

A Review on Graph Neural Network Methods in Financial Applications

JIANIAN WANG¹, SHENG ZHANG¹, YANGHUA XIAO^{2,*}, AND RUI SONG^{1,*}

¹*Department of Statistics, North Carolina State University, Raleigh, United States*

²*School of Computer Science, Fudan University, Shanghai, China*

Abstract

With multiple components and relations, financial data are often presented as graph data, since it could represent both the individual features and the complicated relations. Due to the complexity and volatility of the financial market, the graph constructed on the financial data is often heterogeneous or time-varying, which imposes challenges on modeling technology. Among the graph modeling technologies, graph neural network (GNN) models are able to handle the complex graph structure and achieve great performance and thus could be used to solve financial tasks. In this work, we provide a comprehensive review of GNN models in recent financial context. We first categorize the commonly-used financial graphs and summarize the feature processing step for each node. Then we summarize the GNN methodology for each graph type, application in each area, and propose some potential research areas.

Keywords *deep learning; finance; graph convolutional network; graph representation learning*

1 Introduction

As the data collection techniques grow, graph data are commonly collected in many areas including social sciences, transportation systems, chemistry, and physics (Wu et al., 2020). Representing complex relational data, graph data contain both the individual node information and the structural information. Recently, there are growing interests in developing machine-learning methods to model the graph data of various domains. Among them, graph neural network (GNN) methods could achieve great performance on various tasks, including node classification, edge prediction, and graph classification (Kipf and Welling, 2017; Zhang and Chen, 2018; Xu et al., 2018). Performing node aggregation and updates, graph neural network models extend the deep learning methodology to graphs and are gaining popularity.

A financial system is a complex system with many components and sophisticated relations, which may be frequently updated. To represent the relational data in the financial domain, graphs are commonly constructed, including the transaction network (Weber et al., 2019), user-item review graph (Dou et al., 2020), and stock relation graph (Feng et al., 2019). By converting the financial task into a node classification task, GNN methods are commonly utilized since it performs well among graph modeling methods (Liu et al., 2019). For instance, GNN could be utilized in a stock prediction task, by formulating it as a node classification task, where each node represents a stock and edges represent relations between companies. Figure 1 demonstrates the workflow of a stock prediction task using GNN methods. However, the complex nature

*Corresponding author. Email: shawyh@fudan.edu.cn or rsong@ncsu.edu.

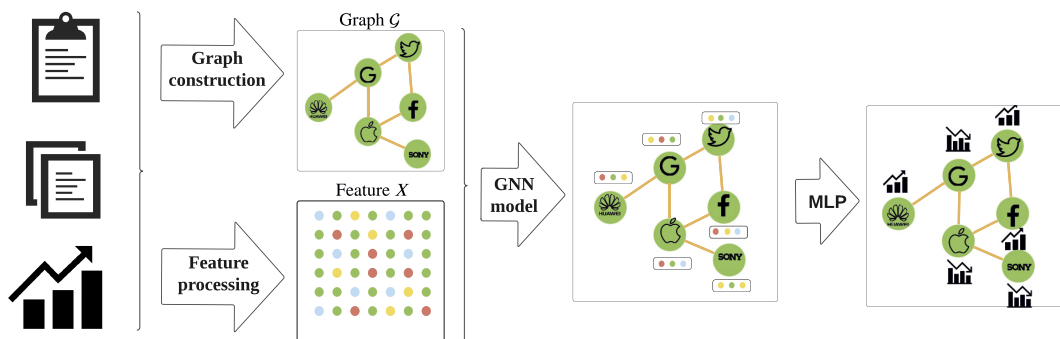


Figure 1: Workflow for stock movement prediction task using GNN methodology. The graph construction and feature processing steps present stock information in a graph and a feature matrix, which is then used as the input for the GNN model. In the graph, nodes are connected if there exist some relationships between stocks, such as supplier, competitor, shake-holder, etc. A multi-layer perception layer (MLP) is used to output the price prediction result.

of financial systems may result in multiple data sources and complicated graph structures, which imposes challenges on feature processing, graph construction, and graph neural network modeling. Represented as numerical sequences or textual information, financial data need to be processed with caution to keep the temporal pattern or semantic meanings. Also, the multi-facet nature of financial relations make it hard to construct a graph to capture the relations. Moreover, the financial-related graph is often heterogeneous or time-varying, which impose challenges on existing graph neural network models. What’s more, to reflect some financial patterns (e.g. device aggregation pattern, see Section 5.4 for details), GNN methods may need to be modified such as changing losses and adding additional layers. Since financial systems process unique characteristics and receive great attention, it is of significant importance to discuss and summary the GNN methodology developed for financial tasks.

There are several recent reviews on graph neural networks. Among them, Wu et al. (2020) present a comprehensive review on graph neural networks and categorize the GNNs into four categories: recurrent graph neural networks, convolutional graph neural networks, graph auto-encoders, and spatial-temporal graph neural networks. Zhou et al. (2020) provide a taxonomy on GNN models based on graph type, training methods, and propagation steps. There is also literature focusing on limited types of GNNs. Zhang et al. (2019) focus on graph convolutional networks (GCN) and introduce two taxonomies to group the existing GCNs. Lee et al. (2019) survey the literature on graph attention models and provided detailed examples on each type of method. However, the aforementioned reviews focus on the general methodology and provide little details for applications, seldom mentioning the financial application. Without covering GNN models developed based on financial contexts, the reviewed models may not be applicable to financial tasks due to the complexity of financial data. On the other side, review papers focusing on the financial domain haven’t covered GNN methodologies in detail yet. Ozbayoglu et al. (2020) summarize the machine learning and deep learning models in the financial field, without mentioning the GNN methodologies. Huang et al. (2020) survey the financial deep learning models in the finance and bank industry, and the GNN models are not covered. Jiang (2021) review stock prediction-related machine-learning mythologies and mention GNN models very briefly. In summary, existing GNN surveys focus on modeling methodology and do not emphasize the financial application of GNN methods, while surveys on financial applications

don't cover the GNN models in detail. To fill the gap, in this survey, we provide a systematic and comprehensive review of graph neural network methods in the financial application.

In this paper, we present a thorough survey on graph neural network models with financial application. We provide a comprehensive review of graph neural networks and summarize the corresponding methods. This survey has contributions as follows.

- We systemically categorize the commonly-used financial graphs based on graph characteristics and provide a thorough list of graphs. Graphs are categorized into five groups: homogeneous graph, directed graph, bipartite graph, multi-relation graph, and dynamic graph. We also present the GNN models according to their graph types, so that this review could serve as a guide for implementing GNNs on real-life datasets.
- We provide a comprehensive list of financial applications that GNN methods are applied. We categorized the applications into five categories: stock movement prediction, loan default risk prediction, recommender system of e-commerce, fraud detection, and event prediction.
- We summarize various aspects of information for each application, including features, graphs, GNN models, and available codes. A GitHub¹ page is built to document the collection of information. This work could be considered as a resource to understand, implement and develop GNN models on multiple financial tasks.
- We identify five challenges and discuss the recent progress. We also suggest future directions for these problems.

The rest of the paper is organized as follows. Section 2 classifies financial graphs into different categories based on its characteristics. Section 3 summarizes the commonly-used feature processing techniques for each node in the graph. Section 4 presents the GNN methodology used for each graph type. Section 5 provides a collection of application areas. Section 6 proposes some challenges that could be future directions of research.

2 Graph Categorization

When preparing the data, how to construct the graph to represent the structural information is essential and the type for the constructed graph could determine the follow-up modeling methodology. In this section, we present the categorization of the graph based on its construction methods and graph types. Table 1 presents a comprehensive list of graphs for financial tasks.

2.1 Graph-Related Definition

In this section, we provide some graph-related definitions for better understanding of this article.

Definition 2.1 (Graph). A graph \mathcal{G} is defined by a pair: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, \dots, v_n\}$ is a set of n nodes and \mathcal{E} is a set of edges, where $e_{ij} = (v_i, v_j) \in \mathcal{E}$ denotes an edge joining node v_i and node v_j .

Definition 2.2 (Adjacency matrix). An adjacency matrix \mathbf{A} is an $n \times n$ matrix, where \mathbf{A}_{ij} represents the connection status between node v_i and node v_j . For an unweighted graph, the adjacency matrix could be a binary matrix where $\mathbf{A}_{ij} = 1$ if $e_{ij} \in \mathcal{E}$ and 0 otherwise.

¹Github link: <https://github.com/jackieD14/Graph-models-in-finance-application>

Table 1: Summary of financial-related graphs.

