

# Temporal Relational Ranking for Stock Prediction

FULI FENG, National University of Singapore, Singapore  
 XIANGNAN HE, University of Science and Technology of China, China  
 XIANG WANG, National University of Singapore, Singapore  
 CHENG LUO, Tsinghua University, China  
 YIQUN LIU, Tsinghua University, China  
 TAT-SENG CHUA, National University of Singapore, Singapore

Stock prediction aims to predict the future trends of a stock in order to help investors to make good investment decisions. Traditional solutions for stock prediction are based on time-series models. With the recent success of deep neural networks in modeling sequential data, deep learning has become a promising choice for stock prediction.

However, most existing deep learning solutions are not optimized towards the target of investment, *i.e.*, selecting the best stock with highest expected revenue. Specifically, they typically formulate stock prediction as a classification (to predict stock trend) or a regression problem (to predict stock price). More importantly, they largely treat the stocks as independent of each other. The valuable signal in the rich relations between stocks (or companies), such as two stocks are in the same sector and two companies have a supplier-customer relation, is not considered.

In this work, we contribute a new deep learning solution, named *Relational Stock Ranking* (RSR), for stock prediction. Our RSR method advances existing solutions in two major aspects: 1) tailoring the deep learning models for stock ranking, and 2) capturing the stock relations in a time-sensitive manner. The key novelty of our work is the proposal of a new component in neural network modeling, named *Temporal Graph Convolution*, which jointly models the temporal evolution and relation network of stocks. To validate our method, we perform back-testing on the historical data of two stock markets, NYSE and NASDAQ. Extensive experiments demonstrate the superiority of our RSR method. It outperforms state-of-the-art stock prediction solutions achieving an average return ratio of 98% and 71% on NYSE and NASDAQ, respectively.

CCS Concepts: • **Information systems** → **Data mining**; • **Computing methodologies** → **Neural networks**; *Machine learning*; *Logical and relational learning*; • **Applied computing** → **Computers in other domains**;

Additional Key Words and Phrases: Stock Prediction, Learning to Rank, Graph-based Learning

## 1 INTRODUCTION

According to the statistics reported by the World Bank in 2017, the overall capitalization of stock markets worldwide has exceeded 64 trillion U.S. dollars<sup>1</sup>. With the continual increase in stock

<sup>1</sup><https://data.worldbank.org/indicator/CM.MKT.LCAP.CD/>.

---

Authors' addresses: Fuli Feng, National University of Singapore, 13 Computing Drive, 117417, Singapore, fulifeng93@gmail.com; Xiangnan He, University of Science and Technology of China, 443 Huangshan Road, Hefei, 230031, China, xiangnanhe@gmail.com; Xiang Wang, National University of Singapore, 13 Computing Drive, 117417, Singapore, xiangwang@u.nus.edu; Cheng Luo, Tsinghua University, 30 Shuangqing Rd, Haidian, Beijing, China, chengluc@tsinghua.edu.cn; Yiqun Liu, Tsinghua University, 30 Shuangqing Rd, Haidian, Beijing, China, yiqunliu@tsinghua.edu.cn; Tat-Seng Chua, National University of Singapore, 13 Computing Drive, 117417, Singapore, dcscts@nus.edu.sg.

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2019 Copyright held by the owner/author(s).

1046-8188/2019/12-ART

[https://doi.org/10.475/123\\_4](https://doi.org/10.475/123_4)

Table 1. An intuitive example that one method predicting the price change of stocks more accurately (i.e., smaller MSE) leads to a less profitable stock selection (i.e., smaller profit). Method 1 selects stock A (30) while Method 2 selects stock B (10).

| Ground Truth |     |     | Method 1   |     |     |             |        |  | Method 2   |            |     |             |        |  |
|--------------|-----|-----|------------|-----|-----|-------------|--------|--|------------|------------|-----|-------------|--------|--|
|              |     |     | Prediction |     |     | Performance |        |  | Prediction |            |     | Performance |        |  |
| A            | B   | C   | A          | B   | C   | MSE         | Profit |  | A          | B          | C   | MSE         | Profit |  |
| +30          | +10 | -50 | <b>+50</b> | -10 | -50 | 266         | 30     |  | +20        | <b>+30</b> | -40 | 200         | 10     |  |

A, B, C denote three stocks; numbers (+20) are the true/predicted price change of stocks; values in bold correspond to suggested selections.

market capitalization, trading stocks has become an attractive investment instrument for many investors. However, whether an investor could earn or lose money depends heavily on whether he/she can make the right stock selection. Stock prediction, which aims to predict the future trend and price of stocks, is one of the most popular techniques to make profitable stock investment [32], although there are still debates about whether the stock market is predictable (*aka.* the Efficient Markets Hypothesis) among financial economists [24, 27]. Some recent evidences indicate the predictability of stock markets, which motivates further exploration of stock prediction techniques [17, 23, 36, 42? ].

Traditional solutions for stock prediction are based on time-series analysis models, such as Kalman Filters [39], Autoregressive Models and their extensions [1]. Given an indicator of a stock (*e.g.*, stock price), this kind of models represents it as a stochastic process and takes the historical data of the indicator to fit the process. We argue that such mainstream solutions for stock prediction have three main drawbacks: 1) The models heavily rely on the selection of indicators, which is usually done manually and is hard to optimize without special knowledge of finance. 2) The hypothesized stochastic processes are not always compatible with the volatile stock in the real world. 3) These models can only consider a few indicators since their inference complexity typically increases exponentially with the number of indicators. As such, they lack the capability to comprehensively describe a stock that could be influenced by a plethora of factors. Towards these drawbacks, advanced techniques like deep neural networks, especially the recurrent neural networks (RNNs), have become a promising solution to substitute the traditional time-series models to predict the future trend or exact price of a stock [4, 42–44].

A state-of-the-art neural network-based solution is the State Frequency Memory (SFM) network [42], which models the historical data in a recurrent fashion and captures temporal patterns in different frequencies. This method achieves promising performance of predicting the daily opening price of fifty U.S. stocks one day ahead with a mean square error (MSE) of less than six dollars. However, we argue that such prediction methods are suboptimal to guide stock selection, since their optimization target is not at selecting the top stocks with the highest expected revenue. To be specific, they typically address stock prediction as either a classification (on price movement direction) or a regression (on price value) task, which would cause a large discrepancy on the investment revenue. Table 1 gives an intuitive example, where a method with better prediction performance (measured by regression MSE) suggests a less profitable stock. This implies the possible discrepancy between the actual target of stock selection and the optimized target of regression (classification), such that an optimal method of regression (classification) does not necessarily select the optimal stock to trade.

Another limitation of existing neural network-based solutions is that they typically treat stocks as independent of each other and ignore the relations between stocks. However, the rich relations between stocks and the corresponding companies may contain valuable clues for stock prediction. For example, stocks under the same sector or industry like GOOGL (Alphabet Inc.) and FB (Facebook

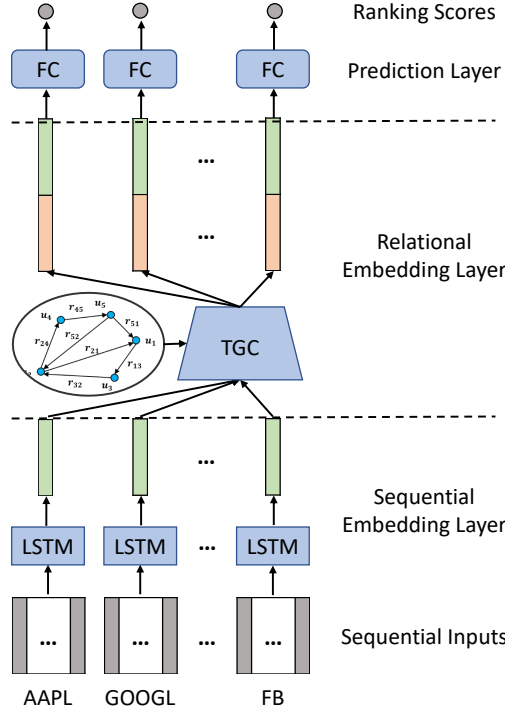


Fig. 1. Relational stock ranking framework. It should be noted that the LSTM cells and FC units (Fully Connected layer) depicted in the same layer share the same parameters.

Inc.) might have similar long-term trends. Besides, the stock of a supplier company might impact the stock of its consumer companies especially when a scandal of the supplier company is reported, such as the falsification of product quality data. To integrate stock relations into prediction, an intuitive solution is to represent the stock relations as a graph and then regularize the prediction of stocks based on the graph (*i.e.*, graph-based learning) [11, 18, 20, 30]. However, conventional graph learning techniques cannot capture the temporal evolution property of stock markets (*e.g.*, the strength of influence between two given stocks may vary quickly), since the graph is fixed at a particular time.

To address the aforementioned limitations of existing solutions, we formulate stock prediction as a ranking task, for which the target is to directly predict a stock list ranked by a desired criteria like return ratio. We then propose an end-to-end framework, named *Relational Stock Ranking* (RSR), to solve the stock ranking problem. An illustration of our framework can be found in Figure 1. Specifically, we first feed the historical time series data of each stock to a Long Short-Term Memory (LSTM) network to capture the sequential dependencies and learn a stock-wise sequential embedding. By devising a new *Temporal Graph Convolution* (TGC), we next revise the sequential embeddings by accounting for stock relations in a time-sensitive way. Finally, we feed the concatenation of sequential embeddings and relational embeddings to a fully connected layer to obtain the ranking score of stocks. To justify our proposed method, we employ it on two real-world markets, New York Stock Exchange (NYSE) and NASDAQ Stock Market (NASDAQ). Extensive back-testing results demonstrate that our RSR significantly outperforms SFM [42] with more than 115% improvements in return ratio.

The key contributions of the paper are summarized as follows.

- We propose a novel neural network-based framework, named *Relational Stock Ranking*, to solve the stock prediction problem in a learning-to-rank fashion.
- We devise a new component in neural network modeling, named *Temporal Graph Convolution*, to explicitly capture the domain knowledge of stock relations in a time-sensitive manner.
- We empirically demonstrate the effectiveness of our proposals on two real-world stock markets, NYSE and NASDAQ.

The remainder of this paper is organized as follows. Section 2 introduces the preliminary knowledge about LSTM and graph-based learning, which forms the building blocks of our method. Section 3 presents our proposed RSR. Section 4 and 5 describe the datasets and experiment, respectively. In Section 6, we review related work, followed by conclusion in Section 7.

## 2 PRELIMINARIES

In this paper, we use bold capital letters (e.g.,  $\mathbf{X}$ ), bold lowercase letters (e.g.,  $\mathbf{x}$ ), and capital script letters (e.g.,  $\mathcal{X}$ ) to denote matrices, vectors, and tensors, respectively. Scalars and hyperparameters are respectively represented as normal lowercase letters (e.g.,  $x$ ) and Greek letters (e.g.,  $\lambda$ ). If not otherwise specified, all vectors are in a column form, and  $X_{ij}$  denotes the entry at the  $i$ -th row and the  $j$ -th column of  $\mathbf{X}$ . The symbols  $\sigma$ ,  $\tanh$ , and  $\odot$  stand for the *sigmoid* function, hyperbolic tangent function, and element-wise production operation, respectively.

### 2.1 Long Short-Term Memory

LSTM [16] networks have been widely used to process sequential data, such as natural language [40], voice [13], and video [35]. LSTM is a special kind of Recurrent Neural Networks (RNNs) [12] that evolve hidden states through time to capture the sequential pattern of input data, e.g., the dependency between words in a sentence. Compared to the vanilla RNN, which is known to suffer from vanishing gradients while trained with Back-Propagation Through Time (BPTT), LSTM adds cell states to store the long-term memory and capture the long-term dependency in a sequence<sup>2</sup>.

