



Company-as-Tribe: Company Financial Risk Assessment on Tribe-Style Graph with Hierarchical Graph Neural Networks

Wendong Bi

Institute of Computing Technology,
University of Chinese Academy of
Sciences
Beijing, China
biwendong20g@ict.ac.cn

Bingbing Xu*

Institute of Computing Technology,
Chinese Academy of Sciences
Beijing, China
xubingbing@ict.ac.cn

Xiaoqian Sun*

Institute of Computing Technology,
Chinese Academy of Sciences
Beijing, China
sunxiaoqian@ict.ac.cn

Zidong Wang

Institute of Computing Technology,
Chinese Academy of Sciences
Beijing, China
wangzidong@ict.ac.cn

Huawei Shen

Institute of Computing Technology,
Chinese Academy of Sciences
Beijing, China
shenhuawei@ict.ac.cn

Xueqi Cheng*

Institute of Computing Technology,
Chinese Academy of Sciences
Beijing, China
cxq@ict.ac.cn

ABSTRACT

Company financial risk is ubiquitous and early risk assessment for listed companies can avoid considerable losses. Traditional methods mainly focus on the financial statements of companies and lack the complex relationships among them. However, the financial statements are often biased and lagged, making it difficult to identify risks accurately and timely. To address the challenges, we redefine the problem as company financial risk assessment on tribe-style graph by taking each listed company and its shareholders as a tribe and leveraging financial news to build inter-tribe connections. Such tribe-style graphs present different patterns to distinguish risky companies from normal ones. However, most nodes in the tribe-style graph lack attributes, making it difficult to directly adopt existing graph learning methods (e.g., Graph Neural Networks (GNNs)). In this paper, we propose a novel Hierarchical Graph Neural Network (TH-GNN) for Tribe-style graphs via two levels, with the first level to encode the structure pattern of the tribes with contrastive learning, and the second level to diffuse information based on the inter-tribe relations, achieving effective and efficient risk assessment. Extensive experiments on the real-world company dataset show that our method achieves significant improvements on financial risk assessment over previous competing methods. Also, the extensive ablation studies and visualization comprehensively show the effectiveness of our method.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; • **Information systems** → **Social networks**.

*Corresponding authors



This work is licensed under a Creative Commons Attribution International 4.0 License.

KDD '22, August 14–18, 2022, Washington, DC, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9385-0/22/08.
<https://doi.org/10.1145/3534678.3539129>

KEYWORDS

company financial risk assessment, tribe-style graph, graph neural network

ACM Reference Format:

Wendong Bi, Bingbing Xu, Xiaoqian Sun, Zidong Wang, Huawei Shen, and Xueqi Cheng. 2022. Company-as-Tribe: Company Financial Risk Assessment on Tribe-Style Graph with Hierarchical Graph Neural Networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*, August 14–18, 2022, Washington, DC, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3534678.3539129>

1 INTRODUCTION

Company financial risk is ubiquitous in the real-world financial market. Early assessment of risks for the listed company can provide decision support for company managers and investment institutions, thereby avoiding considerable losses.

Traditional methods [7, 18], such as financial probability methods, decision tree methods, and Deep Neural Networks (DNNs), treat each company individually and solely leverage the financial statements to assess risk. However, financial statements are often biased and lagged. As Fig. 1 (a) shows, most companies beautify their published financial data, and some companies even commit financial fraud. Besides, these traditional methods ignore the interactions among companies, which is critical because risks can be passed between companies. The above limitations make the traditional methods difficult to identify risks accurately and early.

To effectively assess company financial risks, we found there exist two other types of valuable information: 1) the investment-graph of listed companies, e.g., CATL¹ has more than 200 investors (companies or individuals), which form an investment-graph. As Fig. 1 (b) shows, risky and normal companies often have different investment patterns; 2) The news-graph among listed companies, i.e., there exists an edge between two companies if they concur in at least one piece of news. As Fig. 1 (b) shows, two listed companies connected usually have strong correlations, and risks can spread over this graph. Superior to other information, financial

¹CATL (Contemporary Amperex Technology Co., Limited) is a typical listed company in China, which is a global leader of new energy innovative technologies and committed to providing premier solutions and services for new energy applications worldwide.

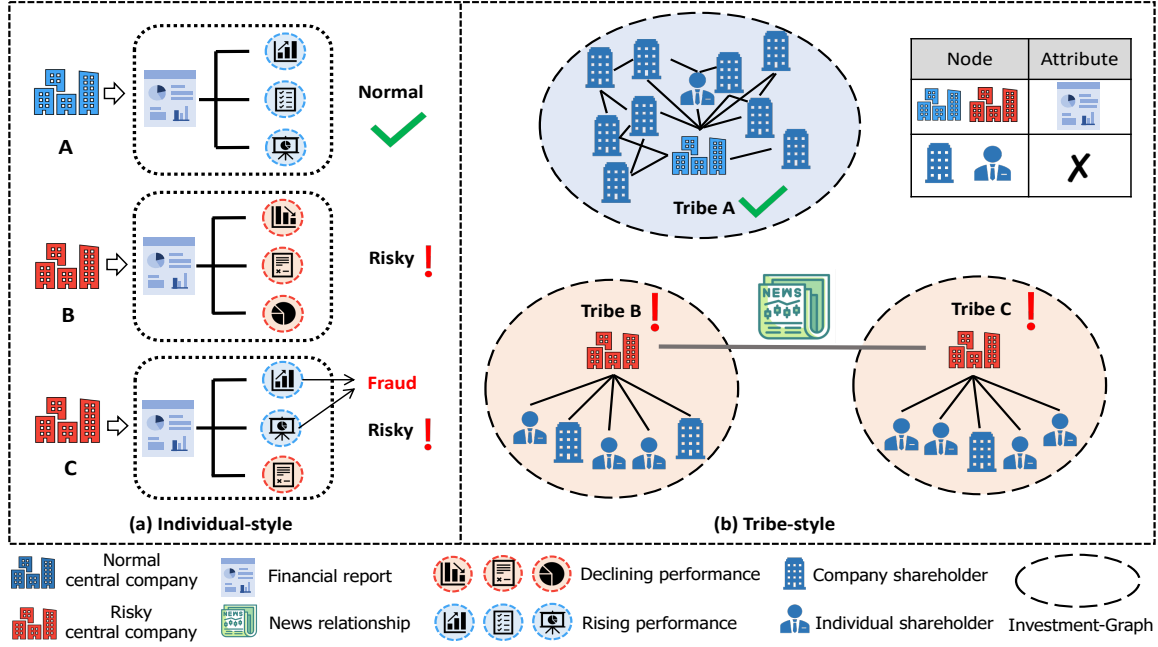


Figure 1: Company financial risk assessment on individual-style vs. tribe-style graph. Each company with its investment-graph in a dotted oval box can be seen as a tribe, and they are further connected by the global relationship (e.g., news relationship).

news is objective and can timely reflect risks. Extensive statistical analyses are provided in Sec. 2.1 to demonstrate the benefit of these data for distinguishing risky companies from the normal ones.

