Transactions on Computational and Scientific Methods | Vo. 4, No. 3, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Subgraph-Aware Graph Representation Learning for Collaborative Risk Scoring and Organized Fraud Detection

Kewei Cao
Columbia University, New York, USA
kc3410@columbia.edu

Abstract: This paper addresses the challenges of high-frequency transactions, complex transaction structures, and covert collaborative crimes in anti-money laundering and anti-fraud scenarios. It investigates a joint modeling method for transaction risk scoring and gang identification, aiming to simultaneously characterize transaction-level risk intensity and subgraph-level group associations within a unified framework. First, transaction data is abstracted into a directed transaction graph, defining risk scoring tasks for transaction nodes and gang identification tasks for suspicious transaction communities (or high-risk subgraphs). A unified embedding representation for transaction nodes and transaction subgraphs/communities is achieved through shared representation learning. Then, a graph representation module based on a graph encoder is constructed to aggregate and fuse transaction attributes and neighborhood structure information. Risk scoring heads and gang identification heads are designed separately for the shared embedding, outputting continuous risk scores and gang probability distributions. A multi-task joint objective is introduced to constrain the consistency and stability of the shared representation between the two tasks. This method can more fully utilize multi-hop fund transmission and local subgraph patterns at the transaction-chain (multi-hop) level, reducing risk fragmentation caused by relying solely on single transaction features, and capture potential collaborative structures at the subgraph/community level to assist in risk handling prioritization and clue closure. Comparative experiments verified the comprehensive advantages of the proposed joint modeling framework across multiple evaluation indicators, demonstrating its effectiveness and application value in transaction monitoring and risk governance tasks.

Keywords: Transaction chain modeling; risk scoring; gang identification; graph representation learning

1. Introduction

With the widespread adoption of digital payments and cross-platform financial services, transaction activity is rapidly increasing in frequency, fragmentation, and cross-domain scope, making fund transfer chains longer, more complex, and more difficult to trace[1]. In anti-money laundering and anti-fraud scenarios, risks are no longer reflected in isolated transactions but are hidden in the combined relationships and chain structures of multiple transactions, such as layered transfers, decentralized aggregation, circular repatriation, and multi-account collaboration. These behaviors are characterized by strong correlation, strong temporality, and strong concealment. Single-point rules or static thresholds are insufficient to cover the constantly evolving patterns of crime and easily create long-term contradictions between compliance pressures and business efficiency. Therefore, there is an urgent need for risk analysis methods that can characterize the transaction chain structure and behavioral semantics[2].

In real-world operations, although anti-money laundering and anti-fraud share the same goal-reducing financial risk and losses-they have long suffered from data fragmentation, indicator fragmentation, and process fragmentation. Anti-money laundering focuses more on the source and destination of funds, the rationality of the chain, and interpretable compliance evidence, while anti-fraud emphasizes real-time performance, accuracy, and rapid response to new types of attacks. When two types of risks intertwine within the same ecosystem, isolated modeling can easily lead to inconsistent risk assessments, a lack of interconnected leads, and confusion in handling priorities. This further increases operational costs due to high false positives and amplifies actual losses from underreporting. Adopting a unified transaction link perspective, incorporating risk signals and related structures into the same framework, helps achieve cross-scenario collaborative governance and improves the continuity of overall risk prevention and control[3].

The core requirement of transaction link risk scoring is to provide measurable, sortable, and traceable risk metrics for transactions, transactions, and link segments in complex networks and dynamic environments, supporting tiered handling and resource allocation[4]. Risk scoring not only needs to integrate multi-source features such as amount, frequency, time interval, geography, and device, but also needs to identify abnormal structural patterns and fund transmission logic at the link level to discover seemingly normal but abnormally combined behaviors. Without a link-level understanding, models often can only capture fluctuations in local features, making it difficult to distinguish between normal business growth and risk diffusion, and also difficult to form stable judgments on key nodes and critical paths of risk propagation[5].

The core challenge of gang identification lies in the organization and disguise of coordinated crimes. Gang members disperse risk exposure through role division and multi-hop transfers, and create noise and obscure through cross-account, cross-merchant, and cross-channel collaborative activities, making it difficult for single-account profiles to fully reflect true intentions. Simultaneously, the dynamic nature of gang structures, with member relationships potentially changing rapidly and significant differences in gang size and behavioral intensity, makes traditional clustering methods based on fixed rules or single similarities prone to failure. Separating gang identification from link risk assessment also leads to the problem of incomplete lead loops; identified related groups are difficult to quantify into actionable risk priorities in a timely manner, and high-risk links are difficult to interpret in reverse to explain their underlying collaborative organizational structure[6,7].

