Contagious Chain Risk Rating for Networked-guarantee Loans

Dawei Cheng
Department of Computer Science and Engineering,
Shanghai Jiao Tong University
Shanghai, China
dawei.cheng@sjtu.edu.cn

Zhibin Niu*
College of Intelligence and Computing,
Tianjin University
Tianjin, China
zniu@tju.edu.cn

Yiyi Zhang
Department of Computer Science and Engineering,
Shanghai Jiao Tong University
Shanghai, China
yi95yi@sjtu.edu.cn

ABSTRACT

The small and medium-sized enterprises (SMEs) are allowed to guarantee each other and form complex loan networks to receive loans from banks during the economic expansion stage. However, external shocks may weaken the robustness, and an accidental default may spread across the network and lead to large-scale defaults, even systemic crisis. Thus, predicting and rating the default contagion chains in the guarantee network in order to reduce or prevent potential systemic financial risk, attracts a grave concern from the Regulatory Authority and the banks. Existing credit risk models in the banking industry utilize machine learning methods to generate a credit score for each customer. Such approaches dismiss the contagion risk from guarantee chains and need extensive feature engineering with deep domain expertise. To this end, we propose a novel approach to rate the risk of contagion chains in the bank industry with the deep neural network. We employed the temporal inter-chain attention network on graph-structured loan behavior data to compute risk scores for the contagion chains. We show that our approach is significantly better than the state-of-the-art baselines on the dataset from a major financial institution in Asia. Besides, we conducted empirical studies on the real-world loan dataset for risk assessment. The proposed approach enabled loan managers to monitor risks in a boarder view and avoid significant financial losses for the financial institution.

CCS CONCEPTS

• **Applied computing → Economics**; • **Information systems → Data mining**.

KEYWORDS

Contagion chain; loan network; risk assessment; data mining

ACM Reference Format:
Dawei Cheng, Zhibin Niu, and Yiyi Zhang. 2020. Contagious Chain Risk Rating for Networked-guarantee Loans. In *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining USB Stick (KDD '20), August 23–27, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3394486.3403322

*Corresponding author.

KDD '20, August 23–27, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7998-4/20/08...$15.00
https://doi.org/10.1145/3394486.3403322

1 INTRODUCTION

Networked-guarantee loans, a widespread economic phenomenon in Asia countries, are attracting increasing attention from the regulators and banks. The small and medium enterprises who lack financial security are allowed to guarantee each other and obtain loans from commercial banks [13, 20]. During the credit expansion period, with more and more enterprises involved, they form complex directed networks [32, 34]. Usually, the guaranteed loan has a debt obligation contract. This means if one corporation fails to repay to the bank, the guarantor has to pay for it, and this leads a risk spreading across the guarantee network [17]. Usually, the accidental default is tolerable while the large-scale defaults or systemic financial crisis is to be firmly prevented. The research of contagion risk for the guaranteed loans is still relatively limited. Monitoring the financial risk status, especially the dynamics, is so difficult and in practice, the financial experts can only study the independent cases after a capital chain rupture. During the economic slowdown period, the need for loan risk management, especially for the monitoring and evaluation of the contagious risk, is more urgent than ever before.

The banking industry has developed credit risk models for each loan applicant since the middle of the twentieth century [4]. The risk rating is also the main business of thousands of worldwide corporations, including dozens of public companies [7]. They have kept these models state-of-the-art by investing millions of dollars [3]. However, global loan default losses still amounted to over 50 billion US dollars in 2018 and are forecasted to continue to increase. This massive amount of losses has increased the importance of risk curbing. Traditionally, the credit scoring models are built using regression algorithms with the temporal credit or loan history as well as some aggregated financial information of the applicants. Shallow learning methods, including the classic logistic regression [27], neural network [1], etc., are extensively utilized to obtain the credit score of the applicants.

Although these approaches are widely used and useful, we also observed that they have some limitations (L) for networked-guarantee loans. L1: the current credit scoring targets on individual-level risk rating, which relies on the borrower's historical data. But for guaranteed loans, the risk may diffuse across the networks. It is still a blank area to estimate the risk of contagion chain risk. L2: the existing model requires extensive feature constructing and specific background knowledge to design representative features. Many features need to be aggregated from historical records, which is highly time-consuming. L3: if the borrower does not have a sufficient credit or loan historical record, it is difficult to estimate the credit level or scores reliably for the company or the chain.

Regarding these issues, we propose a novel risk rating approach, named as TempoRal Attention Contagion chain Enhanced Rating model (TRACER[1]). TRACER is designed to estimate the risk of contagion chains in the bank industry based on deep learning methods. In particular, we design an inter-chain attention network for learning from graph-structured loan behavior data directly, which requires no heavy handcraft feature engineering or in-depth domain knowledge. The attention mechanism is employed to adjust the importance of inter-chains automatically. With the attentional layer, we could allocate optimal rating for chains with limited credit history, by learning from their most relevant neighbors in high-order feature spaces. Afterwards, we introduce a recurrent and rating neural network to infer the risk probability of contagion chain from sequential loan behaviors. Extensive experiments demonstrate the significant performance gain compared with existing baselines. Besides, we conduct empirical studies on over six months' records using the proposed approach. The result shows that TRACER could avoid significant potential financial losses for the bank.

The main contributions of the paper are summarized as follows:

(1) We pioneer the research of contagion chain risk rating in the networked-guarantee loans. Our work fills the blank area of systemic risk estimation in networked-loans, which is urgent than ever before in pessimistic economic situations.
(2) We design and implement the TempoRal Attention Contagion chain Enhanced Rating model (TRACER), which enables the model to learn from networked behavior data directly. We also propose inter-chain attention and recurrent based deep architecture and demonstrate the effectiveness of the model in overcoming three major shortcomings of existing credit model in the banking industry.
(3) We thoroughly evaluate the proposed approach by comparing with the existing benchmarks on the historical loan data and achieved state-of-the-art performance. In addition, we conduct empirical studies in real-world risk control applications, and the result proves our method could prevent major financial losses for the financial institution.

The rest of the paper is organized as: Section 2 describes the business background and data description. Section 3 shows the proposed model in details. We report the experiment results and case studies in Section 4. Section 5 surveys the related work. Conclusion and discussion are described in Section 6.

2 PRELIMINARIES

2.1 Financial Background

The guaranteed-loan, as a special secured loan model, is allowed in some countries to solve the problem of financing difficulty for the SMEs. In general, the borrower needs to provide all the detailed information similar to the fixed asset mortgage to the banks for credit evaluation. However, in practice, most of the small and medium enterprises who are in the stage of rapid expansion, are difficult to obtain loan funds from the banks. The reason might because small businesses usually could not meet the high lending criteria of banks, which are originally designed for large-scale industries.