Based on the above findings, we redefine the problem as **company financial risk assessment on tribe-style graphs**. As illustrated in Fig. 1 (b), we take the investment-graph consisting of a central listed company and its shareholders as a tribe, and leverage the news-graph to construct inter-tribe edges, the financial statements of listed companies are regarded as initial attributes of tribes. However, it is challenging to directly adopt existing graph methods (e.g., Graph Neural Networks (GNNs)) to such tribe-style graphs due to the following serious issues: 1) only the listed companies in a tribe have attributes, and other individuals or companies have no disclosure obligation and therefore do not have attributes, making it difficult to conduct message passing in GNNs; 2) The whole tribe-style graph including both intra-tribe and inter-tribe relationships is large-scale and contains millions of edges, which makes the GNNs inefficient, and traditional node sampling techniques [4, 13, 14] cause the loss of information.

In this paper, we propose a novel **Hierarchical Graph Neural Network** for the financial risk assessment on the tribe-style graph, namely TH-GNN. Specifically, for the first challenge that the individuals and non-central companies in a tribe have no attributes, we find the structure patterns can reflect the company’s risks. Therefore, we design a tribe structure encoder (TSE) based on contrastive learning that learns structural patterns for each tribe (including the scale of the tribe and its investment structure, etc.) without relying on node attributes. For the second challenge, although the whole graph is huge, Fig. 1 (b) shows an important property of the tribe-style graph that the intra-tribe connections (investment-graph) are dense while the inter-tribal connections (news-graph)

are sparse. Inspired by this property, TH-GNN encodes the tribe-style graph through a hierarchical manner, with the first level to encode tribes defined by the investment graphs via the tribe structure encoder and the second level to diffuse the information among tribes based on the news-graph and learn the global representations. Unlike the traditional GNNs that diffuse information over edges on the whole graph, TH-GNN converts a tribe-style graph into parallelly computable local graphs and a smaller global graph. Extensive experiments on the real-world dataset for company financial risk assessment show that our approach achieves significant improvement over previous competing methods. Meanwhile, the ablation studies and visualizations also comprehensively show the effectiveness of our method.

The main contributions of this work are summarized as follows:

- (1) We redefine the previous individual risk assessment problem as **company financial risk assessment on tribe-style graph** and further design a tribe-style graph consisting of financial statements, investment-graphs, and news-graphs rather than solely utilizing financial statements.
- (2) We propose a novel Hierarchical Graph Neural Network named TH-GNN to model the tribe-style graph. To the best of our knowledge, this is the first graph representation learning method for company financial risk assessment on the tribe-style graph.
- (3) We conduct extensive experiments on a real-world company graph dataset with 0.88 million nodes and 1.31 million edges. The results demonstrate the superiority of the proposed model over state-of-the-art methods. The code is available ².

²Our source code is available at <https://github.com/wendongbi/TH-GNN>

2 PRELIMINARY

In this section, we present the detailed statistical analysis of tribe-style graph and the formalized definition of company financial risk assessment on tribe-style graphs.

2.1 Data Analysis

As illustrated in Fig. 1 (b), the tribe-style graph consists of investment-graph (tribe) and news-graph, where the investment-graph presents the intra-tribe connections and the news-graph presents the inter-tribe connections. We then analyze the investment-graphs and news-graph comprehensively to verify their benefits for financial risk assessment.

2.1.1 investment-graph analysis. An investment-graph usually consists of one central listed company and others (companies or individuals) which have investment relationships with the central company. Only the listed company publishes its financial statements as attributes, and the other nodes do not have attributes. Therefore we mainly focus on the structure patterns of tribes. Then we first give a case study of investment-graphs and then present the statistical analysis on all listed companies' investment-graphs. As

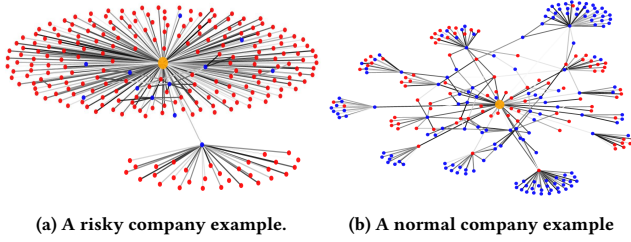


Figure 2: Examples of investment-graphs³. (The bigger orange point denotes listed company, blue points denote unlisted companies and red points denote individuals. And the edges in the graphs represent investment relationships.)

illustrated in Fig. 2, the investment-graphs for risky company and normal company show different patterns. The investment-graph of the risky company shown in Fig. 2 (a) is more similar to a star-like graph, where the single listed company can be viewed as the central node, and its neighboring nodes (investors) consist of more individuals or companies that tend to be disconnected from each other. Different from the risky company, the neighboring nodes of the normal company in Fig. 2 (b) are often popular companies which have more neighbors (investors), and there exist dense connections among these neighbors. Such a pattern with more reliable investors tends to be more stable, and thus the central listed company is less likely to have financial risks. This motivates us to leverage the investment pattern of listed company to identify financial risks.

To further verify this finding, we conduct centrality analysis for the investment-graphs of all listed companies. Five typical metrics are used [2, 22], including degree centrality, eigenvector centrality, clustering centrality, number of bridge, and the central node degree. We calculate the above metrics on each investment-graph and then

³These two examples are from two real-world companies named LongYuan Construction and ShangHai Dragon Corporation respectively.

take the average of risky and normal companies respectively to reflect the differences of centrality between the two classes.

Table 1: Statistical centrality analysis of the investment-graphs

Statistical metric	risky company	normal company
Degree centrality	0.264	0.224
Eigenvector centrality	0.4161	0.3604
Clustering coefficient	0.2102	0.1907
Number of bridge (avg)	125.4	112.3
Central node degree (avg)	58.43	46.71

The results of the centrality analysis are presented in Table 1, which demonstrate that the investment-graphs of risky companies usually have larger graph centrality compared with that of normal companies. And this pattern motivates us to take the structure encoding of investment-graph into consideration for financial risk assessment. We consider the structural pattern of a investment-graph as a tribe individually, where the nodes within a tribe are centered on the centrally-located listed company. We further explain the benefits of tribe-style graphs in Sec. 3.1.2.

