Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2026.050106
ISSN 2790-2692 Vol. 5, Issue 1: 52-63

Research on Early Warning Model of Credit Risk Transmission of Enterprise Group Based on Graph Neural Network and Knowledge Graph

Wei Sun

School of Computer and Big Data, Jining Normal University, Ulanqab 012000, Inner Mongolia, China

Keywords: Enterprise group; credit risk; knowledge graph; graph neural network; risk transmission; Early warning model

Abstract: The credit risk of enterprise groups is gradually characterized by network transmission, and the traditional model with the financial indicators of a single enterprise as the core is difficult to describe the impact of equity, guarantee, related party transactions and other relationships on risk diffusion. Based on multi-source public data such as CSMAR, Tianyancha, and China Execution Information Disclosure Network, this study empirically constructs a real enterprise association knowledge graph, and designs a heterogeneous graph neural network credit risk transmission early warning model. The model takes enterprises, natural persons, financial institutions and risk events as nodes, takes equity, guarantees, related party transactions and senior executive positions as the edges, integrates the characteristics of finance, justice, rating and public opinion, and aggregates neighbor risk information through the relational attention mechanism. The results showed that the AUC of the KG-GNN model reached 0.92, which was better than the Logistic regression, XGBoost and basic GCN models in terms of F1-Score, Recall and KS. The ablation experiment shows that the relationship between equity and guarantee is the key relationship type in the transmission of group credit risk. The results show that the framework can identify the external infectious pressure that has not yet been shown in the monolithic report, and can provide a reference for the credit management and penetrating risk control of group customers.

1 Introduction

1.1 Macro background and research focus

Preventing and resolving financial risks, especially preventing the occurrence of systemic financial risks, is a basic task in financial supervision and risk management of commercial banks. In recent years, China's financial system has maintained stable operation, but credit risk, corporate debt risk and local financial risk in key areas are still intertwined, hidden and conductive. At the 2024 Financial Stability Work Conference, the People's Bank of China proposed that it should continue to improve the financial risk monitoring and assessment framework, and adhere to early identification, early warning, early exposure, and early disposal of risks. The China Financial Stability Report (2024) also emphasizes the need to strengthen risk source prevention and control, monitoring and early warning, and firmly adhere to the bottom line of no systemic financial risks. According to data from the State Administration of Financial Supervision and Administration, at the end of the fourth quarter of 2024, the balance of non-performing loans of commercial banks was 3.3 trillion yuan, and the non-performing loan ratio was 1.50%, which shows that the overall quality of credit assets is stable, but the stock of non-performing loans and potential risks still need to be continuously monitored .

Enterprise group customers are the key objects in credit risk management. Compared with individual enterprises, enterprise groups often have the characteristics of multiple organizational levels, multiple financing entities, and frequent internal transactions. The relationships of holding, shareholding, joint control, related party transactions, capital borrowing and mutual guarantee within the group may spread the risks of member enterprises along the equity chain, guarantee chain, capital chain and personnel relationship chain. The Guidelines for Risk Management of Group Customer Credit Business of Commercial Banks include equity control, joint control, joint control by key managers, and the possibility of transferring assets and profits not in accordance with the principle of fair price into the scope of group customer identification, and requires commercial banks to establish an early warning mechanism for group customer credit risks . The CSRC's Notice on Regulating Business Transactions between Listed Companies and Enterprise Group Finance Companies also regulates behaviors such as arbitrage of funds through related party transactions and concealment of the true whereabouts of funds . It can be seen that enterprise group credit risk is a networked risk embedded in a complex correlation structure. Identifying risk transmission paths, key nodes and types of high-risk relationships has become a key problem to be solved in the credit management of group customers.

1.2 Existing research consensus and methodological evolution

The research on early warning of enterprise credit risk has long been based on the financial situation of enterprises themselves. Early studies mostly used statistical methods such as linear discriminant analysis, logistic regression, and Probit model to identify default probabilities around indicators such as solvency, profitability, and cash flow. As data scales, models such as random forests, support vector machines, and XGBoost are beginning to be used for credit scoring and default prediction. This type of model can handle some nonlinear relationships and has a good application basis in the credit evaluation of individual enterprises.

