HFTCRNet: Hierarchical Fusion Transformer for Interbank Credit Rating and Risk Assessment

Jiangtong Li, Ziyuan Zhou, Jingkai Zhang, Dawei Cheng, and Changjun Jiang

***Abstract*— As a prominent application of deep neural networks in financial literature, bank credit ratings play a pivotal role in safeguarding global economic stability and preventing crises. In the contemporary financial system, interconnectivity among banks has reached unprecedented levels. However, many existing credit risk models continue to assess each bank independently, resulting in inevitable suboptimal performance. Thus, developing advanced neural networks to model intricate temporal dynamics and interconnected relationships in the banking system is essential for an effective credit rating and risk assessment learning system. To this end, we propose a novel hierarchical fusion transformer for interbank credit rating and risk assessment (HFTCRNet), which includes the long-term temporal transformer (LT^3) module, short-term cross-graph transformer (STCGT) module, attentive risk contagion transformer (ARCT) module, and hierarchical fusion transformer (HFT) module to capture the long-term growth trajectories of banks, the short-term interbank network variance, the potential propagation of risks within interbank network, and integrate these information hierarchically. We further develop an interbank credit rating dataset, encompassing quarterly financial data, interbank lending networks, and key indicators such as credit ratings and systemic risk (SRISK) for 4548 banks from 2016Q1 to 2023Q1. Notably, we also adapt the minimum density algorithm to stabilize the interbank loan network over time, aiding in the analysis of long-term and short-term network effects. Our learning system uses semi-supervised learning to handle labels of varying sparsity, integrating credit ratings and SRISK for a comprehensive assessment of individual bank creditworthiness and systemic interbank risk. Extensive experimental results on our interbank dataset show that HFTCRNet not only outperforms all the baselines in terms of credit rating accuracy but also can evaluate the systemic risk within the interbank network. Code will be available at: https://github.com/AI4Risk/HFTCRNet.**

***Index Terms*— Deep learning, graph neural network (GNN), interbank credit rating, temporal transformer.**

Received 12 December 2023; revised 3 July 2024 and 27 August 2024; accepted 1 October 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB4501704, in part by the National Natural Science Foundation of China under Grant 62102287 and Grant 62402341, in part by the Shanghai Science and Technology Innovation Action Plan Project under Grant 22YS1400600 and Grant 22511100700, and in part by the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation (CPSF) under Grant GZC20241225. *(Corresponding authors: Dawei Cheng; Changjun Jiang.)*

Jiangtong Li is with the Department of Computer Science and Technology, Tongji University, Shanghai 200092, China (e-mail: jiangtongli@tongji.edu.cn).

Ziyuan Zhou is with the School of Economics and Management, Tongji University, Shanghai 200092, China (e-mail: 2054250@tongji.edu.cn).

Jingkai Zhang is with the Department of Software Engineering, Tongji University, Shanghai 200092, China (e-mail: 2151396@tongji.edu.cn).

Dawei Cheng and Changjun Jiang are with the Department of Computer Science and Technology, Tongji University, Shanghai 200092, China, and also with Shanghai Artificial Intelligence Laboratory, Shanghai 200233, China (e-mail: dcheng@tongji.edu.cn; cjjiang@tongji.edu.cn).

Digital Object Identifier 10.1109/TNNLS.2024.3475484

I. Introduction

IN TRADITIONAL finance, banks serve as regional resource distributors within specific areas. However, the advent of the digital era and the expansion of global trade have significantly altered global interactions of banks [1]. Furthermore, with new regulations and technological advancements, the financial landscape has evolved, notably with derivatives gaining prominence and altering the role of traditional interbank lending [2]. Consequently, basing credit ratings solely on financial status of banks is now considered outdated [3]. Therefore, exploring hierarchical neural network to model and fuse complex interbank loan relationships is now vital to construct an accurate credit rating and risk assessment learning system, essential for economic stability and preventing systemic financial crises.

Constructing an accurate interbank credit rating system is filled with challenges. The main issue lies in forming interbank loan networks that are consistent over time. Current interbank loan networks [4] are often built using either the maximum entropy [5] or the minimum density [6] methods. Even though these networks depict financial ties effectively [6], they frequently lack continuity and consistency over different time periods, which negatively affect modeling efforts temporally.

Exploring how risks spread within interbank loan networks is another complex task. Many existing models use graph networks to illustrate how risks move from one bank to another [7]. However, these graph networks are usually kept shallow to avoid problems such as oversmoothing [8]. Moreover, it is crucial to note that only a small number of contagion chains have a substantial impact on the entire banking system. Therefore, making the graph network deeper might not necessarily improve efficiency.

Furthermore, the impact of interbank loan networks differs between short-term and long-term timeframes [9]. In the short term, the structure of the interbank loan networks plays a pivotal role in credit ratings. For example, if a bank establishes connections with more highly rated banks recently, it indicates that this bank has gained recognition from these highly rated banks, suggesting that the rating of this bank should be higher as well, and vice versa. Conversely, in the long-term, the financial health of individual banks becomes more significant as it reflects the growth trajectories of each bank. The challenge of incorporating these dynamics into a unified framework has not been extensively investigated. Consequently, gaining a comprehensive understanding of how risks temporally spread within interbank loan networks remains an urgent matter. Moreover, using different bank labels to evaluate the systemic risk (SRISK) in the interbank loan network presents challenges. Various bank label types, such as SRISK [10] and

credit ratings [11], often exhibit varying degrees of annotation sparsity. For instance, SRISK values are available only for a limited subset of banks. Thus, incorporating diverse labels during modeling phase and leveraging them to assess systemic risk are both imperative. To address the previously mentioned challenges, we first curate a credit rating dataset and then introduce an innovative framework for simultaneous credit rating and risk assessment, leveraging interbank loan networks as a foundational component. Our dataset was meticulously collected from BankFocus and V-Lab[1] sources, encompassing quarterly financial data spanning from 2016Q1 to 2023Q1 for a total of 4548 banks. Moreover, we further construct the dynamic interbank loan networks of all banks by quarter that evolve over time. This dataset includes vital information such as credit rating labels and SRISK values.

In terms of our framework, we initiate the process by introducing an interbank credit rating model using a **h**ierarchical **f**usion **t**ransformer for interbank **c**redit rating and **r**isk assessment (HFTCRNet), which effectively captures risk propagation dynamics, facilitating a comprehensive analysis of interbank loan network. Our framework comprises the following components.

1) ***L**ong-**T**erm **T**emporal **T**ransformer (LT^3) Module:* This module is designed to capture the enduring growth trajectories of individual banks, providing insights into their long-term performance.
2) ***S**hort-**T**erm **C**ross-**G**raph **T**ransformer (STCGT) Module:* To grasp the short-term influences of interbank loan networks, we use the STCGT module. It detects changes in the structure of neighboring interbank loan networks over shorter time intervals.
3) ***A**ttentive **R**isk **C**ontagion **T**ransformer (ARCT) Module:* Contagion chains within each interbank loan network are identified and analyzed using the ARCT module. This allows us to capture the potential propagation of risks through interbank relationships.
4) ***H**ierarchical **F**usion **T**ransformer (HFT) Module:* Our framework integrates the findings from the LT^3 module, STCGT module, and ARCT module via the HFT module. This hierarchical integration enhances our understanding of interbank risk by combining long-term trends, short-term changes, and potential risk transmission pathways.

Moreover, our HFTCRNet operates in a semi-supervised manner, making effective use of multiple labels for modeling. This approach further combines various labels to evaluate the systemic risk within the interbank financial system. Extensive experiment on our dataset shows that our HFTCRNet not only outperforms all the baseline in terms of credit rating accuracy but also can evaluate the systemic risk of the interbank loan network.

The main contributions of this work are as follows.

1) We focus on the bank credit rating and risk assessment from the perspective of the temporal interbank dynamics and propose a novel HFTCRNet, leveraging LT^3, STCGT, ARCT, and HFT modules to hierarchically capture and integrate temporal interbank information.
2) We develop an interbank credit rating dataset, encompassing quarterly financial data, interbank lending networks, and key indicators such as credit ratings and SRISK for 4548 banks from 2016Q1 to 2023Q1. Notably, we also adapt the minimum density algorithm to stabilize the interbank loan network over time, aiding in the analysis of long-term and short-term network effects.
3) We optimize HFTCRNet in a semi-supervised manner, which not only uses multiple objective labels annotated with different annotation sparsity but also integrates credit ratings and SRISK for a comprehensive assessment of credit rating and systemic risk.
4) Extensive experimental results on our dataset demonstrate the efficacy of HFTCRNet for credit rating and risk assessment on interbank loan networks.

The rest of this article is organized as follows. In Section II, we briefly review related work about interbank loan network construction, risk propagation in credit rating, long-term and short-term effects in temporal graph, and semi-supervised learning in graph modeling, respectively. In Section III, we present a concise overview of the preliminary content, including the task definition and network construction. In Section IV, we present the hierarchical credit rating and risk assessment model by its four main modules. In Section V, our method is evaluated on our dataset for the credit rating and risk assessment tasks. Finally, the conclusion is presented in Section VI.

II. Related Work

We summarize related work in four areas: 1) interbank loan network characteristics; 2) risk in contagion chain; 3) modeling in temporal graph; and 4) semi-supervised learning in graph.

A. Interbank Loan Network Characteristics

Great importance of study has been attached to the potential threat of contagion defaults and the system risk ever since 2008. There are more and more works concentrating on the construction of the interbank loan network to model the bank system. The bilateral exposures of banks, however, are often not explicitly visible due to confidentially issues. Early work stress that the minimal amount of nonverifiable data from the external should be imposed on the problem. Then comes the maximum entropy estimation matrix [5], which evens the distribution of interbank claims under the constraints across all banks and therefore maximizes the interbank linkages. Various implementations based on the maximum entropy corresponding to certain constraints have also been studied [12], [13]. Moreover, the desired link density is used to generate random graphs, and average values across the graph ensemble contribute to the simulation of constraints observed [14].

In comparison, Anand et al. [6] put forward the minimum density method, considering the economic rationale of engaging in lending relationships with large amounts of banks. An improvement of using a binary index tree is added to the minimum density method, and its time complexity for calculation and sampling achieves notable progress [15]. In a distinct approach, Petropoulos et al. [16] concentrate on a smaller banking system, specifically the Greek banks, and integrate additional financial characteristics and previously observed behaviors into their estimation of interbank relationships. Different from research under static settings, a dynamic

[1]https://bankfocus.bvdinfo.com and https://vlab.stern.nyu.edu

balance sheet framework is designed to generate a dynamic contagion model by Feinstein and Sojmark [17].

All works mentioned above have contributed to the field of interbank loan network construction. Nevertheless, the construction of temporally consistent interbank loan networks is being neglected. In our study, we propose a novel method to implement the temporally consistent loan network and conduct further research into its properties.

B. Risk Contagion

In recent times, the intersection of artificial intelligence, data mining, and the investigation of systemic risk contagion in the banking arena has drawn considerable attention [18]. Contemporary studies often model the propagation of defaults within financial ecosystems by adapting the epidemic spread model from network theory. For instance, Cheng et al. [19] present an imbalanced network risk diffusion model, aiming to predict short-term default risks and introduce the positive weighted k-nearest neighbors (p-wkNN) algorithm for standalone scenarios. Taking a different approach, Barja et al. [20] use the microscopic Markov chain approach (MMCA) to scrutinize the effect of default diffusion in financial structures. On the simulation front, Jager et al. [21] propose a network-centric approach to evaluate interbank risks, offering exhaustive distributions of plausible interbank losses. Some works also study the risk contagion from the traditional machine learning viewpoint. For example, Song and Li [22] introduce a holistic framework leveraging nonlinear time-series analysis, aiming to pinpoint early warning signals (EWSs) indicative of looming phase transitions in banking systems. Laux et al. [23] empirically examine interbank risk contagion and use an long-short term memory (LSTM) to predict future interest variances. Li and Li [24] introduce a split-lending network model using the XGBoost classifier for bank credit risk assessment; however, their study does not address specific interbank risk contagion. Liu and Pun [25] enhance systemic risk measures and propose a dynamic framework for quantifying systemic risk using a two-step supervised learning approach. Kristóf and Virág [26] distinguish between good and bad banks within the EU-27 area, using logistic regression and machine learning techniques to capture nonlinearities and interactions.

