



Contents lists available at ScienceDirect

Information & Management

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



The dynamic predictive power of company comparative networks for stock sector performance

Kun Chen^{a,*}, Peng Luo^b, Dongming Xu^c, Huaiqing Wang^a

^a Department of Finance, South University of Science and Technology, Shenzhen 518055, China

^b School of Management, Harbin Institute of Technology, Harbin 150001, China

^c UQ Business School, The University of Queensland, St Lucia, QLD 4072, Australia

ARTICLE INFO

Article history:

Received 4 July 2015

Received in revised form 13 June 2016

Accepted 19 July 2016

Available online xxx

Keywords:

Company network

Sentiment analysis

Vector autoregression

Stock sector performance

ABSTRACT

As economic integration and business connections increase, companies actively interact with each other in the market in cooperative or competitive relationships. To understand the market network structure with company relationships and to investigate the impacts of market network structure on stock sector performance, we propose the construct of a company comparative network based on public media data and sector interaction metrics based on the company network. All the market network structure metrics are integrated into a vector autoregression model with stock sector return and risk. Several findings demonstrate the dynamic relationships that exist between sector interactions and sector performance. First, sector interaction metrics constructed based on company networks are significant leading indicators of sector performance. Interestingly, the interactions between sectors have greater predictive power than those within sectors. Second, compared with the company closeness network, the company comparative network, which labels the cooperative or competitive relationships between companies, is a better construct to understand and predict sector interactions and performance. Third, competitive company interactions between sectors impact sector performance in a slower manner than cooperative company interactions. The findings enrich financial studies regarding asset pricing by providing additional explanations of company/sector interactions and insights into company management using industry-level strategies.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

As economic integration and business connections increase, companies actively interact with each other in the market in cooperative or competitive relationships. Such relationships often exhibit industry-related features. For example, competitive relationships often exist within an industry because of limited resources and customers. These cooperative relationships usually arise between the supply and demand sides across different industries. Complex interactive business relationships depict the economic market with intra-sector and cross-sector links. These links are helpful for understanding information and shock transfers within and across sectors [1–3]. For example, the spillover effect between sectors was observed during the global financial crisis and the recent Chinese stock market crash. Consider

the manufacturing sector and the utility sector in the Chinese stock market. Between June and July 2015, the manufacturing sector index¹ decreased by 29.96%, and the utility sector index decreased by 24.93%. The manufacturing sector suffered a much heavier loss than the utility sector. In market interactions, companies in the manufacturing sector have more business connections with other companies than companies in the utility sector. To understand market interactive structures and to explain the spillover effect between sectors, we designed this study.

Previous accounting and finance studies have begun to establish the connection between market network structure and stock sector performance [1–3]. They have used trading data to create sector relationship graphs, and they have proposed the theory that sectoral shocks are transmitted to other sectors using networks of input and output linkages. However, the trade flow graphs are rather coarse tools for describing company relationships. In the field of information systems (IS), some studies have

* Corresponding author.

E-mail addresses: chenk@sustc.edu.cn, chenkun_1025@126.com (K. Chen), luopeng_hit@126.com (P. Luo), D.Xu@business.uq.edu.au (D. Xu), wanghq@sustc.edu.cn (H. Wang).

¹ According to the Wind sector index.

constructed company relationship networks based on textual information mining. They have identified the co-occurrences of two companies' names in documents [4–8]. This method helps to measure the closeness of two companies, but it cannot specify the types of comparative relationship, i.e., competitive relations or cooperative relations. In a real-world market, the relationship between Apple and Samsung is definitely different from that between Apple and FoxCom. These different relationships have different spillover effects on stock performance. Therefore, to further investigate the connections between market network structure and stock sector performance, we focus on the following research questions.

- (1) Does the company comparative network provide a stronger market indicator than the company closeness network?
- (2) What are the intra-sector and inter-sector network effects on stock sector performance?
- (3) What are the dynamics of the relationship between company comparative network metrics and stock sector performance?

To answer these questions, we use public news² as a data source to build company networks because we believe that as an easily accessed Web-based data source, news describes richer business relationships between companies than simple trading data. Moreover, to identify network effects, we construct complex network metrics. First, we use comparative analysis, rather than co-occurrence analysis, to investigate the cooperative (positive) and competitive (negative) relationships identified by public information. Second, we construct inter-sector and intra-sector measurements to compare their different effects.

In contrast to previous studies that aimed to detect the static correlations between company network and stock performances, our study uses a vector autoregression with exogenous variables (VARX) model to consider all of the intricate dynamic relationships among network metrics and stock sector performance. The time-series model investigates continuous daily company network effects on stock sector performance, and it captures the dynamics of short- and long-term carryover effects over time.

This study has potential implications for theory and practice. Theoretically, our work confirms and extends financial theories by introducing rich market network structure metrics based on public information. We use a time-series model to investigate the dynamic relationships between company comparative networks and stock sector performance. Our research also provides practical suggestions for sector-level strategies such as industry associations and investments.

We first describe the theoretical background and hypotheses in Section 2. Section 3 introduces the data and the measurements. Section 4 describes the time-series model. The findings are presented in Section 5. The final section discusses the implications.

2. Theoretical background and hypotheses

2.1. Intra-sector and inter-sector network effects on stock sector performance

Stock sector performance has been demonstrated to be related to sector positions in market networks. In the finance domain, Aobdia, Caskey, and Ozel [3] constructed an industry network based on trade flows across different industries, and they found that firms in central industries are more exposed to systemic risks than other firms. Acemoglu et al. [1] argued that sectoral risks can

be transmitted to other sectors through a network of input and output linkages in a system. Ahern and Harford [2] demonstrated that systematic risks constitute the aggregation of idiosyncratic shocks and that more central sectors in a network of intersectoral trade usually have higher returns because they experience greater exposure to systematic risks.

Because of the popularity of social media and Web 2.0, company interactions regarding sales, debts, and other financial or operating activities are reported in public news in real time. Company networks based on keyword co-occurrence have been widely used to explain and predict financial metrics such as company revenue; stock return; and risk. For example; Ma; Sheng; and Pant [6] predicted company revenue relationships based on a company network derived from company citations. Graph-theoretic measurements were used in the classification problem. Jin et al. [5] developed complex longitudinal features for company network evolution and proposed feature selection and prediction models to predict company profit and revenue growth. Focusing on stock market performance; Creamer; Ren; and Nickerson [9] tested the relationships among company positions in networks; company stock returns; and volatility.

We expect that constructing sector-related metrics based on company networks might also provide a useful indicator for predicting sector performance. Compared with trade flow, which has been used in previous financial studies [2,3], company networks encompass broader business relationships between companies.

H1a. Sector interaction metrics constructed based on company networks have significant predictive relationships with sector performance.

To further investigate the sector-interactive characteristics, we construct two metrics: an inter-sector metric and an intra-sector metric. These two metrics have primarily been used in economics to distinguish trades between different industries or within the same industry [10,11]. These sector metrics have also been used in financial studies that have investigated stock performances. Moskowitz and Grinblatt [12] and Aobdia, Caskey, and Ozel [3] demonstrated that inter-sector characteristics have predictive power for assessing firms' stock returns. Conversely, Asness, Porter, and Stevens [13] found that intra-sector momentum is superior to inter-sector momentum in explaining stock returns. Because this study aims to inspect how sector-interactive characteristics affect stock returns, we followed the two popular metrics and proposed two competing hypotheses:

H1b. The inter-sector metric has greater predictive power than the intra-sector metric.

H1b'. The intra-sector metric has greater predictive power than the inter-sector metric.

2.2. Company comparative networks provide a stronger market indicator than closeness networks

In the business world, company comparative analysis refers to evaluating a list of company metrics to compare them. The targets are usually similar companies in the same industry, such as Ford versus Toyota and eBay versus Amazon. In IS and marketing research, comparative analysis has been extended to the analysis of comparative opinions between two entities [14–16]. Taking products as an example, comparative analysis aims to identify the relationship of two products as “product A is better than product B” or “product B is better than product A.” For example, Jindal and Liu [17,18] proposed using rules and naïve Bayes

² “News” in this article refers to a broad range of information from news wires, discussion boards, and blogs.

classifiers to identify comparative sentences and relationships in these sentences. Xu et al. [19] used a conditional random field-based method to extract the comparative relationships between products from a sentence. Zhang et al. [8] proposed a sentiment analysis method for constructing product comparison networks on a coarse-granularity level. In a company analysis scenario, comparative analysis refers to identifying the relationships between companies, including competitive relationships and cooperative relationships. These relations are often hidden in news reports and other public information. Similar to previous sentiment analysis, competitive relationships usually exhibit negative comparative opinions, and cooperative relationships often feature positive comparative opinions.

Although company comparative relationship networks are assumed to constitute a good market indicator, there is little evidence that supports this assumption. Inter-company relationships are currently extracted from textual news based on the co-occurrence of company names. The more frequently the documents mention two companies together, the closer those companies are to each other. This line of reasoning stems from the notion of memory-associative networks [20], and it has strong roots in the co-word analysis literature [21]. Company co-occurrence networks have been used to analyze company financial performance.

On the basis of above analysis, we want to build a company comparative network and compare its market predictive power with that of a company co-occurrence network. Thus, the following hypothesis is posited.

H2a. Company comparative analysis provides a stronger sector interactive indicator than company closeness analysis.

To further investigate the sentiment of comparative opinions, we divide company comparative relationships into two categories: cooperative (or positive) relationships and competitive (or negative) relationships. Sentiment analysis has been widely adopted in financial studies to predict stock prices [22]. One research stream uses the polarity value of news as a predictive measure of stock performance, e.g., Li et al. [23], Yu, Duan, and Cao [24], and Tetlock, Saar-Tsechansky, and Macskassy [25]. The other stream inspects the differential impacts of positive and negative news on stocks. For example, Chan [26] found less drift for stocks with good news than for those with bad news. Van [27] found that arrival of bad news had a greater impact on volatility than did arrival of good news. This study aims to investigate whether a difference exists between the impact of cooperative (positive) relationships and competitive (negative) relationships. Thus, we propose the following competing hypotheses.

