
Exploring Graph Neural Networks for Stock Market Predictions with Rolling Window Analysis

Daiki Matsunaga *

IBM Japan
19-21 Nihonbashi, Hakozaiki-cho, Chuo-ku
Tokyo, Japan
e36957@jp.ibm.com

Toyotaro Suzumura *

IBM T.J. Watson Research Center
1101 Kitchawan Rd, Yorktown Heights
New York, USA
tsuzumura@us.ibm.com

Toshihiro Takahashi

IBM Research - Tokyo
19-21 Nihonbashi, Hakozaiki-cho, Chuo-ku
Tokyo, Japan
e30137@jp.ibm.com

Abstract

Recently, there has been a surge of interest in the use of machine learning to help aid in the accurate predictions of financial markets. Despite the exciting advances in this cross-section of finance and AI, many of the current approaches are limited to using technical analysis to capture historical trends of each stock price and thus limited to certain experimental setups to obtain good prediction results. On the other hand, professional investors additionally use their rich knowledge of inter-market and inter-company relations to map the connectivity of companies and events, and use this map to make better market predictions. For instance, they would predict the movement of a certain company's stock price based not only on its former stock price trends but also on the performance of its suppliers or customers, the overall industry, macroeconomic factors and trade policies. This paper investigates the effectiveness of work at the intersection of market predictions and graph neural networks, which hold the potential to mimic the ways in which investors make decisions by incorporating company knowledge graphs directly into the predictive model. **The main goal of this work is to test the validity of this approach across different markets and longer time horizons for backtesting using rolling window analysis.** In this work, we concentrate on the prediction of individual stock prices in the Japanese Nikkei 225 market over a period of roughly 20 years. For the knowledge graph, we use the Nikkei Value Search data, which is a rich dataset showing mainly supplier relations among Japanese and foreign companies. **Our preliminary results show a 29.5% increase and a 2.2-fold increase in the return ratio and Sharpe ratio, respectively, when compared to the market benchmark, as well as a 6.32% increase and 1.3-fold increase, respectively, compared to the baseline LSTM model.**

*Equal contribution.

1 Introduction

1.1 Traditional investment strategies

Traditionally, investors and analysts in the finance industry have focused on two major approaches in predicting the movement of stock prices: fundamental analysis and technical analysis [1]. The former focuses on indicators which attempt to quantify the intrinsic value of a company, such as macroeconomic analysis (e.g. Gross Domestic Product, Consumer Price Index), industry analysis (i.e. industry-wide trends) and company analysis (e.g. Price-to-Earnings ratio) in order to predict the growth potential of the company in question. On the other hand, technical analysis focuses on various price movement indicators such as momentum, volatility and moving averages, all of which rely upon past trends as future indicators for each respective company [2]. This latter approach presupposes that all of the necessary information regarding each stock in question are included in its stock price. The difference in these two approaches stem from the longstanding debate about whether or not stock markets are inherently predictable and thus profitable, either in the short-term or in the long-term. However, there is a commonality in both approaches in that they use historical indicators to predict each individual stock in question, which is especially the case for most machine learning approaches. Indeed, much of the work in machine learning fits into the category of technical analysis as it is a natural way to formulate stock price prediction as a sequential modeling task [2].

1.2 An increasingly interconnected financial world

In reality, however, the world is increasingly more interconnected through the proliferation of the Internet, the rise of Fintech and global banking conglomerates operating across borders. In this complex financial world, stock prices, market indices, and other crucial indicators work and move in tandem, whereby a multitude of political and economic factors can affect each price. We need to look no further than the 2008 financial crisis to understand this interconnectivity intuitively and understand that individual prices don't move solely from past trends in isolation [3]. As a sidenote, this phenomena makes it increasingly more difficult to make sound investment decisions based solely on past trends from a few predictors, and thus more susceptible to risky investments or even fraud on a massive scale. For instance, the case of Valeant Pharmaceuticals is a good example of a massive fraud case which appeared to have been a sound investment choice if focusing on its linearly soaring stock price trend without understanding the full context of its suspicious business model of acquiring many drug companies and racking up their drug prices for profit. It was a very popular and attractive choice among investors, before it lost north of 60% of its original stock value in 2015, leading to huge losses for both individual and institutional investors. Although this case mostly highlights the malintent of some investors and executives, it also shows that the financial world is quickly becoming too complex for any one person to have a holistic view of the market or dive deeper into the multitude of factors that affect a particular index or stock price [4].