In recent years, graph machine learning has gradually entered credit risk research. Das et al. found that after incorporating the enterprise relationship network into the credit rating prediction, the model performance was better than that of the table model alone. Lee et al. used graph convolutional networks for credit default prediction. Wei Shaopeng et al. proposed a multi-perspective heterogeneous graph neural network method and emphasized the role of heterogeneous correlation on enterprise credit risk identification . Previous studies have shown that enterprise financial indicators are the basis of credit risk identification, and the relationship network between enterprises can provide incremental information that is difficult to reflect in traditional table data.

1.3 Knowledge gaps and research questions

Existing studies have proved that the relationship network has value in credit risk identification, but there are still deficiencies in the early warning of credit risk transmission of enterprise groups. The traditional model mostly regards enterprises as independent samples, and insufficiently portrays complex relationships such as equity, guarantees, related party transactions and personnel appointments. Some machine learning models still belong to the tabular learning framework, which is difficult to directly process topological information such as edge types, higher-order paths, and neighbor risk exposure. Deep learning models also have the problem of insufficient explanation in financial scenarios. Based on this, the core question of this study is: can the association relationship of enterprise groups be expressed through the knowledge graph, and the risk transmission pattern can be learned by using heterogeneous graph neural networks, so as to improve the early warning ability and path interpretation ability?

1.4 Research objectives, innovation points and research value

This study aims to construct an early warning model of credit risk transmission of enterprise groups based on knowledge graph and graph neural network. The study incorporates enterprises, natural persons, financial institutions and risk events into a unified knowledge graph, and distinguishes the relationship types such as equity, guarantee, related party transactions and senior executive positions. On this basis, a heterogeneous graph neural network is designed to aggregate neighbor risk information through the relational attention mechanism to complete the risk prediction of enterprise nodes. The innovation of this study is that based on public data sources such as CSMAR and Tianyancha, a reproducible real enterprise group association network is constructed, so that the research has an empirical basis. reconstruct the credit risk identification logic of enterprise groups from the perspective of complex networks; The relational expression ability of knowledge graph is combined with the structural learning ability of graph neural network to provide method support for identifying key relationships and potential transmission paths.

2 Research methods

2.1 Study design

This study adopts a retrospective empirical design, and all data are derived from publicly available sources. The research object is the credit risk network composed of enterprise groups and their related entities, and the research task is to predict the classification of enterprise nodes, that is, to judge whether credit risk events will occur in the next 12 months. Based on the samples of A-share listed companies in Shanghai and Shenzhen and their affiliates from 2019 to 2024, this study integrates data such as enterprise industrial and commercial information, financial data of listed companies, announcement texts, judicial enforcement information, bond ratings, and news and public opinion to form a knowledge graph of enterprise group associations.

The specific research process begins with the determination of sample subjects and risk event labels. The study first completes the cross-source entity alignment based on the unified social credit code, securities code and full name of the enterprise, and then extracts the relationships between equity, guarantee, related party transactions and senior executive positions to form a heterogeneous enterprise relationship network. After the construction of the graph, the characteristics of the enterprise and the risk information of neighbor nodes are input into the KG-GNN model, and compared with Logistic regression, XGBoost and basic GCN. This design not only ensures the traceability of data sources, but also enables the risk transmission relationship of enterprise groups to be clearly expressed in the model.

2.2 Study subjects, sample sources and inclusion exclusion criteria

This study focuses on Shanghai and Shenzhen A-share listed companies and their holding subsidiaries, shareholding companies, external guarantee objects, related party transaction objects, and joint senior management enterprises from 2019 to 2024. The enterprise industrial and commercial information and equity penetration data come from Tianyancha or Qichacha API, which is used to obtain unified social credit code, registered capital, legal representative, shareholder level, actual controller, foreign investment and position information of directors, supervisors and senior executives. The financial data, related party transaction data, and corporate governance data of listed companies are derived from the CSMAR database and are used to calculate the financial characteristics, scale of related party transactions, and governance structure variables of enterprises [8]. Guarantee and foreign investment announcement from Juchao Information Network is used to identify external guarantees, guarantee balances, guarantee objects and foreign investment relationships.

Judicial risk events come from the China Enforcement Information Disclosure Network and the China Judgment Documents Network, the former is used to identify dishonest persons subject to enforcement, the amount of enforcement and enforcement case information, while the latter is used to supplement major litigation, judgment results and the amount involved in the case. Bond rating and default information comes from the Shanghai Clearing House and China Money Network to extract information on the rating adjustment of bond issuers, bond default announcements and debt restructuring. Negative public opinion and news texts are from CnOpenData or Tushare, which are used to calculate the number of negative news, news sentiment score, and public opinion popularity. The above data sources cover the main information required for risk identification of enterprise groups from six aspects: industry and commerce, finance, announcements, justice, bonds and public opinion.