H2b. Cooperative (positive) sector interactive metrics have greater predictive power than competitive (negative) sector interactive metrics.

H2b'. Competitive (negative) sector interactive metrics have greater predictive power than cooperative (positive) sector interactive metrics.

2.3. The dynamics of the predictive value of company comparative networks

The previous literature has demonstrated the dynamics of stock market responses to word-of-mouth information and social media. Luo, Zhang, and Duan [28] compared the short- and long-term effects of social media with those of conventional online behavioral metrics on a firm's equity values. They found that social media metrics have faster predictive value. Additionally, Tirunillai and

Tellis [29] demonstrated that negative user reviews are related to stock returns, with significant wear-in effects. In dynamic analysis, the wear-in time, which is defined as the time required to reach the peak predictive value, is valuable because it suggests a critical time period for decision-making, whereas the wear-out time, which is defined as the time required before the predictive value reaches asymptotes, indicates the impact duration.

Theoretically, the information diffusion model [30] has been widely used in the finance domain to explain the dynamic effects of information on stock returns. Hong et al. [31] demonstrated that bad news travels slowly through the investing public. Chan [26] also found that prices are slow to reflect bad public news. This study aims to investigate the differences in wear-in and wear-out effects on competitive and cooperative relationships. Thus, we propose the following two groups of competing hypotheses.

H3a. Cooperative (positive) sector interactive metrics have a shorter wear-in time than competitive (negative) sector interactive metrics.

H3a'. Competitive (negative) sector interactive metrics have a shorter wear-in time than cooperative (positive) sector interactive metrics.

H3b. Cooperative (positive) sector interactive metrics have a longer wear-out time than competitive (negative) sector interactive metrics.

H3b'. Competitive (negative) sector interactive metrics have a longer wear-out time than cooperative (positive) sector interactive metrics.

3. Data and measurements

3.1. Data processing

The raw data set consists of one year (2013) of Chinese business news for 300 companies in the Shanghai–Shenzhen 300 Index.³ These companies span 10 sectors⁴ within the Chinese stock market, including materials, finance, energy, and daily consumption, among others. The news stories are collected from a general search portal,⁵ which covers 3000+ online sources, including discussion boards, news wires, and blogs. To obtain a clear overview of the information sources, we focus on the top 100 online news sources ranked by a number of news items. We find that the top 100 online news sources cover 76.89% of the total news items online (the total number of news items is 946,935). Among these sources, we identify 74 news websites, 21 discussion boards, and 5 blogs. Discussion boards have the largest number of news items because anyone can freely post opinions about companies or stocks on discussion boards. From the 74 news websites, we confirm that the major Chinese financial web media is covered. It includes government-operated media, such as renmin.com and xinhua.com, and 4 major security newspapers in China (cs.com.cn, cnstock.com, p52.net, and zqrb.ccstock.cn). The websites also include some popular financial portals such as ifeng.com, hexun.com, jinrongjie.com, eastmony.com, business.sohu.com, and finance.sina.com.cn. In this paper, we want to use a broad range of big data to identify company relationships. Both regular news and

³ Shanghai–Shenzhen 300 Index on July 31, 2013.

⁴ According to the Wind industry classification, schema at the first level.

⁵ The work was completed by a company that specializes in providing information.

rumors are important for investigating the impact of market information. Therefore, we use a variety of news sources.

In the next step, we perform data clearing to delete repeated or forwarded news. According to the efficient market hypothesis (EMH) [32], financial markets respond to market information in an efficient manner. There are three forms of the EMH (the weak form, the semi-strong form, and the strong form), which differ in terms of the information that can be captured in a market (historical public information, current public information, and hidden information). In agreement with the EMH, we must identify the first published news to determine the market time of information. In this step, we use the cosine similarity [33] to compare the similarity of documents. For each news item, we fetch documents within a 30-day time window before and after the news is published. With textual features and the cosine similarity, we compare the similarity of two documents. If the similarity between the two documents is greater than 90%, we assume that the two pieces of news are repeated or forwarded news. The one published later is then deleted. After this step, there are 363,421 news items remaining.

To identify intercompany relationships, we first exclude documents that only mention one company or mention more than five companies because a document that includes many company names is less important than a document that mentions only a few companies [5]. There are 314,475 news items remaining. The next task is to locate target companies in news stories. In contrast to previous studies [6,7,9,24], which assume that news websites clearly label the news with a target company, we believe that labeling news stories is an important task for cases in which news is collected broadly from the Web. Therefore, we define several rules for identifying target companies. If a company name appears in the title, it is the target company. If no company name appears in the title, we determine the most frequently mentioned target companies by counting the number of times that company names occur.

In the subsequent step, we want to identify comparative opinions between companies using machine-learning methods. In the training procedure, we randomly select 3000 news items, which include 8980 sentences containing company names other than the target companies. Then, we manually label these sentences as depicting positive or negative relationships between the appearing company and the target company. The reasons for only using positive and negative labels have previously been summarized [24]. First, a sentence that includes subjective expressions always implies either positive or negative feelings, and “neutral” is a fairly vague concept. Second, no mature methods exist for efficiently and accurately identifying neutral sentiments. Using the labeled dataset, we compute the area under the curve (AUC) [34] of different classifiers based on a 10-fold cross-validation.

The receiver operating characteristic (ROC) curve illustrates the performance of a binary classifier system as its discrimination threshold varies. The curve is created by plotting the true positive rate against the false positive rate at various threshold settings. A ROC curve closer to the top-left corner indicates better dynamic performance. How close the ROC curve is to the top-left corner can be reflected in the AUC measurement, which is also used as an evaluation metric in this paper.

During implementation, we use the bag-of-words feature model,⁶ apply information gain (IG)-based feature selection, and tune the thresholds to test different classifiers’ performances using different feature sizes. We experiment with several popular

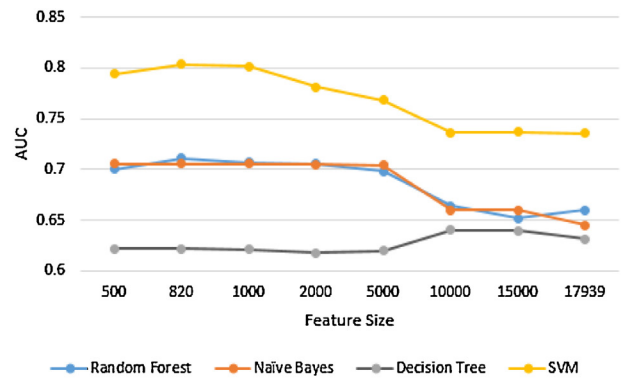
machine learning algorithms including support vector machines (SVM), decision trees, random forest, and naïve Bayes. The results are displayed in Fig. 1. From Fig. 1 (a), we find that classifiers perform best on the top 820 features ordered by IG. These 820 features are selected when the threshold of IG is set to 0. Fig. 1 (b) shows that the performance of SVM is much better than that of the other classifiers ($AUC_{SVM} = 0.8036$, $AUC_{Decision_Tree} = 0.6218$, $AUC_{Naïve_Bayes} = 0.7058$, $AUC_{Random_Forest} = 0.7108$). Therefore, we use the 820 features and train the SVM model to classify company comparative sentences. We identify 13,110 negative relationships and 182,970 positive relationships. The ratio between the positive and negative relationships is supported by previous studies of sentiment classification using user-generated content [29].

3.2. Network construction

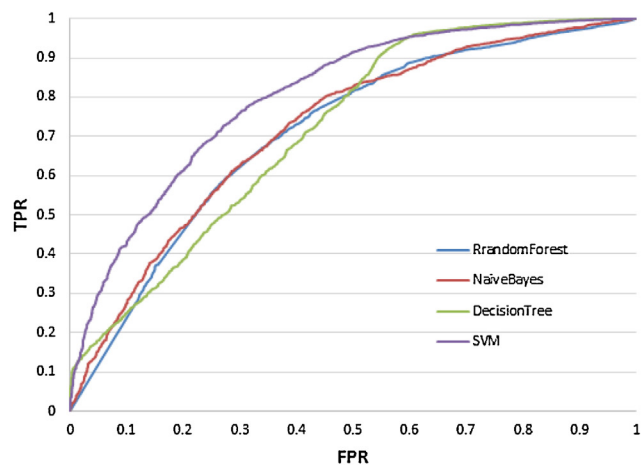
Each node in the network represents a company, and a direct link indicates a comparative relationship between two companies. The corresponding weight of each link indicates the sentiment strength of the comparison relationship. This network is formally defined as follows.

Assume that each sentence in the news for target company $c1$, along with a mention of company $c2$, is mapped into a comparison tuple $t = \{c1, c2, P/N\}$, where P/N indicates that the comparative opinion from $c2$ to $c1$ is positive or negative.

We consider the following methods of network construction.



(a) AUC values for different classifiers and feature sizes



(b) ROC curve for a feature size of 820

Fig. 1. Experimental results for sentiment classification.

⁶ We also test the BOW+POS feature model, but the performance is not improved.

3.2.1. Company closeness (undirected) networks

All n tuples are aggregated to produce a single link with a weight. An edge between nodes c1 and c2 is introduced when $(N_{pd} + N_{nd}) > 0$, and the weight of the link is $w = (N_{pd} + N_{nd})$, where N_{pd} denotes the number of positive sentences and N_{nd} denotes the number of negative sentences.

3.2.2. Company comparative (directed) networks

We construct two categories of directed networks: positive and negative networks. In a positive network, an edge from node c1 to c2 is introduced when $N_{pd} > 0$, and the weight is $w = N_{pd}$. Similarly, when $N_{nd} > 0$, we can introduce a link from c1 to c2 and set the weight as $w = N_{nd}$.

