputation values, this study introduces the information entropy index and likelihood difference based on the rating reputation value. The former is used to describe the rating trend, and the later is used to describe the scoring bias. The information entropy index of each consumer user is displayed equation (7).

$$g_i = (-1)^k \sum_{s=1}^{L} p(t_{is}) \log_e p(t_{is}) \quad (7)$$

In equation (7), k represents the measurement of user attributes. $k = 1$ represents the identification of random malicious users, and $k = -1$ represents extreme malicious users. To correct the scoring bias, probability analysis is used in the study. h_i represents the probability shift of each consumer user, as shown in equation (8).

$$h_1 = \frac{1}{2}\sum_{S=1}^{L-1} \sqrt[3]{|p(t_{i(s+1)}) - p(t_{is})|} \quad (8)$$

In summary, this study elaborates on the reputation technology process based on user rating mechanism and evaluation bias through examples. Figure. 1 shows the bipartite network composed of users and products, involving 18 evaluations. This includes the user product rating table, the average matrix of the product, the user rating matrix, the user information entropy and probability difference, and the final reputation matrix of the user.

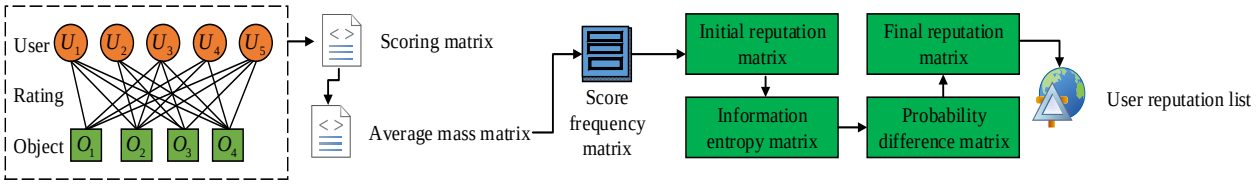


Figure 1: A two-part network diagram of the user and the product

3.2 GNR index for reputation ranking based on evaluation preference and bias

On the basis of reputation ranking based on user evaluation preferences and biases, a comprehensive framework is proposed to improve the current ranking method. This framework not only improves the efficiency of existing technologies, but also reduces the development time and cost. Due to the frequent attacks of malicious users on online rating systems, such as extremely malicious users employed by merchants or randomly rated malicious users, this affects the credibility of the system. There is a significant difference in the rating behavior between normal users and malicious users. Normal users are distributed in a hierarchical manner, while malicious user rating patterns are not. In addition, the study introduces Gini coefficient and range as indicators to measure the distribution differences of ratings, which are combined into GNR metrics, as shown in Figure. 2.

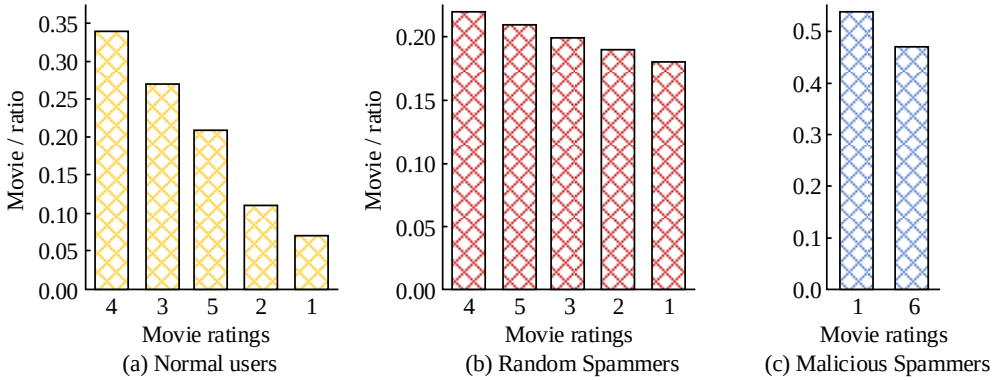


Figure 2: User ratings of GNR metrics

The GNR comprehensive standard is derived from the past evaluations of consumer users, calculated through Gini parameters and extreme differences. According to equation (9), it is possible to standardize the interval between two ratings for a specified user.

$$G_i = \theta\left(\frac{1}{2 * t_i^2 * r_i^{avg}}\right)\sum_{a=1}^{a=t_i}\sum_{\beta=1}^{\beta=t_i}(r_{ia} - r_{i\beta}) \quad (9)$$

In equation (9), r_i^{avg} represents the median of evaluations assigned to each user. t_i represents the frequency of all evaluations by user j. r_{ia} represents the evaluation of user i to item a. The Gini coefficient and range introduced in equation (9) are used to measure differences in rating distribution. This is to capture the inequality in rating distribution and identify users who rate projects in an abnormally biased manner. The extreme difference of each user is shown in equation (10).

$$Range_i = \frac{t_{i\beta}^{\max} - t_{i\beta}^{\min}}{t_{i\beta}^{\max} + t_{i\beta}^{\min}} \quad (10)$$

In equation (10), $t_{i\beta}^{\max}$ represents the highest evaluation frequency of the user x's β class. $t_{i\beta}^{\min}$ represents the lowest evaluation frequency of the user i in the β class. The Deviation-based Ranking (DR) algorithm reveals the reliability of users through user evaluation patterns. This is essentially a difference-based arrangement technique, where user statistical data is defined by scores. This strategy divides each user's credit based on their accuracy in project evaluation [21]. Combining the GNR standard with the DR strategy, this method mainly derives from the fact that most users' evaluations of a certain product are closely related to the actual quality of the product, as shown in equation (11).

$$g_i(q_a^{abs}) = N(f_i(o_a), \sigma_i^2) \quad (11)$$

In equation (11), μ represents the mean and σ refers to the variance. This study illustrates the process of GNR metrics in DR strategy through examples, as shown in Figure. 3. Initially, it is a bidirectional graph network composed of items and users, covering 18 ratings. Subsequently, it is a rating table for users and items, followed by an average quality table. Next is the

Gini value and extreme value matrix for users, followed by a GNR matrix. Finally, the user's reputation ranking table is displayed.

The Iterative Group-based Ranking (IGR) algorithm introduces iterative thinking on the basis of the grouping sorting algorithm to redistribute user reputation. This strategy is that in a large group, user ratings with high credibility are more persuasive. The application of GNR algorithm in IGR is shown in Figure. 4. The beginning is a two-way diagram constructed by the item and the user. Subsequently, there is a user rating table for the item, followed by a weight restructuring table and a rating reward table. Next is the user Gini value and extreme value table, followed by the user GNR table, user first iteration reputation, and final reputation ranking.

The Iterative Balance Ranking (IBR) algorithm is a high-quality reputation ranking method that should reduce bias in user ratings. This law aims to fairly evaluate the reputation of each user and reveal the true performance of the product. The application of GNR metrics in IBR is shown in Figure. 5.