Therefore, conducting joint modeling research on transaction link risk scoring and gang identification for anti-money laundering and anti-fraud purposes is of great significance. On the one hand, joint modeling can simultaneously learn individual risk and group collaborative relationships in a unified representation, making risk scoring closer to real fund flow mechanisms and providing more stable early warning evidence at the link level. On the other hand, it can transform group structure information into actionable risk measurements and handling strategies, helping to identify key hub accounts, key intermediary paths, and potential core members, thereby improving the targeting and consistency of risk handling. More importantly, in the context of balancing regulatory compliance and business growth, a joint perspective oriented towards links and groups is expected to reduce operational costs caused by false alarms, enhance adaptability to new models, and provide more systematic and sustainable technical support for risk governance.

2. Problem Formulation

Let transaction data be represented by a directed graph $G = (V, E)$, where V is a set of transactions, and E denotes directed relations between transactions. For any transaction node $v \in V$, its feature vector is denoted as $x_v \in R^d$, and its context subgraph or path segment $S_v \subseteq G$ in the transaction graph is also given. The first task of this study is to perform transaction-level risk scoring. Specifically, learning a scoring function to output a risk score $s_v \in [0,1]$:

$$s_v = f_\theta(x_v, S_v),\ v \in V$$

The second task is to perform gang identification and affiliation modeling for suspicious transaction communities. For any transaction subgraph $s \in S$, based on its neighborhood subgraph $S \subseteq G$ and node features $x_v \in R^{d*}$, a mapping function g_ϕ is learned to output its gang label $c_v \in \{1, \ldots, K\}$:

$$c_S = g_\phi(z_S)$$

To achieve joint modeling of risk scoring and gang identification, a shared representation learning mechanism is introduced, enabling transactions and transaction subgraphs to be represented in the same embedding space. Let the subgraph embedding embedding and node embedding be z_e and z_v, respectively, generated by the shared encoder h_ψ, and used in the headers of the two tasks.

$$z_v = h_\psi(x_v, N_v)$$

3. Key Challenge

In anti-money laundering and anti-fraud transaction scenarios, risk signals often exhibit strong sparsity and concealment, and evolve continuously with strategic game theory. Many high-risk transactions appear plausible at the individual level; anomalies are only revealed when linking multi-hop paths, fund diversion and convergence structures, and time rhythms. Therefore, relying solely on local features easily leads to both high false positives and high false negatives. Furthermore, transaction data is inherently dynamic and heterogeneous; factors such as transaction type, channel, device, and region cause distribution drift, resulting in the same risk pattern manifesting differently in different time windows or business domains. This makes it difficult for models to generalize stably and places higher demands on real-time performance.

The challenges of gang identification lie in the organization and disguise of collaborative behavior, as well as the ambiguity and variability of gang boundaries. Gang members typically create complex interaction structures through role division and multi-account collaboration, diluting identifiable features with noisy transactions, cross-group connections, or short-term node replacements. This makes simple similarity clustering or static community detection insufficient to accurately characterize the true gang structure. Furthermore, risk scoring and gang identification do not have the same goal. The former emphasizes the orderability of risk intensity and the priority of handling, while the latter emphasizes the consistency of group structure and member affiliation. When the two are combined, there is a risk of task conflict and information leakage. If there is a lack of appropriate shared representation and constraint mechanism, it may improve the effectiveness of one party while damaging the stability and interpretability of the other party.

4. Method

The joint modeling of risk scoring and gang identification for transaction links adopts the overall approach of using the transaction network as a carrier, unifying transaction node and transaction subgraph into the same representation space, and then performing transaction-level risk scoring and node-level gang affiliation prediction on the shared representation. Specifically, the model first constructs a transaction graph and extracts local link segments related to the target transaction or target subgraph. Then, a shared encoder fuses and represents the link structure and attribute features. Finally, two lightweight task heads are used to output the risk score and gang distribution. Its overall model architecture is shown in Figure 1.