Some works make an attempt by exploiting the graph learning, such as graph convolutional network (GCN) [8] and graph attention (GAT) [27] into risk contagion. Balmaseda et al. [28] introduce a graph neural network (GNN)-based method designed to minimize the prelabeling effort required for costly systemic risk metrics. This approach involves categorizing into a limited number of classes while simultaneously predicting continuous risk scores. A distinct approach is adopted by Cao et al. [29], who used a two-tier neural network methodology to deduce the inverse demand function, relying solely on partial data. Cheng et al. [30], [31] introduce temporal attention contagion chain enhanced rating model (TRACER) and stemming the contagion risk for regulation (iConReg), innovative risk-rating mechanisms that deduce the risk probabilities of contagion chains based on sequential loan activities. In a related effort, Liu et al. [7] explore spatial-aware interbank risk contagion and introduce a hyperfeature mapping operator using an attention-based graph representation for interbank credit rating prediction. In addition, Liu et al. [15] investigate the impact of graph attacks on interbank risk contagion and propose a selective-aware GNN (SAGNN) designed to prevent such attacks in interbank credit rating predictions.

A common thread in these endeavors is the risk propagation centered around the graph itself, with an inclination toward shallower layers to circumvent oversmoothing issues. We contend that a contagion chain with extended length is a pertinent consideration, prompting our initiative to architect contagion chains from an individual perspective.

C. Modeling in Temporal Graph or Dynamic Graph

Modeling temporal and dynamic graph [32], [33], [34], [35] is essential for capturing sequential relationships and performing tasks such as prediction. Recurrent neural network (RNN) and its variants [36] are commonly used in this domain. Chen et al. [37] introduce residual recurrent graph neural networks (MRes-RGNN) to capture periodic temporal dependencies. With the prominence of transformer architectures [38], some studies leverage the inherent structure of graphs to characterize the temporal framework. Shao et al. [39] introduce the multi-graph long-term spatio-temporal forecasting graph neural networks (MLSTGCN) framework, which combines the time domain with subgraphs, creating a multi-graph set for their study. Rankformer [40] is presented as an enhancement to transformer for long-term prediction tasks, making use of the rank correlation function and decomposition structure. Spatial–temporal graphs are also studied in the context of traffic flow prediction. Spatio-temporal graph convolutional network (STGCN) [41] develops a model for traffic flow prediction by alternately stacking temporal gated convolutions and spatial graph convolutions. In addition, spatio-temporal multi-graphs convolutional network (STMGCN) [42] introduces a contextual-aware gating mechanism to reweight different historical observations, while hierarchical spatio–temporal graph convolutional network and transformer (HSTGCNT) [43] proposes a hierarchical network that integrates long-term temporal relationships with short-term spatial–temporal relationships. Bidirectional spatial–temporal adaptive transformer (BiSTAT) [44] introduces a bidirectional spatial–temporal adaptive transformer for precise traffic forecasting, and time-enhanced spatio-temporal attention model (TESTAM) [45] separately models recurring and nonrecurring traffic patterns using a mixture-of-experts (MoE) model. Dynamic graph modeling is crucial for capturing the evolving characteristics of graphs. Frequency enhanced continuous-time dynamic graph model (FreeDyG) [46] incorporates frequency-domain analysis into the modeling process to capture more nuanced interaction patterns among different nodes. Memory-based dynamic graph neural networks - PREDICT-TO-SMOOTH (MDGNN-PRES) [47] focuses on entangled temporal and structural dependencies during batch training and proposes an iterative prediction-correction scheme combined with a memory coherence learning objective to mitigate the temporal discontinuity.

In the financial domain, modeling in the time domain finds applications across various contexts. LSTM [48] and temporal graph convolution [49] have been implemented for stock prediction. Notably, Zeng et al. [50] harness CNN and transformer for short-term and long-term influence prediction, respectively. However, there has been limited research into credit ratings and systemic risks within time-series models

and the evolving dynamics of the real financial world. In this article, we introduce an innovative framework designed to delve into these sectors. By investigating the dynamic changes in graph architecture and examining both long-term and short-term influences among different time periods, we aim to achieve insightful analyses of systemic risks and credit ratings.

D. Semi-Supervised Learning in Graph

The nature of graph data labeling, often characterized by its expensiveness or scarcity, has accentuated the significance of semi-supervised learning within the realm of GNNs [51], [52], [53], [54]. This methodology shines in harnessing graph-structured data, facilitating a more extensive understanding of graph data feature representations based on a limited labeled dataset [55]. Several pioneering works in this domain include those by Kong and Yu [56], who introduce gSSC, a semi-supervised feature selection methodology tailored for graph data. Another significant contribution is by Kipf and Welling [8], who pioneered a layerwise propagation rule for neural models, grounded in the first-order approximations of spectral graph convolutions. Further contributions from Zhuang and Ma [57] involve the dual-graph convolutional neural network, which ensures both local and global consistency in semi-supervised learning paradigms. Hao et al. [58] introduced ASGN, a teacher–student framework leveraging information from molecular structures and distributions to mitigate learning loss conflicts. Graph sample and aggregate (GraphSAGE) [59] offers a streamlined representation of hierarchical graphs, embedding diverse graph instances into fixed-length vectors. Dynamic graph convolutional networks by semi-supervised contrastive learning (DGSCL) [60] capitalizes on co-attention to merge embeddings from given topologies and dynamic graphs, enriching the embeddings for classification.

However, semi-supervised learning is rarely applied to credit rating and risk assessment. Given the sparseness intrinsic to SRISK metrics, the imperative to label these data through semi-supervised learning becomes clear. As such, we incorporated annotations of SRISK values on the graph across timeframes to bolster the precision of credit ratings' predictions.

III. Preliminary

In this section, we present the preliminary content, i.e., task definition in Section III-A and network construction in Section III-B.

A. Task Definition

Formally, we define a temporal interbank loan network as $G = [G_1, \ldots, G_T]$, where G_t represents an interbank loan network for one quarter, and the temporal networks span T consecutive quarters. Each interbank loan network G_t is represented as $G_t = (\mathbf{E}_t, \mathbf{X}_t)$, where adjacency matrix $\mathbf{E}_t \in \mathbb{R}^{N\times N}$ encodes the topological structure of N banks in the interbank loan network, where $\mathbf{E}_t^{i,j} \in \{0, 1\}$ in matrix $\mathbf{E}_t$ denotes the relationship between two nodes in position i and j. In addition, the bank node feature matrix $\mathbf{X}_t = \mathbf{x}_t^1, \ldots, \mathbf{x}_t^N \in \mathbb{R}^{N\times D_x}$ comprises M bank feature vectors, each with D_x bank characteristics. It is important to note that the number of banks remains the same across different interbank loan networks within G, but the graph connections vary over time. Given the temporal interbank loan networks denoted as G, we investigate two primary tasks, node classification and node regression. For node classification, we use credit ratings as our guiding supervision, where banks in the current temporal interbank loan network are assigned credit ratings for the upcoming quarter. Conversely, in the node regression task, we use SRISK values for guidance, where only a specific subset of nodes receives the labels.

These tasks form the core of our analysis. Specifically, credit rank represents the ability and willingness of a debtor to repay their obligations in the future, where a higher credit rank indicates a higher willingness and ability to repay, and vice versa. Besides, SRISK reflects the capital shortfall of a firm conditional on a severe market decline. In general, a higher SRISK for a bank indicates a higher importance for financial system, which means if such a bank encounters severe issues, its impact on the financial system is more substantial. By simultaneously predicting credit rank and SRISK, we can not only assess the credit rating of each bank in interbank loan network but also provide an overall evaluation of systemic financial risk in the financial system.

B. Interbank Loan Network

Studying bank risk spread often involves analyzing interbank loan relationships using network analysis. In this approach, each bank becomes a node, with loan relationships forming the edges between them. If a financial system has N banks, a matrix $\mathbf{Z} \in [0, \infty)^{N\times N}$ shows their loan positions. Each X_{ij} entry denotes the loan amount from bank i to bank j. For a bank i, the row sum of $\mathbf{Z}$ indicates the interbank assets $A_i = \sum_{j=1}^{N} \mathbf{Z}_{ij}$, and the column sum of $\mathbf{Z}$ shows the interbank liabilities $L_i = \sum_{j=1}^{N} \mathbf{Z}_{ji}$.

Considering that the real interbank loan network is not publicly available, interbank loan networks are often deduced based on certain assumptions. Therefore, to approximate the interbank loan network in practice, many studies use the maximum entropy principle. This principle posits that given limited information, risk exposure should be distributed evenly among active banks, while ensuring alignment with the available margins. Sheldon et al. [4] first apply this idea in interbank loan network, by maximizing the unpredictability of the interbank loan position matrix with scarce transaction data, which are then revised to the widely accepted method, i.e., maximum entropy estimation [5]. While the maximum entropy principle results in a fully connected interbank loan network, it is not practical given the high costs of processing information, managing risk, and verifying reputations. Thus, Anand et al. [6] introduced the minimum density method to address this problem. This approach targets the most likely links and allocates the biggest possible exposures to them, in line with the overall bank balance sheet, which can be formulated as a constrained optimization problem toward matrix Z

$$\begin{aligned} \min_{\mathbf{Z}} \quad & c \cdot \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{1}_{[\mathbf{Z}_{ij}>0]} \\ \text{s.t.} \quad & \sum_{i=1}^{N} \mathbf{Z}_{ij} = A_j, i = 1, 2, 3, \ldots, N \\ & \sum_{j=1}^{N} \mathbf{Z}_{ij} = L_i, j = 1, 2, 3, \ldots, N. \end{aligned} \tag{1}$$

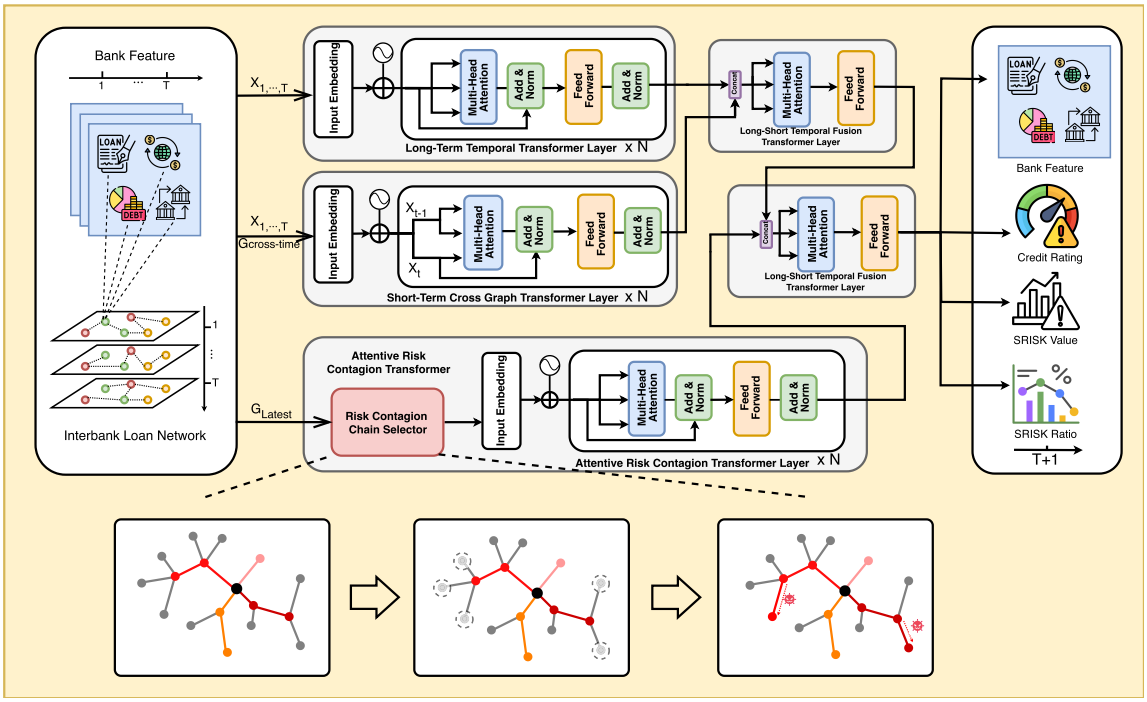


Fig. 1. Illustration of HFTCRNet. Given the temporal bank feature and temporal interbank loan network as input, we first explore LT^3 module to extract the growth trajectories of banks. Then we grasp the short-term influence of the interbank loan network with the STCGT module. At the meanwhile, we construct and select the vital contagion chains with risk contagion chain selector and capture the spatial risk propagation with the ARCT module. After that, the output of these three modules is fused together through the HFT module. Finally, four different losses are exploited to predict the future financial status.