Graph	Construction method ²	Graph type ³	Application ⁴⁵	Reference
Sector-industry stock relation network	Knowledge	Multi	Stock	Feng et al. (2019)
Wiki company-based relation network	Knowledge	Multi	Stock	Feng et al. (2019); Sawhney et al. (2020a); Ying et al. (2020)
Supplier, customer, partner and shareholder relation graph	Knowledge	Multi	Stock	Matsunaga et al. (2019)
Corporation shareholder network	Knowledge	Homo	Stock	Chen et al. (2018)
Stock correlation graph	Similarity	Multi	Stock	Li et al. (2020a)
Stock earning call graph	Knowledge	Bipartite	Stock	Sawhney et al. (2020b)
News co-occurrence graph	Similarity	Homo	Stock	Liou et al. (2021)
User-relation graph	Data	Homo	Loan	Wang et al. (2019)
User-app graph	Data	Bipartite	Loan	Wang et al. (2019)
User-nickname graph	Data	Bipartite	Loan	Wang et al. (2019)
User-address graph	Data	Bipartite	Loan	Wang et al. (2019)
Guarantee network	Data	Directed	Loan	Cheng et al. (2019, 2020)
Temporal guarantee network	Data	Dynamic	Loan	Cheng and Zhang L Wang X (2020b)
Temporal small business entrepreneur network	Data	Dynamic	Loan	Yang et al. (2020)
Alipay user and applet graph	Data	Dynamic	Loan	Hu et al. (2020)
User relationship graph	Data	Multi	Loan	Liang et al. (2021)
Auto relation network	Data	Multi	Loan	Xu et al. (2021)
Loan application event graph	Data	Directed	Loan	Harl et al. (2020)
Borrower relations' network	Similarity	Directed	Loan	Lee et al. (2021)
Xianyu comment graph	Similarity	Homo	E-comm	Li et al. (2019a)
Yelp review network	Data	Bipartite	E-comm	Zhang et al. (2020); Dou et al. (2020)
Amazon review network	Data	Bipartite	E-comm	Zhang et al. (2020); Dou et al. (2020); Kudo et al. (2020)
E-commerce user-item network	Data	Bipartite	E-comm	Li et al. (2019b)
Taobao user-item network	Data	Bipartite	E-comm	Li et al. (2020c)
Device sharing graph	Data	Bipartite	Fraud	Liang et al. (2019)
JD Finance anti-fraud graph	Data	Multi	Fraud	Lv et al. (2019)
Transaction records graph	Data	Multi	Fraud	Rao et al. (2020)
Iqiyi user network	Data	Multi	Fraud	Zhu et al. (2020)
CMU simulated user activity network	Similarity	Homo	Fraud	Jiang et al. (2019)
Alipay one-month account-device network	Data	Bipartite	Fraud	Liu et al. (2018)
Alipay one-week account-device network	Data	Bipartite	Fraud	Liu et al. (2019)
Account-registration graph	Data	Dynamic	Fraud	Rao et al. (2020)
Bitcoin-alpha graph	Data	Directed	Fraud	Zhao et al. (2021)

Definition 2.3 (Undirected graph and Directed graph). A undirected graph is a graph where the edges are undirected. A directed graph is a graph where the edges have orientations. $e_{ij} = (v_i, v_j) \in \mathcal{E}$ denotes an edge pointing from node v_i to node v_j .

Remark. Undirected graph has a symmetric adjacency matrix, i.e., $\mathbf{A}_{ij} = \mathbf{A}_{ji}$.

Definition 2.4 (Bipartite graph). A Bipartite graph is a graph whose nodes could be divided into two non-empty and disjoint sets \mathcal{U} , \mathcal{W} , such that every edge connects a node in \mathcal{U} and a node in \mathcal{W} .

Definition 2.5 (Homogeneous graph and Heterogeneous graph). In a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we can assign a type to each node and edge; in this case, the graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{R})$, where each node $v_i \in \mathcal{V}$ is associated with its type $a_i \in \mathcal{A}$, and each edge $e_{ij} \in \mathcal{E}$ is associated with its type $r_{ij} \in \mathcal{R}$. A homogeneous graph is a graph whose nodes are of the same type and edges are of the same type. Otherwise, the graph is heterogeneous.

Definition 2.6 (Multi-relation graph). A Multi-relation graph is a graph where edges have different types.

Definition 2.7 (Dynamic graph). A dynamic graph is defined as a sequence of graphs $\mathcal{G}^{seq} = \{\mathcal{G}_1, \dots, \mathcal{G}_T\}$, where $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$, for $i = 1, \dots, T$, where \mathcal{V}_i , \mathcal{E}_i are the set of nodes and edges for i th graph in the sequence respectively.

2.2 Graph Categorization by Construction Methods

In this section, we elaborate on frequently-used graph construction methods so that researchers could better understand, choose and construct graph data.

2.2.1 Data-Based Construction

Some types of data could be naturally represented as a graph since they contain relations among data objects. For instance, Wang et al. (2019) construct a user-relationship graph, where users are linked by an edge if they are labeled classmates, friends, or workmates in the data. Liu et al. (2018) construct an account-device network, where nodes are either accounts or devices. Edges connect an account node to a device node if the account has activities on the device. This type of construction method is based on the nature of data and could be used on data where relations are clearly defined.

2.2.2 Knowledge-Based Construction

Sometimes, data may not contain relational information, but relations could be found in knowledge bases. A knowledge base is a collection of descriptive data and contains numerous entities and their relations. For example, Wikidata (Vrandečić and Krötzsch, 2014) is one of the largest open-domain knowledge bases which provides support for Wikipedia, an online encyclopedia. A graph could be built utilizing the relations extracted from knowledge bases. In order to predict stock movement, Feng et al. (2019) extract company relations from Wikidata, such as supplier, provider, partner, etc, and constructed a company-based relation network. This type of graph-construction method brings new information into the graph by utilizing the knowledge bases, which may improve modeling performance. However, it may take some effort to process the complicated data structure and the massive amount of information of knowledge bases.

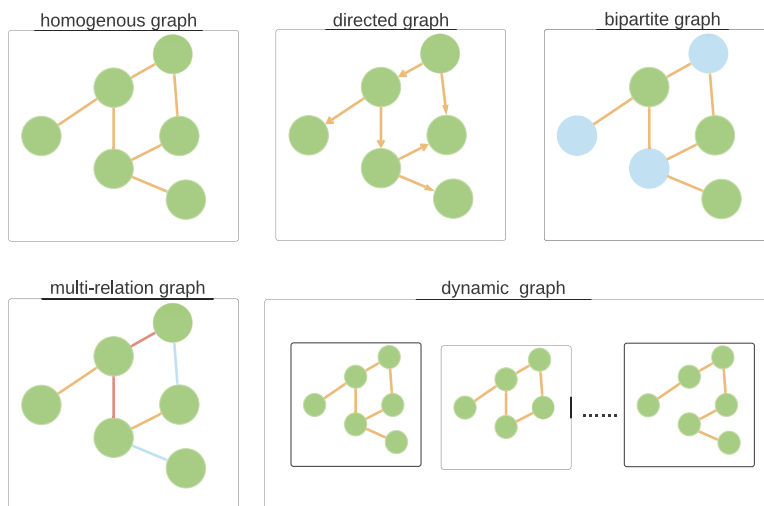


Figure 2: Graph categorization based on graph characteristics. Each color of the circle represents a node type and each color of the line represents an edge type. Arrows represent directed edges. A homogeneous graph is a graph with one type of node and one type of edge. A directed graph is a graph with directed edges. A bipartite graph is a graph with two types of nodes and edges only exist between nodes of different types. A multi-relation graph has edges with different types. A dynamic graph is a sequence of graphs.

2.2.3 Similarity-Based Construction

There also exist cases that neither the data nor knowledge bases contain relations, but there may be some hidden relationships in the data. To mine the underlying relationship, a commonly-used approach is to calculate a similarity measure of the features for different observations and construct relations if the similarity value is greater than a threshold. For instance, Li et al. (2020a) construct a stock correlation graph based on the cosine similarity of stocks' historic market price. Two stocks are then connected if the absolute value of their correlation is larger than a threshold. This type of construction method could be easily implemented and understood. However, in this type of construction, feature information is represented both in the graph adjacency matrix and the feature matrix. The overlapping information may lead to doubts that whether the graph representation is still necessary. Also, how to set the threshold is an issue, and justification may be needed for the selected similarity threshold value.

2.3 Graph Categorization by Graph Types

In this section, we categorize graphs into five categories based on their characteristics and provide examples in the financial context. Since different types of graphs may impose various challenges on modeling technology, we discuss GNN methods for each graph type in section 4 to provide solutions respectively. Figure 2 provides visualization for each graph type: homogeneous graph, directed graph, bipartite graph, multi-relation graph, and dynamic graph. It is also worth mentioning that a graph may be categorized into multiple types.