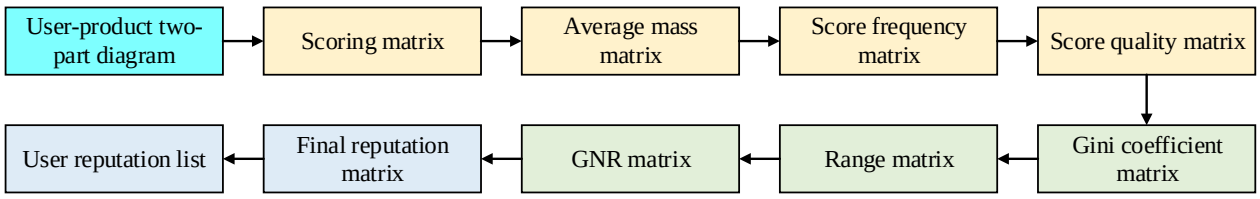


Figure 3: GNR metrics flow in the DR policy

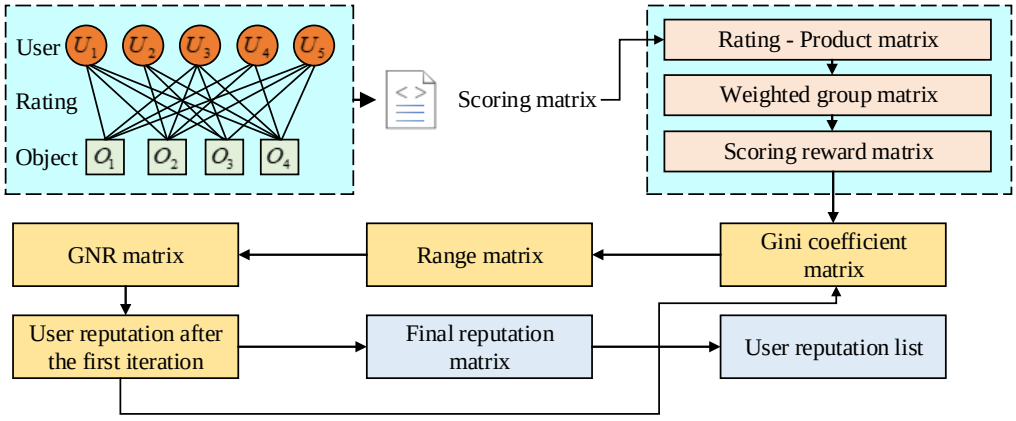


Figure 4: Flow chart of GNR metrics in IGR algorithm

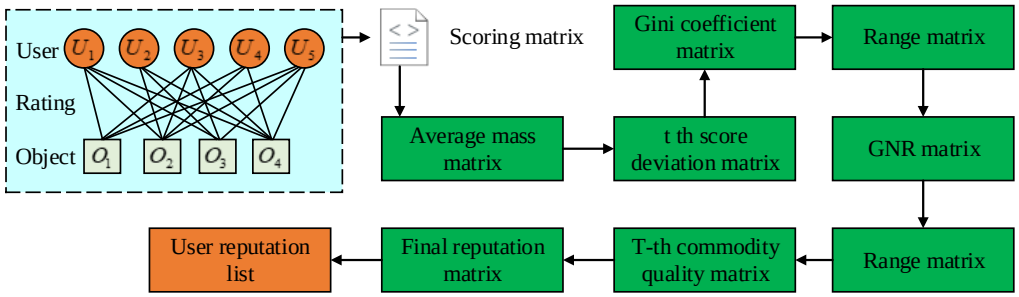


Figure 5: Flow chart of GNR metrics in IGR algorithm

3.3 GNN-based method for identifying malicious e-commerce users

To improve the efficiency of fraud detection, this study introduces the FLAG method and the Graph Convolutional Neural Network (C-GNN) model, which is a GNN-based fraud detectors against camouflaged fraudsters. GNN can effectively capture the complex interaction between users and projects, and identify patterns hidden in the graph structure. After introducing the FLAG method in GNN, GNN can enhance its robustness to malicious data disturbances, and improve the accuracy and stability of the model in fraud detection. GNN can also extract deep features of nodes layer by layer through multi-layer convolution operations, improving the recognition ability of fraudulent behavior. Finally, GNN can evaluate the performance stability of models in large-scale data and malicious user attacks. Therefore, GNN is used as the main structure of the study. The C-GNN can effectively process graph structured data and capture complex relationships between users and projects. The nodes in the graph

network represent users and items, and the weights of edges represent ratings. Through multi-layer graph convolution operation, the information of node neighbors is aggregated layer by layer to learn the representation of nodes. In graph convolution, attention mechanism is introduced to focus on neighboring nodes with strong correlation with the target node. Finally, after introducing the FLAG method, the robustness of the model is enhanced by gradient perturbation of node attributes. GNN can combine Gini coefficient and range as evaluation criteria to optimize the ranking method of user rating preferences and biases, which can significantly improve the performance of existing methods. In the initial GNN, IGR and IBR algorithms are used to iteratively adjust user reputation, reduce rating bias, and improve fairness and accuracy of ratings. Finally, adding FLAG technology to generate adversarial samples can enhance the model generalization ability and performance in data scarcity situations. Under this strategy, a fraud detection network is established and trained. Furthermore, through data augmentation, false nodes and real nodes are generated and mixed. During testing, real nodes are used to predict classification. Model evaluation shows that the FLAG method significantly improves the accuracy of fraud detection. Equation (12) shows the calculation details. In the digital age, it is particularly important to establish the efficient fraud detection model.

$$D^{(l)}(v,u) = \left\| \sigma(MLP^{(l)}(h_v^{(l-1)})) - (MLP^{(l)}(h_u^{(l-1)})) \right\|_1 \quad (12)$$

In equation (12), $D^{(l)}(v,u)$ shows the l_1 interval between core point v and its adjacent point $u \in N(v)$ at the l level. In C-GNN, the similarity measurement between points is shown in equation (13).

$$s^{(l)}(v,u) = 1 - D^{(l)}(v,u) \quad (13)$$

C-GNN evaluates the matching degree between nodes separately at each layer. After the matching evaluation is completed, attention should be paid to nodes with strong correlation. However, as the training continues, the data of the nodes may change. To address this problem, C-GNN integrates the reinforcement learning part and dynamically adjusts the node aggregation degree in each context, which is expressed as $p_r^{(l)} \in [0,1]$. In a specific connected environment, the model uses the primary sampling method to aggregate nodes. If the total distance between nodes in a certain connection decrease, it indicates that the center and adjacent nodes are increasingly matching. At this point, the $p_r^{(l)}$ value should be increased to collect more data, in order to more accurately determine the type of center node [22-23]. This model utilizes the integration of various connections to refresh the embedding of the central node. There are two main steps. The first step is to perform internal integration on the central node within each connection, as shown in equation (14).

$$h_{v,r}^{(l)} = AGG_r^{(l)}(\{h_u^{(l-1)}, u \in N(v), e_{u,v} \in E_r\}) \quad (14)$$

In equation (14), the embedding vector of v is $h_{v,r}^{(l)}$ under the l -layer relationship r. After completing this step, there is a series of embedding vectors. Then, the further integration is carried out based on the weight $p_r^{(l)}$, as shown in equation (15).

$$h_{v,r}^{(l)} = \sigma(h_v^{(l-1)} + AGG^{(l)}(\{p_r^{(l)}, h_{v,r}^{(l)}\}|_{r=1}^{R})) \quad (15)$$

In equation (15), $h_{v,r}^{(l)}$ is the central node update vector passing through layer L. The FLAG strategy increases the stability and accuracy of the model by injecting gradient-based perturbations into node attributes. This involves creating aggressive data points and incorporating them for training, as shown in equation (16).

$$\min E_{(x,y)\sim D}\left[\max_{\|\delta\| p \le \epsilon} L(f_\theta(x+\delta), y)\right] \quad (16)$$

In equation (16), D represents the dispersion of node data, X is the node attribute, and Y is the node label. L is the solving objective. δ is the generated perturbation group. $\|\cdot\|_p$ represents distance measurement of l_p. Equation (16) introduces parameters that control the injected node attribute perturbations. This parameter selection balances the weight between enhancing the model's generalization ability and maintaining the stability of the training process. This problem requires searching for aggressive sample sets to maximize internal losses, while minimizing the losses of the model on this sample set. The projection gradient descent is used to limit the l_p specification, as shown in equation (17).

$$\delta_{t+1} = \prod_{\|\delta\|_{p\le\epsilon}} (\delta_t + a \cdot sign(\nabla_\delta L(f_\theta(x+\delta_t), y))) \quad (17)$$

In equation (17), the projection gradient descent strategy uses continuous iterations, which generate aggressive interference $\delta_{1:M}$ after M rounds of iterations. Figure. 6 shows the general process of this method in node classification. Finally, the sample set and its aggressive interference are integrated into C-GNN as the entry node.