To accelerate the optimization process, Liu et al. [15] store the interbank position matrix with binary index tree, which reduces the complexity of calculation and sampling from $O(n^2)$ and $O(n{\cdot}\log n)$ to $O(n{\cdot}\log n)$ and $O(\log n{\cdot}\log n)$, respectively.

In our study, we use the accelerated minimum density method to construct the temporal interbank loan networks. Whereas we observe significant changes in the network between consecutive quarters due to randomness when optimizing the matrix $\mathbf{Z}$. This might not reflect the variation in interbank loan relationships over quarters, impacting temporal modeling. To address this, we start the optimization of matrix $\mathbf{Z}$ with the interbank position matrix of previous quarter. In addition, in keeping with the overall balance sheet of banks, we introduce two vectors, namely, A_{res} and I_{res}, to track changes in interbank assets and liabilities between quarters. With this approach, our temporal interbank loan network remains consistent for subsequent modeling.

IV. Methodology

In this section, we first briefly present the framework in Section IV-A. Then we introduce LT^3 module in Section IV-B, STCGT module in Section IV-C, ARCT module in Section IV-D, and HFT module in Section IV-E. Finally, we introduce the training objective function and the optimization strategy in Section IV-F. In the remainder of this article, we use regular letters to represent scalar and bold letters to represent vector, matrix, and tensor.

A. Model Framework

To gauge the potential risk in temporal interbank loan network and assign credit ratings, we look at three key factors.

1) The financial performance of each bank over an extended period. Banks showing positive growth typically get a better credit ranking.
2) Variance of the interbank loan network recently. The credit rating of a bank likely improves if it increases ties with top-rated banks.
3) The exposure of banks to high-risk contagion chains. If a bank defaults, this risk can propagate through the contagion chain.

In this study, we construct three modules to capture vital information related to the three aspects previously mentioned, complemented by a fusion module for integration, as depicted in Fig. 1. Taking temporal bank features $\mathbf{X} \in \mathbb{R}^{N\times T\times D_x}$ and interbank loan networks $\mathbf{E} \in \mathbb{R}^{N\times N\times T}$ as input, we first use the LT^3 module to trace the growth patterns of banks as $\mathbf{L}_{\text{out}} \in \mathbb{R}^{N\times T\times D_x}$. Then the STCGT module discerns short-term changes in the interbank loan network with the cross-graph transformer between the interbank loan networks of adjacent quarters as $\mathbf{S}_{\text{out}} \in \mathbb{R}^{N\times T\times D_x}$. Meanwhile, we identify and prioritize high-risk contagion chains using the beam search algorithm. After that, the ARCT module highlights spatial risk propagation within these high-risk contagion chains as $\mathbf{C}_{\text{out}} \in \mathbb{R}^{N\times D_x}$. The outputs of these three modules are combined in stages, from fusing long- and short-term insights to integrating temporal and spatial aspects via the HFT module as $\mathbf{X}_{\text{out}} \in \mathbb{R}^{N\times D_x}$. The framework is optimized using four distinct loss functions in a semi-supervised manner, predicting the financial outlook, credit ranking, SRISK value, and SRISK ratio for the upcoming quarter. By predicting both credit rank and SRISK value together, our approach gauges credit ratings and offers a holistic view of systemic financial risks.

B. Long-Term Temporal Transformer Module

This module processes temporal data to discern long-term growth trajectories in financial health of banks. As depicted in Fig. 1, our LT^3 module integrates temporal position data with the financial status of bank, subsequently extracting enduring temporal insights.

For time-series data, the transformer [38] uses sine and cosine position embedding to capture the temporal order. The traditional position embedding is given by

$$\begin{aligned}\mathbf{PE}(\text{pos}, 2i) &= \sin\left(\frac{\text{pos}}{\text{period}^{2i/d}}\right)\\ \mathbf{PE}(\text{pos}, 2i+1) &= \cos\left(\frac{\text{pos}}{\text{period}^{2i/d}}\right).\end{aligned} \tag{2}$$

Here, pos is the sequence position, i is the feature index, and d is the feature dimension. The "period" parameter adjusts wavelengths, where a smaller period gives a sharper position distinction, suitable for our task. For every sequence element, we use sine and cosine to encode position. By appending this position embedding to the data sequence, the model learns positional context. We then merge the position embedding $\mathbf{PE} \in \mathbb{R}^{N\times T_{\text{in}}\times D_e}$ with bank financial status $\mathbf{X} \in \mathbb{R}^{N\times T_{\text{in}}\times D_x}$ to get input bank features $\mathbf{X}^{\text{LT}^3} \in \mathbb{R}^{N\times T_{\text{in}}\times (D_e+D_x)}$

$$\mathbf{X}^{\text{LT}^3} = \text{Concat}(\mathbf{X}, \mathbf{PE}) \tag{3}$$

where N is the number of banks, T_{in} is the length of the input bank features, D_e is the dimension of position embedding, and D_x is the input dimension of bank features.

To capture the long-term financial health of each bank, we propose an LT^3 module to extract the long-term temporal relationships from the entire financial status feature of banks. The LT^3 module consists of L transformer encoder layer (TEL), where each encoder is constructed by multihead attention (MH), fully connected feedforward layer, along with layer normalization through the residual connection. MH$(\cdot)$ is the core of the transformer, which uses h different linear transformations to analyze the previous layer feature and encourages the high-order feature interaction along the temporal dimension. By regarding $\mathbf{L}_0 = \mathbf{X}^{\text{LT}^3}$, the MH can be formulated as

$$\begin{aligned}\mathbf{H}_l &= \text{LN}(\text{MH}(\mathbf{L}_{l-1}) + \mathbf{L}_{l-1})\\ \text{MH}(\mathbf{L}_{l-1}) &= \text{Concat}\left(\mathbf{h}_1^{\text{mh}}, \ldots, \mathbf{h}_h^{\text{mh}}\right)\mathbf{W}_O\\ \mathbf{h}_i^{\text{mh}} &= \text{softmax}\left(\frac{\mathbf{L}_{l-1}\mathbf{W}_Q^i\left(\mathbf{L}_{l-1}\mathbf{W}_K^i\right)^{\text{T}}}{\sqrt{D_e+D_x}}\right)\mathbf{L}_{l-1}\mathbf{W}_V^i\end{aligned} \tag{4}$$

where $\mathbf{W}_K^i, \mathbf{W}_Q^i, \mathbf{W}_V^i \in \mathbb{R}^{(D_e+D_x)\times D_h}$ and $\mathbf{W}_O \in \mathbb{R}^{h\cdot D_h\times(D_e+D_x)}$ are trainable parameters, and LN [61] is the layer normalization. Therefore, by regarding $\mathbf{L}_l = \text{TEL}(\mathbf{L}_{l-1})$, and $\mathbf{L}_L = \mathbf{L}_{\text{out}}$, the TEL can be formulated as

$$\begin{aligned}\mathbf{L}^l &= \text{TEL}\left(\mathbf{L}^{l-1}\right) = \text{LN}\left(\text{FFN}\left(\mathbf{H}^l\right) + \mathbf{H}^l\right)\\ \text{FFN}\left(\mathbf{H}^l\right) &= \text{ReLU}\left(\mathbf{H}^l\mathbf{W}_{F1} + \mathbf{b}_{F1}\right)\mathbf{W}_{F1} + \mathbf{b}_{F1}\end{aligned} \tag{5}$$

where FFN and LN indicate the feedforward network and layer normalization, respectively, and $\mathbf{W}_{F1} \in \mathbb{R}^{(D_e+D_x)\times D_f}$, $\mathbf{b}_{F1} \in \mathbb{R}^{D_f}$, $\mathbf{W}_{F2} \in \mathbb{R}^{D_f\times(D_e+D_x)}$, and $\mathbf{b}_{F2} \in \mathbb{R}^{D_e+D_x}$ are trainable parameters. The feedforward network $(\cdot)$ contains two fully connected layers and a rectified linear unit (ReLU) activation function, which provides the nonlinear transformation and representation ability in TEL.

In the LT^3 module, we capture the long-term growth trajectories' feature as $\mathbf{L}_{\text{out}} \in \mathbb{R}^{N\times T\times D_x}$ with a linear transformation to adjust the feature dimension, where the receptive field of the self-attention mechanism can cover a long sequence of bank financial status. Therefore, our LT^3 can learn the bidirectional and long-term temporal relationships of the financial sequences.

C. Short-Term Cross-Graph Transformer Module

This module focuses on the variance of the interbank loan network along recent periods. As depicted in Fig. 1, our STCGT module captures the variance of interbank loan network between adjacent quarters through cross-graph transformer [62], [63] layers.

Given the input feature and interbank loan network of adjacent quarter ($\mathbf{X}^{(\mathbf{t})}$-$\mathbf{E}^{(t)}$ and $\mathbf{X}^{(\mathbf{t-1})}$-$\mathbf{E}^{(t-1)}$), to explore the short-term impact of the interbank loan network on bank features, we encourage the bank feature at time $t-1$ to integrate into the bank features at time t through both the interbank loan networks at times t and $t-1$. Using $\mathbf{E}^{(t)}$ as an example, the cross-graph transformer layers (CGTLs) among $\mathbf{X}^{t-1}$, $\mathbf{X}^t$, and $\mathbf{E}^{(t)}$ are formulated as

$$\begin{aligned}&\text{CGTL}(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t)}, \mathbf{E}^{(t)}) = \text{LN}(\text{FFN}(\mathbf{X}_{\text{cmh}}^{(t)}) + \mathbf{X}_{\text{cmh}}^{(t)})\\ &\mathbf{X}_{\text{cmh}}^{(t)} = \text{LN}(\text{CGMH}(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t)}, \mathbf{E}^{(t)}) + \mathbf{X}^{(t)})\\ &\text{CGMH}(\mathbf{X}^{(t-1)}, \mathbf{X}^{(t)}, \mathbf{E}^{(t)}) = \text{Concat}(\mathbf{h}_1^{\text{cmh}}, \ldots, \mathbf{h}_h^{\text{cmh}})\mathbf{W}_O\\ &\mathbf{h}_i^{\text{cmh}} = \text{softmax}\left(\frac{\mathbb{I}_{G^{(t)}}(\mathbf{X}^{(t)}\mathbf{W}_Q^i(\mathbf{X}^{(t-1)}\mathbf{W}_K^i)^{\text{T}})}{\sqrt{D_x}}\right)\mathbf{X}^{(t-1)}\mathbf{W}_V^i\end{aligned} \tag{6}$$

where CGMH is the cross-graph MH network, and LN and FFN are the layer normalization and the feedforward network, respectively, as we have introduced in (5). Besides, $\mathbf{W}_K^i, \mathbf{W}_Q^i$, and $\mathbf{W}_V^i \in \mathbb{R}^{D_x\times D_h}$ and $\mathbf{W}_O \in \mathbb{R}^{h\cdot D_h\times(D_x)}$ are trainable parameters. $\mathbb{I}_{G^{(t)}}(\cdot)$ is the edge indicator function. For edge i, j in interbank load network $\mathbf{E}^{(t)}$, the value of $\mathbb{I}_{G^{(t)}}(\mathbf{M})[i, j]$ is set as the original value of $\mathbf{M}[i, j]$, otherwise, the value of $\mathbb{I}_{G^{(t)}}(\mathbf{M})[i, j]$ is set as -1×10^{12}, resulting in a zero in the attention matrix.