3.3. Measurements

3.3.1. Measurements of stock sector performance

On the basis of previous research [28,35], we use two common measures to determine sector performance: sector return and risk. Return or abnormal return refers to sector stock value beyond what is expected based on the stock market average. Risk, which refers to the vulnerability of sector stock value, can be measured as the standard deviation of the residuals of the returns as follows:

$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it}$, (2) where t is the subscript for the time period, R_{it} is the return of stock i at time t , R_{mt} is the average market return represented by the Shanghai Security Exchange Composite Index, R_{ft} is the risk-free rate of return, α_i is the intercept, and ε_{it} is the model residual. Eq. (2) is processed for a rolling window of 250 trading days before the target day. The abnormal return of stock i (AR_i) is measured as the difference between the observed return and the expected return, and the risk is the standard deviation of the model residuals as indicated below:

$$AR_{it} = (R_{it} - R_{ft}) - (\alpha_i + \beta_i(R_{mt} - R_{ft})). \quad (3)$$

3.3.2. Measurements of sector interaction metrics

The modularity is defined as the fraction of edges that fall within the communities minus the expected value of the same quantity if the edges are assigned at random, conditional on the given community memberships and the degree of the vertices [36]. In previous research, the modularity has primarily been used for evaluating community detection [37,38]. This study introduces the modularity to measure the strength of the connection between the nodes within (or between) groups. We divide stocks into different groups based on the sector to which they belong, and we use the modularity to calculate the intra- and inter-group interactions. When computing the interactions of two groups, we treat the two groups as a whole to yield the modularity value of the entire group.

In the comparative (directed) network, let c_i be the community to which node i is assigned and let $w_i^{in} = \sum_j w_{ji}$, $w_i^{out} = \sum_j w_{ij}$. Then the modularity Q is given by Leicht and Newman [39] as follows:

$$Q = \frac{1}{m} \sum_{ij} \left[w_{ij} - \frac{w_i^{in} w_j^{out}}{m} \right] \delta(c_i, c_j), \quad (4)$$

where the δ -function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise, and $m = \sum_{ij} w_{ij}$ is the sum of the weights in the entire network.

This formula for the modularity is adjusted to measure the intra- and inter-sector interactions as follows:

$$Q_x' = \frac{1}{m} \sum_{ij} \left[w_{ij} - \frac{w_j^{in} w_i^{out}}{m} \right] \delta'(c_i, c_j) \quad (5)$$

$$Q_{xy}'' = \frac{1}{m} \sum_{ij} \left[w_{ij} - \frac{w_j^{in} w_i^{out}}{m} \right] \delta''(c_i, c_j) \quad (6)$$

where $\delta'(c_i, c_j)$ is 1 when the two stocks i and j belong to the same sector; otherwise, the value is 0. The value of function $\delta''(c_i, c_j)$ equals 1 if the two stocks belong to the two target sectors for which we want to calculate the value of inter-sector interactions; otherwise, the value is 0. In the closeness (undirected) network, according to Newman [40], we can measure the intra- and inter-sector closeness by replacing the w_i^{in} or w_i^{out} with the sum of the weights that link node i ($w_i = \sum_j w_{ij}$) in Eqs. (5) and (6). The algorithms are presented as follows:

$$Q_x' = \frac{1}{m} \sum_{ij} \left[w_{ij} - \frac{w_j w_i}{m} \right] \delta'(c_i, c_j) \quad (7)$$

$$Q_{xy}'' = \frac{1}{m} \sum_{ij} \left[w_{ij} - \frac{w_j w_i}{m} \right] \delta''(c_i, c_j) \quad (8)$$

where $m = 0.5 \sum_{ij} w_{ij}$, and the functions of $\delta'(c_i, c_j)$ and $\delta''(c_i, c_j)$ are the same as those in Eqs. (5) and (6).

3.3.3. Measurements of sector news sentiment

According to the EMH [32], stock price reflects all available market information. To control the influences of market momentum on stock performances, we further measure the sentiment of market news [41]. The method of sentiment classification is similar to what we have undertaken in previous comparable relationship mining. We first randomly collect 10,000 documents from the news set for labeling. We then select features and perform the test using the labeled data set.⁷ With the trained classifier model, we perform binary classification of the whole news set. Then, we summarize the daily number of positive news about stock $_i$ as n_p and the daily number of negative news about stock $_i$ as n_n . The sentiment score of stock $_i$ on that day is denoted as $n_p - n_n$. For a sector measurement, the sentiment of individual stock is accumulated. Although other factors that influence sector performance exist, as discussed in the conclusions section, we argue that price and market real-time news have covered the most important and popular parts of the available information in measuring an efficient market.

3.4. An example

Here we provide an example to illustrate the network construction and calculation of sector interaction metrics. First, we focus on stock 00002 (denoted as stock A) and stock 000024 (denoted as stock B). Both are from the finance sector on December 2, 2013. All the target stocks for the 3 negative links are stock A. Among the 14 positive links, the target stocks of 12 links are A, and the target stocks of the other 2 links are B. As indicated in Fig. 2, when constructing a closeness (undirected) network, only one edge exists between the two stocks, and the weight is $17 = 3 + 14$. In the cooperative (positive) network, an edge between B and A exists, and the weight is 12. Simultaneously, an edge runs from A to B, the weight of which is 2. In the competitive (negative) network, the edge between B and A has a weight of 3.

⁷ The IG threshold value is set to 0, and the accuracy reaches 92.42% in a 10-fold cross validation test using SVM.

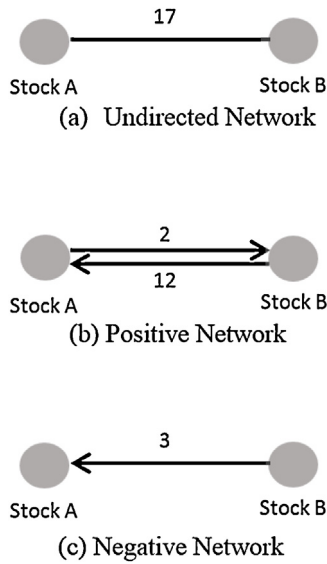


Fig. 2. Examples of constructing social networks.

Second, we use stock A and stock B on December 2, 2013, to calculate the sector interaction metrics. Taking the undirected network as an example, the total sum of weights on the links in the network is 1521; thus, $m = 1521$. Among all of the links, those with stock A at one end are used to calculate W_A , and $W_A = 213$. Similarly, those that have stock B at one end are used to calculate W_B , and $W_B = 23$. The weights of the edges that link both stock A and stock B are used to calculate W_{AB} , and $W_{AB} = 17$. Furthermore, stocks A and B belong to the same sector (finance); thus, $\delta'(c_A, c_B) = 1$. Considering other stocks in the finance sector on the same day, we use Eq. (5) and finally obtain the intra-sector modularity (intra_uq) of finance, which is 0.041.

Given another stock C (601992) that belongs to the materials sector, we can obtain the corresponding values of $m=1521$, $W_A=213$, $W_C=5$, and $W_{AC}=2$ from the network in a similar manner. The value obtained is accumulated in the inter-sector modularity between materials and finance. As indicated in Table 1, the inter-sector modularity between finance and materials is 0.000306. Fig. 3 displays the relationships in the following three sectors: materials (green), daily consumption (blue), and finance (red).

In the positive network, the sum of weights is 1448; thus, $m = 1448$. The links directed to stock A are used for calculating w_A^{in} , and the links directed to stock B are used for calculating w_B^{in} : $w_A^{in} = 172$, $w_B^{in} = 8$. Conversely, the links that start from stock A are used for calculating w_A^{out} , and the links that start from stock B are used for calculating w_B^{out} : $w_A^{out} = 25$, $w_B^{out} = 12$. The weight of the link from A to B is 2 ($w_{AB} = 2$), and the weight of the link from B to A is 12 ($w_{BA} = 12$). Because stocks A and B belong to the same financial sector, the value obtained from Eq. (5) is accumulated in the intra-sector modularity (intra_pq) of finance in the positive network as 0.0686.

Table 1
 Intra- and inter-sector modularities.

Sector	Materials (green)	Daily consumption (blue)	Finance (red)
Materials (green)	0.0045	-0.00025	0.00030
Daily consumption (blue)	-0.00025	0.019	0.00056
Finance (red)	0.00030	0.00056	0.041

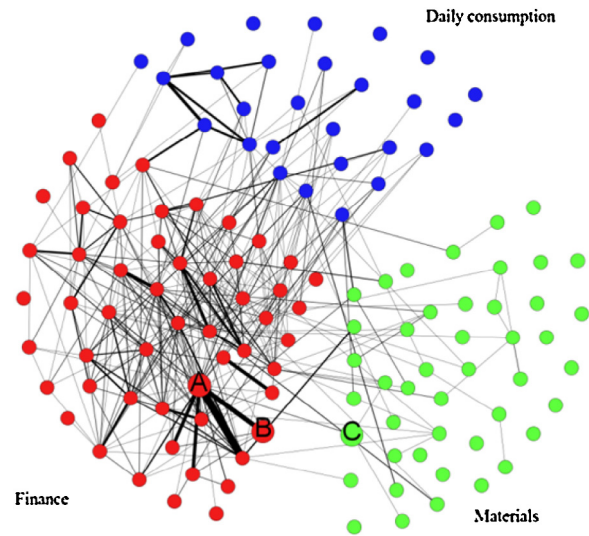


Fig. 3. Undirected network of three industries.

4. Econometric model

4.1. Rationale for VARX

We employ a VARX model, which is a time-series technique, for an empirical investigation. VARX models include exogenous variables, unlike standard VAR models. VARX models are suitable for examining the dynamics of the relationship between the sector interaction measures and sector performance with the following advantages. First, VARX models are particularly useful for describing interaction and feedback effects for forecasting. They allow for more than one evolving variable. All the variables in the model are treated symmetrically in a structural sense; each variable has an equation that explains its evolution based on its own lags (autoregressive carryover effects) and the lags of the other model variables (cross-effects). In this study, the VARX models capture not only the autoregressive carryover and cross-effects on sector interactions and sector performance but also the control effects of market sentiment. Second, VARX models can track the dynamic cumulative effects of the social network in predicting industry value in the short and long terms using *generalized impulse response functions* (GIRFs) [42]. This fact is particularly important because GIRFs can uncover dynamic effects that are not observable with other static models. Third, VARX models can assess the relative contributions of the different metrics of social networks using *generalized forecast error variance decomposition* (GFEVD) [42,43], which is quite helpful for performing hypothesis testing in the study. Recently, VARX models have been broadly adopted in marketing and IS research to investigate the time-series effects of information and economic metrics [28,29,35,44]. Similarly, we use VARX models to estimate complex effects and to determine the full predictive value of social networks.