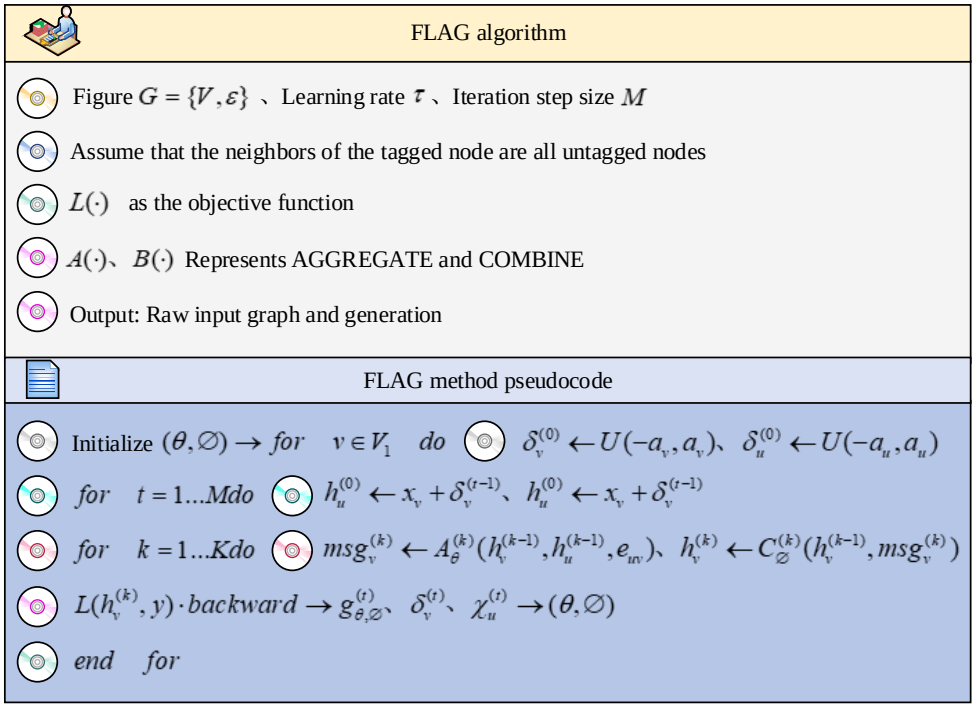


Figure 6: FLAG pseudo-code of the algorithm

4 Malicious e-commerce users' identification and analysis based on user rating behavior and GNN

This experiment first compares the effectiveness and robustness of user recognition among three sample sets. Next, to test the enhancing effect of GNR metrics in reputation ranking technology, F1 and Recall of four algorithms are compared. Finally, the designed method is validated on the fraud detection model of C-GNN. The results show that the model improves the accuracy and robustness.

4.1 Comparative analysis of user identification effectiveness and robustness using adversarial data augmentation

This research experiment used three sample sets, namely Netflix, Movielens2, and Movielens_100, mainly recording users' ratings of movies. The Netflix dataset contains user ratings of movies on a 10-point scale. The Movielens2 dataset contains a 5-point rating system for movies by users. The Movielens_100 dataset contains a 5-point rating of movies by users. The preprocessing process of the model first collects the user rating data for movies from these three datasets (Netflix, Movielens2, Movielens_100) by removing outliers and missing values. Then, the users and items are used as nodes and ratings are used as weights of edges to extract features from user rating behavior. The GNN model consists of 2-5 graph convolutional layers. The hidden units of each layer are set according to the model complexity and the size of the dataset. The activation function is ReLU. The study selects recall rate, F1 value, and Area Under Curve (AUC) as the measurement indicators for the model. The recall can measure the proportion of actual fraudulent users identified by the model. The F1 value can comprehensively consider the accuracy and recall of the model as a comprehensive evaluation indicator. AUC can quantify the classification performance of a model under different thresholds. The Netflix used a 10-point scale, while the other two used a 5-point scale. This study first identified 50 malicious users, and then changed the L value to calculate the recall rates of different strategies to evaluate accuracy. The ideal fraud detection method should accurately identify malicious users in most scenarios and maintain high accuracy even when data is scarce. This experiment examines the performance of four strategies, RPRD, DR, IGR, and IBR, on three databases, as shown in Figure. 7. In Figure. 7, L represents the number of selected users. Different sub-graphs represent different recall rates for malicious user detection.

In the first three sub-graphs, RPRD and IBR performed better in identifying extremely malicious users. In the last three sub-graphs, RPRD was significantly ahead in detecting random malicious users. Especially in sub-graph 7 (e), RPRD outperformed DR, IGR, and IBR even in the context of big data. Preliminary experiments showed that RPRD was particularly effective in identifying small-scale malicious users, especially for random malicious users. Further experiments observe the changes in recall rates of malicious users in different methods, as shown in Figure. 8.