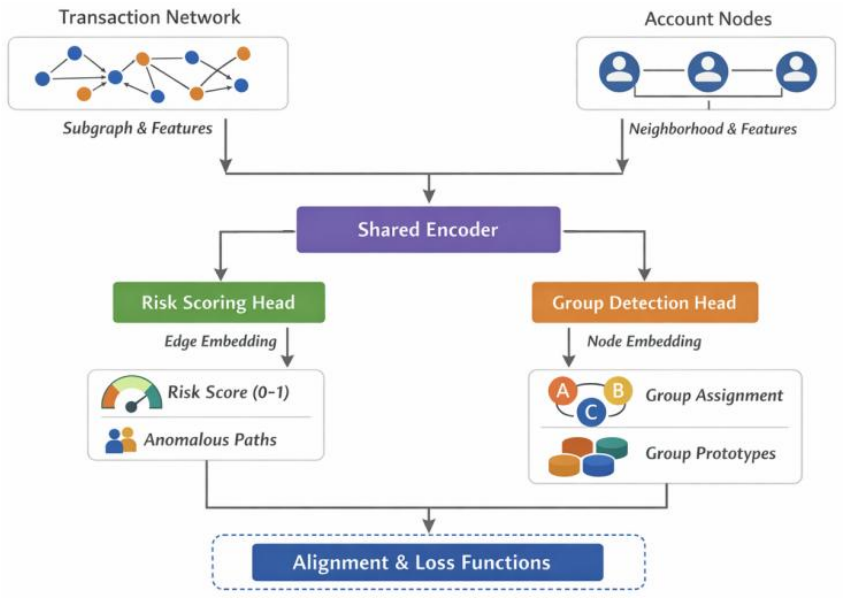

Figure 1. Overall model architecture

To ensure the model simultaneously considers local anomaly clues and global collaborative behavior, the encoder aggregates information through neighborhood message passing and introduces risk heads and gang heads at the output. Simple joint constraints align the two tasks within the same semantic space, avoiding risk fragmentation and gang boundary drift from a single-task perspective.

$$h_v^{(l+1)} = \sigma\left(W_0 h_v^{(l)} + \sum_{u \in N(v)} \alpha_{uv} W_1 h_u^{(l)}\right)$$

In the transaction risk scoring, the representation of each transaction node is composed of the endpoint node representation and the transaction's own attributes, and a normalized risk value is output using a simple scoring function. The edge representation can be obtained by concatenating the embedded endpoint nodes and superimposing a linear mapping of the transaction features. The risk score is mapped to 0 to 1 using Sigmoid, which facilitates sorting and thresholding.

$$z_e = U\left[h_{src(e)}; h_{dst(e)}\right] + V x_e$$

$$s_e = \sigma\left(w^T z_e + b\right)$$

In gang identification, node embeddings are input into a classification head to obtain the gang probability distribution, and further gang-level prototype representations can be obtained to characterize group commonalities, thereby enhancing the stable capture of collaborative crime structures. Gang prediction uses Softmax to output discrete affiliation distributions, and the gang prototype is formed by aggregating the mean of member embeddings, which facilitates subsequent association constraints with risk signals.

$$p_v = softmax\left(Ah_v + c\right)$$

$$g_k = \frac{1}{|V_k|} \sum_{u \in V_k} h_v$$

The optimization objective of the joint modeling adopts weighted multi-task learning, linearly combining the transaction risk scoring loss and gang identification loss, and using a simple regularization term to stabilize

the shared representation, avoiding mutual interference between the two tasks that could lead to representation collapse or excessive bias towards one side. Risk scoring can be represented by binary cross-entropy, and gang identification by multi-class cross-entropy. The overall objective adjusts the importance of the two tasks through weight coefficients, achieving collaborative learning and consistent output within the same model.

$$L = \lambda L_{risk} + (1-\lambda)L_{group} + \beta \|\Theta\|_2^2$$

5. Experimental Results and Analysis

5.1 Dataset Introduction

This paper uses the established open-source Elliptic Bitcoin Transaction Dataset (Elliptic Data Set) as the data source. The dataset is designed for anti-money laundering (AML) research and organizes Bitcoin transactions into a directed temporal transaction graph, enabling reproducible studies on risk characterization and propagation along transaction chains.

Structurally, each node corresponds to a transaction, and each directed edge indicates a flow relationship from an earlier transaction to a subsequent one, forming time-respecting payment chains. The dataset is provided in discrete time steps (time slices), which reflect the dynamic evolution of the transaction network. It contains approximately 203,769 transaction nodes and 234,355 directed edges, and each transaction node is associated with a 166-dimensional feature vector, consisting of transaction-level attributes and aggregated neighborhood statistics. Labels are given in three categories: illicit, licit, and unknown, where unknown denotes unlabeled samples.