The STCGT module enables the propagation of past bank features to the current ones using adjacent interbank loan networks. This module consists of the S STCGT layer (STCGTL). To capture the short-term variance of the interbank loan network from time $t-1$ to time t, we exploit two different CGTL in each STCGT and concatenate the output of these two CGTLs together. Setting $\mathbf{S}_0 = \mathbf{X}$ and $\mathbf{S}_{\text{out}} = \mathbf{S}_S$, the STCGTL for each time step is given by

$$\begin{aligned}\mathbf{S}_s^{(t)} &= \text{STCGTL}\left(\mathbf{S}_{s-1}^{(t-1)}, \mathbf{S}_{s-1}^{(t)}, \mathbf{E}^{(t-1)}, \mathbf{E}^{(t)}\right)\\ &= \text{Concat}\Big(\text{CGTL}\left(\mathbf{S}_{s-1}^{(t-1)}, \mathbf{S}_{s-1}^{(t)}, \mathbf{E}^{(t)}\right)\\ &\qquad \text{CGTL}\left(\mathbf{S}_{s-1}^{(t-1)}, \mathbf{S}_{s-1}^{(t)}, \mathbf{E}^{(t-1)}\right)\Big)\mathbf{W}_s + \mathbf{b}_s\end{aligned} \tag{7}$$

where $\mathbf{W}_s \in \mathbb{R}^{2\cdot D_x\times D_x}$ and $\mathbf{b}_s \in \mathbb{R}^{D_x}$ are the trainable parameters to adjust the feature dimension in the STCGT module.

In the STCGT module, we capture the short-term interbank loan network variance features as $\mathbf{S}_{\text{out}} \in \mathbb{R}^{N\times T\times D_x}$, where the

two CGTLs are exploited in each STCGTL to capture the affect the interbank loan network in term. By stacking S layers of STCGTL together, our STCGT can capture the variance of recent S interbank loan network.

D. Attentive Risk Contagion Transformer Module

This module focuses on capture the potential propagation of risks through interbank loan relationship. As depicted in Fig. 1, our ARCT module first identifies and prioritizes the high-risk contagion chains with beam search algorithm and then highlights the spatial risk propagation within these high-risk contagion chains.

Given the edge matrix $\mathbf{E}^{(T)}$ of interbank loan network, our goal is to establish high-risk contagion chains for every bank. This allows us to model the risk propagation for each bank independently. In Fig. 1 (bottom), we also illustrate the process of high-risk contagion chains' construction. Consider a contagion chain starting from bank A, passing through bank C, and ending at bank G, denoted as $c_A \to c_C \to c_G$. This chain is constructed based on two conditions: 1) a direct loan relationship exists, meaning bank C borrows from bank A, and bank G borrows from bank C and 2) the risk of this chain only measures the risk of bank A through this chain if bank G defaults. The contagion chain is derived from the interbank loan network, and its associated risk corresponds to the smallest loan value within the chain. To build these high-risk chains for each bank, we use the beam search algorithm, initializing the set of chains with each bank. The process iterates as follows until the chain reaches a predetermined length.

1) Expand every existing chain by appending banks that have borrowed from the last node of the chain. In addition, note down the lent amount.
2) Update the risk metric of each chain using the smallest loan value within that chain.
3) For every target bank, retain the top-k chains with the highest risk value. If the risk values of two chains are the same, we compare the length of the two chains first and then compare other lent values in ascending order for both the chains.

In step one, we incorporate an end of chain ("EOC") node, ensuring risk does not dissipate as the chain length increases. To ensure the diversity of the high-risk contagion chains, we encourage the adjacent node of target to be different. Moreover, Algorithm 1 offers a detailed method for constructing high-risk contagion chains. Ultimately, we generate the high-risk contagion chains, denoted as $\mathbf{C} \in \mathbb{R}^{N\times N_c\times L_c}$, using the interbank loan network $\mathbf{E}_T$, where N is the bank numbers, N_c is the contagion chain count of each bank, and L_c is the maximum length of chains.

Given the high-risk contagion chains $\mathbf{C}$, our objective is to model risk considering two perspectives, the chain-aware risk representation and the bank-aware risk representation derived from the chains. To achieve this, we use two ARCT layers (ARCTLs) to obtain representations for chains and banks sequentially. Initially, we transform high-risk contagion chains $\mathbf{C}$ into feature representations $\mathbf{C}_f \in \mathbb{R}^{N\times N_c\times L_c\times D_x}$ using the indexing operation, denoted as $\mathbf{C}_f = \mathbf{X}[\mathbf{C}]$. Subsequently, we apply the TEL equipped with an attention mechanism (ATT) to derive the chain-aware risk representation. For each set of chain features $\mathbf{C}_f^{(n_c)}$ from $\mathbf{C}_f$, the formulations are

$$\begin{aligned}\mathbf{C}_{\text{mid}}^{(n_c)} &= \text{TEL}\left(\mathbf{C}_f^{(n_c)}\right),\\ \mathbf{C}_f^{(n_c)} &= \text{ATT}\left(\mathbf{C}_{\text{mid}}^{(n_c)}, \mathbf{X}^{(T)}\right)\\ &= \text{softmax}\left(\frac{\left(\mathbf{X}^{(T)}\mathbf{W}_Q^1\left(\mathbf{C}_{\text{mid}}^{(n_c)}\mathbf{W}_K^1\right)^{\text{T}}\right)}{\sqrt{D_x}}\right)\mathbf{C}_{\text{mid}}^{(n_c)}\mathbf{W}_V^1 \end{aligned} \tag{8}$$

where $\mathbf{W}_K^1$, $\mathbf{W}_Q^1$, and $\mathbf{W}_V^1$ are trainable parameters. The ATT emphasizes features in chain representation that are similar to the bank features. Next, a secondary ARCTL captures the bank-specific risk representation

$$\begin{aligned}\mathbf{C}_{\text{mid}} &= \text{TEL}(\mathbf{C}_c),\\ \mathbf{C}_{\text{out}} &= \text{ATT}\left(\mathbf{C}_{\text{mid}}, \mathbf{X}^{(T)}\right)\\ &= \text{softmax}\left(\frac{\left(\mathbf{X}^{(T)}\mathbf{W}_Q^2\left(\mathbf{C}_{\text{mid}}\mathbf{W}_K^2\right)^{\text{T}}\right)}{\sqrt{D_x}}\right)\mathbf{C}_{\text{mid}}\mathbf{W}_V^2 \end{aligned} \tag{9}$$

where $\mathbf{W}_K^2$, $\mathbf{W}_Q^2$, and $\mathbf{W}_V^2$ are the trainable parameters.

In the ARCT module, we first derive the high-risk contagion chains. Subsequently, we achieve both chain-aware and bank-aware risk representations. The bank-aware representation, denoted as $\mathbf{C}_{\text{out}} \in \mathbb{R}^{N\times D_x}$, embodies the spatial risk profile for the recent quarter. This is later integrated with temporal attributes for credit rating and risk assessment.

E. Hierarchical Fusion Transformer Module

In LT^3, STCGT, and ARCT modules, we capture the long-term growth trajectories of each bank, short-term variance of interbank relationships, and the spatial risk of contagion chains. To properly integrate these heterogeneous data, we propose the HFT module to learn a better bank representation from long-short temporal fusion to temporal–spatial fusion through a cross-attention mechanism. Given the long-term growth trajectories' feature $\mathbf{L}_{\text{out}} \in \mathbb{R}^{N\times T\times D_x}$ and short-term interbank loan network variance features $\mathbf{S}_{\text{out}} \in \mathbb{R}^{N\times T\times D_x}$, we fuse these two heterogeneous bank feature by MH mechanismATT along the temporal dimension. Specifically, we start by concatenating these two features together. Subsequently, we perform MH along the time dimension to ensure that new information from both long-term and short-term features can be thoroughly integrated over temporal dimension, which can be formulated as

$$\mathbf{T}_{\text{out}} = \text{FS}_1(\mathbf{L}_{\text{out}}, \mathbf{S}_{\text{out}}) = \text{MH}(\text{Concat}(\mathbf{L}_{\text{out}}, \mathbf{S}_{\text{out}})) \tag{10}$$

where $\mathbf{T}_{\text{out}} \in \mathbb{R}^{N\times T\times D_x}$, MH is the MH network in (4), and FS_1 is short for fusion layer 1. Then, we use a pooling operation to compress the temporal information, resulting in the temporal feature $\mathbf{T}_{\text{out}}^f \in \mathbb{R}^{N\times D_x}$. Subsequently, we merge the temporal feature $\mathbf{T}_{\text{out}}^f$ with the spatial bank-aware feature $\mathbf{C}_{\text{out}}$ through the MH, as depicted in (4). This encourages information propagation among distinct bank features in provided graph, which is formulated as

$$\mathbf{X}_{\text{out}} = \text{FS}_2\left(\mathbf{T}_{\text{out}}^f, \mathbf{C}_{\text{out}}\right) = \text{MH}\left(\text{Concat}\left(\mathbf{T}_{\text{out}}^f, \mathbf{C}_{\text{out}}\right)\right) \tag{11}$$

Algorithm 1 High-Risk Contagion Chains

Input: E: Edge matrix. N_c: Contagion chain count of each bank. L_c: Maximum length of chains.
Output: C: The collection of high-risk chains.

Function `FindPath` (*node, path, paths, visited, C*) **:**
  Result: flag
  $visited[node] \leftarrow True$;
  $path \leftarrow path \cup \{node\}$;
  if length of $path = N_c$ or $T(node) = \emptyset$ **then**
    $paths \leftarrow paths \cup \{path\}$;
    if length of $paths < L_c$ **then**
      $C \leftarrow C \cup \{paths\}$;
    $visited[node] \leftarrow False$;
    $path \leftarrow path \setminus \{node\}$;
    return True;
  $flag = False$;
  for t in $T(u)$ **do**
    if *not visited*$[t]$ **then**
      $flag =$ FindPath(t, $path$, $paths$, $visited$, C);
      if *flag* = *True* **then**
        break;
  $visited[u] \leftarrow False$;
  $path \leftarrow path \setminus \{u\}$
  return flag;
$C \leftarrow \emptyset$;
let V is a collection of nodes from E
foreach v in V **do**
  $T(v) \leftarrow \{u \mid u \in V, u$ is the neighbor of $v\}$;
  $visited[v] = False$;
for v in V **do**
  $path \leftarrow \{v\}$;
  $paths \leftarrow \emptyset$;
  for u in $T(v)$ **do**
    FindPath(u, $path$, $paths$, $visited$);
return C;

where FS_2 is short for fusion layer 2. Here, $\mathbf{X}_{\text{out}} \in \mathbb{R}^{N \times D_x}$ consolidates information ranging from long-term growth of banks and short-term fluctuations in interbank relationships to spatial risk in contagion chains.

The HFT module serves as an adaptive attentive fusion mechanism. Instead of using a one-size-fits-all approach, this module dynamically allocates appropriate weights to the long-term temporal, short-term temporal, and spatial risk features for different banks. In doing so, it offers a nuanced analysis of the interbank relationships among banks.