1.3 Graph-based machine learning in the new era of finance

From a machine learning perspective, network structures such as company knowledge graphs were difficult to incorporate due to its non-Euclidean nature. Thus, most statistical approaches, even with recent deep learning methods, did not incorporate interconnectivity into the predictive model. Furthermore, graph analytics traditionally involved hand-engineering features and summary graph statistics using graph algorithms such as PageRank, egonet, degree distribution, among others [5][6]. In stock market prediction, technical analysis falls under this umbrella whereby some of the aforementioned indicators such as moving averages are hand-engineered from raw price data and fed into a classifier as features. This approach, however, is both time-consuming and inflexible, especially in representing the dynamic nature of global financial markets. One possible remedy to these issues is the recent innovation in graph neural network algorithms. These algorithms present an opportunity to make predictive models which better represent the increasingly fast-paced and interconnected nature of the global financial system by incorporating knowledge graphs. This type of research frames the feature extraction process as a learning task rather than a preprocessing step, as is the case with graph summary statistics approaches. We posit that the financial system can best be represented in this network structure where companies are connected through various relation types such as supplier/customer, shareholder and industries. Similar to how a professional investor makes decisions, graph neural networks can utilize the network structure to incorporate the interconnectivity

of the market and make better stock price predictions, rather than relying solely on the historical stock prices of each individual company or on hand-crafted features. Despite its promise, however, the use of these algorithms to the stock market prediction problem has only recently started to be explored.

The key contribution of this paper is as follows.

- We empirically evaluate the effectiveness of graph neural networks on stocks listed in the Nikkei 225 market.
- We conduct backtesting over a period of roughly 20 years (4,632 timesteps) to test the generalizability of graph neural networks as a method for predicting stock prices
- We conduct backtesting using rolling window analysis, which is an effective method for evaluating time series data, especially for historical stock price data.

The remainder of the paper is organized as follows. In Section 2, we review related work in the use of machine learning approaches to making stock market predictions. In section 3, we detail the algorithm and methodology used in our work. In Section 4 we outline the datasets used, evaluation methods and the analysis of preliminary results. We finally discuss some future extensions of this work in Section 5 and conclude in Section 6.

2 Related work

2.1 Deep learning for technical analysis

There have been a number of works in the use of statistical and machine learning methods for analyzing past stock trends. This problem can be framed as either a classification task e.g. UP, DOWN, STAY, or a regression task e.g. prediction of next day’s closing price [1]. A number of approaches [7] [8] have shown that deep learning methods outperform traditional machine learning approaches which use hand-crafted features. [8] combines autoencoders and Long-Short Term Memory (LSTM) networks in order to predict the indices of six different markets across the world and has shown its predictive capabilities across various developing and developed markets. The State Frequency Memory model in [7] extends the LSTM model with inspiration from the signal processing community whereby the model attempts to decompose stock price trends into multiple latent frequency components. This approach outperforms linear Autoregressive (AR) models in the prediction of 50 stocks over a period of 9 years. Although the architectural details of the models vary, they show that deep learning approaches are effective in finding non-linear patterns in the stock market.

2.2 Natural language processing approaches

The use of natural language processing (NLP) approaches have been investigated as another interesting way to incorporate natural language text such as SEC filings [9], patents [10] or financial news [11]. Features such as n-grams can be extracted [9], or word embedding approaches such as Word2vec can be used as an unsupervised feature extractor [7]. The Hybrid Attention Network model in [7] uses Word2vec to extract embeddings for each word in news articles, and aggregates the embeddings to output a fixed size vector that represents the features for each news article. These features can then be fed into a supervised classifier similar to how numerical data would be used as inputs.