Aligned with the joint modeling objective of this paper, Elliptic naturally supports transaction-level risk scoring. Specifically, we interpret risk as a continuous score predicted for each transaction node, where the supervised signal is obtained from the licit/illicit labels and the unknown class is treated as unlabeled data during training. The directed transaction links enable modeling multi-hop risk propagation and aggregation, allowing the model to capture how suspicious behavior may diffuse along temporal payment chains.

In this paper, gang classification does not rely on the real-world "gang" labels provided by the dataset, but rather is a collaborative suspiciousness modeling task based on the transaction graph structure. Specifically, this paper operationally defines a "gang" as a suspicious community or high-risk subgraph in a transaction network that exhibits signs of collaborative behavior; its essence is a type of structural unit characterized by transaction topological relationships. Since the Elliptic dataset only provides transaction-level licit/illicit/unknown labels and not explicit group labels, this paper uses the graph structure itself as the basis for group formation: after constructing the directed graph of transactions, local link segments related to the target transaction or target subgraph are extracted, and connected subgraphs are extracted as candidate groups based on connectivity, thereby expressing the "potential collaborative structure" in the form of a computable subgraph. At the model level, all nodes first fuse link structure and attribute features in the same representation space through a shared encoder. Then, the gang head embeds the nodes into the input classification head to obtain the node-level gang affiliation probability distribution (equation 7), and further forms a group prototype representation (equation 8) by averaging the embeddings of group members, which is used to summarize the common structure and features of the group. At the same time, the risk head outputs a transaction-level risk score (equation 6) on the same shared representation, and imposes consistency constraints on the two tasks through a joint optimization objective (equation 9), so that the risk signal and group affiliation can be learned collaboratively in the same semantic space.

5.2 Experimental setup

The experiment was conducted on a single-machine GPU server with the following hardware configuration: 1 × NVIDIA RTX 4090 with 24GB of VRAM, an Intel Xeon Silver 4314 CPU, 256GB of DDR4 RAM, a 1TB NVMe SSD for the system disk, a 2TB NVMe SSD for the data disk, and Ubuntu 22.04 LTS 64-bit as

the operating system. To ensure stable throughput for link subgraph sampling and batch processing, a fixed number of CPU threads were used for DataLoader and graph sampling during training to avoid the impact of IO jitter on training latency; persistent logging and model checkpoint saving were also enabled to support interruption recovery and reproduction.

In terms of software environment, Python 3.10.13 and Conda were used for dependency management. The deep learning framework was PyTorch 2.2.2, CUDA version 12.1, and cuDNN version 8.9. The graph learning component used PyTorch Geometric 2.5.2, and NumPy 1.26.4, SciPy 1.11.4, scikit-learn 1.4.2, and pandas 2.2.2 were used for data processing and metric calculation. To ensure reproducibility, a fixed random seed of seed=42 was used, and deterministic computation options were enabled during training. Key runtime information, including dependency versions, GPU driver versions, memory usage, and training configuration files, was recorded to ensure stable alignment of results across different machines or under repeated running conditions.

5.3 Experimental Results and Analysis

This paper first presents the experimental results compared with other models, as shown in Table 1.

Table 1: Comparative experimental results

Method	Accuracy	Precision	Recall	F1	AUC	AP	PR-AUC
Johannessen et al.[8]	0.912	0.908	0.914	0.911	0.978	0.903	0.899
Tong et al.[9]	0.927	0.923	0.929	0.926	0.984	0.916	0.911
Wang et al.[10]	0.934	0.931	0.936	0.933	0.987	0.923	0.919
Saldaña-Ulloa et al.[11]	0.941	0.938	0.943	0.940	0.990	0.931	0.926
Kim et al.[12]	0.948	0.944	0.949	0.946	0.993	0.937	0.932
Adloori et al.[13]	0.953	0.950	0.955	0.952	0.994	0.944	0.939
Bakhshinejad et al.[14]	0.958	0.956	0.959	0.957	0.996	0.951	0.947
Ours	0.969	0.967	0.971	0.969	0.998	0.963	0.958

The methods exhibit a relatively consistent improvement gradient across the main classification metrics, indicating that as the modeling and feature representation capabilities of the transaction chain improve, the model's ability to capture suspicious behavior becomes more stable. Early methods often faced a significant trade-off between accuracy and recall, prone to either too many false positives or too many false negatives; subsequent methods have gradually achieved a more balanced performance between these two, leading to improvements in the overall F1 score. In contrast, the method presented in this paper excels in both accuracy and stability, demonstrating that by jointly utilizing the chain structure and risk semantics, the discrimination of boundary samples and complex association patterns is less susceptible to noise interference.

