

Predicting Stock Price Movements via Multi-relational Inter-firm Networks

John Rios
University of Iowa
Iowa City, IA

john-riosrodriguez@uiowa.edu

Kang Zhao
University of Iowa
Iowa City, IA

kang-zhao@uiowa.edu

Nick Street
University of Iowa
Iowa City, IA

nick-street@uiowa.edu

Abstract

Predicting stock movements is challenging, but has attracted tremendous amount of attention from both practitioners and researchers. At the same time, firms are connected with each other in a multi-relational network that consists of multiple types of relationships. Using real-world supply and competition networks among more than one thousand firms, this research predicts a firm's stock movements by leveraging performance of its customers, suppliers, and competitors. We show that features based on network neighbors of a firm significantly contribute to the prediction of its future stock movements. Additional analyses revealed that suppliers' and competitors' performance is more indicative of a firm's stock movement than customers'.

1. Introduction

For many decades, predicting stock returns has attracted investors' interests. The study of such predictions has also led researchers to generate new theories about markets and economics. However, this is a challenging problem because one must find true predictive power beyond the common factors that the structure of the returns may have (Fama and French 1996) (Carhart 1997). Previous research has attempted to explain stock returns by analyzing not only historical financial data, but also additional external information like social media (Luo, et al. 2013) (Rechenthin, et al. 2013) (Nguyen, et al. 2015), search logs (Luo, et al. 2013) (Agarwal, et al. 2017), and supply chains (Hendricks and Singhal 2005) (Agarwal, et al. 2017), among others.

In today's business world, a firm is rarely just operating by itself. Instead, firms are connected with each other in a multi-relational network. A multi-relational network (a.k.a., multi-layer network) represents different types of relationships among the same set of nodes (Zhao, et al. 2016) (Rodriguez and Shinavier 2010). Among firms, one such relationship is captured by supply chain networks. Some supply chain characteristics, such as flexibility (Merschmann and Thonemann 2011) and relationships with trade partners (Vanichchinchai and Igel 2011) have an impact on a firm's performance. In the inter-connected business world, a supply chain system of buyers and sellers is better modeled as a network than as a linear or hierarchical chain, because the latter fails to capture the complexity of the relationships between companies (Kim, et al. 2011). Network analysis provides a unique perspective to depict the characteristics of supply chains (Kim, et al. 2011). Additionally, previous literature has found that, due to information flow across economically linked assets (Cohen and Frazzini 2008) (Menzly and Ozbas 2010), correlations exist between stock returns of a focal company and the lagged returns of its business partners. Then this association can be used to explain the stock return of the focal firm (Cohen and Frazzini 2008) (Menzly and Ozbas 2010) (Agarwal, et al. 2017).

Besides the inter-firm flow of goods or services via supply chain networks, another major type of relationship among firms is competition. As business partners', competitor's information of a focal firm also has an impact on its managerial decisions and performance. Gimeno (2004) found that the likelihood of a firm allying with its competitor's partners depends on the level of co-specialization of its competitor and its competitor's partner. The pattern of competitive actions of

a focal company has an impact not only on the company's performance but also on its competitors' performance (Ferrier and Lee 2002). On one hand, when a firm gains more market shares, its competitors may lose ground and suffer from deteriorating performance. On the other hand, firms competing with each other are often in the same or similar sectors. When a sector as a whole is on the rise or decline, firms in the sector may have similar stock movement. Therefore, stock movements of a focal firm's competitors may carry signals of the focal firm's own stock movement in the future.

While there have been many studies on predicting stock movement and analyzing inter-firm networks, little has been done to integrate multi-relational inter-firm networks into the prediction of stock movement. Thus, this study attempts to answer the following research questions: **Does the stock price movement of neighbors of a focal firm in a multi-relational inter-firm network help to predict the focal firm's stock price movement? If so, what type(s) of inter-firm network neighbors of a focal firm is more predictive of its stock price movement?** Answers to these questions will not only help investors predict stock price movement, but also contributes to the literature by illustrating at a more granular level the relationship among performance of organizations that are connected in a multi-relational network. The latter would provide managers of a focal firm with a better understanding of the impact that business partners have on the focal firm's performance, and enhance their decision making process.

To answer these questions, we built a multi-relational inter-firm network that consists of two networks: a supply chain network and the corresponding competition network. We represented a focal firm's network neighbors' stock movement with a fix-length vector, so that we can predict a focal firm's stock movement by learning a model from other firms' stock movement. Our experiments show that our approach improves the performance of stock prediction. We also ran additional regression analyses to better understand the different impact of business partners and competitors of a firm on its stock price movement.

2. Literature Review

This paper is related to previous literature on stock return cross-predictability as well as multi-relational and inter-firm networks.

