Applied Soft Computing 132 (2023) 109849

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Industry classification based on supply chain network information using Graph Neural Networks

Desheng Wu [a], Quanbin Wang [b], David L. Olson [c,*]

[a] School of Economics and Management, University of Chinese Academy of Sciences, No. 80, Zhongguancun East Road Haidian District, Beijing, 100190, China
[b] Sino Danish College, University of Chinese Academy of Sciences, No. 80, Zhongguancun East Road Haidian District, Beijing, 100190, China
[c] Department of Supply Chain Management & Analytics, University of Nebraska-Lincoln, Lincoln, NE 68588-0491, United States of America

ARTICLE INFO

Article history:
Received 2 February 2022
Received in revised form 22 September 2022
Accepted 14 November 2022
Available online 28 November 2022

Keywords:
Supply chain network
Industry classification
Graph neural network

ABSTRACT

The number and trade volume of Chinese firms are increasing year by year. The resulting variety of complex transactions have made risk control and government supervision difficult. China's listed companies have specific classifications, but most non-listed companies do not have comparable classifications, making it difficult to analyze all companies on the same basis. Supply chain networks have proved to contain rich information, which can more completely reflect transaction relationships. This study mines hidden information obtained from the supply chain network to classify participating companies. We construct the supply chain network data set of listed companies, and use the graph neural network (GNN) algorithm to classify these companies. Experiments show that this method is effective and can produce better results than the commonly used machine learning methods. On average the accuracy of industry classification for listed companies is improved by over 2%, and time required is greatly reduced. In addition, we use economic variables derived from supply chain concepts to try to explain the effectiveness and economic significance of GNN, and find that GNN can also be used to classify companies into multiple industries. Our findings provide new insights, as well as a potential method to label a private company's industry using only public text information, which can be used for the study of smart industry classification and mining implicit information from the perspective of supply chain networks.

1. Introduction

The COVID-19 pandemic has increased the need to pay more attention to supply chain security. It is necessary to examine supply chain risk in the key products such as important semiconductor and high-energy battery industry. Enterprises are a major element in analysis of a supply chain. The industry in which an enterprise is located and its position in the industry have a great impact on supply chain security. Therefore, accurate industry classification for a large number of domestic listed companies, small and medium-sized enterprises and unlisted leading enterprises is a major prerequisite for follow-up work.

The study was motivated by discussion with several supply chain managers from the listed companies as well as staffs of China's securities regulatory authorities. These interviews indicated that some upstream companies located in countries or regions seriously affected by the pandemic were forced to shut down. It was difficult for middle and downstream companies to find alternative suppliers in the same industrial chain within a short period. However, many companies now implement diversification strategies. The diversified business carried out by these companies increased their potential to become upstream suppliers in another industry, but their potentials are unaware by downstream companies which operate in other industries. For example, at the beginning of the pandemic, the production capacity of masks was far from meeting the demand at that time. Because of the similarity of production equipment, the Chinese government designated garment manufacturers and automobile manufacturers to quickly put into mask production. BYD, Shanghai GM and Wuling did so. This example reveals that researchers have an opportunity to find potential suppliers through company information and production information.

Based on the above discussion, we want to construct a company knowledge graph to show the panorama of Chinese companies, including shareholding relationship and supply relationship, which can not only help companies find suppliers and customers, but also provide new means for regulators to regulate the diversified businesses of companies. During this process, we find

* Corresponding author.
E-mail addresses: dash.wu@gmail.com (D. Wu), wangquanbin20@mails.ucas.ac.cn (Q. Wang), david.olson@unl.edu (D.L. Olson).

https://doi.org/10.1016/j.asoc.2022.109849
1568-4946/

that compared to listed companies, tens of thousands of unlisted companies in China do not have any industry classification related to the market except for national economic classification. As we all know, some institutions have classified listed companies according to different rules to facilitate the use of funds by investors and regulators such as CSRC, Wind, Shenwan, MSCI and FTSE. However, the main drivers of these classifications are the proportion of business revenue and profit limited to listed companies. Non-listed companies have no publicly disclosed financial statements, so they cannot be classified directly by the existing rules. However, we found that all companies disclose their business scope, main products and other text information, and listed companies also disclose the transaction amount with key suppliers and customers. We hope to find an effective way to classify a large number of non-listed companies by using this public information using existing industry classification standards. It should be pointed out that we are not establishing a new industry classification scheme but marking a large number of non-listed companies with established standard industry labels on the basis of the existing scheme and exploring the possibility of diversified classification for these companies.

Taking China as an example, China Securities Regulatory Commission (CSRC) has classified more than 3900 listed companies in the Shanghai and Shenzhen stock markets, divided into 19 categories (see Table 1) updated on the website every quarter according to changes in that company's operations. The CSRC classification has become a common classification used by many Chinese scholars in studying Chinese economics. However, this is insufficient for broader research problems and nationwide supply chain research. The total number of companies in China, including listed companies and non-listed companies exceeds tens of thousands, and the vast majority of companies do not have a clear CSRC classification, one of the difficulties we encounter in constructing a knowledge graph of Chinese firms.

As economies have developed, company diversification has become common, further complicating analysis. Commonly used classifications use a single dimension [1]. Although this can meet the needs of economic accounting and government supervision, it inaccurately reflects changes in industry structure under the rapid development of enterprise innovation and confounds academic research on economic and corporate issues [2,3]. In previous industry classification tasks, researchers often focused on single dimension characteristics of one company, such as business scope, business philosophy, goal and mission. With the increasing complexity of economic activities, the rich transaction information contained in the interconnected network of companies can provide more implicit knowledge to help us more accurately classify, to include multiple classification, and also help researchers further identify and evaluate supply chain risks. Therefore, an effective means to classify companies in a supply chain network is possible and necessary. Classification can be by CSRC, national economy or global industry standards. Due to the topology of supply chain networks, each company can be regarded as a node. Upstream and downstream companies are connected through edges, which transmit rich information. Based on the above analysis, we propose a graph neural network (GNN) model, a deep learning algorithm specially designed for graph structure data, to classify companies into CSRC industry classification in supply chain networks. This method has achieved good results on the data set we constructed.

Graph neural networks have achieved remarkable results in various classification tasks in the computer science field. GNN was first used in image classification [4,5], and was then extended to text classification [6], protein function prediction [7], etc. Scholars have proposed specific means to optimize GNN algorithms for specific applications, to include a temporary static graph net (TSGNet) that simultaneously captures time and static interaction patterns, thus reducing the error rate of node classification by an average of 10% [8]. GNN has been also used for text classification and drug classification. Since graph neural networks were developed for graph structure and network structure data, scholars have also used them to enhance visibility and predicting hidden relationships in supply chain networks [9], as well as extracting supply chain map from news [10].

In this paper, we show that the edges in the supply chain network contain a lot of information useful for research. We demonstrate this point with classification of Chinese listed companies. Our contribution is not only to further enrich the application of graph neural networks in supply chain networks, but also to provide a solution for industry to solve practical engineering problems. Firstly, to the best of our knowledge, there is no unified classification standard for all companies in China and even the global market, which brings many difficulties to the supervision and research of non-listed companies. Our proposed framework is helpful to form a unified, open and transparent industry classification method that are more convenient for regulators and researchers. Secondly, Big Name companies have a wealth of supplier and customer data, thus they can construct specific features to accomplish target tasks according to their actual demand, such as grouping suppliers and customers. Doing this can make companies find alternative suppliers or accurately deliver products to classified customer groups, so as to improve the operation efficiency and benefits. Thirdly, we confirm that the supply chain network contains rich economic information that can be used through network structure data calculation, which provides insights for the method research of supply chain network and other networks in the future.