Table 2
 Sector distribution.

Sector ID	Sector Name	No. of Stocks	Missing Days
1	Energy	28	2
2	Materials	45	0
3	Industry	56	0
4	Optional consumption	32	0
5	Daily consumption	27	0
6	Medical care	25	0
7	Finance	54	0
8	Information and technology	12	2
9	Telecom service	2	7
10	Utility	13	34

4.2. Model specification

We estimate a VARX model for each sector. The endogenous variables include the sector performance (return and idiosyncratic risk), undirected network metrics (inter-sector modularity value and intra-sector modularity value), positive network metrics (inter-sector modularity value and intra-sector modularity value), and negative network metrics (inter-sector modularity value and intra-sector modularity value). We include only one exogenous variable to control the market sentiment effects on sector performance. The VARX model is specified as follows:

$$\begin{bmatrix} \text{Return}_t \\ \text{Risk}_t \\ \text{Intra}_U Q_t \\ \text{Inter}_A UQ_t \\ \text{Intra}_P Q_t \\ \text{Inter}_A PQ_t \\ \text{Intra}_N Q_t \\ \text{Inter}_A NQ_t \end{bmatrix} = \begin{bmatrix} \alpha_1 + \delta_1 t \\ \alpha_2 + \delta_2 t \\ \alpha_3 + \delta_3 t \\ \alpha_4 + \delta_4 t \\ \alpha_5 + \delta_5 t \\ \alpha_6 + \delta_6 t \\ \alpha_7 + \delta_7 t \\ \alpha_8 + \delta_8 t \end{bmatrix} + \sum_{k=1}^K \begin{bmatrix} \phi_{1,1}^k \cdots \phi_{1,8}^k \\ \phi_{2,1}^k \cdots \phi_{2,8}^k \\ \phi_{3,1}^k \cdots \phi_{3,8}^k \\ \phi_{4,1}^k \cdots \phi_{4,8}^k \\ \phi_{5,1}^k \cdots \phi_{5,8}^k \\ \phi_{6,1}^k \cdots \phi_{6,8}^k \\ \phi_{7,1}^k \cdots \phi_{7,8}^k \\ \phi_{8,1}^k \cdots \phi_{8,8}^k \end{bmatrix} \cdot \begin{bmatrix} \text{Return}_{t-k} \\ \text{Risk}_{t-k} \\ \text{Intra}_U Q_{t-k} \\ \text{Inter}_A UQ_{t-k} \\ \text{Intra}_P Q_{t-k} \\ \text{Inter}_A PQ_{t-k} \\ \text{Intra}_N Q_{t-k} \\ \text{Inter}_A NQ_{t-k} \end{bmatrix} + \tau_{1,1} x_{1t} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \\ \varepsilon_{5t} \\ \varepsilon_{6t} \\ \varepsilon_{7t} \\ \varepsilon_{8t} \end{bmatrix} \quad (9)$$

where $\text{Intra}_U Q$, $\text{Intra}_P Q$, and $\text{Intra}_N Q$ represent the intra-sector modularity values in the undirected network, positive network, and negative network, respectively; $\text{Inter}_A UQ$, $\text{Inter}_A PQ$, and $\text{Inter}_A NQ$ represent the average inter-sector modularity values in the undirected network, positive network, and negative network, respectively; $\alpha_i (i = 1, 2, \dots, 8)$ are constants; $\delta_i, \phi_{ij}^k (i, j = 1, 2, \dots, 8)$ are coefficients; $\tau_{1,1}$ is the coefficient of the exogenous variable (sector news sentiment) x_{1t} ; K is the lag length, and $\varepsilon_i (i = 1, 2, \dots, 8)$ are white-noise residuals.

The lag order in the VARX model is usually selected using Schwartz's Bayesian information criterion (SIC) and the final prediction error (FPE) [28,35]. Thus, we select the lag order with the minimized SIC and FPE in each model across 10 industries.

5. Estimation results

5.1. Time-series data

To prepare the daily data for the time-series analysis, we investigate the daily sector return and risk, in addition to daily

Table 3
 Statistics regarding daily company networks.

	Undirected network	Positive network	Negative network
Mean	137	133	30
Maximum	276	274	116
Minimum	87	84	12
Median	120	116	21

social networks based on public news. First, we filter out 6 stocks⁸ that have experienced long-term trading suspensions during this period. We divide the remaining 294 stocks into 10 sections (Table 2). We calculate the days that a sector does not appear in the company network, which indicates that the company network metrics for the sector are missing for those days. Fortunately, we find that few data are missing. The utility sector has 204 valid days of a total of 238 days. We replace these missing data with 0, thus indicating no inter- or intra-sector interactions on that day.

To further investigate the network density, we perform simple statistics for daily firm networks. As shown in Table 3, the daily firm network is not quite sparse, considering undirected (company closeness) networks. Even in the smallest network, 87 companies appear. The situation is quite similar for the positive (cooperative company) networks. However, for the negative (competitive company) networks, the nodes are sparse, with a minimum value of 12 companies, because the identified negative relations are much fewer than positive relations. Imbalances between positive and negative opinions have also been found in previous studies of user-generated content [29].

5.2. Tests for stationarity in the time series

We conduct stationary and unit root tests to examine the stability of sector performance metrics and company network metrics. These tests investigate whether the variables entering the system evolve continually or are stationary. We conduct augmented Dickey–Fuller (ADF) tests to assess stationarity [45]. As reported in Table 4, except for the risk and news sentiment, the results of ADF testing of all the metrics across 10 sectors are less than the critical value of -2.87 , thus leading us to reject the null hypothesis of a unit root at the 95% confidence level. We use the first difference for the risk and sector news sentiment. Furthermore, we find that the corrected data series range from -17.83 to -3.29 (Table 3), thereby indicating that the variable series do not co-integrate in equilibrium [28,46].

5.3. Tests for granger causality

The results of the Granger causality test [47] are reported in Tables 5 and 6. According to the results, we can conclude that several social network metrics have significant time-based causal relationships with sector performance. In Table 5, the undirected network metrics, including the average inter-sector modularity value and the intra-sector modularity value, can Granger-cause returns in sectors 2, 5, 7, 8, and 10. Additionally, the positive network metrics have strong effects on the returns in sectors 2, 7, and 10. However, the intra-sector modularity value in the negative network is sufficiently significant to cause a return only in sector 2 ($p=0.005$), and the average inter-sector modularity value in the negative network only causes a return in industries 2 and 8 ($p=0.004$ and 0.08 , respectively).

⁸ The stock codes are 000156, 000527, 002673, 601238, 601800, and 603993 on a 12-share market.

Table 4
Stationarity test of the endogenous variables.

Sector	Return	ΔRisk	Δns	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	-13.24	-14.85	-8.70	-15.13	-15.13	-15.34	-15.30	-13.74	-13.35
2	-17.10	-11.94	-6.25	-15.07	-15.02	-15.69	-15.62	-4.39	-4.39
3	-17.83	-6.94	-5.30	-14.82	-14.8	-14.2	-14.24	-15.77	-15.81
4	-15.34	-10.71	-5.66	-12.65	-12.33	-13.18	-12.85	-14.80	-14.76
5	-14.76	-12.61	-5.14	-12.26	-12.33	-12.49	-12.48	-13.80	-13.86
6	-13.79	-14.40	-3.65	-8.31	-8.23	-8.3	-8.25	-12.76	-12.72
7	-15.63	-8.19	-7.18	-11.7	-12.3	-12.21	-12.49	-14.80	-15.02
8	-14.76	-13.68	-3.31	-12.24	-12.49	-12.17	-12.33	-14.90	-15.08
9	-15.46	-13.64	-10.60	-3.29	-5.00	-14.56	-5.17	-15.53	-15.40
10	-13.80	-13.24	-5.46	-14.77	-14.76	-14.78	-14.78	-15.59	-15.58

Note: Augmented Dickey Fuller (ADF) test statistic critical value: -2.87 (5% level confidence interval).

Table 5
Granger causality tests on returns.

Sector	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	0.43	0.57	0.70	0.79	0.45	0.53
2	0.02**	0.01***	0.02**	0.02**	0.005***	0.004***
3	0.45	0.27	0.17	0.09*	0.97	0.92
4	0.77	0.17	0.71	0.15	0.61	0.21
5	0.01***	0.009***	0.12	0.13	0.76	0.80
6	0.09*	0.13	0.21	0.25	0.61	0.59
7	0.03**	0.08*	0.007***	0.01***	0.24	0.21
8	0.07**	0.09*	0.15	0.17	0.4	0.08*
9	0.95	0.76	0.95	0.83	0.36	0.36
10	0.01***	0.002***	0.01***	0.006***	0.56	0.54

Note: The estimates of Granger causality are the means of the *p*-values of the joint Wald statistics.

p* < 0.1, ** *p* < 0.05, * *p* < 0.01.

Table 6
Granger causality tests on risk.

Sector	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	0.64	0.45	0.51	0.59	0.56	0.71
2	0.58	0.71	0.63	0.75	0.08*	0.07*
3	0.002***	0.0007***	0.01***	0.004***	0.34	0.42
4	0.003***	0.009***	0.001***	0.007***	0.41	0.73
5	0.77	0.67	0.68	0.63	0.86	0.85
6	0.03**	0.01***	0.72	0.60	0.61	0.64
7	0.71	0.95	0.52	0.67	0.09*	0.07*
8	0.65	0.57	0.57	0.80	0.96	0.73
9	0.84	0.67	0.91	0.69	0.57	0.75
10	0.16	0.06*	0.19	0.09*	0.01***	0.01***

Note: The estimates of Granger causality are the mean of the *p*-values of the joint Wald statistics.

p* < 0.1, ** *p* < 0.05, * *p* < 0.01.

Table 7
Duration of the short- and long-term impacts on return.