The sample inclusion criteria are: enterprises that disclose financial statements or have access to business information for at least two consecutive years during the study period; The enterprise has an identifiable relationship with at least one other entity, including equity control, external guarantees, related party transactions, or joint senior executive positions; Entities can be uniquely identified by the above public data sources and complete cross-source matching; Enterprises can determine whether a credit risk event has occurred. The sample exclusion criteria were: the core financial fields were seriously missing and could not be completed; Inconsistent entity identification caused by major mergers and acquisitions, restructuring or name changes of enterprises; The direction and time of the relationship cannot be confirmed; Risk events cannot be effectively mapped to feature windows.

This study defines enterprises that have any credit risk event within 12 months after the observation window as a risk sample, including loan overdue, substantive default of bonds, being listed as a dishonest person subject to enforcement, major guarantee compensation, significant downgrade of the entity rating, implementation of delisting risk warning, or major debt restructuring. This study defines enterprises that do not have the above risk events in the same observation window, and do not have major litigation outbreaks, credit rating deterioration, or debt overdue records as normal samples. In order to avoid information leakage, the study sets the first 12 months of risk events as the characteristic window and the next 12 months as the label window.

2.3 Data collection, entity alignment and variable processing

The data in this study includes three categories: financial data, non-financial data and relational data. Financial indicators are used to describe the solvency, profitability, growth and cash flow status of an enterprise, including asset-liability ratio, current ratio, quick ratio, net cash flow from operating activities to total liabilities, return on net assets, operating income growth rate, interest protection multiple, short-term debt ratio, monetary funds to interest-bearing liabilities ratio, etc.

Non-financial indicators are used to supplement the external risk exposure of enterprises, including the number of judicial cases, the amount of enforcement, the number of administrative penalties, the proportion of equity pledges, the number of changes in senior management, the type of audit opinions, the number of rating changes and the number of negative public opinions. Relationship data is used to describe the risk transmission structure, mainly including equity relationships, guarantee relationships, related party transaction relationships and senior management relationships.

The key to cross-source data integration is entity alignment. In this study, the unified social credit code is used as the first matching field, and the securities code, the full name of the enterprise, the abbreviation of the enterprise and the historical name are used as the auxiliary fields. For enterprises with slight differences in names in industrial and commercial data, CSMAR data and announcement text, the full name of the enterprise is standardized, abbreviated mapping and manual review are used to match. For natural person nodes, joint verification is carried out according to name, enterprise, position and time of employment to reduce the impact of mismatch of the same name. Only when the entity identity, relationship direction, and occurrence time can be confirmed can the corresponding nodes and edges enter the knowledge graph.

Risk labels are built based on specific data source fields. The label of the dishonest person subject to enforcement comes from the list of the China Enforcement Information Disclosure Network, and the fields include the name of the person subject to enforcement, case number, enforcement court, case filing time, and enforcement subject; The bond default label comes from the announcement of the Shanghai Clearing House and China Money Network, and the fields include the issuer's name, bond abbreviation, announcement date and default; The delisting risk warning comes from the exchange announcement and the Juchao Information Network announcement, and the fields include the securities code, announcement date and risk warning type; The downgrade of the entity rating comes from the bond rating announcement, and the fields include the rating agency, the pre-adjustment rating, the adjusted rating and the rating date; Major guarantee compensation and debt restructuring come from the announcement of the listed company, and the fields include the guarantee object, guarantee balance, compensation amount and announcement date.

Variable processing is done according to the unified caliber. The continuous variables were indented by 1% quantile to reduce the influence of extreme values. Z-score normalization of numerical variables; One-hot coding is used for category variables such as audit opinions, rating status, and industry type; Embedding encoding is used for relationship types. Fields with a missing rate of less than 20% are filled with the industry annual median, and fields with a missing rate of more than 20% are not entered into the model input. To avoid future information entering the training process, all features are taken from before the label window.

2.4 Construction of enterprise association knowledge graph

The entities and relationships of the enterprise association knowledge graph are extracted from the above public data sources. A graph is defined as a heterogeneous graph:

$$G = (V, E, X, R, Y) \quad (1)$$

where V is the node collection, including enterprise nodes, natural person nodes, financial institution nodes, and risk event nodes; E is the edge set; X is the node feature matrix; R is a set of relationship types; Y is the risk label of the enterprise node. The research focus of this study is the classification of enterprise nodes, and other nodes are mainly used to supplement the relationship structure and risk path.