2.1.2 news-graph analysis. Besides the investment relationship in tribes, we also use financial news to model the interactions between different tribes. The financial news has evident timeliness and objective authenticity, reporting company financial risks promptly. Specifically, different companies may appear in the same news, reflecting strong correlation among them, e.g., news reported that Company A and Company B jointly invested in a failed project, reflecting that they may have potential financial risks at the same time (news used for graph construction in this paper all describes similarity among companies). Then for companies co-existed in one piece of news, we connect them to construct news-graphs, indicating the risky associations among them.

As illustrated in Fig. 3, we also conduct statistical analysis to validate the benefits of the news-graph. We calculate the proportion of risky companies in neighbors of each node for risky companies and normal companies respectively. For example, the rightmost column in Fig. 3 indicates there are nearly 30% of risky companies (the red bar) with more than 80% of neighbors at-risk, while no more than 10% of normal companies (the blue bar) with more than 80% of neighbors at-risk. The results show that the probability of risky companies with high-proportion risky neighbors is much larger than that of normal companies, that is, the risky nodes in the news-graph have higher-proportional-risk neighbors. This finding also suggests that news-graph can benefit the company financial risk assessment. As a result, we use the news-graph to construct inter-tribe edges, modeling the relations among listed companies.

2.2 Problem Definition

Our company financial risk assessment problem is defined on the tribe-style graph, which consists of a set of tribes and a global news-graph connecting different tribes. For company financial risk assessment, we need to classify each company into binary classes:

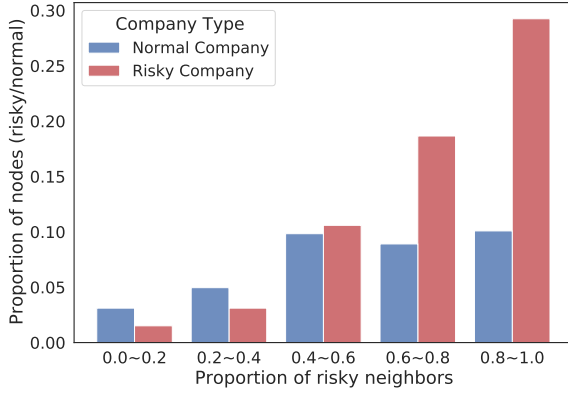


Figure 3: Analysis on proportion of risky neighbors. The x-axis is the proportion of risky companies in all neighbors of each node, and the y-axis is the corresponding proportion of nodes in a certain node type (risky/normal). Note that we ignore the 0-degree nodes when calculating these results.

risky or normal. We give a formal definition of the task as follows. Let $\mathcal{G}^T = \{\mathcal{G}^{global}, \mathcal{G}^{tribe}\}$ denote a tribe-style graph, where $\mathcal{G}^{global} = (V^G, E^G, X^G)$ is the global graph which taking all the listed companies as nodes. Considering that each listed company is the center of a tribe, we name the listed company node as the central node for shortly. V^G denotes the central node set, N^c is the number of central nodes, and E^G is the set of edges connecting central nodes. $X^G \in \mathbb{R}^{N^c \times D}$ is the attribute matrix, and the i -th row of X^G denoted by X_i^G is the D -dimensional attribute vector of central node v_i^G . For each central node $v_i^G \in V^G$, there exists a tribe $g_i^T \in \mathcal{G}^{tribe}$ corresponding to v_i^G , where $g_i^T = (V_i^T, E_i^T)$ and $\mathcal{G}^{tribe} = \{g_i^T \mid i = 1 \dots N^c\}$. V_i^T and E_i^T are the node set and edge set of tribe g_i^T . Each tribe have one central node. Each central node v_i^G is associated with a binary label $y_i = \{0, 1\}$, using 1 for risky companies and 0 for normal companies. Then the listed company's financial risks assessment problem on tribe-style graph can be described as: given a tribe-style graph \mathcal{G}^T , and the goal is to classify each listed company node into binary classes: risky or normal.

3 METHODS

In this section, we first explain how and why we design the tribe-style graph. Then we introduce the proposed TH-GNN model for company financial risk assessment on tribe-style graphs in detail.

3.1 Tribe-style Graph Construction

3.1.1 How to construct the tribe-style graph? As illustrated in Fig. 1, the tribe-style graph consists of company financial statements, investment-graphs, and financial news. Specifically, listed companies are viewed as central target nodes in the tribe-style graph, and the investment-graph of each listed company is viewed as a tribe (i.e., a super node on the tribe-style graph), where only the central listed company nodes have attributes extracted from their financial statements. The global graph connecting central companies is constructed by financial news, where two central listed

companies connected if they co-existed in at least one piece of financial news. Note that we only use news that describes the risky linkages among listed companies in China. If there are multiple companies appearing in the same news, we connect all possible pairs of companies. More details about the dataset information and preprocessing are presented in Sec. 4.1 and Sec. 4.2.

3.1.2 Why we construct the tribe-style graph? We summarize the following advantages of constructing a tribe-style graph.

- (1) Based on the analysis in Sec. 2.1, it is the structural pattern of tribes that benefits the identification of risky companies.
- (2) Considering that only the central node of a tribe has attributes, the central and non-central nodes should be treated separately. Therefore we make the tribe-style graph a hierarchical graph.
- (3) To improve model efficiency, we treat each tribe independently and obtain the representation of each tribe by graph pooling (regardless of overlap among tribes) instead of merging them into one graph, which actually truncates the original large-scale graph.

3.2 Model Overview

We give an overview of TH-GNN in Fig. 4. TH-GNN includes two main components, including the Tribe Structure Encoder (TSE) and Global Graph Representation Learning (GGRL) module. TH-GNN encodes the tribe-style graph in bottom-up order. TH-GNN first learns the structural representation for each tribe with the TSE. Then the learned structural representation of tribes and the financial statements are fused into the embedding of central node (listed company) by an attention-based fusion module. Next, the embedding is diffused over the global news-graph to learn the final representation of central nodes for financial risks assessment.

3.3 Tribe Structure Encoder (TSE)

The Tribe Structure Encoder (TSE) is used to learn the structural representation for each tribe based on contrastive learning, including a structure embedding module and a graph encoder module. Considering that nodes in the tribe have no attributes, we first initialize the node attributes according to their position in the tribe. Then we transform the structural attributes into learnable embedding with a structure embedding module for each node in the tribe. Finally, with the tribe (investment graph) and the node structure embedding, a GIN model is used to get the representation of tribes.