Sector	Wear-in						Wear-out					
	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	1	1	1	1	1	1	4	4	4	3	5	5
2	1	2	1	1	1	1	4	6	4	4	6	7
3	2	2	1	1	1	1	7	7	6	6	8	8
4	2	2	2	2	1	1	4	4	4	4	5	5
5	5	5	5	5	2	2	8	8	7	8	8	9
6	4	4	2	2	1	1	6	6	5	5	8	8
7	3	3	3	3	1	1	4	5	5	5	6	6
8	2	3	3	3	1	2	5	6	5	5	7	7
9	1	1	1	1	2	2	4	4	5	6	6	6
10	7	7	7	7	4	4	10	9	9	8	9	10
Average	2.8	3	2.6	2.6	1.5	1.6	5.6	5.9	5.4	5.4	6.8	7.1
Test	Intra_pq + Inter_apq > Intra_nq + Inter_anq						Intra_pq + Inter_apq < Intra_nq + Inter_anq					
F-test	6.14***						37.77***					

Notes: * *p* < 0.1, ** *p* < 0.05, *** *p* < 0.01.

As indicated in Table 6, the results suggest that the undirected network metrics can cause risk in sectors 3, 4, and 6, followed by the negative network metrics in industries 2, 7, and 10 and the positive network metrics in sectors 3 and 4. These results support H1a in that the sector's interactive metrics in company networks have predictive power for sector performance.

5.4. Short- and long-term relationships between company comparative networks and sector performance

We model the variable dynamics based on GIRFs. In this step, we use the estimated parameters of the VARX model ϕ_{ij}^k to generate the GIRFs with $\psi_{ij}(t)$, measuring the net effects of one unit of unexpected change in the social network metrics *i* on the industry value metric *j* at time *t* without assuming a causal ordering [45,48]. We obtain the standard errors by simulating the fitted VARX model using a Monte Carlo method with 1000 runs, and the statistical significance of the parameters is tested. The short-term (immediate predictive value) and long-term (cumulative predictive value) effects are also derived from the GIRFs. We can also assess the dynamics of parameters relative to wear-in time by gauging the number of periods before the peak predictive value is reached and quantify the wear-out time by gauging the number of periods before the stable predictive value is reached.

We first investigate the wear-in and wear-out effects on sector performance. Tables 7 and 8 present the results and averages of the outcomes of the time effects between social networks and sector values across 10 industries. From the results, we find that negative comparative relationships have a shorter wear-in time on return than positive comparative relationships (*F* = 6.14, *p* < 0.01). Simultaneously, negative comparative relationships have a longer wear-out time on return than positive comparative relationships

Table 8
 Duration of the short- and long-term impacts on risk.

Sector	Wear-in						Wear-out					
	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	1	1	1	1	3	3	6	5	5	5	6	6
2	1	1	2	2	1	1	2	1	3	3	5	6
3	3	3	5	5	3	3	5	6	6	6	8	9
4	2	2	2	2	1	1	5	6	5	5	6	6
5	4	4	4	4	3	3	8	9	8	8	9	9
6	3	3	2	2	1	1	6	7	7	7	9	9
7	1	2	2	2	3	3	5	6	5	4	6	6
8	1	3	1	1	1	3	4	5	6	5	7	6
9	2	2	2	2	4	4	5	6	6	6	7	7
10	8	8	8	8	6	6	9	8	9	9	10	10
Average	2.6	2.9	2.9	2.9	2.6	2.8	5.5	5.9	6	5.8	7.3	7.4
Test	Intra_pq + Inter_apq > Intra_nq + Inter_anq						Intra_pq + Inter_apq < Intra_nq + Inter_anq					
F-test	0.16						50.79***					

Notes. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

($F = 37.77, p < 0.01$). Regarding risk measurement, the wear-in time exhibits no significant differences between positive and negative relationships. However, negative comparative relationships do have a longer wear-out time on risk ($F = 50.79, p < 0.01$). This finding is consistent with previous financial studies (Hong et al. [31]) that reported that bad news travels slowly across the public domain and has a longer impact duration. Thus, H3a' is partially supported by the sector return, and H3b' is well supported by both the sector return and risk.

To further investigate the immediate and cumulative impulsive response elasticities, we calculate the change in basis points (one basis point is one-hundredth of a percentage) of sector return or as a percentage of sector risk in response to one unit of unexpected change in sector interactive metrics [28,29]. Taking the finance sector (labeled as 7) as an example, Fig. 4 presents the accumulated impulse responses to sector interactive metrics. From the results presented in Tables 9 and 10, we observe that in the undirected network analysis, an unexpected increase in intra-sector closeness will predict a surge in daily sector return by 9.33 basis points in the short term and the accumulated impact of 12.52 basis points in

20 days. However, an unexpected increase in the inter-sector closeness will immediately predict a decrease in the daily sector return by 9.06 basis points ($p < 0.01$) and accumulated impact of 12.08 basis points ($p < 0.1$). In the positive network, the intra-sector relationship has positive predictive value with returns both in the short term (11.19 basis points, $p < 0.1$) and the long term (14.52 basis points, $p < 0.1$). In the negative network, the intra-sector relationship is positively related immediately with risk (0.062 basis point, $p < 0.1$); however, the inter-sector relationship is negatively related immediately with risk (-0.070 basis point, $p < 0.1$). Although these effects seem to be small in terms of the number of basis points, they have a substantial impact in terms of the dollar value. In monetary terms, the relationships between company network and sector performance could translate into a significant impact on the market capitalization of the sector [29]. For example, holding other factors constant, for the finance sector, one unit of unexpected increase in positive intra-sector could add approximately \$11.19 million to the average market capitalization in the short term and could accumulate approximately \$14.52 million over a 20-day period.

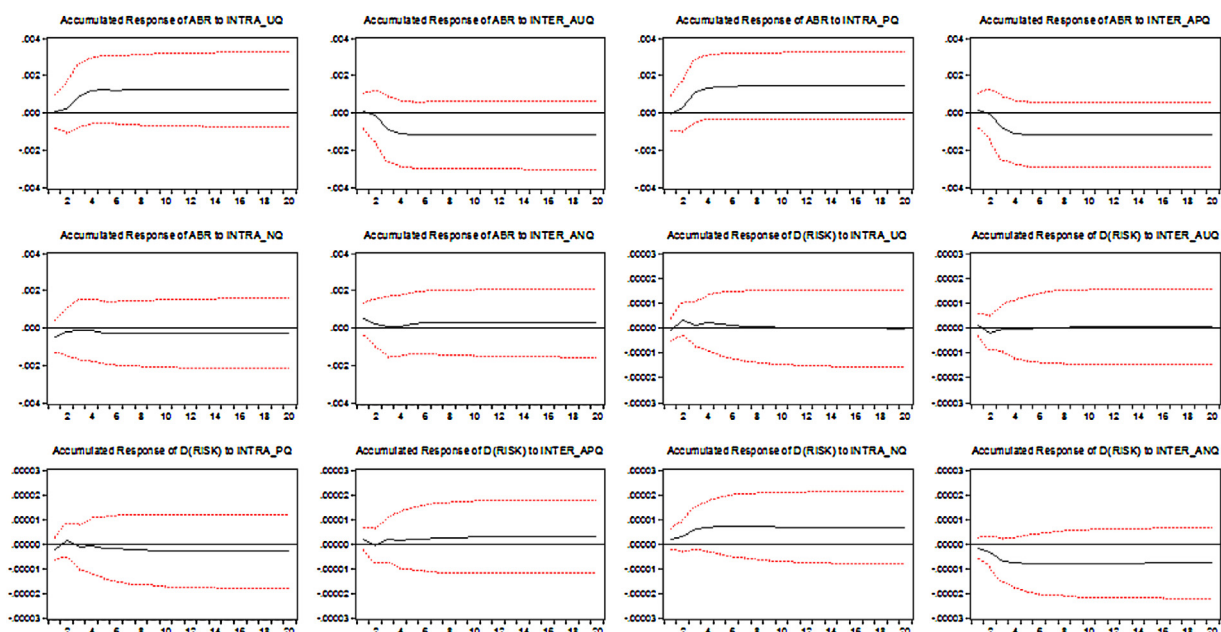


Fig. 4. Accumulated impulse response functions of social network metrics.

Table 9
Impulse response of return to company network metrics.

Sector	Immediate						Accumulate					
	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	12.66*	-12.59*	14.39*	-14.59*	-3.84	2.89	12.75*	-12.79*	14.86*	-15.16*	-4.89	4.13
2	-2.32	2.69	-3.06	3.16	2.46	-2.01	-3.54	3.14	-4.26	3.63	1.01	-1.44
3	0.573	-0.852	2.21	-2.82	-3.44	3.19	2.32	-3.28	4.58	-6.13	-7.96	7.85
4	-0.105	2.85	-2.14	5.07	4.75	-4.42	-1.33	4.34	-3.32	6.51	4.53	-4.00
5	-25.35**	26.06**	-20.16*	19.84*	-0.196	-0.67	-28.62*	29.33*	-26.66*	26.36*	-0.73	-0.34
6	23.02*	-21.94*	14.84	-14.31	24.23	-23.39	27.83	-26.41	19.56	-18.84	28.21	-27.24
7	9.33	-9.06***	11.19*	-8.62	-1.06	0.544	12.52	-12.08*	14.52*	-11.92	-2.90	2.60
8	-8.02	8.03	-7.86	8.89	-7.32	16.11	-5.61	5.93	-5.19	6.41	-11.42	20.58
9	-3.42	8.49	-3.69	7.78	10.49	-4.84	-1.19	7.25	-1.53	7.47	15.52	-6.32
10	21.48*	-21.89*	21.78*	-22.5*	8.84	-9.11	21.38	-21.88	22.4	-23.31	4.27	-4.43

Notes: The coefficients of returns are in basis points (1 basis point=hundredth of a percentage). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10
Impulse response of risk to company network metrics.