2.1. Stock Return Prediction

Predicting stock movements has led to the generation of numerous theories and factors associated with the Capital Asset Price Model (CAPM). While CAPM explains the relationship between systematic risk and expected return for stocks, anomalies in this model offer an opportunity to make money. The extended Fama-French model provides three factors (market, small minus big, and high minus low) from Fama & French (1996) and the momentum factor from Carhart (1997), which explain the anomalies in the CAPM model. To prove true predictive power of new features, these factors are usually included as control variables in regressions for cross-sectional stock return prediction. Other variables like average daily return or price in previous time periods are also included as control variables or baseline models for stock return prediction in previous literature (Goyal and Welch 2003) (Nguyen, et al. 2015).

In addition to historical financial data, researchers have discovered that information outside a focal firm, such as social media and investor attention, can improve the prediction of stock returns. Luo, et al. (2013) found that social media metrics (Web blogs and consumer ratings) were significant at explaining future equity values, while conventional online behavioral metrics (Google searches and Web traffic) have weaker effects. However, their model is explanatory,

instead of predictive, in nature. Rechenhain, et al. (2013) used supervised learning algorithms to predict future stock price direction and found slight predictability in sentiment of posts of a stock in Yahoo Finance. Additionally, Nguyen, et al. (2015) took a step further by including not only the overall sentiments of a stock in message boards, but also topic-sentiment features, which represent sentiments about specific topics of a company.

2.2. Multi-relational and Inter-firm Networks

Networks as collections of nodes and edges (links between nodes) can be used to represent a variety of complex interconnected systems. Depending on the nature of the connections between their nodes, networks can be classified as *single-relational* or *multi-relational*. As opposed to single-relational networks, multi-relational networks explicitly contain at least two channels of connectivity between the same set of nodes. Each channel is defined by a separate set of edges representing different kinds of interactions (Rodriguez and Shinavier 2010) (Boccaletti, et al. 2014). In a multi-relational *inter-firm network*, relationships can vary from partnership to rivalry.

A network approach to analyze supply chains provides unique insights. Compared with the linear and hierarchical model for supply chains, a supply chain network consists of a system of inter-connected firms that engage in procurement, use, and transformation of raw materials to provide goods and services (Kim, et al. 2011) (Lamming, et al. 2000) (Harland, et al. 2001). Based on this approach, researchers started to adopt network analysis tools to study supply chains. For example, Kim et al. (2011) related key social network analysis metrics to the three automotive supply chain networks reported in Choi and Hong (2002), showing how to apply social network analysis to investigate the structural characteristics of supply chain networks. In addition, Hearnshaw and Wilson (2013) mirrored the properties of complex network models with real-world supply chain networks. The authors identified the key properties (properties of a “scale-free” network) for a supply chain network to be efficient. Zhao, et al. (2011) studied disruptions in supply chain networks from a topological perspective.

While supply chain networks represent partnership among firms, competition networks are about rivalry or competition. Yao, et al. (2007) built competition networks and study rivalrous behaviors. However, it was until Skilton and Bernardes (2015) that social network theory was shown to be suitable for settings where cooperative ties are absent. The authors also showed that the structure of the whole competition network of a firm, instead of just dyadic or triadic rivalry, impacts the firm’s competition behavior—expressed as the rate the firm enters new product markets. Researchers have also tried to leverage competition networks in the model of disruption propagations in supply chain networks (Zuo, et al. 2018).

Embedded in a multi-relational inter-firm network, a firm’s overall performance is associated with the performance of its competitors and business partners. Ferrier and Lee (2002) explored how the pattern of competitive actions that a focal company deploys over time has an impact not only on the stock price of the company but also on its competitors’ stock price. A firm’s own supply chain performance is related to its stock price (Hendricks and Singhal 2005), but its price is also related to the performance of the firm’s business partners. The lack of investor’s attention hinders the information diffusion in supply chains, leading to stock returns predictability. Cohen and Frazzini (2008) found that future monthly return of a focal company can be explained by lagged returns of its customers. According to Menzly and Ozbas (2010), at the sector level, the lagged performance of a firm’s supplier and customer sectors explains the future stock return of the focal firm. Finally, Agarwal, et al. (2017) measured the weekly online co-search attention that

investors pay to a focal firm and its business partners. The authors found that partners that have a low co-search attention explain future weekly stock returns of the focal company.

These findings motivate us to explore if we can do a better job predicting stock movement for individual firms by incorporating the stock performance of the firm’s neighbors in a multi-relational inter-firm network. Such improvements in the prediction suggest that the performance of a firm’s business partners holds valuable information that would be useful for managerial and investment decisions. Unlike most previous research, our prediction uses supervised learning techniques evaluated with out-of-sample data. We also use information about the daily stock return, which allows us to identify short-lived signals that previous methods would miss.