2.3.1 Homogeneous Graph

A homogeneous graph is a graph with one type of node and one type of edge. For instance, Liou et al. (2021) construct a financial news co-occurrence graph, where two companies are connected if they are tagged in the same news articles. Li et al. (2020b) build a transaction network, where nodes are accounts and are connected when there exist transactions between them. This type of graph has a relatively simple structure and the majority of the GNN methods could be applied to model this type of graph.

2.3.2 Directed Graph

A directed graph is a graph where edges have orientations. For instance, Cheng et al. (2019) construct a guarantee network where nodes are the companies and edges represent the guarantee relationship. Since the guarantor has the obligation to pay the debt for the borrower, but not the other way round, this type of guarantee relationship is one-sided and could be represented in a directed edge. In general, a directed graph may have an asymmetric adjacency matrix and thus cannot be semi-definite. Since some GNN mythologies are developed for semi-definite adjacency matrices, they may not be suitable for the directed graph.

2.3.3 Bipartite Graph

A bipartite graph is a graph with two types of nodes and edges only exist between nodes of different types. For instance, Liu et al. (2018) construct an account-device network in a fraud detection task. Nodes could be either accounts or devices, with edges connecting them if the account has activities on the device. Li et al. (2019a) extract a user-item network using the rating data, where nodes could be either a user or an item. An edge exists if a user has rated the item. A bipartite graph is commonly used, when data can be divided into two groups and the interaction between two groups matters. A bipartite graph could be seen as a specific case of a multi-relation graph discussed in the next session and GNN methods developed for the multi-relation graph are applicable on a bipartite graph as well.

2.3.4 Multi-Relation Graph

Sometimes, edges may have multiple types to represent different relations between nodes. For instance, Wang et al. (2019) construct a user-relation graph, where nodes are the users of an e-commerce platform. There are multiple edge types representing various relationships including friendship, workmates and classmates. Dou et al. (2020) build a review graph where users are represented as nodes in the graph. Three types of relations between users are defined to capture their behavioral patterns: reviewing the same product, having the same star rating, and having similar texts. With multiple edge types, this type of graph contains more information about the relationship between nodes, and thus how to capture this information is critical when developing the GNN methodology.

2.3.5 Dynamic Graph

A dynamic graph is a sequence of graphs, where each graph could have an adjacency matrix and feature matrix. For example, Cheng and Zhang L Wang X (2020b) construct a temporal guarantee network, representing the guarantee relationship in each time step. This type of graph is commonly used to represent the changes in both relations and features, as time goes. Since

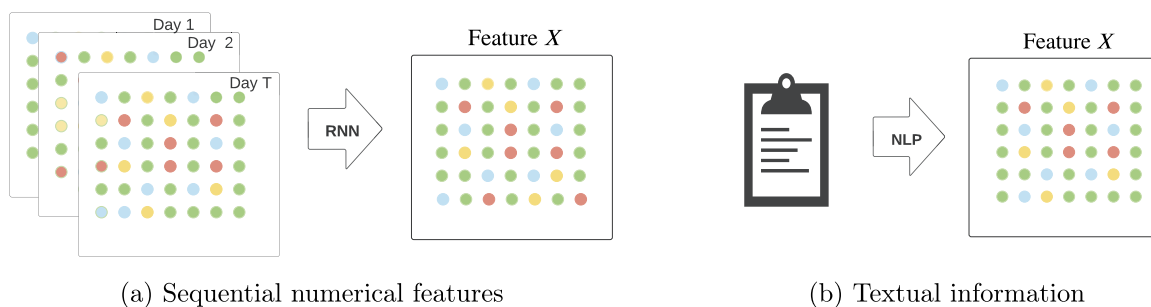


Figure 3: Feature processing for sequential features and textual information. For sequential numerical features, recurrent neural network (RNN) based approaches are commonly used to capture the temporal dependencies. For text features, it is often processed utilizing natural language processing (NLP) methods including word embedding, sentence embedding, and language models, to convert the unstructured data to structured ones.

both nodes and edges could appear and disappear, it is hard to perform some graph operations that require fixed dimensions of matrices. Thus, capturing the dynamically of the graph is challenging and requires a more sophisticated methodology.

3 Feature Processing

With diverse data sources in the financial field, node features are commonly formatted as sequential numerical features or textual information. These data formats impose challenges on the feature processing step since GNN methods could not be directly applied to these data formats. In this section, we summarize the commonly-used feature processing techniques and how they solve these challenges.

3.1 Sequential Numerical Data

Updating information as time goes, the financial industry has a rich source of time-series data. Indexed using timestamps, features could be seen as sequential which requires appropriate modeling. Consider the feature matrix at time s , $X^s \in \mathbb{R}^{n \times p \times l}$, where n is the number of nodes, p is the dimension of features at each time point, l is the length of the sequence. For example, in a stock prediction task, we have a stock relation graph, with n stocks as nodes. The feature matrix X^s represents the p features in the past l days from time s for these n stocks. To encode the numerical sequence for each node, a recurrent neural network (RNN) is frequently used due to its superior performance in predicting time-series data. The literature could be summarized into two lines of work, long short-term memory (LSTM) based approach and gated recurrent unit (GRU) based approach.

3.1.1 LSTM Based Approach

As a special form of recurrent neural network, long short-term memory(LSTM) (Hochreiter and Schmidhuber, 1997) is capable to capture the long-term dependencies and avoid the vanishing

gradient problem. Using memory cells and gate units, it has the following expression:

$$\begin{aligned}
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f), \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o), \\
\tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c), \\
c_t &= f_c \circ c_{t-1} + i_t \circ \tilde{c}_{t-1}, \\
h_t &= o_t \circ \tanh(c_t),
\end{aligned}$$

where $x_t \in \mathbb{R}^D$ is the input vector at time t and D is the number of features, $f_t, i_t, o_t, \tilde{c}_t, c_t, h_t$ denotes the forget gate, output gate, cell input, cell state and hidden state vectors respectively, $W_f, W_i, W_o, W_c, U_f, U_i, U_o, U_c$ are trainable weight matrices and b_f, b_i, b_o, b_c are trainable bias vectors, $\sigma(\cdot)$ represents the sigmoid activation function, and \circ denotes the element-wise product. The hidden state of the LSTM on day t is denoted by: $h_t = LSTM(x_t, h_{t-1})$, $s - 1 \leq t \leq s$.

Since the LSTM updates the hidden state to capture the structural information, a common approach to encode the historical data is generating sequential embedding E^s using the last hidden state of LSTM, $E^s = LSTM(X^s) \in \mathbb{R}^{n \times u}$, where u is the dimension of the output feature. Then the encoded pricing information is used as input to the graph neural network. For instance, Chen et al. (2018) used the generated sequential embedding E^s as the input feature matrix for graph convolutional network and Feng et al. (2019) utilized E^s as the input feature matrix for their proposed temporal graph convolutional layer. Using the last hidden state as an input feature, this type of method could capture the information in the past days while having an appropriate format to feed into the GNN model.

3.1.2 GRU Based Approach

Gated recurrent unit (GRU) (Cho et al., 2014) is another variant of RNN models. It also applies gating mechanism and has fewer parameters. It has the following structure:

$$\begin{aligned}
r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r), \\
z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z), \\
\tilde{h}_t &= \tanh(W_h x_t + U_h (r_t \circ h_{t-1}) + b_h), \\
h_t &= (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t,
\end{aligned}$$

where $x_t \in \mathbb{R}^D$ is the input vector at time t for stock i and D is the number of features, $r_t, z_t, \tilde{h}_t, h_t$ denotes the reset gate, update gate, candidate activation and hidden state vectors respectively, $W_r, W_z, W_h, U_r, U_z, U_h$ are trainable weight matrices and b_r, b_z, b_h are trainable bias vectors, $\sigma(\cdot)$ represents the sigmoid activation function, and \circ denotes the element-wise product. The hidden states of the GRU on day t is denoted by: $h_t = GRU(x_t, h_{t-1})$.

Utilizing GRU to encode the past numerical information, we could obtain the hidden state for each day. Since past days' impact on the current-day representation may differ, an attention mechanism is frequently used to assign weights differently. For instance, Sawhney et al. (2020a) use an additive attention mechanism to aggregate the hidden states across time. Cheng et al. (2020) utilize the concatenated attention method to incorporate different importance of time. The attention mechanism aggregates the hidden states of past days and assigns different weights across time. To get the feature representation at time s , it rewards the influential days when

aggregating the hidden states from time $s - l$ to time s , and thus take temporal dependencies into account. The obtained node representation is then used as a feature in the graph neural network model.