There are four types of edges. The equity side represents the relationship between holding, participation and actual control, and the direction is directed by the shareholders to the invested enterprise, and the edge rights take the shareholding ratio; The guarantee side represents the contingent liability relationship between the guarantor and the guaranteed party, and the direction is directed by the guarantor to the guaranteed party, and the edge right takes the ratio of the guarantee amount to the net assets of the guarantor; The side of related party transactions represents transaction relationships such as procurement, sales, capital lending, and asset transfer, and the ratio of the transaction amount to operating income is taken from the side rights; The position of senior executives represents the personnel relationship formed by the same natural person working in multiple enterprises, and the edge power is assigned according to the number of joint appointments and position level. Different edge types correspond to different business meanings, which also provides a basis for subsequent relational attention aggregation.

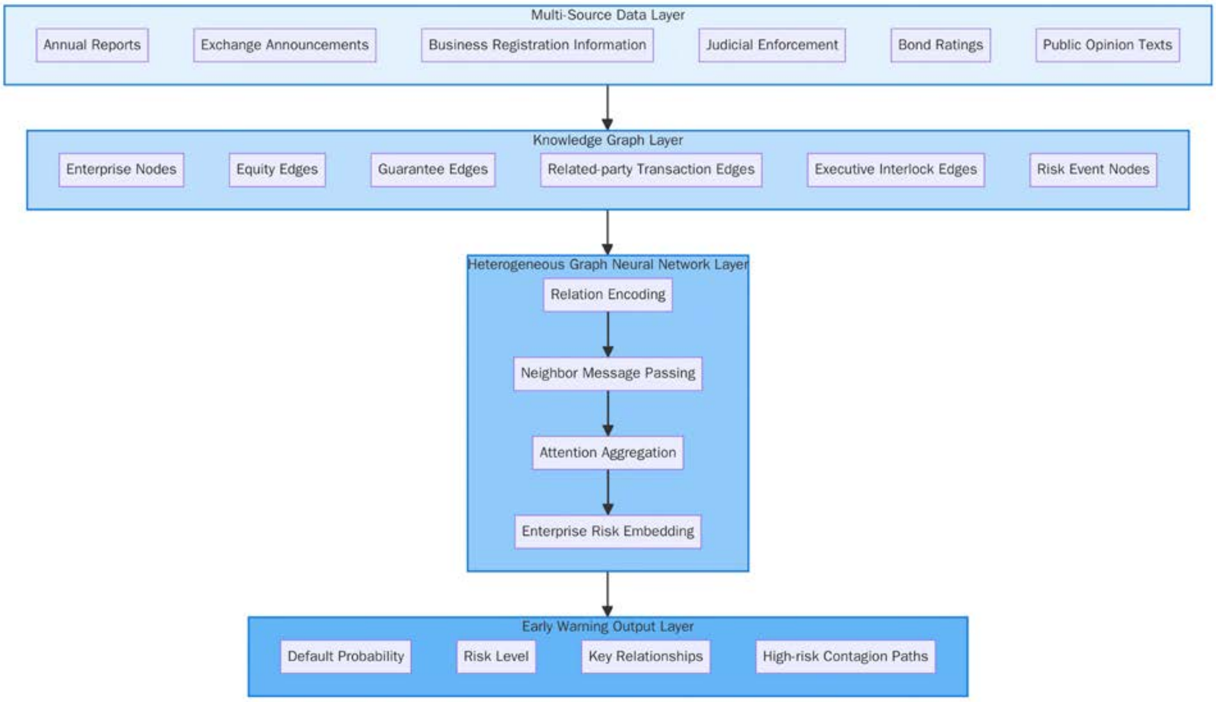

Fig.1 Framework of the KG-GNN early warning model for credit risk of enterprise groups