3. Data Collection and Networks Construction

Our study focused on predicting stock movement for 27 firms from 4 sectors. All of them are public companies traded in U.S. stock markets, as well as top companies according to Supply Chain Gartner Top 50 (Gartner Inc. 2015). Data was collected from Mergent Horizon (<http://www.mergentonline.com>) in June of 2016. This database provides companies’ basic information (e.g., Ticker Symbol, Industry Sector, and Country), financial data (e.g., Net Income and Revenue), lists of suppliers, lists of customers, lists of competitors, and product overlap with competitors, among others. Using the 27 firms as seeds, we scraped web pages for companies following a snowball sampling approach. We started by retrieving information about customers and suppliers of each seed company. Then, we parsed the web page of each of the companies identified as Tier-1 suppliers of the seed companies in order to include their customers and their suppliers (Tier-2 suppliers of the seed companies) into the analysis. Once the total set of companies in the supply chain network was identified, we collected the list of competitors of each of these companies to build our competition network. Figure 1 presents the structure of the web scraping process for the supply network (Figure 1(a)) and competition network (Figure 1(b)).

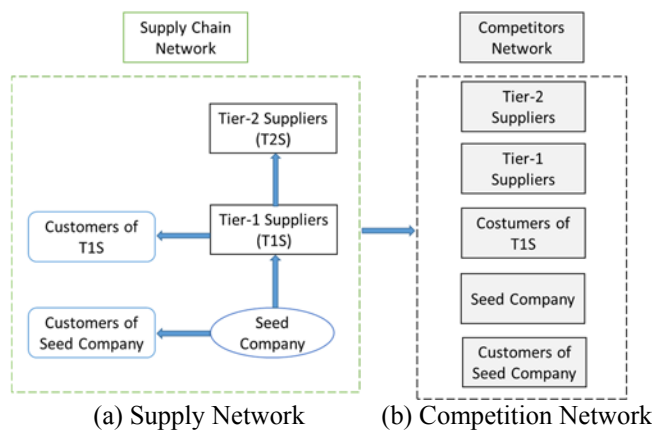


Figure 1. Structure of the complete web-scraping process for the two networks

With the data from Mergent Horizon, we constructed a multi-relational network with 1,150 nodes (firms) that combines the supply chain network of all the seed companies. Note that we excluded nodes whose stock data is unavailable, including non-public firms, and government agencies. The two networks share the same set V of nodes (firms) of size $|V| = 1,150$ and different sets of edges. The first one is the supply chain network where the edge $e_{i,j}$ represents a directed supply edge from v_i to v_j , which means v_i is a supplier of v_j . Because we collected data using a snowball sampling approach, there are no isolated nodes in this network. The second network is the competition network. It was built with exactly the same set of companies in the supply chain network, but the edges $e_{i,j}$ in this network indicate a competition relationship between v_i and v_j .

Figure 2 visualizes these two networks and Figure 3 presents the distribution of the total degree for both network. Both networks feature highly skewed degree distributions with many nodes with low degrees, but few nodes with very high degrees.

Table 1 presents a summary of the main network metrics including number of nodes, number of edges, average degree, diameter (maximum of the shortest path length between any two nodes), and size of the largest connected component (LCC). A connected component is a subnetwork with a path (edge directions ignored) between any pair of nodes.

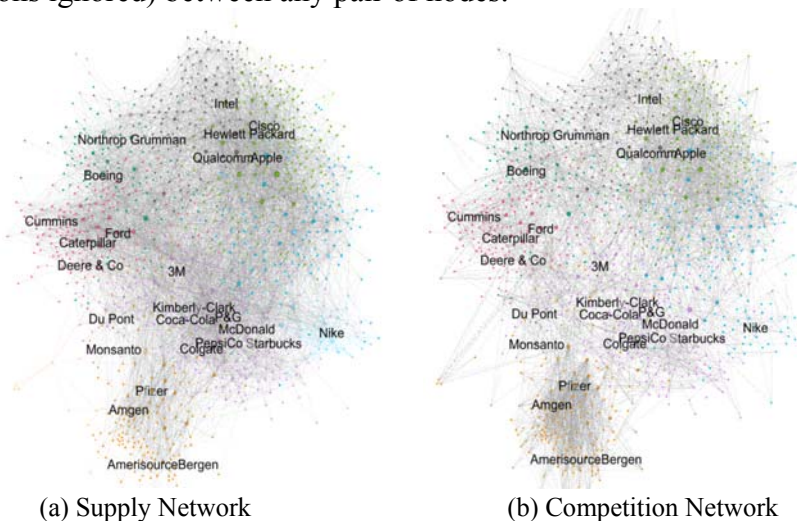


Figure 2. Supply and competition networks. Node colors represent clusters discovered via Gephi¹