Before providing the specific formulation of LSTM, we first describe the terms associated with LSTM. At each time-step  $t$ ,  $\mathbf{x}^t \in \mathbb{R}^D$  denotes an input vector (e.g., embedding vector of the  $t$ -th word in a given sentence), where  $D$  is the input dimension. Vectors  $\mathbf{c}^t$  and  $\mathbf{h}^t \in \mathbb{R}^U$  denote the cell (memory) state vector and the hidden state vector, respectively, where  $U$  is the number of hidden units. Vector  $\mathbf{z}^t \in \mathbb{R}^U$  is an information transformation module. Vectors  $\mathbf{i}^t$ ,  $\mathbf{o}^t$ , and  $\mathbf{f}^t \in \mathbb{R}^U$  denote the input, output, and forget gate, respectively. Formally, the transformation module, state vectors, and controlling gates are defined via the following equations:

$$\begin{aligned}
 \mathbf{z}^t &= \tanh(\mathbf{W}_z \mathbf{x}^t + \mathbf{Q}_z \mathbf{h}^{t-1} + \mathbf{b}_z) \\
 \mathbf{i}^t &= \sigma(\mathbf{W}_i \mathbf{x}^t + \mathbf{Q}_i \mathbf{h}^{t-1} + \mathbf{b}_i) \\
 \mathbf{f}^t &= \sigma(\mathbf{W}_f \mathbf{x}^t + \mathbf{Q}_f \mathbf{h}^{t-1} + \mathbf{b}_f) \\
 \mathbf{c}^t &= \mathbf{f}^t \odot \mathbf{c}^{t-1} + \mathbf{i}^t \odot \mathbf{z}^t \\
 \mathbf{o}^t &= \sigma(\mathbf{W}_o \mathbf{x}^t + \mathbf{W}_h \mathbf{h}^{t-1} + \mathbf{b}_o) \\
 \mathbf{h}^t &= \mathbf{o}^t \odot \tanh(\mathbf{c}^t),
 \end{aligned} \tag{1}$$

where  $\mathbf{W}_z, \mathbf{W}_i, \mathbf{W}_f, \mathbf{W}_o \in \mathbb{R}^{U \times D}$ , and  $\mathbf{Q}_z, \mathbf{Q}_i, \mathbf{Q}_f \in \mathbb{R}^{U \times U}$  are mapping matrices;  $\mathbf{b}_z, \mathbf{b}_i, \mathbf{b}_f$ , and  $\mathbf{b}_o \in \mathbb{R}^U$  are bias vectors. The updating formulation can be understood as performing the following procedures: (1) calculate the information to be transformed from the input  $\mathbf{x}^t$  to the memory states

<sup>2</sup>Detailed illustration of LSTM and its comparison against vanilla RNN are referred to: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.

$c^t$  by updating  $z^t$ ; (2) update the input gate  $i^t$  to control the information from  $z^t$  to  $c^t$ ; (3) update the forget gate  $f^t$  to decide how much information should be kept in the memory state; (4) refresh the memory state  $c^t$  by fusing the information flows from the input gate and memory gate; (5) update the output gate  $o^t$  to regulate the amount of information that can be outputted; (6) update the hidden state  $h^t$ . As can be seen, the memory state  $h^t$  only has linear adding interactions, which allows the information to be unchanged during the BPTT. Benefited by the memory state, LSTM is capable of capturing the long-term dependency in the sequential data.

## 2.2 Graph-based Learning

Graph-based learning has been applied to various machine learning tasks to utilize entity relations [2, 11, 26, 41? ]. The general problem setting is to learn a prediction function  $\hat{y} = f(\mathbf{x})$ , which maps an entity from the feature space to the target label space. It is usually achieved by minimizing an objective function abstracted as:

$$\Gamma = \Omega + \lambda\Phi, \quad (2)$$

where  $\Omega$  is a task-specific loss that measures the error between prediction  $\hat{y}$  and ground-truth  $y$ ,  $\Phi$  is a graph regularization term that smooths the prediction over the graph, and  $\lambda$  is a hyperparameter to balance the two terms. The regularization term typically implements the *smoothness* assumption that similar vertices tend to have similar predictions. A widely used  $\Phi$  is defined as:

$$\Phi = \sum_{i=1}^N \sum_{j=1}^N \underbrace{g(\mathbf{x}_i, \mathbf{x}_j)}_{\text{strength of smoothness}} \underbrace{\left\| \frac{f(\mathbf{x}_i)}{\sqrt{D_{ii}}} - \frac{f(\mathbf{x}_j)}{\sqrt{D_{jj}}} \right\|^2}_{\text{smoothness}}, \quad (3)$$

where  $g(\mathbf{x}_i, \mathbf{x}_j)$  is the similarity between the feature vectors of an entity pair (e.g., the edge weight between the corresponding vertices);  $D_{ii} = \sum_{j=1}^N g(\mathbf{x}_i, \mathbf{x}_j)$  is the degree of vertex  $i$ . The regularization term operates smoothness on each pair of entities, enforcing their predictions (after normalized by their degrees) to be close to each other. The strength of smoothness is determined by the similarity over their feature vectors  $g(\mathbf{x}_i, \mathbf{x}_j)$ . It can be equivalently written in a more concise matrix form:

$$\mathcal{G} = \text{trace}(\hat{\mathbf{Y}}\mathbf{L}\hat{\mathbf{Y}}^T), \quad (4)$$

where  $\hat{\mathbf{Y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N]$ ,  $\mathbf{L}$  is defined as  $\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{A})\mathbf{D}^{-1/2}$ , also known as the *graph Laplacian matrix*, and each element of  $\mathbf{A}$  is  $A_{ij} = g(\mathbf{x}_i, \mathbf{x}_j)$ .

**2.2.1 Graph Convolutional Networks.** Graph Convolutional Network (GCN) is a special kind of graph-based learning methods, which integrates the core idea of graph-based learning (i.e., the *smoothness* assumption over graphs) with advanced convolutional neural networks (CNNs) [8, 10, 14, 20]. The core idea of standard CNNs [21] is using convolutions (e.g.,  $3 \times 3$  filter matrices) to capture the local patterns in input data (e.g., oblique lines in an image). Following the idea of CNNs, the aim of GCN is to capture the local connection patterns on graphs. However, intuitive solutions like directly applying convolution operations on the adjacency matrix of a graph are not feasible. Because the filtering output of convolutions might change when we switch two rows of the adjacency matrix, while the switched adjacency matrix still represent the same graph structure. An alternative solution is to use spectral convolutions to capture the local connections in the Fourier domain, such as:

$$f(\mathbf{F}, \mathbf{X}) = \mathbf{U}\mathbf{F}\mathbf{U}^T\mathbf{X}, \quad (5)$$

where  $f$  denotes the filtering operation of a convolution parameterized by a diagonal matrix  $\mathbf{F}$ , and  $\mathbf{U}$  is the eigenvector matrix of the graph Laplacian matrix, i.e.,  $\mathbf{L} = \mathbf{U}\mathbf{A}\mathbf{U}^T$ .

Table 2. **Terms and notations.**

| Symbol  | Definition   |
|---|--|
| $\mathcal{X}^t \in \mathbb{R}^{N \times S \times D} = [\mathbf{X}_1^t, \dots, \mathbf{X}_N^t]^T$  | historical prices of $N$ stocks on trading day $t$ .                               |
| $\mathcal{A} \in \mathbb{R}^{N \times N \times K}$  | binary encoding of stock relations.  |
| $\mathbf{E}^t = [\mathbf{e}_1^t, \dots, \mathbf{e}_N^t]^T \in \mathbb{R}^{N \times U}$            | sequential embedding of $N$ stocks learned from historical prices.                 |
| $\overline{\mathbf{E}}^t = [\mathbf{e}_1^t, \dots, \mathbf{e}_N^t]^T \in \mathbb{R}^{N \times U}$ | relational embedding of all stocks learned from $\mathbf{E}^t$ and $\mathcal{A}$ . |
| $\mathbf{r}^{t+1}, \hat{\mathbf{r}}^{t+1} \in \mathbb{R}^N$                                       | ground-truth and predicted ranking scores of $N$ stocks.                           |
| $\mathbf{w}, b$   | weights and bias to be learned.  |

Suffering from the overhead of computing the eigendecomposition of  $\mathbf{L}$ , it is suggested to treat  $\mathbf{F}$  as a function of  $\mathbf{A}$ . Then it can be approximated by the Chebyshev polynomials  $T_k(x)$  of up to the  $K$ -th order,

$$\mathbf{F} \approx \sum_{k=0}^K \theta_k T_k(\hat{\Lambda}), \quad (6)$$

where  $\hat{\Lambda} = \frac{2}{\lambda_{max}} \mathbf{A} - \mathbf{I}$  with  $\lambda_{max}$  denotes the largest eigenvalue of  $\mathbf{L}$ ;  $\theta_k$  represents the Chebyshev coefficient;  $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$  with  $T_1x = x$  and  $T_0x = 0$ . In [20], the authors proved that the GCN performed well enough while setting  $K$  to 1. As such, they reduced Equation (5) to  $f(\mathbf{F}, \mathbf{X}) = \mathbf{A}\mathbf{X}$  and injected the convolution into a fully connected layer as  $\mathbf{A}(\mathbf{X}\mathbf{W} + \mathbf{b})$ , which is the state-of-the-art formulation of GCN<sup>3</sup>.

### 3 RELATIONAL STOCK RANKING

The typical problem setting of stock prediction (*i.e.*, price movement classification and price regression) is to learn a prediction function  $\hat{y}^{t+1} = f(\mathbf{X}^t)$  which maps a stock from the feature space to the target label space at time-step  $t$ . Matrix  $\mathbf{X}^t = [\mathbf{x}^{t-S+1}, \dots, \mathbf{x}^t]^T \in \mathbb{R}^{S \times D}$  represents the sequential input features, where  $D$  is the dimension of features at each time-step and  $S$  is the length of the sequence. Distinct from the typical problem setting of stock prediction, which treats different stocks as independent sequences, our target is to learn a ranking function  $\hat{\mathbf{r}}^{t+1} = f(\mathcal{X}^t)$ , which simultaneously maps a bunch of stocks to a ranking list. In the learned ranking list, stocks with higher ranking scores are expected to achieve higher investment revenue at time-step  $t + 1$ . Assuming we have  $N$  stocks, then  $\mathcal{X}^t \in \mathbb{R}^{N \times S \times D} = [\mathbf{X}_1^t, \dots, \mathbf{X}_N^t]^T$  is the collected features. In addition, we further associate the problem with a set of explicit stock relations (*e.g.*, supplier-consumer relations), which reflect the potential influence between different stocks. Given  $K$  types of relations, we encode the pairwise relation between two stocks as a multi-hot binary vector  $\mathbf{a}_{ij} \in \mathbb{R}^K$  and represent the relation of all stocks as a tensor  $\mathcal{A} \in \mathbb{R}^{N \times N \times K}$ , of which the entry at the  $i$ -th row and  $j$ -th column is  $\mathbf{a}_{ij}$ .

In what follows, we first present the overall solution. We then elaborate our proposed *Temporal Graph Convolution* for handling stock relations, followed by discussing its connections to existing graph-based learning methods. In Table 2, we summarize some of the terms and notations.

#### 3.1 Framework

As illustrated in Figure 1, RSR contains three layers, named a sequential embedding layer, a relational embedding layer, and a prediction layer, which are elaborated as follows.

**Sequential Embedding Layer.** Considering the strong temporal dynamics of stock markets, it is intuitive to regard the historical status of a stock as the most influential factor to predict its future trend. As such, we first apply a sequential embedding layer to capture the sequential dependencies

<sup>3</sup>Note that in the reduced form of GCN, the input diagonal matrix  $\mathbf{F}$  is omitted due to the Chebyshev approximation.



in the historical data. Since RNN has achieved significant performance to process sequential data [13, 35, 40] and demonstrated to be effective in recent stock prediction research [4, 42], we opt for RNN to learn the sequential embeddings. Among various RNN models, such as vanilla RNN, LSTM, and Gated Recurrent Unit (GRU) [7], we choose LSTM owing to its ability to capture long-term dependency, which is of great importance to stock prediction. This is because that many factors have long-term effects on a stock, such as the rise of interest rates, the release of annual reports, a rapid drop in its price, among others. For example, if a stock has experienced a very rapid drop in its price, after that, the stock's price tends to exhibit an upward trend in the following days or weeks (*aka.* the mean reversion phenomenon). As such, we feed the historical time series data of stock  $i$  at time-step  $t$  ( $X_i^t$ ) to a LSTM network and take the last hidden state ( $h_i^t$ ) as the sequential embedding ( $e_i^t$ ) of a stock (note that  $e_i^t = h_i^t$ ), *i.e.*, we have,

$$E^t = LSTM(X^t), \quad (7)$$

where  $E^t = [e_1^t, \dots, e_N^t]^T \in \mathbb{R}^{N \times U}$  denotes the sequential embeddings of all stocks, and  $U$  denotes the embedding size (*i.e.*,  $U$  is the number of hidden units in LSTM).