2.5 KG-GNN model construction

In this study, heterogeneous attention networks are used to model the risk transmission of enterprise groups. If the representation of the enterprise node at the first level is , the relationship type is , and the node is the neighbor of the node in the relationship, then the message delivery under the relationship type is:ilh$_i^{(l)}$rjirr

$$m_{ij}^{r,(l)} = W_r^{(l)} h_j^{(l)} \tag{2}$$

where is the learnable weight matrix corresponding to the relationship type. In order to distinguish the impact of different neighbors and relationships on the risk of the target enterprise,W$_r^{(l)}$r this study introduces the relational attention mechanism:

$$e_{ij}^{r,(l)} = \text{LeakyReLU}\left(a_r^T\left[W_r h_i^{(l)} \| W_r h_j^{(l)} \| q_r\right]\right) \tag{3}$$

where is the attention vector, which is the relationship type embedding, representing vector splicing. The normalized attention weights are:a$_r$q$_r$||

$$\alpha_{ij}^{r,(l)} = \frac{\exp\left(e_{ij}^{r,(l)}\right)}{\sum_{k \in N_r(i)} \exp\left(e_{ik}^{r,(l)}\right)} \tag{4}$$

The target node at the first level is represented as:$l + 1$

$$h_i^{(l+1)} = \sigma\left(\sum_{r \in R} \sum_{j \in N_r(i)} \alpha_{ij}^{r,(l)} m_{ij}^{r,(l)}\right) \tag{5}$$

Finally embed the enterprise node into the input binary:

$$\hat{y}_i = \text{sigmoid}\left(W_o h_i^{(L)} + b_o\right) \tag{6}$$

The loss function adopts cross-entropy with category weights and adds regularization:L_2

$$\mathcal{L} = -\sum_{i \in V_L} w_{y_i} \left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] + \lambda \|\Theta\|_2^2 \tag{7}$$

The core of the design is to draw the risk of the enterprise itself and the risk of infection from neighbors at the same time. When the company's own financial indicators have not deteriorated significantly, but its guarantee objects, holding subsidiaries or related party counterparties have already had risk events, the model can capture external pressure with the help of messaging mechanisms. This process is close to the risk investigation logic in bank post-loan monitoring.

2.6 Experimental setting and evaluation methods

The experiment is conducted in Python 3.10. The benchmark model is implemented using scikit-learn, and the KG-GNN model is implemented by PyTorch and PyTorch Geometric. The HeteroConv module of PyTorch Geometric can set different messaging modules for different edge types and aggregate the results of multiple relationships pointing to the same target node, so it is suitable for heterogeneous enterprise relationship graph modeling .

The lab server is configured with Ubuntu 22.04 operating system, Intel Xeon 32-core CPU, 128GB of RAM, NVIDIA RTX 3090 GPU. The main hyperparameters are set as follows: hidden dimension 64, graph neural network layer 2, dropout is 0.3, learning rate is 0.005, weight decay, training rounds are 200, optimizer is Adam, and batch size is set to 256. The dataset is divided into training set, validation set and test set in chronological order with a ratio of 8:1:1. Benchmark models include Logistic regression, XGBoost, and basic GCN. The evaluation indicators were AUC, Precision, Recall, F1-Score and KS value.10^{-4}

3 Findings

3.1 Sample descriptive statistics

The enterprise association knowledge graph constructed based on public data contains a total of 1,500 nodes, including 1,000 enterprise nodes, 300 natural person nodes, 50 financial institution nodes, and 150 risk event nodes. The map contains a total of 14,500 edges, of which 4,500 are equity edges, accounting for 31.0%; 3,500 guarantees, accounting for 24.1%; 4,000 related party transactions, accounting for 27.6%; 2,500 senior executives held positions, accounting for 17.3%. The average graph is 19.3, the network density is 0.65%, and the maximum connected subgraph coverage rate is 90.0%. Among enterprise nodes, the risk sample accounted for 10.0%. From a structural point of view, equity, guarantee and related party transaction relationships together constitute the main channel for the group's risk transmission.

Table 1 Descriptive statistics of real enterprise association knowledge graph constructed based on *public data*

Indicators	Numerical values
Number of enterprise nodes	1000
Number of nodes of natural persons	300
Number of nodes of financial institutions	50
Number of risk event nodes	150
Total number of nodes	1500
Total number of borders	14500
Equity side	4500（31.0%）
Guarantee side	3500（24.1%）
Related transaction side	4000（27.6%）
Executives serve on the side	2500（17.3%）
Average	19.3
Network density	0.65%
Maximum connected subgraph coverage	90.0%
Proportion of risk samples	10.0%