2.3 Complex networks and network science

A related area of research in stock prediction using graphs comes from the network science community, especially with the study of complex networks. Complex networks study a specific type of random graphs which can broadly be categorized as small-world or scale-free. Small-world networks randomly connect two distant nodes, shortening the diameter of a graph. Scale-free networks follow a power-law connectivity distribution, where a small number of hub nodes have a large number of connections [12]. In the stock prediction setting, complex networks are used to study the major influencer companies (network centrality) or clustered communities (modularity), without specifying any explicit edge relationships. For example, [13] makes random graphs at each time interval by connecting companies through the mutual information of price movements, in order to predict the S&P 500 index. For each random graph, they extract graph information such as the node

strength distributions, network centrality and modularity. They show that incorporating these graph characteristics can improve upon ARIMA models, in predicting the one-minute interval fluctuations of the S&P 500 market index.

2.4 Company knowledge graphs for market prediction

Recently, there have been interesting works [14][15] which incorporate open source knowledge graphs (e.g. Wikidata) into their stock prediction models for individual companies by combining graph neural networks and sequential models such as Recurrent Neural Networks (RNN). Both works follow four major steps with some differences in the algorithmic details. First, historical stock price data are fed into an RNN layer as features to output node embeddings for each stock in question. The features can be a combination of the previous day’s closing price and moving averages [14] or price change rates [15]. These features are used to capture the short term and/or mid term trend of each stock price. Second, an adjacency matrix is created through open knowledge graph sources such as Wikidata, to connect companies through industry, subsidiary, people and product relations. Third, the node (company) embeddings created by the RNN layer and the adjacency matrix are combined and fed into a graph neural network layer to update the node embeddings. Finally, the updated embeddings are fed into a fully connected layer to make price predictions. Furthermore, [15] feeds the updated node embeddings into a separate graph pooling layer to predict the movement of market indices e.g. S&P 500 Index, making the overall architecture a multi-task one. We revisit the algorithm in [14] in more detail in Section 3.

3 Generalizability of GNNs for stock prediction

In this paper we examine the following research question:

Are knowledge graph datasets e.g. company relations data, useful in making stock predictions across different markets and longer time horizons?

To answer this question, we have adopted and extended the work in [14] and the corresponding open source codebase.

3.1 Graph convolutional network

The basis for the work in [14] is the Graph Convolutional Network (GCN) algorithm proposed in [16]. The algorithm requires as input (a) features for each node and (b) an adjacency matrix showing relations among nodes. The features for each node i are updated by using the given adjacency matrix to aggregate feature representations of its neighbors:

$$\bar{e}_i = \sum_j^N \left(\frac{\phi(W^T \mathbb{A}_{ij} + b)}{d_j} \times e_j \right) \quad (1)$$

where for N companies and K relation types, $\mathbb{A} \in \mathbb{R}^{N \times N \times K}$ is the adjacency matrix, d_j is the degree of node j acting as a normalization factor, $W \in \mathbb{R}^K$ and b are learnable parameters and ϕ is a non-linear activation. Along the diagonal of the adjacency matrix, \mathbb{A}_{ii} always equals 1 to ensure that a node’s own embeddings are being aggregated as well. The updated embeddings can be fed into either additional GCN layers for further neighborhood sampling or into a fully connected layer for the specific learning task such as classification or regression. For the stock market prediction setting, the features for node e_i can be the output of an RNN layer, where the sequential data is mapped into a vector of fixed length i.e.

$$e_i = RNN(\chi_i) \quad (2)$$

where χ_i is the features calculated from historical stock price data such as closing price or moving averages. Equation 2 is calculated for each relevant stock at each timestep. In this work, we follow the algorithm in [14] and use an LSTM layer rather than the vanilla RNN layer. As for hyperparameters, we followed the hyperparameters used in [14] to train on NYSE stock prices, and set the number of LSTM units at 32 and the sequence length at 8 timesteps.