More importantly, the trends in ranking metrics correspond to those of classification metrics. This typically means that the model not only provides right or wrong judgments but also prioritizes risky samples, facilitating tiered handling and resource priority allocation in real-world scenarios. Metrics such as AUC, AP, and PR-AUC are more sensitive to class imbalance and better reflect the model's screening ability when high-risk samples are scarce. The method presented in this paper maintains a leading position on these metrics, indicating that its output risk scores are more discriminative. In particular, when it is necessary to quickly

identify a small number of high-risk transactions from a large number of normal transactions, the advantages brought by the joint modeling of link-level information and gang association are more easily demonstrated.

The learning rate directly affects the step size and convergence trajectory of the optimization process, thus changing how the model captures effective patterns during the training phase. For structured tasks such as joint modeling of transaction chains, excessively large or small learning rates can lead to unstable representation learning or insufficient updates. Therefore, learning rate sensitivity experiments are needed to verify the stability and controllability of the method under different learning rate settings. The experimental results are shown in Figure 2.

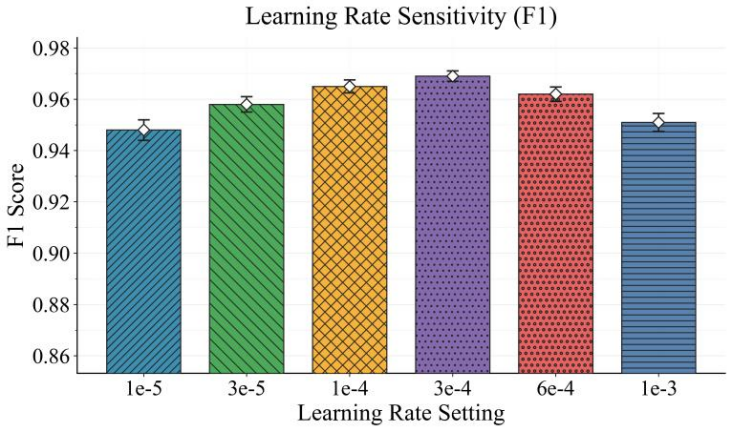

Figure 2. Sensitivity experiment of learning rate to F1

As shown in the figure, the F1 score initially increases and then decreases with the learning rate, indicating that there is a relatively suitable range for step size selection in this method. When the learning rate is too small, parameter updates are more cautious, and the model is more prone to repeated local trials, leading to a slower absorption of key link patterns and insufficient overall performance. As the learning rate increases to a suitable range, the update magnitude and convergence rhythm are better matched, and the model can more stably balance accuracy and recall, thus achieving a more ideal F1 score.

When the learning rate continues to increase, the F1 score decreases, reflecting that training instability caused by overly aggressive updates begins to dominate. For joint modeling of transaction links, representation learning not only needs to fit local transaction features but also aggregate and align link structures and potential collaborative behaviors. An excessively large step size amplifies noise gradients and batch fluctuations, causing shared representations to oscillate back and forth in different directions, ultimately affecting the consistency of boundary sample discrimination. The overall trend suggests that this method achieves a more robust optimization balance near a moderate learning rate, and also indicates that subsequent fine-tuning within this range should be prioritized under different data distributions or time windows to maintain stable output.

The shared representation dimension determines the upper limit of the model's capacity to accommodate link structure information and transaction semantic information within the same embedding space. Too small a dimension may limit the expression of complex association patterns, while too large a dimension may introduce redundancy and amplify training fluctuations. Therefore, it is necessary to examine the changes in the model's risk scoring capability as the shared representation dimension changes, thus providing a basis for subsequent configuration. The experimental results are shown in Figure 3.

The AUC initially improves gradually with increasing shared representation dimensionality, then declines in larger dimensional ranges, reflecting the balance between representational capacity and generalization ability. At lower dimensionalities, the encoder needs to compress transaction attributes and link structure signals simultaneously within a more compact space, potentially squeezing out key discriminative information and

resulting in insufficient characterization of complex link patterns. As the dimensionality increases to an appropriate range, the model can more completely retain multi-source risk cues, making the ranking of normal and suspicious patterns clearer, thus leading to more stable overall discrimination ability.