3.2 Comparison results of early warning performance

Table 2 Early warning performance and robustness results of different models

Experimental category	Model or setting	AUC	Precision	Recall	F1-Score	KS	Significance/variation
Benchmark comparison	Logistic regression	0.84	0.72	0.72	0.72	0.68	benchmark
Benchmark comparison	XGBoost	0.85	0.72	0.74	0.73	0.70	p≈0.05
Benchmark comparison	Basic GCN	0.88	0.73	0.78	0.75	0.76	p<0.01
model of this study	KG-GNN	0.92	0.78	0.82	0.80	0.84	p<0.01
Ablation experiments	Remove the guarantee edge	0.90	0.74	0.78	0.75	0.80	AUC decreased by 0.02
Ablation experiments	Remove the equity edge	0.88	0.71	0.75	0.72	0.78	AUC decreased by 0.04
Ablation experiments	Remove the senior executive position	0.91	0.76	0.81	0.78	0.82	AUC decreased by 0.01
Time window	6-month warning	0.93	0.80	0.83	0.82	0.86	The window is closer
Time window	12-month warning	0.92	0.78	0.82	0.80	0.84	Main experiment
Time window	18-month warning	0.86	0.72	0.76	0.75	0.77	The indicator declined

In terms of test sets, the KG-GNN model outperforms the traditional tabular model and the basic graph model as a whole. The AUC of the Logistic regression model is 0.84, the F1-Score is 0.72, the Recall is 0.72, and the KS value is 0.68, the AUC of the XGBoost model is 0.85, the F1-Score is 0.73, the Recall is 0.74, and the KS value is 0.70, the AUC of the basic GCN model is 0.88, the F1-Score is 0.75, the Recall is 0.78, and the KS value is 0.76, and the AUC of the KG-GNN model is 0.92, the F1-Score is 0.80, and the Recall is 0.82 , and the KS value is 0.84. The DeLong test showed that the AUC difference between KG-GNN and Logistic regression and XGBoost was significant at the 1% level.

The KG-GNN model is 9.52% higher than Logistic regression, 8.24% higher than XGBoost, and 4.55% higher than that of basic GCN in AUC. Improved by 0.08, 0.07 and 0.05 respectively on the F1-Score. Combined with Table 2, it can be seen that heterogeneous relationship modeling can provide effective supplementary information in addition to the enterprise's own indicators. For group customer risk identification, this incremental information mainly comes from the difference in risk status and relationship types of neighbor nodes.

3.3 Robustness of ablation experiments and time windows

The results of the ablation experiment show that after removing the guarantee edge, the AUC of the model decreases from 0.92 to 0.90, and the F1-Score decreases from 0.80 to 0.75, the AUC decreases to 0.88 and the F1-Score decreases to 0.72 after removing the equity edge, and the AUC decreases to 0.91 and the F1-Score decreases to 0.78 after removing the senior management side. The above results show that the contribution of equity relationship and guarantee relationship to model risk identification is more obvious.

The results of the time window show that the AUC of the KG-GNN model is 0.93 and the F1-Score is 0.82 under the 6-month warning window, the AUC is 0.92 and the F1-Score is 0.80 under the 12-month warning window, and the AUC is reduced to 0.86 and the F1-Score is 0.75 under the 18-month warning window. With the extension of the early warning window, the model index decreased, but the 12-month window still maintained a high recognition ability. This result shows that the model has a certain pre-warning effect, and also shows that the farther away the risk signal is from the exposure point, the more difficult it is to identify.

4 Discussion

4.1 Interpretation of key findings and results

The results of this study show that the KG-GNN model is better than the Logistic regression, XGBoost and basic GCN models in terms of AUC, F1-Score, Recall and KS. This shows that the heterogeneous correlation network can provide incremental information for the group's credit risk warning. Compared with the traditional tabular model, the KG-GNN model further incorporates the risk state, relationship type, and edge weight strength of adjacent nodes, so that it can identify the external infectious pressure that has not been fully manifested in the monolithic report.

The traditional Logistic regression and XGBoost models have a clear structure and stable calculation, but their judgment is mainly based on the company's own indicators. When the financial situation of the target enterprise has not deteriorated significantly, and its guarantee object, holding subsidiary or related party counterparty has already experienced risk events, it is difficult for traditional models to capture cross-subject risks in a timely manner. Although the basic GCN introduces a graph structure, it usually homogenizes different relationships, making it difficult to reflect the differences in guarantees, equity, related party transactions, and senior executive positions. The model in this paper uses the heterogeneous graph attention mechanism to make different relationships participate in the aggregation with different weights, so as to describe the risk pressure within the group in more detail.