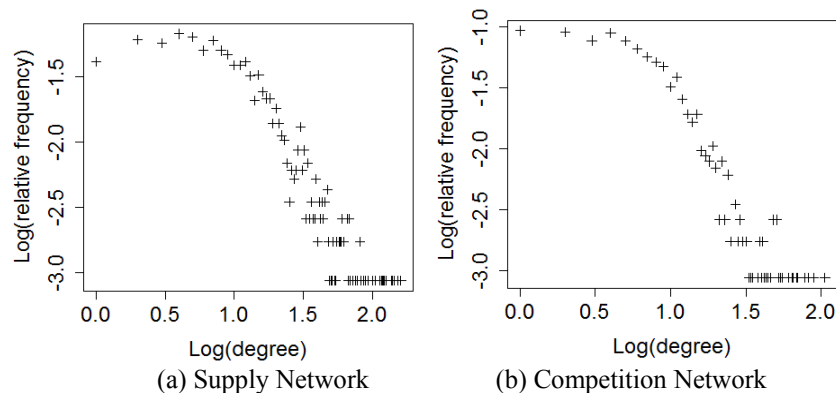


Figure 3. Degree distribution of the supply and competition networks.

	Number of Nodes	Number of Edges	Average Degree	Diameter	Size of LCC (%)
Supply Network	1,150	8,604	14.96	12	100
Competition Network	1,150	4,421	7.69	11	87.7

Table 1. Summary of basic network metrics.

Besides firms' basic information and networks, we also collected daily stock prices and returns for each of the 1,150 companies in the network during a period of 18 months, between July 1st, 2015 and December 31st, 2016. The first 12 months (July 2015-June 2016) are referred to as Period 1, and the last 6 months (July 2016-December 2016) are considered Period 2. Such information was collected from both the Center for Research in Security Prices (CRSP) and scraped from the

¹ Community detection via modularity maximization uses the heuristic method proposed by Blondel, et al. (2008).

Yahoo Finance website. Companies were identified by their ticker symbols collected from Mergent Horizon, and their matching with the information from CRSP and Yahoo Finance was crosschecked by using other attributes like company name.

4. Predictive Model

With the multi-relational network among the 1,150 firms, and their stock prices over 18 months, we built and evaluated a set of supervised classification models for the 27 firms to illustrate the predictive power from a focal firm's network neighbors.

4.1. Model Setup

We defined the target variable to predict as a three-class categorical variable (going up, down, or staying the same) indicating the stock price movement of a particular company from one day to the next. However, predicting if a stock will move up or down by a tiny amount (e.g., 0.001%) is extremely challenging yet not very helpful for investors. Therefore, we decided to raise values of between-class thresholds. We first calculated the standard deviation (sd) of the daily stock returns of the firm during Period 1. Then, we used those estimates to discretize the observed numeric daily stock returns during Period 2 into three classes. If the daily stock return of a company from day t to day $t+1$ is within $\pm 0.25sd$, that instance was labeled as *Class 0* (i.e., Neutral). This means that the stock price does not change much from day t to day $t+1$. By contrast, if the change in stock return is lower than $-0.25sd$, that instance was labeled as *Class -1*, which means that the stock price goes down from day t to day $t+1$. Similarly, if the stock return change is higher than $0.25sd$, that instance was labeled as *Class 1*, which means that the stock price is going up. This discretization yields a balanced distribution of 0.3, 0.37, and 0.33 for *Class -1*, *Class 0*, and *Class 1*, respectively.

4.2. Feature Engineering

Features used in our predictive models consist of baseline features and proposed features. Baseline features are related to financial performance of a focal firm. We included three baseline features that are known to have predictive power for stock returns (Fama and French 1996) (Carhart, 1997) (Goyal and Welch 2003).

- Stock Price: the adjusted close price of the stock i at time t .
- Stock Return: the percentage of stock i 's price change from time $t-1$ to t .
- Market Return: the return of S&P Composite Index (provided by CRSP) from time $t-1$ to time t .

Proposed features are based on stock movement of each of the three types of neighbors (suppliers, customers, and competitors) of the seed companies in the multi-relational inter-firm network. If we build one predictive model for each firm, we can directly use the stock movement of all that firm's network neighbors as features. However, this means each model could have different number of features as each firm could have a different number of network neighbors. To build one unified model that can be applied to different firms once it is learned, we need a way to represent each firm's network neighbors' stock movement in a fixed-length feature space.