[1]It indicates the approach enables risk managers to trace potential default losses across loan networks, named "tracer" by the meaning of "tracing risk from contagion chains".

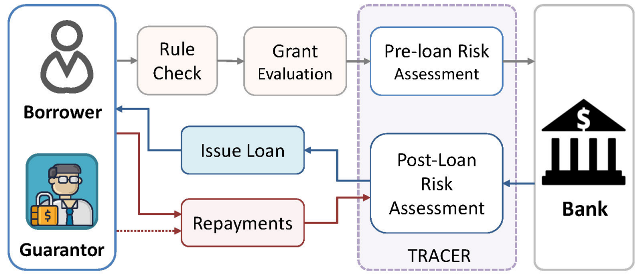


Figure 1: The business procedure of the loan management.

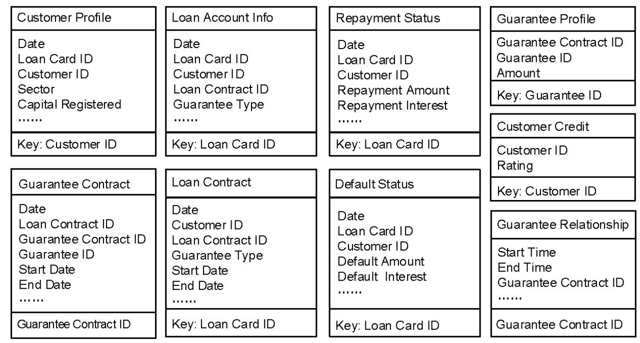


Figure 2: The description of loan data. There are nine table records covering from loan records, guarantee status, user basic profile to default events.

Several approaches, including the network-guaranteed loans, are designed and implemented to perfect the credit supporting system.

Figure 1 shows a typical procedure for network-guaranteed loans. It includes five modules: 1) the borrower finds several guarantors to provide credit guarantee and signs the contracts with the banks; 2) the banks perform a pre-loan risk assessment, and if passed 3) the borrower receives the funds and repay the interests and principal (or partial) regularly according to the loan contract; 4) the bank monitors the repayment status and conducts a post-loan risk assessment; 5) suppose the borrower fails to repay the rest part of the loan, its guarantors have to pay as the contract address. In this procedure, our proposed TRACER are employed to measure the potential systemic risk of contagion chains, which is proceeded in both the pre-loan assessment and post-loan control. The alarming high-risk chains will be escalated to the chief account managers for taking appropriate measures to prevent large-scale default.

2.2 Data Description

Figure 2 shows an overview of the data utilized in our approach, which is stored tabularly in the database system. There are nine tables, including customer profile, loan account information, guarantee contract, guarantee profile, loan contract, guarantee relationship, customer credit, repayment status and default status. As strict compliance with the requirements of data protection legislation was a top priority for the financial institution in the implementation of the project, we cannot describe our data with specific elements and make them available to readers.

It should be noted that our approach does not require handcraft feature engineering. We construct the raw tabular loan behaviors into graph-structured sequential formats, which can be learned

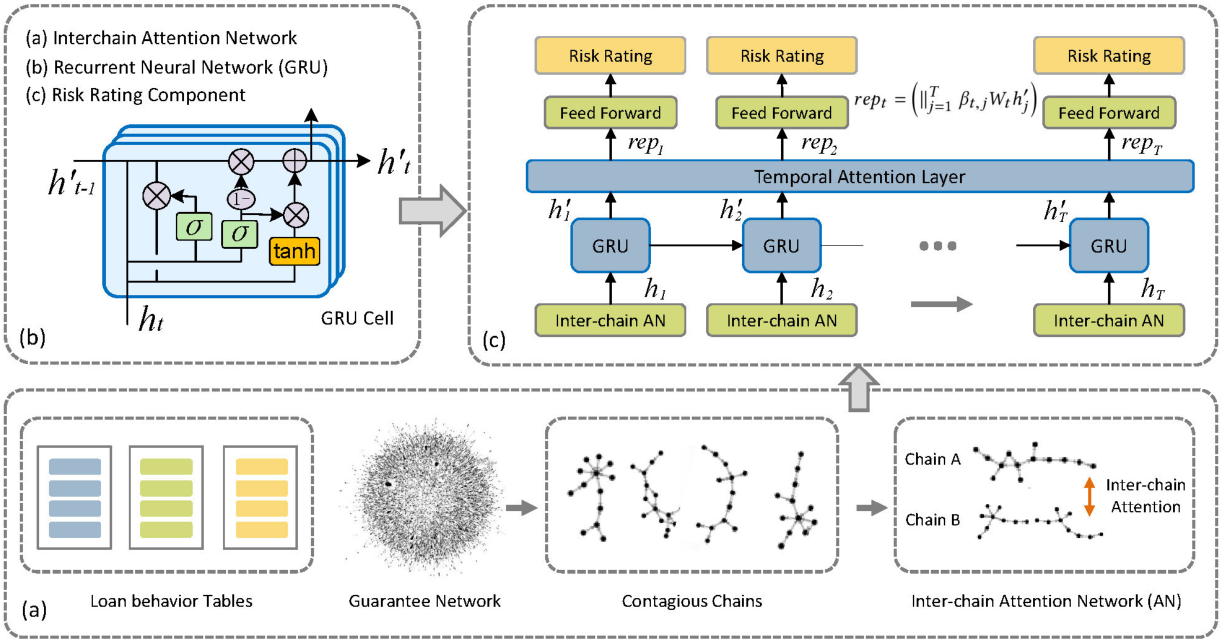


Figure 3: TRACER: The architecture of the proposed contagion chain risk rating model. (a) illustrates procedure of contagion chain construction and the inter-chain attention network; (b) shows GRU based recurrent neural network; (c) displays the risk rating component which takes the high-order representations learned from GRU layer and inter-chain attention network.

directly by graph neural network layer of the proposed TRACER. This graph construction step, the risk rating procedure, and the prediction through deep graph neural network (does not include the training time), can be processed in near real-time, which is one of the advantages of our proposed approach.

3 THE PROPOSED TRACER

In this section, we introduce the architecture of the proposed approach first and then present the procedure to construct the contagion chains from raw tabular data. Finally, we introduce each component of the model and the loss function of risk rating module.