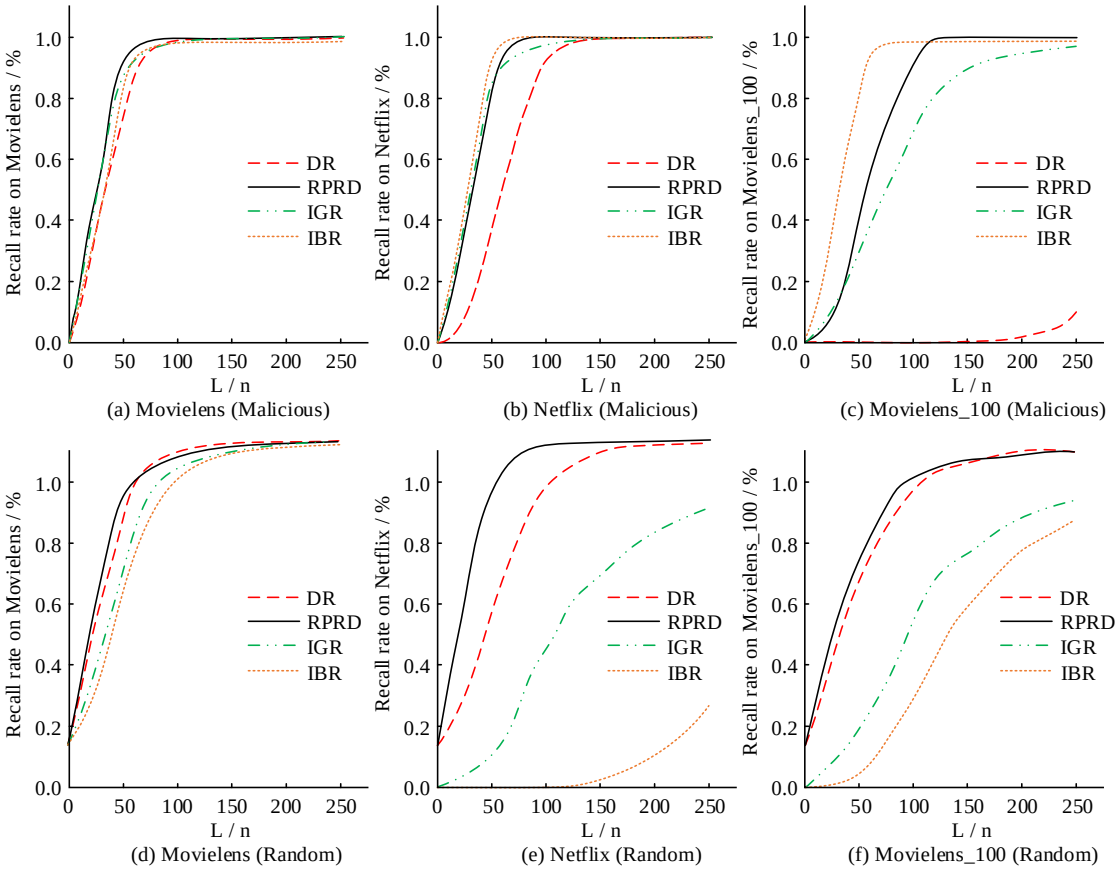


Figure 7: Comparison effect of user recognition

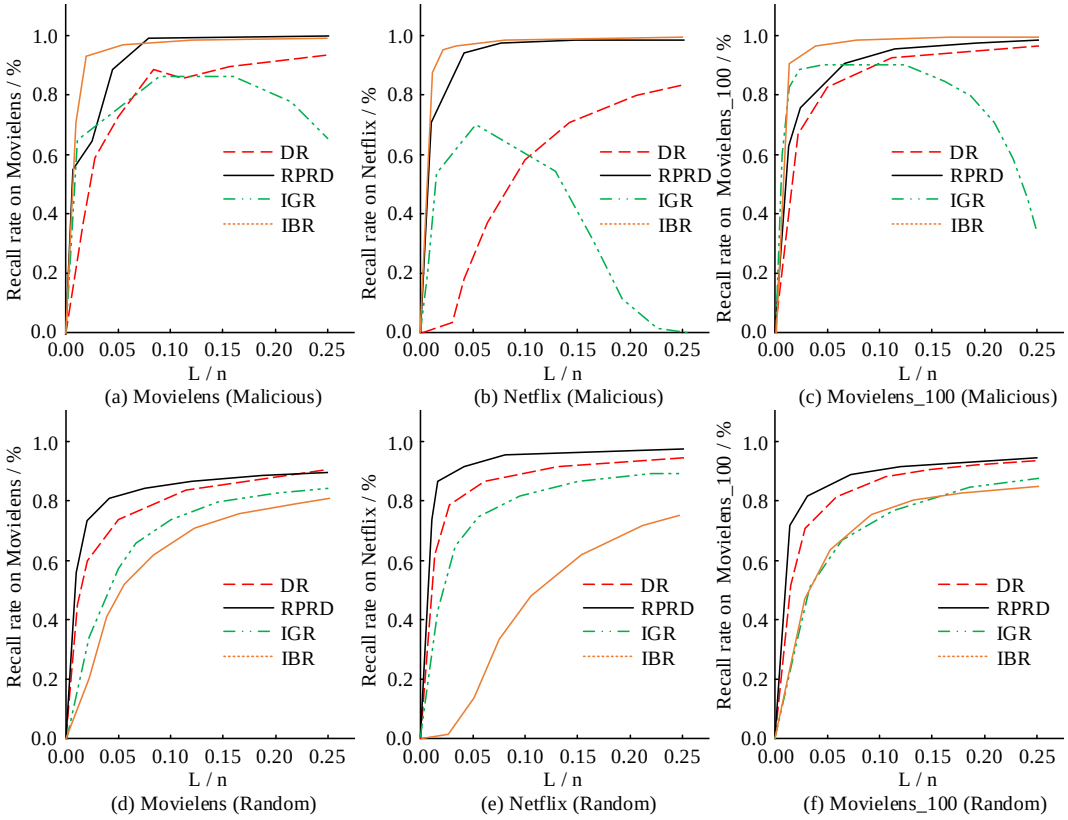


Figure 8: Variation of recall rate by different methods

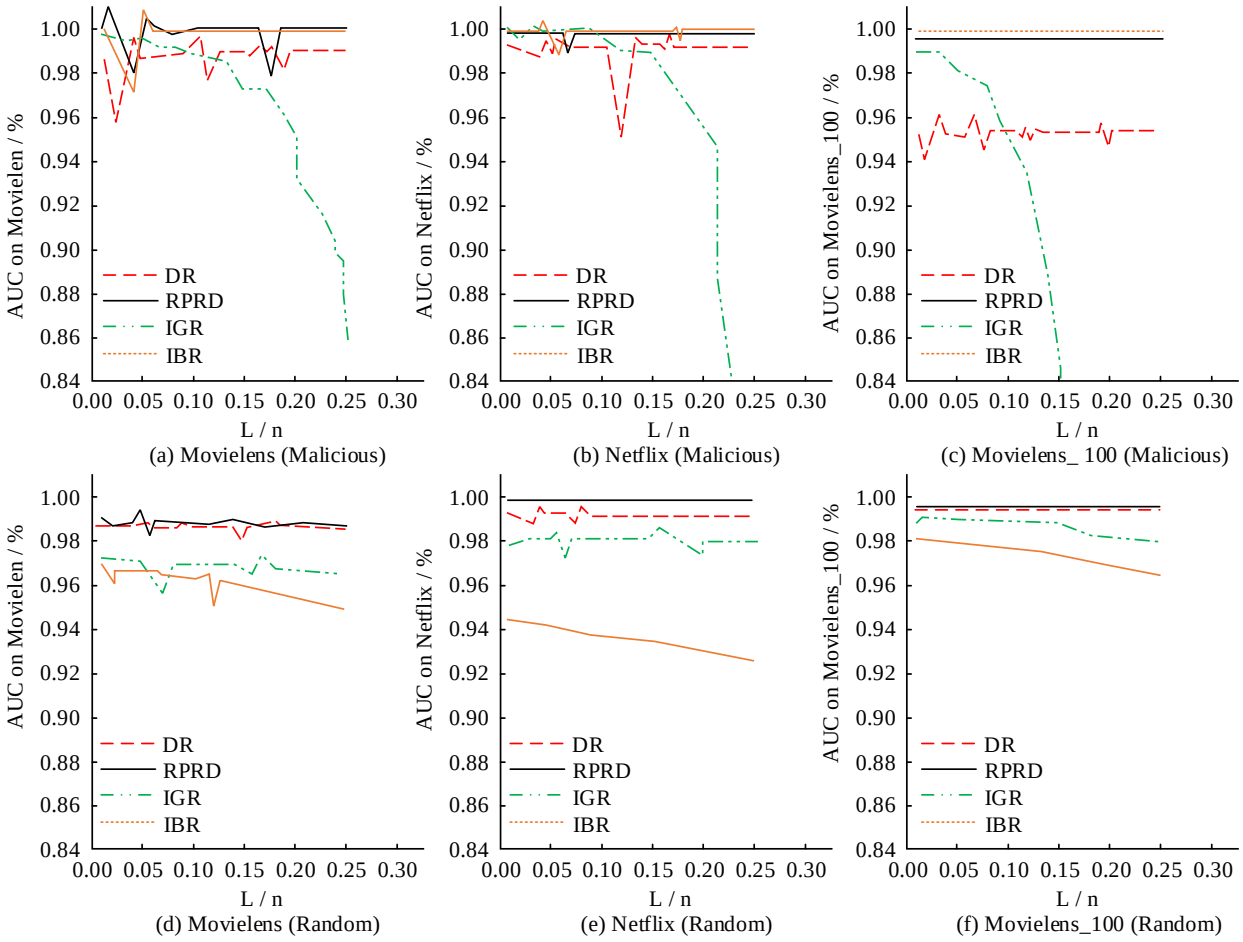


Figure 9: Comparative robustness of user recognition

In the three sub-graphs (a-c) of Figure. 8, RPRD and several other methods showed good results in detecting malicious users. Especially, when the proportion of fraudulent users was 0.035, RPRD and IBR performed well, with a recall value of 0.79. However, after the proportion of malicious users increased, the performance of the IGR method began to decline, reaching only 0.89, indicating that it may have shortcomings in large-scale malicious user attacks. This indicates that introducing RPRD strategy can effectively distinguish between real users and malicious users. Detecting malicious users is effective, especially when the proportion of malicious users is very small. For random malicious users, RPRD always led in the last three sub-graphs (d-f) of Figure. 7, which had significant advantages over DR, IBR, and IGR, especially in identifying random malicious users. RPRD used specific algorithmic equations to distinguish user rating patterns, avoiding algorithmic issues when there were a large number of malicious users. Next, the robustness analysis of each method is conducted, as shown in Figure. 9.

