Information Sciences 618 (2022) 317–335

Contents lists available at ScienceDirect

Information Sciences

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

FinHGNN: A conditional heterogeneous graph learning to address relational attributes for stock predictions

Jinghua Tan [a], Qing Li [b,*], Jun Wang [c], Junxiao Chen [a]

[a] School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China
[b] Research Institute for Digital Economy and Interdisciplinary Sciences, Southwestern University of Finance and Economics, Chengdu 611130, China
[c] School of Management Science and Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China

ARTICLE INFO

Article history:
Received 6 December 2021
Received in revised form 2 November 2022
Accepted 5 November 2022
Available online 11 November 2022

Keywords:
Heterogeneous graph learning
Conditional messaging
Multiple spillovers
Stock prediction

ABSTRACT

Recent financial studies have shown that spillover effects of some market factors play a significant role in stock fluctuations. Previous studies, however, were incapable of capturing the spillovers of these relational market factors because they relied on a homogeneous graph that condenses these factors into firm node attributes and requires their spillover effects to follow firm relationship instead of themselves. This fact brings up a heterogeneous graph learning problem that requires multiple node types to transport different spillover effects. This study proposes a novel conditional heterogeneous graph neural network (FinHGNN) to capture multiple spillover effects in asset pricing with two uniquely designed mechanisms. First, it presents an efficient way to preserve the connectivity of relational attributes in graph learning, which is achieved by converting relational attributes into node variables to form a heterogeneous graph. Second, a conditional message-passing mechanism is proposed to handle multiple spillover effects simultaneously by messaging conditioned on different types of nodes and node attributes. This study paves the way for addressing the relational attributes in graph learning. Experiments on two real-world datasets demonstrate the advantages of the proposed framework over three classic and four state-of-the-art algorithms, including LSTM, GCN, HGNN, eLSTM, TGC, FinGAT, and AD-GAT.

1. Introduction

The stock market is a complex dynamic system in which various market factors mingle together to cause stock fluctuations. Previous studies in both finance and computer science have been devoted to capturing a stock movement via its historical fundamentals, including prices [1], volume [2], and PE/PB ratios [3], while ignoring the influence of relevant stocks (known as momentum spillover effects in finance) [4,5].

Later on, some computer scientists took a further step by utilizing graph neural networks (GNNs) to capture momentum spillovers in the stock market [6–9]. In these pilot studies, to capture momentum spillovers, the entire stock market is considered as a graph in which one node represents a stock (firm), and the edge between two nodes is defined by some predefined or estimated firm relations for converging and aggregating the momentum.

* Corresponding author.
E-mail addresses: jinghuatan.swufe@gmail.com (J. Tan), liq_t@swufe.edu.cn (Q. Li), wangjun1987@swufe.edu.cn (J. Wang), 41711055@smail.swufe.edu.cn (J. Chen).

https://doi.org/10.1016/j.ins.2022.11.013
0020-0255/

In fact, previous studies have found that stock movements are affected by various spillover effects in the market, including momentum spillovers [4], media spillovers [1], and analyst spillovers [5]. In particular, the momentum spillover indicates that the fluctuation of a firm is affected by its relevant firms [4]. The media (analyst) spillover suggests that the impact of one news article (analyst) on the stock market can be strengthened or weakened by its related news (analysts) [1,5]. Unfortunately, previous studies only focused on the momentum spillover, ignoring other spillover effects [6–9]. The common strategy of these studies is to condense market information, such as fundamentals, media news or analyst information, as the attributes of a firm node and apply traditional GNN models into stock forecasting, which treats the market information as a non-relational attribute of a node and hence inevitably disconnects their interactions and fails to capture the spillover effects. This essentially poses the problem of how to properly handle relational attributes in GNN-based prediction tasks to capture the spillover effect(s) from the relevance of relational attribute(s).

To address relational attributes in a homogeneous graph, we argue that a relational attribute of a graph node should be separated out and upgraded as a node to form a heterogeneous graph, because if it is treated as a node attribute in the original homogeneous graph as in previous studies, its relational information will be lost. As illustrated in Fig. 1, to model the market information by a homogeneous graph, news articles are condensed and added into the node attributes of the firms. This forces media spillovers to follow firm linkages, inevitably distorting the way of media spillovers. Here, the original relation between news article n_1 and article n_3 disappears, because there is no direct connection between firm f_1 and firm f_3. Meanwhile, two error paths (n_1-n_2 and n_2-n_3) for media spillovers are generated by deferring news relations to the firm relations in the homogeneous graph. In fact, solving the relational attribute problem essentially requires extending homogeneous graph learning to heterogeneous graph learning, leading to the following two unique challenges.

- The first challenge lies in finding an efficient way to preserve the connectivity of relational attributes in graph learning. A natural approach is to separate them into different graph node types and model these multiple relationships in a heterogeneous graph. Therefore, in stock predictions, to capture the media and analyst spillover effects, the news attribute and analyst attribute should be converted into graph nodes to form a heterogeneous market graph. However, the liquidity of media nodes brings up an important issue in the converted heterogeneous graph. Specifically, most heterogeneous graphs are typically made up of static and slow nodes [10,11]. That is, the graph nodes are relatively fixed, which are rarely added or removed from the graph. A good example of the slow variable is the analyst node because the analysts are relatively stable even though the market has a mechanism by which analysts can exit or enter. Different from the analyst node, however, the media node is a fast variable as its influence on the market is approximately one month [12,13]. The vanishing and the appearance of these fast node variables makes the handling of updated learning in heterogeneous graphs considerably challenging. Therefore, designing a proper heterogeneous graph network to incorporate the slow and fast nodes simultaneously for capturing multiple spillovers in asset pricing is imperative.
- The second challenge lies in finding reasonable messaging mechanisms to accommodate messaging within a certain node type or between different node types. Unlike homogeneous graphs that rely on one type of message passing, converting relational attributes to heterogeneous graph nodes requires several messaging channels, each of which controls the message passing for a certain spillover effect. Just as in the market graph (Fig. 1), momentum spillovers of listed firms converge and transport among the firm nodes; the media influence spreads through the media nodes; and the effect of media on the stocks is activated between the firm and media node. Apparently, the node types matter for message transmission in heterogeneous graphs and their roles to control message transportation should be distinguished properly. This is essentially a conditional messaging mechanism in which the transmission or convergence of information is based on the types of two connected nodes. In addition, zooming into the attribute level of the nodes, the information transmission between two nodes is attribute-aware. Let suppose two firm attributes, namely trading volume and profits. In a bull market, sharp changes in the trading volume of a stock tend to ripple through its related firms while the profits of this stock are relatively stable and hard to spread. This fact indicates that the spillover effect of each attribute should be distinguished and dynamically estimated. In summary, to capture multiple spillover effects, the message transmission should be dynamically adjusted in terms of both node types and node attributes in heterogeneous graphs.

To address the above two challenges, we propose a novel conditional heterogeneous graph neural network (FinHGNN). Essentially, the proposed heterograph framework shed lights on addressing rational attributes of a node in homogeneous graphs that is a typical challenge when modeling the stock market as a graph. Although financial research has demonstrated the importance of several spillover effects in stock volatility, previous studies in computer science lacked an efficient way to incorporate them into stock forecasting. To the best of our knowledge, this is a pilot study to apply a heterogeneous graph network to capture multiple types of spillover effects in stock movements, which bridges the gap between finance and computer science research. The study makes the following three unique contributions.

- To preserve the spillover effects of relational attributes, we propose a heterogeneous graph solution by converting relational node features into a new graph node type.
- The conventional heterogeneous graph consists of slow node variables. To incorporate fast node variables obtained from some relational attributes, we propose an intermediate layer mechanism. With this design, the proposed method can transform fast nodes into relatively slow nodes and ensure the handling of both slow and fast node types simultaneously.

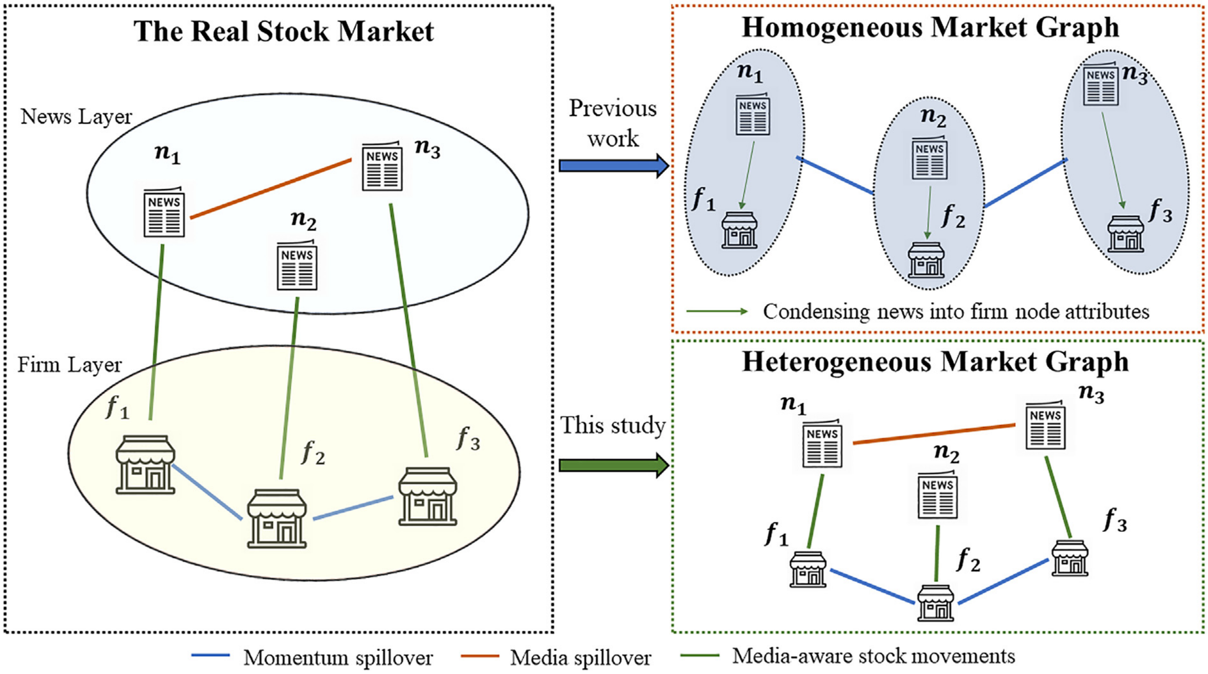


Fig. 1. The market information graph. The left part of the figure presents the paths in the market for spillovers. The upper right figure shows the use of a homogeneous graph to model market information. This forces all spillovers to be subservient to firm relationships and inevitably distorts the spillover paths of media, e.g., lost path (n_1-n_3) and wrong paths (n_1-n_2, n_2-n_3). The lower right figure shows the use of a heterogeneous graph to preserve the original spillover paths.