3.1 Architecture Overview

Figure 3 shows the general architecture of the proposed TRACER for contagion chain risk prediction in networked-guarantee loans. Generally, the model includes three parts: 1) Inter-chain attention network, which takes tabular loan behavior records as inputs and proceeds a high-order representative feature of contagion chains in each timestamp as outputs. In particular, we construct the attributed loan networks from loan behaviors and then mining the corresponding contagion chains. If two chains share the same connected network, which means they are produced from the same network, we treat them as affinal chains so that we could produce attention mechanism on each other. It means that the chain's representation is not only relied on the attributes of inner nodes but also depended on the nodes from affinal chains. The learned representations from inter-chain attention network are fed as inputs for down steam components. 2) The recurrent neural network takes the high-order inter-chain attentional feature as input sequences and leverage temporal self-attention to learn sequential representations as the outputs. 3) The risk rating module takes the structural chain and temporal self-attentional representations as inputs to infer the probability of risks for a given chain in each timestamp. Recall that in these three modules, our proposed method directly learn from raw records and produce the risk probability of contagion chains, without requiring handcraft feature engineering.

3.2 Contagion Chain Construction

Given the input loan behaviors, we construct loan networks $G = (V, E)$, first. SMEs are treated as nodes $V = \{v_1, v_2, \cdots, v_{|V|}\}$ and guarantee relationships are denoted as edges $E = \{e_1, e_2, \cdots, e_{|E|}\}$. $|V|$ denotes the number of nodes and $|E|$ is the number of edges. We embed the profile of SMEs as the attribute of nodes, such as registered capital, number of employees, financial status, etc. The guarantee and loan information is preserved in the edge attributes, such as guarantee amount, loan amount, loan interest, etc. It should be remarked that the edge is valid from the guarantee start time until the guarantee end time, which means we update the network dynamically in training and predicting phase.

In the data preprocessing, the initial networks are constructed by existing records in offline. Then, we add a new edge e to the loan network G if the guarantee is issued and remove it when expired during the learning process. Meanwhile, the contagion chain is also updated $C = \{c_1, c_2, \cdots, c_{|C|}\}$. The contagion chain c_i of node v_i can be seemed as the maximum diffused nodes and edges $\mathrm{Diff}(v_i, G)$ that directed from v_i across the network G, denoted as $c_i = \{(v, e) | v \in \mathrm{Diff}(v_i, G) \cup e \in \mathrm{Diff}(v_i, G)\}$. In the implementation, we employ directed breadth-first search (BFS) [2] algorithm to produce contagion chains for each node. We then aggregate the attributes of corresponding nodes and edges within the collection as the feature of chains. So far, once the initial network and chains are constructed, it is very efficient to update them along with each loan and guarantee behavior. Besides, we only leverage the node and edge attributes. Thus the contagion chain construction can be computed efficiently and meet the low-legacy demands in the industry.

3.3 Inter-chain Attention Network

As presented above, if two chains are extracted from the same connected network, we mark them as affinal chains. Because its affinal chains also influence the risk of a contagion chain, we leverage the inter-chain attention network to learn the latent representations of each chain, which could preserve affinal information. Given the contagion chains with inner features of nodes and edges $C = \{c_1, c_2, \cdots, c_{|C|}\}$, and $c_i \in \mathbb{R}^F$, where F is the dimension of chain features, and $|C|$ is the number of chains. The inter-chain attention network produces hidden representation of chains $C' = \{c'_1, c'_2, \cdots, c'_{|C|}\}$, $c_i \in \mathbb{R}^{F'}$.

In the implementation, we conduct non-linear transformation of LeakyReLU on the feature concatenation of contagion chain c_i and each of its affinal chain, denoted as $c_j \in \mathcal{N}_i$. Then, a shared attention layer is performed on the transformed representations of c_i and c_i, which is defined as:

$$\begin{aligned} \alpha_{i,j} =& \mathrm{softmax}_j \left(\mathrm{NN_c}(W_c c_i, W_c c_j)\right) \\ =& \frac{\exp(\mathrm{LeakyReLU}(W_c[c_i \parallel c_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\mathrm{LeakyReLU}(W_c[c_i \parallel c_k]))} \end{aligned} \tag{1}$$

where $\alpha_{i,j}$ means the shared attention weights. $\mathrm{NN_c}$ denotes the non-linear transformation and W_c is a weighted matrix.

Then, the inter-attentional representation of a contagion chain c_i can be computed by the mathematical aggregation of its affinal chains. In our implementation, we leverage sigmoid function on the weighted summarization of each inputs, described as:

$$c'_i = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{i,j} \mathrm{LeakyReLU}(W_a c_j)\right) \tag{2}$$

where W_a denotes the weight matrix and $\alpha_{i,j}$ is the attentional coefficient. σ means the sigmoid function. The output of inter-chain attention layer $C' = \{c'_1, c'_2, \cdots, c'_{|C|}\}$ are employed as the input of downstream tasks.

3.4 Recurrent Neural Network

As shown in Figure 3b, the recurrent layer takes the output of inter-chain attention network as input. We construct the temporal feature of contagion chains in time stamp t as $h_t = \{c'^t_1, c'^t_2, \cdots, c'^t_{|C|}\}$, where c'^t_i is the learned representation of i-th chain in time stamp t. $t \in 1, 2, \cdots, T$ and T is the number of time stamp.

Given the sequential input features $h = \{h_1, h_2, \cdots, h_T\}$, follow the work of Chung et al.,[14], we leverage gated recurrent unit (GRU) to learn the temporal patterns in the contagion chain risk rating task. Compared with Long-Short Term Memory (LSTM) network, we choose to use the GRU in our method because it is computational efficiency. In detail, the formulation of GRU block in TRACER is defined as:

$$\begin{aligned} r_t &= \sigma\left(W_r h_t + U_r h'_{t-1} + b_r\right) \\ z_t &= \sigma\left(W_z h_t + U_z h'_{t-1} + b_z\right) \\ \widetilde{h}_t &= \tanh\left(W_h h_t + U_h(r_t \circ h't - 1) + b_h\right) \\ h'_t &= (1 - z_t) \circ h'_{t-1} + z_t \circ \widetilde{h}_t \end{aligned} \tag{3}$$

where r_t is the reset gate vector, z_t is update gate vector. σ is the sigmoid function and tanh denotes the hyperbolic tangent. W_r, W_z, U_r and U_z are the weight matrix. b_r and b_z denote bias vectors in the reset and update gates, respectively. $\circ$ means the element-wise manipulation.