We also introduce in detail how to construct a supply chain network dataset suitable for GNN, which can serve as a basic work for future research. This paper is organized as follows: Section 2 discusses work that inspired this paper. Section 3 introduces the research design, including detailed introduction of the GNN model. Section 4 presents the data used in this paper with sources, as well as the construction process for data sets. Section 5 shows the performance of two algorithms Graph Convolutional Network (GCN)/Graph Attention Network (GAT) of graph neural network in industry classification of listed companies in supply chain networks. Results are compared with the traditional methods of machine learning, indicating more accurate classification. Section 6 discusses the limitations of our work and prospects for future research and summarizes the paper.

2. Related work

2.1. Supply chain network

Strader et al. [11] first introduced network science into supply chain research and put forward the concept of a supply chain network. Suppliers, manufacturers and distributors have interdependent strategies, tasks, resources and capacity, thus forming a complex "supply-production-sales" network. Roscoe et al. [12] defined a supply chain network as "a collection of supply chains, a network formed by the flow of services or goods from resource origin to final customers". Surana et al. [13] identified contemporary supply chains as very complex networks, including a large number of interactions and interdependencies generated by different firm entities sharing processes and resources. The network is nonlinear, indicating the scale-free characteristics of the network, and the network has a structure spanning several scales. This structure is self-organized, evolving through complex interactions and functions. With continuous refinement of the division of production, core companies in one supply chain carry

Table 1
CSRC industry classification.

CSRC industry classification	
Agriculture, forestry, animal husbandry and fishery	Finance
Mining	Real estate
Manufacturing	Leasing and commercial service
Electricity, Heat, Gas & Water production and supply	Scientific research and technology services
Construction industry	Water conservancy, environment and public facilities management
Wholesale and retail trade	Residential services, repair and other services
Transport, Storage and postal service	Education
Catering and accommodation	Health and social work
Information transmission, Software and information technology service	Culture, sports and entertainment
comprehensive	

out transactions in cooperation with core companies in another supply chain. The supply chain evolves from a chain structure to a network structure, with collaborative organization of production as the goal, and multiple economic entities forming a large-scale, interrelated and mutually influencing complex heterogeneous network through the relationship between supply and demand. Relevant studies also show that the supply network structure is directly related to supply chain risk elements, and the classical binary method commonly used in supply chain risk management cannot meet this challenge. Therefore, researchers are constantly looking for appropriate methods or developing new methods to analyze more complex supply chain risks from the perspective of network science. In terms of research methods about supply chain network. Bier et al. [14] concluded that these methods can be divided into qualitative (conceptual methods) and quantitative (direct risk determination and agency) methods. Hallikas et al. [15] presented a pure theoretical general supply chain network risk management method. Peck [16] proposed drivers of vulnerability in complex supply networks. Cheng and Kam [17] extended Peck's work and gave a conceptual framework based on the principal–agent concept. Some methods combine grey theory and directed graph methods. Although these research studies include quantitative methods, the main contribution is conceptual. Applying quantitative methods suggests analyzing the impact of risk through the probability distribution of negative events in a supply network. This can be achieved by calculating random performance index targets or entering the probability distribution into a simulation model, such as using value at risk to analyze the probability distribution of risk events and supply chain risk exposures [18–20]. Basole et al. [21] used Z-score, which is an indicator to measure the financial distress of enterprises. Kauppi et al. [22] used analysis of variance (ANOVA) to analyze the changes and fluctuations of countries involved in the supply process to develop a measurement standard of country risk.

2.2. Industry classification

Industry classification involves sorting companies with similar characteristics in the market into the same category [23]. As stated in the introduction, many scholars have demonstrated that present industry classification methods have not attained consensus [24]. Existing methods cannot adapt to the rapid change of economic environments [1] and fail to capture changes in industry structure [2,3], and are not sufficiently accurate [25]. In order to solve these problems, Fan and Lang [1] used an input–output (IO) table to measure the correlation between companies. Lenard et al. [26] used a fuzzy clustering algorithm to classify companies according to potential risks. Chong and Zhu [27] used labels to group companies in XBRL files. Their results show that the grouping results are inconsistent with NAICS classification, and show that the amount of information of NAICS scheme is not fruitful. Lee et al. [28] suggested using the Internet traffic pattern observed in the SEC EDGAR system (the official company financing platform managed by the SEC) to identify similar companies, which provides an alternative method of peer recognition. Fang et al. [29] applied text mining technology using the potential Dirichlet assignment algorithm to search business description text of companies. Companies were clustered according to the relevance of the business description. A recent text mining study on industry classification uses clustering method to check the similarity of words disclosed in the business description part of the annual report, and studies the changes in the industry based on the text, demonstrating that their industry classification method was more informative than SiC and NAICS codes [30]. Financial reporting structure was found to be an important attribute in review of comparability, reflecting the company operating results and indirectly reflecting the output of the selected resource allocation and strategy. Recently, scholars have used a graph similarity measure combined with spectral clustering algorithm to quantify the similarity of financial disclosure to classify companies accordingly [31]. In addition, distinguishable features that are extracted from the business description in financial reports have been used to obtain industry clusters in combination with machine learning technology [32].

2.3. Graph neural network

A graph neural network (GNN) is an improvement based on a convolutional neural network (CNN) and Long short-term memory (LSTM) commonly used on Euclidean data in order to apply deep learning to non-Euclidean data. Kipf and Welling [33] first proposed graph convolutionary networks (GCNs), the pioneering work of GNN. A series of subsequent work including the Graph Auto-Encoder (GAE) [34], Graph Attention Networks (GAT) [35], and Dynamic Graph Convolutionary Network (DyGCN) [36] can be viewed as optimization based on GCN. GNN was first applied to computer vision. The method of using graph structure has been widely applied, to include scene graph generation, point cloud classification and segmentation, and action recognition. Recent advances include human–object interaction, image classification, semantic segmentation, visual reasoning and question answering. GNNs application performance in natural language processing is encouraging. It is primarily used for semantic role annotation [37], neural machine translation [38], relationship extraction [39], question answering systems [40] and the construction of knowledge graphs. In the field of data mining, GNN has been used for social network analysis [41], recommendation systems [42], traffic flow prediction [43], malicious account detection and false news detection [44]. In the field of Biochemistry and medical health, GNN has been used for molecular representation learning [45], protein interaction interface prediction [46], drug target binding affinity prediction, drug similarity integration, drug side effects prediction [47], and disease prediction. In terms of classification tasks, GNN was originally used for image classification, especially in settings involving few samples and multiple

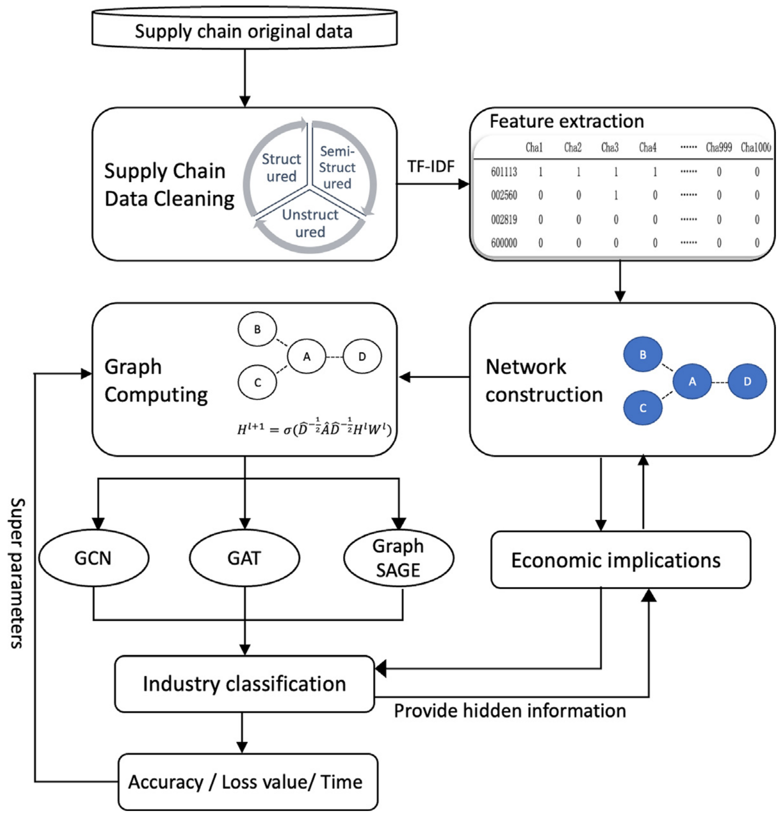

Fig. 1. Conceptual model of integrated method.