To address this issue, we adopted a binning approach. For a focal firm, whose stock movement we want to predict, we created three sets of bins, each for one type of network neighbors. Then for network neighbors of the same type, we put them into one of K bins based on their standardized stock return from day $t-1$ to day t . We use suppliers of Caterpillar as an example to illustrate how the binning works with $K=11$. For each supplier of Caterpillar, we first calculated the standard deviation of its stock movement during Period 1, and then the supplier's movement during Period 2 is standardized using its own historical standard deviation. The creation of bins is similar to the discretization of the target variable: we defined a middle bin that goes between $\pm 0.25sd$ around the

mean, having a width of $0.5sd$. We created four more bins with width $0.5sd$ for daily stock returns that are below $-0.25sd$ ($[-2.25sd, -1.75sd]$; $[-1.75sd, -1.25sd]$; $[-1.25sd, -0.75sd]$; $[-0.75sd, -0.25sd]$) and one more for the far-left tail ($-\infty, -2.25sd$). In a similar fashion, we created 5 bins for daily stock returns that are above $0.25sd$, 4 of them with a width of $0.5sd$ and one more for the far-right tail. Each of the 11 bins counts the number of Caterpillar's suppliers with the corresponding stock movement during the previous day. For instance, having 2 companies in the 11th bin means that among Caterpillar's suppliers, two firms' stock prices have very high increase (higher than $2.25sd$) during the previous day. Then the numbers of suppliers in each of the 11 bins will serve as features for predicting Caterpillar's stock movement. In other words, for each day during Period 2, Caterpillar will have 11 features to represent its suppliers' performance during the previous day. The sum of the 11 features is the total number of suppliers for Caterpillar. Figure 4 presents an example distribution of suppliers for Caterpillar on particular day. The same process was applied to competitors and customers of each focal firm, leading to a total of 33 features (11 per type of network neighbors) to represent stock movements of its suppliers, customers and competitors.

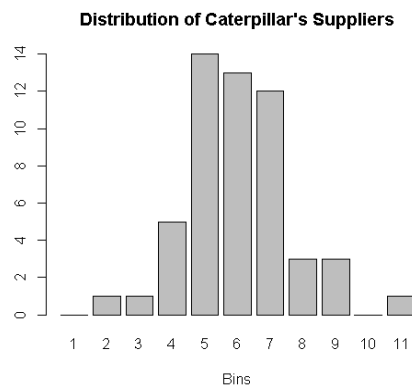


Figure 4. Distribution of the Caterpillar's suppliers among the 11 bins.

4.3. Results

To measure the predictive power of the proposed network-based features, we compared the performance of two predictive models: the baseline model has only the 3 baseline features, while the complete model adds 33 network-based features for each focal firm. Both models are predicting stock price movement from time t to $t+1$ of the 27 focal companies (supply chain leaders), using features defined from time $t-1$ to t . To evaluate the models, we implemented a “leave-one-company-out” cross-validation: during Period 2, data of one company was held out as testing data, and the predictive model was trained and tuned on the data of the remaining 26 companies (training data). The model's three-class classification performance was then evaluated on the held-out company's stock movement. This process was repeated for each of the 27 focal companies.

We also implemented eight different classifiers to predict the future stock movement of a focal firm, including linear models, decision trees, and ensemble methods. As part of the training process, we tuned the parameters of the classifiers using 10 times 10-folds cross-validation for a total of 100 iterations. We implemented all the models in python using the SciKit-learn library. To measure the performance of classifiers, we used area under the curve ROC (AUC). Since we defined this prediction as a three-class classification problem, we ran our models to predict each of the classes separately in a one-versus-the-others fashion. Then the AUCs were combined by calculating a macro-average of the AUC (macro AUC). For this, we averaged the three ROC

curves and calculated the AUC for the resulting curve. This is a modified version of the HT3 approximation proposed by Ferri, et al. (2003). The models for each class were jointly trained to minimize the macro AUC. To check if differences in performance between the baseline and complete models are statistically significant, we calculated the macro AUC for the prediction of each of the 27 firms and then built a paired t-test to compare the two models for each classifier.

Table 2 presents a comparison of AUCs obtained by the baseline and complete models for our three-class prediction. The “P-value” column summarizes the results of the t-tests comparing the performance of the two models for each classifier. The “Count” column lists the number of firms (out of 27), for which the complete model outperforms the baseline model. The table also includes additional results of predicting *Class 1*, because it is more interesting to predict whether or not the stock price will greatly move up from time t to time $t+1$, than predicting if a stock will stay pretty much the same or go down in the next day.

Model	Three-class Prediction				Class "1" Prediction			
	Baseline	Complete	P-value	Count	Baseline	Complete	P-value	Count
LASSO	0.545	0.566	***	21	0.557	0.579	**	19
Ridge Regression	0.545	0.566	**	19	0.557	0.578	*	19
SVM	0.534	0.566	***	23	0.543	0.576	**	20
Decision Tree	0.533	0.565	**	22	0.533	0.597	***	22
Random Forest	0.567	0.583	**	17	0.584	0.597	*	17
AdaBoost (Log. Regression)	0.544	0.565	**	20	0.556	0.577	*	19
Artificial Neural Networks	0.563	0.561		13	0.572	0.562		13
K-NN	0.555	0.533	*	11	0.564	0.549		12

*p-value<0.05, **p-value<0.01, ***p-value<0.001.