Inspired by the importance of centrality patterns in our scenario discussed in Sec. 2.1, we also consider the encoding of centrality when designing the TSE. For a tribe without node attributes, we first assign each node with a structure embedding as its initial attributes. Specifically, each node $v_j^T \in V_i^T$ on the investment-graph g_i^T has the three properties: (1) node degree (in-degree denoted by deg_j^+ and out-degree denoted by deg_j^- for a directed graph); (2) node type denoted by ϕ_j (listed company, unlisted company or human); (3) distance of shortest path (SPD) to the central node denoted by SPD_j . With these structure attributes, we further transform them into learnable embedding by an Embedding Layer:

$$Emb(v_j^T) = Embedding_Layer(deg_j^+, deg_j^-, \phi_j, SPD_j) \quad (1)$$

Following [9, 25], we also use the top Laplacian eigenvector with the largest eigenvalue of each tribe as the nodes positional encoding

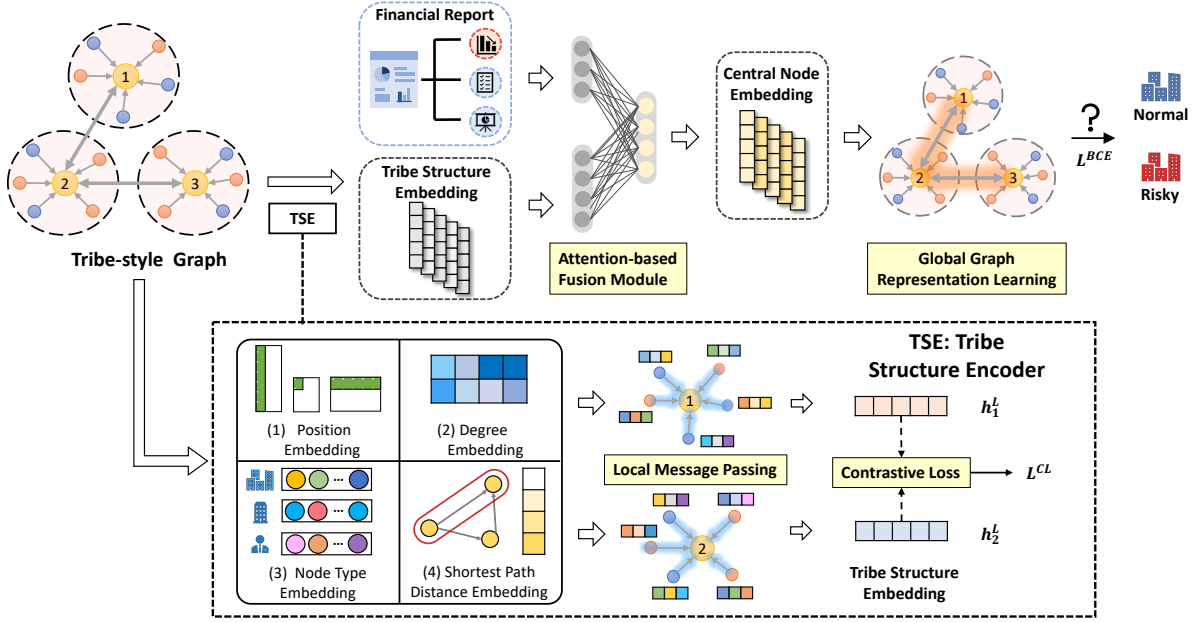


Figure 4: Overview of our proposed TH-GNN model. TH-GNN includes two main components, including the Tribe Structure Encoder (TSE) and Global Graph Representation Learning (GGRL) module.

besides the learnable structure embedding, and the j -th value of the top Laplacian eigenvector is exactly the eigenvector centrality [2] of node $v_j^T \in V_i^T$. Then the final structure embedding for each node on the tribe can be represented as:

$$Z(g_i^T) = [Emb(v_j^T) \parallel u_0(g_i^T)] \quad (2)$$

where $u_0(g_i^T)$ is the top Laplacian eigenvector of g_i^T . Next, we use the structure embedding (Eq. 2) for local message passing on tribes. In this paper, we use GIN with SUM pooling as the graph encoder for tribes. The GIN updates the node representations by:

$$h_{v_i^T}^{(l)} = MLP^{(l)} \left((1 + \epsilon^{(l)}) \cdot h_{v_i^T}^{(l-1)} + \sum_{v_j^T \in \mathcal{N}(v_i^T)} h_{v_j^T}^{(l-1)} \right) \quad (3)$$

where ϵ is a learnable parameter, $h_{v_i^T}^{(l)}$ is the learned representation of node v_i^T at the l -th layer of GIN. Then we get the global graph representation by performing SUM Pooling for all nodes in each tribe and taking an average over all layers of the model:

$$h_{g_i^T} = \frac{1}{L+1} \sum_{l=0}^L \text{SUM}(\{h_{v_i^T}^{(l)} | v_i^T \in g_i^T\}) \quad (4)$$

Then the final tribe representation is $H_{\mathcal{G}^{tribe}} = \{h_{g_i^T} | g_i^T \in \mathcal{G}^{tribe}\}$.

Considering the high cost of data labeling, we introduce contrastive learning to guide the training of TSE. Specifically, we design a graph instance discrimination tasks and use InfoNCE [24] as the objective to optimize the model parameters. The contrastive task treats each tribe as a distinct class and learns to discriminate between different tribes through a self-supervised way, and help the TSE module learn the structural dissimilarity between different tribes.

Specifically, we first prepare one positive sample pair and $N - 1$ negative sample pairs for each training batch with N samples and the tribes of each batch are fed into the TSE twice to obtain the query representation $H_{\mathcal{G}^{tribe}}^q$ and key representation $H_{\mathcal{G}^{tribe}}^k$ of tribes. Due to the randomness of dropout, there is a certain difference between the obtained two sets of tribe representations. Then we use the query and key representations of the same tribe as positive pairs denoted by $\{\langle q_i, k_i^+ \rangle, g_i^T \in \mathcal{G}^{tribe}\}$, and use the representations of different tribes to construct negative pairs denoted by $\{\langle q_i, k_j^- \rangle, j \neq i \text{ and } g_i^T, g_j^T \in \mathcal{G}^{tribe}\}$. Then we compute InfoNCE Loss as a regular term besides the supervised classification loss to optimize parameters of the Tribe Structure Encoder module:

$$\mathcal{L}^{CL} = -\frac{1}{N} \sum_{i=1}^N \log \frac{q_i^T \cdot k_i^+}{q_i^T \cdot k_i^+ + \sum_{j=1}^{N-1} q_i^T \cdot k_j^-} \quad (5)$$

3.4 Global-Graph Representation Learning (GGRL)

With TSE, we obtain the representations of tribes. And then for each central company, its node features come from two parts: the tribe representation and financial statements, which can be represented as $\{(h_{g_i^T}, X_i^G) | v_i^G \in V^G\}$. We further use an attention-based fusion module to integrate the tribe representations $h_{g_i^T}$ and financial statements feature X_i^G into one central node embedding $h_i^{(0)}$. Finally, the fused central node embedding is used for message passing on the global news-graph to learn the final representations of central companies.