labels. For graph node classification, semi supervised graph node classification integrating node characteristics and node labels is widely used, that is, the labels of some nodes are given and the labels of unspecified nodes are predicted. Text classification can also be regarded as a special form of node classification. Text classification uses graph neural networks to analyze text. First, text data is transformed into graph data, then the corresponding GNN is designed, and a loss optimization model is designed for text classification. If the article is regarded as a node, text classification equals node classification. GNN has performed well in many fields, but it has not been widely applied in the field of economics and management.

3. Research design

As shown in Fig. 1, the proposed approach to perform the experiments involves four steps: Data collection & cleaning, Feature Extraction, Network Construction, and Graph Computing.

The graph neural network methods used in this paper are GCN, GAT and GraphSAGE [48]. Several machine learning models are used for comparison. Graph Convolution Networks (GCN) was the earliest algorithm to open the graph neural network field. Its essential purpose is to extract the spatial features of topology. Since then, work has focused on the optimization of the GCN model, resulting in a series of GCN variants such as GAT, GAE, GraphSAGE and DyGCN, which have upgraded and optimized the algorithm according to different sample characteristics. According to the static and small graph characteristics of our dataset, we select common algorithms GCN and GAT for verification and comparison, which is also convenient for further optimization of dataset characteristics. In this article, we will introduce these two models as briefly as possible.

3.1. Graph Convolution Networks

Convolution core formula of GCN is displayed in Formula (1).

$$H^{l+1} = \sigma(\widehat{D}^{-\frac{1}{2}}\hat{A}\widehat{D}^{-\frac{1}{2}}H^lW^l) \tag{1}$$

where H^l, H^{l+1} represents the node matrices of layer l and layer $l+1$ respectively, $\widehat{D}$ represents the diagonal node degree matrix of $\hat{A}$, where A represents adjacency matrix and $\hat{A} = A + I$. W_l is a weight matrix for the l_{th} neural network layer and $\sigma(\cdot)$ is a non-linear activation function like the ReLU. Fig. 2 gives an example of GCN.

GCN calculation is essentially a weighted summation process (Fig. 3), that is, neighbor nodes pass through the degree matrix and its adjacency matrix to calculate the weight of each edge, and then calculate weighted summation. D is responsible for providing the matrix of weights, the adjacency matrix A controls which points should be fused, and H represents the embedding parameters of the upper layer.

Fig. 2, we select a subgraph for operation, and the weight of edges is calculated by $\widehat{D}^{-\frac{1}{2}}\hat{A}\widehat{D}^{-\frac{1}{2}}$ in formula (1). The larger d, the smaller the amount of information and the smaller the weight $\frac{1}{\sqrt{d}}$

3.2. Graph Attention Networks

GAT adds an attention mechanism in the propagation process to overcome the disadvantage that the weight value of edges is fixed when using GCN model. The purpose of the attention layer is to enable calculation of the attention coefficient of node pairs (i, j) according to the different characteristics of neighbor nodes:

$$\alpha_{ij} = \frac{\exp\left(Leaky\ ReLU\left(a^T\left[Wh_i \| Wh_j\right]\right)\right)}{\sum_{k\in N_i}\exp\left(Leaky\ ReLU\left(a^T\left[Wh_i \| Wh_j\right]\right)\right)} \tag{2}$$

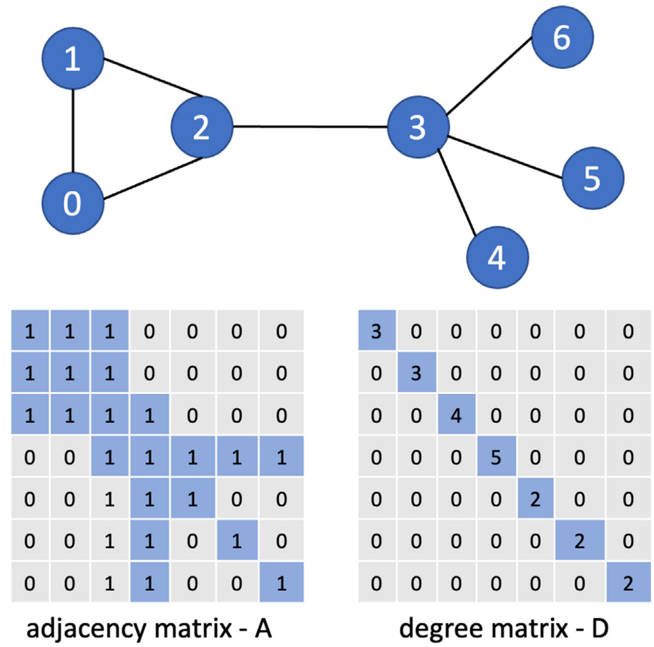

Fig. 2. An example of GCN calculating adjacency and degree matrix. This example considers self-circulation.

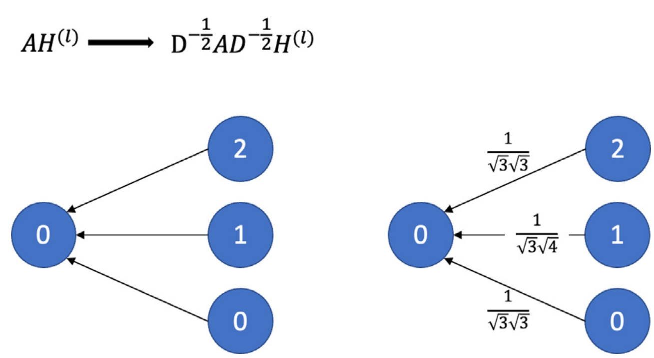

Fig. 3. Edge weights.

Here α_{ij} is the attention coefficient from node j to i, N_i represents the neighbor node of node i. The node input characteristics are $h = \{h_1, h_2, \ldots, h_N\}, h_i \in \mathbb{R}^F$, where N and F represent the number of nodes and characteristic dimension respectively. $W \in \mathbb{R}^{F' \times F}$ is the linear transformation weight matrix applied on each node, $\alpha \in \mathbb{R}^{2F'}$ represents a weight vector that can map input to R. Finally, SoftMax is used for normalization and LeakyReLU is added to provide nonlinearity. The input of the final node feature is $h' = \left\{h'_1, h'_2, \ldots, h'_N\right\}, h'_i \in R^{F'}$. h'_i calculated by:

$$h'_i = \sigma\left(\sum_{j \in N_i} \alpha_{ij} W h_j\right) \tag{3}$$

In addition, this layer also uses multi head attention to stabilize the learning process. It applies K_{th} independent attention mechanisms to calculate the hidden state, and then connects its features (or calculates the average value) to obtain the following two output representations:

$$h'_i = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{j \in N_i} \alpha_{ij}^k W^k h_j\right) \tag{4}$$

$$h'_i = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in N_i} \alpha_{ij}^k W^k h_j\right) \tag{5}$$

In Formula (4) and (5), α_{ij}^k represents the normalized attention coefficient of the K_{th} attention head, | | represents the splicing operation. The model details are displayed in Fig. 4.