Sector	Immediate						Accumulate					
	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	0.116*	-0.118*	0.113*	-0.114*	0.012	-0.016	0.131**	-0.133**	0.125**	-0.127**	0.027	-0.031
2	0.015	-0.017	-0.026	0.022	0.122***	-0.119***	0.011	-0.015	-0.041	0.037	0.172**	-0.175**
3	-0.046	0.055	-0.004	0.011	-0.088**	0.090**	-0.075	0.082	-0.004	0.009	-0.169*	0.169*
4	-0.096***	0.088***	-0.105***	0.089***	0.010	-0.0008	-0.163***	0.159***	-0.173***	0.157***	-0.015	0.028
5	-0.049	0.048	-0.025	0.021	0.016	-0.021	-0.056	0.055	-0.036	0.031	0.039	-0.048
6	0.05	-0.045	0.074	-0.072	-0.015	0.017	0.025	-0.019	0.068	-0.065	-0.081	0.084
7	0.012	-0.003	-0.014	0.016	0.062*	-0.070*	-0.002	0.007	-0.030	0.029	0.065	-0.078
8	0.027	-0.015	0.042	-0.037	-0.203*	0.181*	0.032	-0.016	0.048	-0.044	-0.220**	0.213*
9	-0.027	0.057	-0.065	0.087	0.122	-0.150	0.103	-0.091	0.079	-0.054	0.121	-0.144
10	-0.003	0.001	0.005	-0.008	-0.084	0.081	-0.031	0.027	-0.014	0.008	-0.143	0.139

Notes: The coefficients of risk are in basis points (1 basis point=hundredth of a percentage). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.5. Relative importance of sector-interactive metrics

We assess the relative impact of the company network metrics on sector performance using GFEVD. The GFEVD estimates are derived using the following algorithm:

$$\theta_{ij}(t) = \frac{\sum_{k=0}^t (\psi_{ij}(k))^2}{\sum_{k=0}^t \sum_{j=0}^m (\psi_{ij}(t))^2}, \quad i, j = 1, \dots, m. \quad (8)$$

GFEVD can identify the relative predictive value of all the company network metrics. It is appropriate to test the hypotheses proposed in our article. The relative value of the endogenous

Table 11
Variance decomposition of return explained by company network metrics.

Sector	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	0.26	0.58	0.12	0.09	0.21	0.41
2	0.09	2.18	0.29	1.93	0.76	3.20
3	0.39	1.40	0.95	2.32	2.25	0.30
4	0.10	0.08	0.36	0.29	1.10	0.65
5	3.38	2.52	1.57	2.61	4.83	0.57
6	1.82	2.05	1.01	1.67	0.17	0.93
7	1.03	0.37	0.50	3.77	0.13	1.25
8	0.99	0.73	1.13	0.47	0.17	2.60
9	0.38	0.56	0.18	0.67	0.44	1.81
10	2.25	8.78	2.46	8.30	2.54	2.14
Average	1.07	1.93	0.86	2.21	1.26	1.39
Testing	intra_uq+	inter_auq<	intra_pq+	inter_apq+	intra_nq+	inter_anq
F-test	26.43***					
Testing	intra_pq+	inter_apq<	intra_nq+	inter_anq		
F-test	-0.29					
Testing	intra_uq+	intra_pq+	intra_nq<	inter_auq+	inter_apq+	inter_anq
F-test	2.93*					

Notes: The coefficients of return are percentage values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

variables is established based on GFEVD over 20 days, which is intended to reduce the short-term functions, as suggested in previous research [28,29].

The GFEVD of return and risk is used to assess the importance of sector interactive metrics, and Tables 11 and 12 provide the results. The results suggest the order of contributions in predicting sector return to be inter_apq (2.21%), inter_auq (1.93%), inter_anq (1.39%), intra_nq (1.26%), intra_uq (1.07%), and intra_pq (0.86%). Similarly, in predicting sector risk, the results of the contributions of the sector interactive metrics are ordered as inter_apq (1.87%), inter_auq (1.79%), inter_anq (1.44%), intra_uq (1.38%), intra_pq (1.27%), and intra_nq (1.01%). On the basis of these results, we acknowledge that the total directed network metrics (the positive

Table 12
Variance decomposition of risk explained by company network metrics.

Sector	intra_uq	inter_auq	intra_pq	inter_apq	intra_nq	inter_anq
1	0.92	1.02	0.25	0.64	1.50	0.09
2	0.08	0.12	0.89	0.42	0.36	2.38
3	0.34	1.46	2.41	3.38	1.45	0.64
4	3.80	0.17	0.21	3.93	0.90	0.22
5	0.93	0.30	1.20	0.94	0.78	0.78
6	1.35	5.30	4.00	0.90	0.60	0.63
7	1.80	0.87	0.51	0.49	0.43	2.19
8	0.30	2.88	0.61	0.87	0.27	0.51
9	1.28	1.08	0.48	1.53	1.29	0.97
10	2.96	4.65	2.17	5.62	2.54	5.99
Average	1.38	1.79	1.27	1.87	1.01	1.44
Testing	intra_uq+	inter_auq<	intra_pq+	inter_apq+	intra_nq+	inter_anq
F-test	6.48***					
Testing	intra_pq+	inter_apq<	intra_nq+	inter_anq		
F-test	-1.14					
Testing	intra_uq+	intra_pq+	intra_nq<	inter_auq+	inter_apq+	inter_anq
F-test	2.63*					

Notes: The coefficients of return are percentage values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

and negative network metrics) contribute toward a greater proportion of the variance than the total undirected network metrics (5.72% versus 3.00% for return and 5.59% versus 3.17% for risk). According to the *F* statistics, the differences are statistically significant ($F=26.43, p < 0.01$ for return and $F=6.48, p < 0.01$ for risk), thereby supporting H2a because comparative analysis provides a stronger network indicator than closeness metrics.

Furthermore, the total inter-sector metrics, including *inter_auq*, *inter_apq*, and *inter_anq*, consist of a greater proportion of the variance than the total intra-sector metrics, including *intra_uq*, *intra_pq*, and *intra_nq* (5.53% versus 3.19% for return and 5.10% versus 3.66% for risk). These differences are statistically significant according to the *F* statistics ($F=2.93, p < 0.1$ for return and $F=2.63, p < 0.1$ for risk). Thus, these results support H1b in that the inter-sector metrics have greater predictive power than the intra-sector metrics.

However, the relationship between the positive and negative network metrics is not supported. In a variance decomposition of return, the total negative network metrics account for a larger proportion of variance than the average total positive network metrics across the sectors, and adverse results occur in the variance decomposition of risk. The results are not statistically significant.

5.6. Robustness testing

We conduct several tests to ascertain the robustness of the results. We use alternative measurements of the inter-sector and intra-sector interactions, in addition to different subsamples of industries for the robustness tests. First, we replace the modularity measurement with the weighted link number to gauge the sector interaction. The intra-sector interaction is measured by the weighted link number among stocks within a sector. The metrics *intra_ulin*, *intra_pln*, and *intra_nln* represent the intra-sector weighted link numbers for the undirected network, positive network, and negative network, respectively. Similarly, the inter-sector interaction is gauged by the weighted link number among stocks that belong to different sectors. The metrics *inter_aulin*, *inter_apln*, and *inter_anln* are the inter-sector weighted link numbers for the undirected network, positive network, and negative network, respectively. Because the negative links are less than the positive links, the metrics of the undirected network variables (*intra_ulin* and *inter_aulin*) are strongly correlated with the metrics of the positive network variables (*intra_pln* and *inter_apln*). In this case, we cannot place all the variables into one VARX model. Therefore, we construct two models: model 1 for the undirected company network and model 2 for the directed

company network. This construction enables us to compare two models using the adjusted R^2 [35,48].

Model 1

$$\begin{bmatrix} Return_t \\ Risk_t \\ Intra_{UJ}LN_t \\ Inter_{A}ULN_t \end{bmatrix} = \begin{bmatrix} \alpha_1 + \delta_1 t \\ \alpha_2 + \delta_2 t \\ \alpha_3 + \delta_3 t \\ \alpha_4 + \delta_4 t \end{bmatrix} + \sum_{k=1}^K \begin{bmatrix} \phi_{1,1}^k \cdots \phi_{1,4}^k \\ \phi_{2,1}^k \cdots \phi_{2,4}^k \\ \phi_{3,1}^k \cdots \phi_{3,4}^k \\ \phi_{4,1}^k \cdots \phi_{4,4}^k \end{bmatrix} \begin{bmatrix} Return_{t-k} \\ Risk_{t-k} \\ Intra_{UJ}LN_{t-k} \\ Inter_{A}ULN_{t-k} \end{bmatrix} + \tau_{1,1}x_{1t} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix} \quad (11)$$

Model 2

$$\begin{bmatrix} Return_t \\ Risk_t \\ Intra_{p}LN_t \\ Inter_{A}PLN_t \\ Intra_{N}LN_t \\ Inter_{A}NLN_t \end{bmatrix} = \begin{bmatrix} \alpha_1 + \delta_1 t \\ \alpha_2 + \delta_2 t \\ \alpha_3 + \delta_3 t \\ \alpha_4 + \delta_4 t \\ \alpha_5 + \delta_5 t \\ \alpha_6 + \delta_6 t \end{bmatrix} + \sum_{k=1}^K \begin{bmatrix} \phi_{1,1}^k \cdots \phi_{1,6}^k \\ \phi_{2,1}^k \cdots \phi_{2,6}^k \\ \phi_{3,1}^k \cdots \phi_{3,6}^k \\ \phi_{4,1}^k \cdots \phi_{4,6}^k \\ \phi_{5,1}^k \cdots \phi_{5,6}^k \\ \phi_{6,1}^k \cdots \phi_{6,6}^k \end{bmatrix} \begin{bmatrix} Return_{t-k} \\ Risk_{t-k} \\ Intra_{p}LN_{t-k} \\ Inter_{A}PLN_{t-k} \\ Intra_{N}LN_{t-k} \\ Inter_{A}NLN_{t-k} \end{bmatrix} + \tau_{1,1}x_{1t} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \\ \varepsilon_{5t} \\ \varepsilon_{6t} \end{bmatrix} \quad (12)$$

Using the two models, we obtain the following results. As indicated in Table 13, the R^2 value of model 2 is statistically significantly greater than the R^2 of model 1 ($F=9.59, p < 0.01$ for return and $F=5.90, p < 0.01$ for risk), thus supporting H2a in that the competitive analysis provides a stronger network indicator than the closeness metrics. Additionally, the inter-sector metrics (*inter_apln* and *inter_anln*) account for significantly greater proportions of the variance than the intra-sector metrics (*intra_pln* and *intra_nln*) in model 2: 2.53% versus 1.81% for return ($F=8.03, p < 0.01$) and 5.57% versus 4.23% for risk ($F=2.87, p < 0.1$). To further test the dynamic effects of the company comparative network, we calculate the wear-in and wear-out times in model 2. As indicated in Tables 14 and 15, the negative network metrics (*intra_nln* and *inter_anln*) have significantly shorter wear-in times than the positive network metrics (*intra_pln* and *inter_apln*): 3.1 days versus 4.0 days for return ($F=4.31, p < 0.05$) and 3.5 days versus 4.7 days for risk ($F=3.27, p < 0.05$). The negative network metrics have significantly longer wear-out times than the positive network

Table 13
Results of the VARX model with network link metrics.