3.2 Temporal graph convolution

The implementation of GCNs, while able to incorporate graph data into the predictive model, only uses a static version of the adjacency matrix. In other words, the weighing factor which updates e_i to \bar{e}_i is static across time. In reality, however, the propagation of stock trends among connected companies evolves dynamically. For instance, the stock price between car manufacturer Company X and car parts supplier Company Y might be highly correlated as a default, but there could be instances where, for example, a new car model release could strengthen this relationship to an even greater degree. **To incorporate this time-sensitivity into the relations among nodes (companies), the dot product of the embeddings for nodes i and j are multiplied as a weighing factor for the relationship strength.**

$$\bar{e}_i = \sum_j^N (e_i^T e_j \times \frac{\phi(W^T \mathbb{A}_{ij} + b)}{d_j}) \times e_j \quad (3)$$

Intuitively, the dot product represents the similarity between the embeddings of nodes i and j ; the more similar the recent trends in stock prices, the stronger the relation strength will be. Again, equation 3 is calculated for each relevant company at each timestep. These embeddings are then fed into a fully-connected layer for regression.

As one minor adjustment, as opposed to the implementation for the paper optimizing the RNN layers and TGC layers separately, we’ve instead optimized the whole network in an end-to-end manner. This is to avoid unnecessary bugs and speed up training [17], especially in a relatively simple network with one RNN layer, one TGC layer and one fully connected layer.

3.3 Rolling window analysis

In the financial sector, professionals are often concerned with models which reliably can predict the market over long time horizons. Therefore, we have implemented the rolling window analysis method [18] for splitting the full dataset into the training and test set. We use 2000 timesteps for training and 200 for testing as a fixed window, and slide that window over the full number of timesteps. For example, index 0 to 2000 and 2000 to 2200 will be used in the first iteration for training and testing, respectively. The second iteration will have indices 200 to 2200 and 2200 to 2400, and so on. This method not only ensures that the model is generalizeable across long time horizons but also represents the idea that recent stock prices have better predictive power than older ones. For instance, in order to make future predictions for a particular stock, it makes more intuitive sense to use the historical prices from the last five years rather than from the 1980s. As a further note, it is also possible to implement the growing window method, whereby the window size for the training set continues to increase while the test set window continues to slide. For instance, the first iteration contains indices 0 to 2000 as the training set, the second iteration contains 0 to 2200, and so on. This method can be useful especially in cases where the data is sparse with wider time intervals or if there are difficulties in obtaining a comprehensive number of timesteps for backtesting.

4 Evaluation

4.1 Historical stock price data

We test our model on Japanese companies listed in the Nikkei 225 market. The number of companies, however, is filtered down to 176 from 225 listed companies. This is to ensure that there are enough timesteps to conduct long-term backtesting, since some companies were either founded or listed only recently and thus contains fewer timesteps. For features to be fed into the LSTM layer, we follow [14] and calculate 5, 10, 20, 30-day moving averages at each timestep in addition to the adjusted closing prices. We also calculate the 1-day return ratio at timestep t , $R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$ as the ground truth, where p_t represents the price at timestep t .