F. Objective Function and Optimization Strategy

Given $\mathbf{X}_{\text{out}}$ from our HFTCRNet, we aim to predict the credit rank and SRISK of the next quarter for every bank. Therefore, we first map $\mathbf{X}_{\text{out}}$ with four different linear transformations to predict the future financial status, credit rank, SRISK value, and SRISK rate as

$$\begin{aligned} \hat{\mathbf{X}}_{\text{fs}} &= \mathbf{X}_{\text{out}}\mathbf{W}_{\text{fs}} + \mathbf{b}_{\text{fs}}, \quad \hat{\mathbf{X}}_{\text{cr}} = \mathbf{X}_{\text{out}}\mathbf{W}_{\text{cr}} + \mathbf{b}_{\text{cr}} \\ \hat{\mathbf{X}}_{\text{sv}} &= \mathbf{X}_{\text{out}}\mathbf{W}_{\text{sv}} + \mathbf{b}_{\text{sv}}, \quad \hat{\mathbf{X}}_{\text{sr}} = \mathbf{X}_{\text{out}}\mathbf{W}_{\text{sr}} + \mathbf{b}_{\text{sr}} \end{aligned} \tag{12}$$

where $\mathbf{W}_{\text{fs}} \in \mathbb{R}^{D_x \times D_{\text{fs}}}$, $\mathbf{W}_{\text{cr}}$, $\mathbf{W}_{\text{sv}}$, $\mathbf{W}_{\text{sr}} \in \mathbb{R}^{D_x \times 1}$, $\mathbf{b}_{\text{fs}} \in \mathbb{R}^{D_{\text{fs}}}$, $\mathbf{b}_{\text{cr}}$, $\mathbf{b}_{\text{sv}}$, and $\mathbf{b}_{\text{sr}} \in \mathbb{R}^1$ are the trainable parameters. For future financial status and credit rating, we supervised them with the ground-truth financial status and the credit rating through mean square error and cross-entropy. For SRISK value and rate, we exploit the semi-supervised learning by first supervising on the annotated banks and then inferring the unannotated banks and finally supervising all the banks. The loss function can be formulated as

$$\begin{aligned} \mathcal{L}_{\text{fs}} &= ||\hat{\mathbf{X}}_{\text{fs}} - \mathbf{X}_{\text{fs}}||_2, \quad \mathcal{L}_{\text{cr}} = \text{CE}\big(\hat{\mathbf{X}}_{\text{cr}}, \mathbf{X}_{\text{cr}}\big) \\ \mathcal{L}_{\text{sv}} &= ||\hat{\mathbf{X}}_{\text{sv}} - \mathbf{X}_{\text{sv}}||_2, \quad \mathcal{L}_{\text{sr}} = ||\hat{\mathbf{X}}_{\text{sr}} - \mathbf{X}_{\text{sr}}||_2 \\ \mathcal{L} &= \mathcal{L}_{\text{cr}} + \lambda_1\mathcal{L}_{\text{fs}} + \lambda_2\mathcal{L}_{\text{sv}} + \lambda_3\mathcal{L}_{\text{sr}} \end{aligned} \tag{13}$$

where λ_1, λ_2, and λ_3 are the hyperparameters balancing the importance of difference modules.

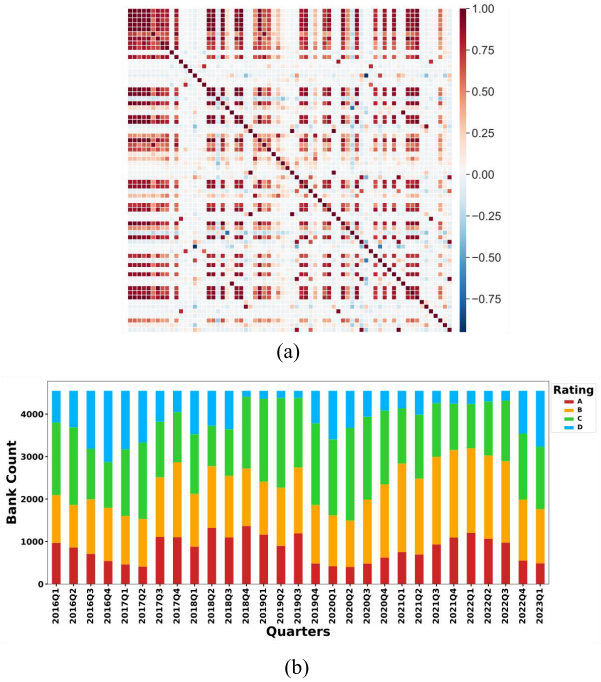
G. Time Complexity

The HFTCRNet consists of four key components: LT^3, STCGT, ARCT, and HFT modules. To determine the time complexity, we define the number of nodes as N, the number of quarters as T, the hidden dimension of HFTCRNet as D, each node connected to K nodes on average, the maximum length of the risk contagion chain as L, and the number of risk contagion chains as R. We calculate the time complexity per layer as follows: For the LT^3 module, which uses a transformer layer in the temporal dimension, the time complexity is $O(NT^2D)$. The STCGT module, involving a graph transformer layer between temporally adjacent graphs, has a time complexity of $O(TNKD)$. In the ARCT module, where the transformer layer operates within the risk contagion chain, the time complexity is $O(NRL^2D)$. For the HFT module, which uses a transformer layer first in the temporal level and then in the graph level, the time complexity is $O(NT^2D + NKD)$. Given that $N \gg T \approx K \approx L \approx R$, the overall time complexity of HFTCRNet is $O(NRL^2D)$, indicating that the time complexity is primarily determined by the ARCT module. In comparison, the time complexity of HSTGCNT [43] is $O(NT^2D + NTKD) = O(NTD \times (T + K))$, showing that HFTCRNet is more complex. However, given the relatively small scales of R and L, our time complexity is acceptable.

V. Experiments

A. Experimental Settings

1) Dataset: In our study to devise the interbank credit rating and risk assessment, we have amassed an extensive dataset spanning 29 quarters, from 2016Q1 to 2023Q1, incorporating records from a total of 4548 banks worldwide, which is named as **Interbank**. These banks encompass various types including commercial, savings, cooperative, real estate & mortgage, investment, Islamic, and central banks. To specifically evaluate the financial crisis in the first quarter of 2023, prominent entities such as Silicon Valley Bank (SVB), Signature Bank (SBNY), First Republic Bank (FRCB), and Credit Suisse (CS) Group, alongside their closely affiliated banks, have been



(a)

(b)

Fig. 2. (a) Correlation heat-map of selected features. The correlation is measured through cosine similarity between two features averaged on all quarters. (b) Rank distribution among different quarters.

incorporated into our dataset. For each quarter, our dataset includes four parts of content, the bank financial status matrix, the interbank loan network adjacency matrix, the credit rank of banks, and the SRISK metrics of banks. The bank financial status matrix is collected from the BankFocus, where we collect 383 features related to the bank finance status initially. After feature pruning with variance threshold and correlation coefficient, we identified 70 features as the finally feature set. The correlation heat-map of selected feature is shown in Fig. 2(a). The interbank loan network adjacency matrix is construct by our improved minimum density method (refer to Section III-B) to generate the temporally consistent interbank loan networks. For data ratings, we collect existing ranks given by Moody's Analytics [64] and transfer original bank ranks of each bank to relative ranks with "A," "B," "C," and "D" to facilitate further analysis on the systemic risk. The distribution of the relative ranks is shown in Fig. 2(b). The objective of the SRISK is to measure the capital shortfall of financial firms conditional on a systemic event, and the SRISK is calculated by its size, leverage, and risk. Therefore, we collect SRISK value and SRISK rate from vLab, where the SRISK value measures the potential impact of a bank in the interbank system, and SRISK ratio is the ratio between SRISK value and total SRISK value.

To further verify the efficacy of our model, we conduct additional comparative experiments using a related financial dynamic graph dataset, DGraph-Fin [65]. DGraph-Fin comprises a directed, unweighted dynamic graph that encapsulates a social network among users of the Finvolution Group. Within this network, each node represents an individual user, and a directed edge from one user to another signifies that the originating user has nominated the target user as their emergency contact. The core challenge of DGraph-Fin is to identify fraudulent users, a task that requires analyzing both node

TABLE I

ACCURACY AND MACRO-F_1 WITH DIFFERENT METHODS FOR CREDIT RATING AND THE ROC-AUC WITH DIFFERENT METHODS ON DGRAPH-FIN. BEST AND SECOND BEST PERFORMANCE ARE HIGHLIGHTED IN BOLD AND UNDERLINE, RESPECTIVELY

Method	Interbank		DGraph-Fin
	Accuracy	Macro-F_1	ROC-AUC
GCN	0.537 ± 0.060	0.435 ± 0.043	0.751 ± 0.007
GAT	0.540 ± 0.087	0.429 ± 0.123	0.752 ± 0.007
SAGNN	0.546 ± 0.059	0.436 ± 0.046	0.747 ± 0.011
TGAR	0.628 ± 0.084	0.569 ± 0.094	0.721 ± 0.030
STGCN	0.562 ± 0.129	0.469 ± 0.126	0.756 ± 0.006
STMGCN	0.614 ± 0.081	0.553 ± 0.086	0.762 ± 0.010
HSTGCNT	<u>0.659</u> ± 0.099	<u>0.598</u> ± 0.106	<u>0.775</u> ± 0.016
TESTAM	0.647 ± 0.095	0.603 ± 0.102	0.765 ± 0.009
FreeDyG	0.615 ± 0.075	0.557 ± 0.089	0.770 ± 0.012
MDGNN_bs	0.648 ± 0.089	0.590 ± 0.102	0.772 ± 0.013
BiSTAT	0.622 ± 0.088	0.544 ± 0.082	0.757 ± 0.012
HFTCRNet	**0.678** ± 0.104	**0.630** ± 0.110	**0.784** ± 0.014

features and the structure of the graph elements in financial applications. To align this dataset with our specifications of our models, we reformat the data into segments consisting of ten time steps each, i.e., ten days, for our experimental analysis. It is important to note that since DGraph-Fin lacks risk contagion chains, we do not use the ARCT module of our model in these experiments. This adjustment ensures that application of our model to DGraph-Fin is both appropriate and optimized for the structure of this dataset.

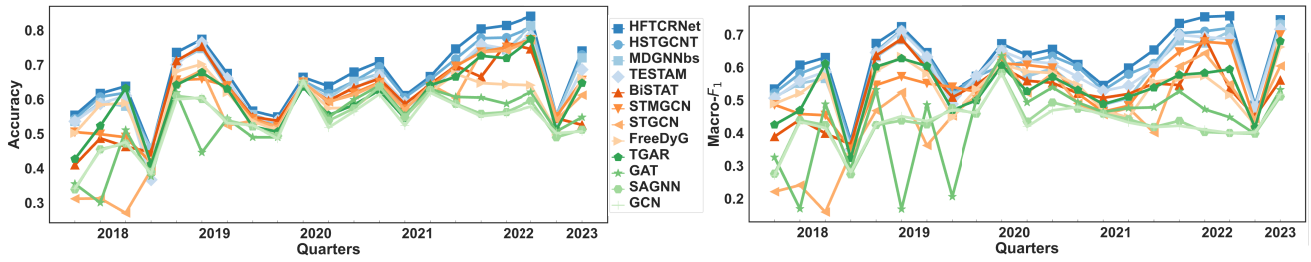
2) Compared Method: To reveal the effectiveness of our HFCTRNet, we use several state-of-the-art methods on our credit rating and risk assessment dataset. These baselines can be divided into two parts, the spatial methods and spatial–temporal methods.

The spatial methods include the following.

1) *GCN [8]:* The original GCN model which defines the graph convolution as aggregating features from neighboring.
2) *GAT [27]:* This model is composed of attention layers that learn different weights to different nodes in neighborhood.
3) *SAGNN [15]:* This model explores selective representation layer with GCN and GAT to explore the graph attack in interbank credit rating.
4) *Transformed Graph Attention for Credit Rating (TGAR) [7]:* This model proposes a hyperfeature mapping operator based on attention-based knowledge representation to build a knowledge system for interbank credit rating.

The spatial–temporal methods include the following.

1) *STGCN [41]:* This model formulates the traffic flow prediction on graph and builds the model by stacking the temporal gated convolution and spatial graph convolution in turn.
2) *STMGCN [42]:* This model proposes contextual gated RNN which augments RNN with a contextual-aware gating mechanism to reweights different historical observations.
3) *HSTGCNT [43]:* This model proposes a hierarchical network to incorporate the long-term temporal relationships and short-term spatial–temporal relationships through attention.
4) *BiSTAT [44]:* This model introduces an encoder–decoder architecture, featuring both spatial-adaptive and



(a) (b)

Fig. 3. (a) Accuracy for credit rating among different quarters. (b) Macro-F_1 of credit rating among different quarters.

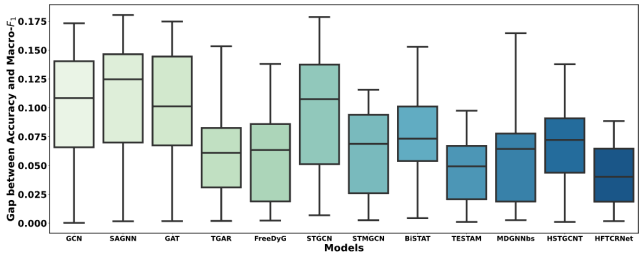


Fig. 4. Gap between accuracy and macro-F_1 in boxplot.