3.2 Textual Information

In the financial industry, a large amount of information is of textual form, including financial news, financial statements, and customer reviews. How to translate the texts into vector representation while preserving semantic information, is essential. The following section summarizes the commonly-used natural language processing (NLP) methodology to convert the unstructured texts into a vector form.

3.2.1 Word Embedding and Sentence Embedding

Word embedding methods are widely used to represent a word as a fixed-length vector. Then, to learn the sentence representation, a recurrent neural network model is often utilized to capture local semantic information. For instance, to embed the news headlines, Li et al. (2020a) encode the word as word embedding using GloVe (Pennington et al., 2014). Then LSTM method is applied with an attention mechanism to create the sentence representation. However, since a word may have multiple meanings, word embedding methods may cause problems by assigning the same vector to words with different meanings. Thus, there are also approaches to embed at a sentence level in order to alleviate the problem. For instance, Sawhney et al. (2020a) generate sentence-level embedding for each Tweet using Universal sentence encoders (Cer et al., 2018).

3.2.2 Language Model

Instead of focusing on generating vectors for words, language models focus on capturing the pattern of languages to predict the word based on its surrounding words. Taking the contexts into account, language models achieve the state of art performance on many NLP tasks and are widely used in the literature. For instance, Liou et al. (2021) use bidirectional encoder representations from transformers (BERT) model (Devlin et al., 2019) to encode the entire news article and generated a news embedding. Without modifying the model architecture, the pre-trained BERT model is able to be fine-tuned and produce state-of-art performance. For example, FinBERT (Araci, 2019), a BERT model pre-trained specific to the financial domain, is utilized to encode the text scripts in company earning calls (Sawhney et al., 2020b).

4 Graph Neural Network Models

Proposed by Gori et al. (2005), a graph neural network is a neural network model capable of processing graphs. Unlike network embedding methods whose major aim is to generate a vector to represent each node, graph neural network models are designed for a variety of tasks, including node classification, edge prediction, and graph classification. Due to its wide application and superior performance, graph neural network models have drawn great attention. In this section, we present the commonly-used GNN models for each type of graphs, since the methodology may vary with different graph characteristics. In the supplementary materials, we present a figure demonstrating the major GNN methodology used for each graph type.

4.1 Homogeneous Graph

Proposed by Kipf and Welling (2017), graph convolutional network (GCN) is a widely-used graph neural network model and could encode both local graph structure and node features. Extending the convolution concepts to graphs, graph convolution could be seen as message passing and information propagation. Aggregating neighbors' feature information, graph convolutional networks could be represented with the following layer-wise propagation rule:

$$H^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l),$$

where \tilde{A} is the adjacency matrix with added self-connections, $\tilde{D} = \text{diag}(\sum_j \tilde{A}_{ij})$ W^l is trainable weight matrix of l^{th} layer, H^l is the node hidden feature matrix in the l^{th} layer and $\sigma(\cdot)$ is the activation function. With its relatively simple model structure and great performance, the GCN model is often used as a benchmark method to compare with. Among the reviewed literature, over half of them have applied the GCN method as a benchmark method.

While GCN equally treats the neighbors of the target node, it often occurs that some neighbors may be more influential than others. Considering various impacts of the neighbor nodes, Veličković et al. (2018) propose graph attention networks (GAT) and it is able to assign different weights to nodes in the same neighborhood as follows:

$$h_i^{l+1} = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^l W^l h_j^l\right),$$

$$\alpha_{ij}^l = \frac{\exp(\text{LReLU}(a^T [W^l h_i^l \| W^l h_j^l]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LReLU}(a^T [W^l h_i^l \| W^l h_k^l]))},$$

where h^l is the hidden feature vector for node i in the l^{th} layer, W^l is trainable weight matrix, a is a learnable vector, \mathcal{N}_i is the neighborhoods of node i , α_{ij}^l represents attention coefficient of node j to i at l^{th} layer, $\sigma(\cdot)$ is the activation function, $\|$ denotes vector concatenation, and LReLU denotes the leaky ReLU activation function. It is also worth mentioning that, since GAT is able to learn the weights of the neighboring node, we could interpret the learned attention weights as a relative importance measure, to better understand the model. Similar to GCN, GAT is also often used as a benchmark method in the reviewed papers with about 40% coverage.

4.2 Directed Graph

A undirected graph has a symmetric adjacency matrix that guarantees a semi-definite Laplacian matrix, which lays the foundation for applying GCN. The directed graph, on the other hand, may have asymmetric adjacency matrix and could be handled by spatial-based GNN methods (Wu et al., 2020), such as GAT. In practice, there is not much work developing methodologies for directed graph, since it could be processed by spatial-based GNN methods or making the adjacency matrix symmetric. However, there is also a line of work developing GNN methodology to predict sequences of events represented as a directed graph.

A sequence of events could be naturally formed as a directed graph, where each node is one type of event and an edge points from one event to the following one. Given the graph, which could be seen as a partial sequence of events, how to encode the features and predict the rest of the sequence is a challenge. The aforementioned GCN and GAT models aim at representation learning and are used to produce a single output instead of outputting a sequence. To fill the gap, Li et al. (2016) propose a gated graph neural network (GGNN) that could produce sequential

outputs. It applies gated recurrent unit (GRU) as a recurrent function and is constructed as follows:

$$\begin{aligned}
a_t &= A^T h_{t-1} + b, \\
r_t &= \sigma(W_r a_t + U_r h_{t-1}), \\
z_t &= \sigma(W_z a_t + U_z h_{t-1}), \\
\tilde{h}_t &= \tanh(W_h a_t + U_h (r_t \circ h_{t-1})), \\
h_t &= (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t,
\end{aligned}$$

where h_t is the updated event representation at t^{th} step, a_t contains information transferred from both directions' edges, r_t , z_t , \tilde{h}_t denotes the reset gate, update gate, and candidate activation vectors at t^{th} respectively, W_r , W_z , W_h , U_r , U_z , U_h are trainable weight matrices, b is a trainable bias vector, $\sigma(\cdot)$ is the sigmoid function, \circ denotes the element-wise product, and \tanh denotes the hyperbolic tangent function. Incorporating the adjacency matrix A , GGNN aggregates the structural information in every propagation step. Unrolling the recurrence function to a fixed number, the GGNN ensures convergence without constraining the parameters.

4.3 Bipartite Graph

The aforementioned methods focus on homogeneous graphs whose nodes and edges are all of one type. In real-life applications, graphs could be heterogeneous. As a running example, in a spam detection task, we could have a user-item network, where nodes are either users or items. An edge e_{ij} denotes that user i has rated on item j . This type of graph is well known as a bipartite graph \mathcal{G} with the following notation: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{U, W\}$ is a set of nodes and could be divided into two non-empty and disjoint sets \mathcal{U} , \mathcal{W} . \mathcal{E} is a set of edges, where every edge joins a node in \mathcal{U} and a node in \mathcal{W} .

Based on the characteristics of a bipartite graph, the commonly-used methodologies could be categorized into a framework as follows. In each iteration, the edge information h_{uw}^l is updated aggregating its previous hidden state h_{uw}^{l-1} and hidden states of the two nodes it links (h_u^{l-1} , h_w^{l-1}), as shown in equation (1). At each iteration, a user node u aggregates information from its rated items $h_{\mathcal{N}(u)}^l$ and its past hidden state h_u^{l-1} , as in equation (3). While the representation of the rated items $h_{\mathcal{N}(u)}^l$ is updated aggregating the information from the linking edges h_{uw}^l and the item nodes h_w^{l-1} , as in equation (2). Then, the item node w is updated respectively as shown in the following equations:

$$h_{uw}^l = \sigma(W_E^l \cdot \text{AGG}_E(h_{uw}^{l-1}, h_u^{l-1}, h_w^{l-1})), \quad (1)$$

$$h_{\mathcal{N}(u)}^l = \sigma(W_{\mathcal{N}(U)}^l \cdot \text{AGG}_U(\text{AGG}_{UW}(h_w^{l-1}, h_{uw}^l))), \quad \forall w \in \{\mathcal{N}(u)\}, \quad (2)$$

$$h_u^l = \text{concat}(W_U^l \cdot h_u^{l-1}, h_{\mathcal{N}(u)}^l), \quad (3)$$

$$h_{\mathcal{N}(w)}^l = \sigma(W_{\mathcal{N}(W)}^l \cdot \text{AGG}_W(\text{AGG}_{WU}(h_u^{l-1}, h_{uw}^l))), \quad \forall u \in \{\mathcal{N}(w)\}, \quad (4)$$

$$h_w^l = \text{concat}(W_W^{l-1} \cdot h_w^{l-1}, h_{\mathcal{N}(w)}^l), \quad (5)$$

where e_{uw} represents the edge representation for edge linking node u and w , h_{uw}^l , h_u^l , h_w^l , $h_{\mathcal{N}(u)}^l$, $h_{\mathcal{N}(w)}^l$ are hidden states at l^{th} layer, W_E^l , W_U^l , $W_{\mathcal{N}(U)}^l$, $W_{\mathcal{N}(W)}^l$, W_W^l are trainable weight matrices at l^{th} layer, $\text{AGG}_E(\cdot)$, $\text{AGG}_U(\cdot)$, $\text{AGG}_{UW}(\cdot)$, $\text{AGG}_W(\cdot)$, $\text{AGG}_{WU}(\cdot)$ are user-chosen aggregation functions, $\sigma(\cdot)$ is the activation function, and concat denotes the vector concatenation.