- A conditional message passing mechanism is proposed to handle hybrid information flows, each of which represents one type of spillover effects in a heterogeneous graph. This is achieved by extending the attention mechanism in graph learning for node-type-aware messaging, and seamlessly integrating a novel non-linear relevance mapping mechanism for attribute-aware messaging.

Experiments on two real-world datasets (S&P 500 and NASDAQ) demonstrate the advantages of the proposed framework over four state-of-the-art algorithms, including eLSTM, TGC, FinGAT, and AD-GAT. This finding further consolidates that modeling multiple spillover effects is a promising way for stock predictions. In particular, the proposed framework achieved statistical enhancements in terms of various evaluation metrics including directional accuracy (at least 2.94%), area under the curve (at least 4.77%), and Matthews correlation coefficient (at least 17.53%).

2. Related work

This study is closely related to two types of previous studies. One is asset pricing via machine learning, and the other is heterogeneous graph learning.

2.1. Asset pricing via machine learning

The study of stock market volatility is a long-standing topic in the field of financial risk. Several studies in both finance and computer science have captured the stock fluctuations via advanced algorithms, including statistical models [14,15], econometric regression models [16,17], conventional machine learning models [2,12] and deep learning models [18–21]. The detailed summary of related literature can be referred to Table A.1 in our appendix.

Statistical models emphasize the correlations between a single feature and stock markets. Econometric regression models focus on the causal relationships between specific features and market movements. However, both statistical and econometric models often have difficulty preserving the interconnections among multiple data sources and thus fail to capture their joint effects on stock performance. Computer scientists have taken a further step by utilizing machine learning algorithms, including support vector machines (SVMs), decision tree (DT), and artificial neural networks (ANNs), to capture such joint effects, and complex nonlinear relationships [12,22,23].

Unfortunately, all of these studies on stock prediction ignored the influences of other firms. That is, they simplified the stock movement problem and assumed that a stock is independent of other stocks and its volatility is determined only by its own market information. In fact, such spillover effect is acknowledged as a momentum spillover effect [4]. To capture the momentum spillover, some pilot studies have utilized graph neural networks for asset pricing [6,9,24,25]. The common strategy of these studies is to model the entire stock market as a graph in which a node represents a firm, and the edge between two nodes for momentum spillovers is denoted by some pre-defined firm relationships, like industry [6], shareholders [25] or price comovement [24].

Recent financial studies, however, show that there are various spillover effects in the market, including analyst spillovers [5] and media spillovers [1]. It is of great interest to consider multiple spillovers for asset pricing. Some researchers have attempted to capture multiple spillover effects by condensing media information as node attributes. For instance, Cheng and Li [7] treated news articles as attributes of firm nodes by counting the sentiment words, and built a homogeneous graph for stock forecasting. Li et al. [24] proposed an LSTM-RGCN prediction model to capture the influence of related stocks based on a homogeneous graph, where each firm is considered as a node and news embedding is considered as a node attribute. However, such an attribution-based approach models media relationships by deferring to firm relationships, which inevitably disconnects and distorts the way that media spills over.

In this work, we present a conditional heterogeneous graph to capture multiple spillovers in stock prediction. This is essentially a challenge on how to deal with the relational attributes in graph learning. Our empirical studies (Section 5) show that a relational attribute of graph nodes should be separated out and upgraded as a node to form a heterogeneous graph to preserve its inherent relationships.

2.2. Heterogeneous graph learning

Another related research topic is the study on heterogeneous graph learning. Unlike homogeneous graphs, heterogeneous graphs exhibit a more complex structure with nodes from different domains [26,27]. The challenge is to properly aggregate feature information from heterogeneous neighboring nodes to incorporate complex structural details in graph learning [28]. A common strategy in previous studies has been to transfer messages in heterogeneous graphs via meta-paths [29,30]. Specially, a meta-path is a sequence of node types in the heterogeneous graph, which is used to guide graph learning to generate node embeddings. For example, Dong et al. [31] manually selected "A-P-A" (representing co-authorship relationship) and "A-P-C-P-A" (representing two papers published by two authors in the same conference venue) as two meta-paths to aggregate information in a bibliographic graph consisting of three heterogeneous nodes, namely, "Author" (A), "Paper" (P) and "Conference" (C). Shi et al. [32] defined eight meta-paths to learn effective heterogeneous network representations for movie recommendation, such as "M-A-M" (indicating two movies with the same actor), and "M-D-M" (indicating two movies with the same director). Both studies treated the information obtained from each meta-path equally. However, because meta-paths contain different relations with different semantics, the importance of each meta-path should be distinguished and weighted accordingly. Wang et al. [33] took a further step by proposing a HAN model to aggregate the attributes of neighboring nodes from several meta-paths hierarchically. This hierarchical approach used semantic-level attention to distinguish the importance of different meta-paths.

Essentially, a meta-path focuses on describing a specific or meaningful relationship of nodes to select neighboring nodes for information aggregation in graph learning. Apparently, the selection of meta-paths requires some prior knowledge about specific learning problems. One must know they want to explore co-authorship before they select the "A-P-A" meta-path in the example above. More importantly, the meta-path approach is unable to explore message flows that are hard to specify clearly, like the amplification or ripple effect among news articles for media-aware stock movements.

In this study, we present a novel method to model different message flows in heterogeneous graphs without any prior knowledge to define or select meta-paths. This is achieved by differentiating neighboring nodes and aggregating information conditional on the states of the connected nodes. Specifically, both node types and attributes are considered to control the spillover effects in a unified learning process. In this way, multiple spillovers are captured in asset pricing. The proposed conditional message-passing mechanism shed lights on the heterogeneous graph learning without pre-defined meta-paths.

3. Preliminary

Before providing details of our proposed model, we briefly introduce the formal definition of heterogeneous graph networks.

Definition 1 (*Heterogeneous Graph Network*). A heterogeneous graph network is a special kind of graph network. Unlike traditional graph neural networks (homogeneous networks), it comprises different types of nodes and links. It can be denoted as an undirected graph $\mathscr{G} = (\mathscr{V}, \varepsilon)$. Here, $\mathscr{V}$ is the set of nodes, and each node $v_i \in \mathscr{V}$ belongs to one of node types $n \in \{n_1, n_2 \ldots, n_N\}$, where N is the total number of node types. ε denotes the set of links, and each link $e_{i,j} \in \varepsilon$ belongs to one of link types $m \in \{m_1, m_2 \ldots, m_M\}$, where M is the total number of edge types.

Problem Statement: As indicated in recent financial studies [5,17], the market information to shape stock movements includes firm's fundamentals ($\mathbf{f}$), analyst information ($\mathbf{s}$), and firm-specific news ($\mathbf{p}$). In this study, the heterogeneous graph network consists of three node types ($\mathscr{V}$) and five link types (ε), and more details are presented in Section 4.1. We treat stock prediction as a binary classification problem, i.e., each firm node needs to be classified as one of the upward or downward classes ($y \in \{-1, 1\}$). The stock trend of a firm node is determined by these three kinds of market factors along with their spillover effects, namely, momentum, analyst and media spillovers.

4. Model architecture

In this study, we propose a heterogeneous graph learning to simultaneously address multiple types (momentum spillovers, media spillovers and analyst spillovers) of spillover effects in stock movements, each of which has been proven to be effective in shaping stock trends in financial research. Previous studies either ignore the spillover effects using non-graph ML methods or rely on homogeneous graphs to capture the momentum spillovers. To model the spillovers among attributes, i.e., news and analysts, we separated these attributes into graph nodes and modeled their relations in a heterogeneous market graph. The challenge lies in incorporating slow and fast nodes simultaneously, and designing a conditional message passing mechanism to distinguish different spillovers in heterogeneous graphs. Therefore, we propose a novel conditional heterogeneous graph neural network (FinHGNN) to deal with relational attributes for capturing multiple spillover effects in asset pricing. Fig. 2 is an overview of the proposed framework.

4.1. Heterogeneous market graph

As aforementioned, most previous studies state that the price of a stock is driven by its historical fundamentals. Recent financial studies have shown that stock movements are also associated with multiple spillover effects. Therefore, all the firm, analyst, and news information are considered in this study. In the following, we present the embedding of each type of node, and the relationship between nodes as the inputs of the proposed heterogeneous market graph learning.

4.1.1. Firm embedding

The stock market is a dynamic system in which market signals are time-series. The embedding vector of firm i is typically represented as its fundamentals of the past T days, and these fundamentals are the main concerns of technical analysts in traditional quantitative investing. Specifically, a single layer LSTM [34] is adopted to capture the time dependency, as it is efficient enough in practice, and denoted as

$$\mathbf{f}_i = LSTM\left(\sum_{t=1}^{T}\mathbf{m}_{i,t}\right), \quad \mathbf{f}_i \in \mathbb{R}^F. \tag{1}$$

where vector $\mathbf{m}_{i,t}$ denotes the fundamental information on day t, and its dimension F indicates the number of fundamental features including trading volume, turnover, PB ratio, PE ratio, opening price and so forth.

4.1.2. Analyst embedding

Previous financial research has demonstrated that the firms co-commented by the same analyst (analyst co-coverage) is a good way (proxy) to measure economic connections between stocks that essentially cause the effect of analyst spillover in stock movements [5]. In other words, the vector of analyst nodes should have common coverage information for analysts to track spillover effects. To reach this goal, we first initialize the analyst‘s embedding vector as a one-hot vector that can simply and efficiently distinguish analysts. To capture their co-coverage for stock volatility, we update the analysts‘ embedding vectors in a manner similar to word embedding. Specifically, we consider each analyst as a basic unit (word), and the analysts commenting on the same firm as the context (sentence) of these analysts. For example, suppose analyst A, B, and C comment on firm 1, the context of these analysts is the collection consisting of the one-hot vectors A, B, and C. Then, we can take A and C as inputs and B as the output by feeding them into neural networks learned by CBOW [1]. In this way, the embedding ($\mathbf{s}_i$) of analyst i enriched by the co-coverage information of analysts can be learned by maximizing probability $p(\mathbf{s}_i|C(\mathbf{s}_i, \mathbb{C}))$ as

$$max \prod_{j \in C(\mathbf{s}_i, \mathbb{C})} \frac{\exp(\mathbf{s}_i \mathbf{s}_j)}{\Sigma_{k \in C(\mathbf{s}_i, \mathbb{C})} \exp(\mathbf{s}_i \mathbf{s}_k)}. \tag{2}$$

where $\mathbf{s}_i \in \mathbb{R}^S$ is the initial one-hot embedding of the analyst node, and dimension S indicates the number of analysts in the market. The context $C(\mathbf{s}_i, \mathbb{C})$ of analyst i is defined as the contextual information of analyst nodes, which is the partial collection of analysts commenting on the same firm selected for CBOW learning.

4.1.3. Media embedding

In addition to stock co-movement caused by firm fundamentals (Section 4.1.1) and stock analysts (Section 4.1.2), there is an important spillover effect caused by the release of news articles. In finance, stock movements are media-aware movements that can be enhanced by a ripple or amplification effect from relevant news. That is, stock volatility caused by a news

[1] The continuous bag of words (CBOW) model is a predictive deep learning-based model to compute and generate high-quality, distributed, and continuous dense vector representations of words, which can consider the surrounding words in a sequence of words. Due to its predictive structure, it is often used to learn embeddings of sequential entities [35].