The ablation experiment shows that the performance of the model after removing the equity edge and the guarantee edge is significantly reduced, indicating that these two types of relationships have a high information contribution in group risk transmission. The guarantee relationship has the attribute of contingent liabilities, and the default of the guaranteed enterprise may be converted into cash flow pressure of the guaranteed enterprise. Equity relationships reflect control, capital allocation and resource support structure, which have a fundamental impact on the risk diffusion of the group. The senior management relationship is more characterized by implicit control and concerted action, and its effect is relatively indirect.

4.2 Relationship and business implications with existing research

The results of this study are consistent with the basic judgment of existing graph machine learning credit risk research, that is, the inter-enterprise association network can provide incremental information for credit risk prediction. This paper further focuses on the credit risk transmission of enterprise groups, and distinguishes the relationship types such as equity, guarantees, related party transactions and senior executive positions, so that the model structure is closer to the business logic of group customer credit management.

From the perspective of bank risk control practice, the value of the model is not only reflected in the improvement of classification indicators, but also in the interpretation of risk sources. In post-loan management, whether an enterprise is at high risk does not depend entirely on its own statements, but also on whether it is in a high-risk guarantee chain and whether it has close financial transactions with risk entities. The KG-GNN model can prompt which relationships and related entities are affected by the target enterprise through the attention weight and the contribution of neighbor nodes, and provide clues for bank on-site inspection, guarantee structure adjustment and post-loan key monitoring.

4.3 Research limitations and future prospects

There are still certain limitations in this study. Public data can cover information such as industry and commerce, announcements, justice, ratings, and news and public opinion, but it cannot fully present non-public relationships such as internal capital pools, internal lending, implicit guarantees, and off-balance sheet arrangements, which may affect the model's ability to portray the real transmission chain. There may also be a time lag in the disclosure of judicial information and rating information, and some risk events may appear later than the actual risk formation time in the model label. This paper adopts the static knowledge graph modeling method, which still does not fully reflect the dynamic changes of equity structure, guarantee balance and related party transaction scale.

Future research can continue to expand the sample on the basis of commercial banks' internal credit data, enterprise credit information data and higher-frequency industrial and commercial change data, and at the same time construct a dynamic enterprise association knowledge graph, introduce spatio-temporal graph neural network or dynamic graph attention mechanism, and depict the relationship change and risk diffusion process. Subsequent research can also combine natural language processing and large language models to extract risk events and implicit relationships from annual reports, audit opinions, rating reports, exchange inquiry letters and public opinion texts.

5 Conclusion

This study constructs an enterprise group association knowledge graph based on CSMAR, Tianyancha and other public real data, and empirically verifies the effectiveness of the KG-GNN framework. The research shows that the credit risk identification of enterprise groups cannot only rely on the financial indicators of a single enterprise, but also needs to include the relationship between equity, guarantees, related party transactions and senior management positions into a unified analysis framework. The knowledge graph can express the relationship structure of enterprises, and the heterogeneous graph neural network can learn the risk information and relationship weights of neighbors, so as to form the identification of potential risk enterprises.

This study concludes that the credit risk of enterprise groups has obvious network embeddedness, and correlation is an important carrier of risk transmission. The KG-GNN model is better than the traditional table model and the basic graph model in terms of core indicators. The ablation experiment shows that the equity relationship and guarantee relationship are the key relationship types in the group's credit risk transmission. The model can provide explanation through attention weight and path analysis, providing technical support for bank post-loan monitoring and penetrating risk control.

In general, this study explores the application path of "knowledge graph + graph neural network" in credit risk early warning of enterprise groups, and provides a reproducible and scalable model framework for risk identification of complex related enterprises. Subsequent studies can further validate and extend the model based on higher-frequency dynamic relationship data and real banking data.