There exist many modeling methodologies for bipartite graphs that could fit into the above framework. For instance, Zhang et al. (2020) propose a similar model structure as the framework and apply the attention mechanism as the aggregation function, since different items may have different impacts when learning users' representations. Instead of using all neighbors, Li et al. (2019a) utilize a sampling technique when aggregating neighbors' information in each iteration. There are also literature using the above framework as a building block and combining clustering methodology to learn a hierarchical representation of the graph, since hierarchical representation with various GNN models could achieve satisfactory performance. For instance, Li et al. (2019b) utilize the node embedding generated from the framework to cluster users into different communities and make a recommendation based on both community information and user information. Specifically, the user information is decomposed into two orthogonal spaces representing community-level information and individualized user preferences. Li et al. (2020c) treat the framework as a GNN module and stack it in a hierarchical fashion. With the embedding generated from the framework, clustering algorithms are performed to generate a coarsened graph which is used as an input for the next GNN layer.

4.4 Multi-Relation Graph

Instead of a simple homogeneous graph where all nodes and edges have the same type, in real life, there may exist multiple relations between nodes. For example, in a malicious account detection task, we could construct an Amazon review network. Users are the nodes in the graph and there are three relations between users: reviewing the same product, having the same star rating, and having similar texts. The multi-relation graph is denoted as \mathcal{G} : $\mathcal{G} = (\mathcal{V}, \mathcal{E}_{1:R})$, where \mathcal{V} is the set of nodes, $\mathcal{E}_{1:R}$ is the set of edges, R is the number of node types, $e_{i,j}^r \in \mathcal{E}_r$ is an edge between node i, j with a relation $r \in \{1, \dots, R\}$.

A frequently used approach is to transform the heterogeneous graph into multiple homogeneous graphs by extracting R subgraphs $\{\mathcal{G}^r = (\mathcal{V}, \mathcal{E}_r), r = 1, \dots, R\}$ from it. Each subgraph \mathcal{G}^r only preserves one type of edge and thus is homogeneous. This line of work could be unified in a two-step framework. The first step implements the sub-graph aggregation to aggregate neighbor information in each sub-graph as shown in equation (6). The second step is to conduct the inter-relation aggregation to aggregate relation-specific embeddings as equation (7),

$$h_{i,r}^l = f(\text{AGG}_r\{h_{j,r}^{l-1}\}), \quad \forall j \text{ s.t. } (i, j) \in \mathcal{E}_r, \quad (6)$$

$$h_i^l = g(\text{AGG}\{h_{i,(1:R)}^l, h_i^{l-1}\}), \quad (7)$$

where $h_{i,r}^l$ is the subgraph-specific embedding of node i in subgraph r in l^{th} layer, h_i^l is the general embedding of node i in l^{th} layer, $\text{AGG}_r(\cdot)$ is the aggregation function in subgraph r , $\text{AGG}(\cdot)$ is the inter-relation aggregation function, and $f(\cdot), g(\cdot)$ are user-defined functions.

There are multiple methods that could be categorized into the above two-step framework. For instance, in a fraud classification task, Liu et al. (2018) observe that fraudsters tend to congregate in topology and thus use weighted sum for within-relation aggregation to capture this congregation pattern. They then apply an attention mechanism for inter-relation aggregation to learn the significance for each sub-graph as follows:

$$h_{i,r}^l = \sigma(\text{Weighted mean}\{h_{j,r}^{l-1}\}), \quad \forall j \text{ s.t. } (i, j) \in \mathcal{E}_r,$$

$$h_i^l = \sigma(X_i W + \text{Attention}\{h_{i,(1:R)}^l\}),$$

where X_i is the feature vector for node i , W is a trainable matrix, $\sigma(\cdot)$ is the activation function, and Attention denotes the attention aggregator.

To incorporate neighbors' information, Dou et al. (2020) use the mean aggregator for within-relation aggregation. To reduce the computational cost and keep the relational importance information, they apply a pre-calculated parameter p_r^l as the weight in the intra-relation aggregation step. They also test several aggregating functions when aggregating relation-specific embeddings with the following structure:

$$\begin{aligned} h_{i,r}^l &= \sigma(\text{Mean}\{h_{j,r}^{l-1}\}), & \forall j \text{ s.t. } (i, j) \in \mathcal{E}_r, \\ h_i^l &= \sigma(h_i^{l-1} + \text{AGG}\{h_{i,r}^l \cdot p_r^l\}), & \forall r \in (1, \dots, R), \end{aligned}$$

where p_r^l is a pre-trained weight.

Since different relations provide various facets of user characteristics, relationship-specific embedding may have different statistical properties, which may cause trouble when aggregating them in a lower-level space. To deal with that, Wang et al. (2019) project the relation-specific node embedding to higher spaces using multi-layer perception (MLP) and concatenate them with relation-level attention:

$$\begin{aligned} h_{i,r}^l &= \text{MLP}\{h_{i,r}^{l-1}\}, \text{ where } h_{i,r}^{l-1} = \text{Attention}\{x_{j,r} : \forall j \text{ s.t. } (i, j) \in \mathcal{E}_r\}, \\ h_i^l &= \text{Concatenation with attention}\{h_{i,(1:R)}^l\}, \end{aligned}$$

where Attention denotes the attention aggregator.

4.5 Dynamic Graph

The previously mentioned neural network models generally focus on a static graph. However, in real-life settings, a graph may be dynamically evolving since relations may be updated with time. For example, in order to predict the loan default risk, a guarantee network needs to be updated, adding newly-constructed guarantee relationships and removing companies that have fully paid the loan. With the rapid development of graph neural network methodologies on static graphs, there emerges a trend to extend GNN models to a dynamic setting. Consider a sequence of T graphs $\mathcal{G}^{seq} = \{\mathcal{G}_1, \dots, \mathcal{G}_T\}$, where $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$ represent the graph at i th time point. The feature matrices are represented as $X = \{\mathbf{X}_1, \dots, \mathbf{X}_T\}$ and the adjacency matrices are $A = \{\mathbf{A}_1, \dots, \mathbf{A}_T\}$.

To capture the sequential pattern in the dynamic graph, a common approach is to train a GNN to generate the node embedding at each time stamp and then utilize a recurrent neural network to aggregate the information. For example, Cheng and Zhang L Wang X (2020b) obtain the node embeddings at each time step by training a GCN with multi-head attention. Then, they utilize the GRU to capture the sequential pattern with a temporal attention layer to capture the temporal variation over timestamps. Similarly, Yang et al. (2020) first aggregate node and edge information in each snapshot and then employ a LSTM operator to capture the temporal variations in the node embeddings.

In the aforementioned methods, a graph neural network is learned for feature aggregation and an RNN model is trained to capture the sequential pattern of the node embeddings. However, in reality, a node may appear and disappear, which may worsen the performance of the RNN model when updating the node representation. In a guarantee network, for example, a company that has borrowed loans could disappear from the graph after it pays all the debts and could

appear again backing up other companies' loans. To overcome the limitation, Pareja et al. (2020) proposed EvolveGCN which utilizes a recurrent neural network to evolve the GCN parameters instead of updating the node embeddings. For each time point t , a GCN model is constructed as follows to fit the graph \mathcal{G}_t :

$$H_t^{l+1} = \sigma(\tilde{D}_t^{-\frac{1}{2}} \tilde{A}_t \tilde{D}_t^{-\frac{1}{2}} H_t^l W_t^l),$$

where \tilde{A}_t is the adjacency matrix with added self-connections, $\tilde{D}_t = \text{diag}(\sum_j \tilde{A}_{ij})$ is the degree matrix, W_t^l is trainable weight matrix of l^{th} layer, H_t^l is the matrix of activation in the l^{th} layer, and $\sigma(\cdot)$ is the activation function.