4. Data

4.1. Data collection

Our data is taken from Wind, an authoritative financial data service provider in China (similar to Bloomberg in the United States), and CSMAR, an economic data service provider. These are recognized authoritative databases for the study of China's economic and management issues. The information of listed companies, including stock code, securities name, business scope, company profile, main products, etc., was downloaded from Wind. Supply chain information, including upstream and downstream companies, major customers and suppliers, and sales amount, etc., were downloaded from CSMAR. We manually selected a certain number of company information and financial statements for proofreading to ensure the reliability and accuracy of the data. Through observation of the data we found that the top five suppliers and customers of a large number of listed companies are non-listed enterprises, and the main suppliers and customers have not changed much within five years, which is in line with our expectations. Collecting complete information from a large number of unlisted enterprises is a huge project. In order to efficiently and accurately complete our research purposes, we retained supply chain information of listed companies both upstream and downstream from 2016 to 2021 to build supply chain networks and conform to the application paradigm of the graph neural network algorithm.

4.2. Dataset construction

The supply chain network dataset we constructed is based on the Cora dataset [49], which is usually used in the graph neural network and deep learning domain. The dataset is divided into two parts: feature construction and network construction. For ease of understanding, the following introduction is also divided into these two parts.

4.2.1. Feature extraction

The first step is to extract company characteristics. First, we mix the profiles, business scope, main business products and main product types of more than 900 companies to form an information base. We use Jieba word segmentation and TFIDF algorithm to sort words by the number of occurrences, and extract the top 1000 words with the highest frequency as features. We then compare the information of a specific company with the characteristics in the information base. If company information contains a word (feature) in the information base, it will be marked as 1, otherwise it will be marked as 0. Each company has 1000 information features. After marking all 904 companies, we obtain a 904 $*$ 1000 matrix of company features, the node feature table.

Step 1. Data cleaning and word segmentation. First, all wired symbols in Chinese text are deleted to ensure the accuracy of subsequent word segmentation. Then we use Jieba word segmentation library to segment sentences, and segment all sentences into words.

Step 2. The TF-IDF algorithm is used to calculate the weight of keywords after text segmentation in the company information

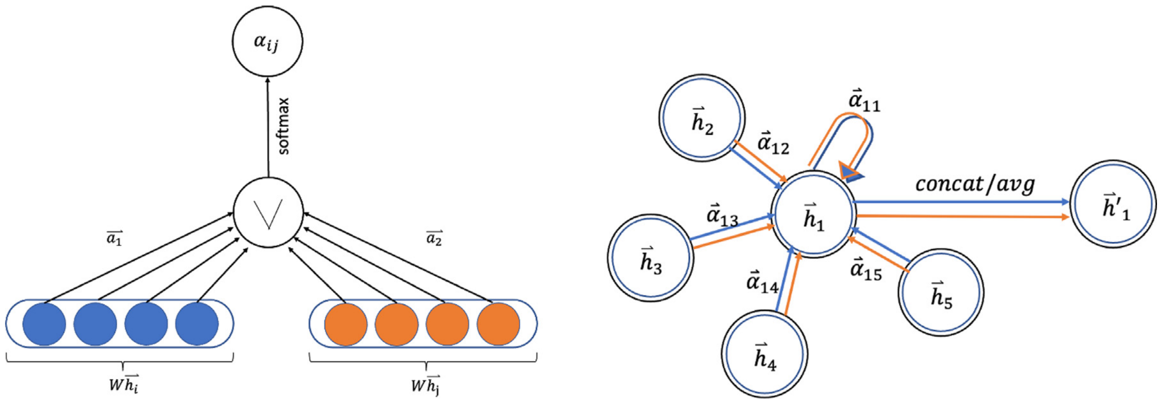

Fig. 4. The schematic diagram of GAT model.

Table 2
Examples of company characteristics.

ID	$c_{i,1}$	$c_{i,2}$	$c_{i,3}$	$c_{i,4}$	……	$c_{i,999}$	$c_{i,1000}$	Industry
601113	1	1	1	1	……	0	0	Manufacturing
002560	0	0	1	0	……	0	0	Manufacturing
002819	0	0	0	0	……	0	0	Wholesale and retail trade
600000	0	0	0	0	……	0	0	Finance
000796	0	0	1	1	……	0	0	Leasing and commercial service
000688	0	0	1	0	……	0	1	Mining
300811	1	0	1	0	……	0	1	Real estate
300117	0	1	1	1	……	0	0	Construction industry

base. TF-IDF is a statistic that measures the importance of a word to a document. It is widely used in natural language processing. Given a word or phrase t_i and a document d_z. Its TF-IDF value can be calculated as:

$$TFIDF_{i,z} = TF_{i,z} \times IDF_i = \frac{n_{i,z}}{\sum_{t_k \in d_z} n_{k,z}} \times \log \frac{|D|}{1 + |\{d_z \in D : t_i \in d_z\}|} \tag{6}$$

$n_{i,z}$ denotes the number of occurrences of the word t_i in document d_z. $|D|$ represents the total number of documents, $|\{d_z \in D : t_i \in d_z\}|$ represents the number of documents containing the word t_i. *Step 3* The TFIDF value calculated in *Step 2* is used to sort all word segmentation. The first N high-frequency words are taken as the feature information base C_N. The information for each company $c_{i,n}$ and feature information base C_N for comparison. If $c_{i,n} \in C$, then $c_{i,n}$ is marked as 1, otherwise marked as 0. This yields an $n \times N$ dimension characteristic matrix, completing feature extraction. Some examples of feature data extracted in this paper are shown in Table 2.

4.2.2. Construction of supply chain network

We use Python to process the supply chain information downloaded from CSMAR, screen the top five suppliers and top five customers of listed companies in the database. Non-listed companies are eliminated along with duplicates consistent with upstream and downstream within five years as well as companies that cannot match the characteristic data set. Finally, 904 listed companies (the number of nodes) are retained. Referring to the Cora dataset, we put the seller's stock code on the left side of the text and the buyer's stock code on the right, forming two columns. Each line represents connection of two nodes, and purchase and sales relationship forms an edge. There are 850 links in total, that is, 850 edges (Fig. 5). At least one company in the data set is both a seller and a customer of other companies, so the supply chain network constructed includes a secondary supply chain. This constitutes the supply chain diagram $\mathcal{G}_s = (\mathcal{V}, \mathcal{E})$, $\mathcal{V}$ represents the node in the graph, and $\mathcal{E}$ represents the edge. Further, we define each node i as x_i. The number of nodes is N.

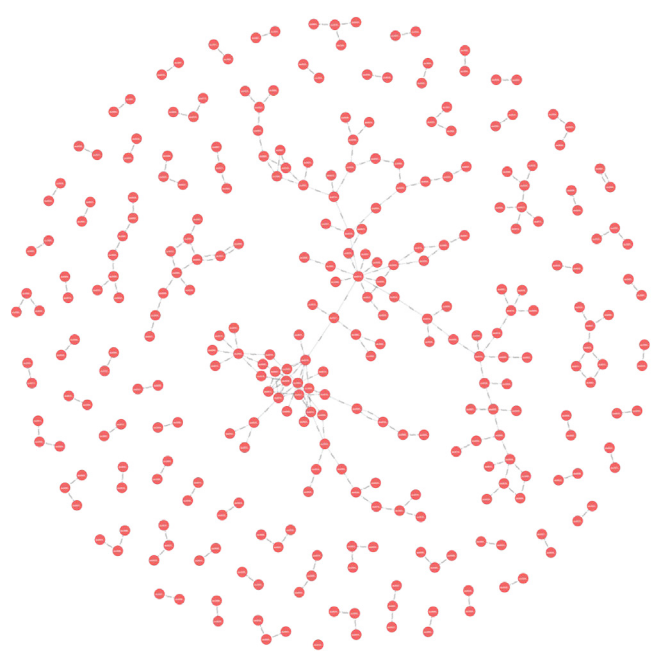

Fig. 5. Visualization of supply chain network dataset in Neo4j.