On the top of the graph recurrent unit layer, we utilize temporal attention neural network to learn the different importance in dynamic contagion chains. In particular, the temporal attention layer takes $h' = \{h'_1, h'_2, \cdots, h'_T\}$ as inputs and produce a new representation as outputs, denoted as $rep = \{rep_1, rep_2, rep_T\}$. The process is formulated as follows:

$$\begin{aligned} \beta_{t,j} =& \frac{\exp(\mathrm{LeakyReLU}(W_n[h'_t \parallel h'_j]))}{\sum_{i=1}^{T} \exp(\mathrm{LeakyReLU}(W_n[h'_t \parallel h'_i]))} \\ rep_t =& \left(\parallel_{j=1}^{T} \beta_{t,j} W_t h'_j\right) \end{aligned} \tag{4}$$

where W_n is the trainable weights in temporal attention network. $\parallel$ denotes the vector concatenation operation. $\beta_{t,j}$ means attentional coefficients of time t and j.

3.5 Risk Rating Component and Loss Function

Given the representation of contagion chains rep, the risk rating component aims to infer the likelihood of which risk level will the chain located. For example, in this paper, $r = \{r_1, r_2, \cdots, r_l\}$ denotes the l classes of risk rating level, each r_i indicates a risk range. r_1 is the lowest range and r_l is the highest range. Based on the widely-used rating levels in financial literature [6], we set the $l = 4$ in our implementation. Finally, we define the loss function of the rating component as follows:

$$\begin{aligned} &\mathcal{L}(\sigma) = -\log(\sigma_y(z)) \\ &\mathrm{s.t.}\sigma_i(z) = \frac{\exp(z_i)}{\sum_{j=1}^{l} \exp(z_j)}, i = 1, \cdots, l \\ &z_i = \mathrm{NN_r}(rep, i; \theta) \end{aligned} \tag{5}$$

where $y \in \{1, 2, \cdots, l\}$ is the risk label. rep denotes the representation of contagion chain learned from attentional recurrent neural network. z_i takes the rep as inputs and denotes the predicted score of class i. $\mathrm{NN_r}(rep, i; \theta)$ is a shallow feed forward neural network that maps the chain representation rep into a real valued score and θ is the weights be learned. In our implementation, $\mathrm{NN_r}$ is consisted of two-layer ReLU(Rectified Linear Units) and one layer sigmoid. $\sigma_i(z)$ indicates the likelihood that the chain in risk level i, which employ the sigmoid function to scale the probability between 0 and 1. During the optimization process, we set the batch size to 64, learning ratio to 0.001, the dropout ratio to 0.2.

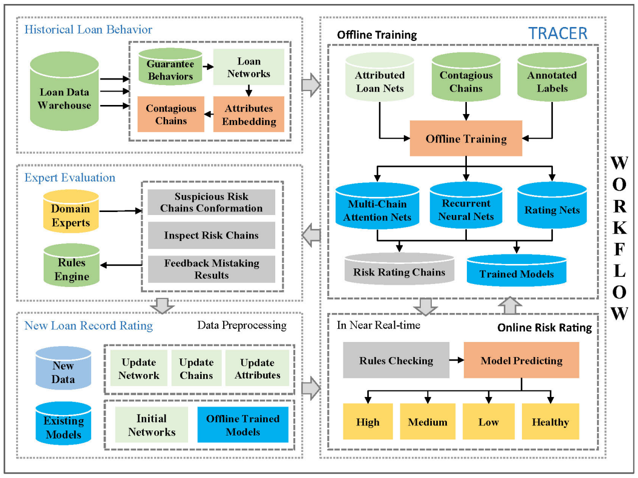


Figure 4: The workflow of contagion chain risk rating.

3.6 Workflow for Contagion chain Rating

Figure 4 shows the workflow for contagion chain risk rating. It contains four modules: historical loan behavior preprocessing, TRACER risk rating, expert evaluation and new loan record rating.

- **Historical loan behavior preprocessing** produces from raw data described in Figure 2 in three main steps: loan networks, attributes embedding and contagion chain construction.
- **TRACER risk rating module** trains the above data in offline by inter-chain attention network, recurrent neural layer and rating network. It produces trained models and predicts high-risk chains. The model is employed for online rating and retrained by results feedbacked from the expert evaluation module.
- **Expert evaluation** contains suspicious risk confirmation, inspecting the risk chains and feedback the mistake results to a database for model retrain. Domain experts also discover new risk control rules from predicted results to update the rules.
- **New loan record rating module** dynamically construct updated chains and attributed network from new records for online risk rating, which is processed in near real-time. Online module process the new records by risk rule checker and model predictor in turn for appropriate level allocation, which contain four levels: high, medium, low risk and healthy.

4 EXPERIMENTS

In this section, we describe the extensive experiments for evaluating the effectiveness of our proposed methods. We first describe the experimental settings. Then, present the experimental results of contagion chain risk ratings compared with other baselines, which is the main task of this paper. After that, the effects of alleviating cold start problem by the proposed method are tested. Finally, we report the case study.

4.1 Experimental Settings

Datasets. We collect the loan dataset from a major commercial bank in China, during 01/01/2013 and 31/12/2016. It includes 112,872 nodes(SMEs), with 124,957 edges (guarantee relationships). In data preprocessing, to exclude the trivial noise, we filter connected guarantee networks with over six vertices and contagion chains with three and more nodes. Finally, we construct 26,195 contagion chains. During the experiment, we utilize the data of the first year as the historical training set to initialize the model, and then predict the new records in sequence on the next three years. We retrain the model every day and predict the new records in the following days. We observe the ground-truth label of risk levels by the reported default loans in the testing period. The risk level is set according to collaborated financial domain experts as 1) Healthy, no loan defaults in the contagion chain. 2) Low, there is loans default, and the amount is smaller than 5%. 3) Medium, amount of 5%∼15% loans default. 4) High, an amount over 15% of loan defaults in a chain.

Compared Methods and Parameter Setting. We use the following widely used approaches in the banking industry as baselines to highlight the effectiveness of our proposed methods:

- *LR:* Logistic regression (LR) model [30] with one-vs-the-rest for multi-label classification. We apply L2 normalization and set $\lambda = 1$, tolerance for stopping criteria to 1e-4 and max iteration to 1000 and Follow the-regularized-leader (FTRL) for optimization.
- *GBDT:* Gradient boosting decision tree [24] is a popular ensemble learning method for classification, which has proved effective in credit risk rating procedures. We set the max depth to 3, The number of boosting stages to 500, learning rate to 0.1.
- *DNN:* [35] A deep neural network-based model for with 5 layers, in which the hidden layer are set as ReLU(256), ReLU(256) and ReLU(128). The output is a softmax layer, and Adam [26] is utilized for optimization. The dropout rates of first four hidden layers are 0.5, 0.5, 0.4, and 0.4, respectively. The batch size is 128. The maximum of epochs is 1000 with the early stopping of patience of 50 epochs.