From Figure. 9, as malicious users increased, RPRD remained at 0.98 to 1, significantly leading other models. Even when dealing with millions of ratings, the AUC was still maintained above 0.990, showing strong stability. After incorporating information entropy and probability difference, RPRD effectively reflected the difference in ratings between normal and malicious users, surpassing DR, IGR, and IBR, especially in identifying random malicious users. The core concept of this method is based on fundamental assumptions, utilizing specific equation characteristics to ensure the accuracy of user classification and avoid algorithm failure caused by a large number of malicious users. RPRD also efficiently utilized a triplet structure to store data, improving operational and spatial efficiency. It has been proven that RPRD exhibits better stability and robustness, which can effectively resist malicious user attacks.

4.2 Comparative analysis of the performance of GNR metrics model using adversarial data augmentation

This study tests the complementary ability of GNR metrics to reputation ranking techniques. The study uses a fixed number of 50 malicious users to observe the changes in recall rate. Each experiment is independently repeated 100 times. The comparative effect based on the GNR metrics is shown in Figure. 10.

In Figure. 10, the accuracy of the DR, IGR, and IBR methods in identifying malicious users improved after applying GNR metrics. Specifically, the peak user count of DR method on Movielens and Netflix data decreased from 95-100 to 55-60. Especially in

Movielens_100, the accuracy increased from 0.07 to 0.88, resulting in an overall performance improvement of 9.40%. Although IGR and IBR already had strong detection capabilities, they also slightly improved after applying GNR, with the accuracy increasing by 2.89% and 2.54%, respectively. These indicate that the GNR significantly enhances the ability to measure score distribution differences and identify malicious users. The last three sub-graphs (d-f) of Figure. 10 showed that the efficiency of these three methods in identifying random malicious users increased by 5.52%, 17.12%, and 32.24%, respectively, proving that GNR enhanced the accuracy of these methods. The next step is to conduct a comparative robustness analysis based on the GNR, as shown in Figure. 11.

In Figure. 11, after applying the GNR universal indicators, the baseline method was enhanced in detecting both extreme and random malicious users, especially the IGR and DR methods. The IGR method also solved the problem caused by the increase in fraudulent users. It can be proven that these enhanced methods still maintain higher robustness and accuracy in a large number of malicious attacks.