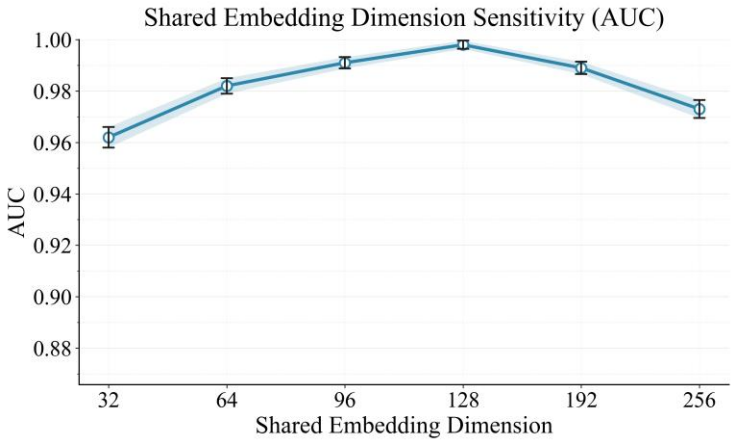

Figure 3. Sensitivity experiment of shared representation dimension to AUC

As the dimensionality continues to increase, performance degradation appears more like a side effect of redundant representations. Excessively high dimensionality makes the representation space more susceptible to noise and accidental correlations. Especially in joint modeling, where risk leaders and gang leaders share the same foundation, if the representation becomes too free, inconsistent traction directions between tasks may occur, resulting in less compact ranking boundaries. This trend suggests that larger shared dimensions are not always better. A more reasonable approach is to prioritize stable intervals around medium dimensionality and make minor adjustments when data distribution changes or links become denser to maintain consistency and robustness in risk ranking.

The hidden dimension of a graph encoder determines the available representation capacity during message passing, thus affecting the fusion quality of link structure information and transaction attribute information. A hidden dimension that is too small may limit the expression of key patterns, while a hidden dimension that is too large may introduce redundancy and amplify training fluctuations. To evaluate the stability of the method under different capacity configurations, sensitivity experiments on the hidden dimension are needed to observe the performance response to capacity changes. The experimental results are shown in Figure 4.

From an overall perspective, changes in hidden dimensions do indeed affect multiple metrics simultaneously, but it's not simply a matter of the larger the better. In low-dimensional settings, the graph encoder needs to simultaneously carry key information about link structures and transaction attributes within a more compact space, which can easily lead to over-compression of information, resulting in insufficient characterization of complex link patterns and causing some metrics to be inherently disadvantaged. As the capacity increases to an appropriate range, the model can more completely retain structural clues and neighborhood aggregation features of suspicious paths, leading to more stable overall judgments, especially those capabilities dependent on ranking and threshold boundaries.

Looking at the trends of Precision and Recall, they don't change completely synchronously under different dimensions, indicating that hidden dimensions influence the shape of the model's decision boundary. Some configurations tend to capture more suspicious samples, but are also more likely to introduce additional false positives; other configurations are more cautious, improving accuracy but sacrificing coverage. The change in F1 is therefore more like a compromise between the two, suggesting that this method is more likely to achieve a more balanced risk assessment near medium capacity, rather than amplifying one extreme.

AUC and PR-related metrics reflect the ranking quality of risk scores and their screening ability in imbalanced scenarios. As shown in the graph, they are more sensitive to hidden dimensions. When capacity is insufficient, ranking is more easily interfered with by noise and localized random correlations, resulting in insufficient concentration of risk score discrimination. When capacity is too large, some irrelevant details may be encoded, making the shared representation looser, and the risk head output is more prone to fluctuation across different batches or subgraph contexts. Especially in the joint modeling framework, where risk scores and gang clues share a common base representation, if the base is too flexible, the two learning signals are more likely to have inconsistent directions, thus weakening the stability of the ranking metrics.

Finally, looking at the trend of Avg Score, it provides a perspective closer to engineering choices: when multiple objectives exist simultaneously, capacity configuration should prioritize a more stable overall range rather than chasing individual peak values. A comprehensive analysis of the differences in the eight subgraphs leads to an intuitive conclusion: hidden dimensions should primarily focus on stably carrying structural information while avoiding overfitting due to excessive redundancy. A more reasonable strategy in practice is to lock the hidden dimension in the medium range as the default configuration, and then make minor adjustments based on graph density, time window length, or subgraph sampling scale to maintain consistency between ranking ability and classification boundary.