To better integrate the node features from financial statements and tribes on the global graph, we design an attention-based fusion module to fuse the two features to common space. We first calculate

the weights for the financial statements and tribe representations:

$$\begin{cases} e_i^g = \sigma([h_{g_i^T} \cdot W^g || X_i^G \cdot W^X] \cdot a_g^T), & a_g \in \mathbb{R}^{1 \times 2D} \\ e_i^X = \sigma([h_{g_i^T} \cdot W^g || X_i^G \cdot W^X] \cdot a_x^T), & a_x \in \mathbb{R}^{1 \times 2D} \end{cases} \quad (6)$$

where σ is the LeakyReLU activation, W^g and W^X are transformation matrix to project $h_{g_i^T}$ and X_i^G into common hidden dimension. Then we normalize them by the softmax function:

$$\alpha_i^g = \frac{\exp(e_i^g)}{\exp(e_i^g) + \exp(e_i^X)}, \quad \alpha_i^X = \frac{\exp(e_i^X)}{\exp(e_i^g) + \exp(e_i^X)} \quad (7)$$

Next, we can get the fused central node embedding:

$$h_i^{(0)} = \alpha_i^g \cdot (h_{g_i^T} \cdot W^g) + \alpha_i^X \cdot (X_i^G \cdot W^X) \quad (8)$$

Then we use the fused central node embedding as node features and further perform message passing on the global news-graph to learn the final representations of each central node:

$$h_i^{(l)} = \sigma \left(\left(\sum_{j \in \{\mathcal{N}(v_i^G) \cup v_i^G\}} \frac{1}{d_i} \cdot h_j^{(l-1)} \right) \cdot W^l \right) \quad (9)$$

where d_i is the in-degree of v_i^G (including the self-loop). And the learned representations can be further used for the company financial risk assessment task.

3.5 Model Optimization Methods

After aggregating the information from neighbors on the graph, the obtained representation $\hat{h}_i^{(L)}$ is fed into a final fully connected neural network with a sigmoid activation function, as follows:

$$p_i = \text{Sigmoid}(h_i^{(L)} \cdot W_p + b_p) \quad (10)$$

where p_i is the probability of company node v_i suffering risks in the further. Then we compute binary cross entropy (BCE) loss to utilize the supervised information of labels:

$$\mathcal{L}^{BCE} = \frac{1}{N} \sum_{i=1}^N y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i) \quad (11)$$

Then, the final loss function is composed of \mathcal{L}^{BCE} and \mathcal{L}^{CL} (Eq. 5):

$$\mathcal{L} = \mathcal{L}^{BCE} + \alpha \cdot \mathcal{L}^{CL} \quad (12)$$

where α is a hyper-parameter to control the weight of \mathcal{L}^{CL} .

4 EXPERIMENTS

In this section, we compare TH-GNN with other state-of-the-art methods on a real-world dataset for company risk assessment.

4.1 Dataset

The company dataset used in this paper comes from the real-world data of 4040 listed companies in China from 2019 to 2020, i.e., the listed company's financial statements, investment-graph, and financial news related to these companies. The financial statements and the company's investment-graph data are provided by TianYanCha (an authority enterprise credit institute for company information inquiry in China). The annual financial statements reflect a listed company's industry information and its financial and business situation in a year. The investment-graph of a company describes

Table 2: Information of the Hierarchical Graph

Graph	#Nodes	#Edges
Whole graph \mathcal{G}^T	879252	1311364
News-graph \mathcal{G}^{global}	4040	16330
Investment-graphs (total) \mathcal{G}^{local}	879252	1295034
Investment-graphs (average) \mathcal{G}^{local}	217.6	320.5

the relationship between the central company and its shareholders, including other companies and humans. The **financial news** data are provided by Wind (an authority China finance database), and these news are obtained from more than 800 authority news websites in China, which have extremely wide coverage and timeliness to capture the risk information of companies. Note that the financial news used in this paper, which has already been preprocessed by Wind, all describes the risk linkages among companies. Then we construct the tribe-style graph and more specific information of this graph is illustrated in Table 2. Based on the real-world risk events of companies happened in 2020 provided by Wind, the positive (risky) and negative (normal) labels can be naturally generated, and we use all companies marked as high-risk as positive samples and others as negative samples. To prevent information leakage, the part of the dataset in 2019, including financial and operating information, investment-graph and news-graph, is used as training data. There are 1698 positive samples and 2342 negative samples among all listed companies.

4.2 Experimental Setup

We conduct experiments on the real-world dataset with different configurations. We design experiments with different training ratios (percent of nodes in the training set) ranging from 20% to 40%. And for each training ratio, we use three different random partitions of the dataset and ten random seeds for the model parameter initialization, a total of 30 trials for each model. For all attributes of the dataset used in this paper, we preprocess the categorical or discrete attributes into one-hot vectors, and then we use binning methods to divide continuous numerical attributes into 50 bins and use the index of bin as their feature. For fairness, we perform a hyper-parameter search for all models, and the size of searching space for each model is the same. The hidden dimension of all models are searched in {32, 64, 128} and we choose the number of training epoch from {100, 200, 300}. We use the Adam optimizer for all experiments and the learning rate is searched in {1e-2, 1e-3, 1e-4}, weight decay is searched in {1e-4, 1e-3, 5e-3}, and α (the coefficient of \mathcal{L}^{CL}) is searched in {0.01, 0.05, 0.1, 0.5, 1.0} for all experiments. The number of layers for GNN models, including TH-GNN and other baseline GNN models except for GCNII and DAGNN, are set to be two layers in this paper. The number of layers for GCNII and DAGNN, which are designed with deeper depth, are set to 64 and 20 respectively according to their papers [5, 16]. All models used in this paper were trained on Nvidia Tesla V100 (32G) GPU.

4.3 Compared Methods

4.3.1 Baseline methods. We compare our model with two classical machine learning models (XGBoost [6], DNN), three baseline GNN

Table 3: Binary node classification results (%) of models on our dataset. The model with * means using the concatenation of tribe structural embedding and financial statements data as inputs (see Sec. 4.3.2). Otherwise, only financial statement data is used.