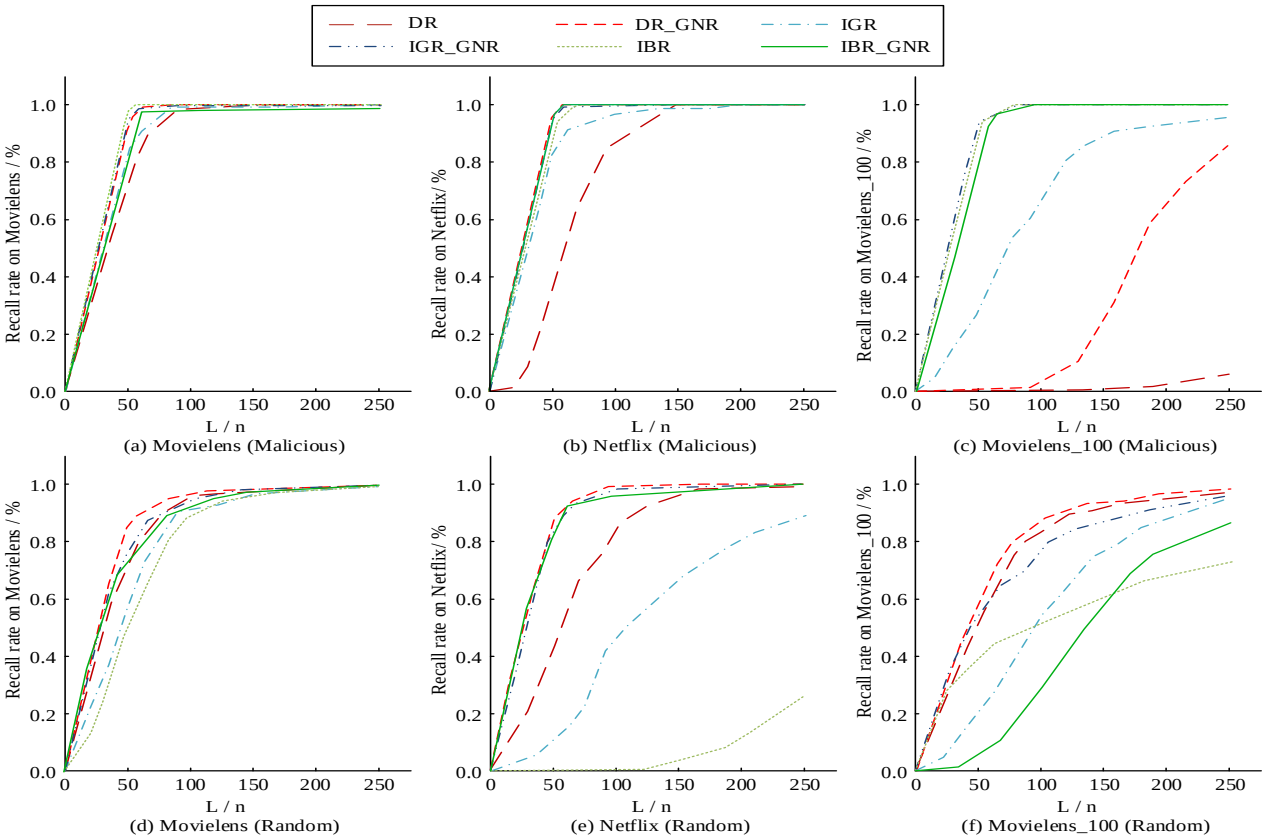


Figure 10: Comparison effect based on GNR

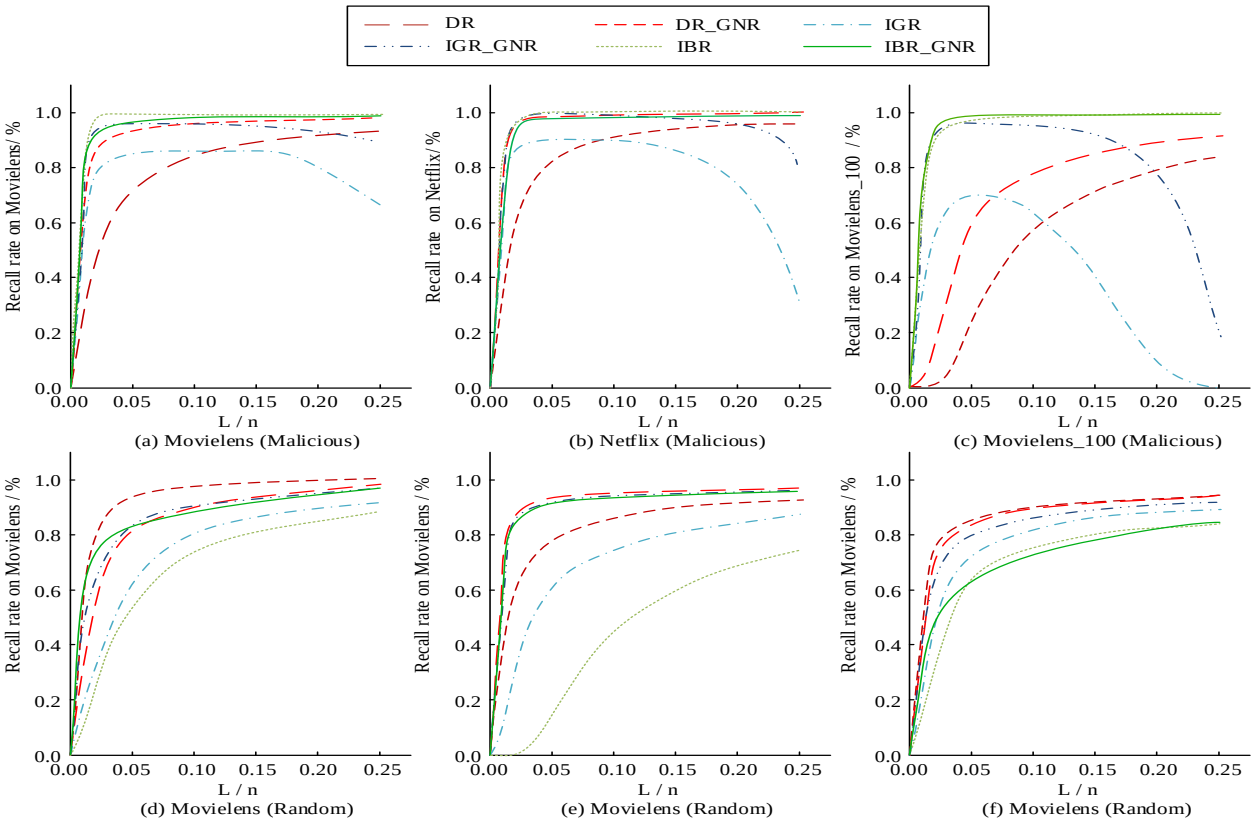


Figure 11: Comparative robustness analysis based on GNR index model

4.3 Comparative performance analysis of C-GNN fraud detection models augmented with adversarial data

This study introduces the FLAG method to the C-GNN model on the C-Yelp and C-Amazon datasets to observe changes in accuracy. In Figure. 12, the accuracy of C-GNN significantly improved using the FLAG method, especially when the training data was limited. For example, when the training set was 5%, the recall and F1 improved by 1.20% and 1.34%. When the training set was 30%, the improvement was relatively small. When the training data increased to 40%, the accuracy improved significantly again. This proves that the FLAG method effectively enhances the performance of the C-GNN model.

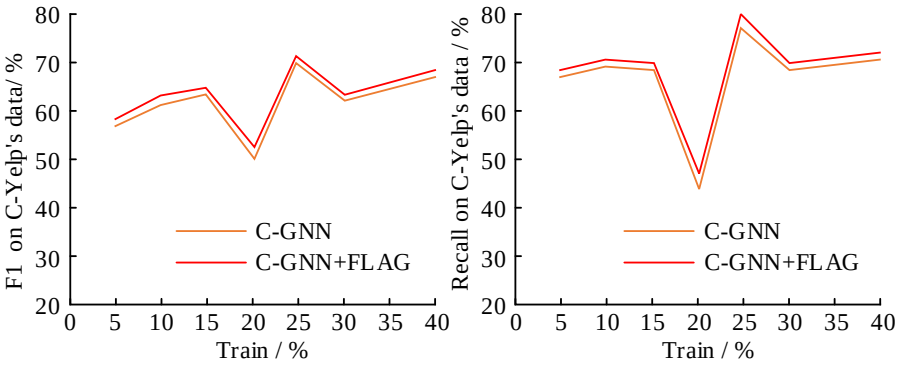
To further explore the improvement of the FLAG method on the C-GNN model, this study specifically introduces the AUC to quantitatively describe how the accuracy of the model changes under different training conditions. The detailed experimental results are shown in Figure. 13. When the FLAG method was applied to the C-GNN, the AUC significantly increased. Especially on the C-Yelp dataset and C-Amazon dataset, the growth range of the ACU was 0.06%-1.54% and 0.19%-1.14%, respectively. This indicates the effectiveness of the FLAG method in improving model performance and robustness. These data clearly demonstrate that the FLAG method effectively enhances the model's classification ability and improves its accuracy in detecting fraudulent users.

As shown in Table 2, the standard deviation and confidence interval of different methods tested on different datasets are analyzed.

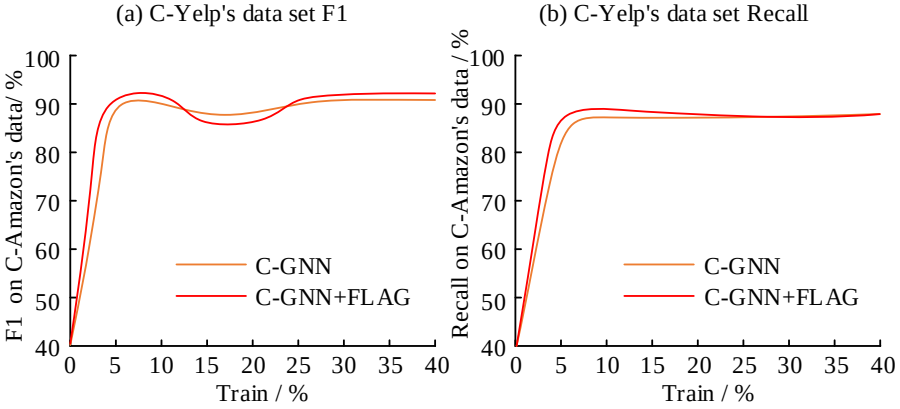
In Table 2, the confidence intervals for comparing different datasets and methods used for research are 95% or higher. The standard deviation used is between 0.01-0.02.

To compare the effectiveness of different methods for identifying malicious fraudulent behavior among e-commerce users, ML algorithm, behavioral biometric identification, and Multi-factor Authentication method (MFA) are compared with existing methods. The recognition performance of the algorithm is obtained using the dataset Movielens2, as shown in Table 3.

From Table 3, among the four methods, the research method achieved the highest recall rate of 91.2% in the Movielens2 dataset. Compared with the lowest ML algorithm, its recall rate increased by 3.6%. At the same time, the F1 and AUC values of the model showed good performance, with the F1 value increasing by 3.8% compared with the ML algorithm. The AUC value increased by 3.2% compared with the ML algorithm. When comparing the accuracy and specificity of different models, this study uses more accurate and specific methods, and processed each segment with better processing speed. The improved GNN model has better data processing performance and the ability to identify malicious fraudulent behavior of e-commerce users.