**Relational Embedding Layer.** We now consider how to model the influence between different stocks, especially the ones with explicit relations. Note that it can be seen as an *injection of explicit domain knowledge* (*i.e.*, stock relations) into the data-driven approach for sequential embedding learning. Here we give two cases for illustration:

- If two companies are in the same sector or industry, they may exhibit similar trends in their stock prices, since they tend to be influenced by similar external events. Figure 2(a) shows two example stocks, MSFT (Microsoft Inc.) and GOOGL (Alphabet Inc.), both of which are in the same sector (Technology) and industry (Computer Software)<sup>4</sup>. As can be seen in Figure 2(a), the two stocks exhibit quite similar trends in terms of the change on price in 2017. Note that we normalize the prices of each stock separately by calculating the increase ratio at each trading day according to the price on the first day to reflect the price changes.
- If two companies are partners in a supply chain, then the events of the upstream company may affect the stock price of the downstream company. Figure 2(b) shows an example to demonstrate the impact of such supplier-consumer relation, which shows the stock price change of Lens Technology Co Ltd *after the release of iPhone 8 (09/22/2017)*<sup>5</sup>. Since the Lens Technology Co Ltd is the supplier of the screen of iPhone, which was expected to be selling well, its stock price kept increasing in the following several weeks of 09/22/2017.

To capture such patterns in stock historical data, we devise a new component of neural network modeling, named *Temporal Graph Convolution* to revise the sequential embeddings according to stock relations. It generates the relational embeddings  $\bar{E}^t \in \mathbb{R}^{N \times U}$  in a time-sensitive (dynamic) way, which is a key technical contribution of this work and will be elaborated later in Section 3.2.

**Prediction Layer.** Lastly, we feed the sequential embeddings and revised relational embeddings to a fully connected layer to predict the ranking score of each stock; the ranked list of stocks recommended to buy is then generated based on the prediction scores.

To optimize the model, we propose an objective function that combines both pointwise regression loss and pairwise ranking-aware loss:

$$l(\hat{\mathbf{r}}^{t+1}, \mathbf{r}^{t+1}) = \|\hat{\mathbf{r}}^{t+1} - \mathbf{r}^{t+1}\|^2 + \alpha \sum_{i=0}^N \sum_{j=0}^N \max(0, -(\hat{r}_i^{t+1} - \hat{r}_j^{t+1})(r_i^{t+1} - r_j^{t+1})), \quad (8)$$

<sup>4</sup><http://www.nasdaq.com/screening/companies-by-industry.aspx>

<sup>5</sup><https://www.techradar.com/reviews/iphone-8-review>

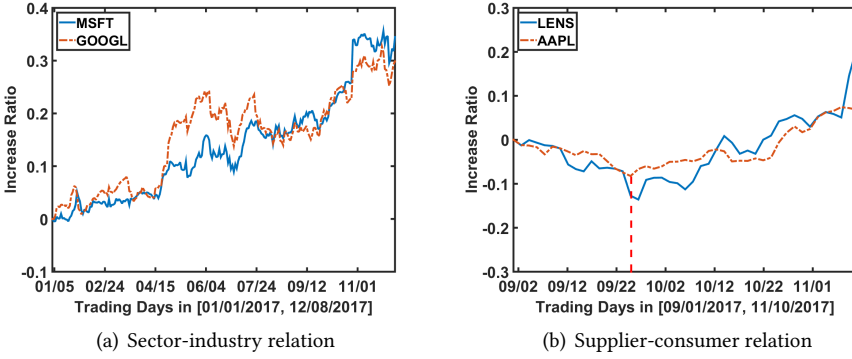


Fig. 2. Two examples of stock price history (normalized as increase ratio as compared to the first depicted trading day) to illustrate the impact of company relations on the stock price.

where  $\mathbf{r}^{t+1} = [r_1^{t+1}, \dots, r_N^{t+1}]$  and  $\hat{\mathbf{r}}^{t+1} = [\hat{r}_1^{t+1}, \dots, \hat{r}_N^{t+1}] \in \mathbb{R}^N$  are ground-truth and predicted ranking scores, respectively, and  $\alpha$  is a hyperparameter to balance the two loss terms. Since we focus on identifying the most profitable stock to trade, we use the 1-day return ratio of a stock as the ground-truth rather than the normalized price used in previous work [42]. We will provide more details on computing the ground-truth in Section 4.1 in our data collection.

The first regression term punishes the difference between the scores of ground-truth and prediction. The second term is pair-wise max-margin loss [45], which encourages the predicted ranking scores of a stock pair to have the same relative order as the ground-truth. The similar max-margin loss has been used in several applications such as recommendation [38] and knowledge based completion [34] and demonstrated good performance in ranking tasks. Minimizing our proposed combined loss will force the prediction ranking scores to be close to both 1) the return ratios of stocks in terms of absolute values, and 2) the relative orders of return ratios among stocks, so as to facilitate investors making better investment decisions. On one hand, correct relative order of stocks could help to select the investment targets (i.e., the top ranked stocks). On the other hand, the accurate prediction of return ratio would facilitate deciding the timing of investment since the top ranked stocks are valuable targets only when the return ratios would largely increase.

### 3.2 Temporal Graph Convolution

Given  $N$  stocks with their sequential embeddings  $\mathbf{E}^t \in \mathbb{R}^{N \times U}$  (i.e., the output of sequential embedding layer) and their multi-hot binary relation encodings  $\mathcal{A} \in \mathbb{R}^{N \times N \times K}$ , the aim of *Temporal Graph Convolution* is to learn revised embeddings  $\bar{\mathbf{E}}^t \in \mathbb{R}^{N \times U}$  that encode the relation information. Instead of directly presenting the formulation of TGC, we detail how we design the component to shed some lights on its rationale. Lastly, we discuss its connection with existing graph-based learning methods.

**a) Uniform Embedding Propagation.** Our first inspiration comes from the link analysis research, where in a graph, the impact of a vertex on another one can be captured by propagating information on the graph. A well-known example is the PageRank [31] method that propagates the importance score of a vertex to its connected vertices. Since a stock relation encodes certain similarity information between two connected stocks, we consider relating their embeddings through the similar



propagation process as in link analysis:

$$\bar{\mathbf{e}}_i^t = \sum_{\{j | \text{sum}(\mathbf{a}_{ji}) > 0\}} \frac{1}{d_j} \mathbf{e}_j^t, \quad (9)$$

where  $\text{sum}(\mathbf{a}_{ji})$  is the sum of all elements in the relation vector  $\mathbf{a}_{ji}$  (recall that  $\mathbf{a}_{ji}$  is a multi-hot binary vector where each element denotes whether the corresponding **type of** relation exists between  $j$  and  $i$ ). The condition  $\text{sum}(\mathbf{a}_{ji}) > 0$  ensures that only stocks have at least one relation will be considered.  $d_j$  is the number of stocks satisfying the condition  $\text{sum}(\mathbf{a}_{ji}) > 0$ . After such a propagation in the embedding space, the relational embedding  $\bar{\mathbf{e}}_i^t$  encodes the impacts coming from other stocks that have relations with stock  $i$  at time  $t$ .

**b) Weighted Embedding Propagation.** Consider that different relations between two stocks may have varying impacts on their prices, we apply a non-uniform coefficient when propagating the embeddings:

$$\bar{\mathbf{e}}_i^t = \sum_{\{j | \text{sum}(\mathbf{a}_{ji}) > 0\}} \frac{g(\mathbf{a}_{ji})}{d_j} \mathbf{e}_j^t, \quad (10)$$

where  $g(\mathbf{a}_{ji})$  is a mapping function that aims to learn the impact strength of the relations in  $\mathbf{a}_{ji}$ , and we term it as the *relation-strength function*. As an example, suppose we have two relations named *supplier\_customer* and *same\_industry*, and three stocks  $j$ ,  $i$ , and  $k$ . Given that stock  $j$  is a supplier of stock  $i$  while stock  $k$  is in the same industry as stock  $i$ , we can encode their relations as two different vectors:  $\mathbf{a}_{ji} = [1, 0]$  and  $\mathbf{a}_{ki} = [0, 1]$ . We can see that by feeding different relation vectors into a learnable relation-strength function for different stock pairs, we allow the embedding propagation process to account for both the topology of relation graph and the semantics of relations.

**c) Time-aware Embedding Propagation.** A limitation of the above weighted propagation process is that the *relation-strength function* returns a fixed weight for a given relation vector  $\mathbf{a}_{ji}$  regardless the evolution across different time-steps. As stock market is highly dynamic such that the status of a stock and the strength of a relation are continuously evolving, assuming a relation vector to have a static weight limits the modeling fidelity. For instance, in the previous example of Figure 2(b), the *supplier\_customer* relation between Apple Inc. and Lens Technology Co Ltd has a larger impact on the Lens's stock price in the period of releasing new version of iPhone than usual. To address this limitation, we propose to encode the temporal information into the relation-strength function and define the *Time-aware Embedding Propagation* process as follows:

$$\bar{\mathbf{e}}_i^t = \sum_{\{j | \text{sum}(\mathbf{a}_{ji}) > 0\}} \frac{g(\mathbf{a}_{ji}, \mathbf{e}_i^t, \mathbf{e}_j^t)}{d_j} \mathbf{e}_j^t, \quad (11)$$

which takes the sequential embeddings (note that they are time-sensitive) into account to estimate the strength of a relation. Besides encoding the temporal information, another benefit of such a design is that sequential embedding also encodes the stock information. This allows the relation-strength function to estimate the impact of a relation vector based on the stocks of concern, which is very desirable.

Next we describe two designs of the time-sensitive relation-strength function, which differ in whether to model the interaction between two stocks in an explicit or implicit manner.

- **Explicit Modeling.** For the explicit way, we define the relation strength function as:

$$g(\mathbf{a}_{ji}, \mathbf{e}_i^t, \mathbf{e}_j^t) = \underbrace{\mathbf{e}_i^t \mathbf{e}_j^t}_{\text{similarity}} \times \underbrace{\phi(\mathbf{w}^T \mathbf{a}_{ji} + b)}_{\text{relation importance}}, \quad (12)$$

where  $\mathbf{w} \in \mathbb{R}^K$  and  $b$  are model parameters to be learned;  $\phi$  is an activation function<sup>6</sup>. The relation strength of  $\mathbf{a}_{ji}$  is determined by two terms – *similarity* and *relation importance*. Specifically, the first term measures the similarity between the two stocks at the current time-step. The intuition is that the more similar the two stocks are at the current time, it is more likely that their relations will impact their prices in the near future. We use inner product to estimate the similarity, inspired by its effectiveness in modeling the similarity (interaction) between two entities (embeddings) in Collaborative Filtering [15]. The second term is a nonlinear regression model on the relations, where each element in  $\mathbf{w}$  denotes the weight of a relation in general and  $b$  is a bias term. Since both terms of this function are directly interpretable, we call it as *Explicit Modeling*.

- **Implicit Modeling.** In this design, we feed the sequential embeddings and the relation vector into a fully connected layer to estimate the relation strength:

$$g(\mathbf{a}_{ji}, \mathbf{e}_i^t, \mathbf{e}_j^t) = \phi(\mathbf{w}^T [\mathbf{e}_i^t, \mathbf{e}_j^t, \mathbf{a}_{ji}^T]^T + b), \quad (13)$$

where  $\mathbf{w} \in \mathbb{R}^{2U+K}$  and  $b$  are model parameters to be learned;  $\phi$  is an activation function same as the one in Equation 12. Then we normalize the outputs using a softmax function, which also endows it with more non-linearities. Since this way of interaction is implicitly captured by the parameters, we call it as *Implicit Modeling*.

**3.2.1 Connection with Graph-based Learning.** The embedding propagation is equivalent to the graph convolutional network (GCN). To show the relation, let us first construct a graph based on the stock relation encodings  $\mathcal{A}^t$ , where vertices represent stocks and edges connect vertices with at least one relation, i.e., we connect vertex  $i$  and  $j$  if they satisfy the condition  $\text{sum}(\mathbf{a}_{ij}) > 0$ . If we represent the graph with an adjacency matrix  $\mathbf{A}$  and normalize it by column, the Uniform Embedding Propagation (i.e., Equation 10) has exactly the same effect as the state-of-the-art graph convolutional operation (i.e.,  $f(\mathbf{F}, \mathbf{X}) = \mathbf{AX}$ ; details see Section 2.2.1). However, GCN cannot capture the temporal evolution properties as designed in our TGC, since the adjacency matrix  $\mathbf{A}$  has to be fixed in GCN. As such, our proposed operation can be seen as generalizing the GCN by specifically modeling the temporal patterns, thus we term it as the *Temporal Graph Convolution*.