Table 1: An example company relation data for a particular Company X

Relation Type	Company Name	Sector	Industry	Country
Supplier	Company A	Producer Manufacturing	Auto Parts: OEM	JPN
Customer	Company B	Finance	Finance/Retail/Leasing	USA
Partner	Company C	Consumer Durables	Motor Vehicles	USA

Table 2: Knowledge graph summary

# of Nodes	12, 473
# of Edges	38, 252
Average Node Degree	6.13

4.2 Knowledge graph dataset

We used the Nikkei Value Search dataset, which consists of supplier, customer, partner and shareholder relations for each of the 176 relevant companies, as shown in Table 1. This data is then used to create a comprehensive knowledge graph which includes companies outside of Nikkei 225 index as well as Japan. The summary statistics for this comprehensive knowledge graph is shown in Table 2. The nodes have a "node_type" attribute with values "company", "sector", "industry" or "country", while edges have an "edge_type" attribute with values "in-country", "in-industry", "in-sector", "parent-company-of", "partner-with", "related-to", "same-company", or "shareholder". The nodes also have a boolean attribute called "in_nikkei" to distinguish the nodes which are listed in the Nikkei 225 market. Also, "same-company" relations were used for entity resolution purposes, where English and Japanese names for the same company were concatenated as identical nodes. These preprocessing steps allow for the extraction of first-order and second-order relations, listed in Table 3. For example, if two nodes listed in the Nikkei 225 market both had an "in-industry" edge connected to a common "industry" node, the two nodes can be connected via second-order connections.

First-order relations are direct relations among the 176 companies. This shows an apparent relationship where connected stocks could be related. However, these relationships are quite sparse and also an area where investors are already actively using for analyzing individual stocks. To alleviate the sparsity problem and make better use of the capabilities of graph neural networks, we've extracted second-order relations, which are indirect connections between any two of the 176 companies in question. The result of the preprocessing step are seven different types of adjacency matrices, $\mathbb{A} \in \mathbb{R}^{N \times N}$ where N is the number of companies. We've also tested a relation type called "all" where we stack the seven adjacency matrices into a three dimensional matrix, $\mathbb{A} \in \mathbb{R}^{N \times N \times K}$ where K is the number of relation types.

4.3 Evaluation Metrics

As an evaluation metric, we use the return ratio and the Sharpe ratio, as they are both used frequently in the finance industry. The return ratio R_t is calculated as $\frac{p_t - p_{t-1}}{p_{t-1}}$, which is simply the percentage increase in the stock price or the return when we buy a stock at timestep $t - 1$ and sell at timestep t .

Table 3: First- and second-order relations extracted from the created company knowledge graph

Relation Type	# of edges	Description
supplies-from	145	First-order relation between companies connected by Supplier relation
customer-of	145	First-order relation between companies connected by Customer relation
partner-with	555	First-order relation between companies connected by Partner relation
shares-owned-by	220	First-order relation between companies connected by Shareholder relation
common-industry	438	Second-order relation between companies with common industry
common-customer	3634	Second-order relation between companies with common customers
common-supplier	8262	Second-order relation between companies with common supplier

Table 4: Performance comparison among various graph types against LSTM and Buy-Hold benchmarks

Edge Type	Return Ratio (Yearly Rates)	Sharpe Ratio
Benchmark (Buy-Hold)	4.30%	0.16
LSTM	23.63%	0.29
supplier-of	21.10%	0.30
customer-of	27.52%	0.35
partner-with	15.48%	0.12
shares-owned-by	23.54%	0.22
common-industry	18.09%	0.24
common-customer	18.61%	0.17
common-supplier	16.27%	0.14
all	29.95%	0.37

Table 5: Return ratio for the prediction of various timesteps (customer-of)

# of time steps	Return Ratio (Yearly Rates)
1-day	27.52%
5-day	21.93%
10-day	23.327%
20-day	17.36%

The Sharpe ratio is calculated as $\frac{R_t - R_f}{\sigma}$ where the return R_t is subtracted by the risk-free rate R_f which is then divided by the standard deviation of the returns σ . The standard deviation σ represents the risk associated with a stock, which is important in avoiding the evaluation of models based purely on returns. It is important to note that the Sharpe Ratio should not be evaluated as an absolute value, but rather viewed as a performance indicator in comparison with the benchmark.

4.4 Evaluation Results

Our preliminary results in Table 4 shows the following results.