To update the weight matrix W_t^l , Pareja et al. (2020) propose two methods. The first method considers W_t^l as a hidden state of the dynamics and update it using a GRU model, as shown in equation (8). The second method treats W_t^l as an output state which is updated using a LSTM method, as shown in equation (9). The structure for both methods is as follows:

$$W_t^l = \text{GRU}(H_t^l, W_{t-1}^l), \quad (8)$$

$$W_t^l = \text{LSTM}(W_{t-1}^l). \quad (9)$$

Compared to the second method, the first method incorporates the updated node embedding in the recurrent neural network and it may lead to better performance when node features are informative.

5 Application

In this section, we have detailed some financial applications that the GNN methods have been commonly applied on. We have also summarized features, graphs, methods, evaluation metrics, and baselines used in each financial application in the supplementary materials.

5.1 Stock Movement Prediction

Though there are still debates on whether stocks are predictable, stock prediction receives great attention and there are rich literature on predicting stock movements using machine learning methods. However, the task of stock prediction is challenging due to the volatile and non-linear nature of the stock market. Traditionally, there are two major approaches to handle the task: technical analysis and fundamental analysis (Sawhney et al., 2020a). Technical analysis utilizes numerical features such as closing prices and trading volumes, while the fundamental analysis approach includes non-numerical information, such as news and earning calls. The limitation of these non-graph approaches is that they often have a hidden assumption that the stocks are independent. To take the dependence into account, there is an increasing trend to represent the stock relations in a graph where each stock is represented as a node and an edge would exist if there are relations between two stocks. Predicting multiple stocks' movements could then be formed as a node classification task and GNN models could be utilized to make the prediction.

However, there also exist challenges to apply the GNN methods in the stock prediction task. Unlike other fields where the benchmark graphs are available, to the best of our knowledge, there is no off-the-shelf graph representing inter-stock relations. With abundant relations existing in the financial system, it becomes challenging to obtain and select the relation for graph construction. Moreover, owing to the volatility of the stock market, how to model the sequential features

and capture the temporal patterns are also critical. Also, the financial industry has rich data sources including financial statements, news and pricing information, which impose difficulty on modeling the data.

There are multiple ways to construct the stock relational graph. For instance, believing that correlation on historical prices reflects the inter-stock relation, Li et al. (2020a) construct the graph using the correlation matrix of historic data to predict the movement of Tokyo stock price index. On the other hand, Matsunaga et al. (2019) borrow information from knowledge bases and construct supplier, customer, partner, and shareholder relational graphs. With multiple ways of graph construction, there doesn't exist a "best" graph due to the lack of graph evaluation methods. Future work could be done to design a graph evaluation method to help researchers better construct a relational graph.

To effectively process the sequential data and incorporate related corporations' information Chen et al. (2018) propose a joint model using LSTM and GCN to predict the stock movement. However, Chen et al. (2018)'s approach assumes that the relations between stocks are static, which may not reflect the reality. Instead, Feng et al. (2019) propose a temporal graph convolution layer to capture the stock relations in a time-sensitive manner, so that the strength of relation could be evolving over time. The relations are then updated based on historical pricing sequences and the proposed method obtained better performance compared to GCN. Believing that stock description documents also contain information reflecting the changes in companies' effect, Ying et al. (2020) capture the temporal relation by both sequential features and stock document attributes with a time-aware relational attention network.

The aforementioned methods focus on capturing the temporal dependencies, while Sawhney et al. (2020a) focus on fusing data from different sources. Sawhney et al. (2020a) propose a multipronged attention network to jointly learn from historical price, social media, and inter stock relations. Encoded pricing and textual information are used as node feature inputs to GAT, where the graph information comes from the Wiki company-based relations. The attention mechanisms is applied to allocate different weights on various data sources and latent correlations may learned via the attention layers.

5.2 Loan Default Risk Prediction

For commercial banks and financial regularity institutions, monitoring and assessing the default risk is at the heart of risk controlling process. As one of the credit risks, default risk is the probability that the borrower fails to pay the interest and principal on time. With a binary outcome, loan default prediction could be seen as a classification problem and is commonly addressed utilizing user-related features with classifiers including neural network (Turiel and Aste, 2020) and gradient boosted trees (Ma et al., 2018). Since the probability that a borrower defaults may be influenced by other related individuals, there is plenty of literature forming a graph to reflect the interactions between borrowers. With the rapid growth of GNN methods, GNN methods are widely applied on the graph structure for loan default predicting problems. There are currently three lines of work focusing on various types of loans: guarantee loans, e-commerce loans, and other loans.

The guarantee loan allows small entrepreneurs to back each other in order to increase their credibility. It has a debt obligation contract that specifies that if one corporation fails to pay the debt, its guarantor needs to pay for it. A guarantee network naturally arises where each node is a company and directed edges represent the guarantee relationship. To learn a better representation of the network, Cheng et al. (2019) utilize the graph attention layer and design

a objective function, so that vertices with similar structures will be closer in the learned feature space. Since the guarantee relations changes with time, Cheng and Zhang L Wang X (2020b) forms a dynamic guarantee network to represent the dynamics. A recurrent graph neural network layer is developed to learn the temporal pattern and attentional weights are learned for each time point via an attention architecture.

Unlike guarantee loans that the loan information could be naturally represented in a directed graph, other loan types may not have a clear graph structure and researchers need to construct the graph based on interactive information. For instance, Xu et al. (2021) construct a user relation graph where users are connected by various relationships, such as social connections, transactions, and device usage. However, the interactive graph may also contain noisy data, which may be irrelevant. Since the massive interactive information may be noisy and the impacting supply-chain information is deficient, Yang et al. (2020) extract supply-chain relations while predicting loan defaults. Forming the interaction data as a graph, Yang et al. (2020) formulate the supply chain mining task as a link prediction task and thus construct a supply chain network, which is then used to predict the default probability with GNN methodology.

With the rise of e-commerce, e-commerce consumer lending service is gaining popularity to enhance consumers' purchasing power. Able to obtain information from multiple facets, the e-commerce platform could have multi-view data and multi-relation networks, which may require sophisticated modeling methodology. For instance, in order to predict the default probability for each consumer with multi-view data, Liang et al. (2021) utilize a hierarchical attention mechanism to encode the features on each view. Exploring multiplex relations, Hu et al. (2019) propose an attributed multiplex graph-based model with relation-specific layer and attention mechanism to jointly model multiple relations. To simultaneously model the labeled and unlabeled data, Wang et al. (2019) proposed a semi-supervised graph neural network approach and obtain interpretable results.

5.3 Recommender System of e-Commerce

With the rapid growth of e-commerce, customers gradually get used to shopping online and are exposed to a numerous range of products. To alleviate the burden of users choosing the appropriate item, the recommender system is developed to suggest products to users based on predicted item ratings. Presenting the user and item information in a graph, GNN methodologies are widely used in recommender system related tasks including click rate prediction and fake review detection.

To accurately predict users' preferences and recommend the appropriate item, it is vital to exploit information for users, items, and their interactions. A widely-used representation of that information is a bipartite graph, where nodes are of two types, user and item, and edges represent that there are relationships between user and item nodes. Noticing that the community the user belongs to may affect the shopping decision, for example, a user belongs to the traveler group may purchase travel-related items, Li et al. (2019b) combine the bipartite graph modeling algorithm with clustering techniques to reflect the community impact and the individual preference. However, Li et al. (2019b)'s approach only considers the hierarchy in the user side, while items may also have hierarchical information. To fill the gap, Li et al. (2020c) stack several GNN modules hierarchically to capture the hierarchical structure in both user and item perspectives. Embeddings learned from the GNN layer are clustered and used as an input for the next GNN layer, which could preserve the high-order hierarchical connections.

The recommender system is mainly based on the past history of the user, including its rating and reviews on the item. However, fake ratings and feedback may be posted by the fraudsters to seek financial benefits. To detect fraudulent reviews on the e-commerce platform, Kudo et al. (2020) construct a directed and signed comment graph with a signed graph convolutional network approach. Compared to Kudo et al. (2020)'s approach which only considers the comment graph, Li et al. (2019a) integrate both the bipartite user-item graph and a comment graph to capture the local and global context of the comments. Noticing that camouflage behaviors of the fraudsters may deteriorate the performance of fraud detection mechanism and has seldom been considered by prior works, Dou et al. (2020) propose a model against both feature and relation camouflage. For each node, only informative neighbors are selected for the next aggregation step, utilizing a similarity measure and a reinforcement learning mechanism. While the above literature focus on the fraud review detection side, there are also works that accomplish both fraud review detection and item recommendation tasks. For example, Zhang et al. (2020) propose a GCN-based framework that performs both item recommendation and fraud detection in an end-to-end manner, while each of the tasks is beneficial for the other one.