## 4 DATA COLLECTION

Most existing works evaluate stock prediction on dozens of stocks, and there lacks a large stock dataset for an extensive evaluation. As such, we consider constructing data by ourselves, which is accessible through: [https://github.com/hennande/Temporal\\_Relational\\_Stock\\_Ranking](https://github.com/hennande/Temporal_Relational_Stock_Ranking). Specifically, we collect the stocks from the NASDAQ and NYSE markets that have transaction records between 01/02/2013 and 12/08/2017, obtaining 3, 274 and 3, 163 stocks respectively. Note that we select these two markets for their representative properties that NASDAQ is more volatile whereas NYSE is more stable [33]. Furthermore, we perform a filtering on the stocks by retaining the stocks satisfying the two conditions: 1) have been traded on more than 98% of trading days since 01/02/2013; 2) have never been traded at less than five dollars per share during the collection period. It should be noted that the first condition is based on concerns that intermittent sequences may bring abnormal patterns; the second condition ensures that the selected stocks are not penny stocks<sup>7</sup>, which are

<sup>6</sup>Note that we employ the *leaky rectifier* [25] with a slope of 0.2 as the activation function in our implementation.

<sup>7</sup><https://www.sec.gov/fast-answers/answerspenny.htm>

Table 3. Statistics of the sequential data.

| Market | Stocks# | Training Days#           | Validation Days#         | Testing Days#            |
|--------|---------|--------------------------|--------------------------|--------------------------|
|        |         | 01/02/2013<br>12/31/2015 | 01/04/2016<br>12/30/2016 | 01/03/2017<br>12/08/2017 |
| NASDAQ | 1,026   | 756                      | 252                      | 237                      |
| NYSE   | 1,737   | 756                      | 252                      | 237                      |

too risky for general investors as suggested by the U.S. Securities and Exchange Commission. This results in 1,026 NASDAQ and 1,737 NYSE stocks for our experiments. For these stocks, we collect three kinds of data: 1) historical price data, 2) sector-industry relations, and 3) Wiki relations between their companies such as supplier-consumer relation and ownership relation. Next, we present the details of these data.

#### 4.1 Sequential Data

Following [42], we set the prediction frequency as daily-level. Under our problem formulation, we aim to predict a ranking list of stocks for the following trading day, based on the daily historical data in the last  $S$  trading days. As the return ratio of a stock indicates the expected revenue of the stock, we set the ground-truth ranking score of stock  $i$  as its 1-day return ratio  $r_i^{t+1} = (p_i^{t+1} - p_i^t) / p_i^t$  where  $p_i^t$  is the closing price at day  $t$ . To calculate the ground-truth, we first collect the daily closing price of each stock ranging from 01/02/2013 and 12/08/2017. After the collection, we normalize the price of each stock via dividing it by its maximum value throughout the entire 2013-2017 dataset. In addition to the normalized closing price, we calculate four more sequential features: 5, 10, 20, and 30 days moving averages which represent the weekly and monthly trends. Following the existing work of stock prediction [42], we chronologically separate the sequential data into three time periods for training (2013-2015), validation (2016), and evaluation (2017), respectively, and summarize the basic statistics in Table 3. As can be seen, there are 756, 252, and 237 trading days in training, validation, and evaluation, respectively.

#### 4.2 Stock Relation Data

**4.2.1 Sector-Industry relations.** Observing the trends that stocks under the same industry are similarly influenced by the prospect of the industry, we collect the sector-industry relation between stocks. In NASDAQ and NYSE, each stock is classified into a sector and an industry as illustrated in Figure 3 in which stocks in the same industry are organized under the corresponding industry node. We collect the hierarchy structure of NASDAQ and NYSE stocks from the official company list maintained by NASDAQ Inc.<sup>8</sup> and extract industry relations for each stock pair under the same industry node, such as (GOOGL; Computer Software: Programming, Data Processing; FB). The specific relations extracted are detailed in the Appendix A.1 at the end of this paper. After the extraction, we count the number of industry (i.e., relation) types occurred in each market, and the ratio of stock pairs having industry relations and summarize them in Table 4. As can be seen, there are 112 and 130 types of industry relations between stock pairs in NASDAQ and NYSE, respectively. Moreover, the industry relation data is sparse since less than 10% of stock pairs have at least one type of industry relation in both stock markets.

**4.2.2 Wiki Company-based Relations.** As rich sources of entity relations, knowledge bases contain company entities and company relations, which might reflect the impact across stocks. As such, we extract the first-order and second-order company relations from Wikidata [37], one of the biggest and most active open domain knowledge bases with more than 42 million items (e.g.,

<sup>8</sup><https://www.nasdaq.com/screening/industries.aspx>



Fig. 3. Illustration of sector-industry hierarchy of companies in NASDAQ and NYSE.

Table 4. Statistics of sector-industry relation and Wiki relation data in the NASDAQ and NYSE datasets.

|               | Sector-Industry Relation |                           | Wiki Relation   |                           |
|---------------|--------------------------|---------------------------|-----------------|---------------------------|
|               | Relation Types#          | Relation Ratio (Pairwise) | Relation Types# | Relation Ratio (Pairwise) |
| <b>NASDAQ</b> | 112                      | 5.00%                     | 42              | 0.21%                     |
| <b>NYSE</b>   | 130                      | 9.37%                     | 32              | 0.30%                     |

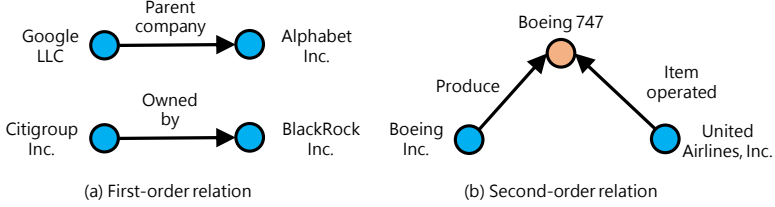


Fig. 4. Examples of the first-order and second-order company relations extracted from Wikidata.

*Alphabet Inc.*) and 367 million statements (e.g., *Alphabet Inc.; founded by; Larry Page*) in the format of (*subject; predicate; object*)<sup>9</sup>. As shown in Figure 4, company  $i$  has a first-order relation with  $j$  if there is a statement that has  $i$  and  $j$  as the subject and object, respectively. Companies  $i$  and  $j$  have a second-order relation if they have statements sharing the same object, such as *Boeing Inc.* and *United Airlines, Inc.* have different statements towards *Boeing 747*. After an exhausted exploration of a recent dump of Wikidata (01/05/2018), we obtain 5 and 53 types of first-order and second-order relations, respectively<sup>10</sup>. The detailed description of these relations is elaborated in the Appendix A.2 at the end of this paper. We then summarize the count of relation types and the ratio of stock pairs with at least one Wiki company-based relation in Table 4. As can be seen, there are 42 and 32 types of company relations occurring between stock pairs in NASDAQ and NYSE, respectively.

## 5 EXPERIMENT

To the best of our knowledge, our work is the first one to incorporate stock relations into the models for stock prediction, especially neural network-based ones. As such, in this section, we conduct experiments with the aim of answering the following research questions:

- (1) **RQ1:** How is the utility of formulating the stock prediction as a ranking task? Can our proposed RSR solution outperform state-of-the-art stock prediction solutions?
- (2) **RQ2:** Do stock relations enhance the neural network-based solution for stock prediction? How is the effectiveness of our proposed TGC component compared to conventional graph-based learning?
- (3) **RQ3:** How does our proposed RSR solution perform under different back-testing strategies?

<sup>9</sup><https://www.mediawiki.org/wiki/Wikibase/DataModel/JSON>

<sup>10</sup>We manually filter out less informative relations such as *located at the same timezone*.

In what follows, we first present the experimental settings, followed by answering the above three research questions.

## 5.1 Experimental Setting

**5.1.1 Evaluation Protocols.** Following [9], we adopt a daily buy-hold-sell trading strategy to evaluate the performance of stock prediction methods regarding the revenue. On each trading day  $t + 1$  during the testing period (from 01/03/2017 to 12/08/2017), we simulate a trader using a stock prediction method to trade in the following way:

- (1) **When the market closes at trading day  $t$ :** The trader uses the method to get the prediction, a ranking list with predicted return ratio of each stock. The trader buys the stock with the highest expected revenue (*i.e.*, ranked at the top).
- (2) **When the market closes at trading day  $t + 1$ :** The trader sells the stock purchased at day  $t$ .

In calculating the cumulative investment return ratio, we follow several simple assumptions: (1) The trader spends the same amount of money (*e.g.*, 50 thousand dollars) on every trading day. We make this assumption to eliminate the temporal dependency of the testing procedure for a fair comparison. (2) The market is always sufficiently liquid such that the buying order gets filled at the closing price of day  $t$  and the selling price is the closing price of day  $t + 1$ . (3) The transaction costs are ignored since the costs for trading US stocks through brokers are quite cheap no matter charging by trades or shares. For instance, Fidelity Investments and Interactive Brokers charge only 4.95 dollars per trade and 0.005 dollar per share, respectively<sup>11</sup>.

Since the target is to accurately predict the return ratio of stocks and appropriately rank the relative order of stocks, we employ three metrics, Mean Square Error (MSE), Mean Reciprocal Rank (MRR), and the cumulative investment return ratio (IRR), to report model performance. MSE has been widely used for evaluating regression tasks such as stock price prediction [22, 28, 42]. We thus calculate the MSE over all stocks on every trading day within the testing period. MRR [?] is a widely used metric for ranking performance evaluation. Here, we calculate the average reciprocal rank of the selected stock over the testing days. Since directly reflecting the effect of stock investment, IRR is our main metric, which is calculated by summing over the return ratios of the selected stock on each testing day. Smaller value of MSE ( $\geq 0$ ) and larger value of MRR ( $[0, 1]$ ) and IRR indicate better performance. For each method, we repeat the testing procedure five times and report the average performance to eliminate the fluctuations caused by different initializations.

**5.1.2 Methods.** We compare with the following stock price prediction baselines with *regression* formulation:

- **SFM [42]:** This method is the state-of-the-art stock price prediction method. It takes the historical closing prices as input and decomposes the prices into signals of different frequencies with a Discrete Fourier Transform (DFT). It then feeds the DFT coefficients into an extended LSTM with separate memory states for different frequencies to learn the frequency-aware sequential embeddings, which are fed into a FC layer to make the prediction.
- **LSTM [4]:** This method is the vanilla LSTM, which operates on the sequential data including closing prices and moving averages of 5, 10, 20, and 30 days, to obtain a sequential embedding; and then a FC layer is used to make prediction of the return ratio.

It should be noted that we ignore the potential baselines based on time-series models and shallow machine learning models, since they have been reported to be less effective than SFM and LSTM in

<sup>11</sup><https://www.stockbrokers.com/guides/commissions-fees>

Table 5. Performance comparison between the solutions with regression formulation (**SFM** and **LSTM**) and ranking formulation (**Rank\_LSTM**).

|                  | NASDAQ                 |                        |                  | NYSE                   |                        |                  |
|------------------|------------------------|------------------------|------------------|------------------------|------------------------|------------------|
|                  | MSE                    | MRR                    | IRR              | MSE                    | MRR                    | IRR              |
| <b>SFM</b>       | 5.20e-4±5.77e-5        | 2.33e-2±1.07e-2        | -0.25±0.52       | 3.81e-4±9.30e-5        | <b>4.82e-2±4.95e-3</b> | 0.49±0.47        |
| <b>LSTM</b>      | 3.81e-4±2.20e-6        | 3.64e-2±1.04e-2        | 0.13±0.62        | 2.31e-4±1.43e-6        | 2.75e-2±1.09e-2        | -0.90±0.73       |
| <b>Rank_LSTM</b> | <b>3.79e-4±1.11e-6</b> | <b>4.17e-2±7.50e-3</b> | <b>0.68±0.60</b> | <b>2.28e-4±1.16e-6</b> | 3.79e-2±8.82e-3        | <b>0.56±0.68</b> |

several previous works [4, 17, 42]. Moreover, we also compare with several methods with *ranking* formulation:

- **Rank\_LSTM**: We remove the relational embedding layer of the proposed RSR to obtain this method, *i.e.*, this method ignores stock relations.
- **Graph-based ranking (GBR)**: According to Equation 2, we add the graph regularization term to the loss function of **Rank\_LSTM**, which smooths predicted return ratios over the graph of *stock relations*. In the graph, we connect a pair of vertices (*i.e.*, stocks) having at least one type of relations.
- **GCN [20]**: GCN is the state-of-the-art graph-based learning method. We obtain this method by replacing the TGC layer of our proposed RSR with a GCN layer. The graph of *stock relations* in GBR is fed into the GCN layer.
- **RSR\_E**: Our proposed RSR with explicit modeling in the TGC.
- **RSR\_I**: Our proposed RSR with implicit modeling in the TGC.