- (a) Both the LSTM model and the various graph neural network models outperform the market benchmark.
- (b) The LSTM model is outperformed by graph neural networks for "customer-of" and "all" in terms of both the Sharpe and return ratio, and is outperformed by "supplier-of" in terms of the Sharpe ratio.
- (c) "all" relations have the best performance based on both the return ratio and Sharpe ratio.
- (d) "customer-of" relations had a relatively high return and Sharpe ratio despite its sparsity, which suggests that customer relations are highly effective in improving predictive performance.

The results for (d) makes intuitive sense since supply chain analysis is an important industry practice among investors. However, it is important to note that while customer relations seemed to be effective in the aggregate, its effectiveness varies by time periods. For example, [14] concluded that partner relationships e.g. product collaborations were the most effective relation type for the NASDAQ market. As the authors mention, it does make intuitive sense for companies who collaborated on a common product to have connected revenue. However, the low performance of the partner relations in this study suggest three possible theories. First, since the testing period for the paper in [14] was limited to 200 timesteps between 2017 and 2018, the results were inconclusive. Second, partner relations could indeed have high predictive power in the NASDAQ market even with extended backtesting, but not generalizable to the Nikkei 225 market. Finally, it could also be the case that since work in [14] used a Wikipedia-based knowledge graph, the obtained partner relationships were filtered to very important or well-known ones, while the Nikkei Value Search dataset was more comprehensive.

Figure 1: Comparison of stock price movements between Company A and its customer: Company B.



4.5 Qualitative Analysis

The nature of the graph data also allows us to investigate post-hoc whether or not related companies have similar stock trends. For example Figure 1 shows a snapshot of a stock trend chart for two companies of which the model predicted strong edge relations and are connected by customer relations. In this scenario, Company B (Company A's customer) has a trough point on 6/4/2009 and Company A has a similar lagging trough point on 6/8/2009, which suggests that the stock price of a customer contributed to the prediction of stock prices for the target company. It is also important to note that the golden cross point has a time lag in a similar fashion. The golden cross occurs at the intersection of the mid-term (30-day) moving average and the closing price, and this indicates a bullish sentiment as the closing price crosses the moving average on an upward trend. As previously mentioned, however, the effectiveness of customer relations don't always hold across all time-spans, and other factors can affect each stock price in different directions and magnitudes. Nonetheless, the intuitive nature of knowledge graph data holds interesting opportunities in terms of explainability of stock predictions.

5 Possible extensions of current work

The scope of this paper was limited to the prediction of 1-day returns for 176 companies listed in the Nikkei 225 market. However, much of the finance industry is concerned with longer prediction timesteps due to the attractiveness of lower transaction fees in passive funds. Therefore, it can be worthwhile to explore the different relation types with different prediction timesteps to find an effective long-term predictive model. With the current model, the direct customer relation proved to be an effective indicator only for 1-day predictions, where the effectiveness wanes as the timesteps increase, as shown in Table 5. However, it is possible that the accuracy for predictions of longer time horizons will increase as we increase the number of hops. For example, the price of a customer two hops away can be a strong indicator for predicting the 5 or 10-day future returns of the target company, if the correlations propagate. Of course, the prediction of longer time horizons becomes more and more difficult as the timesteps increase due to the multivariate and non-linear nature of the stock market leading to higher error propagation [19]. In this scenario, it can also be worthwhile to incorporate other information such as macroeconomic data or news articles rather than solely relying on supply chain relationships, which can be regarded as an important predictor variable in a family of factors which affect the market, rather than the sole indicator. Finally, it is also worthwhile to explore the method of feeding the entire knowledge graph, rather than extracting relations only among the relevant companies listed in the Nikkei 225 market. For instance, if a company X listed in the Nikkei 225 had a relation with an American Company Y, we were only able to connect other listed companies if they also had the same relation with Company Y. This will be possible once we can collect stock price data for Company Y and other non-listed companies.