Table 2. Comparing AUCs of predictive models.

Comparison results revealed that random forest outperforms the other 7 classifiers for both the baseline and the complete models and for both predictions (three-class and *Class 1* vs others). One can note that the complete model is significantly better than the baseline model in 6 of the 8 experiments for both prediction problems. In the three-class prediction, the KNN classifier is better with the baseline model than when using the complete set of features. This could be caused by model overfitting, since KNN would be affected by noisy or non-relevant features included in the prediction. Overall, the complete model predicts significantly better for 17 to 23 (out of 27) firms for 6 out of 8 classifiers we tested. The baseline model significantly outperforms the complete model in only one classifier for the three-class prediction and but in no case for the *Class-1* prediction.

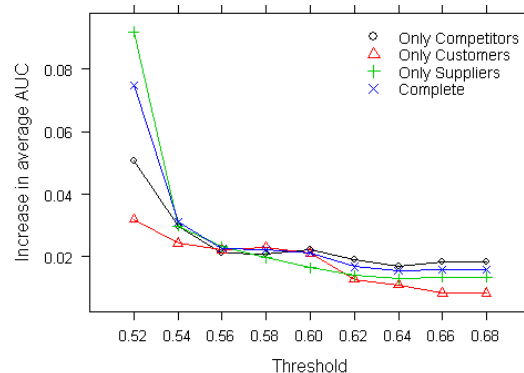
In order to illustrate the contribution from different groups of neighbors in the multi-relational network (suppliers, customers, and competitors), we ran the random forest classifier (our best classifier) with only one group of proposed features, as well as with two groups of proposed features. Figure 5(a) lists their performance (macro AUC), along with their p-values when they are compared with the baseline model. Compared with the baseline model, adding network neighbor features associated with either only competitors or with only suppliers is enough for a statistically significant improvement, with competitors providing the highest improvement. This is not the case for network features associated with customers, whose improvement over the baseline model is not statistically significant. When we included two groups of neighbors as features, adding competitors plus either customers or suppliers produces a statistically significant improvement.

However, if both customers and suppliers are included, the improvement is not statistically significant. In summary, adding either only suppliers or any combination that includes competitors will produce a statistically significant improvement in the prediction of stock price movement.

Even though we found an improvement in the average performance of our prediction when the complete model is used, for some stocks the baseline model is predicting better than the complete model. As we discussed in Section 2, baseline features Stock Price, Stock Return, and Market Return could be good predictors of the future movement of some stocks. In that case there may be little room for improvement when using the proposed network features. Following Nguyen (2015), we hypothesized that network features may not considerably improve the prediction of the stock price movement when the baseline model has a good prediction performance. To explore this hypothesis, we made another comparison of predictive models with only one group of network features. With a threshold α , we excluded companies for which the baseline model predicts well, with an AUC greater than α . Then we compared the prediction performance only for the remaining companies, by calculating the average of their AUC. Figure 5(b) presents the results of this analysis for different values of the threshold α . The complete model's increase in AUC compared to the baseline model drops as α increases. With a threshold of 0.52, the maximum difference in the average AUC is more than 0.09. By contrast, when $\alpha=0.68$, all the companies are included and the maximum increase is only 0.02. This supports our conjecture that the effectiveness of the network features decreases when the baseline model already has a high predictive power.

Set of Features	AUC	P-value
Baseline	0.567	--
Only Competitors	0.585	0.001 **
Only Suppliers	0.580	0.013 *
Only Customers	0.575	0.124
Competitors + Customers	0.586	0.002 **
Competitors + Suppliers	0.584	0.005 **
Customers + Supplier	0.576	0.076
Complete	0.583	0.007 **

(a) Prediction performance with various groups of proposed features.



(b) Comparison of the classifiers with different AUC thresholds of the baseline model.

Figure 5. Performance of the best classifier with different feature subsets.

5. Discussions

Going beyond predictions, we performed regression analyses to better understand how different types of network neighbors' performance is indicative of a focal firm's stock price movement. We focused on the binary task of prediction of *Class 1* vs others, which measures whether or not the stock price will go up from time t to $t+1$.

As for covariates, for the purpose of simplicity, we consolidated the 11 bins we created for each type of network neighbors (competitors, suppliers and customers) into three: the Down bin includes 5 original bins representing movement below $-0.25sd$; the Up bin has 5 original bins representing movement above $0.25sd$. We also included the three baseline features as covariates. Additionally, we calculated the Pearson correlation coefficient between features, finding that the highest correlation was 0.62, and 91% of the correlation coefficients were no greater than 0.5.

We performed a logistic regression analysis with backward elimination. Starting with all the 12 covariates (9 network features and 3 baseline features), the analysis runs all the models that exclude exactly one of the features and picks the best model based on the highest log-likelihood. This process was repeated until all the features in the model are statistically significant. The pseudocode of the process is presented in Figure 6. We also calculated the Variance Inflation Factor (VIF) of the features to identify any potential multicollinearity issue in the final model. Table 3 presents the regression results of the final model after the running the algorithm.