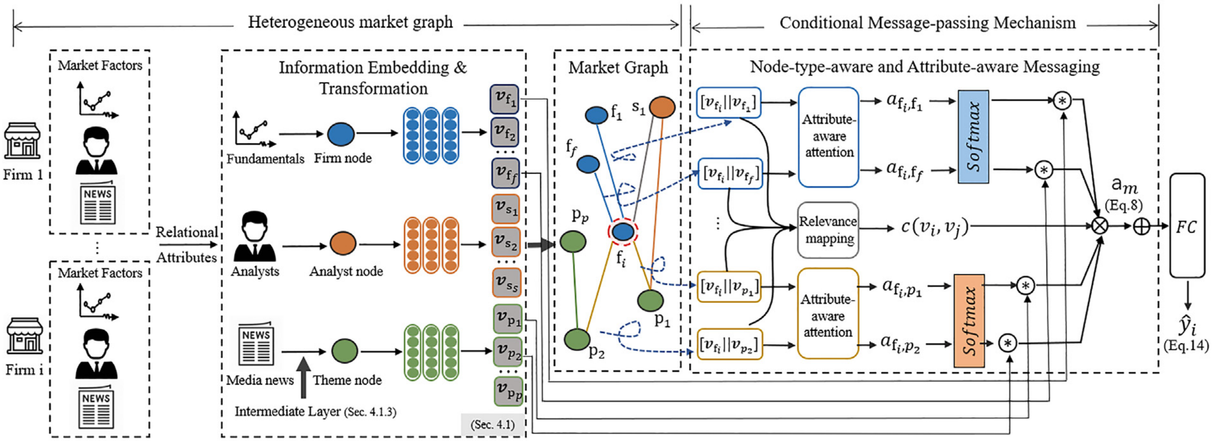


Fig. 2. The overview of the proposed FinHGNN.

article can be exacerbated by the release of related news. Therefore, we should also include news as a node type in the heterogeneous market graph, along with firm fundamentals and stock analysts' nodes. However, the rapid appearance and vanishing of daily news articles make them inappropriate as a node type for heterogeneous graph learning. This fact indeed poses a challenge on how to mix fast and slow node variables in heterogeneous graph learning. In this study, we propose the concept of a proxy node that converts fast node variables into relatively slow node variables. Recall that the purpose of news nodes is to capture the ripple or amplification effect of related news, which is caused by the superimposed effect of news with similar content or by the herding effect caused by related news with little content relevance. A good example of the herding effect with different news content is that the panic caused by the news of a firm being sued accelerated the downward pressure on the stock caused by the news of an earlier product recall. Therefore, we pre-classify news into different news themes based on the Bert model to capture the related news with similar contents (the superimposed effect) and then calculate the co-occurrence of news themes to capture the related news with non-textual similarity (the herding effect). By doing this, we can replace the news nodes (the entities with fast-changing characteristics) with the theme nodes (the entities with slow-changing characteristics) to form the heterogeneous graph. In this study, the initial embedding of theme node, $\mathbf{p}'_i$, is defined as

$$\mathbf{p}'_i = \sum_{k=1}^{D_i} \mathbf{n}_i^k / D_i, \quad \mathbf{p}'_i \in \mathbb{R}^P, \tag{3}$$

where $\mathbf{n}_i^k$ is the k-th news article during the past T days that is mapped into theme i, and D_i is the total number of news articles in theme i. Each news article is given a vector of P dimensions by the BERT model [36]. Here, k-means and hierarchical clustering algorithms are adopted to obtain themes. In such a way, theme nodes are representative of news with similar content. To capture the herding effect of the related news with less content relevance, we updated the initial theme embeddings with CBOW learning, in which each theme is considered as a basic unit (word), and the context (sentence) is a collection of themes ($C(\mathbf{p}'_i)$) to which the news articles of a firm published within T days belong. Therefore, the final theme embedding $\mathbf{p}_i$ is obtained as

$$max \prod_{j \in C(\mathbf{p}'_i, \mathbb{C})} \frac{\exp(\mathbf{p}'_i \mathbf{p}'_j)}{\Sigma_{k \in C(\mathbf{p}'_i, \mathbb{C})} \exp(\mathbf{p}'_i \mathbf{p}'_k)}, \tag{4}$$

where $\mathbb{C}$ is a control parameter to select partial $C(\mathbf{p}'_i)$ for CBOW learning.

4.1.4. Embedding transformation

In general, a graph node aggregates information from its neighboring nodes to incorporate graph structure information. A common strategy in traditional graph learning (homogeneous graph learning) is to sum-average node embeddings directly, as these graphs only contain nodes of the same kind with the same embedding vector size. However, in heterogeneous graphs, different node types typically have different numbers of attributes, and thus their embedding vectors are of different sizes. In our proposed heterogeneous market graph, the vector dimensions of firm nodes, analyst nodes, and theme nodes are F, S, and P, respectively. Such heterogeneity makes direct aggregation of information challenging, as in the case of the homogeneous graph. Therefore, as suggested by related studies [28,37], a linear transformation method is proposed to project the embeddings of various node types into a unified node representation with the same dimensionality for further manipulation in graph learning. For a certain node i, its unified node embedding $\mathbf{v_i}$ is denoted as

$$\mathbf{v}_i = \begin{cases} W_f \times \mathbf{f}_i & \text{if} \quad node \quad i \in \mathscr{V}_f, \\ W_s \times \mathbf{s}_i & \text{if} \quad node \quad i \in \mathscr{V}_s, \\ W_p \times \mathbf{p}_i & \text{if} \quad node \quad i \in \mathscr{V}_p, \end{cases} \tag{5}$$

where $W_f \in \mathbb{R}^{L\times F}$, $W_s \in \mathbb{R}^{L\times S}$, and $W_p \in \mathbb{R}^{L\times P}$ represent the linear transformations for the firm node, the analyst node, and the theme node, respectively. Essentially, W_f is the parameter matrix to convert the F dimensional firm node embedding $\mathbf{f}_i$ into vector $\mathbf{v}_i$ with L dimensions. Similarly, W_s and W_p are the transform matrixes to convert the embeddings of analyst and theme nodes into L dimensional vectors, respectively. These parameters are estimated in the learning task simultaneously.

4.1.5. Node links

In the proposed heterogeneous market graph, there are three types of nodes, including firms, analysts, and news. To incorporate multiple spillovers, five link types are considered in this study. Specifically, **firm-firm** links are built to capture the influence of related firms. Even though various firm relations guide such momentum spillovers in the market, recent financial studies show that media co-exposure is dominant for momentum spillovers [5,38]. Therefore, we use the number of themes to which the news of the two connected nodes belong to build the firm-firm edge. **Theme-theme** links are established for media spillovers. As aforementioned, the purpose of theme linkage is to capture the herding effect caused by related news with less relevant content (Section 4.1.3). Therefore, we calculate the number of firms containing news articles belonging to two connected themes in a given period as the weight of the theme-theme link. **Analyst-analyst** links are determined by the number of firms covered by the connected analysts. This link type aims to reflect the similar foci of analysts on some stocks (analyst spillovers). **Theme-firm** links are determined by the number of news articles belonging to the connected theme. This link type essentially reflects the media-aware stock movement. **Analyst-firm** links are determined by the coverage of the analyst on this firm. Similar to theme-firm links, this link type captures the impact of concerns from analysts on stock movements. All edges are normalized for the following processing. Using these node links to establish connections of different types of nodes essentially constructs a way to aggregate and transmit multiple spillover effects for stock predictions.

4.2. Conditional messaging in FinHGNN

A common strategy in previous studies defines meta-paths for information transmission and aggregation using a priori knowledge. For example, meta-path M-A-M is used to explore movie-director-movie message flows for recommending movies with graph learning [32]. In this study, we propose a conditional messaging mechanism in heterogeneous graph learning without any pre-defined meta-paths by learning the probability of information flow conditional on both node type and node attributes. Fig. 3 is the overview of the proposed conditional messaging mechanism.

4.2.1. Messaging conditional on node types

In a traditional graph network (homogeneous graph), the information transferred between nodes is independent of the node types, as all nodes are of the same type. However, for the heterogeneous market graph, the messaging of different node types has different messaging functionalities. For example, the messaging between firm nodes and theme nodes captures the media-aware stock movement, the messaging between firm nodes reveals momentum spillovers of the listed firms, and the messaging between theme nodes simulates the ripple or amplification effect of news (media spillovers). Therefore, for node i, its aggregated information from a certain type of spillover effect can be indicated as

$$\begin{aligned} \mathbf{h}_{i,m} &= \sum_n \sum_m \sum_{j\in U} G(\mathbf{v}_i, \mathbf{v}_j, \theta | n, m) \\ &= \sigma(\sum_{j\in U} e_{i,j} W_m \mathbf{v}_j + W_m \mathbf{v}_i + \mathbf{b}_m), \end{aligned} \tag{6}$$

where $\mathbf{v}_i$ and $\mathbf{v}_j$ are the unified node embeddings for node i and node j, respectively. $n \in \{n_1, n_2, \ldots, n_N\}$ is the node type that node i belongs to, $m \in \{m_1, m_2, \ldots, m_M\}$ is the link type that the edge between node i and j belongs to, and U denotes the neighboring nodes of the target node i with link type m. Given node type n and link type m, the aggregation function $G(\cdot)$ with a set of parameters θ is utilized to aggregate the information in the heterogeneous graph. Specifically, The learning parameter θ of function $G(\cdot)$ is defined as $\theta = [W_m, \mathbf{b}_m]$, in which $W_m \in \mathbb{R}^{L\prime\times L}$ is a weight matrix, and $\mathbf{b}_m \in \mathbb{R}^{L\prime}$ is the bias vector. $e_{i,j}$ is the link between node i and j. σ is the sigmoid function. By doing this, the L-dimension $\mathbf{h}_{i,m}$ is converted into $L\prime$ dimension via the aggregation function.