Evaluation Metric	binary F1-score			AUC score		
	60%	40%	20%	60%	40%	20%
Training ratio						
XGBoost	57.4 ± 1.63	56.1 ± 1.55	55.9 ± 1.05	65.3 ± 2.23	64.3 ± 2.07	62.7 ± 2.23
DNN	55.8 ± 2.77	54.9 ± 2.56	53.5 ± 2.89	65.1 ± 2.77	65.6 ± 2.67	65.3 ± 2.86
GCN	58.9 ± 3.54	58.3 ± 1.74	56.8 ± 2.87	68.1 ± 1.64	67.2 ± 1.81	67.8 ± 1.63
GAT	57.6 ± 3.86	57.2 ± 3.75	57.2 ± 3.96	67.6 ± 3.11	67.2 ± 3.52	66.5 ± 4.21
GraphSAGE	59.4 ± 2.36	59.5 ± 2.36	58.5 ± 2.38	68.48 ± 3.84	67.14 ± 3.25	65.6 ± 1.78
GCNII	59.9 ± 2.04	59.7 ± 1.94	59.1 ± 1.88	69.6 ± 1.81	69.1 ± 1.72	68.8 ± 2.01
DAGNN	60.6 ± 1.96	60.7 ± 1.30	<u>60.0 ± 1.13</u>	70.6 ± 1.17	70.4 ± 1.17	69.7 ± 1.17
XGBoost*	58.2 ± 2.19	56.8 ± 2.05	56.2 ± 1.92	67.2 ± 2.56	64.9 ± 2.25	64.7 ± 2.33
DNN*	57.2 ± 2.11	56.8 ± 2.32	55.1 ± 2.13	66.3 ± 1.66	66.4 ± 1.88	65.3 ± 2.15
GCN*	59.2 ± 1.30	58.6 ± 2.01	57.3 ± 2.01	70.3 ± 0.95	68.8 ± 1.31	69.1 ± 1.55
GAT*	58.4 ± 2.01	57.4 ± 2.15	56.8 ± 2.34	67.8 ± 2.26	67.0 ± 2.34	66.1 ± 3.12
GraphSAGE*	59.7 ± 2.08	58.3 ± 2.0	57.3 ± 3.76	70.6 ± 2.15	69.8 ± 1.04	68.2 ± 1.23
GCNII*	60.5 ± 1.23	60.6 ± 1.30	<u>60.0 ± 1.17</u>	70.9 ± 1.73	<u>70.6 ± 1.85</u>	<u>70.9 ± 2.09</u>
DAGNN*	<u>61.1 ± 1.06</u>	<u>60.9 ± 1.12</u>	59.9 ± 3.69	<u>71.1 ± 1.74</u>	70.6 ± 1.38	70.2 ± 0.84
TH-GNN	63.2 ± 0.75	62.8 ± 0.95	62.2 ± 1.13	73.5 ± 0.54	72.8 ± 0.54	72.5 ± 0.54

models (GCN [15], GAT [30], GraphSAGE [13]), and two state-of-the-art GNN models (GCNII [5], DAGNN [16]) to demonstrate the superiority of our TH-GNN model. Furthermore, six variants of TH-GNN are designed for ablation studies: TH-GNN_{\Attribute} indicates removing node attributes (financial statements) from TH-GNN and only using the graph structure information; TH-GNN_{\TSE} indicates removing Tribe Structure Encoder from TH-GNN, without leveraging tribes; TH-GNN_{\GGRL} indicates removing the GGRL module from TH-GNN, without leveraging the news-graph; TH-GNN_{\Fusion} indicates removing attention-based fusion module from TH-GNN; TH-GNN_{\Emb} indicates removing the structure embedding used in TSE; TH-GNN_{\CL} indicates removing the contrastive loss term used to optimize the Tribe Structure Encoder from TH-GNN. In our experiments, we select two widely used metrics as performance measurement, i.e., AUC (the area enclosed by the coordinate axis under the ROC curve) and F1-score (the harmonic average of the precision and recall) on the test set.

4.3.2 Two implementations for base GNN models. Note that GNN models except TH-GNN cannot directly encode the tribe-style graph with hierarchical structures. For fair comparisons, we design two implementations for each baseline GNN model.

(1) **A basic implementation** is to directly train the vanilla baseline GNNs on the news-graph, because nodes except the central node (listed company) in a tribe do not have attributes (financial statements), which are necessary for vanilla GNN models.

(2) **A two-stage implementation** is to learn structure representation of tribes (investment-graphs) first and then train baseline GNNs on the news-graph. In this paper, we use GCC [25], a graph contrastive learning based method, to learn structure representation for tribes, which are concatenated with the financial statement feature as attributes of nodes on the news-graph.

4.4 Main Results

The main results of different models are presented in Table 3 and the major findings are summarized as follows:

(1) We observe that our TH-GNN model significantly outperforms other competing methods. Its binary F1-score, with the reported value of 63.2, is at least 5% higher than the tree-based model and traditional DNN model, and AUC gets 6.3% higher at the same time. Furthermore, TH-GNN is more advanced than the state-of-the-art GNN-based methods, i.e., GCNII and DAGNN, with about 2.1% increased F1-score and 2.4% increased AUC. Besides, the lower standard deviations of the results of TH-GNN indicate that our proposed model is more robust with different dataset splits.

(2) The results on the upper and lower sides of the middle line in Tab. 3 show the comparison of the two implementations of baseline models (see Sec. 4.3.2). Generally, the performance of base models with tribe structure representations as additional input (the model with *) are improved with varying degrees, which demonstrate the effectiveness of tribe structure encoding. Besides, TH-GNN outperforms the state-of-the-art GNN methods with tribe structure encoding as additional input. These improvements are mainly brought by the TSE module of TH-GNN that considers encoding of graph centrality and the end-to-end training strategy under the guidance of the supervised loss and the contrastive loss jointly.

4.5 Ablation Study

Then, we perform ablation studies to demonstrate the effectiveness of every component in our model. As shown in Table 4, the binary F1 and AUC of all variants deteriorate to some extent.

4.5.1 The effects of tribe-style graph. First, we demonstrate the effects of the tribe-style graph by removing information of certain type (i.e., the financial statements, global news-graph or local tribes)

Table 4: Binary node classification results (%) of TH-GNN and its variants (see Sec. 4.3.1) for ablation study.