6 Conclusion

In this paper, we have conducted preliminary experiments on the effectiveness of incorporating company relations knowledge graphs in predicting the stock market. We have shown that certain relations (customer-of) are effective predictors individually as well as combined with other weaker

relations. Our preliminary results suggest that knowledge graph data and graph neural networks holds strong promise in creating a more generalizable and practical stock market prediction mechanism. Furthermore, there are many opportunities for extending this work to the field of model interpretability and explainability as the intuitive nature of graph data makes it easier to visualize results. We encourage the finance and AI research community to further explore the use of graph neural networks for the challenging problem of making accurate market predictions and as a consequence, improved investment decisions.

Acknowledgments

We’d like to thank JLB and the GBS Vision Fund (IBM Japan) for supporting this work. We’d also like to thank Yoshiki Minowa, Tomokazu Takeda and Tomohiro Hayashi for many useful discussions and advice.

References

- [1] E. Beyaz, F. Tekiner, X. Zeng, and J. Keane. Comparing technical and fundamental indicators in stock price forecasting. In *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pages 1607–1613, June 2018.
- [2] Dev Shah, Haruna Isah, and Farhana Zulkernine. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *International Journal of Financial Studies*, 7(2):1–22, May 2019.
- [3] OECD. ‘the 2008 financial crisis – a crisis of globalisation?’, in economic globalisation: Origins and consequences. *OECD Publishing, Paris*, 2013.
- [4] Hugh Grove and Mac Clouse. Forensic accounting procedures applied to valeant: Where were the gatekeepers? *Journal of Forensic and Investigative Accounting*, 9(2):836–848, 2017.
- [5] Mark Weber, Jie Chen, Toyotaro Suzumura, Aldo Pareja, Tengfei Ma, Hiroki Kanezashi, Tim Kaler, Charles E. Leiserson, and Tao B. Schardl. Scalable graph learning for anti-money laundering: A first look. *CoRR*, abs/1812.00076, 2018.
- [6] William L. Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. *IEEE Data Eng. Bull.*, 40:52–74, 2017.
- [7] Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM ’18*, pages 261–269, New York, NY, USA, 2018. ACM.
- [8] Wei Bao, Jun Yue, and Yulei Rao. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS ONE*, 12, 07 2017.
- [9] Heeyoung Lee, Mihai Surdeanu, Bill MacCartney, and Dan Jurafsky. On the importance of text analysis for stock price prediction. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, pages 1170–1175, Reykjavik, Iceland, May 2014. European Languages Resources Association (ELRA).
- [10] S. Wu, C. Tsao, P. Chang, C. Fan, M. Chen, and X. Zhang. A study of patent analysis for stock price prediction. In *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, pages 115–119, July 2017.
- [11] Adam Atkins, Mahesan Niranjan, and Enrico Gerding. Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science*, 4(2):120 – 137, 2018.
- [12] Edward Hearnshaw and Mark Wilson. A complex network approach to supply chain network theory. *International Journal of Operations & Production Management*, 33, 03 2013.
- [13] Minjun Kim and Hiroki Sayama. Predicting stock market movements using network science: An information theoretic approach. *CoRR*, abs/1705.07980, 2017.

- [14] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)*, 37(2):27, 2019.
- [15] Raehyun Kim, Chan Ho So, Minbyul Jeong, Sanghoon Lee, Jinkyu Kim, and Jaewoo Kang. Hats: A hierarchical graph attention network for stock movement prediction. *arXiv preprint arXiv:1908.07999*, 2019.
- [16] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *CoRR*, abs/1609.02907, 2016.
- [17] Tobias Glasmachers. Limits of end-to-end learning. *CoRR*, abs/1704.08305, 2017.
- [18] Eric Zivot and Jiahui Wang. Rolling analysis of time series. *Modeling Financial Time Series with S-PLUS®*, pages 313–360, 2006.
- [19] Rose Yu, Stephan Zheng, Anima Anandkumar, and Yisong Yue. Long-term forecasting using tensor-train rnns. *CoRR*, abs/1711.00073, 2017.