In the heterogeneous market graph, the aggregated information $\mathbf{h}_i$ of node i is the combination of aggregated information obtained from each link type m. Here, the importance of $\mathbf{h}_{i,m}$ is calculated by the concat product that is a classic non-linear transformation adopted in graph attention network [37]. Therefore, the aggregated information $\mathbf{h}_i$ is defined as

$$\mathbf{h}_i = \sum_m a_m \mathbf{h}_{i,m}, \tag{7}$$

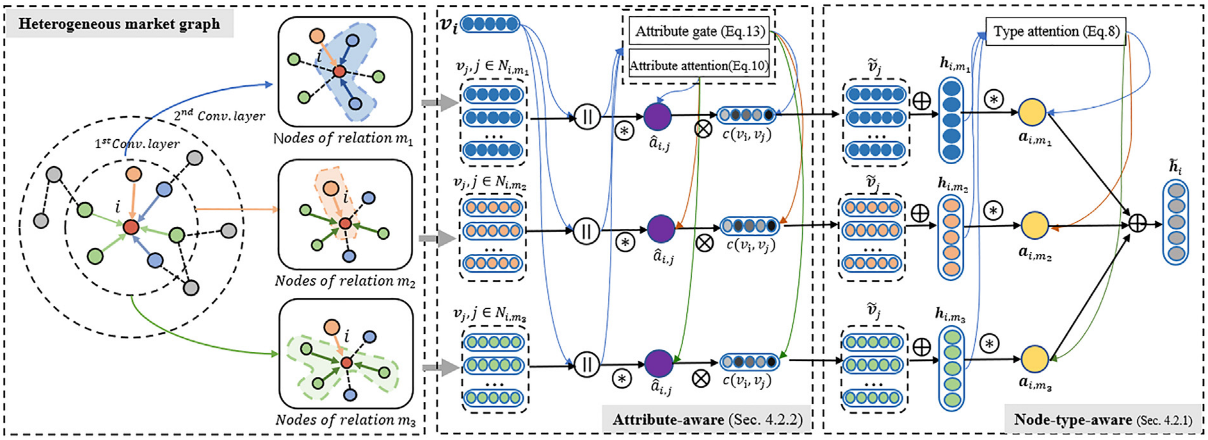


Fig. 3. The process of conditional messaging mechanism.

$$a_m = \frac{\exp(LeakyReLU(\mathbf{u}^T[\mathbf{h}_i'||\mathbf{h}_{i,m}])}{\sum_m \exp(LeakyReLU(\mathbf{u}^T[\mathbf{h}_i'||\mathbf{h}_{i,m}])}. \tag{8}$$

Here, $||$ denotes the concatenation operation. $\mathbf{u}^T \in \mathbb{R}^{2L'}$ is the trainable attention parameter shared by different link types. In this way, messaging in the heterogeneous graph is conditional on link types, which is tailored to capture multiple spillovers in the stock market.

4.2.2. Messaging conditional on node attributes

As aforementioned, the information propagation is attribute-aware. The meaning of attribute-awareness in this study is twofold as follows.

a) For a target node, conditional on the link type m, the difference in the amount of information transmitted by its neighboring nodes is closely related to the state of those neighbors. Apparently, each neighboring node has a different impact on the target node. Suppose that both firm i and j are related to target firm k due to the supply chain relationship. Since firm i is a major supplier to target firm k, the effect of firm i on target firm k will be more significant. To distinguish the different effects of neighboring nodes, we introduce an attention function, $a(\mathbf{v}_i, \mathbf{v}_j)$. Therefore, the aggregated information $\mathbf{h}_{i,m}$, is further redefined as

$$\mathbf{h}_{i,m} = \sigma(\sum_{j\in U} a_{i,j} W_m \mathbf{v}_j + W_m \mathbf{v}_i + \mathbf{b}_m), \tag{9}$$

where $a_{i,j}$ is the weight score generated by $a(\mathbf{v}_i, \mathbf{v}_j)$, indicating the importance of each neighbor. $a(\cdot)$ is essentially a non-linear scoring function obtained by a single-layer feed-forward neural network with *LeakyReLU* activation function. Specifically,

$$a_{i,j} = a(\mathbf{v}_i, \mathbf{v}_j) = LeakyReLU([\mathbf{v}_i^T||\mathbf{v}_j^T]\ \mathbf{a}_r). \tag{10}$$

Here, $||$ denotes the concatenation operation. $\mathbf{a}_r \in \mathbb{R}^{2L}$ is the trainable attention parameter shared by the neighboring nodes with the same link type. Note that, with $\mathbf{a}_r$, the most representative node pairs within the given link type can be distinguished by $a_{i,j}$. To make $a_{i,j}$ comparable over all the nodes, it can be further normalized as

$$\hat{a}_{i,j} = \exp(a_{i,j})/\sum_{j\in U}\exp(a_{i,j}). \tag{11}$$

b) Each attribute of the target node overflows in a different way. In real-world stock markets, the spillover effects are sensitive to the attributes. For example, an abnormal price decline in one firm may not spill over if this price decline is traded in small volumes or if the linked firms are undervalued. This fact indicates that the spillover effect of each attribute should be distinguished and dynamically estimated. Therefore, different from the traditional approach, in which all attributes are transmitted together with the same weight, this study introduces a non-linear relevance mapping mechanism $c(\cdot)$ to distinguish the spillover of each attribute. In this way, the aggregated information $\mathbf{h}_{i,m}$for node i can be reformulated as

$$\sigma(\sum_{j\in U}\hat{a}_{i,j} W_m \mathbf{v}_j \otimes c(\mathbf{v}_i, \mathbf{v}_j) + W_m \mathbf{v}_i + \mathbf{b}_m), \tag{12}$$

where $\otimes$ denotes the element-wisely multiplication. $c(\mathbf{v}_i, \mathbf{v}_j)$ essentially indicates a gate mechanism that is designed to control message passing conditional on the attributes of node i and j, denoted as

$$c(\mathbf{v}_i, \mathbf{v}_j) = sigmoid(W_c[\mathbf{v}_i||\mathbf{v}_j] + \mathbf{b}_c), \tag{13}$$

where $W_c \in \mathbb{R}^{L' \times 2L}$ and $\mathbf{b}_c \in \mathbb{R}^{L'}$ are the parameters to be estimated in the learning process. With this design, the proposed method is able to control how much information each attribute should spill over to the target node under the condition of the attributive states of the two connected nodes.

4.3. Output Mapping Module

Finally, a single-layer feed-forward neural network with the softmax function is applied to generate the probability of future stock trends, denoted as

$$\hat{y}_i = O_i(\mathbf{h}_i) = softmax(W_o\mathbf{h}_i + \mathbf{b}_o), \tag{14}$$

where $W_o \in \mathbb{R}^{C \times L'}$ is the weight matrix, $\mathbf{b}_o \in \mathbb{R}^C$is the bias vector. C is the number of classes. Essentially, the stock prediction task is a binary classification task. By minimizing the binary cross-entropy loss between $\hat{y}_i$ and y_i, the parameters of the proposed framework can be learned. The pseudocode for this proposed algorithm is presented in Algorithm 1.

Algorithm 1: Conditional Messaging Procedure of FinHGNN

Input: A heterogeneous market graph $\mathcal{G} = (\mathcal{V}, \varepsilon)$, target firm node and its associated stock trend y_i.

Output: The trained model for stock prediction.

for *node* i *in* $\mathcal{V}$ **do**
  for *link type* $m = 1$ *to* M **do**
    Obtain the neighboring nodes (U) of the target node i with link type m;
    for j *in* U **do**
      Calculate the attention weight $a_{i,j}$ at the node level via Equations (10) and (11);
      Calculate the attribute gate $c(\mathbf{v}_i, \mathbf{v}_j)$ at the attribute level via Equation (13);
    end
    Obtain aggregated information $\mathbf{h}_{i,m}$ via Equation (12);
  end
  Calculate the score a_m for each link type via Equation (8);
  Obtain aggregated information $\mathbf{h}_i$ via Equation (7);
  Obtain the output via Equation (14);
end

---

5. Experiments

To the best of our knowledge, this study is the first to explore a heterogeneous graph to address relational attributes for multiple spillovers in asset pricing. To gauge the effectiveness and robustness of the proposed framework, we conducted a series of experiments from the perspective of predictability and profitability.

5.1. Experiment settings

5.1.1. Datasets

Although stock transaction data is accessible, there are few public stock datasets containing firm-specific news and analyst information. In this study, two real-word datasets are collected to evaluate the proposed framework. Table 1 shows the data statistics. In particular,

- **S&P 500**: This dataset has been widely adopted in previous studies as a benchmark [6,7,18]. It contains the market transaction data of S&P 500 from February 8, 2011 to November 18, 2013, and the financial news published by Reuters and Bloomberg. There are 198 stocks with 371,579 news articles during the period of 700 transaction days. Note that, this

Table 1
Descriptive statistics of the data.

Datasets	Nodes			Edges					Labels	Periods
	Firm	Analyst	Theme	F-F	A-A	T-T	F-A	F-T		
S&P 500	198	3,539	40	3,218	6,726	482	5,915	557	1: 52.09% −1: 47.91%	08/02/2011 −18/11/2013
NASDAQ	1048	4,259	40	11,336	8,296	699	8,252	4,019	1: 50.82% −1: 49.18%	01/10/2019 −29/03/2021

dataset focuses on partial market factors. We collect the analyst information from Wharton Research Data Services[2]. The analyst dataset named Institutional Brokers' Estimate System [3], covers consensus and details forecasts of the securities analysts from more than 22,000 listed firms in 100 countries. Here, we select the analysts related with the firms listed in S&P 500 during the same trading period mentioned above.

- **NASDAQ**: Even though S&P 500 dataset is well-acknowledged in the field of asset pricing via machine learning, it only contains 198 stocks. To gauge the robustness of the proposed framework, we construct a larger dataset based on the firms listed in NASDAQ. This dataset contains 1,048 stocks with 457,076 news articles selected from October 1, 2019 to March 29, 2021. It is 5 times larger than S&P 500 dataset, and has one-fifth of the total firms listed in NASDAQ. These 1,048 stocks were selected according to the filtering rules adopted in [7], which requires that each stock should not miss any transaction record from the service provider and has at least 100 related news articles during the selected period. Here, the transaction record and analyst information are also provided by WRDS. The firm-specific news articles posted on the Nasdaq website[4] are collected by our focused crawler [39].

In addition, both datasets were divided sequentially into three periods. The first 70% of transaction days were used for training, the following 10% of days for validation, and the last 20% of the days for testing. The evaluation was conducted in a rolling-window fashion, as suggested by Cheng and Li [7]. That is, the market signals during the past T transaction days were utilized to predict future stock trends. These datasets and key algorithms are delivered to investors via security analyzer (stock++)[5].

5.1.2. Evaluation metrics

Most asset pricing studies via machine learning treat the prediction task as a binary classification problem. Only a few researches regard it as a regression task [40]. To make the proposed framework comparable, we also treated this as a binary classification problem. Specifically, if the closing price on the next day is higher than the closing price today, the sample is labeled with "upward" ($y_i = 1$), otherwise labeled with "downward" ($y_i = -1$). In Table 1, there are 52.09% "upward" samples and 47.91% "downward" samples in the S&P 500 dataset, and 50.82% "upward" samples and 49.18% "downward" samples in the NASDAQ dataset, respectively. To gauge the performance of the proposed framework, the three most popular metrics adopted in the previous studies [3,41] were utilized as our evaluation metrics, namely, directional accuracy (DA), area under the curve (AUC) and the Matthews correlation coefficient (MCC). Specifically,

$$DA = \frac{TP + TN}{TP + TN + FN + FP}, \tag{15}$$

$$AUC = \int_0^1 \frac{TP}{TP + FN} d\frac{FP}{FP + TN}, \tag{16}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \tag{17}$$

where $TP(TN)$ denotes the number of positive (negative) samples correctly classified as positive (negative). $FP(FN)$ is the number of negative (positive) samples falsely classified as positive (negative) samples. The DA and AUC are widely adopted as the evaluation metrics for stock classification tasks. DA measures the difference between predicted trends and actual changes in stock prices, and AUC obtained from the receiver operating characteristic (ROC) curve depicts the area under the ROC curve. In addition, MCC is used to avoid evaluation bias caused by skewed data. The higher the values of these metrics, the better the predictive performance of the model.