Our model has several variations: TRACER-noICAN, in which inter-chain attention network is not employed. We directly fed chain attributes into the recurrent neural network. TRACER-noRNN, in which we remove the recurrent neural network and temporal attention mechanism. TRACER-all denoted the full model proposed in this paper. In our implementation, we apply the Adam optimizer [26] to update the parameters.

Evaluation Metrics. We evaluate the performance of the proposed approach by Precision@k, Micro-F1 and Macro-F1 score as evaluation metrics. In detail, we count the number of correct identification of positive labels in True Positives TP, incorrect identification of positive labels in False Positives FP and incorrect identification of negative labels in False Negative FN. Then, we calculate the Micro-F1 and Macro-F1 score as:

$$
\begin{aligned}
\mathrm{Micro-F1} &= \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l}(TP_i + FP_i)} \\
\mathrm{Macro-F1} &= \sum_{i=1}^{l} \frac{TP_i}{TP_i + FP_i}
\end{aligned}
\tag{6}
$$

where l denotes the number of levels in contagion chain risk rating. In our experiment, l is set to 4.

Precision@k is employed to evaluate the newly constructed chains in the cold start problem. The risk level assessment for contagion chains, unlike the classic corporation or personal risk level evaluation, has no credit or loan history for the chains of contagions. In this work, Precision@k is employed to measure the rank confidence of predicted chains. It is defined as:

$$\text{Precision@}k = \frac{|\{i|i \in C_p \cap C_o\}|}{|C_o|} \tag{7}$$

where C_p is the predicted top k confidence chain levels, C_o is the observed chain risk levels, and $|\cdot|$ represents the size of the set.

4.2 Risk Rating of Contagion Chain

In this section, we evaluate the risk rating accuracy of the contagion chains, which is the main task of this paper. We also demonstrate the superior performance of our work that does not require handcraft feature engineering with in-depth domain knowledge. We report the Macro-F1 and Micro-F1 first and then present the F1 score of each risk category in detail.

Table 1 shows the Macro-F1, Micro-F1 and number of features (N Features) of different baselines (** indicates that the improvements are statistically significant for $p < 0.01$ judged by paired t-test). We manually construct approximately 250 handcraft features by existing financial domain knowledge for logistic regression, and 1500 features for gradient boosting decision trees (GBDT). For deep neural network (DNN), our proposed TRACER and its variations (TRACER-noICAN and TRACER-noRNN), we employ the original data attributes as the input of the model, which means we do not conduct feature engineering for deep learning-based approaches. The first three rows of Table 1 report the result of baseline methods. As we can see, DNN is close to GBDT, both of them are better than logistic regression, indicating that contagion chain risk rating problem is too complicated for a swallow model (logistic regression) to address. GBDT and DNN improve the F1 score by enlarging model capacity. It should also be remarked that DNN achieves comparable results with GBDT of 1500 features by only 17 attributes, proves that the effectiveness of deep learning model in automatic feature engineering. TRACER-noICAN and TRACER-noRNN perform much better than GBDT and DNN, which strongly show the effects of each component in our proposed inter-chain attention network and recurrent network. The proposed TRACER-all significantly outperforms all baselines in both Macro-F1 and Micro-F1. We observe that TRACER-all achieves remarkable improvements by only 17 original loan attributes, compared with existing widely used risk rating model by about 250 handcraft features of logistic regression and 1500 features of GBDT. The result proves the significant advantages of our proposed method.

We then conduct experiments on each individual risk levels and report their F1-scores in Table 2. We observe a similar result here that GBDT and DNN are better than logistic regression. GBDT performs improvements slightly with DNN. The superior performance of TRACER is much more significant in the risk level of high and low. The F1-score of high chins in risk improves from 69.593% to 82.820%, and the F1-score of low-risk level boosts from 42.147% of logistic regression to 67.981%. We conduct the experiment 10 times and report the averaged result in the table. The improvements of our proposed method are statistically significant for $p < 0.01$ judged by paired t-test.

Table 1: The results of contagion chain risk rating.

Methods	Macro-F1	Micro-F1	N Features
Logistic Regression	0.76342	0.82266	~250
GBDT	0.79335	0.83937	~1500
DNN	0.78461	0.83314	17
TRACER-noICAN	0.83231	0.86528	17
TRACER-noRNN	0.83559	0.86807	17
TRACER-all	**0.85118****	**0.88191****	17

Table 2: The F1-score of each risk category.

Methods	Low	Medium	High
Logistic Regression	0.42147	0.93785	0.69593
GBDT	0.50389	0.94017	0.74130
DNN	0.48639	0.93949	0.73187
TRACER-noICAN	0.60966	0.95139	0.81307
TRACER-noRNN	0.61832	0.95141	0.81512
TRACER-all	**0.67981****	**0.95421****	**0.82820****

4.3 Cold Start Experiment

As described above, another major limitation of the previous work (L3) is that most of these works require a comprehensive credit or loan history for reliable risk rating. In this experiment, we evaluate the performance of the newly generated contagion chains; in other words, unlike the previous corporation or personal level analysis tasks, there is no credit or loan history for the chains of contagions. We utilize the precision of predicted top k confident chains for evaluating the cold start experiment.

Figure 5 shows the risk level assessment results of the cold start experiments on various level of contagion risk. The x-axis denotes the number of top k-th confident results and the y-axis denotes the precision. As can be seen from the chart, the precision for the top results keeps sufficient precision. The top 200 confident results receive better precision than the lower 300 confident results. In general, the precision gradually decreases according to the increase of k, that is because the more samples predicted, the less confidence of the model. From the three subgraphs of Figure 5, we observe that the precision of logistic regression that is extensively used in the risk assessment in the banking industry is not comparable to other methods. It is probably because the conventional model requires loan history, but in this experiment, there is no historical information for the newly constructed chains of contagion. GBDT and DNN are very close to each other and perform better than logistic regression. The ensemble learning approach and deep learning are proved to be effective, especially for deep learning, which achieves comparable performance with only 11.3% number of features of GBDT. TRACER performs significantly better than GBDT and DNN across all risk levels. The improvements vary from 5% to 15%, which are remarkable and constant. The reason might be that with inter-chain attention and temporal recurrent network structures, TRACER could effectively learn from the structure and temporal patterns from node-wise (the small and medium enterprises who borrow money from the banks) historical loan behaviors even the current chain does not preserve loan history. The essential and effectiveness of inter-chain attention and recurrent neural network are demonstrated in addressing the cold start problem.