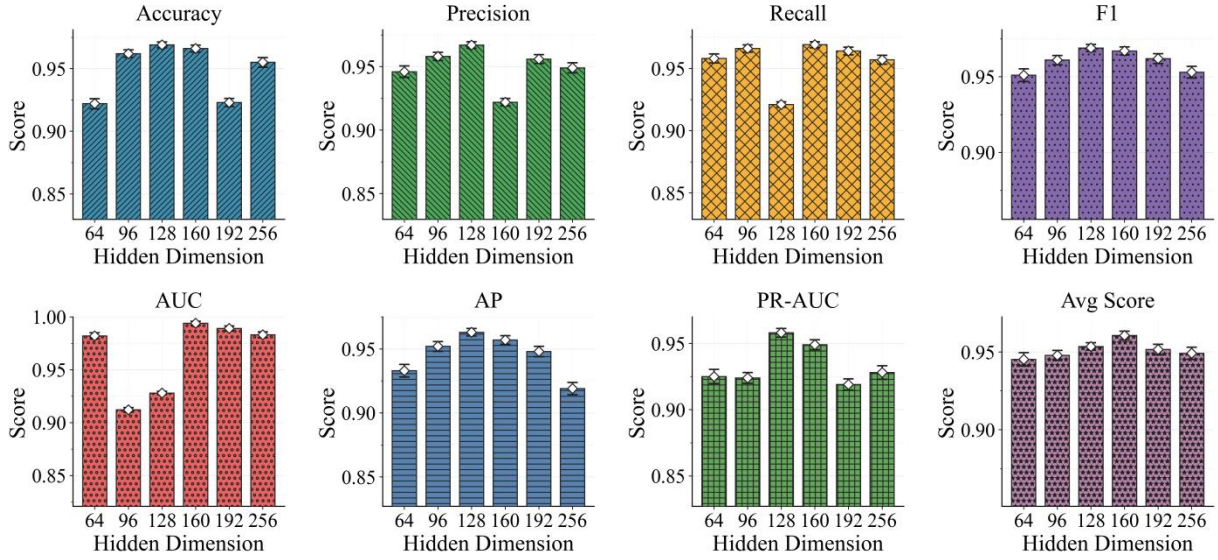

Figure 4. The impact of hidden dimensions of the graph encoder on experimental results

6. Conclusion

This paper, focusing on the risk governance needs of transaction links in anti-money laundering and anti-fraud efforts, proposes a research approach that jointly models transaction link risk scoring and gang identification, providing a unified task definition and methodological framework. This research emphasizes using the transaction network as the core carrier, placing transaction subgraph and transaction node in the same representation space for collaborative expression of structure and attributes, thus more closely reflecting the propagation patterns of risk and the organizational characteristics of collaborative behavior in real fund flows. Compared to the traditional approach of treating risk assessment and group clues separately, the joint perspective can achieve a ranked output of risk intensity and a linked characterization of suspicious group structures within the same system, providing a more consistent basis and a more coherent closed loop of clues for risk assessment and disposal.

From an application value perspective, the significance of this research lies not only in improving identification capabilities but also in providing a feasible organizational method for actual risk control

processes. Transaction link risk scoring can support tiered early warning, resource priority allocation, and disposal threshold setting, while gang identification can help locate potential collaborative networks, key hubs, and suspicious fund channels. The collaboration of both within a unified framework helps reduce information gaps in risk disposal, making the model output more aligned with the collaborative chain of compliance review and business response. In financial institutions' transaction monitoring, suspicious activity screening, risk list maintenance and continuous operation, this research is expected to reduce the cost of repetitive analysis and shorten the path from anomaly discovery to clue network formation and disposal execution, thereby improving the efficiency and consistency of risk governance.

Simultaneously, this research also has implications for the broader field of digital financial security. With the development of cross-platform payments, digital asset transfers, and multi-ecosystem account systems, risky behaviors increasingly rely on multi-entity collaboration and multi-hop link transfers to evade regulation and detection. Link-centric modeling is naturally suited to these complex scenarios and provides a unified technical paradigm for regulatory technology, anti-fraud platforms, payment clearing and settlement risk control, and transaction compliance monitoring. More importantly, collaborative modeling can connect individual-level risk signals with group-level collaborative structures, enabling the system not only to discover suspicious points but also to form an actionable risk network view, providing clearer decision support for subsequent strategy formulation and manual review.