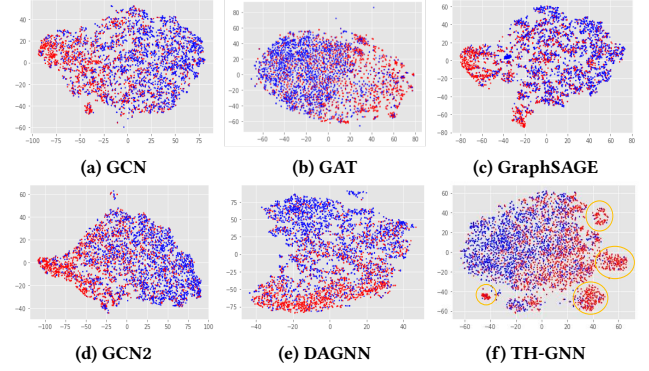
Graph	binary F1-score	AUC
TH-GNN _{\Attribute}	56.7 _{\-6.5}	62.4 _{\-10.5}
TH-GNN _{\GGRL}	58.2 _{\-5.0}	67.4 _{\-6.1}
TH-GNN _{\TSE}	60.2 _{\-3.0}	69.5 _{\-4.0}
TH-GNN _{\CL}	62.0 _{\-1.2}	71.9 _{\-1.6}
TH-GNN _{\Emb}	62.6 _{\-0.6}	72.1 _{\-1.4}
TH-GNN _{\Fusion}	62.5 _{\-0.7}	71.6 _{\-1.9}
TH-GNN	63.2 _{\0.0}	73.5 _{\0.0}

respectively. TH-GNN_{\Attribute}, as a variant without financial statements as node attributes, obtains the worst performance among all the variants with 6.5% decreased in F1-score and 10.5% decreased in AUC. Moreover, the reasonable score shows that even without the guidance of financial statements, only the graph structure information can still guarantee a certain classification accuracy, which demonstrates the effectiveness of our proposed tribe-style structure. Moreover, the performance degradation of TH-GNN_{\GGRL} and TH-GNN_{\TSE} indicates that the global news-graph and local tribes (investment-graphs) both benefit the downstream task.

4.5.2 The effects of different model components. To further verify the importance of different components in our model, another three variants of TH-GNN are designed for ablation study. The performance degradation of TH-GNN_{\CL}, which removes the contrastive loss term, shows that the contrastive loss L^{CL} plays an essential role in learning the discriminative structural information of tribes. Besides, the whole TH-GNN model yields a good performance boost in comparison to TH-GNN_{\Emb}, a variant without the structure embedding in TSE and using randomly initialized attributes for nodes on tribes, which indicates that the proposed structure embedding is beneficial for modeling graphs without node attributes. From the results of TH-GNN_{\Fusion}, we find that the fusion module, which integrates the information from financial statements and tribes comprehensively, brings significant improvements. Overall, the whole TH-GNN achieves the best result compared with all variants.

4.6 Visualization of Learned Representations

We analyse the model’s interpretability by visualizing the learned representations. We first extract the hidden outputs of the penultimate layer for different GNN models. Next, we reduce the hidden vectors to 2-dimensional vectors with T-SNE algorithm and plot their the scatter plot. As shown in Fig. 5, the representations learned by GCN, GAT and GraphSAGE have weak discrimination between risky and normal nodes. Despite the representations learned by GCNII and DAGNN have improved node classification ability, our TH-GNN has the best discriminative power and provides interpretability that other GNN models do not have. Fig. 5 (f) illustrates the visualization of the representations learned by TH-GNN which have evident outlier clusters, and most of points in these clusters are risky companies. This unique phenomenon indicates that some risky companies have similar local investment-patterns (i.e., the example presented in Fig. 2 (a)), which is easily affected by their

**Figure 5: T-SNE visualization of the representations learned by models, where red dots represent samples of risky companies and blue dots represent samples of normal companies.**

risky neighbors. And this observation further demonstrate the indispensable effects of the Tribe Structure Encoder (TSE).

5 RELATED WORK

In this paper, we mainly focus on using graph learning methods to solve the company financial risk assessment problem.

Graph Representation Learning. Graph Neural Networks (GNNs) [13, 15, 30, 31, 33, 34] have shown excellent performance on graph representation learning. Graph Convolution Network (GCN) [15] was first proposed with aggregation-based graph convolution operation. After that, its improved models have been proposed. GAT [30] introduced an attention mechanism to distinguish the importance of neighbors. GIN [36] investigate the effect of different aggregation functions for graph representation ability. GraphSAGE [13] proposed to use graph sampling for inductive learning on graphs. Recently, some studies [5, 16] also focus on deeper GCN with larger receptive filed. GCNII [5] is an extension of vanilla GCN with initial residual and identity mapping. DAGNN [16] is a deeper GNN model that decouples the transformation and aggregation of message passing. However, most of them [12, 19, 20, 32, 33] are designed without taking the unique domain knowledge into consideration and are not aimed at tackling the company financial risk assessment on financial networks.

Company Financial Risk Assessment. Financial risk has been a major concern for financial companies and governments, and extensive works have been studied to assess company financial risks. With the rapid development of machine learning methods in recent years, researchers have developed machine learning-based company financial risk prediction models using financial statements or privately-owned data provided by financial institutions [7, 10, 18, 38]. After the global financial crisis in 2008, the crisis brought by the collapse of Lehman Brothers spread rapidly and widely to related companies, and graph theory began to attract the attention of researchers [1, 3, 17, 21, 23, 26–29, 35]. Recently, some studies attempt to construct different financial graphs and design domain-specific GNNs for company financial risk assessment [8, 11, 37, 39]. For example, [8] proposed a High-order Graph Attention Networks to assess risks for companies on guarantee

loans networks considering higher-order neighbors. Yang et al. [37] proposed a dynamic GNN for supply chain mining. Zheng et al. [39] designed a heterogeneous graph attention network to predict the bankruptcy risks of small and medium-sized companies. However, all these works only consider one type of financial graph, and the data is specific and private. Models trained on such domain-specific data may have bias and have poor generalization ability.

6 CONCLUSION AND FURTHER WORK

In this paper, we investigate the listed company's financial data in real words and find that it is far from sufficient to assess the risk of listed companies solely through financial statements. To provide more comprehensive representations of companies, we design a tribe-style network and propose TH-GNN for company financial risk assessment task. Experiments on a real-world large-scale dataset show that our proposed model is effective in company financial risk assessment task and can provide brilliant interpretability of results. Furthermore, much future work is focused on the scalability of the method to investigate problems such as credit evaluation, risk transmission and so on. In addition, TH-GNN provides valuable tools to analyse the information of companies and their risks. As future work, we will try more complicated GNN backbones rather than simply GIN and GCN, and combine real-time information with tribe-style graph to further improve the dynamic risk identification.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No.61902380, No.61802370, No.U21B2046 and No.U1911401) and the Beijing Nova Program (No. Z201100006820061).