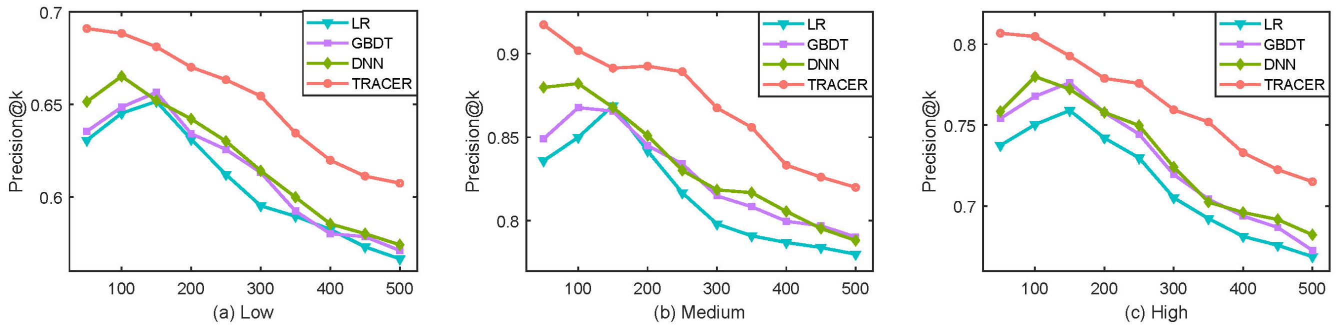


Figure 5: The precision@k of each risk category (low, medium and high) in the cold start experiment. The contagion chains for model prediction are newly constructed in test times, which have not any loan or credit history before.

Table 3: Statistical information of contagion chains in the GN263 loan network.

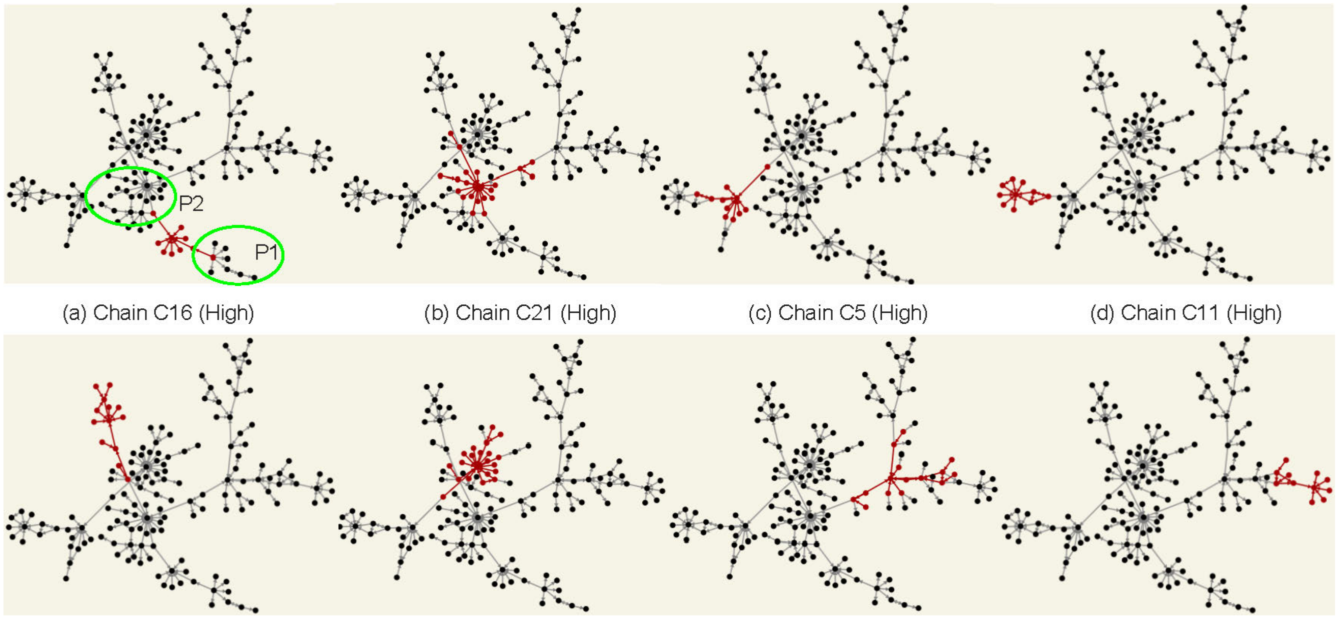
Chain #ID	C16	C21	C5	C11	C8	C10	C12	C3	C25	C9	C15	C14	C1	C17
Risk Level	H	H	H	H	H	H	H	M	M	L	L	L	L	L
Avg. Degree	1.8	1.9	2.0	2.6	2.0	2.0	2.1	2.0	2.2	2.0	2.0	2.0	1.5	1.5
Max. Degree	8	16	9	8	7	17	7	6	4	3	4	3	2	3
Firms	10	21	11	10	13	21	15	11	9	6	7	5	4	4
Default firms	2	4	3	1	2	3	2	1	2	1	1	2	1	1
Default firm ratio(%)	20.0	19.0	27.3	10.0	15.4	14.3	13.3	9.1	22.2	16.7	14.3	40.0	25.0	25.0
Total loan amount	360.1	295.7	610.7	426.8	621.1	401.5	149.3	140.3	146.5	55.0	69.9	148.1	26.2	139.9
Total default amount	117.7	89.5	157.5	92.6	120.1	77.2	27.4	14.9	10.4	2.3	1.5	1.8	0.1	0.4
Default amount ratio(%)	32.7	30.3	25.8	21.7	19.3	19.2	18.3	10.6	7.1	4.3	2.2	1.2	0.4	0.3

4.4 Case Studies

In this section, we report the case studies applying our methods on a real-world loan management system in our collaborated financial institution. We select a real-world loan network encoded as GN263 and report the statistical information of the predicted high-risk contagion chains and visualize them in the empirical study.

Table 3 present the detail information of typical contagion chains in GN263 network. As we can see, the first seven columns are detected as high-risk level (H), the next two chains as medium level (M) and the rest five columns are predicted as low risk (L). We report the average (Avg.) and maximum (Max.) degree of each chain in the second and third row. It is clear that the high and medium risk level chains preserve a larger number of the maximum degree. However, the average degree does not show a similar phenomenon. Line 4-6 reports the number of firms (nodes), the number of firms which fail to repay the loan (default), and the corresponding default ratio. As we can see, the contagion chain with more firms is more likely to be at higher risk, compared with those with a small number of firms. Besides, the number of default firms is also more significant in the high-level risk. Thus, chains in high, medium and low level generally share similar default firm ratio of around 15%. The last three rows show the loan amount within each chain, the amount of defaulted loans (in million dollars) and the default ratio. Normally, the risk contagion chain involves a larger amount of loans as they also preserve more companies, and thereby cause significant losses for financial institutions. In this pilot test, our proposed method successfully predict the high-risk level chains, in which over 30% of the loans are default, such as C16 and C21.