5.4 Fraud Detection

Including payment fraud, identity theft, financial scam and insurance fraud, financial fraud has a variety of types and has an increasing trend (Kurshan and Shen, 2020). Observing that fraudsters tend to have abnormal connectivity with other users, there is a trend to present users' relations in a graph and thus, the fraud detection task could be formulated as a node classification task.

Aiming to detect the malicious accounts, who may attack the online services to seek excessive profits, Liu et al. (2018) find out that fraudsters have two patterns: device aggregation and activity aggregation. Due to economic constraints, attackers tend to use limited number of devices and perform activities in a limited time, which may be reflected in the local graph structure. With this observation, Liu et al. (2018) propose a variant of GCN and use sum operators to capture the aggregation pattern. While the malicious accounts may be aggregated together, Liu et al. (2019) argue that the normal accounts could also be connected with the malicious account and may be mislabeled as malicious, which adds noisy signals to the graph. Taking into account that the graph could be noisy and nodes may have different impact, Liu et al. (2019) propose an adaptive path layer to adaptively select the important neighbor nodes that contribute most to the target node. Applying the GNN methodologies proposed by Liu et al. (2018), Liang et al. (2019) stack multiple adaptive path layers to aggregate neighbors' features and have great performance in a insurance fraud detection task.

The above literature focus on a bipartite graph, where nodes are either account or devices, and there are also literature utilizing other types of graphs. For instance, Jiang et al. (2019) construct a homogeneous user network connecting user nodes based on their similarity of behaviors and apply the GCN model. Rao et al. (2020) focus on a dynamic graph of registration records and implement GCN layers on structural and temporal subgraphs.

Besides the literature developing GCN-based methodologies, there are also fraud-detection related literature applying GNN models with other structures. Since GCN may suffer from over-smoothing problem and have shallow model structure, Lv et al. (2019) propose to replace the graph convolutional matrix with auto-encoder to increase the depth of the neural network. Zhao et al. (2021) argues that GNN models with random-walk based losses have poor performance in anomaly detection task, since nodes of the same label may not be closer. They then propose a loss function which leads to better performance and has bounded prediction error.

5.5 Event Prediction

Financial events, including revenue growth, acquisition and bankruptcy, could provide valuable information on market trends and could be used to predict future stock movement. Therefore, it draws great attention on how to predict next financial event based on past events and currently GGNN model is often used to accomplish the task. For example, given a sequence of financial events of Chinese listed companies, Yang et al. (2019) aim to predict the next event type and construct an event graph where each node is a financial event and edges are weighted using the frequency of the event pairs. Harl et al. (2020) transform a binary classification problem into an event prediction task by dividing the loan application process into several events and predicting whether the next event would be accepting or rejecting the application. Utilizing the GGNN model, they obtain the predictions with high accuracy.

6 Challenges

6.1 Graph Evaluation Methods

To justify the inclusion of a graph, a commonly-used evaluation method is to compare the outcome for a graph-based machine learning method with a graph-free machine learning method (Feng et al., 2019; Li et al., 2020a). However, there is little discussion on comparing different graphs' effects and quality, while existing literature discussing graph comparisons are often inadequate. For example, Liang et al. (2019) visualize the structural patterns in different graphs, concluding that the device sharing graph is more appropriate based on the observed patterns. Without presenting the evaluation metric for each graph, the graph comparison based on the visualization may not be adequate. The problem may be more severe in similarity-based graph construction since a threshold needs to be set to determine whether an edge exists. Different threshold values may lead to completely different graphs and thus affect the model performance. Thus, justification on threshold setup during graph construction is of great importance. Some efforts have been made to tackle this problem, such as utilizing a reinforcement learning approach to automatically select the optimal threshold (Dou et al., 2020). More attention may need to be drawn to develop a framework to assess the graph quality systematically.

6.2 Explainability

Combining both graph structural information and feature information, GNN models are often complicated and it is challenging to make an interpretation. Recently, there are some literature focusing on the explainability of GNN models. GNNExplainer (Ying et al., 2019), for example, is proposed to provide an interpretable explanation on trained GNN models such as GCN and GAT. Model explainability in financial tasks is of great importance, since understanding the model could benefit decision-making and reduce economic losses. However, there is little literature studying the explainability of GNN models in a financial application, which often accompanies by heterogeneous and dynamic graphs. Current literature focuses on relatively simpler graphs. For example, Li et al. (2019c) extend the GNNExplainer to a weighted directed graph and apply it on a Bitcoin transaction graph. Rao et al. (2020) propose an explainable fraud prediction system that could operate on heterogeneous graphs consisting of different node and edge types. More work could be done on the explainability of GNN models with edge-attributed graphs and dynamic graphs, which are not yet considered.

6.3 Task Type

Tasks of GNN models are commonly classified into three categories: node-level task, edge-level task, and graph-level task (Wu et al., 2020). However, the GNN methods applied on financial applications are mostly focused on the node-level task. For example, the stock movement prediction task is often formulated as a node classification task, where stocks are represented as nodes. There exist some literature focus on other types of task, for instance, Yang et al. (2020) aim to mine the underlying supply chain network and formulate it as a link prediction task, but this line of work is rare. Since there are rich literature developing GNN methodologies on edge prediction or graph classification tasks, there are rich opportunities on applying recently-developed GNN methods on financial fields if financial tasks could be formulated into an edge or graph level task.

6.4 Data Availability

To pursue reproducibility, it is a common practice to release both data sources and codes when publishing the paper. However, in the reviewed papers, only about 24% of them release the code. Since the financial tasks are commonly based on real-world problems, data may have some restrictions due to the privacy obligation of the related corporations. Lack of open-source codes and datasets, it is hard to reproduce previous works and compare their methodologies in the latter literature. Thus, it is of great value to construct benchmark datasets, so that methods could be compared on the same data.

6.5 Scalability

In the real-world financial scenario, commercial data are often of large scales. For instance, Yang (2019) utilize data from a popular e-commerce platform and it contains about 483 million nodes with 231 million edges. How to improve the scalability of GNNs is vital but challenging. Computing the Laplacian matrix becomes hard with millions of nodes and for a graph of irregular Euclidean space, optimizing the algorithm is also difficult. Sampling techniques may partially solve the problem with the cost of losing structural information. Thus, how to maintain the graph structure and improve the efficiency of GNN algorithms are worth further exploration.

Supplementary Material

In the supplementary materials, we present materials that are not covered in the main text. The supplementary materials contain the summary table for each financial application, figures categorizing major GNN methodologies for each graph type and acronyms used in the text.

References

- Araci D (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint: <https://arxiv.org/abs/1908.10063>*.
- Cer D, Yang Y, Kong S, Hua N, Limtiaco N, John RS et al. (2018). Universal sentence encoder for english. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 169–174.

- Chen Y, Wei Z, Huang X (2018). Incorporating corporation relationship via graph convolutional neural networks for stock price prediction. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 1655–1658.
- Cheng D, Niu Z, Zhang Y (2020). Contagious chain risk rating for networked-guarantee loans. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2715–2723.
- Cheng D, Tu Y, Ma Z, Niu Z, Zhang L (2019). Risk assessment for networked-guarantee loans using high-order graph attention representation. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 5822–5828. International Joint Conferences on Artificial Intelligence Organization.
- Cheng D, Wang X, Zhang Y, Zhang L (2020). Risk guarantee prediction in networked-loans. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20* (C Bessiere, ed.), 4483–4489. International Joint Conferences on Artificial Intelligence Organization. Special Track on AI in FinTech.
- Cho K, van Merriënboer B, Gulcehre C, Bougares F, Schwenk H, Bengio Y (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. In: *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*.
- Devlin J, Chang MW, Lee K, Toutanova K (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota.
- Dou Y, Liu Z, Sun L, Deng Y, Peng H, Yu PS (2020). Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 315–324.
- Feng F, He X, Wang X, Luo C, Liu Y, Chua TS (2019). Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)*, 37(2): 1–30.
- Gori M, Monfardini G, Scarselli F (2005). A new model for learning in graph domains. In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005*, volume 2, 729–734. IEEE.
- Harl M, Weinzierl S, Stierle M, Matzner M (2020). Explainable predictive business process monitoring using gated graph neural networks. *Journal of Decision Systems*, 1–16.
- Hochreiter S, Schmidhuber J (1997). Long short-term memory. *Neural computation*, 9(8): 1735–1780.
- Hu B, Zhang Z, Shi C, Zhou J, Li X, Qi Y (2019). Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 946–953.
- Hu B, Zhang Z, Zhou J, Fang J, Jia Q, Fang Y, et al. (2020). Loan default analysis with multiplex graph learning. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2525–2532.
- Huang J, Chai J, Cho S (2020). Deep learning in finance and banking: A literature review and classification. *Frontiers of Business Research in China*, 14: 1–24.
- Jiang J, Chen J, Gu T, Choo KKR, Liu C, Yu M, et al. (2019). Anomaly detection with graph convolutional networks for insider threat and fraud detection. In: *MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM)*, 109–114. IEEE.
- Jiang W (2021). Applications of deep learning in stock market prediction: Recent progress.