REFERENCES

- [1] Franklin Allen and Ana Babus. 2009. Networks in finance. *The network challenge: strategy, profit, and risk in an interlinked world* 367 (2009).
- [2] Phillip Bonacich. 1987. Power and centrality: A family of measures. *American journal of sociology* 92, 5 (1987), 1170–1182.
- [3] Spiros Boughreas and Alan Kirman. 2015. Complex financial networks and systemic risk: A review. *Complexity and geographical economics* (2015), 115–139.
- [4] Jianfei Chen, Jun Zhu, and Le Song. 2017. Stochastic training of graph convolutional networks with variance reduction. *arXiv preprint arXiv:1710.10568* (2017).
- [5] Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. 2020. Simple and deep graph convolutional networks. In *International Conference on Machine Learning*. PMLR, 1725–1735.
- [6] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.
- [7] Zhensong Chen, Wei Chen, and Yong Shi. 2020. Ensemble learning with label proportions for bankruptcy prediction. *Expert Systems with Applications* 146 (2020), 113155.
- [8] Dawei Cheng, Yi Tu, Zhen-Wei Ma, Zhibin Niu, and Liqing Zhang. 2019. Risk Assessment for Networked-guarantee Loans Using High-order Graph Attention Representation.. In *IJCAI*. 5822–5828.
- [9] Vijay Prakash Dwivedi and Xavier Bresson. 2020. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699* (2020).
- [10] Birsan Eygi Erdogan. 2013. Prediction of bankruptcy using support vector machines: an application to bank bankruptcy. *Journal of Statistical Computation and Simulation* 83, 8 (2013), 1543–1555.
- [11] Bojing Feng, Haonan Xu, Wenfang Xue, and Bindang Xue. 2020. Every Corporation Owns Its Structure: Corporate Credit Ratings via Graph Neural Networks. *arXiv preprint arXiv:2012.01933* (2020).
- [12] Qiang Fu, Lun Du, Haitao Mao, Xu Chen, Wei Fang, Shi Han, and Dongmei Zhang. 2021. Neuron with Steady Response Leads to Better Generalization. *arXiv preprint arXiv:2111.15414* (2021).
- [13] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 1025–1035.
- [14] Wenbing Huang, Tong Zhang, Yu Rong, and Junzhou Huang. 2018. Adaptive sampling towards fast graph representation learning. *arXiv preprint arXiv:1809.05343* (2018).
- [15] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
- [16] Meng Liu, Hongyang Gao, and Shuiwang Ji. 2020. Towards deeper graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 338–348.
- [17] Yang Liu, Xiang Ao, Zidi Qin, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. 2021. Pick and choose: a GNN-based imbalanced learning approach for fraud detection. In *Proceedings of the Web Conference 2021*. 3168–3177.
- [18] Feng Mai, Shaonan Tian, Chihoon Lee, and Ling Ma. 2019. Deep learning models for bankruptcy prediction using textual disclosures. *European journal of operational research* 274, 2 (2019), 743–758.
- [19] Haitao Mao, Xu Chen, Qiang Fu, Lun Du, Shi Han, and Dongmei Zhang. 2021. Neuron Campaign for Initialization Guided by Information Bottleneck Theory. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 3328–3332.
- [20] Haitao Mao, Lun Du, Yujia Zheng, Qiang Fu, Zelin Li, Xu Chen, Han Shi, and Dongmei Zhang. 2021. Source Free Unsupervised Graph Domain Adaptation. *arXiv preprint arXiv:2112.00955* (2021).
- [21] Xiangfeng Meng, Yunhai Tong, Xinhai Liu, Yiren Chen, and Shaohua Tan. 2017. Netrating: Credit risk evaluation for loan guarantee chain in china. In *Pacific-Asia Workshop on Intelligence and Security Informatics*. Springer, 99–108.
- [22] Mark Newman. 2018. *Networks*. Oxford university press.
- [23] 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.
- [24] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).
- [25] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. 2020. Gcc: Graph contrastive coding for graph neural network pre-training. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1150–1160.
- [26] Shirin Rezaei, Sajjad Shokouhyar, and Mostafa Zandieh. 2019. A neural network approach for retailer risk assessment in the aftermarket industry. *Benchmarking: An International Journal* (2019).
- [27] Xiaoqian Sun, Xueqi Cheng, and Huawei Shen. 2014. Trading network predicts stock price. *Scientific Reports* 4, 1 (2014), 313–317.
- [28] Xiaoqian Sun, Xueqi Cheng, Huawei Shen, and Zhaoyang Wang. 2011. Distinguishing manipulated stocks via trading network analysis. *Physica A* 390 (2011), 3427–3434.
- [29] Peter Trkman and Kevin McCormack. 2009. Supply chain risk in turbulent environments—A conceptual model for managing supply chain network risk. *International Journal of Production Economics* 119, 2 (2009), 247–258.
- [30] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).
- [31] Bingbing Xu, Keyan Cen, Junjie Huang, Huawei Shen, and Xueqi Cheng. 2020. A survey on graph convolutional neural network. *Chinese Journal of Computers* 43, 5 (2020), 755–780.
- [32] Bingbing Xu, Junjie Huang, Liang Hou, Huawei Shen, Jinhua Gao, and Xueqi Cheng. 2020. Label-consistency based graph neural networks for semi-supervised node classification. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*. 1897–1900.
- [33] Bingbing Xu, Huawei Shen, Qi Cao, Keting Cen, and Xueqi Cheng. 2020. Graph convolutional networks using heat kernel for semi-supervised learning. *arXiv preprint arXiv:2007.16002* (2020).
- [34] Bingbing Xu, Huawei Shen, Qi Cao, Yunqi Qiu, and Xueqi Cheng. 2019. Graph wavelet neural network. *arXiv preprint arXiv:1904.07785* (2019).
- [35] Bingbing Xu, Huawei Shen, Bingjie Sun, Rong An, Qi Cao, and Xueqi Cheng. 2021. Towards consumer loan fraud detection: Graph neural networks with role-constrained conditional random field.
- [36] Keyulu Yu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018).
- [37] Shuo Yang, Zhiqiang Zhang, Jun Zhou, Yang Wang, Wang Sun, Xingyu Zhong, Yanming Fang, Quan Yu, and Yuan Qi. 2020. Financial Risk Analysis for SMEs with Graph-based Supply Chain Mining.. In *IJCAI*. 4661–4667.
- [38] Yanci Zhang, Tianming Du, Yujie Sun, Lawrence Donohue, and Rui Dai. 2021. Form 10-q itemization. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4817–4822.
- [39] Yizhen Zheng, Vincent Lee, Zonghan Wu, and Shirui Pan. 2021. Heterogeneous Graph Attention Network for Small and Medium-Sized Enterprises Bankruptcy Prediction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 140–151.