Looking to the future, several directions still warrant further exploration. First, data distribution and attack strategies in real-world business environments will continue to change. How to ensure stability while rapidly adapting to new types of groups and new link patterns will be key to continuously improving practicality. Secondly, risk governance emphasizes explainability and traceability. Future efforts can further strengthen the ability to organize evidence at the chain level, enabling model outputs to more naturally correspond to auditable funding paths and group structure clues, thus enhancing compatibility with compliance processes. Finally, cross-institutional and cross-platform collaboration will gradually become a trend. How to achieve federated or collaborative chain modeling and group identification within the boundaries of privacy protection and compliance will determine the potential for wider adoption of this approach in the broader financial ecosystem. Overall, this paper provides a clear framework and scalable foundation for integrated modeling of transaction chain risk scoring and group identification, which plays a positive role in improving the intelligence level of digital financial security and risk governance systems.

References

[1] M. Cardoso, P. Saleiro and P. Bizarro, "Laundrograph: Self-Supervised Graph Representation Learning for Anti-Money Laundering", Proceedings of the Third ACM International Conference on AI in Finance, pp. 130-138, 2022.

[2] H. S. Assumpção, F. Souza, L. L. Campos, et al., "Delator: Money Laundering Detection via Multi-Task Learning on Large Transaction Graphs", 2022 IEEE International Conference on Big Data (Big Data), pp. 709-714, 2022.

[3] Y. Tian, G. Liu, J. Wang, et al., "Transaction Fraud Detection via an Adaptive Graph Neural Network", arXiv preprint arXiv:2307.05633, 2023.

[4] H. Tariq and M. Hassani, "Topology-Agnostic Detection of Temporal Money Laundering Flows in Billion-Scale Transactions", Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 402-419, 2023.

[5] Í. D. G. Silva, L. H. A. Correia and E. G. Maziero, "Graph Neural Networks Applied to Money Laundering Detection in Intelligent Information Systems", Proceedings of the XIX Brazilian Symposium on Information Systems, pp. 252-259, 2023.

[6] M. Weber, J. Chen, T. Suzumura, A. Pareja, T. Ma, H. Kanezashi and T. B. Schardl, "Scalable Graph Learning for Anti-Money Laundering: A First Look", arXiv preprint arXiv:1812.00076, 2018.

[7] I. Alarab, S. Prakoonwit and M. I. Nacer, "Competence of Graph Convolutional Networks for Anti-Money Laundering in Bitcoin Blockchain", Proceedings of the 2020 5th International Conference on Machine Learning Technologies, pp. 23-27, 2020.

[8] A. Mohan, K. PV, P. Sankar, K. Maya Manohar and A. Peter, "Improving Anti-Money Laundering in Bitcoin Using Evolving Graph Convolutions and Deep Neural Decision Forest", Data Technologies and Applications, vol. 57, no. 3, pp. 313-329, 2023.

[9] G. Tong and J. Shen, "Financial Transaction Fraud Detector Based on Imbalance Learning and Graph Neural Network", Applied Soft Computing, vol. 149, p. 110984, 2023.

[10] H. Milner, R. Mahmud, M. Afrin, S. G. Siddhartha, S. Mistry and A. Krishna, "On-Graph Machine Learning-Based Fraud Detection in Ethereum Cryptocurrency Transactions", 2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 1279-1285, 2023.

[11] D. Saldaña-Ulloa, G. De Ita Luna and J. R. Marcial-Romero, "A Temporal Graph Network Algorithm for Detecting Fraudulent Transactions on Online Payment Platforms", Algorithms, vol. 17, no. 12, p. 552, 2024.

[12] Y. Kim, Y. Lee, M. Choe, et al., "Temporal Graph Networks for Graph Anomaly Detection in Financial Networks", arXiv preprint arXiv:2404.00060, 2024.

[13] H. Adloori, V. Dasanapu and A. C. Mergu, "Graph Network Models to Detect Illicit Transactions in Block Chain", arXiv preprint arXiv:2410.07150, 2024.

[14] N. Bakhshinejad, U. T. Nguyen, S. Ghahremani and R. Soltani, "A Graph-Based Deep Learning Model for the Anti-Money Laundering Task of Transaction Monitoring", IJCCI, pp. 496-507, 2024.