**5.1.3 Parameter Settings.** We implement the models with TensorFlow<sup>12</sup> except **SFM** of which we use the original implementation<sup>13</sup>. It is worth mentioning that the implementations can be accessed through: [https://github.com/hennande/Temporal\\_Relational\\_Stock\\_Ranking](https://github.com/hennande/Temporal_Relational_Stock_Ranking). We employ grid search to select the optimal hyperparameters regarding IRR for all methods. For **SFM**, we follow the original setting in [42], optimizing it by RMSProp with a learning rate of 0.5, and tuning the number of frequencies and hidden units within {5, 10, 15} and {10, 20, 30}, respectively. For all other methods, we apply the Adam [19] optimizer with a learning rate of 0.001. We tune two hyperparameters for **LSTM**, the length of sequential input  $S$  and the number of hidden units  $U$ , within {2, 4, 8, 16} and {16, 32, 64, 128}, respectively. Besides  $S$  and  $U$ , we further tune  $\alpha$  in Equation 8, which balances the point-wise and pair-wise terms; specifically, we tune  $\alpha$  within {0.1, 1, 10} for **Rank\_LSTM**, **GCN**, **RSR\_E**, and **RSR\_I**. We further tune the  $\lambda$  of the regularization term in GBR within {0.1, 1, 10}.

## 5.2 Study of Stock Ranking Formulation (RQ1)

Table 5 summarizes the performance of baselines in *regression* fashion and **Rank\_LSTM** our basic solution of stock ranking *w.r.t.* MSE, MRR, and IRR, from which we have the following observations:

- **Rank\_LSTM** outperforms both **SFM** and **LSTM** on the two markets with great improvement *w.r.t.* IRR (>14%). It verifies the advantage of the stock ranking solutions and answers **RQ1** that stock ranking is a promising formulation of stock prediction. Moreover, it indicates the potential of advanced learning-to-rank techniques in solving the stock prediction task.
- However, **Rank\_LSTM** fails to consistently beat **SFM** and **LSTM** regarding all evaluation measures, its performance on NYSE *w.r.t.* MRR is worse than **SFM**. The reason could be attributed to minimizing the combination of point-wise and pair-wise losses, which would lead to a tradeoff between accurately predicting absolute value of return ratios and their relative order.

<sup>12</sup><https://www.tensorflow.org/>

<sup>13</sup><https://github.com/z331565360/State-Frequency-Memory-stock-prediction>



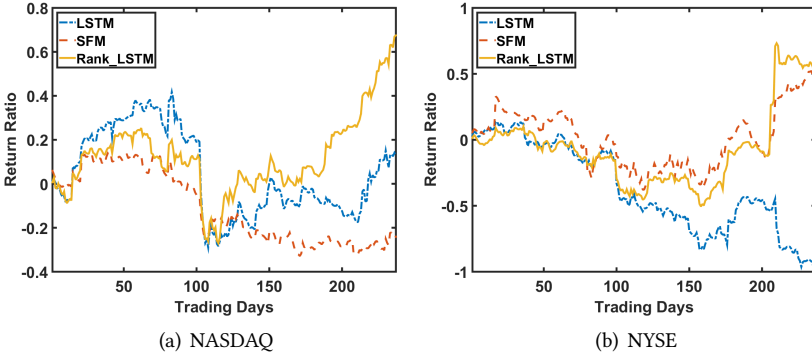
Fig. 5. Performance comparison of **Rank\_LSTM**, **SFM**, and **LSTM** regarding IRR.

Table 6. Performance comparison among relational ranking methods with industry relations.

|                  | NASDAQ                 |                        |                  | NYSE                   |                        |                  |
|------------------|------------------------|------------------------|------------------|------------------------|------------------------|------------------|
|                  | MSE                    | MRR                    | IRR              | MSE                    | MRR                    | IRR              |
| <b>Rank_LSTM</b> | <b>3.79e-4±1.11e-6</b> | 4.17e-2±7.50e-3        | <b>0.68±0.60</b> | 2.28e-4±1.16e-6        | 3.79e-2±8.82e-3        | 0.56±0.68        |
| <b>GBR</b>       | 5.80e-3±1.20e-3        | <b>4.46e-2±5.20e-3</b> | 0.57±0.29        | 2.29e-4±2.02e-6        | 3.43e-2±6.26e-3        | 0.68±0.31        |
| <b>GCN</b>       | 3.80e-4±2.24e-6        | 3.45e-2±8.36e-3        | 0.24±0.32        | 2.27e-4±1.30e-7        | <b>5.01e-2±5.56e-3</b> | 0.97±0.56        |
| <b>RSR_E</b>     | 3.82e-4±2.69e-6        | 3.16e-2±3.45e-3        | 0.20±0.22        | 2.29e-4±2.77e-6        | 4.28e-2±6.18e-3        | 1.00±0.58        |
| <b>RSR_I</b>     | 3.80e-4±7.90e-7        | 3.17e-2±5.09e-3        | 0.23±0.27        | <b>2.26e-4±5.30e-7</b> | 4.51e-2±2.41e-3        | <b>1.06±0.27</b> |

- The performance *w.r.t.* IRR varies a lot under different runs of a method. It is reasonable since the absolute value of daily return ratio varies from 0 to 0.98 in our dataset, which means that a tiny switch of the top 2 ranked stocks may lead to a huge change of the IRR. Such results also indicate that learning to rank techniques emphasizing the top-ranked stocks is worthwhile to be explored in the future.
- The performance of **LSTM** on the NYSE market *w.r.t.* IRR is unexpectedly bad. We repeat the parameter tuning and testing procedure several times and find that **LSTM** could achieve better performance (with IRR value between 0.1 and 0.2) with other settings of hyperparameters. However, the selected setting always beats the others on the validation. **This result indicates the potential difference between the validation and testing.**

Figure 5 illustrates the procedure of back-testing regarding the cumulative return ratios. As can be seen, **in all cases, the curves are volatile, which indicates that selecting only one stock from more than 1,000 is a highly risk operation.** Consequently, it also suggests the worth of introducing risk-oriented criteria into stock ranking tasks in the future.

### 5.3 Impact of Stock Relations (RQ2)

**Effect of Industry Relations:** Table 6 shows the performance of methods considering the industry relation of stocks. We can see that:

- **Considering industry relations is more beneficial to stock ranking on NYSE as compared to NASDAQ.** It could be attributed to that the industry relations reflect more of long-term correlations between stocks, since NASDAQ is considered as a much more volatile market as compared to NYSE and dominated by short-term factors [33].
- **On NYSE, all methods considering stock relations, i.e., GBR, GCN, RSR\_E, and RSR\_I, outperform Rank\_LSTM w.r.t. IRR.** Considering that all these methods take Rank\_LSTM as the building block, this result verifies the effectiveness of encoding stock relations in stock prediction.

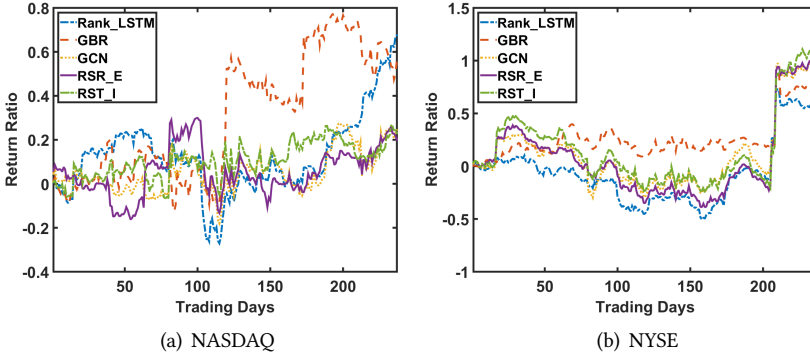


Fig. 6. Back-testing procedure of relational ranking methods with industry relations regarding IRR.

- Moreover, **RSR\_E** and **RSR\_I** achieve improvement over **GCN** and **GBR**. This result verifies the effectiveness of the proposed *Temporal Graph Convolution* as compared to the traditional modeling of relational data. Considering that **GCN** and **GBR** utilize a static graph to represent stock relations, the result also indicates the rationale of considering temporal property in stock relation modeling.
- Again, the performance regarding different evaluation measures is inconsistent. We speculate the reason is that we tune the hyperparameters regarding IRR, which focuses more on correct ranking on testing days with high return ratios. For instance, correct prediction on a trading day with ground truth return ratio of 0.5 would lead to higher IRR than correct predictions in ten trading days with return ratio of 0.01. As such, a model achieves better IRR could achieve suboptimal MSE and MRR.

Figure 6 illustrates the IRR curve of the compared methods in the back-testing. Again, the curves are volatile of which the reason has been discussed in Section 5.2. On NYSE, the IRR of the methods presents huge increases on the 206-th and 209-th trading days when the best-performed stock exhibits return ratios larger than 0.6. This result further highlights the importance of accurately predicting both the return ratio of single stock and the relative order of stocks. Note that capturing rare opportunities by precisely ranking the stocks on trading days with huge change of return ratios would lead to satisfied IRR.

**Effect of Wiki Relations:** Similarly, Table 7 shows the results of considering the Wiki relation of stocks. We observe that:

- In all cases, our proposed **RSR\_E** and **RSR\_I** achieve the best performance *w.r.t.* IRR. It further demonstrates the effectiveness of the approach we model stock relations, that is, the *TGC* component.
- All methods considering Wiki relations outperform **Rank\_LSTM** with a significant improvement ( $>0.09$ ) *w.r.t.* IRR on NYSE. It again verifies the merit of encoding stock relations in stock prediction and the effectiveness of the RSR framework.

Figure 7 shows the associated back-testing procedure, which presents similar trends as the results of considering industry relations (Figure 6).

**Sector-wise Performance:** Taking **RSR\_I** as an example, we then investigate whether the performance is sensitive to sectors via evaluating its performance over the stocks in each sector, *i.e.*, separately conducting back-testing for each sector. Recall that, on NASDAQ, the performance of **RSR\_I** considering industry relations is not promising (with an IRR of 0.23). We take this case to investigate the sector-wise performance, which is presented in Table 8. Note that we only show the performance on sectors with the top-5 most stocks. We can see that, the method only achieves

Table 7. Performance comparison among relational ranking methods with Wiki relations.

|                  | NASDAQ                 |                        |                  | NYSE                   |                 |                  |
|------------------|------------------------|------------------------|------------------|------------------------|-----------------|------------------|
|                  | MSE                    | MRR                    | IRR              | MSE                    | MRR             | IRR              |
| <b>Rank_LSTM</b> | <b>3.79e-4±1.11e-6</b> | 4.17e-2±7.50e-3        | <b>0.68±0.60</b> | 2.28e-4±1.16e-6        | 3.79e-2±8.82e-3 | <b>0.56±0.68</b> |
| <b>GBR</b>       | 3.80e-4±2.40e-7        | 3.32e-2±4.50e-3        | 0.33±0.34        | <b>2.26e-4±4.20e-7</b> | 3.64e-2±5.35e-3 | 0.65±0.27        |
| <b>GCN</b>       | <b>3.79e-4±9.70e-7</b> | 3.24e-2±3.21e-3        | 0.11±0.06        | <b>2.26e-4±6.60e-7</b> | 3.99e-2±1.03e-2 | 0.74±0.30        |
| <b>RSR_E</b>     | 3.80e-4±7.20e-7        | 3.94e-2±8.15e-3        | 0.81±0.85        | 2.29e-4±2.77e-6        | 4.28e-2±6.18e-3 | <b>0.96±0.47</b> |
| <b>RSR_I</b>     | <b>3.79e-4±6.60e-7</b> | <b>4.09e-2±5.18e-3</b> | <b>1.19±0.55</b> | <b>2.26e-4±1.37e-6</b> | 4.58e-2±5.55e-3 | 0.79±0.34        |

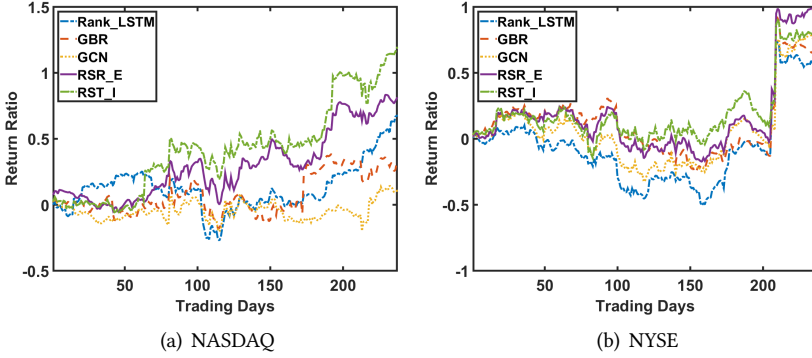


Fig. 7. Performance comparison of relational ranking methods with Wiki relations regarding IRR.