TABLE II

ACCURACY AND MACRO-F_1 OF HFTCRNET WITH DIFFERENT VARIANTS FOR CREDIT RATING PREDICTING

Method	Accuracy	Macro-F_1
HFTCRNet	0.678 ± 0.104	0.630 ± 0.110
(w/o) LT^3	0.594 ± 0.066	0.512 ± 0.065
(w/o) STCGT	0.642 ± 0.094	0.571 ± 0.096
(w/o) ARCT	0.638 ± 0.091	0.565 ± 0.094
(w/o) HFT	0.662 ± 0.101	0.617 ± 0.111
(w/o) Semi	0.669 ± 0.109	0.620 ± 0.120

temporal-adaptive transformers in the encoder and decoder.

5) *TESTAM [45]:* This model presents an MoE model with three experts on temporal modeling, static spatiotemporal modeling, and dynamic spatiotemporal modeling.
6) *FreeDyG [46]:* This model incorporates a node interaction frequency encoding module to capture the proportion of neighboring nodes and the frequency of node interactions.
7) *MDGNN-PRES [47]:* This model introduces an iterative prediction-correction scheme, with a memory coherence learning objective, to address temporal discontinuity.

To measure the performance of all the methods in credit rating, two common metrics are chosen, i.e., accuracy and macro-F_1, where macro-F_1 is calculated by first calculating F_1 in each credit rank, (i.e., "A," "B," "C," and "D"), and then averaging these four F_1 as macro-F_1. Besides, we use R^2 to measure the performance of SRISK prediction, which measures the proportion of the variance in the dependent variable that is predictable from independent variables, i.e., how well the model fits the data. The R^2 is calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{14}$$

where N is the total number of banks, y_i is the ground truth of the ith bank, $\hat{y}_i$ is the prediction of the ith bank, and $\bar{y}$ is the average of the ground truth.

TABLE III

ACCURACY AND MACRO-F_1 OF HFTCRNET WITH DIFFERENT VARIANTS FOR CREDIT RATING PREDICTING FROM 2021Q1 TO 2023Q1

Method	Accuracy	Macro-F_1
HFTCRNet	0.678 ± 0.104	0.630 ± 0.110
STCGT-GCN	0.669 ± 0.093	0.609 ± 0.118
STCGT-GAT	0.670 ± 0.113	0.620 ± 0.108
ARCT (largest)	0.663 ± 0.096	0.611 ± 0.110
Graph (MinDen)	0.668 ± 0.092	0.616 ± 0.096
(w/o) SRISKR	0.670 ± 0.107	0.620 ± 0.100
(w/o) SRISKV	0.673 ± 0.104	0.624 ± 0.099

3) Implementation Details: We implement our method using PyTorch [66], which is trained on one RTX 8000 GPU. We use Adam [67] optimizer $\beta_1 = 0.5$ and $\beta_2 = 0.999$ to train our model with fixed learning rate 0.01 for 2000 epochs. For each epoch, we directly send the whole temporal interbank loan network to our HFCTRNet, and therefore, the batch size equals to the bank size, i.e., 4548.

For input data, we exploit seven consecutive quarters to formulate the temporal interbank loan network and bank features. For the input bank feature, we first exploit a linear transformation to project the feature dimension from 70 to 256 ($D_x = 256$). For the LT^3 module, we stack two LT^3 layers with four head in TEL. The dimension for positional embedding is set as the same dimension as the feature dimension 256. For the STCGT module, we stack two STCGTLs with two heads in CGTL. For the ARCT module, we extract at most eight contagion chains for each bank with the maximum length as eight, and we set the number of head in the TEL as four. For the HFT module, we exploit max-pooling operation to compress the temporal information and the number of head in the TEL is also set as four. For optimization, we set λ_1, λ_2, and λ_3 as 0.1, 0.5, and 0.5, respectively. Moreover, for the first 1000 epochs, we only optimize $\mathcal{L}_{sv}$ and $\mathcal{L}_{sr}$ on the annotated SRISK metrics; for the last 1000 epochs, we optimize these two SRISK losses on both annotated and generated SRISK metrics.

For baseline comparisons, we use PyTorch to implement all the baseline models and train them using an RTX 8000 GPU. Each model is trained with a learning rate of 0.01 over 2000 epochs and a batch size of 4548, which corresponds to the size of the dataset (bank size), which is consistent with our HFTCRNet. Regarding the hyperparameters, we configure both the GCN [8] and GAT [27] with two layers. To ensure a fair comparison, the hidden and embedding dimensions for all the baseline models are uniformly set to 256, matching the configuration of our HFTCRNet. In addition, we adhere to the original hyperparameter settings specified in

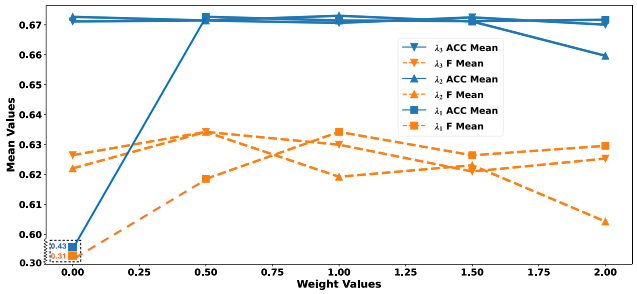
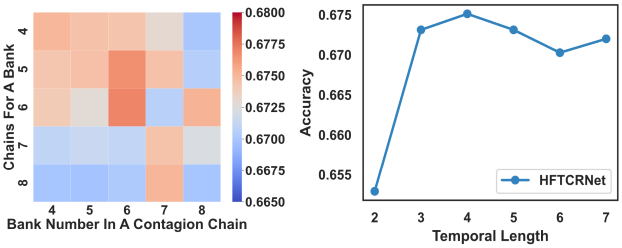


Fig. 5. Variance of accuracy and macro-F_1 along with the change in λ_1, λ_2, and λ_3.



(a) (b)

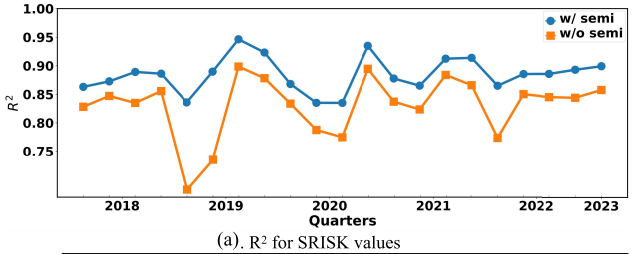
Fig. 6. (a) Accuracy heat-map with different contagion chains for a bank N_c and the maximum bank number in each contagion chain L_c. (b) Credit rating accuracy with different temporal lengths.

each corresponding study for other variables, maintaining the integrity of the analysis.

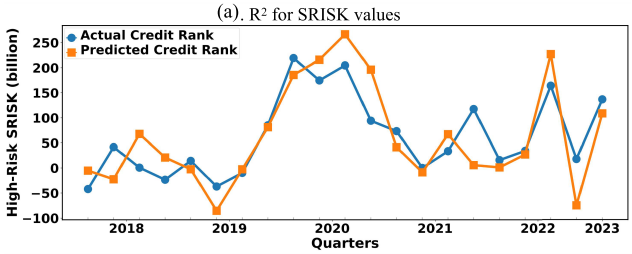
B. Credit Rating

We undertook an evaluation of diverse models in credit rating predictions using our dataset. The credit rating task is structured such that content from n consecutive quarters is leveraged as input to forecast the credit rank for the subsequent quarter. Initially, the dataset comprising bank features and the interbank loan network, spanning from 2016Q1 to 2017Q3, was used as input to train models to predict the credit rank for 2017Q4. For evaluation purposes, the observation window was slid forward by one quarter, allowing us to predict the credit rank for 2018Q1 using the dataset covering the period from 2016Q2 to 2017Q4. Note that to spatial-only methods, we explore the data of the last quarter as input to predict the credit rank of next quarter. Adopting this modeling approach, we then assessed both the accuracy and macro-F_1 of credit rating predictions from 2018Q1 to 2023Q1. Moreover, we evaluate the performance of all the methods by calculating the gap between accuracy and macro-F_1 and show the performance in boxplot. The results of this evaluation are depicted in Table I and Figs. 3 and 4.

From the experimental results, it is evident that our HFTCRNet surpasses other baseline models in terms of accuracy and macro-F_1 scores across most quarters. Moreover, when compared with the second-best method in Table I, i.e., HSTGCNT, our HFTCRNet demonstrates improvements of 1.9% and 3.2% in accuracy and macro-F_1 scores, respectively, while maintaining a comparable standard deviation. These results highlight the robustness and stability of our model, which is specifically tailored for interbank risk assessment. HFTCRNet effectively captures the characteristics of bank risk, leading to more precise credit ratings. In addition, by comparing the



(a). R^2 for SRISK values



(b). High-Risk SRISK

Fig. 7. (a) Effect of semi-supervised learning toward R^2 score. (b) Variance of high-risk SRISK along different quarters.

performance gap between accuracy and macro-f1, we can also observe that our HFTCRNet is more stable than other baselines. This indicates the balanced performance of our method across different credit ranks, especially when confronted with an imbalanced credit rank distribution. For spatial-only methods, we observe that TGAR surpasses other methods (GCN, GAT, and SAGNN). This superiority can be attributed to the transformed graph attention and hyper context extraction used in the representation of the interbank loan network. Regarding temporal–spatial methods, HSTGCNT stands out against other baselines. The reason for this is that HSTGCNT integrates both long-term temporal features and short-term spatial–temporal features. Our HFTCRNet not only extracts temporal information from both long-term and short-term perspectives but also incorporates the contagion chain in the spatial domain. This makes it particularly well-suited for interbank credit rating leading to our superior performance.

C. DGraph-Fin

In Table I, we provide a detailed comparison of the experimental results obtained using the DGraph-Fin dataset. We adhere to the original evaluation metrics specified by DGraph-Fin, focusing on the receiver operating characteristic-area under curve (ROC-AUC) scores to benchmark against other baseline models. Our analysis reveals that our model outperforms all the baselines in terms of ROC-AUC scores. This is particularly noteworthy considering that DGraph-Fin lacks risk contagion chains, which are a significant component in the interbank dataset we have studied. Despite this, our model demonstrates a robust ability to detect and analyze the structural nuances and temporal dynamics inherent in the graph, leading to effective identification of fraudulent users within financial contexts. This capability is largely attributable to the sophisticated architectural design of our model, particularly the LT^3 module and the STCGT module, which excel in capturing complex patterns and changes over time. The performance of these modules underscores their potential in enhancing fraud detection mechanisms in dynamic temporal graph, highlighting the effectiveness of our approach in tackling real-world challenges.

TABLE IV

AVERAGE R^2 OF SRISK VALUE WITH DIFFERENT SUPERVISED DATA USAGE RATIOS OF SRISK DATA WITH AND WITHOUT SEMI-SUPERVISED LEARNING

Ratio	100%	80%	60%	40%	20%
w/ semi	0.901	0.895	0.890	0.879	0.841
w/o semi	0.826	0.812	0.792	0.741	0.682

TABLE V

ACCURACY WITH REGARDING DIFFERENT SRISK VALUE. ACC IS THE ACCURACY. GAP IS THE ACCURACY GAP BETWEEN EACH SETTING AND ALL BANKS

Setting	All	Top-10%	Top-20%	Top-30%	Top-40%
Acc	0.678	0.712	0.704	0.699	0.691
Gap	0.00	0.034	0.026	0.021	0.013

D. Ablation Study

To better understand the contribution of different components to credit rating, we conduct ablative studies on our proposed model, which includes four primary components: LT^3, STCGT, ARCT, and HFT. We also integrate semi-supervised learning during the training phase. We evaluate five model variants to assess the influence of each component: (w/o) LT^3, (w/o) STCGT, (w/o) ARCT, (w/o) HFT, and (w/o) Semi, as detailed in Table II. The removal of each component tests its individual impact: (w/o) LT^3, (w/o) STCGT, and (w/o) ARCT involve bypassing their outputs directly to the HFT module, whereas (w/o) HFT aggregates outputs of LT^3, STCGT, and ARCT directly. The (w/o) semi variant uses only annotated labels for optimizing $\mathcal{L}_{sv}$ and $\mathcal{L}_{sr}$. The results show that LT^3 is the most influential, highlighting the significance of understanding bank growth trajectories. Both STCGT and ARCT are crucial for capturing short-term variances and contagion risks in the interbank loan network, enhancing credit rating predictions. While HFT does not introduce new features, it significantly boosts performance by effectively integrating diverse data sources. Semi-supervised learning provides marginal gains but offers valuable insights in systemic risks, refer to Section V-E.