Then, we visualize the top eight risk chains of GN263 network for empirical study. As presented in Figure 6, the network layout is partially centralized and globally bridged. Specifically, the network is consisted by a set of sub-community, and most sub-community have a star node and various leaf nodes, such as C21 and C10. These chains are connected by one or multiple structure hole nodes [34]. In Figure 6, we highlight the risk contagion chains in red and layout them in each sub-figure. As we can see, these chains in risk can be categorized into two class, named tail chain and bridge chain. For example, C21 and C12 are typical bridge chains, while the C16, C21, C5, C3 and C11 are tail chains. Based on the empirical study with domain experts, we find that bridge chain is extremely important for the stability of the whole financial network. Taking C16 as an example, which is the highest risk chains detected by TRACER, if the inner-nodes (marked in red) of chain C16 are exposed in risk, the connected two parts (P1 and P2) will also lose protection. In practice, the part P2 has already detected as high-level risk chain C21. For P1, the red node in C16 guarantees most firms. If C16 is in risk, even though the firms in P1 will not be diffused, but they will lose guarantor's protection, which may lead to potential losses for financial institutions. This means that the network structure will be reorganized active or passive, and this will be strong hint information for the banks. Our proposed TRACER successfully rates the risk level of contagion chains in networked-loans, enabling loan issuer to take actions in advance for preventing financial losses.



(a) Chain C16 (High) (b) Chain C21 (High) (c) Chain C5 (High) (d) Chain C11 (High)

(e) Chain C8 (High) (f) Chain C10 (High) (g) Chain C12 (High) (h) Chain C3 (Medium)

Figure 6: Visualization of contagion chains in the GN263 loan network. We entitle each sub-figure with the Chain #ID and corresponding risk level. Highlighted chains are detected in the risk level of "High" and "Medium".

5 RELATED WORK

This section presents a review of recent literature on consumer credit rating and financial networks risk evaluation.

Consumer credit rating. A large number of existing studies in the broader literature have examined the data-driven approach based consumer credit rating (assessment) [18]. Seminal contributions have been made by Geoff on the "Partial Credit" model [29] about forty years ago. Later on, various statistical approaches are introduced for rating the credit level problem [22, 33]. A recent study by Dmitrii Babaev et al concluded that learned risk knowledge embedded in the neural networks are interpretable[4]. The rules and learnt hidden information are visualised to help analysis the semantic meanings. Rich consumer information such as debt-to-income ratios, consumer banking transactions is fed into the linear regression models to predict the delinquent rates in the near future, benefiting an accuracy of 85% and a cost-saving between 6% and 25% [25]. Babaev et. al, [3] proposed recurrent neural network model for loan applicant assessment. Afterwards, The authors in [11] employ graph attention neural network for networked loans and report considerable performance gains. However, despite from the naive averaging individual credit scoring in the diffusion path, there is still a blank area to evaluate the risk of contagion chains, the key task in preventing systemic risks in networked loans.

Risk evaluation in financial networks. Systemic financial crisis dealt the most severe threat to the country, even the world economy and a significant concern for financial companies and governments [8, 19]. The network structures exist in modern financial systems. For example, the listed companies prefer to adopt a cross-shareholding strategy to stabilize the stock price, and the banking institutions also rely heavily on the network structure to process some of the critical and sensitive tasks (i.e. overnight exchange and keep liquidity bank funds). Such complex dependencies and relationships inside may induce a risk of various kinds [9, 12, 28, 36]. The relationships between network structure and financial system risk attracted lots of attention from both the academic and industry. Some insights on the risk evaluation in the financial networks include: the network structure plays an important role in determining systemic risk and welfare in short-term debt than impacting the well-being of the system [10]. Regarding the financial network nature, the risk of default propagation attracted lots of attention as the crisis may be amplified quickly across the inter-linked institutes. The epidemic spreading models in networks were introduced into the financial network diffusion process assessment [5]. Examined and performed empirical study and give the conclusion that the market shock will affect all the capital of the networked financial institutes and makes all the members are vulnerable to potential losses and thus increases the possibility of a large scale default cascade [15, 23]. Rama et al pointed out that the two factors contagion exclosure and the institute are main concerns for the contagion and systemic risk in financial networks [16, 21]. The network-guaranteed network consisted of multiple enterprises in secured loans, was brought into the research community from the work [31]. Later on, the risk management framework for the network-guaranteed network was introduced, and a visual analytics approach was presented [34]. Extensive researches on default risk prediction have been performed since then [11, 13]. However, as far as we know, the risk evaluation problem of contagion chains in the networked-guarantee loans present in this work is the first attempt to the financial and computing research community.

6 CONCLUSION AND DISCUSSION

In this paper, we propose a novel method entitled TRACER for contagion chain risk rating. Our method addresses the three major limitations of existing approaches in the banking industry: L1: lacking systemic risk rating methods for contagion chains. Existing methods are for individual measuring. L2: requiring heavy handcraft feature engineering with deep domain knowledge. L3: cold start problem, which needs credit history for a reliable rating. We thoroughly evaluate our method comparing with the benchmarks on the historical records and achieve superior performance. The ability of the proposed method in addressing the above three limitations are demonstrated in turn. Besides, the empirical study result proves that our work could avoid significant financial losses.

In conclusion, the significant advantages of our method are: 1) even complex networked loan behaviors can be directly employed for risk rating without requiring feature engineering. 2) Attention mechanisms (inter-chain attention and temporal attention) could learn the optimal representations for unrecorded data without the need for loan or credit history. These mean that our method reduced the requirement of significant domain expertise for feature engineering, or sufficient historical records for the positive rating. The graph-based neural network learns meaningful and discriminative internal representations for the input networked-records with the training procedure, and the attention mechanism infers reliable features by attending on the relevant neighbors and temporal histories. In future work, we will integrate dynamic network information in our model more effectively and produce more efficient systems for various types of risk rating in networked-guarantee loans.

ACKNOWLEDGMENTS

The work is supported by the National Key R&D Program of China (2018AAA0100704), the China Postdoctoral Science Foundation (2019M651499), the National Natural Science Foundation of China (61802278) and the MoE Key Laboratory Foundation of Artificial Intelligence (AI2019004).