[2] Wharton Research Data Services (WRDS): https://wrds-www.wharton.upenn.edu
[3] Institutional Brokers' Estimate System (IBES): https://wrds-www.wharton.upenn.edu/pages/get-data/ibes-thomson-reuters
[4] https://www.nasdaq.com/market-activity
[5] The key techniques of this study were implemented and deployed at stock++. This system can be accessed athttps://fic.swufe.edu.cn/info/1228/1257.htm.

Note that, to ensure the robustness of the evaluation, for each comparison, we trained it 10 times with different initializations. As suggested by Feng et al. [6], the average performance of these runs in the testing period is reported to eliminate fluctuations caused by different initializations.

5.1.3. Parameter settings

Here, we report the parameter settings as follows:

- In this study, the firm attributes include the volume, turnover, PB ratio, PE ratio, opening, highest, lowest, and closing prices. Considering the time-series of these attributes, the representations of firm attributes on t-th day are obtained from the records of the past T days via the LSTM (Eq. 1) with the dimension of the firm node embedding (F). Here, the optimal T is set to 20 by searching the grid in the range of $\{5, 10, 15, 20, 25, 30\}$, and F is set to 32 searching within $\{8, 16, 32, 64, 128, 256\}$. Fig. 4 and Fig. 5 illustrate the prediction performance in terms of DA and AUC with different T and F, respectively.
- Similarly, for the node embedding of analyst and theme, the optimal value is set to 64 and 256 via grid searching within $\{8, 16, 32, 64, 128, 256\}$. Note that with the BERT model, each news article is represented as a dimensional vector of 256. In fact, the BERT model initially produces a dimensional vector of 768. To reduce the computational complexity, we further mapped it to obtain a dimensional vector of 256. In our preliminary experiments, there was no noticeable difference when projecting the article embedding dimension from 768 to 256 with statistical significance.
- Recall that we convert the entities with fast-changing characteristics (daily news) into the entities with slow-changing characteristics (themes) to capture media spillover effects in stock markets. Fig. 6 shows the prediction performance with different number of themes. It can be observed that the optimal theme number is around 40. In addition, the theme boundaries with the different number of themes on S&P 500 are presented in Fig. 7 for better understanding.
- Recall that the embedding size of different numbers should be the same for free messaging in heterogeneous graphs (Section 4.1.4). In this study, we unify the size of firm embedding (F), theme embedding (P) and analyst embedding (S) into dimension L via a transformation operation (Eq. 5). In Fig. 8, the optimal L is obtained at 64.
- For the neural network, we test different layers ($L\prime$) within $\{8, 16; 32; 64; 128; 256\}$ and find the optimal value is 32. All network parameters are initialized using Glorot [42] and trained using the Adam optimizer [43] with an initial learning rate of 0.0005. We train a maximum of 300 epochs.

5.2. Comparison

To evaluate the overall performance of the proposed framework, we compared it with several baselines, including three classic and four state-of-the-art models. In particular,

5.2.1. Classic classifiers

.

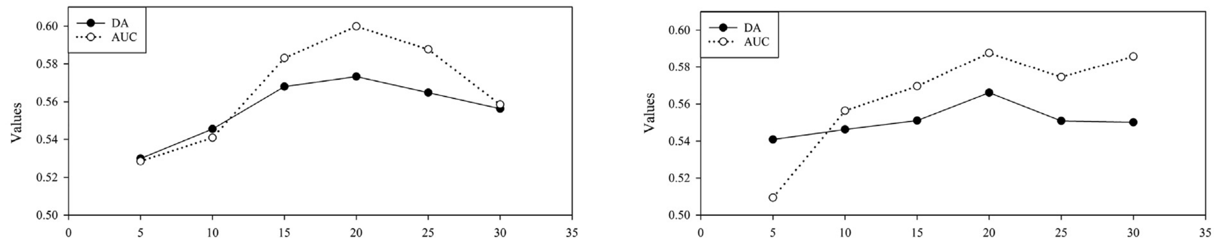
- LSTM: This is one of the most powerful deep learning models for time-series data. The LSTM with 2 layers is adopted.
- GCN: This aggregates the attributes of neighboring nodes to central nodes in a linear way based on the normalized connected relations. The GCN with 2 convolution layers is implemented for a homogeneous firm graph.
- HGNN_{Fin}: This is derived from a conventional HGNN, where relational attributes (news and analysts) are converted to graph nodes. Different spillover effects are separately transported in pre-defined meta-paths, as in the classical HGNN [29].

5.2.2. State-of-the-art models

.

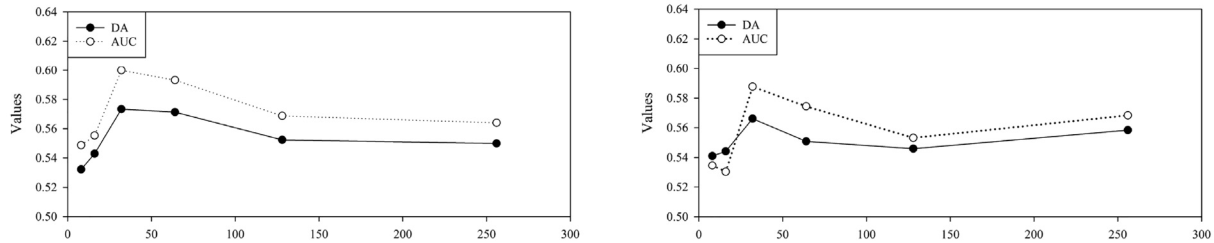
- eLSTM: Li et al. [3] proposed an enhanced LSTM model, which is capable of handling extra shock from media and momentum spillovers to some degree.
- TGC: Feng et al. [6] proposed a temporal graph convolution framework (TGC) to capture momentum spillover effects for predicting stock movements, which constructed a homogeneous market graph in terms of the listed firms.
- FinGAT: Hsu et al. [9] proposed the FinGAT model to estimate the firm relations within each industry category for capturing momentum spillovers.
- AD-GAT: Cheng and Li [7] proposed a homogeneous graph method, which designs an attribute-matter (AM) aggregation mechanism to incorporate momentum spillover effects from related firms for capturing stock movements.

Here, LSTM and eLSTM are nongraph-based methods. In contrast, the GCN, TGC, FinGAT and AD-GAT are graph-based methods. All baselines use fundamental indicators and news features as inputs, whereas graph-based baselines require additional pre-defined firm relationships. Here, graph-based baselines are homogeneous graphs, where the graph node link is determined by firm relations. Specifically, for GCN and TGC, we used the media co-exposure relation to construct a market graph, as the media co-exposure has proven to be the best of existing relations [3,38]. Both FinGAT and AD-GAT estimate



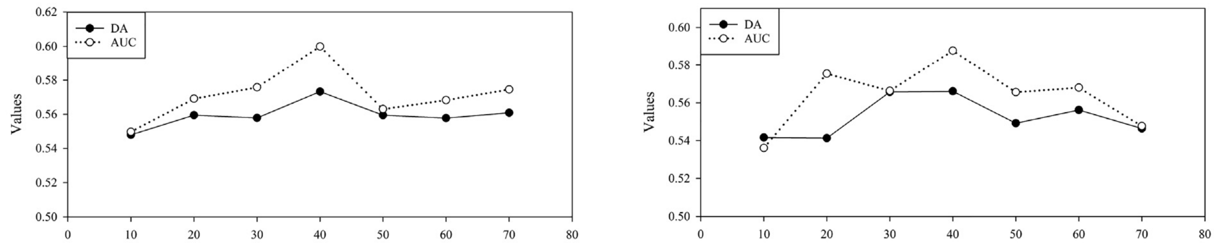
(a) S&P 500   (b) NASDAQ

Fig. 4. The variations of DA and AUC with different T.



(a) S&P 500   (b) NASDAQ

Fig. 5. The variations of DA and AUC with different F.



(a) S&P 500   (b) NASDAQ

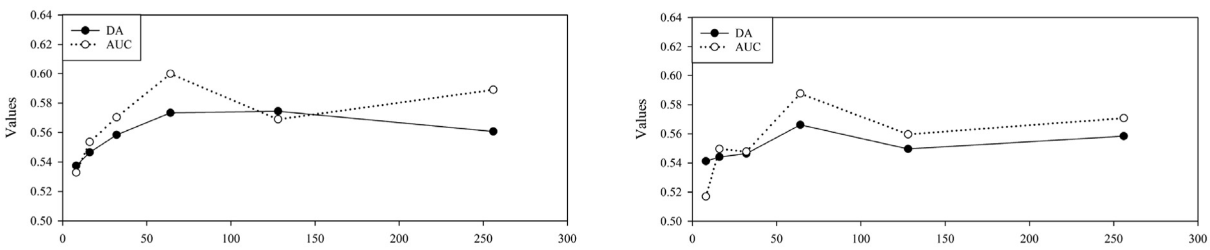
Fig. 6. The variations of DA and AUC with different number of themes.

firm relations with the observed firm attributes. The difference is that the estimation of FinGAT is restricted to each industry category, while AD-GAT estimates the relationships for the entire graph. Note that, the network parameters of all baselines were selected by grid search as described in Section 5.1.3, and the detailed parameter settings can be referred to Appendix B.

Table 2 presents the prediction results of baselines and the proposed FinHGNN. Almost all methods perform better on the S&P 500 than on the NASDAQ dataset, indicating that stocks in the S&P 500 are more predictable than those in the NASDAQ. A good explanation is that NASDAQ stocks are more volatile than S&P 500 stocks. This fact has also been observed in a previous study [6]. In both datasets, GCN that is capable of handling momentum spillovers performs slightly better than LSTM that captures the time dependence of market factors. This indicates that to improve the stock predictions, capturing momentum spillovers is as important, if not more so, than dealing with time-series market data. When the models (eLSTM, TGC, FinGAT and AD-GAT), that are capable of dealing with time-series and momentum spillover issues, are utilized, stock predictions can be further improved. When further incorporating multiple spillovers, HGNN_{Fin} and the proposed FinHGNN outperformed all the baselines. Specifically, for the S&P 500 dataset, the proposed method achieved the best performance with enhancements of at least $2.96\%, 4.77\%$ and 17.53% in terms of DA, AUC, and MCC, respectively. For the NASDAQ dataset, enhancements of the proposed method are at least $2.94\%, 5.64\%$ and 41.59% in terms of DA, AUC, and MCC, respectively.