We further conducted an ablation study to evaluate the performance of various modules from 2021Q1 to 2023Q1, as detailed in Table III. STCGT-GCN and STCGT-GAT refer to modifications where we replace the graph transformer in STCGT with GCN and GAT, respectively. ARCT (largest) denotes using the largest loan value in each contagion chain as the chain's risk value. Graph (MinDen) uses the minimum density algorithm to generate interbank relationships. Removing the SRISK value and SRISK ratio from the training objectives is indicated by (w/o) SRISKV and (w/o) SRISKR, respectively. The results highlight the importance of the graph transformer for capturing short-term temporal relationships due to its robust representational ability, achieving improvements of 2.1% and 1.0% in macro-F_1 over GCN and GAT. Using the largest loan value as the risk measure in ARCT (largest) leads to a performance drop, as our original strategy effectively captures risk from the start to the end of the chain, introducing varied risk patterns into the ARCT. Conversely, ARCT (largest) focuses on the risk from any node to the end, merging different patterns, which confuses the learning process. Besides, both the SRISK value and ratio prove vital for credit rating prediction, as they help measure systemic risk through absolute and relative risk metrics.

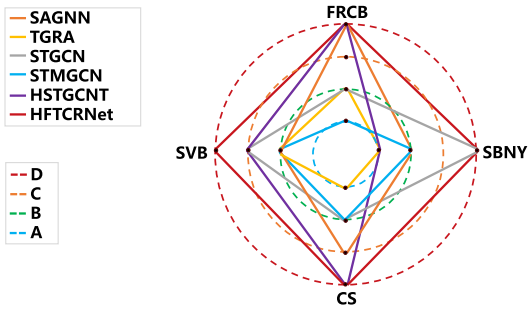


Fig. 8. Predicted credit rank of SVB, SBNY, FRCB, and CS on 2023Q1.

During training, we use four distinct training objectives, managed by three hyperparameters, λ_1, λ_2, and λ_3, to maintain balance among them. To explore the impact of these training objectives further, we varied λ_1, λ_2, and λ_3 across a set of values {0, 0.5, 1.0, 1.5, 2.0} and documented the results in Fig. 5. The experimental results reveal that $\mathcal{L}_{fs}$ is the most critical objective, and selecting appropriate values for λ_1, λ_2, and λ_3 significantly enhances performance, particularly in terms of macro-F_1 score.

In addition, we explore the impact of high-risk contagion chains and the dynamics of temporal interbank loan networks on credit rating accuracy. For high-risk contagion chains, we adjust the number of chains N_c and their maximum length L_c within a range of 4–8. The influence of these parameters on credit rating accuracy is illustrated in a heat-map shown in Fig. 6(a), which indicates that using six contagion chains of up to six links each optimizes accuracy. Longer chains tend to introduce noise, reducing risk assessment effectiveness. Regarding the temporal interbank loan networks, we varied the sequence length from 2 to 8, with the resulting accuracy changes depicted in Fig. 6(b). We observe that increasing the sequence length improves credit rating performance up to a length of 4, beyond which the performance levels off. This finding highlights the importance of capturing growth trajectories within a year for precise credit rating assessments.

E. Risk Assessment

In our training process, we optimize the SRSIK metrics along with the credit rating. To assess the effectiveness of our approach in predicting SRISK, we first compute the R^2 score comparing the predicted and actual SRISK values for the subsequent quarter. The resulting scores, plotted in Fig. 7(a), demonstrate that all the R^2 values exceed 0.85, indicating the superior performance of our method. To assess the impact of semi-supervised learning on SRISK prediction, we conduct experiments with varying ratios of annotated SRISK data, both with and without semi-supervised learning, and present the results in Table IV. The experimental findings show that the R^2 metric improves as the ratio of supervised data usage increases from 20% to 100%, highlighting the value of annotated data. In addition, across different supervised data usage ratios, the semi-supervised learning approach consistently enhances R^2 significantly, confirming its effectiveness. Higher R^2 values facilitate a more precise estimation of systemic risk, which can further contribute to mitigating risk propagation. Furthermore, a comparison of credit rating accuracy across all banks and those with high SRISK reveals an interesting pattern, as shown in Table V. Our methodology consistently achieves superior

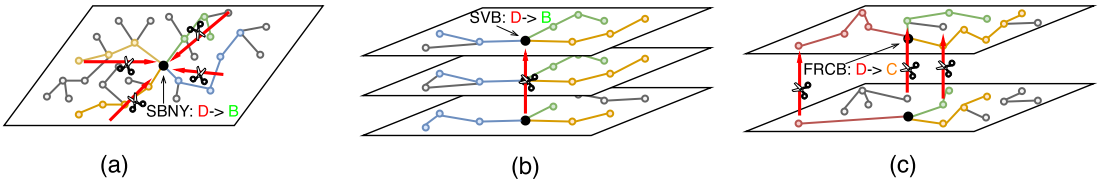


Fig. 9. Visualization of the effect of how different modules of our HFTCRNet affect the credit rank prediction on 2023Q1. (a) HFTCRNet w/o ARCT. (b) HFTCRNet w/o LT^3. (c) HFTCRNet w/o STCGT.

TABLE VI
ACCURACY AND MACRO-F_1 OF HFTCRNET WITH DIFFERENT ASSETS AND LIABILITIES OF CREDIT RATING PREDICTING

Setting	Accuracy	Macro-F_1
All	0.678 ± 0.104	0.630 ± 0.110
[0, 25%]	0.698 ± 0.098	0.648 ± 0.101
(25%, 50%]	0.688 ± 0.112	0.638 ± 0.112
(50%, 75%]	0.667 ± 0.109	0.624 ± 0.115
(75%, 100%]	0.659 ± 0.108	0.612 ± 0.114

accuracy rates with high SRISK banks. Considering that credit ratings mirror the default risk of banks, and SRISK values gauge the potential impact on the financial market in the event of bank default, a combination of these two metrics offers a robust evaluation of systemic risk. For a holistic assessment of the systemic risk in the interbank financial ecosystem, we propose a metric termed "high-risk SRISK." The computation for this metric involves the following steps: 1) identifying banks that fall within the top 10% bracket in terms of SRISK and having a predicted credit rank of "D" and 2) aggregating the predicted SRISK values of these identified banks. The high-risk SRISK thus derived, charted over consecutive quarters, is illustrated in Fig. 7. This metric, when juxtaposed against both predicted and actual credit ratings, showcases congruent variance patterns. Notably, heightened levels of high-risk SRISK are observed in 2018Q3 and 2023Q1, aligning with real-world.

F. Case Study

In this empirical evaluation, we delve deep into the performance of our proposed HFTCRNet with other baseline models. Notably, we focus on insights unearthed through pertinent case studies centered around the 2023 banking crisis. As part of this endeavor, we spotlight four banks: SVB, SBNY, FRCB, and CS. In the first quarter of 2023, all these banks declared bankruptcy, plummeting their credit rating to the nadir, i.e., "D." This is especially intriguing as just the previous quarter, in 2022Q4, these banks boasted substantially higher credit ranks. The pertinent question then arises: Can various modeling techniques accurately predict this drastic downturn in 2023Q1? The answers are collated in Fig. 8. It is important to highlight that to minimize the influence of randomness from a single experiment on the presented credit ratings, we use different random seeds to train all the models. Specifically, each model involved in the comparison is trained ten times with different random seeds, and the credit ratings are determined using a majority voting strategy, i.e., the risk rank that is predicted the most times are the predicted rank. Fig. 8 reveals that while baselines manage to partially anticipate the credit ranks of these banks, our HFTCRNet is the only model that forecasts the trajectories of four banks.

To unravel the driving factors behind superior performance of our HFTCRNet, we conduct an evaluation akin to that in Section V-D, dissecting the contribution of its intrinsic modules. In our analysis presented in Fig. 9, we investigate the influence of the ARCT, LT^3, and STCGT modules on the credit ratings of SBNY, SVB, FRCB, and CS. We observe that omitting any of the ARCT, LT^3, or STCGT modules leads to inaccurate predictions for SBNY, SVB, and FRCB, respectively, while the credit rank prediction for CS remains unaffected. Delving deeper into the unique attributes of these banks, we identify distinct characteristics. For SBNY, a higher number of contagion chains and longer chain lengths are notable. In contrast, SVB demonstrates stable connections over time, yet exhibits significant variations in its financial status. FRCB, on the other hand, shows considerable changes in connection status between 2022Q4 and 2023Q1. These observations highlight the specific contributions of each module: the pivotal role of ARCT in capturing risk within contagion chains is more pronounced for SBNY, LT^3's focus on financial variability is crucial for SVB, and the capability of STCGT to discern variances in adjacent temporal graphs is relevant for FRCB.

G. Limitation

During the analysis of experimental results, we rank all banks based on the total sum of their assets and liabilities. We then segment them into four quartiles, i.e., [0%, 25%], (25%, 50%], (50%, 75%], and (75%, 100%], and calculate the accuracy and Macro-F_1 of credit rating predictions for each quartile, as depicted in Table VI. The experimental results indicate that our model provides accurate credit rating for banks in the higher asset and liability brackets, i.e., [0%, 50%]. In contrast, the predictions for banks in the lower brackets, i.e., [50%, 100%], are less accurate, highlighting areas where credit rating predictions could be enhanced for banks with lower assets and liabilities. The performance discrepancies observed are largely attributable to the greater impact that fundamental economic indicators and interbank relationships have on the risk assessments of banks with lower asset and liability totals. Looking ahead, key research directions will include enhancing the dataset with more comprehensive economic indicators, developing a more effective interbank network, and improving the robustness and accuracy of credit rating prediction.

VI. CONCLUSION

In this article, we have focused on the bank credit rating and risk assessment from the perspective of temporal interbank dynamics. To accomplish this task, we have proposed a novel credit rating and risk assessment learning system, named as HFTCRNet, which includes the LT^3 module, STCGT module, ARCT module, and HFT module to capture and integrate the interbank relationship from long-term growth trajectories, short-term interbank relationship variance, and spatial

high-risk contagion chain. Besides, we have proposed a novel interbank credit rating dataset, which incorporates quarterly financial data of banks, temporal interbank loan networks, and some vital labels (e.g., credit rating and SRISK) from 2016Q1 to 2023Q1 for a total of 4548 banks. To the best of our knowledge, this is the first work that addresses the bank credit rating and risk assessment by the temporal interbank dynamic. Moreover, we optimize our framework in a semi-supervised manner, which not only uses multiple objective labels annotated with different annotation sparsity but also demonstrates a new method to evaluate the systemic risk of the interbank financial system. Extensive experiments on our interbank dataset demonstrate the efficacy of HFTCRNet on temporal interbank dynamics.

REFERENCES

[1] V. Macchiati, G. Brandi, T. Di Matteo, D. Paolotti, G. Caldarelli, and G. Cimini, "Systemic liquidity contagion in the European interbank market," *J. Econ. Interact. Coordination*, vol. 17, no. 2, pp. 443–474, Apr. 2022.

[2] E. Nier, J. Yang, T. Yorulmazer, and A. Alentorn, "Network models and financial stability," *J. Econ. Dyn. Control*, vol. 31, no. 6, pp. 2033–2060, 2007.

[3] M. K. Brunnermeier, "Deciphering the liquidity and credit crunch 2007–2008," *J. Econ. Perspect.*, vol. 23, no. 1, pp. 77–100, 2007.

[4] G. Sheldon and M. Maurer, "Interbank lending and systemic risk: An empirical analysis for Switzerland," *Revue Suisse D Economie Politique Et De Statistique*, vol. 134, pp. 685–704, Dec. 1998.

[5] H. Elsinger, A. Lehar, and M. Summer, "Network models and systemic risk assessment," *Handbook Systemic Risk*, vol. 1, no. 1, pp. 287–305, 2013.

[6] K. Anand, B. Craig, and G. von Peter, "Filling in the blanks: Network structure and interbank contagion," *Quant. Finance*, vol. 15, no. 4, pp. 625–636, Apr. 2015.