In the final model, significant covariates include “Market Return,” “Competitors Down” (number of competitors with stock price significantly moving down during time t), “Competitors Neutral” (number of competitors with stock price that is not moving during time t), and “Suppliers Up” (number of suppliers with stock price significantly moving up during time t). Specifically, for a focal firm, the number of suppliers going up and the number of competitors going down in the previous day are associated with an increased probability that its stock price moves up in the next day. These results complement our previous findings by showing more specifically the subgroup of competitors and suppliers that are more relevant in the prediction of stock price movement.

-
1. Let M_p denote the full model, which contains all p predictors.
 2. For $k = p, p-1, \dots, 1$, do until all the features in M_{k-1} are statistically significant:
 - o Consider all k models that contain all but one of the predictors in M_k , for a total of $k - 1$ predictors in each of the models.
 - o Choose the “best” among these k models, and call it M_{k-1} . “best” is defined as having highest log-likelihood.
 3. Analyze the coefficients of the model M_k .
-

Figure 6. Pseudocode for the backward feature elimination.

Features	Final Model		
	Coefficient	Standard Error	VIF
Market Return	-0.232 ***	0.046	1.503
Competitors Down	0.159 ***	0.045	1.551
Competitors Neutral	-0.225 ***	0.045	1.345
Suppliers Up	0.147 ***	0.042	1.316

Table 3. Regression results after backward elimination.

6. Conclusions and Future Work

This paper contributes to the literature on stock price prediction by leveraging information from multi-relational networks that a firm is in. Previous studies had shown that companies are now more connected and their performance is related to the performance of the other companies in an inter-firm network. Based on those findings we defined our main thesis, which states that analyzing the performance of a focal firm’s neighbors in its multi-relational inter-firm networks improves the prediction of its stock price movement. Based on data we scraped from a database, we constructed a large-scale multi-relational inter-firm network that incorporate a supply chain network and the corresponding competition network. We proposed and designed network-based features that reflect the performance of different types of network neighbors of a focal firm. Experimental results suggest that these network features can significantly improve the prediction of stock movement compared to baseline features used in the financial literature. In addition to the improvement in prediction, we also identified a more specific subgroup of competitors and suppliers that are associated with a focal firm’s upward movement in stock price. The results

highlight the importance of business partners for a focal firm's performance in different ways and provide managers with additional knowledge to make better decisions.

This study has its limitations. First, our prediction models focused only on the 27 companies that are supply chain leaders and trade in US stock markets. Future work could include a larger pool of companies in the stock market, for example S&P100. Also, additional sources of reliable financial information can be explored to build a much larger inter-firm network beyond the US. Second, one day (24 hours) is a somewhat arbitrary unit of time lag for information flow and stock-price adjusting. Potential extensions of the study could explore if our approach would work differently with different time lags (e.g., hours or more than one day). Also, our current model only considers a focal firm's immediate neighbors in a multi-relational network. A logical next step would explore neighborhoods of firms beyond 1-hop neighbors to identify the most relevant peers of a focal firm. Third, our approach only considers two types of relationship (partnership and competition) in the multi-relational network. Other relationships, such as co-location and investors' attention (Agarwal, et al. 2017) are interesting extensions to our multi-relational network. Finally, future work could include the design of a trading strategy based on our findings to highlight their practical relevance for the finance industry too.

ACKNOWLEDGMENTS

The authors thank anonymous reviewers for their valuable comments for the present paper and future work.

References

- Agarwal, Ashish, Alvin Chung Man Leung, Prabhudev Konana, and Alok Kumar. 2017. "Co-search attention and stock return predictability in supply chains." *Information Systems Research* (September): 1-62.
- Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment* (IOP Publishing) 2008 (10): P10008.
- Boccaletti, S, G Bianconi, R Criado, Ci del Genio, J Gómez-Gardeñes, M Romance, I Sendiña-Nadal, Z Wang, and M Zanin. 2014. "The structure and dynamics of multilayer networks." *Physics Reports* 544: 1-122.
- Carhart, Mark M. 1997. "On persistence in mutual fund performance." *The Journal of Finance* 52 (1): 57-82.
- Choi, Thomas Y., and Yunsook Hong. 2002. "Unveiling the structure of supply networks: Case studies in Honda, Acura, and DaimlerChrysler." *Journal of Operations Management* 20 (5): 469-493.
- Cohen, Lauren, and Andrea Frazzini. 2008. "Economic links and predictable returns." *Journal of Finance* 63 (4): 1977-2011.
- Downing, Stephen Thomas, Jin-Su Kang, and Gideon D Markman. 2017. "Competitive dynamics: Of whom should you be aware?" *Academy of Management Proceedings* 2017 (1): 16381.
- Fama, Eugene F, and Kenneth R French. 1996. "Multifactor explanations of asset pricing anomalies." *The Journal of Finance* 51 (1): 55-84.
- Ferri, César, José Hernández-Orallo, and Miguel Angel Salido. 2003. "Volume under the ROC surface for multi-class problems." *European Conference on Machine Learning*. 108-120.
- Ferrier, Walter J, and Hun Lee. 2002. "Strategic aggressiveness, variation, and surprise: How the sequential pattern of competitive rivalry influences stock market returns." *Source: Journal of Managerial Issues* 14 (2): 162-180.
- Gartner Inc. 2015. *The Gartner supply chain top 25 for 2015*.
<https://www.gartner.com/newsroom/id/3053118>.
- Gimeno, Javier. 2004. "Competition within and between networks: The contingent effect of competitive embeddedness on alliance formation." *Academy of Management Journal* 47 (6): 820-842.
- Goyal, Amit, and Ivo Welch. 2003. "Predicting the equity premium with dividend ratios." *Management Science* 49 (5): 639-654.