(a) 20 themes (b) 40 themes (c) 60 themes

Fig. 7. The clustering performance for different number of themes.



(a) S&P 500 (b) NASDAQ

Fig. 8. The variations of DA and AUC with different *L*.

Table 2
Comparison of different baselines.

Method	S&P 500			NASDAQ			Time consumption (second)
	DA	AUC	MCC	DA	AUC	MCC	
LSTM	0.5136	0.5139	0.0677	0.5025	0.5040	0.0081	36.1582
GCN	0.5182	0.5183	0.0714	0.5143	0.5094	0.0199	179.1056
$HGNN_{Fin}$	0.5616	0.5613	0.1586	0.5503	0.5516	0.1147	7671.5369
eLSTM	0.5318	0.5320	0.0639	0.5344	0.5355	0.0716	71.8947
TGC	0.5573	0.5584	0.1180	0.5213	0.5259	0.0543	633.7758
AD-GAT	0.5614	0.5607	0.1329	0.5483	0.5491	0.0989	4897.6484
FinGAT	0.5455	0.5474	0.0979	0.5414	0.5407	0.0817	4143.8079
FinHGNN	**0.5782**	**0.5881**	**0.1864**	**0.5665**	**0.5827**	**0.1624**	3154.4442

The *p*-values for the *t*-tests were all less than the critical confidence value (0.05), indicating that the superior performance of the proposed approach is statistically significant.

In addition, we further compared the time consumption of the proposed method with that of all the baselines. Two Tesla V100 GPU cards were used for the training. The average training costs for each model are presented in Table 2. It can be observed that the computing time of the graph-based models is generally longer than that of the nongraph-based models. This is well understood because in the graph-based methods, there were additional estimated parameters to capture the market spillovers, which inevitably increases the computing complexity. Among all the graph models, although the proposed FinHGNN is not the most efficient in terms of computing time, it is generally better than heterogeneous methods ($HGNN_{Fin}$) and even some homogeneous methods (FinGAT and AD-GAT). In addition, the closure time of the securities market is around 18 h, which provides sufficient time for training the graph-based models offline. This indicates that the computing cost of our method is affordable and worthwhile in practice.

5.3. Investment simulation

To further measure the robustness of the proposed framework, we perform backtesting by simulating stock investments. As suggested by Xu et al. [20], the top-k strategy is adopted in this study. Specifically, the top-k stocks with the highest predicted ranking scores in each model are bought and the holding shares of the stocks ranked below k are sold. The number of stocks (k) is set to $5, 10, 15$, and 20, respectively. Here, we assume zero transaction costs, as in previous work [2,3], and use two financial metrics to evaluate simulation results, i.e., cumulative returns and Sharpe ratio.

Fig. 9(a) and (b) show the cumulative returns for the last 70 days of the S&P 500 and NASDAQ datasets. These results were promising, i.e., the proposed method outperforms all baselines, even when the S&P 500 has a downward pressure, during which time the index decreases by 1.6% (from 2096.92 to 2063.36). In addition, Fig. 9 (c) and (d) present the Sharpe ratios of all the methods. Note that the Sharpe ratio is used to assess the volatility of portfolio returns. The higher the Sharpe ratio, the greater the return on investment relative to the same risk level. For the S&P 500 dataset, the Sharpe ratio of the proposed method was 1.64, higher than those of LSTM (0.16), GCN (0.47), eLSTM (0.72), HGNN_{Fin} (1.34), FinGAT (1.31), TGC (1.26), and AD-GAT (1.42). For the NASDAQ dataset, the proposed method also achieves the best performance (1.62). These results further demonstrate the superiority of the proposed method in terms of the trade-off between return and risk.

5.4. Effectiveness of the proposed approach

In this section, we examine the effectiveness of sub-functions of the proposed framework in addressing relational attributes for asset pricing. Specifically, we evaluate the performance improvement brought by building a heterogeneous graph (graph structure) and addressing conditional message-passing (graph learning).

5.4.1. Effectiveness of the graph structure

As aforementioned, multiple spillovers are salient for capturing stock movements, which is a challenge when dealing with relational attributes in graph learning. For this purpose, we argue that extending the homogeneous graph with the relational attributes to a heterogeneous graph is a promising way to preserve the connectivity of the relational attributes in the graph. 8 track experiments are carried out to examine the effectiveness of converting a homogeneous graph to a heterogeneous graph for capturing multiple spillover effects. Specifically,

- homo_F is a homogeneous graph, and each node represents a firm with fundamentals.
- homo_{F+N_s} is a homogeneous graph, and each firm node contains both fundamentals and news attributes. Here, news attributes are extracted in terms of sentiment words as the way in the work of Cheng and Li [7].
- homo_{F+N_b} is similar to homo_{F+N_s} method where news attributes are represented by embedding vectors via BERT instead of sentiment words.
- heter_{F+T_s} is a heterogeneous graph with both firm and theme nodes. Especially, the theme-theme edge is determined by its semantic similarity based on top-300 keywords [44].
- heter_{F+T_l} is a heterogeneous graph with both firm and theme nodes. Especially, the theme-theme edge is determined by the state-of-the-art semantic similarity method (LSTM-CNN) [45].
- heter_{F+T_c} is similar to heter_{F+T_s} method where the theme-theme edge is determined by the firms that co-existed in connected themes.
- heter_{F+A} is a heterogeneous graph with both firm and analyst nodes. Essentially, this graph can be seen as a subgraph of the proposed heterogeneous market graph.
- heter_{F+T_c+A} (the proposed approach) is a heterogeneous graph with all types of nodes (i.e., firms, themes, and analysts).

Table 3 presents comparison results. It can be observed that heterogeneous-based methods ($\text{heter}_{\#}$) outperformed the homogeneous-based methods ($\text{homo}_{\#}$). This finding consolidates our argument that relational attributes of the homogeneous graph should be transformed to graph nodes to form a heterogeneous graph, and thus preserve their relationships. For a heterogeneous graph, when incorporating richer spillovers, the proposed method (heter_{F+T_c+A}) performs better than heter_{F+T_c}, heter_{F+T_s} and heter_{F+A}. In addition, heter_{F+A} was superior to heter_{F+T_c} and heter_{F+T_s}. A good explanation is that analyst attributes can be directly converted into slow node variables, whereas the news attributes must be converted into relatively slow node variables via an intermediate layer mechanism, which inevitably results in the loss of some information.

For them-them edges, heter_{F+T_c} that connects the two themes in terms of firms, outperformed heter_{F+T_s} and heter_{F+T_l} that bridges these two themes based on their textual similarity. In this study, calculating the semantic similarity between news articles helped us track the amplification of media caused by similar news articles. However, there is also a media amplification effect caused by dissimilar news articles. For example, the media-aware stock movement caused by a news story about firm A's lawsuit can be enhanced by the release of a report about firm A's product recall. The contents of these two news articles were unrelated. However, we can observe that many firms involved in both types of news (lawsuit and product flaws) face downward pressure. Therefore, to capture the ripple and amplification effect of media on stock movements, theme links not only need to preserve the textual similarity of news, but also record co-occurring news with less textual

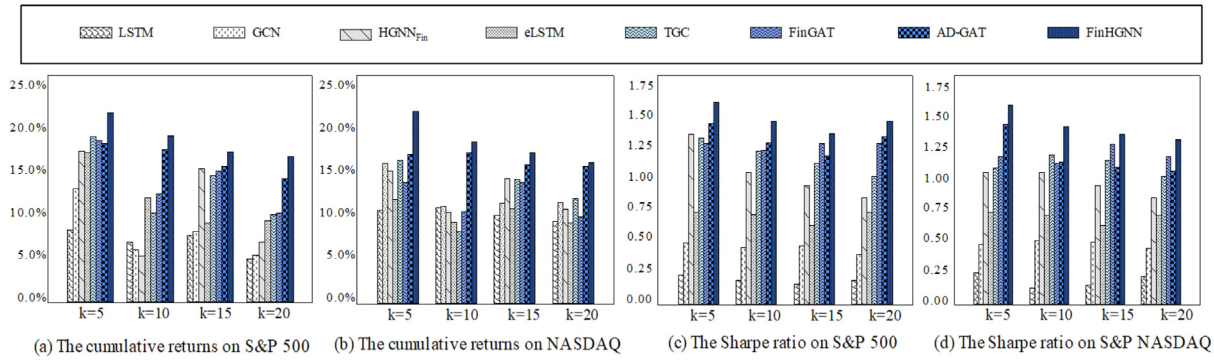


Fig. 9. The results of investment simulation.

Table 3
Comparison of different graph methods.

Method	S&P 500			NASDAQ		
	DA	AUC	MCC	DA	AUC	MCC
homo_F	0.5014	0.5171	0.0076	0.5008	0.5129	0.0070
homo_{F+N_s}	0.5251	0.5313	0.0577	0.5178	0.5221	0.0461
homo_{F+N_b}	0.5285	0.5196	0.0592	0.5244	0.5252	0.0504
heter_{F+T_s}	0.5436	0.5472	0.0889	0.5281	0.5309	0.0666
heter_{F+T_l}	0.5440	0.5503	0.0825	0.5236	0.5298	0.0493
heter_{F+T_c}	0.5476	0.5483	0.0907	0.5328	0.5390	0.0762
heter_{F+A}	0.5508	0.5582	0.1001	0.5379	0.5366	0.0794
heter_{F+T_c+A}	**0.5782**	**0.5881**	**0.1864**	**0.5665**	**0.5827**	**0.1624**

similarity. Because the semantic similarity of news articles has been captured by news clustering (theme modeling), the co-occurrence of news is given more importance when building theme-theme links.

For a homogeneous graph, it can be observed that homo_{F+N_b} outperformed homo_{F+N_s}, which shows that representing news with word embeddings is more efficient than the way of quantifying texts as sentiment words. Both methods that incorporate news information beat homo_F that contains only firm fundamentals. This shows that richer attributes also help improve predictive performance.

5.4.2. Effectiveness of graph learning

Different from traditional homogeneous graph network, heterogeneous graph network comprises multiple node types that requires a flexible message-passing mechanism to converge and transport information in graph learning.

Here, we investigate two unique features of the proposed conditional message-passing mechanism (Section 4.2) to understand its functionality in terms of the node types (FinHGNN_{type}) or the attributes (FinHGNN_{attr}). Specifically, the FinHGNN_{type} differentiated the messaging in terms of node types, while FinHGNN_{attr} transports information only in terms of node attributes without considering the interference from node types. Here, FinHGNN_{none} denotes message flows without considering the difference of node types and attributes. This is essentially the basic messaging mechanism in traditional graph learning.

In Table 4, it can be observed that FinHGNN_{attr} performed better than FinHGNN_{none}. This finding proves that the basic messaging mechanism in traditional graph learning is inappropriate for dealing with the attribute-sensitive momentum spillovers of listed firms in stock markets. This is because traditional graph learning based on linear aggregators transfers and aggregates peer influences without distinguishing different influences of connected firms' attributes.