Sector	Return		Risk		Variance Decomposition of Return				Variance Decomposition of Risk			
	R^2_1	R^2_2	R^2_1	R^2_2	<i>intra_pln</i>	<i>inter_apln</i>	<i>intra_nln</i>	<i>inter_anln</i>	<i>intra_pln</i>	<i>inter_apln</i>	<i>intra_nln</i>	<i>inter_anln</i>
1	0.021	0.024	0.043	0.045	0.050	0.119	0.135	0.371	0.580	0.733	0.074	0.142
2	0.068	0.095	0.153	0.177	1.448	4.002	1.875	2.597	1.667	4.462	1.278	1.371
3	0.050	0.072	0.247	0.264	1.156	0.673	0.982	2.075	1.471	2.649	0.950	1.915
4	0.112	0.176	0.300	0.384	2.977	2.628	2.441	4.701	4.585	4.144	5.850	4.450
5	0.043	0.052	0.106	0.124	0.129	0.978	0.497	0.225	0.936	0.624	1.094	1.276
6	0.038	0.041	0.012	0.014	0.157	0.087	0.005	0.358	0.088	0.340	0.156	0.112
7	0.058	0.084	0.174	0.244	3.055	3.071	2.059	1.868	14.136	20.947	6.994	10.271
8	0.026	0.029	0.029	0.033	0.202	0.146	0.203	0.044	0.787	0.241	0.092	0.033
9	0.004	0.006	0.029	0.030	0.115	0.357	0.077	0.054	1.381	1.322	0.031	0.022
10	0.048	0.084	0.064	0.071	0.220	0.577	0.312	0.322	0.103	0.487	0.044	0.125
Ave.	0.047	0.066	0.116	0.139	0.951	1.264	0.859	1.262	2.573	3.595	1.656	1.972
Testing	$R^2_1 < R^2_2$		$R^2_1 < R^2_2$		<i>intra_pln</i> + <i>intra_nln</i> < <i>inter_apln</i> + <i>inter_anln</i>				<i>intra_pln</i> + <i>intra_nln</i> < <i>inter_apln</i> + <i>inter_anln</i>			
F-test	9.59***		5.90***		8.03***				2.87*			

Notes: The coefficients of return are percentage values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 14
Duration of the short- and long-term impacts on return.

Sector	Wear-in				Wear-out			
	intra_pln	inter_apln	intra_nln	inter_anln	intra_pln	inter_apln	intra_nln	inter_anln
1	2	1	1	2	4	3	4	5
2	1	1	1	1	8	8	8	9
3	3	3	2	2	7	6	8	8
4	1	1	1	1	4	5	5	5
5	1	3	1	2	7	8	8	9
6	2	1	1	2	5	5	6	5
7	3	4	1	2	7	8	9	9
8	1	3	3	1	6	8	8	9
9	1	1	1	1	5	6	6	7
10	3	4	1	4	6	7	7	9
Average	1.8	2.2	1.3	1.8	5.9	6.4	6.9	7.5
Test	Intra_pln + inter_apln > Intra_nln + inter_anln				Intra_pln + inter_apln < Intra_nln + inter_anln			
F-test	4.31**				57.45***			

Table 15
Duration of the short- and long-term impacts on risk.

Sector	Wear-in				Wear-out			
	intra_pln	inter_apln	intra_nln	inter_anln	intra_pln	inter_apln	intra_nln	inter_anln
1	3	2	1	1	5	5	6	6
2	1	5	1	1	6	8	8	9
3	3	5	3	2	7	8	8	8
4	1	1	1	1	4	4	5	5
5	2	1	3	1	6	6	7	7
6	1	1	1	2	6	5	7	7
7	3	3	1	1	8	8	8	9
8	1	5	1	5	7	9	9	9
9	2	3	1	3	6	6	7	7
10	1	3	3	2	7	8	9	9
Average	1.8	2.9	1.6	1.9	6.2	6.7	7.4	7.6
Test	Intra_pln + inter_apln > Intra_nln + inter_anln				Intra_pln + inter_apln < Intra_nln + inter_anln			
F-test	3.27**				81.00***			

metrics: 14.4 days versus 12.3 days for return ($F=57.45, p < 0.01$) and 15 days versus 12.9 days for risk ($F=81.00, p < 0.01$).

To control outliers and to determine that our results are not driven by one particular sector, we eliminate one sector at a time on a rolling basis and examine the results. The new results remain similar to the original results. Table 16 presents the consistent variance decomposition results for the data excluding sector 1. UNM refers to the undirected network metrics (intra_uq and inter_aq), and DNM denotes the directed network metrics (intra_pq, inter_apq, intra_nq, and inter_anq). IRAM is the intra-sector metrics (intra_uq, intra_pq, and intra_nq), and IERM refers to the inter-sector metrics (inter_aq, inter_apq, and inter_anq).

Table 16
Variance decomposition of return explained by company network metrics.

Sector	Variance Decomposition of Return				Variance Decomposition of Risk			
	UNM	DNM	IRAM	IERM	UNM	DNM	IRAM	IERM
2	2.28	6.18	1.15	7.31	0.19	4.04	1.33	2.91
3	1.79	5.81	3.59	4.02	1.80	7.88	4.20	5.48
4	0.18	2.40	1.56	1.02	3.97	5.26	4.91	4.31
5	5.90	9.58	9.79	5.70	1.24	3.70	2.91	2.03
6	3.87	3.77	2.99	4.65	6.66	6.12	5.95	6.83
7	1.40	5.66	1.66	5.39	2.67	3.62	2.73	3.55
8	1.72	4.37	2.29	3.80	3.18	2.27	1.19	4.26
9	0.94	3.09	0.99	3.04	2.37	4.27	3.05	3.59
10	11.03	15.44	7.24	19.22	7.61	16.32	7.67	16.26
Average	3.23	6.26	3.47	6.02	3.30	5.94	3.77	5.47
Test	UNM < DNM		IRAM < IERM		UNM < DNM		IRAM < IERM	
F-test	38.77***		2.82*		6.44***		3.20*	

Notes: The coefficients of return are percentage values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6. Discussion and conclusions

This study aims to construct an effective company relationship network using big data and to investigate the dynamic relationships between sector interactions and stock sector performance. The results suggest that company networks constructed based on public news provide predictive indicators for sector performance and that inter-sector interaction has a stronger predictive power than intra-sector interaction. Moreover, in the network construction, comparative analysis provides a better method than closeness analysis. The negative interactions have a shorter reaction time than the positive interactions for return, and they have longer effects for both sector return and risk. These findings are also confirmed using the links as alternative metrics to reflect the interactions between sectors. Collectively, these findings provide important implications for research regarding market structure and stock sector performance.

6.1. Theoretical implications

This study contributes to the IS and finance literature in several aspects. First, the network analysis method has been widely used in IS, focusing on the relationships among social entities, and it is an important addition to standard social and behavioral research. For example, the network effects and personal influences relevant to product sales have been investigated [49,50]. Social communication and mood influences have been used to study information effects on stock prices [4,51]. In contrast to these studies of social influence, the present study focuses on the structure of company comparative networks and demonstrates how sector interactions

have a predictive relationship with stock sector performance. The constructed company network is quite different from previous social networks. It describes the relationships between objective entities. The links between nodes are built based on a machine-learning algorithm instead of using observations. The network construction and analysis method inspires social influence research from a technical perspective.

Second, we present comparative analysis in network construction. In contrast to previous marketing studies that used comparative analysis for sales predictions [15,52], we examine the predictive power of the company comparative network for stock sector performance. Our study first unveils the correlations between the positive (negative) sector interactions and sector performance. Although more positive than negative interactions are found, we observe that the negative interactions have more rapid effects on returns and that they have longer impacts on both returns and risk. Thus, this study motivates us to explore sentiment analysis between sector interactions in IS and finance.

Finally, previous finance studies have demonstrated that the network structure between sectors affects sector performance [1–3]. We agree with this finding and extend the breadth of research by introducing sector interaction metrics and time-series models. This study investigates both short-term effects and long-term and cumulative effects. Furthermore, we evaluate the dynamic effects of multiple interaction relationships (inter-, intra-, positive, and negative) with VARX models. Thus, this study provides a comprehensive and dynamic approach for both market structure and financial research.

6.2. Practical implications

This study contributes to sector-level strategies. First, both inter- and intra-sector interactions have predictive power for stock sector performance. This finding suggests that companies should strengthen their ties within an industry. For example, they can establish industry associations and frequently hold domain conferences. Simultaneously, companies should also encourage interactions between sectors, such as cooperation with companies in upstream or downstream industries.

Second, because the constructed company comparative network significantly influences sector performance, companies should pay attention to public media information. They should strengthen efforts to promote public propaganda for improving exposure and should also monitor the company interactive dynamics reported by various media outlets. The shortest wear-in time can provide an early warning signal to companies regarding future damage to sector performance, particularly when competitive or negative interactions occur. The company network also provides a good visualization method for understanding the market network structure.

Third, the predictive model contributes to portfolio and risk management. Investors can apply the company comparative analysis and sector interactive analysis methods to predict sector returns and risks on a daily basis.

6.3. Limitations and future research

Nevertheless, this study has several limitations that should be addressed in future research. First, we control for few exogenous variables. In this study, we use only news sentiment to control for market environment. In fact, there are many other factors that can have impact on sector performance. For example, the web search volume concerning a stock could indicate a dynamic “hot spot” in the market. Other likely control variables include sector productivity and profits. Second, we have noted that different sectors exhibit different reactions, potentially due to sector properties.