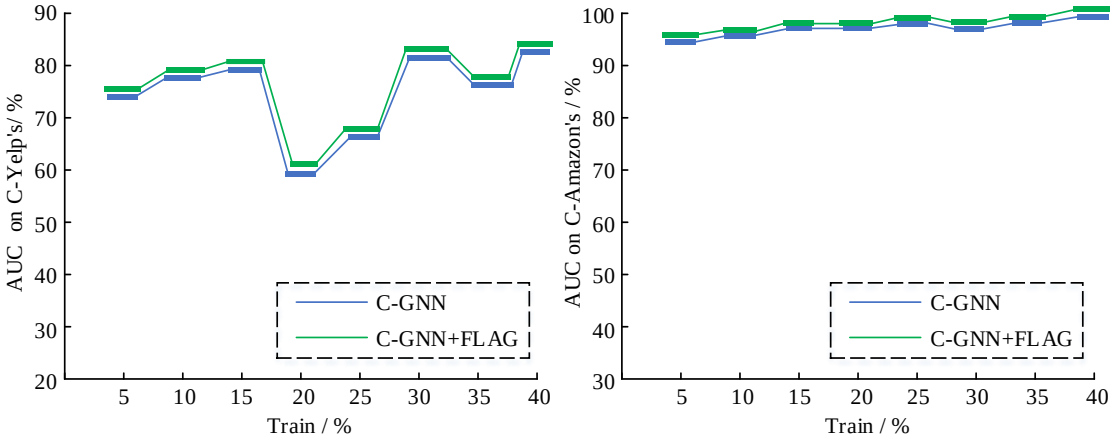


(a) C-Yelp's data set F1 (b) C-Yelp's data set Recall



(a) C-Amazon's data set F1 (d) C-Amazon's data set Recall

Figure 12: Comparison F1 and Recall of C-GNN model before and after FLAG method



(a) C-Yelp's AUC comparison (b) C-Amazon's AUC comparison

Figure 13: AUC of C-GNN model before and after FLAG method

Table 2: Confidence level and standard deviation of different methods under different data sets

Dataset	Method	Recall (%)	F1 (%)	AUC	Confidence Interval (95%)	Standard Deviation
Netflix	RPRD	0.79	0.85	0.96	(0.78,0.80)	0.01
	DR	0.75	0.81	0.94	(0.74,0.76)	0.02
	IGR	0.77	0.83	0.95	(0.76,0.78)	0.01
	IBR	0.76	0.82	0.95	(0.75,0.77)	0.02
Movielens2	RPRD	0.81	0.87	0.97	(0.80,0.82)	0.01
	DR	0.78	0.84	0.96	(0.77,0.79)	0.02
	IGR	0.8	0.86	0.97	(0.79,0.81)	0.01
	IBR	0.79	0.85	0.96	(0.78,0.80)	0.02

Movielens_100	RPRD	0.82	0.88	0.98	(0.81,0.83)	0.01
	DR	0.79	0.85	0.97	(0.78,0.80)	0.02
	IGR	0.81	0.87	0.98	(0.80,0.82)	0.01
	IBR	0.8	0.86	0.97	(0.79,0.81)	0.02

Table 3: Comparison of recognition performance of different methods

Method	Recall (%)	F1 (%)	AUC	Accuracy (%)	Idiosyncrasy (%)	Number of graph nodes processed per second(nps)
ML	87.6	88.6	92.4	86.8	83.4	68
Behavioral biometric recognition	89.5	89.7	93.5	89.4	84.6	73
MFA	88.4	90.3	92.1	91.3	92.4	83
Improved GNN	91.2	92.4	95.6	94.8	95.6	103

5 Discussion

The study compares the model used with the current state-of-the-art methods. The current research uses a method that combines user rating behavior analysis with graphical neural networks and adversarial data augmentation to better identify user fraudulent behavior. For example, the recall of the current method was 0.79 at 0.035. This indicates that the method has good performance, possibly because it can analyze user rating behavior in more detail and includes the GNN framework. Meanwhile, the stability of the research method in model comparison can exceeded 0.990. This indicates that the research method also has model stability and good reliability when dealing with large amounts of malicious user data. After adding graphical neural network to the model, the peak number of users in the model was significantly reduced and the overall performance was improved by 9.40%. User interactions and ratings can be effectively captured by adding GNNR, but the traditional SOTA method cannot capture the details of user interaction ratings. This study integrates adversarial data augmentation technology. In the entire study using this model, the accuracy of the IGR and the IBR improved by 2.89% and 2.54%, respectively. This indicates the need to enhance the application of the model by integrating the data network during model training and enhancement.

When comparing the research method with other advanced methods such as machine learning algorithms, behavioral biometrics and multi-factor authentication algorithms, the recall of the research method in the Movielens2 dataset reached a maximum of 91.2%, which was 3.6% higher than that of the machine learning algorithm. The F1 value and the AUC were also 3.8% and 3.2% higher than that of the ML algorithm, respectively. This may be due to the inclusion of an authentication module in the research method.

Although the currently used method has better application performance, there are still some drawbacks in the research method. For example, in some complex and diverse fraudulent behaviors, the performance of the research method may decrease, which may be due to the limitations of GNN in capturing subtle data changes in complex environments. For this reason, more complex environments will be explored in subsequent studies, while continuously improving the performance of the model to cope with complex environments to adapt to more diverse and changing fraud strategies. Meanwhile, the computational efficiency of the method used in the study still needs to be improved, and the model performance is also enhanced by reducing the computational complexity of the model in subsequent studies.

For extreme malicious users, adversarial data augmentation technology can improve the model's recognition accuracy by simulating user fraudulent behavior and helping the model learn how to recognize and defend against these extreme malicious behaviors. In the case of data scarcity, adversarial data augmentation enhances the generalization ability of the model by generating adversarial samples to simulate malicious user behavior. Faced with large-scale malicious user attacks, adversarial data augmentation improves the stability and accuracy of the model by continuously introducing perturbations during the training process, so that the model can adapt to and recognize malicious behaviors in large-scale attacks. Adversarial data augmentation can also help the model adapt to complex and changing fraud strategies, and improve its ability to recognize emerging fraudulent behaviors by learning more malicious behavior patterns. When the graph layers of the convolutional network are large, the computational complexity of the whole model is elevated. During the model training process, adversarial samples are incorporated into the model training, which leads to an increase in the computational complexity of the model. Secondly, as the model dataset increases, the size and complexity of the graph also increase, resulting in a significant increase in training and inference time for GNN models. In the given experiment, the training time of the model was not explicitly given, but it can be expected that as the size of the dataset increases, the training time will significantly increase. Therefore, to reduce the complexity of the

model in subsequent research, it is possible to train the model by reducing the dataset used and using lightweight models.

GNN-based methods have better advantages in handling complex relationships, large-scale data, robustness, and feature learning. This includes better model performance and data processing capabilities, such as higher performance testing of model accuracy, F1 score, and specificity, which makes GNN based methods superior to non-GNN-based techniques. These advantages make GNN particularly suitable for malicious user identification tasks in the e-commerce field, providing more accurate, efficient, and adaptable solutions.

In summary, compared with current state-of-the-art methods, the method used in the study shows significant improvement in detecting fraudulent user evaluations in e-commerce. By combining user evaluation behavior analysis with GNN and adversarial data augmentation, the practical application of the model can be significantly improved.