In addition, Table 4 shows that FinHGNN_{type} performed better than FinHGNN_{attr}, indicating that the differentiation of messaging in heterogeneous graph networks relies more on node types than on node attributes. When considering both the node types and node attributes, performance of the proposed FinHGNN on the S&P 500 and NASDAQ datasets is further improved. This finding suggests that information transfer relying on type and attribute conditions should be addressed in heterogeneous graph networks.

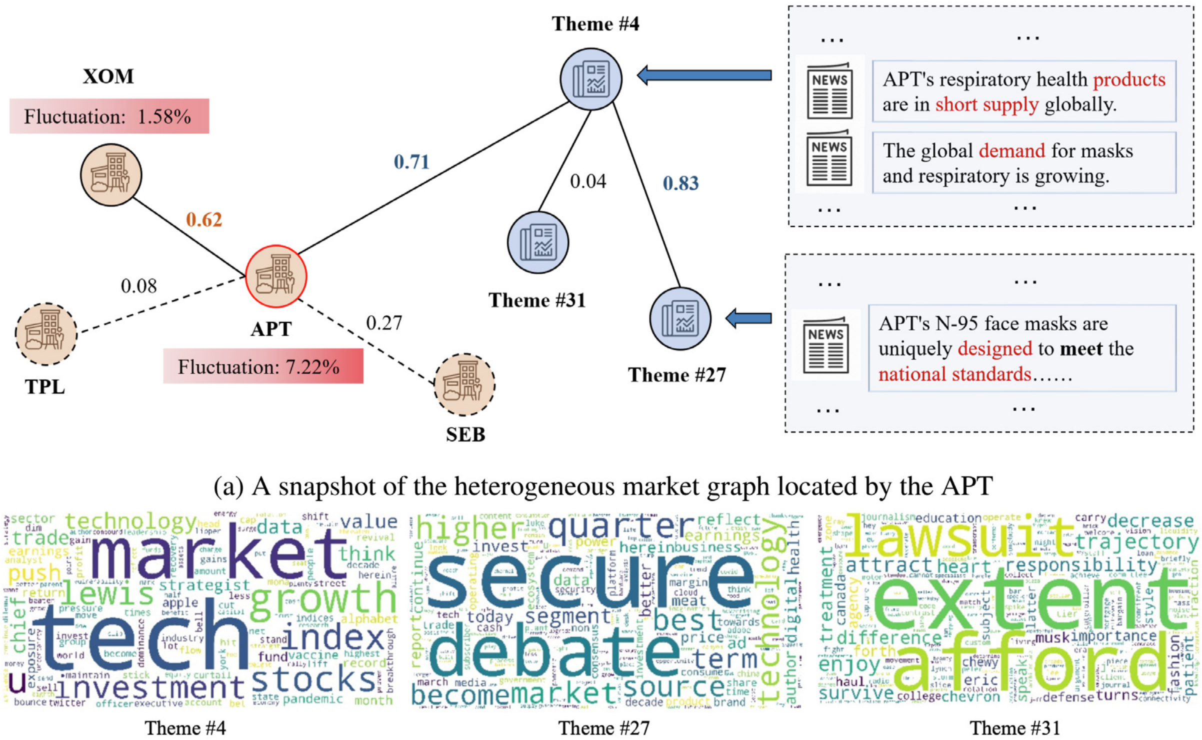
5.5. Case study

This section presents a case study for a better understanding of the proposed FinHGNN. Here, we take Alpha Pro Tech Ltd (APT) as our example. This company manufactures and markets high-value protective apparel garments, face masks, face shields, and a line of construction weatherization building products for the housing market.

Table 4
Comparison of different messaging mechanisms.

Method	S&P 500 DA	S&P 500 AUC	S&P 500 MCC	NASDAQ DA	NASDAQ AUC	NASDAQ MCC
FinHGNN$_{none}$	0.5211	0.5127	0.0060	0.5099	0.5043	0.0154
FinHGNN$_{attr}$	0.5461	0.5485	0.0826	0.5507	0.5515	0.1037
FinHGNN$_{type}$	0.5528	0.5673	0.1064	0.5575	0.5579	0.1158
FinHGNN	**0.5782**	**0.5881**	**0.1864**	**0.5665**	**0.5827**	**0.1624**

In Fig. 10, APT is linked with XOM (Exxon Mobil Corp) in terms of Eq. 11. At first glance, these two firms appear irrelevant. However, when we dig a little more, we find that XOM has mastered the melt-blown fabric technology, which is one of the key processes for mask and respirator materials. From this perspective, the downward or upward pressures of the XOM can cause stock fluctuations in the APT. For APT, during the period between January 3rd, 2020 and January 30th, 2020, most of influential news falls into theme 4. Theme 27 is one of the nearest neighbours of theme 4 (weighting 0.83), while theme 31 is far from theme 4 (weighting 0.04) in terms of Eq. 11. When we look into two news articles in theme 4, the contents of these two are related to a certain degree. Specially, one concerns the respiratory health products of APT being in short supply globally as the coronavirus spread. The other is the story that APT stock is on fire, as the global demand for masks and respirators increases. This fact suggests that the news articles partitioned into the same theme tend to have a similar impact on stock movements as they have similar content. In addition, we also show high-frequency word clouds for these three themes. It can be observed that themes 4 and 27 share the "tech". In addition, even though the words "growth" and "investment" in theme 4 and the words "debate", "secure", and "quarter" in theme 27 are different, the news about the APT's N-95 face masks are uniquely designed to meet the national recommended level of protection consisting of "meet" and "uniquely design" can enhance the market impact of the news with the words "demand" and "product". In contrast, words in themes 4 and 31 are difficult to connect. This means that both themes are considerably different with each other. Based on the above facts, we can understand that news articles within the same theme or close themes cause media spillovers in stock markets.



(a) A snapshot of the heterogeneous market graph located by the APT

(b) Word clouds of three different themes

Fig. 10. Case study of the proposed FinHGNN.

6. Conclusion

Heterogeneous graph neural networks provide an interesting probe of asset pricing when there are multiple spillovers. It essentially challenges the traditional method of handling relational node attributes. The findings of this study indicate that if a node attribute is relational in a homogeneous graph, it should be separated out and upgraded as a node to form a heterogeneous graph. This further clarifies the fact that information transfer in heterogeneous graphs should be conditional on both the node types and attributes. Additionally, many other domains, including social recommendations and financial derivative investments, follow this vein when relational attributes exist. More relevant studies are expected to be conducted in the near future.

CRediT authorship contribution statement

Jinghua Tan: Conceptualization, Methodology, Software, Writing - original draft, Visualization. **Qing Li:** Conceptualization, Methodology, Software, Visualization, Writing - original draft. **Jun Wang:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Funding acquisition. **Junxiao Chen:** Conceptualization, Methodology, Software, Data curation, Visualization.

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.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) (71671141, 71873108, 62072379), Fundamental Research Funds for Central Universities (JBK 1806003, JBK2202050, KJCX 20210103), Sichuan Province Science and Technology Department (2019YJ0250), the Fintech Innovation Center of Southwestern University of Finance and Economics, and Key Laboratory of Financial Intelligence and Financial Engineering of Sichuan Province.

Appendix A. Literature comparison in asset pricing

Table A.1.

Table A.1
Literature comparison in terms of the analysis model used.

		Analysis Model			Experiment			Focus	
Category	Literature	Predictor	Model	Response	Spillover	Scale	Metrics	Market	Period
Statistical and econometric models	[46]	Media Info.	Statistical model	Index trend	–	Week	Returns	DJIA	12/10/2007 –04/30/2012
	[17]	Media Info.	Linear model	Returns	–	Day	Standard errors, t-statistics	S&P 500	01/01/1980 –12/31/2004
ML-based models	[1]	Fundamentals, media Info.	SVR	Stock price	–	Minute	Accuracy, MSE	S&P 500	10/26/2005 –11/28/2005
	[2]	Fundamentals, media Info.	TeSIA	Stock price	–	Minute	Accuracy, RMSE	CSI 100	01/01/2011 –12/31/2011
	[3]	Fundamentals, media Info.	eLSTM	stock trend	Momentum spillover	Day	Accuracy, MCC	CSI 100	01/01/2016 –12/31/2016
	[18]	Media Info.	CNN	Index trend, stock trend	–	Day	Accuracy, MCC	S&P 500	10/02/2006 –11/21/2013
	[47]	DJIA Index values, media Info.	SOFNN	Index trend	–	Day	Accuracy, MAPE	DJIA	12/28/2008 –12/20/2008
	[6]	Fundamentals, stock sector relationships	TGC (homogeneous graph)	Return ratio	Momentum spillover	Day	MSE, MRR, IRR	NASDAQ, NYSE	01/02/2013 –12/08/2017

(continued on next page)

Table A.1 (*continued*)

Category	Literature	Analysis Model			Experiment			Focus	
		Predictor	Model	Response	Spillover	Scale	Metrics	Market	Period
	[7]	Fundamentals, media Info., firm relationships	AD-GAT (homogeneous graph)	Stock trend	Momentum spillover	Day	Accuracy, AUC	S&P 500	02/08/2011 –11/18/2013
	[9]	Fundamentals, stock sector relationships	FinGAT (homogeneous graph)	Return ratio	Momentum spillover	Day	Accuracy, precision, MRR	Taiwan, S&P 500, NASDAQ	965/965/ 1245 transaction days
Graph-based models	[24]	Media Info., firm relationships	LSTM-RGCN (homogeneous graph)	Index trend, stock trend	Momentum spillover	Day	Accuracy	TPX 500, TPX 100	01/01/2013 –09/28/2018
Ours	–	Fundamentals, media Info.	FinHGNN (heterogeneous graph)	Stock trend	Multiple spillovers	Day	Accuracy, AUC	S&P 500, NASDAQ	02/08/2011 –11/18/ 2013, 10/01/2019 –03/29/2021

Appendix B. Parameter settings of the benchmark models used for the comparison study

Here, we elaborates the parameter settings of the benchmark models used for the comparison study in Section 5.2. There are 2 non-graph-based methods (LSTM and eLSTM) and 5 graph-based methods (GCN, TGC, FinGAT, AD-GAT and HGNN_{Fin}). Specifically,