Table 8. Performance of **RSR\_I** on ranking stocks in different sectors of NASDAQ w.r.t. IRR.

| Sector  | Finance | Technology | N/A   | Consumer Services | Health Care |
|---------|---------|------------|-------|-------------------|-------------|
| #Stocks | 222     | 182        | 156   | 117               | 91          |
| IRR     | 0.33    | 1.12       | -0.70 | 0.57              | -0.85       |

Table 9. Impacts of different types of Wiki relation regarding the Relative Performance Decrease (RPD) of **RSR\_I** on NASDAQ as removing the selected relation.

| Relation       | P1056_P1056 | P463_P463 | P452_P452 | P361_P361 | P1056_P452 |
|----------------|-------------|-----------|-----------|-----------|------------|
| ID in Table 14 | 46          | R38       | 35        | 31        | 45         |
| #Occurrences   | 130         | 58        | 506       | 1,194     | 10         |
| <b>RPD</b>     | -144.00%    | -70.13%   | -21.66%   | -17.52%   | -15.71%    |

acceptable performance with an IRR of 1.12 on the *Technology* sector. This result further indicates the less effectiveness of considering industry relation on the NASDAQ market, which is coherent with the results in Table 6. In addition, it also suggests the separate consideration of stocks in each single sector.

**Importance of Each Type of Wiki relation:** By comparing the performance of **RSR\_I** when a type of relation is removed, we investigate the importance of different types of Wiki relations. Table 9 shows the relative performance decrease w.r.t. IRR as compared to **RSR\_I** with all Wiki relations as input (NASDAQ). Note that we only present the relations with the top-5 largest performance decreases. We can see that the most importance relation is the *P1056\_P1056* of which *P1056* denotes the predicate of *product or material produced*. Mainly, two stocks have the relation of *P1056\_P1056* means the associated companies collaborate on producing the same product. The high impact is reasonable, considering that collaborated companies have closely connected revenues and would be affected by similar factors.

**Brief Conclusion:** a) Considering stock relations is helpful for stock ranking, especially on the stable markets (e.g., NYSE). 2) The proposed *TGC* is a promising solution for encoding stock relations. 3) It is important to consider appropriate relations suitable for the target market, for example, encoding industry relations on NASDAQ is a suboptimal choice.

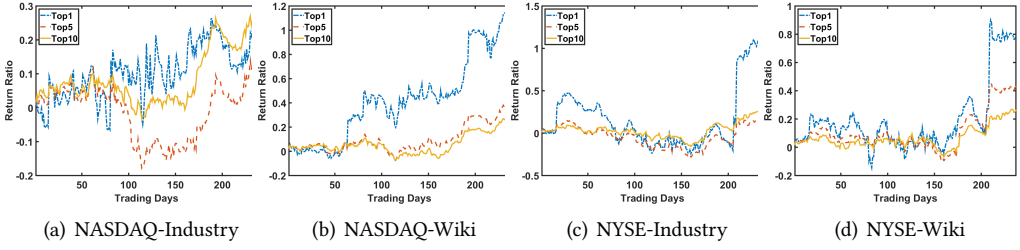


Fig. 8. Comparison on back-testing strategies (**Top1**, **Top5**, and **Top10**) w.r.t. IRR based on prediction of **RSR\_I**.

#### 5.4 Study on Back-testing Strategies (RQ3)

We then investigate the performance of our proposed methods under three different back-testing strategies, named **Top1**, **Top5**, and **Top10**, buying stocks with top-1, 5, 10 highest expected revenue, respectively. For instance, with the back-testing strategy of **Top10**, we equally split our budget to trade the top-10 ranked stocks on each testing day. Note that we accordingly calculate the IRR by summing the mean return ratio of the 10 selected stocks on each testing day. Figure 8 illustrates the performance comparison of these strategies with the predictions of **RSR\_I**. Similar trends are observed on the predictions of **RSR\_E**, which are omitted for the consideration of saving space. From the figure, we have the following observations:

- **RSR\_I** (Figure 8(a)) fails to achieve expected performance with different back-testing strategies under the NASDAQ-Industry setting (*i.e.*, ranking stocks in NASDAQ and modeling their industry relations). It further indicates the less effectiveness of industry relations on NASDAQ.
- In the other cases, the performance of **Top1**, **Top5**, and **Top10** on most testing days follows the order of **Top1** > **Top5** > **Top10**, *i.e.*, the **Top1** and **Top10** achieve the highest and lowest IRR, respectively. The reason could be that the ranking algorithm could accurately rank the relative order of stocks regarding future return ratios. Once the order is accurate, buying and selling the stock with higher expected profit (*e.g.*, the top-1 ranked one) would achieve higher cumulative return ratio.

Considering that it would easier to achieve better performance regarding IRR in a bullish market, we further compare the performance of our method with two market indices, S&P 500 Index and Dow Jones Industrial Average Index (DJI). Moreover, in order to better judge the achieved performance, we compare two more ideal investment strategies: a) selecting the stocks with highest return ratio (*e.g.*, Top10) in the testing period from the whole market; and b) among the stocks traded by the proposed method, selecting the stocks with highest return ratio in the testing period. Table 10 shows the performance of the compared investment strategies w.r.t. IRR. From which, we have the following observations:

- In the testing period, the stock market is bullish, which suggests future exploration of the proposed method in bearish market. In addition, noting that market indices are competitive portfolios (investment strategies)<sup>14</sup>, achieving IRR higher than market indices justifies the effectiveness of the proposed method.
- When trading the same number of stocks (*e.g.*, **Top5**), the performance of the proposed method presents a significant gap towards the ideal investment strategies. This result is acceptable since accurately selecting the stock performing best in the range of almost one year is non-trivial, but reflects the huge improvement space for stock prediction methods.

<sup>14</sup>From 2008 to 2017, the S&P 500 achieved a return ratio of 125.8%, beating most of the portfolios of funds of hedge funds.

Table 10. Performance of RSR\_I as compared to market indices and ideal portfolios.

|                    | NASDAQ      |             |             | NYSE        |             |             |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                    | Top1        | Top5        | Top10       | Top1        | Top5        | Top10       |
| <b>Market</b>      | <b>3.40</b> | <b>2.36</b> | <b>1.99</b> | <b>2.42</b> | <b>1.90</b> | <b>1.47</b> |
| <b>Selected</b>    | 1.63        | 0.81        | 1.10        | 2.24        | 1.78        | 1.39        |
| <b>RSR_I</b>       | 1.19        | 0.40        | 0.27        | 1.06        | 0.18        | 0.26        |
| <b>S&amp;P 500</b> | 0.17        |             |             |             |             |             |
| <b>DJI</b>         | 0.22        |             |             |             |             |             |

- The **Top1** version of our method, *i.e.*, trading the top-1 ranked stock on each trading day, achieves an IRR comparable to the investment strategy **Selected** under **Top10**. This further justifies the competitiveness of our proposed method.

## 6 RELATED WORK

Our work is directly related to the recent work on *stock prediction*, *graph-based learning*, and *knowledge graph embedding*.

### 6.1 Stock Prediction

Recent work on stock prediction can be separated into two main categories: *stock price regression* and *stock trend classification*. On one hand, Bao *et al.* predicted the 1-day ahead closing price of stocks with historical prices as input. The authors viewed the historical price as a signal and decomposed the historical price into multiple frequencies with a Wavelet Transform. They then filtered out noises in the frequency domain with a Stacked Autoencoder (SAE) and fed the output of SAE to an LSTM to make prediction [4]. Zhang *et al.* devised an extension of LSTM, which decomposes the historical prices into frequency domain with a Discrete Fourier Transform and equips each frequency with a memory state to capture the patterns in different frequencies [42]. Instead of directly modeling the stock prices, Alberg and Lipton used a combination of an LSTM and Multi-Layer Perception to predict the future trend of fundamental indicators of a company and trade the corresponding stock based on the predicted indicators [3].

On the other hand, Nguyen and Shirai proposed a stock trend classification solution, which learns a topic distribution representation of each stock from posts mentioning it on stock message boards, and fed the topic representation into a Support Vector Machine to make the trend classification [29]. Under a similar classification setting, another line of research is classifying the trend of a stock from relevant financial news reports [17, 44 ? ? ]. For instance, Zhao *et al.* achieve it via constructing a event causality network of news reports and learning news embeddings from the causality networks, which is fed into a classification layer [44]. Taking financial news as input as well, Hu *et al.* devised a neural network-based solution, named *Hybrid Attention Networks*, which leverages a hybrid attention mechanism to attentively fuse multiple news reports mentioning a stock into a joint representation [17]. In addition, textual contents mentioning stocks in social media are also used to forecast the movement of stocks [? ].

However, none of the existing work is able to incorporate the rank/relative order among stocks regarding the expected revenue, tending to lead to suboptimal stock selections. Moreover, the existing work either totally ignores stock relations or heuristically models such relations. For instance, an intuitive consideration of sector-industry relation is to separately train a predictor for stocks under each sector [? ]. To the best of our knowledge, our work is to first one to leverage techniques of learning-to-rank to solve the stock prediction task and inject the stock relations into the learning framework with a new neural network component.

## 6.2 Graph-based Learning

In the literature of graph-based learning, it has been intensively studied that incorporating the relationship among entities into the learning procedure of the target task to achieve better performance. Work on graph-based learning are mainly in two fashions: *graph regularization* and *graph convolution*. On one hand, Zhu *et al.* proposed a regularization term based on directed and undirected simple graphs with pair-wise entity relations to smooth the predictions across the topology of the graph [47]. Zhou *et al.* regularized the learning procedure of the target task with a hypergraph that captures the higher-order relations among entities [46]. On the other hand, Bruna *et al.* proposed spectral graph convolutions to capture the local connection patterns in graphs and propagate information of locally connected vertices for better representations [6]. Upon the spectral graph convolution, several fast approximations have been proposed for accelerating [8, 20]. However, most of the works of graph-based learning fail to handle the temporal evolution property of stock market. **Consequently, they would suffer from severe information loss and achieve limited improvement when directly applied to model stock relations.**

## 6.3 Knowledge Graph Embedding

In a similar line, modeling the relations of two entities (a.k.a. knowledge graph embedding) has been intensively studied in the literature of knowledge graphs. Bordes, *et al.* represented entities and relations with embedding vectors and transferred entity embeddings through adding relation embedding [5]. Similarly, Socher *et al.* represented relations as matrices and transferred entity embeddings via matrix multiplication [34]. Such techniques mainly focus on solving knowledge graph-oriented problems such as knowledge graph completion, while we target on a different problem setting of stock ranking. The idea of embedding propagation that revises stock sequential embeddings through stock relations is partially inspired by TransE, our TGC is more generic in terms of jointly capturing temporal properties and topologies of relations.

## 7 CONCLUSIONS

In this paper, we formulated stock prediction as a ranking task and demonstrated the potential of learning-to-rank methods for predicting stocks. To solve the problem, we proposed a *Relational Stock Ranking* framework. The core of the framework is a neural network modeling component, named *Temporal Graph Convolution*, which can handle the impact between different stocks by encoding stock relations in a time-sensitive way. Experimental results on NASDAQ and NYSE demonstrate the effectiveness of our solution — with three different back-testing strategies, the RSR framework outperforms the S&P 500 Index with significantly higher return ratio.

As mentioned in Section 5, we will explore the potential of emphasizing top-ranked entities with more advanced learning-to-rank techniques. In addition, we will integrate risk management techniques in finance into the RSR framework to force the predictions to be risk sensitive. Furthermore, we will investigate the performance of RSR under multiple investment operations such as *buy-hold-sell* (*aka.* long position) and *borrow-sell-buy* (*aka.* short position). Moreover, we will integrate alternative data such as financial news and social media contents into the predictive model. **Lastly, considering that the proposed TGC is a general component to model relational data, especially structured domain knowledge, we would like to explore the potential of TGC in enhancing the neural network solutions for tasks with such relational data, such as recommender system and product search.**



## A STOCK RELATION

In this appendix, we describe the details of stock relations (*i.e.*, *sector-industry relations* and *Wiki company-based relations*) in our collected data (Section 4).

### A.1 Sector-Industry Relation

We extract 112 and 130 types of industry relations from the [company classification](#) hierarchy structure of NASDAQ and NYSE stocks, respectively. Table 11 and 12 illustrates the specific industry relations in NYSDAQ and NYSE markets, respectively.