[7] C. Z. Liu, S. Xiang, D. Cheng, J. Liu, Y. Zhang, and L. Qin, "Transformed graph attention for credit rating," in *Proc. IEEE 18th Conf. Ind. Electron. Appl. (ICIEA)*, Aug. 2023, pp. 1011–1016.

[8] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2016, pp. 1–14.

[9] P. Zappa and D. Q. Vu, "Markets as networks evolving step by step: Relational event models for the interbank market," *Phys. A, Stat. Mech. Appl.*, vol. 565, Mar. 2021, Art. no. 125557.

[10] C. Brownlees and R. F. Engle, "SRISK: A conditional capital shortfall measure of systemic risk," *Rev. Financial Stud.*, vol. 30, no. 1, pp. 48–79, Jan. 2017.

[11] C.-H. Shen, Y.-L. Huang, and I. Hasan, "Asymmetric benchmarking in bank credit rating," *J. Int. Financial Markets, Institutions Money*, vol. 22, no. 1, pp. 171–193, Feb. 2012.

[12] T. Squartini, G. Cimini, A. Gabrielli, and D. Garlaschelli, "Network reconstruction via density sampling," *Appl. Netw. Sci.*, vol. 2, no. 1, pp. 1–13, Dec. 2017.

[13] T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli, and D. Garlaschelli, "Reconstruction methods for networks: The case of economic and financial systems," *Phys. Rep.*, vol. 757, pp. 1–47, Oct. 2018.

[14] A. Gabrielli, V. Macchiati, and D. Garlaschelli, "Critical density for network reconstruction," in *From Computational Logic to Computational Biology*. Cham, Switzerland: Springer, 2023.

[15] J. Liu, D. Cheng, and C. Jiang, "Preventing attacks in interbank credit rating with selective-aware graph neural network," in *Proc. 32nd Int. Joint Conf. Artif. Intell.*, Aug. 2023, pp. 6085–6093.

[16] A. Petropoulos, V. Siakoulis, P. Lazaris, and S. Chatzis, "Re-constructing the interbank links using machine learning techniques. An application to the Greek interbank market," *Intell. Syst. Appl.*, vol. 12, Nov. 2021, Art. no. 200055.

[17] Z. Feinstein and A. Sojmark, "Dynamic default contagion in heterogeneous interbank systems," *SIAM J. Financial Math.*, vol. 12, no. 4, pp. 83–97, 2021.

[18] S. Xiao, T. Bai, X. Cui, B. Wu, X. Meng, and B. Wang, "A graph-based contrastive learning framework for medicare insurance fraud detection," *Frontiers Comput. Sci.*, vol. 17, no. 2, Apr. 2023, Art. no. 172341.

[19] D. Cheng, Z. Niu, Y. Tu, and L. Zhang, "Prediction defaults for networked-guarantee loans," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 361–366.

[20] A. Barja et al., "Assessing the risk of default propagation in interconnected sectoral financial networks," *EPJ Data Sci.*, vol. 8, no. 1, p. 32, Dec. 2019.

[21] M. Jager, T. Siemsen, and J. Vilsmeier, "Interbank risk assessment—A simulation approach," *Deutsche Bundesbank Discuss.*, vol. 4, no. 1, pp. 1–18, Apr. 2020.

[22] S. Song and H. Li, "Unveiling early warning signals of systemic risks in banks: A recurrence network-based approach," 2023, *arXiv:2310.10283*.

[23] P. Laux, W. Qian, and H. Zhang, "Learning from lending in the interbank network," *Data Sci. Sci.*, vol. 2, no. 1, Dec. 2023, Art. no. 2151949.

[24] X. Li and L. Li, "A deep learning model-based approach to financial risk assessment and prediction," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, pp. 1–13, Jan. 2024.

[25] R. Liu and C. S. Pun, "Machine-learning-enhanced systemic risk measure: A two-step supervised learning approach," *J. Banking Finance*, vol. 136, Mar. 2022, Art. no. 106416.

[26] T. Kristóf and M. Virág, "EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks," *Res. Int. Bus. Finance*, vol. 61, Oct. 2022, Art. no. 101644.

[27] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[28] V. Balmaseda, M. Coronado, and G. de Cadenas-Santiago, "Predicting systemic risk in financial systems using deep graph learning," *Intell. Syst. Appl.*, vol. 19, Sep. 2023, Art. no. 200240.

[29] Z. Cao, Z. Chen, P. Mishra, H. Amini, and Z. Feinstein, "Modeling inverse demand function with explainable dual neural networks," 2023, *arXiv:2307.14322*.

[30] D. Cheng, Z. Niu, and Y. Zhang, "Contagious chain risk rating for networked-guarantee loans," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 2715–2723.

[31] D. Cheng, Z. Niu, J. Li, and C. Jiang, "Regulating systemic crises: Stemming the contagion risk in networked-loans through deep graph learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 6, pp. 6278–6289, Jun. 2023.

[32] Y. Zheng et al., "Correlation-aware spatial–temporal graph learning for multivariate time-series anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 9, pp. 11802–11816, Sep. 2024.

[33] M. Saffari, M. Khodayar, M. E. Khodayar, and M. Shahidehpour, "Behind-the-meter load and PV disaggregation via deep spatiotemporal graph generative sparse coding with capsule network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 10, pp. 14573–14587, Oct. 2024.

[34] I. Marisca, C. Alippi, and F. M. Bianchi, "Graph-based forecasting with missing data through spatiotemporal downsampling," in *Proc. 41st Int. Conf. Mach. Learn.*, 2024, pp. 1–20.

[35] A. Cini, D. Zambon, and C. Alippi, "Sparse graph learning from spatiotemporal time series," *J. Mach. Learn. Res.*, vol. 24, no. 242, pp. 1–36, 2023.

[36] F. Chen, Y. Zhang, L. Chen, X. Meng, Y. Qi, and J. Wang, "Dynamic traveling time forecasting based on spatial-temporal graph convolutional networks," *Frontiers Comput. Sci.*, vol. 17, no. 6, Dec. 2023, Art. no. 176615.

[37] C. Chen, K. Li, S. G. Teo, X. Zou, and Z. Zeng, "Gated residual recurrent graph neural networks for traffic prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 485–492.

[38] A. Vaswani et al., "Attention is all you need," in *Proc. NeurIPS*, 2017, pp. 1–11.

[39] W. Shao et al., "Long-term spatio-temporal forecasting via dynamic multiple-graph attention," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, Jul. 2022, pp. 2225–2232.

[40] Z. Ouyang, M. Jabloun, and P. Ravier, "Rankformer: Leveraging rank correlation for transformer-based time series forecasting," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jul. 2023, pp. 85–89.

[41] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[42] X. Geng et al., "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proc. AAAI*, 2019, pp. 3656–3663.

[43] G. Huo, Y. Zhang, B. Wang, J. Gao, Y. Hu, and B. Yin, "Hierarchical spatio–temporal graph convolutional networks and transformer network for traffic flow forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 3855–3867, Apr. 2023.

[44] C. Chen, Y. Liu, L. Chen, and C. Zhang, "Bidirectional spatial–temporal adaptive transformer for urban traffic flow forecasting," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 6913–6925, Oct. 2023.
[45] H. Lee and S. Ko, "TESTAM: A time-enhanced spatio-temporal attention model with mixture of experts," in *Proc. ICLR*, 2024, pp. 1–19.
[46] Y. Tian, Y. Qi, and F. Guo, "FreeDyG: Frequency enhanced continuous-time dynamic graph model for link prediction," in *Proc. ICLR*, 2023, pp. 1–20.
[47] J. Su, D. Zou, and C. Wu, "PRES: Toward scalable memory-based dynamic graph neural networks," in *Proc. ICLR*, 2024, pp. 1–29.
[48] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9713–9729, Jul. 2020.
[49] D. Cheng, F. Yang, S. Xiang, and J. Liu, "Financial time series forecasting with multi-modality graph neural network," *Pattern Recognit.*, vol. 121, Jan. 2022, Art. no. 108218.
[50] Z. Zeng, R. Kaur, S. Siddagangappa, S. Rahimi, T. Balch, and M. Veloso, "Financial time series forecasting using CNN and transformer," 2023, *arXiv:2304.04912*.
[51] Q. Tao, R. Cai, Z. Lin, and Y. Tang, "Automatic design of deep graph neural networks with decoupled mode," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 14, 2024, doi: 10.1109/TNNLS.2024.3438609.
[52] Z. Ke, H. Yu, J. Li, and H. Zhang, "DUPLEX: Dual GAT for complex embedding of directed graphs," in *Proc. 41st Int. Conf. Mach. Learn.*, 2024, pp. 1–19.
[53] J. Hu, B. Hooi, and B. He, "Efficient heterogeneous graph learning via random projection," *IEEE Trans. Knowl. Data Eng.*, early access, Aug. 21, 2024, doi: 10.1109/TKDE.2024.3434956.
[54] B. Jiang, B. Wang, S. Chen, J. Tang, and B. Luo, "Graph neural network meets sparse representation: Graph sparse neural networks via exclusive group lasso," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 12692–12698, Oct. 2023.
[55] Y. Chong, Y. Ding, Q. Yan, and S. Pan, "Graph-based semi-supervised learning: A review," *Neurocomputing*, vol. 408, pp. 216–230, Sep. 2020.
[56] X. Kong and P. S. Yu, "Semi-supervised feature selection for graph classification," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2010, pp. 793–802.
[57] C. Zhuang and Q. Ma, "Dual graph convolutional networks for graph-based semi-supervised classification," in *Proc. World Wide Web Conf.*, 2018, pp. 499–508.
[58] Z. Hao et al., "ASGN: An active semi-supervised graph neural network for molecular property prediction," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 731–752.
[59] J. Li, Y. Rong, H. Cheng, H. Meng, W. Huang, and J. Huang, "Semi-supervised graph classification: A hierarchical graph perspective," in *Proc. World Wide Web Conf.*, May 2019, pp. 972–982.
[60] G. Zhang, Z. Hu, G. Wen, J. Ma, and X. Zhu, "Dynamic graph convolutional networks by semi-supervised contrastive learning," *Pattern Recognit.*, vol. 139, Jul. 2023, Art. no. 109486.
[61] J. Lei Ba, J. Ryan Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.
[62] Y. Shi, Z. Huang, S. Feng, H. Zhong, W. Wang, and Y. Sun, "Masked label prediction: Unified message passing model for semi-supervised classification," 2020, *arXiv:2009.03509*.
[63] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous graph transformer," in *Proc. WWW*, Apr. 2020, pp. 2704–2710.
[64] J. R. Aronson and J. R. Marsden, "Duplicating Moody's municipal credit ratings," *Public Finance Quart.*, vol. 8, no. 1, pp. 97–106, Jan. 1980.
[65] X. Huang et al., "DGraph: A large-scale financial dataset for graph anomaly detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 22765–22777.
[66] A. Paszke et al., "Automatic differentiation in PyTorch," in *Proc. NeurIPS*, 2017, pp. 1–4.
[67] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

Jiangtong Li received the B.E. and Ph.D. degrees from Shanghai Jiao Tong University, Shanghai, China, in 2019 and 2024, respectively.

He is currently a Post-Doctoral Associate with the School of Computer Science and Technology, Tongji University, Shanghai. His current research interests include multimodal modeling, large language model, graph learning, and big data in finance.

Ziyuan Zhou is currently pursuing the bachelor's degree with the School of Economics and Management, Tongji University, Shanghai, China.

Her research interests include data mining, operation research, optimization theory, and graph learning.

Jingkai Zhang is currently pursuing the bachelor's degree in software engineering with the School of Software, Tongji University, Shanghai, China.

His research interests include graph learning, data mining, and big data.

Dawei Cheng received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2018.

He was a Post-Doctoral Associate with the MoE Key Laboratory of Artificial Intelligence, Department of Computer Science, Shanghai Jiao Tong University. He is currently an Associate Professor with the Department of Computer Science and Technology, Tongji University, Shanghai. His research fields include data mining, graph learning, and big data in finance.

Changjun Jiang received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1995.

He is currently a Professor with the Department of Computer Science and Technology, Tongji University, Shanghai, China. His current research interests include concurrence theory, formal verification of software, service-oriented computing, big data in finance, intelligent systems, financial risk management, and big data computing.