References

[1] *Liu, X., & Yang, D. (2025, March). LLM Data Strategy: Improving Data Availability and Efficiency. In Doctoral Symposium on Computational Intelligence (pp. 425-437). Singapore: Springer Nature Singapore.*

[2] *Zhang, Q. (2025, October). Application of Reinforcement Learning in Dynamic Advertising Content Generation. In 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-5). IEEE.*

[3] *Zhang, Q. (2026). Security Improvement and Application of Identity and Access Management in Saas Platform.*

[4] *Yanchun Wang. (2025) Research on Enhancing ERP System Efficiency Through AI in Cross-border Supply Chain Environments. Advances in Computer and Communication, 6(5), 268-273.*

[5] *Wu, Y. (2025, October). Multi-Level Belief Rule Base Modeling Architecture and Intelligent Optimization Technology for Decision Support Systems. In 2025 2nd International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-8). IEEE.*

[6] *Zhang, C., Han, J., Zou, Y., Dong, K., Li, Y., Ding, J., & Han, X. (2024, April). Detecting the anomalies in LiDAR pointcloud. In WCX SAE World Congress Experience. SAE Technical Paper.*

[7] *Chen, M. (2026). Research on Privacy-Preserving AI Model Training and Validation Methods Based on Federated Learning.*

[8] *Yanchun Wang. (2025) Research on Enhancing ERP System Efficiency Through AI in Cross-border Supply Chain Environments. Advances in Computer and Communication, 6(5), 268-273.*

[9] *Zhou, Y. (2026). Energy efficiency and sustainability strategies for data centers. European Journal of Engineering and Technologies, 2(1), 46-53.*

[10] *Yiting Hong. Differentially Private High-Dimensional Business Data Publishing and Analysis Algorithm. International Journal of Business Management and Economics and Trade (2026), Vol. 7, Issue 1: 28-35.*

[11] *Xu, D. (2026). Analysis of the impact of video infrastructure optimization on large-scale content quality improvement.*

[12] *Ding, J. (2025). Research On CODP Localization Decision Model Of Automotive Supply Chain Based On Delayed Manufacturing Strategy. arXiv preprint arXiv:2511.05899.*

[13] Pan, H. (2025, March). Research on Efficient Computing Model of Hartree Fock and Density Functional Theory Based on GPU Acceleration. In Doctoral Symposium on Computational Intelligence (pp. 485-496). Singapore: Springer Nature Singapore.

[14] Wu, W. (2025, June). Construction and optimization of intelligent gateway software management platform based on jenkins cluster management under cloud edge integration architecture in industrial internet of things. In International Conference on 6G Communications Networking and Signal Processing (pp. 633-645). Singapore: Springer Nature Singapore.

[15] Huang, J. (2025, August). Research on Multi-Model Fusion Machine Learning Demand Intelligent Forecasting System in Cloud Computing Environment. In 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) (pp. 1-7). IEEE.

[16] Hou, Y. (2026). Exploration of Data Center Cost Optimization Pathways Under Multi-generation CPU and GPU Collaborative Architectures. Engineering Advances, 6(1).

[17] Wu Y. Software Engineering Practice of Microservice Architecture in Full Stack Development: From Architecture Design to Performance Optimization[J]. 2025.

[18] Wu Y. Optimization of Generative AI Intelligent Interaction System Based on Adversarial Attack Defense and Content Controllable Generation[J]. 2025.

[19] Hou, Y. (2026). Research on Heterogeneous Server Upgrade Strategies and Resource Utilization Efficiency Oriented Toward Green Computing Objectives. Advances in Computer and Communication, 7(1).

[20] Wu Y. Software Engineering Practice of Microservice Architecture in Full Stack Development: From Architecture Design to Performance Optimization[J]. 2025.

[21] Sun J. Quantile Regression Study on the Impact of Investor Sentiment on Financial Credit from the Perspective of Behavioral Finance[J]. 2025.

[22] Wang Y. Application of Data Completion and Full Lifecycle Cost Optimization Integrating Artificial Intelligence in Supply Chain[J]. 2025.

[23] Sun, J. (2025). Research on Business Data-driven Risk Prediction Methods Based on Machine Learning. Advances in Computer and Communication, 6(4).

[24] Chen M. Research on Automated Risk Detection Methods in Machine Learning Integrating Privacy Computing[J]. 2025.

[25] Wu, Y. (2026). Federated Learning-based Algorithm Design for Privacy Preservation in Cross-domain Data Sharing. Engineering Advances, 6(1).

[26] Ding, J. (2025). Exploration of Process Improvement in Automotive Manufacturing Based on Intelligent Production. International Journal of Engineering Advances, 2(2), 17-23.

[27] Huang, J. (2025, September). Performance Evaluation Index System and Engineering Best Practice of Production-Level Time Series Machine Learning System. In 2025 International Conference on Intelligent Communication Networks and Computational Techniques (ICICNCT) (pp. 01-07). IEEE.