4.3. Economic variables

Table 3 lists some variables related to economic interpretation in our supply chain data set. Although they do not appear in the data set used by graph neural network, a deep understanding of these economic variables in the supply chain will help

Table 3
Economic variables in supply chain network.

Variable	Description	Pattern
Symbol	Securities codes published by Shanghai Stock Exchange and Shenzhen Stock Exchange	Character type number
InstitutionName	Name of listed company	Character
IsListed	Whether the customer is a listed company	Y/N
SalesAmount	Amount of sales of listed companies to customers in the current period	Amount of RMB
ProportionOfTotalValue	Proportion of sales to customers of Listed Companies in the total annual sales in the current period	Proportion
PurchaseAmount	Purchase amount of listed companies from suppliers in the current period	Amount of RMB
ProportionOfTotalValue	Proportion of purchase amount from suppliers of Listed Companies in total annual purchase amount in current period	Proportion

Table 4
Initial parameter setting.

Algorithm	Epochs	Learning rate	Training data sample	Verification data sample	Test data sample
GCN	800	0.01	300	200	404
GAT	Auto iteration	0.01			
Device: 2.7 GHz CPU, dual core Intel Core i5, and graphics card is Intel iris graphics 6100 1536 MB.					

us better understand the economic and management implications contained in the supply chain network to better enable us improve and explore the deep-seated principles and problems behind industry classification.

5. Experiments

5.1. Experimental design

This chapter will introduce the parameters we use in GCN and GAT models. Through many previous experiments, we set more reasonable parameters for researchers' reference, and the effect of the model is relatively better under these parameters.

For the data set, after randomly disrupting all samples, we selected 300 as the training set, 200 (300–500) as the verification data set and the remaining samples as the test data set. We set the number of samples in the training set to be less than the number of samples in the test set, which can ensure that the model does not rely too much on the training samples.

For GCN, on each cross-validation split, we train for 800 epochs (without early stopping) with a learning rate of 0.01.

For GAT, we followed the author's original model setting [35], initialization with Glorot method and training with AdamSGD optimizer. The initial learning rate of PubMed is 0.01, that of other data sets is 0.01 (the original parameter is 0.005). In both cases, we use the early stop strategy of cross entropy loss and verify the accuracy of nodes (transduction learning), and use the stop strategy of $micro - F1$ (inductive learning) score to train 100 epochs until the optimal result is learned. All settings are displayed in Table 4.

5.2. Experimental results

Results of the experiment are given in Table 5 and Fig. 6. The definition of *accuracy* is the ratio of the number of samples correctly classified by the classifier to the total number of samples for a given test data set. In this paper, *accuracy* refers to the number of companies that the classifier correctly divides them into their real industry divided by the total number of all companies in the test dataset. *Time* refers to the time consumed in the whole process of the model from the beginning of training to the end of testing. The experimental results show that the effect of GCN and GAT on the supply chain network dataset is better than the several traditional machine learning and neural network methods. The specific performance is that the accuracy is improved significantly, and the time consumed is greatly reduced. In addition, compared with the traditional classification methods, the accuracy of graph neural network is more stable and less affected by the change of random seed number. We also use another graph neural network algorithm GraphSAGE to find more clues. The classification effect of GraphSAGE is slightly worse than that of several other algorithms. This may be because GraphSAGE aggregates multiple nodes so that the model learns an aggregator rather than a representation for each node, which can improve the flexibility and generalization ability of the model. However, because the supply chain graph we constructed has small sample characteristics, multiple classification and relative sparsity, the aggregation effect of GraphSAGE is not good. Therefore the learning effect is not good and the accuracy is reduced.

Table 5
Results of different models.

No.	Model	Accuracy	Time (s)
1	Support Vector Machine (SVM)	0.6374	–[a]
2	DecisionTree (DT)	0.5728	–
3	Information Gain DT	0.6389	1189.95 s
4	KNeighbors	0.6523	–
5	Random Forest (RF)	0.6556	–
6	Bagging-CRAT	0.6716	297.76 s
7	GraphSAGE	0.6367	**5.85 s**
8	GCN	**0.6881**	6.86 s
9	GAT	0.6807	157.85 s

[a]These models are directly invoked in Python, which takes less than 1 s, but should not be directly compared.

We measure the performance of the models under different test ratio and visualize the result in the plots shown in Fig. 7. It shows that graph algorithms outperform all of the other algorithms in terms of accuracy rate, across different test ratio.

Other methods do not utilize the relationship between nodes and only classifies all companies based on company characteristics (see Fig. 8). Both GCN and GAT make use of the information transmitted by the edges between nodes, which is captured and used by graph neural network algorithm, so as to enrich the company (node) characteristics, therefore achieve better results in the industry classification task.

As can be seen from Fig. 9, when the GCN model is used to train the data set constructed in this paper, the loss value decreases rapidly in a few epochs, and then gradually decreases and tends to be flat with the increase of epoch times. The accuracy

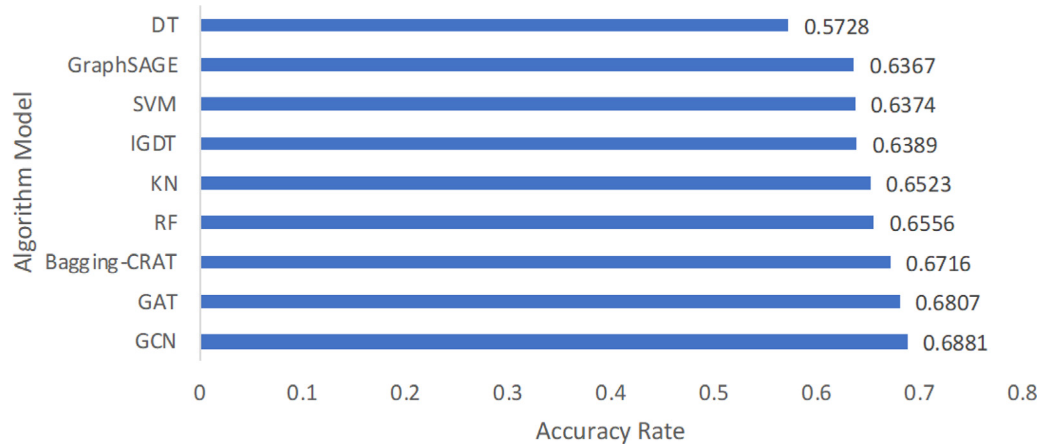

Fig. 6. The results of main experiment.

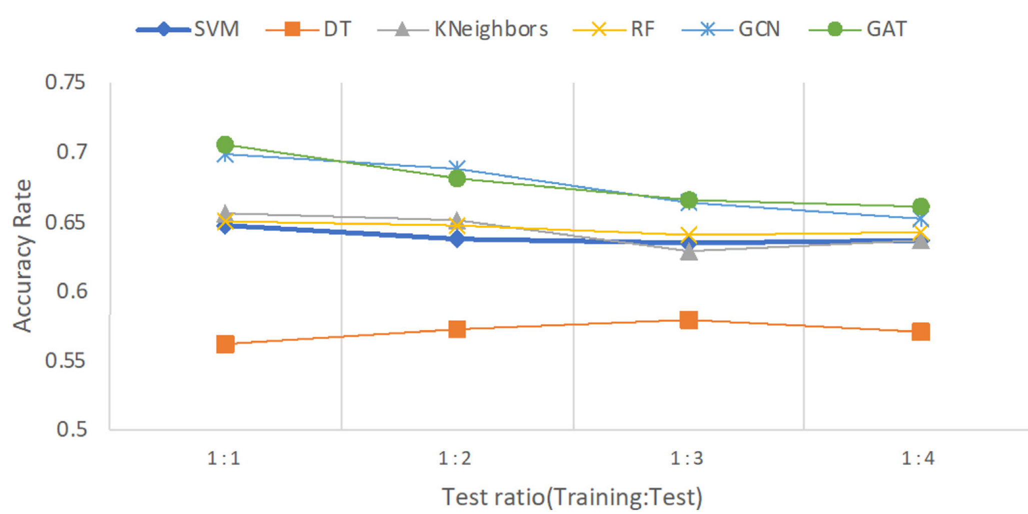

Fig. 7. Average accuracy across different test ratio.