Table 11. Industry relations among 1,026 selected stocks from the NASDAQ market.

| Sectors               | Industries  | Count of Industries |
|-----------------------|---|---------------------|
| Consumer Durables     | Office Equipment/Supplies/Services, Consumer Specialties, Specialty Chemicals, Metal Fabrications, Consumer Electronics/Appliances, Building Products, Containers/Packaging, Miscellaneous manufacturing industries, Automotive Aftermarket   | 9                   |
| Transportation        | Transportation Services, Air Freight/Delivery Services, Trucking Freight/Courier Services, Oil Refining/Marketing   | 4                   |
| Finance               | Specialty Insurers, Commercial Banks, Savings Institutions, Real Estate, Major Banks, Investment Managers, Investment Bankers/Brokers/Service, Life Insurance, Finance: Consumer Services, Banks, Property-Casualty Insurers, Finance Companies   | 12                  |
| Public Utilities      | Telecommunications Equipment, Environmental Services, Natural Gas Distribution  | 3                   |
| Energy                | Electric Utilities: Central, Coal Mining, Oil & Gas Production  | 3                   |
| Miscellaneous         | Business Services, Publishing, Multi-Sector Companies   | 3                   |
| Consumer Non-Durables | Plastic Products, Meat/Poultry/Fish, Beverages (Production/Distribution), Shoe Manufacturing, Packaged Foods, Package Goods/Cosmetics, Apparel, Farming/Seeds/Milling, Food Distributors, Specialty Foods, Recreational Products/Toys   | 11                  |
| Health Care           | Ophthalmic Goods, Medical/Nursing Services, Hospital/Nursing Management, Biotechnology: In Vitro & In Vivo Diagnostic Substances, Biotechnology: Commercial Physical & Biological Research, Other Pharmaceuticals, Major Pharmaceuticals, Medical Specialties, Medical Electronics, Biotechnology: Electromedical & Electrotherapeutic Apparatus, Biotechnology: Biological Products, Medical/Dental Instruments, Industrial Specialties  | 13                  |
| Consumer Services     | Other Consumer Services, Restaurants, Clothing/Shoe/Accessory Stores, Marine Transportation, Television Services, Consumer Electronics/Video Chains, Other Specialty Stores, Home Furnishings, Diversified Commercial Services, Paper, Professional Services, Hotels/Resorts, Rental/Leasing Companies, Real Estate Investment Trusts, Food Chains, Broadcasting, Books, Motor Vehicles, Movies/Entertainment, RETAIL: Building Materials, Advertising, Catalog/Specialty Distribution, Services-Misc. Amusement & Recreation, Department/Specialty Retail Stores | 24                  |
| Basic Industries      | Water Supply, Miscellaneous, Forest Products, Precious Metals, Mining & Quarrying of Nonmetallic Minerals, Engineering & Construction, Major Chemicals  | 7                   |

|               |  |    |
|---------------|--|----|
| Capital Goods | Military/Government/Technical, Biotechnology: Laboratory Analytical Instruments, Electrical Products, Building Materials, Railroads, Ordnance And Accessories, Homebuilding, Electronic Components, Aerospace, Industrial Machinery/Components, Construction/Ag Equipment/Trucks, Auto Parts: O.E.M., Steel/Iron Ore, Auto Manufacturing | 14 |
| Technology    | Computer Manufacturing, Radio And Television Broadcasting And Communications Equipment, Computer Communications Equipment, Computer peripheral equipment, EDP Services, Computer Software: Programming, Data Processing, Computer Software: Prepackaged Software, Semiconductors, Retail: Computer Software & Peripheral Equipment       | 9  |
| N/A           | N/A  | 1  |

Table 12. Industry relations among 1,737 selected stocks from the NYSE market.

| Sectors               | Industries   | Count of Industries |
|-----------------------|--|---------------------|
| Consumer Durables     | Electrical Products, Home Furnishings, Specialty Chemicals, Metal Fabrications, Consumer Electronics/Appliances, Building Products, Miscellaneous manufacturing industries, Containers/Packaging, Publishing, Automotive Aftermarket, Industrial Specialties   | 11                  |
| Transportation        | Transportation Services, Air Freight/Delivery Services, Trucking Freight/Courier Services, Railroads, Oil Refining/Marketing   | 5                   |
| Finance               | Finance/Investors Services, Specialty Insurers, Commercial Banks, Savings Institutions, Real Estate, Major Banks, Investment Managers, Investment Bankers/Brokers/Service, Life Insurance, Diversified Financial Services, Accident &Health Insurance, Finance: Consumer Services, Banks, Property-Casualty Insurers, Finance Companies  | 15                  |
| Public Utilities      | Electric Utilities: Central, Telecommunications Equipment, Oil/Gas Transmission, Water Supply, Power Generation  | 5                   |
| Energy                | Coal Mining, Oil & Gas Production, Integrated oil Companies, Oilfield Services/Equipment, Natural Gas Distribution   | 5                   |
| Miscellaneous         | Business Services, Office Equipment/Supplies/Services, Multi-Sector Companies  | 3                   |
| Consumer Non-Durables | Electronic Components, Plastic Products, Meat/Poultry/Fish, Shoe Manufacturing, Beverages, Packaged Foods, Consumer Specialties, Apparel, Farming/Seeds/Milling, Food Distributors, Specialty Foods, Motor Vehicles, Recreational Products/Toys  | 13                  |
| Health Care           | Ophthalmic Goods, Medical/Nursing Services, Hospital/Nursing Management, Major Pharmaceuticals, Biotechnology: Commercial Physical Research, Biotechnology: Electromedical Apparatus, Other Pharmaceuticals, Medical/Dental Instruments, Medical Specialties   | 9                   |
| Consumer Services     | Other Consumer Services, Restaurants, Clothing/Shoe/Accessory Stores, Electronics/Video Chains, Other Specialty Stores, Home Furnishings, Diversified Commercial Services, Paper, Professional Services, Hotels/Resorts, Rental/Leasing Companies, Real Estate Investment Trusts, Food Chains, Broadcasting, Books, Motor Vehicles, Movies/Entertainment, RETAIL: Building Materials, Advertising, Catalog/Specialty Distribution, Services-Misc. Amusement & Recreation, Department/Specialty Retail Stores | 24                  |
| Basic Industries      | Water Supply, Miscellaneous, Forest Products, Precious Metals, Mining & Quarrying of Nonmetallic Minerals, Engineering & Construction, Major Chemicals   | 7                   |

|               |  |    |
|---------------|--|----|
| Capital Goods | Package Goods/Cosmetics, Forest Products, Precious Metals, Environmental Services, Paper, Agricultural Chemicals, Mining & Quarrying of Nonmetallic Minerals, Engineering & Construction, General Bldg Contractors - Nonresidential Bldgs, Aluminum, Major Chemicals, Paints/Coatings, Steel/Iron Ore, Textiles  | 13 |
| Technology    | Computer Manufacturing, Computer peripheral equipment, Computer Software: Programming, Semiconductors, Data Processing, Computer Software: Prepackaged Software, Diversified Commercial Services, Professional Services, Computer Communications Equipment, EDP Services, Retail: Computer Software & Peripheral Equipment, Radio And Television Broadcasting Equipment, Advertising | 12 |
| N/A           | N/A  | 1  |

## A.2 Wiki Company-based Relations

From Wikidata, one of the biggest and most active open domain knowledge bases, we obtain 5 and 53 types of first-order (in the format of  $\textcircled{A} \xrightarrow{R} \textcircled{B}$ ) and second-order relations (in the format of  $\textcircled{A} \xrightarrow{R_1} \textcircled{C} \xleftarrow{R_2} \textcircled{B}$ ) between companies corresponding to the selected stocks in NASDAQ and NYSE markets, respectively. Note that  $A$  and  $B$  denote entities in Wikidata corresponding to two companies;  $C$  denotes another entity bridging two company-entities in a second-order relation;  $R$ ,  $R_1$ , and  $R_2$  denotes different types of entity relation defined in Wikidata<sup>15</sup>. In Table 13 and 14, we summarize the extracted first-order and second-order relations, respectively.

Table 13. First-order Wiki company-based relations in the format of  $\textcircled{A} \xrightarrow{R} \textcircled{B}$ .

|   | Wikidata Relation ( $R$ ) | Relation Description   |
|---|---------------------------|--|
| 1 | <a href="#">P127</a>      | <i>Owned by</i> : owner of the subject.  |
| 2 | <a href="#">P155</a>      | <i>Follows</i> : immediately prior item in a series of which the subject is a part.            |
| 3 | <a href="#">P156</a>      | <i>Followed by</i> : immediately following item in a series of which the subject is a part.    |
| 4 | <a href="#">P355</a>      | <i>Subsidiary</i> : subsidiary of a company or organization.                                   |
| 5 | <a href="#">P749</a>      | <i>Parent organization</i> : parent organization of an organisation, opposite of subsidiaries. |

Table 14. Second-order Wiki company-based relations in the format of  $\textcircled{A} \xrightarrow{R_1} \textcircled{C} \xleftarrow{R_2} \textcircled{B}$ .

|   | Wikidata Relations   | Relation Descriptions   |
|---|----------------------|---|
| 1 | $R_1 = \text{P31}$   | <i>Instance of</i> : that class of which this subject is a particular example and member. |
|   | $R_2 = \text{P366}$  | <i>Use</i> : main use of the subject.   |
| 2 | $R_1 = \text{P31}$   | <i>Instance of</i> : that class of which this subject is a particular example and member. |
|   | $R_2 = \text{P452}$  | <i>Industry</i> : industry of company or organization.                                    |
| 3 | $R_1 = \text{P31}$   | <i>Instance of</i> : that class of which this subject is a particular example and member. |
|   | $R_2 = \text{P1056}$ | <i>Product or material produced</i> : material or product produced by an agency.          |
| 4 | $R_1 = \text{P112}$  | <i>Founded by</i> : founder or co-founder of this organization.                           |
|   | $R_2 = \text{P112}$  | <i>Founded by</i> : founder or co-founder of this organization.                           |
| 5 | $R_1 = \text{P112}$  | <i>Founded by</i> : founder or co-founder of this organization.                           |
|   | $R_2 = \text{P127}$  | <i>Owned by</i> : owner of the subject.   |
| 6 | $R_1 = \text{P112}$  | <i>Founded by</i> : founder or co-founder of this organization.                           |
|   | $R_2 = \text{P169}$  | <i>Chief executive officer</i> : the CEO within an organization.                          |
| 7 | $R_1 = \text{P113}$  | <i>Airline hub</i> : airport that serves as a hub for an airline.                         |
|   | $R_2 = \text{P113}$  | <i>Airline hub</i> : airport that serves as a hub for an airline.                         |

<sup>15</sup>[https://www.wikidata.org/wiki/Wikidata:List\\_of\\_properties/all](https://www.wikidata.org/wiki/Wikidata:List_of_properties/all)