- Harland, Christine M., Richard C. Lamming, Jurong Zheng, and Thomas E. Johnsen. 2001. "A taxonomy of supply networks." *Journal of Supply Chain Management* 37 (3): 21-27.
- Hearnshaw, Edward J.S., and Mark M.J. Wilson. 2013. "A complex network approach to supply chain network theory." *International Journal of Operations & Production Management* 33 (4): 442-469.
- Hendricks, Kevin B, and Vinod R Singhal. 2005. "An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm." 14 (1): 35-52.
- Kim, Yusoon, Thomas Y. Choi, Tingting Yan, and Kevin Dooley. 2011. "Structural investigation of supply networks: A social network analysis approach." *Journal of Operations Management* 29 (3): 194-211.
- Lamming, Richard, Thomas Johnsen, Jurong Zheng, and Christine Harland. 2000. "Classification of supply networks An initial classification of supply networks." *International Journal of Operations & Production Management* 20 (6): 675-691.
- Luo, Xueming, Jie Zhang, and Wenjing Duan. 2013. "Social media and firm equity value." *Information Systems Research* 24 (1): 146-163.
- Menzly, Lior, and Oguzhan Ozbas. 2010. "Market segmentation and cross-predictability of returns." *Journal of Finance* 65 (4): 1555-1580.
- Merschmann, Ulf, and Ulrich W. Thonemann. 2011. "Supply chain flexibility, uncertainty and firm performance: An empirical analysis of German manufacturing firms." *International Journal of Production Economics* (Elsevier) 130 (1): 43-53.
- Nguyen, Thien Hai, Kiyooki Shirai, and Julien Velcin. 2015. "Sentiment analysis on social media for stock movement prediction." *Expert Systems with Applications* (Elsevier Ltd.) 42 (24): 9603-9611.
- Rechenthin, Michael, W. Nick Street, and Padmini Srinivasan. 2013. "Stock chatter: Using stock sentiment to predict price direction." *Algorithmic Finance* 2 (3-4): 169-196.
- Rodriguez, Marko A., and Joshua Shinavier. 2010. "Exposing multi-relational networks to single-relational network analysis algorithms." *Journal of Informetrics* (Elsevier) 4 (1): 29-41.
- Sanders, Nada R. 2016. "How to use big data to drive your supply chain." *California Management Review* 58 (3): 26-48.
- Skilton, Paul F., and Ednilson Bernardes. 2015. "Competition network structure and product market entry." *Strategic Management Journal* (Wiley-Blackwell) 36 (11): 1688-1696.
- Vanichchinchai, Assadej, and Barbara Igel. 2011. "The impact of total quality management on supply chain management and firm's supply performance." *International Journal of Production Research* 49 (11): 3405-3424.

- Yao, Emery, Walter J Ferrier, and Yu Tieying. 2007. "Rivals by association: Competition networks and firm performance." *The Academy of Management Meeting*. Philadelphia, PA.
- Zhao, Kang, Akhil Kumar, and John Yen. 2011. "Achieving high robustness in supply distribution networks by rewiring." *IEEE Transactions on Engineering Management* 58 (2): 347-362.
- Zhao, Kang, Xi Wang, Sarah Cha, Amy M Cohn, George D Papandonatos, Michael S Amato, Jennifer L Pearson, and Amanda L Graham. 2016. "A multirelational social network analysis of an online health community for smoking cessation." *Journal of Medical Internet Research* (JMIR Publications Inc.) 18 (8): e233.
- Zuo, Zhiya, Kang Zhao, and Jennifer V Blackhurst. 2018. "Network analysis of supply chain adaptation for disruption recovery: An empirically grounded complex adaptive systems approach." *Journal of Operations Management* (forthcoming).