- Expert Systems with Applications*, 115537.
- Kipf TN, Welling M (2017). Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations (ICLR)*.
- Kudo W, Nishiguchi M, Toriumi F (2020). Gcnnext: Graph convolutional network with expanded balance theory for fraudulent user detection. *Social Network Analysis and Mining*, 10(1): 1–12.
- Kurshan E, Shen H (2020). Graph computing for financial crime and fraud detection: Trends, challenges and outlook. *International Journal of Semantic Computing*, 14(04): 565–589.
- Lee JB, Rossi RA, Kim S, Ahmed NK, Koh E (2019). Attention models in graphs: A survey. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 13(6): 1–25.
- Lee JW, Lee WK, Sohn SY (2021). Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. *Expert Systems with Applications*, 168: 114411.
- Li A, Qin Z, Liu R, Yang Y, Li D (2019a). Spam review detection with graph convolutional networks. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2703–2711.
- Li C, Jia K, Shen D, Shi CJR, Yang H (2019b). Hierarchical representation learning for bipartite graphs. In: *IJCAI*, 2873–2879.
- Li W, Bao R, Harimoto K, Chen D, Xu J, Su Q (2020a). Modeling the stock relation with graph network for overnight stock movement prediction. In: *no. CONF*, 4541–4547.
- Li X, Liu S, Li Z, Han X, Shi C, Hooi B, et al. (2020b). Flowscope: Spotting money laundering based on graphs. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 4731–4738.
- Li X, Saúde J, Reddy P, Veloso M (2019c). Classifying and understanding financial data using graph neural network. In: *AAAI Workshop on Knowledge Discovery from Unstructured Data in Financial Services 2020*.
- Li Y, Zemel R, Brockschmidt M, Tarlow D (2016). Gated graph sequence neural networks. In: *Proceedings of ICLR'16*.
- Li Z, Shen X, Jiao Y, Pan X, Zou P, Meng X, et al. (2020c). Hierarchical bipartite graph neural networks: Towards large-scale e-commerce applications. In: *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, 1677–1688. IEEE.
- Liang C, Liu Z, Liu B, Zhou J, Li X, Yang S, et al. (2019). Uncovering insurance fraud conspiracy with network learning. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1181–1184.
- Liang T, Zeng G, Zhong Q, Chi J, Feng J, Ao X, et al. (2021). Credit risk and limits forecasting in e-commerce consumer lending service via multi-view-aware mixture-of-experts nets. In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 229–237.
- Liou YT, Chen CC, Tang TH, Huang HH, Chen HH (2021). Finsense: An assistant system for financial journalists and investors. In: *Proceedings of the 14th International Conference on Web Search and Data Mining*.
- Liu Z, Chen C, Li L, Zhou J, Li X, Song L, et al. (2019). Geniepath: Graph neural networks with adaptive receptive paths. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 4424–4431.
- Liu Z, Chen C, Yang X, Zhou J, Li X, Song L (2018). Heterogeneous graph neural networks for malicious account detection. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2077–2085.

- Lv L, Cheng J, Peng N, Fan M, Zhao D, Zhang J (2019). Auto-encoder based graph convolutional networks for online financial anti-fraud. In: *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER)*, 1–6. IEEE.
- Ma X, Sha J, Wang D, Yu Y, Yang Q, Niu X (2018). Study on a prediction of p2p network loan default based on the machine learning lightgbm and xgboost algorithms according to different high dimensional data cleaning. *Electronic Commerce Research and Applications*, 31: 24–39.
- Matsunaga D, Suzumura T, Takahashi T (2019). Exploring graph neural networks for stock market predictions with rolling window analysis. *arXiv preprint: <https://arxiv.org/abs/1909.10660>*.
- Ozbayoglu AM, Gudelek MU, Sezer OB (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, 106384.
- Pareja A, Domeniconi G, Chen J, Ma T, Suzumura T, Kanezashi H, et al. (2020). Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 5363–5370.
- Pennington J, Socher R, Manning CD (2014). Glove: Global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- Rao SX, Zhang S, Han Z, Zhang Z, Min W, Cheng M, Shan Y, Zhao Y, Zhang C (2020). Suspicious massive registration detection via dynamic heterogeneous graph neural networks. *arXiv preprint: <https://arxiv.org/abs/2012.10831>*.
- Rao SX, Zhang S, Han Z, Zhang Z, Min W, Chen Z, et al. (2020). xfraud: Explainable fraud transaction detection on heterogeneous graphs. *arXiv preprint: <https://arxiv.org/abs/2011.12193>*.
- Sawhney R, Agarwal S, Wadhwa A, Shah R (2020a). Deep attentive learning for stock movement prediction from social media text and company correlations. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8415–8426.
- Sawhney R, Khanna P, Aggarwal A, Jain T, Mathur P, Shah R (2020b). Voltage: Volatility forecasting via text-audio fusion with graph convolution networks for earnings calls. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8001–8013.
- Turiel J, Aste T (2020). Peer-to-peer loan acceptance and default prediction with artificial intelligence. *Royal Society Open Science*, 7(6): 191649.
- Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y (2018). Graph attention networks. In: *International Conference on Learning Representations*.
- Vrandečić D, Krötzsch M (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 57(10): 78–85.
- Wang D, Lin J, Cui P, Jia Q, Wang Z, Fang Y, et al. (2019). A semi-supervised graph attentive network for financial fraud detection. In: *2019 IEEE International Conference on Data Mining (ICDM)*, 598–607. IEEE Computer Society.
- Weber M, Domeniconi G, Chen J, Weideler DKI, Bellei C, Robinson T, et al. (2019). Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint: <https://arxiv.org/abs/1908.02591>*.
- Wu Z, Pan S, Chen F, Long G, Zhang C, Philip SY (2020). A comprehensive survey on graph neural networks. In: *IEEE Transactions on Neural Networks and Learning Systems*, 4–24. IEEE.
- Xu B, Shen H, Sun B, An R, Cao Q, Cheng X (2021). Towards consumer loan fraud detection:

- Graph neural networks with role-constrained conditional random field. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 4537–4545.
- Xu K, Hu W, Leskovec J, Jegelka S (2018). How powerful are graph neural networks? *arXiv preprint: <https://arxiv.org/abs/1810.00826>*.
- Yang H (2019). Aligraph: A comprehensive graph neural network platform. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 3165–3166.
- Yang S, Zhang Z, Zhou J, Wang Y, Sun W, Zhong X, et al. (2020). Financial risk analysis for smes with graph-based supply chain mining. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 4661–4667.
- Yang Y, Wei Z, Chen Q, Wu L (2019). Using external knowledge for financial event prediction based on graph neural networks. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2161–2164.
- Ying R, Bourgeois D, You J, Zitnik M, Leskovec J (2019). GNN explainer: A tool for post-hoc explanation of graph neural networks. *CoRR*, arXiv preprint: <https://arxiv.org/abs/1903.03894>.
- Ying X, Xu C, Gao J, Wang J, Li Z (2020). Time-aware graph relational attention network for stock recommendation. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2281–2284.
- Zhang M, Chen Y (2018). Link prediction based on graph neural networks. *Advances in Neural Information Processing Systems*, 31.
- Zhang S, Tong H, Xu J, Maciejewski R (2019). Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1): 1–23.
- Zhang S, Yin H, Chen T, Hung QVN, Huang Z, Cui L (2020). Gcn-based user representation learning for unifying robust recommendation and fraudster detection. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 689–698.
- Zhao T, Deng C, Yu K, Jiang T, Wang D, Jiang M (2021). Gnn-based graph anomaly detection with graph anomaly loss. In: *The Second International Workshop on Deep Learning on Graphs: Methods and Applications (DLG-KDD'20)*. Available at: <https://deep-learning-graphs.bitbucket.io/dlg-kdd20/>.
- Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1: 57–81.
- Zhu YN, Luo X, Li YF, Bu B, Zhou K, Zhang W, et al. (2020). Heterogeneous mini-graph neural network and its application to fraud invitation detection. In: *2020 IEEE International Conference on Data Mining (ICDM)*, 891–899. IEEE.