|    |               |   |
|----|---------------|---|
| 8  | $R_1 = P114$  | <i>Airline alliance</i> : alliance the airline belongs to.                                |
|    | $R_2 = P114$  | <i>Airline alliance</i> : alliance the airline belongs to.                                |
| 9  | $R_1 = P121$  | <i>Item operated</i> : equipment, installation or service operated by the subject.        |
|    | $R_2 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
| 10 | $R_1 = P121$  | <i>Item operated</i> : equipment, installation or service operated by the subject.        |
|    | $R_2 = P121$  | <i>Item operated</i> : equipment, installation or service operated by the subject.        |
| 11 | $R_1 = P127$  | <i>Owned by</i> : owner of the subject.   |
|    | $R_2 = P112$  | <i>Founded by</i> : founder or co-founder of this organization.                           |
| 12 | $R_1 = P127$  | <i>Owned by</i> : owner of the subject.   |
|    | $R_2 = P127$  | <i>Owned by</i> : owner of the subject.   |
| 13 | $R_1 = P127$  | <i>Owned by</i> : owner of the subject.   |
|    | $R_2 = P169$  | <i>Chief executive officer</i> : the CEO within an organization.                          |
| 14 | $R_1 = P127$  | <i>Owned by</i> : owner of the subject.   |
|    | $R_2 = P355$  | <i>Subsidiary</i> : subsidiary of a company or organization.                              |
| 15 | $R_1 = P127$  | <i>Owned by</i> : owner of the subject.   |
|    | $R_2 = P749$  | <i>Parent organization</i> : parent organization of an organisation.                      |
| 16 | $R_1 = P127$  | <i>Owned by</i> : owner of the subject.   |
|    | $R_2 = P1830$ | <i>Owner of</i> : entities owned by the subject.  |
| 17 | $R_1 = P127$  | <i>Owned by</i> : owner of the subject.   |
|    | $R_2 = P3320$ | <i>Board member</i> : member(s) of the board for the organization.                        |
| 18 | $R_1 = P155$  | <i>Follows</i> : immediately prior item in a series of which the subject is a part.       |
|    | $R_2 = P155$  | <i>Follows</i> : immediately prior item in a series of which the subject is a part.       |
| 19 | $R_1 = P155$  | <i>Follows</i> : immediately prior item in a series of which the subject is a part.       |
|    | $R_2 = P355$  | <i>Subsidiary</i> : subsidiary of a company or organization.                              |
| 20 | $R_1 = P166$  | <i>Award received</i> : award or recognition received by a person, organisation.          |
|    | $R_2 = P166$  | <i>Award received</i> : award or recognition received by a person, organisation.          |
| 21 | $R_1 = P169$  | <i>Chief executive officer</i> : the CEO within an organization.                          |
|    | $R_2 = P112$  | <i>Founded by</i> : founder or co-founder of this organization.                           |
| 22 | $R_1 = P169$  | <i>Chief executive officer</i> : the CEO within an organization.                          |
|    | $R_2 = P127$  | <i>Owned by</i> : owner of the subject.   |
| 23 | $R_1 = P169$  | <i>Chief executive officer</i> : the CEO within an organization.                          |
|    | $R_2 = P169$  | <i>Chief executive officer</i> : the CEO within an organization.                          |
| 24 | $R_1 = P169$  | <i>Chief executive officer</i> : the CEO within an organization.                          |
|    | $R_2 = P3320$ | <i>Board member</i> : member(s) of the board for the organization.                        |
| 25 | $R_1 = P199$  | <i>Business division</i> : divisions of this organization.                                |
|    | $R_2 = P355$  | <i>Subsidiary</i> : subsidiary of a company or organization.                              |
| 26 | $R_1 = P306$  | <i>Operating system</i> : operating system (OS) on which a software works.                |
|    | $R_2 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
| 27 | $R_1 = P355$  | <i>Subsidiary</i> : subsidiary of a company or organization.                              |
|    | $R_2 = P127$  | <i>Owned by</i> : owner of the subject.   |
| 28 | $R_1 = P355$  | <i>Subsidiary</i> : subsidiary of a company or organization.                              |
|    | $R_2 = P155$  | <i>Follows</i> : immediately prior item in a series of which the subject is a part.       |
| 29 | $R_1 = P355$  | <i>Subsidiary</i> : subsidiary of a company or organization.                              |
|    | $R_2 = P199$  | <i>Business division</i> : divisions of this organization.                                |
| 30 | $R_1 = P355$  | <i>Subsidiary</i> : subsidiary of a company or organization.                              |
|    | $R_2 = P355$  | <i>Subsidiary</i> : subsidiary of a company or organization.                              |
| 31 | $R_1 = P361$  | <i>Part of</i> : object of which the subject is a part.                                   |
|    | $R_2 = P361$  | <i>Part of</i> : object of which the subject is a part.                                   |
| 32 | $R_1 = P366$  | <i>Use</i> : main use of the subject.   |
|    | $R_2 = P31$   | <i>Instance of</i> : that class of which this subject is a particular example and member. |
| 33 | $R_1 = P400$  | <i>Platform</i> : platform for which a work was developed or released.                    |
|    | $R_2 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
| 34 | $R_1 = P452$  | <i>Industry</i> : industry of company or organization.                                    |
|    | $R_2 = P31$   | <i>Instance of</i> : that class of which this subject is a particular example and member. |
| 35 | $R_1 = P452$  | <i>Industry</i> : industry of company or organization.                                    |

|    |               |   |
|----|---------------|---|
|    | $R_2 = P452$  | <i>Industry</i> : industry of company or organization.                                    |
| 36 | $R_1 = P452$  | <i>Industry</i> : industry of company or organization.                                    |
|    | $R_2 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
| 37 | $R_1 = P452$  | <i>Industry</i> : industry of company or organization.                                    |
|    | $R_2 = P2770$ | <i>Source of income</i> : source of income of an organization or person.                  |
| 38 | $R_1 = P463$  | <i>Member of</i> : organization or club to which the subject belongs.                     |
|    | $R_2 = P463$  | <i>Member of</i> : organization or club to which the subject belongs.                     |
| 39 | $R_1 = P749$  | <i>Parent organization</i> : parent organization of an organisation.                      |
|    | $R_2 = P127$  | <i>Owned by</i> : owner of the subject.   |
| 40 | $R_1 = P749$  | <i>Parent organization</i> : parent organization of an organisation.                      |
|    | $R_2 = P1830$ | <i>Owner of</i> : entities owned by the subject.  |
| 41 | $R_1 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
|    | $R_2 = P31$   | <i>Instance of</i> : that class of which this subject is a particular example and member. |
| 42 | $R_1 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
|    | $R_2 = P121$  | <i>Item operated</i> : equipment, installation or service operated by the subject.        |
| 43 | $R_1 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
|    | $R_2 = P306$  | <i>Operating system</i> : operating system (OS) on which a software works.                |
| 44 | $R_1 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
|    | $R_2 = P400$  | <i>Platform</i> : platform for which a work was developed or released.                    |
| 45 | $R_1 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
|    | $R_2 = P452$  | <i>Industry</i> : industry of company or organization.                                    |
| 46 | $R_1 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
|    | $R_2 = P1056$ | <i>Product or material produced</i> : material or product produced by an agency.          |
| 47 | $R_1 = P1344$ | <i>Participant of</i> : event a person or an organization was a participant in.           |
|    | $R_2 = P1344$ | <i>Participant of</i> : event a person or an organization was a participant in.           |
| 48 | $R_1 = P1830$ | <i>Owner of</i> : entities owned by the subject.  |
|    | $R_2 = P127$  | <i>Owned by</i> : owner of the subject.   |
| 49 | $R_1 = P1830$ | <i>Owner of</i> : entities owned by the subject.  |
|    | $R_2 = P749$  | <i>Parent organization</i> : parent organization of an organisation.                      |
| 50 | $R_1 = P2770$ | <i>Source of income</i> : source of income of an organization or person.                  |
|    | $R_2 = P452$  | <i>Industry</i> : industry of company or organization.                                    |
| 51 | $R_1 = P3320$ | <i>Board member</i> : member(s) of the board for the organization.                        |
|    | $R_2 = P127$  | <i>Owned by</i> : owner of the subject.   |
| 52 | $R_1 = P3320$ | <i>Board member</i> : member(s) of the board for the organization.                        |
|    | $R_2 = P169$  | <i>Chief executive officer</i> : the CEO within an organization.                          |

## ACKNOWLEDGMENTS

This research is part of NExT++ project, supported by the National Research Foundation, Prime Minister's Office, Singapore under its IRC@Singapore Funding Initiative.

## REFERENCES

- [1] Ayodele Ariyo et al. Adebisi. 2014. Comparison of ARIMA and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics* 2014 (2014).
- [2] Charu C Aggarwal and Chandan K Reddy. 2013. *Data clustering: algorithms and applications*. CRC press.
- [3] John Alberg and Zachary C Lipton. 2017. Improving Factor-Based Quantitative Investing by Forecasting Company Fundamentals. *arXiv preprint arXiv:1711.04837* (2017).
- [4] Wei et al. Bao. 2017. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one* 12, 7 (2017), e0180944.
- [5] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*. 2787–2795.
- [6] Joan et al. Bruna. 2014. Spectral networks and locally connected networks on graphs. In *ICLR*.
- [7] Kyunghyun et al. Cho. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).

- [8] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *NIPS*. 3844–3852.
- [9] Matthew et al. Dixon. 2016. Classification-based financial markets prediction using deep neural networks. *Algorithmic Finance Preprint* (2016), 1–11.
- [10] Claire Donnat, Marinka Zitnik, David Hallac, and Jure Leskovec. 2017. Spectral Graph Wavelets for Structural Role Similarity in Networks. *arXiv preprint arXiv:1710.10321* (2017).
- [11] Fuli et al. Feng. 2017. Computational social indicators: a case study of chinese university ranking. In *SIGIR*. ACM, 455–464.
- [12] Christoph Goller and Andreas Kuchler. 1996. Learning task-dependent distributed representations by backpropagation through structure. In *International Conference on Neural Networks*, Vol. 1. IEEE, 347–352.
- [13] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In *ICASSP*. IEEE, 6645–6649.
- [14] David K Hammond, Pierre Vandergheynst, and Rémi Gribonval. 2011. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis* 30, 2 (2011), 129–150.
- [15] Xiangnan et al. He. 2017. Neural collaborative filtering. In *WWW*. International World Wide Web Conferences Steering Committee, 173–182.
- [16] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [17] Ziniu et al. Hu. 2018. Listening to Chaotic Whispers: A Deep Learning Framework for News-oriented Stock Trend Prediction. In *WSDM*. ACM, 403–412.
- [18] Shan et al. Jiang. 2016. Learning query and document relevance from a web-scale click graph. In *SIGIR*. ACM, 185–194.
- [19] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [20] Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. *ICLR* (2017).
- [21] Alex et al. Krizhevsky. 2012. Imagenet classification with deep convolutional neural networks. In *NIPS*. 1097–1105.
- [22] B Shravan Kumar and Vadlamani Ravi. 2016. A survey of the applications of text mining in financial domain. *Knowledge-Based Systems* 114 (2016), 128–147.
- [23] Qing et al. Li. 2016. A tensor-based information framework for predicting the stock market. *TOIS* 34, 2 (2016), 11.
- [24] Andrew W Lo and A Craig MacKinlay. 2002. *A non-random walk down Wall Street*. Princeton University Press.
- [25] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, Vol. 30. 3.
- [26] Tao Mei, Yong Rui, Shipeng Li, and Qi Tian. 2014. Multimedia search reranking: A literature survey. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 38.
- [27] Gerald L Musgrave. 1997. A Random Walk Down Wall Street. *Business Economics* 32, 2 (1997), 74–76.
- [28] Arman Khadjeh et al. Nassirtoussi. 2014. Text mining for market prediction: A systematic review. *Expert Systems with Applications* 41, 16 (2014), 7653–7670.
- [29] Thien Hai Nguyen and Kiyoaki Shirai. 2015. Topic modeling based sentiment analysis on social media for stock market prediction. In *ACL*, Vol. 1. 1354–1364.
- [30] Adi et al. Omari. 2016. Novelty Based Ranking of Human Answers for Community Questions. In *SIGIR*. ACM, 215–224.
- [31] Lawrence et al. Page. 1999. *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford InfoLab.
- [32] G Preethi and B Santhi. 2012. STOCK MARKET FORECASTING TECHNIQUES: A SURVEY. *Journal of Theoretical & Applied Information Technology* 46, 1 (2012).
- [33] G William Schwert. 2002. Stock volatility in the new millennium: how wacky is Nasdaq? *Journal of Monetary Economics* 49, 1 (2002), 3–26.
- [34] Richard et al. Socher. 2013. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*. 926–934.
- [35] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. 2015. Unsupervised learning of video representations using lstms. In *ICML*. 843–852.
- [36] Wenting et al. Tu. 2016. Investment recommendation using investor opinions in social media. In *SIGIR*. ACM, 881–884.
- [37] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.
- [38] Markus et al. Weimer. 2007. COFIRANK Maximum Margin Matrix Factorization for Collaborative Ranking. In *NIPS*. 1593–1600.
- [39] Yan Xu and Guosheng Zhang. 2015. Application of Kalman Filter in the Prediction of Stock Price. In *KAM*.
- [40] Rui Yan, Yiping Song, and Hua Wu. 2016. Learning to respond with deep neural networks for retrieval-based human-computer conversation system. In *Proceedings of the 39th ACM SIGIR*. ACM, 55–64.



- [41] Rose Yu, Huida Qiu, Zhen Wen, ChingYung Lin, and Yan Liu. 2016. A survey on social media anomaly detection. *SIGKDD* 18, 1 (2016), 1–14.
- [42] Liheng et al. Zhang. 2017. Stock Price Prediction via Discovering Multi-Frequency Trading Patterns. In *SIGKDD*. ACM, 2141–2149.
- [43] Yao et al. Zhang. 2017. Learning Node Embeddings in Interaction Graphs. In *CIKM*. ACM, 397–406.
- [44] Sendong et al. Zhao. 2017. Constructing and embedding abstract event causality networks from text snippets. In *WSDM*. ACM, 335–344.
- [45] Zhaohui et al. Zheng. 2007. A regression framework for learning ranking functions using relative relevance judgments. In *SIGIR*. ACM, 287–294.
- [46] Denny et al. Zhou. 2007. Learning with hypergraphs: Clustering, classification, and embedding. In *NIPS*. 1601–1608.
- [47] Xiaojin et al. Zhu. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*. 912–919.