6 Conclusion

Malicious users in the e-commerce industry forging ratings have become a major pain point, disrupting the purchasing decisions of real consumers and weakening the credibility of rating systems. The research mainly proposed a new reputation strategy that combined RPRD, which effectively distinguished between real users and malicious users. The study further enhanced the effectiveness of the detection method by introducing the GNR and measuring the differences in score distribution through Gini coefficient and range. The study also adopted adversarial data augmentation technology to improve the robustness and accuracy of GNN model in data scarcity situations. The research innovatively combined multiple advanced technologies and proposed a comprehensive solution to address the malicious user identification in e-commerce. This provides an effective solution for improving the integrity and health of the e-commerce environment, which has important theoretical and practical value. The experiment showed that the RPRD strategy performed well, with a recall value of 0.79 when the proportion of fraudulent users was 0.035, which significantly surpassed other conventional methods. Despite facing a large number of malicious users, the stability of the RPRD strategy was still better, reaching over 0.990. When the GNR was introduced into the research method, existing fraud detection methods performed better. Taking Movielens and Netflix as datasets for testing, the DR method showed a particularly significant reduction in peak user count, with an overall performance improvement of 9.40%. The IGR and IBR methods also made progress, with accuracy improved by 2.89% and 2.54%, respectively. In addition, the FLAG technology further improved the performance of the C-GNN model. In situations where data was scarce, the recall value and F1 score of the C-GNN model increased by 1.20% and 1.34%, respectively. This improvement was also validated on the C-Yelp and C-Amazon datasets. The results indicate that this study provides an effective method for detecting user evaluation fraud in e-commerce. However, despite the positive results of this study, there are still some potential shortcomings. For example, the method used in this study may experience a decrease in effectiveness under certain complex attack strategies. Further exploration is necessary to address potential new fraud strategies that have not yet emerged. Meanwhile, the verification will also be conducted on more different types of e-commerce platforms in the future, analyzing the differences in user behavior on different platforms, and further verifying the widespread applicability and robustness of the methods. Future research will also explore how to optimize the combination of multiple strategies and technologies, improve computational efficiency and resource management, and ensure high operability in practical applications. How to deal with new complex fraud strategies will be further explored to enhance the adaptability and robustness of the model. To address the applicability of the model in complex fraud scenarios and cross-platform, model applicability can be improved by combining multiple data sources such as user ratings, comments, and behavioral logs to capture more comprehensive user behavioral features in subsequent research. Secondly, more advanced GNN architectures will be developed in subsequent research, such as GNNs with enhanced attention mechanisms to better capture complex interactions between users and fraud patterns.

Conflict of interest

The authors declare that they have no conflicts of interest.

References

[1] Zhang XF, Liu HJ, Chen XY, Zhong JB, Wang D. A novel hybrid deep recommendation system to differentiate user's preference and item's attractiveness. Information Sciences, 2020, 519(1):306-316. DOI: 10.1016/j.ins.2020.01.044.

[2] Zhu Y, Lu H, Qiu P, Shi K, Chambua J, Niu Z. Heterogeneous teaching evaluation network based offline course recommendation with graph learning and tensor factorization – Science Direct. Neurocomputing, 2020, 415(20):84-95. DOI: 10.1016/j.neucom.2020.07.064

[3] Mishra, A., Abhishek Kumar GARG, Honey PURWAR, Pushpendra RANA and Huchang LIAO. An Extended Intuitionistic Fuzzy Multi-Attributive Border Approximation Area Comparison Approach for Smartphone Selection Using Discrimination Measures. Informatica, 2020,32(1):119-143. DOI:10.15388/20-infor430

[4] Khan Z Y, Niu Z, Yousif A. Joint Deep Recommendation Model Exploiting Reviews and Metadata Information. Neurocomputing, 2020, 402(18):256-265. DOI: 10.1016/j.neucom.2020.03.075

[5] Bi Z, Dou S, Liu Z, Li Y. A recommendations model with multiaspect awareness and hierarchical user-product attention mechanisms. Computer Science and Information Systems, 2020, 17(3):849-865. DOI:10.2298/CSIS190925024B

[6] Ning X, Yac L, Wang X, Benatallah B, Dong M, Zhang S. Rating prediction via generative convolutional neural networks-based regression - ScienceDirect. Pattern Recognition Letters, 2020, 132(4):12-20. DOI: 10.1016/j.patrec.2018.07.028

[7] Wang Z, Xia H, Du B, Chen S, Chun G. Joint Representation Learning with Ratings and Reviews for Recommendation. Neurocomputing, 2020, 425(2/3):181-190. DOI: 10.1016/j.neucom.2020.04.033

[8] Wang C D, Chen Y H, Xi W D, Huang L, Xie G. Cross-Domain Explicit–Implicit-Mixed Collaborative Filtering Neural Network. IEEE transactions on systems, man, and cybernetics. Systems, 2022, 52(11):6983-6997. DOI:10.1109/tsmc.2021.3129261

[9] Guo Z, Wang H. A Deep Graph Neural Network-Based Mechanism for Social Recommendations. IEEE Transactions on Industrial Informatics, 2020, 11(4):2776-2783. DOI:10.1109/TII.2020.2986316

[10] Liu H, Wang Y, Peng Q, Wu Q, Gan L, Pan L, Jiao P. Hybrid neural recommendation with joint deep representation learning of ratings and reviews. Neurocomputing, 2020, 374(21):77-85. DOI: 10.1016/j.neucom.2019.09.052.

[11] Da'U A, Salim N, Idris R. Multi-level attentive deep user-item representation learning for recommendation system. Neurocomputing, 2021, 433(14):119-130. DOI: 10.1016/j.neucom.2020.12.043.

[12] Tang J, Zhang X, Zhang M, Wu X, Jiang M. A neural joint model for rating prediction recommendation. Journal of Computational Methods in Sciences and Engineering, 2020, 20(1):1-16. DOI:10.3233/JCM-204226.

[13] Wang J, Liu L. A multi-attention deep neural network model base on embedding and matrix factorization for recommendation. International Journal of Cognitive Computing in Engineering, 2020, 1(8):70-77. DOI: 10.1016/j.ijcce.2020.11.002.

[14] Shi L, Wu W, Guo W, Hu W, Chen J, Zheng W, He L. SENGR: Sentiment-Enhanced Neural Graph Recommender. Information Sciences: An International Journal, 2022, 589(12):655-669. DOI: 10.1016/j.ins.2021.12.120.

[15] Liang Y, Qian T, Yu H. ARTAN: Align Reviews with Topics in Attention Network for Rating Prediction. Neurocomputing, 2020, 403(1):337-347. DOI: 10.1016/j.neucom.2020.04.054.

[16] Li P, Yu H, Luo X, Wu J. LGM-GNN: A local and global aware memory-based graph neural network for fraud detection. IEEE Transactions on Big Data. 2023 6(1):1116-1127. DOI:10.1109/tbdata.2023.3234529

[17] Zhao W, Liu X. Detection of E-Commerce Fraud Review via Self-Paced Graph Contrast Learning. The Computer Journal. 2023 18(11):2054-2065. DOI:10.1093/comjnl/bxad123

[18] Yu J, Wang H, Wang X, Li Z, Qin L, Zhang W, Liao J, Zhang Y. Group-based fraud detection network on e-commerce platforms. InProceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2023 6(6):5463-5475. DOI:10.1145/3580305.3599836

[19] Khan Z, Iltaf N, Afzal H, Abbas H. Enriching Non-Negative Matrix Factorization with Contextual Embeddings for Recommender Systems. Neurocomputing, 2020, 380(7):246-258. DOI: 10.1016/j.neucom.2019.09.080.

[20] Feng L, Cai Y, Wei E, Li J. Graph neural networks with global noise filtering for session-based recommendation. Neurocomputing, 2022, 472(1):113-123. DOI: 10.1016/j.neucom.2021.11.068

[21] Chen Z. Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm. Journal of Computational and Cognitive Engineering, 2022, 1(3):103-108. DOI:10.47852/bonviewjcce149145205514

[22] Wang X, Cheng M, Eaton J. Fake node attacks on graph convolutional networks. Journal of Computational and Cognitive Engineering, 2022, 1(4):165-173. DOI:10.47852/bonviewjcce2202321

[23] Nimrah S, Saifullah S. Context-Free Word Importance Scores for Attacking Neural Networks. Journal of Computational and Cognitive Engineering, 2022, 1(4):187-192. DOI:10.47852/bonviewjcce2202406