- LSTM: This is one of the most powerful deep learning models for time-series data and is widely used in stock prediction tasks. As suggested by [6,7,9], the LSTM with 2 layers is adopted and two hyperparameters are tuned, i.e., the length of sequential input, and the number of hidden units. Here, the length of sequential input is set to 20, and the number of hidden units is set to 64 by searching the grid in the range of $\{5, 10, 15, 20, 25, 30\}$ and $\{8, 16, 32, 64, 128, 256\}$, respectively.
- GCN: This aggregates the attributes of neighboring nodes to central nodes in a linear way based on the normalized connected relations. The GCN with 2 convolution layers (one is 64-dim and one is 32-dim) is implemented for a homogeneous firm graph. Both are obtained by grid searching within the range of $\{8, 16, 32, 64, 128, 256\}$.
- HGNN_{Fin}: This model is derived from a conventional HGNN, where relational attributes (news and analysts) are converted to graph nodes. The size of firm embedding (F), theme embedding (P) and analyst embedding (S) are the same to the transported embedding in the proposed FinHGNN. All of them are set to 32 after grid searching. The hidden size of HGNN are set to 64 by searching the grid in the range of $\{8, 16, 32, 64, 128, 256\}$. In addition, different spillover effects are separately transported in predefined meta-paths, as in the classical HGNN [29]. There are five pre-defined meta-paths, namely, Firm-Firm (F-F), Firm-Theme-Firm (F-T-F), Firm-Analyst-Firm (F-A-F), Firm-Theme-Theme-Firm (F-T-T-F) and Firm-Analyst-Analyst-Firm (F-A-A-F).
- eLSTM: Li et al. [3] proposed an enhanced LSTM model which is capable of handling extra shock from media and momentum spillovers in some degree. Here, the length of sequential input is set to 20, and the number of hidden units is set to 64 by searching the grid in the range of $\{5, 10, 15, 20, 25, 30\}$ and $\{8, 16, 32, 64, 128, 256\}$, respectively.
- TGC: Feng et al. [6] proposed a temporal graph convolution framework (TGC) to capture momentum spillover effects for predicting stock movements, which constructed a homogeneous market graph in terms of listed firms. The parameter setting of TGC is same with the GCN. It consists of 2 convolution layers, one is 64-dim and one is 32-dim. These values are set by searching within the range of $\{8, 16, 32, 64, 128, 256\}$.
- FinGAT: Hsu et al. [9] proposed the FinGAT model to estimate firm relations within each industry category for capturing momentum spillovers. Here, instead of setting the hidden layers to 16 that is adopted in the original work, we set all the hidden size of RNN and GAT to 32 by grid searching within the range of $\{8, 16, 32, 64, 128, 256\}$.
- AD-GAT: Cheng and Li [7] proposed a homogeneous graph method which designs an attribute-matter (AM) aggregation mechanism to incorporate momentum spillover effects from related firms for capturing stock movements. Specifically, the length of sequential input is set to 30, which is the same to the original AD-GAT. And the hidden size of RNN and GAT are all set to 32 by searching within the range of $\{8, 16, 32, 64, 128, 256\}$.

References

[1] R.P. Schumaker, H. Chen, Textual analysis of stock market prediction using breaking financial news: The AZFin text system, ACM Trans. Inform. Syst. 27 (2) (2009) 1–19.

[2] Q. Li, Y. Chen, L.L. Jiang, P. Li, H. Chen, A tensor-based information framework for predicting the stock market, ACM Trans. Inform. Syst. 34 (2) (2016) 1–11.
[3] Q. Li, J. Tan, J. Wang, H. Chen, A Multimodal Event-Driven LSTM Model for Stock Prediction Using Online News, IEEE Trans. Knowl. Data Eng. 33 (10) (2021) 3323–3337.
[4] T.J. Moskowitz, M. Grinblatt, Do Industries Explain Momentum?, J Finance 54 (4) (1999) 1249–1290.
[5] U. Ali, D.A. Hirshleifer, Shared analyst coverage: Unifying momentum spillover effects, J. Financ. Econ. 136 (3) (2020) 649–675.
[6] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, T.-S. Chua, Temporal relational ranking for stock prediction, ACM Trans. Inform. Syst. 2 (37) (2019) 1–30.
[7] R. Cheng, Q. Li, Modeling the momentum spillover effect for stock prediction via attribute-driven graph attention networks, in: Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI '21), 55–62, 2021.
[8] R. Sawhney, S. Agarwal, A. Wadhwa, R. Shah, Exploring the scale-free nature of stock markets: Hyperbolic graph learning for algorithmic trading, in: Proceedings of the Web Conference (WWW '21), 11–22, 2021.
[9] Y.-L. Hsu, Y.-C. Tsai, C.-T. Li, FinGAT: Financial graph attention networks for recommending top-k profitable stocks, IEEE Transactions on Knowledge and Data Engineering Doi:10.1109/TKDE.2021.3079496.
[10] S. Fan, J. Zhu, X. Han, C. Shi, L. Hu, B. Ma, Y. Li, Metapath-guided heterogeneous graph neural network for intent recommendation, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '19), 2478–2486, 2019.
[11] H. Ji, J. Zhu, X. Wang, C. Shi, B. Wang, X. Tan, Y. Li, S. He, Who you would like to share with? A study of share recommendation in social e-commerce, in: Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI '21), 232–239, 2021.
[12] Q. Li, T. Wang, P. Li, L. Liu, Q. Gong, Y. Chen, The effect of news and public mood on stock movements, Inf. Sci. 278 (2014) 826–840.
[13] P.C. Tetlock, Giving content to investor sentiment: The role of media in the stock market, J. Finance 62 (3) (2007) 1139–1168.
[14] B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, J. Zhang, Daily stock market forecast from textual web data, in: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2720–2725, 1998.
[15] I. Zheludev, R. Smith, T. Aste, When can social media lead financial markets?, Sci Rep. 4 (2014) 4213.
[16] E.F. Fama, K.R. French, Common risk factors in the returns on stocks and bonds, J. Financ. Econ. 33 (1) (1993) 3–56.
[17] P.C. Tetlock, M. Saar, Tsechansky, S. Macskassy, More than words: Quantifying language to measure firms' fundamentals, J. Finance 63 (3) (2008) 1437–1467.
[18] X. Ding, Y. Zhang, T. Liu, J. Duan, Deep learning for event-driven stock prediction, in: Proceedings of the 24th International Joint Conferences on Artificial Intelligence (IJCAI '15), 2327–2333, 2015.
[19] H. Wang, T. Wang, Y. Li, Incorporating expert-based investment opinion signals in stock prediction: A deep learning framework, in: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '20), 971–978, 2020.
[20] W. Xu, W. Liu, C. Xu, J. Bian, J. Yin, T.-Y. Liu, REST: Relational event-driven stock trend forecasting, in: Proceedings of the Web Conference (WWW '21), 1–10, 2021.
[21] C. Xu, H. Huang, X. Ying, J. Gao, Z. Li, P. Zhang, J. Xiao, J. Zhang, J. Luo, HGNN: Hierarchical graph neural network for predicting the classification of price-limit-hitting stocks, Inf. Sci. 607 (2022) 783–798.
[22] W. Chen, M. Jiang, W.-G. Zhang, Z. Chen, A novel graph convolutional feature based convolutional neural network for stock trend prediction, Inf. Sci. 556 (2021) 67–94.
[23] C. Xie, D. Rajan, Q. Chai, An interpretable Neural Fuzzy Hammerstein-Wiener network for stock price prediction, Inf. Sci. 577 (2021) 324–335.
[24] W. Li, R. Bao, K. Harimoto, D. Chen, J. Xu, Q. Su, Modeling the stock relation with graph network for overnight stock movement prediction, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI '20), 4541–4547, 2020.
[25] Y. Chen, Z. Wei, X. Huang, 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 (CIKM '18), 1655–1658, 2018.
[26] Y. Fang, W. Lin, V.W. Zheng, M. Wu, J. Shi, K.C.-C. Chang, X.-L. Li, Metagraph-based learning on heterogeneous graphs, IEEE Trans. Knowl. Data Eng. 33 (1) (2021) 154–168.
[27] Y. Zhang, Y. Xiong, X. Kong, S. Li, J. Mi, Y. Zhu, Deep collective classification in heterogeneous information networks, in: Proceedings of the Web Conference (WWW '18), 399–408, 2018.
[28] C. Zhang, D. Song, C. Huang, A. Swami, N.V. Chawla, Heterogeneous graph neural network, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '19), 793–803, 2019.
[29] R. Hussein, D. Yang, P. Cudré-Mauroux, Are meta-paths necessary? Revisiting heterogeneous graph embeddings, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18), 437–446, 2018.
[30] S. Yun, M. Jeong, R. Kim, J. Kang, H.J. Kim, Graph transformer networks, Adv. Neural Inform. Process. Syst. 32 (2019) 11983–11993.
[31] Y. Dong, N.V. Chawla, A. Swami, Metapath2vec: Scalable representation learning for heterogeneous networks, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17), 135–144, 2017.
[32] C. Shi, B. Hu, W.X. Zhao, P.S. Yu, Heterogeneous information network embedding for recommendation, IEEE Trans. Knowl. Data Eng. 31 (2) (2019) 357–370.
[33] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, P.S. Yu, Heterogeneous graph attention network, in: Proceedings of the Web Conference (WWW '19), 2022–2032, 2019.
[34] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[35] Y. Wu, D. Lian, S. Jin, E. Chen, Graph Convolutional Networks on User Mobility Heterogeneous Graphs for Social Relationship Inference, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI '19), 3898–3904, 2019.
[36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need ArXiv:1901.03838.
[37] H. Hong, H. Guo, Y. Lin, X. Yang, Z. Li, J. Ye, An Attention-based Graph Neural Network for Heterogeneous Structural Learning, in: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '20), 4, 55–62, 2020.
[38] R. Xing, Q. Li, J. Zhao, X. Xu, Media-based corporate network and its effects on stock market, Emerging Markets Finance and Trade 2 (2019) 1–26.
[39] G. Yu, Q. Li, J. Wang, D. Zhang, Y. Liu, A multimodal generative and fusion framework for recognizing faculty homepages, Inf. Sci. 525 (2020) 205–220.
[40] Q. Li, C. Yan, J. Wang, Y. Chen, H.C. Chen, Web media and stock markets: A survey and future directions from a big data perspective, IEEE Trans. Knowl. Data Eng. 30 (2) (2018) 381–399.
[41] H. Wang, S. Li, T. Wang, J. Zheng, Hierarchical adaptive temporal-relational modeling for stock trend prediction, in: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI '21), 3691–3698, 2021.
[42] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the 13rd International Conference on Artificial Intelligence and Statistics (AISTATS '10), 249–256, 2010.
[43] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Proceedings of the 3rd International Conference on Learning Representations (ICLR '15), arXiv:1901.03838, 2015.
[44] K. Xiao, Z. Qian, B. Qin, A graphical decomposition and similarity measurement approach for topic detection from online news, Inf. Sci. 570 (2021) 262–277.
[45] T. Nguyen, L. Nguyen Huy, T. Minh, T. Izuha Yamasaki, Sentence modeling via multiple word embeddings and multi-level comparison for semantic textual similarity, Inform. Process. Manage. 56 (6) (2019) 102090.
[46] H.S. Moat, C. Curme, A. Avakian, D.Y. Kenett, H.E. Stanley, T. Preis, Quantifying Wikipedia usage patterns before stock market moves, Sci. Rep. 3 (2013) 1801.
[47] J. Bollen, H. Mao, X. Zeng, Twitter Mood Predicts the Stock Market, J. Comput. Sci. 2 (1) (2011) 1–8.