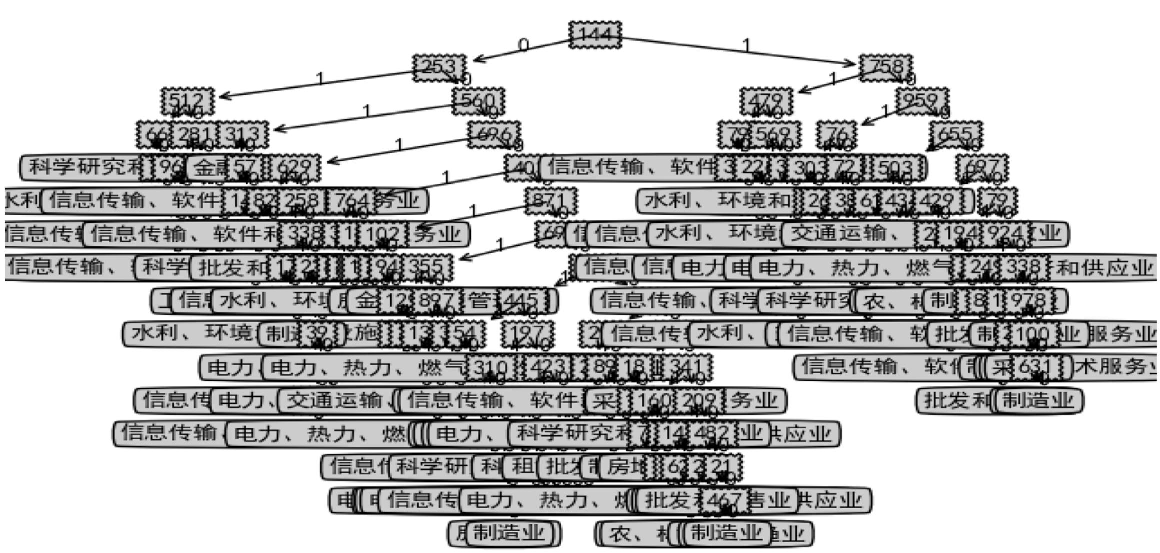

Fig. 8. Visualization of Decision Tree classification process. Chinese tags are reserved here.

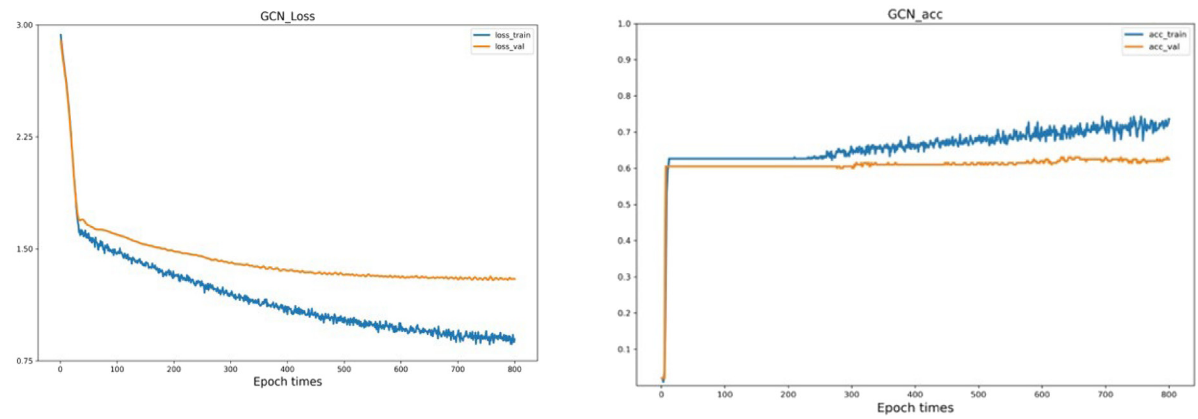

Fig. 9. The loss value and accuracy rate of running GCN on our data.

Table 6
Robustness check results.

Hidden units	LR	Accuracy	Time
8	0.1	0.6559	5.12 s
8	0.01	0.6559	5.38 s
8	0.001	0.6287	5.58 s
16	0.1	0.6881	5.43 s
16	0.01	0.6658	5.96 s
16	0.001	0.6287	5.34 s
32	0.1	0.6658	6.36 s
32	0.01	0.6881	6.73 s
32	0.001	0.6411	6.97 s

rate shows the same trend in the opposite direction. This result is in line with our expectations and confirms the effectiveness of our data set and the chosen method.

5.3. Robustness check

We followed the research article of Li et al. [50] and used different GCN superparametric combinations to test the robustness of the prediction model. Mainly by changing two super parameters — hidden units [8, 16, 32] and learning rate (LR) [0.001, 0.01, 0.1]. The reason why we did not choose GAT is that it uses the stop strategy $micro - F1$ method to train 100 epochs until the optimal result is learned. Its optimal result is fixed. The results for robustness check are presented in Table 6.

The values in Table 5 shows that the training time increases with the increase of the number of hidden units. Changing the values of the two super parameters of GCN can produce different model prediction accuracy rates, but they both fluctuate in a narrow range, indicating that the model we use is stable for our data set.

5.4. Economic explanation

Due to the "Black Box" property of graph neural networks [9], it is difficult to explain its effectiveness from the perspective of mathematical derivation. But we can try to explain why GNN can achieve better results in the industry classification task of companies under the supply chain network in combination with management and production practice. In the previous study, supplier–customer relationships have been proved to be a stable cooperative relationships among companies [51]. Coindustry relations are often considered to represent competitor relations [52]. Companies in the supply chain network cooperate with each other through transactions and credit to form a complete industrial operation. We randomly selected several pairs of upstream and downstream companies in the supply chain network data set to view their industry and relevant business information. We found that companies in the same supply chain network have competition and cooperation relationships, corresponding to previous research [53]. The two companies with trading relationships are often cooperative relationships, and their business activities are interrelated and inseparable. For example, YITAI GROUP(900948) belongs to the mining industry, HENGYUN GROUP(000531) belongs to electricity, heat, gas & water production and supply. HENGYUN's main customers (unlisted) belong to manufacturing and residential services. Their respective business activities are reflected in the business scope, company profile and main products collected in our data set, that is, the constructed characteristics.

However, due to the intersection of business scope and main products in different industries, if only their own characteristics are used, there will be too much noise and poor classification effect. Using GNN can capture the relationship between companies in the same supply chain network. Combined with the company characteristics and the transactions between upstream and downstream companies, GNN can judge the company's position in the supply chain network, and thus improve the effect of industry classification tasks. If a company in one industry accounts for a high share of transactions with companies in other industries, and is the main source of income for one company and the main source of cost for the other company, we call it a typical product (e.g. the sale of power produced by power plants to manufacturers or residential service providers). Combined with the economic variables in Table 2, the transaction volume and transaction amount of typical products and other information, the industry relationship can be indirectly reflected in the transactions of upstream and downstream companies. Through the results in Table 4, we have reason to believe that this economic information is implicitly included in the relationship between upstream and downstream companies. However, if we consider explicitly integrating product and transaction information into the supply chain network as the weight of heterogeneous nodes or edges, it may yield better results. This is also one of the future research directions. A recent article shows that adding an implicit concept based on the predefined concept can better predict the stock trend. This kind of implicit concept includes the implicit relationship between companies, which also supports the conclusion of our paper [54].