REFERENCES

[1] Eliana Angelini, Giacomo di Tollo, and Andrea Roli. 2008. A neural network approach for credit risk evaluation. *The quarterly review of economics and finance* 48, 4 (2008), 733–755.

[2] Baruch Awerbuch. 1985. Complexity of network synchronization. *Journal of the ACM (JACM)* 32, 4 (1985), 804–823.

[3] Dmitrii Babaev, Maxim Savchenko, Alexander Tuzhilin, and Dmitrii Umerenkov. 2019. E.T.-RNN: Applying Deep Learning to Credit Loan Applications. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2183–2190.

[4] Bart Baesens, Rudy Setiono, Christophe Mues, and Jan Vanthienen. 2003. Using neural network rule extraction and decision tables for credit-risk evaluation. *Management science* 49, 3 (2003), 312–329.

[5] Adrià Barja, Alejandro Martínez, Alex Arenas, Pablo Fleurquin, Jordi Nin, José J Ramasco, and Elena Tomás. 2019. Assessing the risk of default propagation in interconnected sectoral financial networks. *EPJ Data Science* 8, 1 (2019), 32.

[6] Diana Bonfim. 2009. Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics. *Journal of Banking & Finance* 33, 2 (2009), 281–299.

[7] Cristián Bravo, Lyn C Thomas, and Richard Weber. 2015. Improving credit scoring by differentiating defaulter behaviour. *Journal of the Operational Research Society* 66, 5 (2015), 771–781.

[8] Markus K Brunnermeier and Martin Oehmke. 2012. *Bubbles, financial crises, and systemic risk.* Technical Report. National Bureau of Economic Research.

[9] C Bayan Bruss, Anish Khazane, Jonathan Rider, Richard Serpe, Antonia Gogoglou, and Keegan E Hines. 2019. Deeptrax: Embedding graphs of financial transactions. *arXiv preprint arXiv:1907.07225* (2019).

[10] Mark Carey, Anil Kashyap, Raghuram Rajan, and RenÃĪ Stulz. 2012. *Market Institutions and Financial Market Risk.* Elsevier, Journal of Financial Economics. http://www.nber.org/books/care10-1

[11] Dawei Cheng, Yi Tu, Zhenwei Ma, Zhibin Niu, and Liqing Zhang. 2019. Risk assessment for networked-guarantee loans using high-order graph attention representation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 5822–5828.

[12] Dawei Cheng, Sheng Xiang, Chencheng Shang, Yiyi Zhang, Fangzhou Yang, and Liqing Zhang. 2020. Spatio-Temporal Attention-Based Neural Network for Credit Card Fraud Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 362–369.

[13] Dawei Cheng, Yiyi Zhang, Fangzhou Yang, Yi Tu, Zhibin Niu, and Liqing Zhang. 2019. A Dynamic Default Prediction Framework for Networked-guarantee Loans. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management.* 2547–2555.

[14] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014.*

[15] Rama Cont, Amal Moussa, et al. 2010. Network structure and systemic risk in banking systems. *Edson Bastos e, Network Structure and Systemic Risk in Banking Systems (December 1, 2010)* (2010).

[16] Rama Cont, Amal Moussa, Andreea Minca, and Edson Basto. 2009. Too interconnected to fail: contagion and systemic risk in financial networks. *Lecture presented at the IMF, May* (2009).

[17] Elena Dumitrescu, Sullivan Hue, Christophe Hurlin, and Sessi Tokpavi. 2018. Machine Learning for Credit Scoring: Improving Logistic Regression with Non Linear Decision Tree Effects. (2018).

[18] Eva Maria Falkner and Martin RW Hiebl. 2015. Risk management in SMEs: a systematic review of available evidence. *The Journal of Risk Finance* 16, 2 (2015), 122–144.

[19] Gerhard H Fischer and Ivo W Molenaar. 2012. *Rasch models: Foundations, recent developments, and applications.* Springer Science & Business Media.

[20] Trevor Fitzpatrick and Christophe Mues. 2016. An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market. *European Journal of Operational Research* 249, 2 (2016), 427–439.

[21] Jean-Pierre Fouque and Joseph A Langsam. 2013. *Handbook on systemic risk*. Cambridge University Press.

[22] Cheng-Lung Huang, Mu-Chen Chen, and Chieh-Jen Wang. 2007. Credit scoring with a data mining approach based on support vector machines. *Expert systems with applications* 33, 4 (2007), 847–856.

[23] Maximilian Jager, Thomas Siemsen, and Johannes Vilsmeier. 2020. Interbank Risk Assessment–A Simulation Approach. (2020).

[24] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *NeurIPS*. 3146–3154.

[25] Amir E Khandani, Adlar J Kim, and Andrew W Lo. 2010. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance* 34, 11 (2010), 2767–2787.

[26] Diederik P Kingma and Jimmy Lei Ba. 2014. Adam: Amethod for stochastic optimization. In *Proc. 3rd Int. Conf. Learn. Representations.*

[27] Erkki K Laitinen. 1999. Predicting a corporate credit analyst's risk estimate by logistic and linear models. *International review of financial analysis* 8, 2 (1999), 97–121.

[28] Alejandro Martínez, Jordi Nin, Elena Tomás, and Alberto Rubio. 2019. Graph convolutional networks on customer/supplier graph data to improve default prediction. In *International Workshop on Complex Networks*. Springer, 135–146.

[29] Geoff N Masters. 1982. A Rasch model for partial credit scoring. *Psychometrika* 47, 2 (1982), 149–174.

[30] H Brendan McMahan. 2011. Follow-the-regularized-leader and mirror descent: Equivalence theorems and l1 regularization. (2011).

[31] Xinhai Liu Xiangfeng Meng. 2015. Credit risk evaluation for loan guarantee chain in China. (2015).

[32] Atif Mian and Amir Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *The Quarterly Journal of Economics* 124, 4 (2009), 1449–1496.

[33] Yabo Ni, Dan Ou, Shichen Liu, Xiang Li, Wenwu Ou, Anxiang Zeng, and Luo Si. 2018. Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 596–605.

[34] Zhibin Niu, Dawei Cheng, Liqing Zhang, and Jiawan Zhang. 2018. Visual analytics for networked-guarantee loans risk management. In *2018 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 160–169.

[35] Fei Tan, Xiurui Hou, Jie Zhang, Zhi Wei, and Zhenyu Yan. 2018. A deep learning approach to competing risks representation in peer-to-peer lending. *IEEE transactions on neural networks and learning systems* (2018).

[36] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph attention networks. In *ICLR*.