Therefore, analyzing the sector-specific results could be an important undertaking. Third, we propose that our results can be applied to portfolio and risk management. We intend to conduct future experiments using real-world data to test the effectiveness of the model for investing.

Acknowledgement

This paper was supported by the Shenzhen Fundamental Research Grant (No.: JCYJ2010417105742712).

References

- [1] D. Acemoglu, et al., The network origins of aggregate fluctuations, *Econometrica* 80 (5) (2012) 1977–2016.
- [2] K.R. Ahern, J. Harford, The importance of industry links in merger waves, *J. Finance* 69 (2) (2014) 527–576.
- [3] D. Aobdia, J. Caskey, N.B. Ozel, Inter-industry network structure and the cross-predictability of earnings and stock returns, *Rev. Account. Stud.* 19 (3) (2013) 1191–1224.
- [4] B. Han, L. Yang, Social networks, information acquisition, and asset prices, *Manage. Sci.* 59 (6) (2013) 1444–1457.
- [5] Y. Jin, et al., Mining dynamic social networks from public news articles for company value prediction, *Soc. Netw. Anal. Min.* 2 (3) (2012) 217–228.
- [6] Z. Ma, O.R. Sheng, G. Pant, Discovering company revenue relations from news: a network approach, *Decis. Support Syst.* 47 (4) (2009) 408–414.
- [7] Z.P. Ma, Gautam Sheng, R.L. Olivia, Mining competitor relationships from online news: a network-based approach, *Electr. Comm. Res. Appl.* 10 (2011) 418–427.
- [8] Z.G. Zhang, Chenhui Guo, Paulo Goes, Product comparison networks for competitive analysis of online word-of-mouth, *ACM Trans. Manage. Inform. Syst. (TMIS)* 3 (4) (2013) 20:1–20:22.
- [9] G.G. Creamer, Y. Ren, J.V. Nickerson, Impact of dynamic corporate news networks on asset return and volatility, *Social Computing (SocialCom), 2013 International Conference On*. 2013. IEEE (2016).
- [10] B. Handjiski, Enhancing Regional Trade Integration in Southeast Europe, *World Bank Publications*, 2010.
- [11] R.J. Ruffin, The Nature and Significance of Intra-industry Trade, 4, *Economic and financial review-federal reserve Bank of Dallas*, 1999, pp. 2–16.
- [12] T.J. Moskowitz, G. Mark, Do industries explain momentum? *J. Finance* 54 (4) (1999) 1249–1290.
- [13] C.S. Asnes, R.B. Porter, R.L. Stevens, Predicting stock returns using industry-relative firm characteristics. Available at SSRN 213872, (2000).
- [14] Y. Chen, Q. Wang, J. Xie, Online social interactions: a natural experiment on word of mouth versus observational learning, *J. Market. Res.* 48 (2) (2011) 238–254.
- [15] Z. Zhang, X. Li, Y. Chen, Deciphering word-of-mouth in social media: text-based metrics of consumer reviews, *ACM Trans. Manage. Inform. Syst. (TMIS)* 3 (1) (2012) 5.
- [16] W. He, et al., A novel social media competitive analytics framework with sentiment benchmarks, *Inform. Manage.* 52 (7) (2015) 801–812.
- [17] N. Jindal, B. Liu, Identifying comparative sentences in text documents, *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, 2006.
- [18] N. Jindal, B. Liu, Mining comparative sentences and relations, *AAAI* (2006).
- [19] K. Xu, et al., Mining comparative opinions from customer reviews for Competitive Intelligence, *Decis. Support Syst.* 50 (4) (2011) 743–754.
- [20] J.R. Anderson, G.H. Bower, *Human Associative Memory*, Psychology press, 1973.
- [21] Q. He, Knowledge discovery through Co-Word analysis, *Library Trends* 48 (1) (1999) 133–159.
- [22] S. Goel, H.A. Shawky, Estimating the market impact of security breach announcements on firm values, *Inform. Manage.* 46 (7) (2009) 404–410.
- [23] X. Li, et al., News impact on stock price return via sentiment analysis, *Knowledge-Based Syst.* 69 (2014) 14–23.
- [24] Y. Yu, W. Duan, Q. Cao, The impact of social and conventional media on firm equity value: a sentiment analysis approach, *Decis. Support Syst.* 55 (4) (2013) 919–926.
- [25] P.C. Tetlock, M. Saar-Tsechansky, S. Macskassy, More than words: quantifying language to measure firms' fundamentals, *J. Finance* 63 (3) (2008) 1437–1467.
- [26] W.S. Chan, Stock price reaction to news and no-news: drift and reversal after headlines, *J. Financial Econ.* 70 (2) (2003) 223–260.
- [27] P.N. Van, A Good News or Bad News Has Greater Impact on the Vietnamese Stock Market? *Banking Academy of Vietnam*, State Bank of Vietnam, 2015.
- [28] X. Luo, J. Zhang, W. Duan, Social media and firm equity value, *Inform. Syst. Res.* 24 (1) (2013) 146–163.
- [29] S. Tirunillai, G. Tellis, Does chatter matter? The impact of online consumer generated content on a firm's financial performance, *Market. Sci.* 31 (2) (2012) 198–215.
- [30] H. Hong, J.C. Stein, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *J. Finance* 54 (1999) 2143–2184.

- [31] H. Hong, T. Lim, J.C. Stein, Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies, *J. Finance* 55 (1) (2000) 265–295.
- [32] E. Fama, Efficient capital markets: a review of theory and empirical work, *J. Finance* 25 (2) (1970) 383–417.
- [33] A. Singhal, Modern information retrieval: a brief overview, *Bull. IEEE Comput. Soc. Tech. Comm. Data Eng.* 24 (4) (2001) 35–43.
- [34] K. Dejaeger, T. Verbraken, B. Baesens, Towards comprehensible software fault prediction models using Bayesian network classifiers, *IEEE Trans. Software Eng.* 39 (2) (2013) 237–257.
- [35] X. Luo, J. Zhang, How do consumer buzz and traffic in social media marketing predict the value of the firm? *J. Manage. Inform. Syst.* 30 (2) (2013) 213–238.
- [36] M.E. Newman, M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E* 69 (2) (2004) 026113.
- [37] Z. Bu, et al., A fast parallel modularity optimization algorithm (FPMQA) for community detection in online social network, *Knowledge-Based Syst.* 50 (2013) 246–259.
- [38] M.E. Newman, Modularity and community structure in networks, *Proc. Natl. Acad. Sci.* 103 (23) (2006) 8577–8582.
- [39] E.A. Leicht, M.E. Newman, Community structure in directed networks, *Phys. Rev. Lett.* 100 (11) (2008) 118703.
- [40] M.E. Newman, Analysis of weighted networks, *Phys. Rev. E* 70 (5) (2004) 056131.
- [41] E. Rubin, A. Rubin, The impact of business intelligence systems on stock return, *Inform. Manage.* 50 (2–3) (2013) 67–75.
- [42] H.H. Pesaran, Y. Shin, Generalized impulse response analysis in linear multivariate models, *Econ. Lett.* 58 (1) (1998) 17–29.
- [43] G. Koop, M. Pesaran, S. Potter, Impulse response analysis in nonlinear multivariate models, *J. Econom.* 74 (1996) 119–147.
- [44] G. Adomavicius, J. Bockstedt, A. Gupta, Modeling supply-side dynamics of IT components, products, and infrastructure: an empirical analysis using vector autoregression, *Inform. Syst. Res.* 23 (2) (2012) 397–417.
- [45] M.G. Dekimpe, D.M. Hanssens, Sustained spending and persistent response: a new look at long-term marketing profitability, *J. Market. Res.* 36 (4) (1999) 397–412.
- [46] J.D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, NJ, 1994.
- [47] C. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* 37 (3) (1969) 424–438.
- [48] S.V. Srinivasan, Koen Marc; Pauwels, Minset metrics in market response models: an integrative approach, *J. Market. Res.* 47 (3) (2010) 672–684.
- [49] G.R. Gonzalez, D.P. Claro, R.W. Palmatier, Synergistic effects of relationship managers' social networks on sales performance, *J. Market.* 78 (1) (2014) 76–94.
- [50] E. Moretti, Social learning and peer effects in consumption: evidence from movie sales, *Rev. Econ. Stud.* 78 (1) (2011) 356–393.
- [51] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, *J. Comput. Sci.* 2 (1) (2011) 1–8.
- [52] N. Archak, A. Ghose, P.G. Ipeirotis, Deriving the pricing power of product features by mining consumer reviewers, *Manage. Sci.* 57 (8) (2011) 1485–1509.

Kun Chen is an assistant professor in the Department of Finance at South University of Science and Technology of China. She received her Ph.D. from the Department of Information Systems at the City University of Hong Kong. Dr Chen's research deals with business intelligence, text mining, and big data analytics. She has published in academic journals such as *INFORMS Journal on Computing* and *Journal of Management Information Systems*.

Peng Luo is a Ph.D. student in Harbin Institute of Technology. His research focuses on network topology and social networks. Mr. Luo has published in academic journals such as *Physica A*, *Journal of Informetrics*, and *Management Decisions*.

Dongming Xu is a senior lecturer in Business of Information Systems at the University of Queensland Business School and has a Ph.D. in the area of information systems from the City University of Hong Kong. Her interests include research knowledge management, eLearning effectiveness, financial monitoring management systems, electronic commerce, and intelligent agent business applications. She has published in academic journals such as *Information & Management* and *Decision Support Systems*.

Huaiqing Wang is a professor in the Department of Finance at South University of Science and Technology of China. He is also the Honorary Dean and a Guest Professor of the School of Information Engineering, Wuhan University of Technology, China. He received his Ph.D. from University of Manchester, UK, in 1987. Dr. Wang specializes in research on financial intelligence and intelligent systems (such as intelligent financial systems, intelligent learning systems, business process management systems, knowledge management systems, conceptual modeling, and ontology). He has published more than 70 international refereed SCI/SSCI journal articles and received more than 700 SCI citations.