For the diversified classification of the company, as mentioned above, the industries marked in HENGYUN belong to electricity, heat, gas & water production and supply. Nevertheless, its financial report shows that in addition to the power business, more than 10% of its revenue comes from the real estate business. These businesses are expected to be reflected through the supply chain network relationship, so is labeled a diversified industry. However, it should be noted that suppliers and customers do not require mandatory disclosure in China's financial reports, and the disclosure is only the top five suppliers and top five customers, which may not fully reflect the company's business relationship. Despite all this, our method can provide reference for institutions with rich data.

6. Discussion and conclusions

A hundred years ago, most of the world's industries were embodied in vertical structures, and all companies were centrally coordinated. In the era of economic globalization, countries and regions all over the world participate in chains sharing the division of labor. The complexity of world economic activities makes global supply chains more complex and to some extent more fragile. Especially in the post-epidemic period, countries around the world are concerned about supply chain security. Due to division of labor inherent in today's supply chains, many multinational companies are more familiar with their primary suppliers and have limited understanding of secondary and tertiary suppliers. If a black swan incident occurs, such as the current global epidemic, supply chain resilience will be greatly reduced.

This paper provides a new perspective for the study of companies in supplier networks. The graph neural network method applied here has achieved good results in industry classification compared with traditional machine learning methods. This is accomplished through data mining utilizing the implicit relationship between upstream and downstream companies. This work not only enriches the application of deep learning in the field of supply chains, it also enriches the connotation of risk intelligence and provides a means for industry to solve practical engineering problems. Combined with the work of Kosasih & Brintrup and others, we believe that the application of a graph neural network to supply chains is not limited to node classification and link prediction, but may be applied in supply chain network risk management in the future, to include supply chain risk factor mining, multi-dimensional risk assessment, etc. For example, risk analysis is often based on surface data, and the hidden risks are difficult to identify and measure. Our work suggests that researchers can construct a network structure between entities enabling mining of implicit associations and implicit attributes by using deep learning in the process of risk measurement.

There are several possible extensions to this work. From the perspective of data, the data set we constructed only contains Chinese listed companies. After excluding non-listed companies, the graph structure becomes sparse. Although this also proves the effectiveness of GNN, the larger and denser data after adding the information of non-listed companies may improve results. From the perspective of graph structure, we construct a static graph without taking into account changes in a company's supply chain, which is sufficient for the task of classification. However, if we want to study other supply chain problems in depth, we should take into account supply chain timing changes, especially when the company faces adverse economic conditions. Second, all the nodes in the graph are companies without considering other stakeholders. When the model includes retailers, consumers or other external stakeholders, it is necessary to construct heterogeneous graphs containing different node attributes. Current graph neural network algorithms have evolved towards more complex graph structures using tools such as dynamic graphs and heterogeneous graphs, although some research has pointed out that the effectiveness of these more complex algorithms may not compare well with GCN, GAT and other basic graph neural network algorithms. Therefore, the rapid development of GNN also makes it interesting to further study more complex supply chain networks and utilize more hidden information. Optimization of the above steps can not only make GNN play a greater role in the supply chain network, but also improve the accuracy and efficiency of the task of this paper — industry classification. The complex supply chain network structure contains rich space–time, transaction, investment and other information. Integrating multi factor information can help us better understand the company's business, enabling more accurate classification. In the future, we are going to apply our methods in large-scale knowledge graphs and supply chain networks, and continue to overcome the aforementioned limitations in the application process. We will also further explore the hidden information behind supply chain networks, and try to use a visuals to present it intuitively, as well as combine the quantitative data in the supply chain to conduct economic and management analysis. We also intend to use causal inference to reveal why graph neural network is effective in the analysis of supply chain networks.

CRediT authorship contribution statement

Desheng Wu: Oversight, Idea, Management. **Quanbin Wang:** Draft, Computations. **David L. Olson:** Review, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported in part by the Ministry of Science and Technology of China under Grant 2020AAA0108400 and Grant 2020AAA0108402 and the National Natural Science Foundation of China (NSFC) under Grant 71825007, in part by the National Natural Science Foundation of China (NSFC) under Grant 72210107001.

References

[1] J.P.H. Fan, L.H.P. Lang, The measurement of relatedness: An application to corporate diversification, J. Bus. 73 (4) (2000) 629–660, Univ Chicago Press, Chicago.

[2] J.E. Carrillo, C. Druehl, J. Hsuan, Introduction to innovation WITHIN and ACROSS borders: A review and future directions, Decis. Sci. 46 (2) (2015) 225–265, Wiley, Hoboken.

[3] A.H. Segars, V. Grover, The industry-level impact of information technology: An empirical analysis of three industries, Decis. Sci. 26 (3) (1995) 337–368.

[4] J. Wu, J. He, J. Xu, DEMO-Net: Degree-specific graph neural networks for node and graph classification, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery, New York, NY, USA, 2019, pp. 406–415.

[5] Y. Xie, C. Yao, M. Gong, C. Chen, A.K. Qin, Graph convolutional networks with multi-level coarsening for graph classification, Knowl.-Based Syst. 194 (2020) 105578, Elsevier, Amsterdam.

[6] L. Huang, D. Ma, S. Li, X. Zhang, H. Wang, Text level graph neural network for text classification, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019, 2019, pp. 3442–3448.

[7] R. You, S. Yao, H. Mamitsuka, S. Zhu, DeepGraphGO: graph neural network for large-scale, multispecies protein function prediction, Bioinformatics 37 (2021) I262–I271, Oxford Univ Press, Oxford.
[8] H. Park, J. Neville, Exploiting interaction links for node classification with deep graph neural networks, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Presented at the Twenty-Eighth International Joint Conference on Artificial Intelligence {IJCAI — 19}, International Joint Conferences on Artificial Intelligence Organization, Macao, China, 2019, pp. 3223–3230.
[9] E.E. Kosasih, A. Brintrup, A machine learning approach for predicting hidden links in supply chain with graph neural networks, Int. J. Prod. Res. (2021) 1–14.
[10] P. Wichmann, A. Brintrup, S. Baker, P. Woodall, D. McFarlane, Extracting supply chain maps from news articles using deep neural networks, Int. J. Prod. Res. 58 (17) (2020) 5320–5336, Taylor & Francis Ltd, Abingdon.
[11] T.J. Strader, F.-R. Lin, M.J. Shaw, Information infrastructure for electronic virtual organization management, Decis. Support Syst. 23 (1) (1998) 75–94.
[12] S. Roscoe, P.D. Cousins, R.C. Lamming, Developing eco-innovations: a three-stage typology of supply networks, J. Clean. Prod. 16 (pt.3) (2016) 1948–1959.
[13] A. Surana, S. Kumara, M. Greaves, Supply chain networks: A complex adaptive systems perspectrive, Int. J. Prod. Res. 43 (20) (2005) 4235–4265.
[14] T. Bier, A. Lange, C.H. Glock, Methods for mitigating disruptions in complex supply chain structures: a systematic literature review, Int. J. Prod. Res. 58 (6) (2020) 1835–1856, Taylor & Francis Ltd, Abingdon.
[15] J. Hallikas, I. Karvonen, U. Pulkkinen, V. Virolainen, M. Tuominen, Risk management processes in supplier networks, Int. J. Prod. Econ. 90 (1) (2004) 47–58.
[16] H. Peck, Driveers of supply chain vulnerability: An integrated framework, Int. J. Phys. Distrib. Logist. Manage. 35 (4) (2005) 210–232.
[17] S.K. Cheng, B.H. Kam, A conceptual framework for analysing risk in supply networks, J. Enterp. Inf. Manag. 21 (4) (2008) 345–+.
[18] A. Madadi, M.E. Kurz, K.M. Taaffe, J.L. Sharp, S.J. Mason, Supply network design: Risk-averse or risk-neutral? Comput. Ind. Eng. 78 (2014) 55–65.
[19] K.J. Mizgier, Global sensitivity analysis and aggregation of risk in multi-product supply chain networks, Int. J. Prod. Res. 55 (1) (2017) 130–144.
[20] K.J. Mizgier, S.M. Wagner, M.P. Juettner, Disentangling diversification in supply chain networks, Int. J. Prod. Econ. 162 (2015) 115–124.
[21] R.C. Basole, M.A. Bellamy, H. Park, J. Putrevu, Computational analysis and visualization of global supply network risks, IEEE Trans. Ind. Inform. 12 (3) (2016) 1206–1213.
[22] K. Kauppi, A. Longoni, F. Caniato, M. Kuula, Managing country disruption risks and improving operational performance: risk management along integrated supply chains, Int. J. Prod. Econ. 182 (2016) 484–495.
[23] R.N. Clarke, SICs as delineators of economic markets, J. Bus. 62 (1) (1989) 17–31, University of Chicago Press.
[24] K.M. Kahle, R.A. Walkling, The impact of industry classifications on financial research, J. Financ. Quant. Anal. 31 (3) (1996) 309–335, Univ Washington, Seattle.
[25] M. Dalziel, A systems-based approach to industry classification, Res. Policy 36 (10) (2007) 1559–1574, Elsevier Science Bv, Amsterdam.
[26] M.J. Lenard, P. Alam, D. Booth, An analysis of fuzzy clustering and a hybrid model for the auditor's going concern assessment, Decis. Sci. 31 (4) (2000) 861–884, Decision Sciences Inst, Georgia State Univ, Atlanta.
[27] D. Chong, H. Harry Zhu, Firm clustering based on financial statements, in: 22nd Workshop on Information Technology and Information Systems (WITS'12), December 15-16, 2012, Orlando, Florida, 2012, Available at SSRN: https://ssrn.com/abstract=2319313.
[28] C.M.C. Lee, P. Ma, C.C.Y. Wang, Search-based peer firms: Aggregating investor perceptions through internet co-searches, J. Financ. Econ. 116 (2) (2015) 410–431, Elsevier Science Sa, Lausanne.
[29] F. Fang, K. Dutta, A. Datta, LDA-based industry classification, in: Proceedings of the International Conference on Information Systems, ICIS 2013, Milano, Italy, December 15-18, 2013, 2013, available at: http://aisel.aisnet.org/icis2013/proceedings/ResearchInProgress/36.
[30] G. Hoberg, G. Phillips, Text-based network industries and endogenous product differentiation, J. Polit. Econ. 124 (5) (2016) 1423–1465, Univ Chicago Press, Chicago.
[31] S.Y. Yang, F. Liu, X. Zhu, D.C. Yen, A graph mining approach to identify financial reporting patterns: An empirical examination of industry classifications, Decis. Sci. 50 (4) (2019) 847–876.
[32] D. Kim, H.-G. Kang, K. Bae, S. Jeon, An artificial intelligence-enabled industry classification and its interpretation, Internet Res. (2021) ahead-of-print No. ahead-of-print, available at: https://doi.org/10.1108/INTR-05-2020-0299.
[33] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, 2017, pp. 1–14.
[34] T.N. Kipf, M. Welling, Variational graph auto-encoders, 2016, arXiv:1611.07308 [Cs, Stat], available at: http://arxiv.org/abs/1611.07308 (accessed 31 December 2021).
[35] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, 2018, available at: https://openreview.net/forum?id=rJXMpikCZ.
[36] Z. Cui, Z. Li, S. Wu, X. Zhang, Q. Liu, L. Wang, M. Ai, Dygcn: Dynamic graph embedding with graph convolutional network, 2021, arXiv:2104.02962 [Cs], available at: http://arxiv.org/abs/2104.02962 (accessed 31 December 2021).
[37] D. Marcheggiani, I. Titov, Encoding sentences with graph convolutional networks for semantic role labeling, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017, 2017, pp. 1506–1515.
[38] D. Marcheggiani, J. Bastings, I. Titov, Exploiting semantics in neural machine translation with graph convolutional networks., in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers), 2018, pp. 486–492.
[39] T.-J. Fu, P.-H. Li, W.-Y. Ma, GraphRel: Modeling text as relational graphs for joint entity and relation extraction, in: Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, 2019, pp. 1409–1418.
[40] N.D. Cao, W. Aziz, I. Titov, Question answering by reasoning across documents with graph convolutional networks, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), 2019, pp. 2306–2317.
[41] C. Li, D. Goldwasser, Encoding social information with graph convolutional networks forpolitical perspective detection in news media., in: Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers, 2019, pp. 2594–2604.
[42] R. Ying, R. He, K. Chen, P. Eksombatchai, W.L. Hamilton, J. Leskovec, Graph convolutional neural networks for web-scale recommender systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Presented at the KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, London United Kingdom, 2018, pp. 974–983.
[43] X. Wang, Y. Ma, Y. Wang, W. Jin, X. Wang, J. Tang, C. Jia, et al., Traffic flow prediction via spatial temporal graph neural network, in: Web Conference 2020: Proceedings of the World Wide Web Conference (Www 2020), Assoc Computing Machinery, New York, 2020, pp. 1082–1092.
[44] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (6380) (2018) 1146–1151.
[45] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, R.P. Adams, Convolutional networks on graphs for learning molecular fingerprints, in: Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, 2015, pp. 2224–2232.
[46] A. Fout, J. Byrd, B. Shariat, A. Ben-Hur, Protein interface prediction using graph convolutional networks, in: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, 2017, pp. 6530–6539.
[47] Z. Marinka, A. Monica, L. Jure, Modeling polypharmacy side effects with graph convolutional networks, Bioinformatics (13) (2018) i457–i466.
[48] W.L. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs., in: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, 2017, pp. 1024–1034.
[49] L. Getoor, Link-based classification, in: S. Bandyopadhyay, U. Maulik, L.B. Holder, D.J. Cook (Eds.), Advanced Methods for Knowledge Discovery from Complex Data, Springer, London, 2005, pp. 189–207.
[50] Y. Li, S. Jiang, Y. Wei, S. Wang, Take bitcoin into your portfolio: a novel ensemble portfolio optimization framework for broad commodity assets, Financ. Innov. 7 (1) (2021) 63, Springer, New York.

[51] Di, Shock Spillover and Financial Response in Supply Chain Networks: Evidence from Firm-Level Data, Working paper, University of Chicago, Chicago, 2016, Available at: http://www.usc.edu/schools/business/FBE/seminars/papers/RECR_1-29-16_F_WUjmp.pdf (accessed 25 December 2021).

[52] D.R. Gnyawali, J. He, R. Madhavan, Impact of co-opetition on firm competitive behavior: An empirical examination, J. Manag. 32 (4) (2006) 507–530, Sage Publications Inc, Thousand Oaks.

[53] K. Chen, X. Li, P. Luo, J.L. Zhao, News-induced dynamic networks for market signaling: Understanding the impact of news on firm equity value, Inf. Syst. Res. 32 (2) (2021) 356–377.

[54] W. Xu, W. Liu, L. Wang, Y. Xia, J. Bian, J. Yin, T.-Y. Liu, HIST: A graph-based framework for stock trend forecasting via mining concept-oriented shared information, 2021, arXiv:2110.13716 [Cs, q-Fin], available at: http://arxiv.org/abs/2110.13716 (accessed 24 December 2021).