THE DESIGN OF A NETWORK-BASED MODEL FOR BUSINESS PERFORMANCE PREDICTION

Completed Research Paper

Wenping Zhang
Department of Information Systems
City University of Hong Kong
Hong Kong
wzhang23-c@my.cityu.edu.hk

Raymond Y. K. Lau
Department of Information Systems
City University of Hong Kong
Hong Kong
raylau@cityu.edu.hk

Abstract

While much research work has been devoted to analysis and prediction of individuals' behavior in social networks, very few studies about the analysis of business networks are conducted. Empowered by recent research on automated mining of business networks, this paper illustrates the design of a novel business network-based model called Energy Cascading Model (ECM) for the analysis and prediction of business performance using the proxies of stock prices. More specifically, the proposed prediction model takes into account both influential business relationships and twitter sentiments of firms to infer their stock price movements. Our empirical experiments based on a publicly available financial corpus and social media postings reveal that the proposed ECM model is effective for the prediction of directional stock price movements. The business implication of our research is that business managers can apply our design artifacts to more effectively analyze and predict the potential business performance of targeted firms.

Keywords: Business Networks, Twitter Sentiments, Business Performance Prediction, Stock Volatility, Design Science

Introduction

Recent empirical studies have revealed that a firm's relationships with other firms may have direct influence on its strategic competitiveness, and hence its business performance (Bernstein 2003; Lau et al. 2012; Ma et al. 2009). Such an influence may be explained with reference to the structural embeddedness theory (Gnyawali and Madhavan 2001) and Porter's five forces model (Porter 1980). These theories provide the theoretical ground for the design of a network-based model for the analysis and prediction of business performance. For example, the structural embeddedness theory posits that a firm's specific position in an inter-firm network influences its competitive behavior due to the firm's unique way of access to external assets and information through the network (Gnyawali and Madhavan 2001). Although a large body of work has been devoted to business performance analysis based on the proxies of stock prices, existing prediction models are mainly based on the analysis of macro-level industrial and economic factors (King 1966; Knack and Keefer 1995; Schumaker and Chen 2010) or the micro-level management style and business strategy factors (Chang and Liu 2008; Johnson and Magee 1985; Loomis 1985). This paper illustrates the design of a novel business network-based model called Energy Cascading Model (ECM) for the analysis and prediction of firms' business performance through the proxies of their stock prices.

The first modern stock exchange market was established in the 17th century (Petram 2011). Since then, the operations of stock markets have been widely studied (Johnson and Magee 1985; King 1966; Ofer and Thakor 1987; Schumaker and Chen 2009; Vega 2006) due to the considerable growth of capital and economic development in the 20th century (Case et al. 2005; Levine and Zervos 1996; Ludvigson and Charles 1999; Poterba 2000; Summers 1986). Our research begins with analyzing firms' performance based on their stock prices because historical stock data are readily available at major stock exchange and financial services web sites. In the past, researchers believed that stock price movements could be explained according to the random walk theory or the efficient-market hypothesis (Cootner 1964; Fama 1965; Fama 1991). Accordingly, stock price movements of a firm are considered random and unpredictable. However, recent research findings reveal that stock price movements may not simply follow a random walk process, and stock volatility is predictable to a certain extent (Chaudhuri and Wu 2003; Lo and MacKinlay 2011).

Though a lot of research work has been devoted to the analysis of individuals' behavior in social networks, very few studies about analyzing firms' performance with respect to business networks are reported in the literature. Our research work reported in this paper aims to fill such a research gap. In particular, we employ stock prices as the proxies to measure firms' performance. It is generally believed that stock prices can reveal traders' expectations about a firm's business performance and other economic variables (Elton et al 1981; Feldman et al. 1997). According to the extended five forces model (Brandenburger and Nalebuff 1996), a firm's rival force, supplier force, and complementary force are among the most significant factors that influence its competitiveness, and hence affect its ultimate business success. Accordingly, it is an intuitive assumption that a firm's relationships (e.g., collaborative supply-chain relationships) with other firms in an inter-firm network can be used as the basis to infer its stock price movements. To illustrate the intuition of our business network-based approach for predicting directional stock price movements, the stock price movements of two fierce rivals, Nokia (the former cellphone giant) and Apple Inc. (the current giant in the cellphone market) pertaining to the period from 2008 to 2012 are plotted in Figure 1. It is easy to observe that the stock price trends of Nokia and Apple Inc. are almost the opposite. More specifically, when the stock price of Apple Inc. rose due to the release of its new products, there was a corresponding slump of Nokia's stock price. In contrast, when the stock price of Apple Inc. fell (e.g., due to the release of the disappointing iPhone 5), the stock price of Nokia climbed up in the corresponding period.

One of the reasons why little research work about analyzing a firm's performance with respect to a business network is that information about business networks is not readily available. Although there are abundant social networks for individuals (e.g. Twitter, Facebook, Linkedin, etc.), to the best of our knowledge, there are not such counterparts for commercial firms such that business managers can easily identify the explicit and latent business relationships of targeted firms. Business consulting agencies such as Hoover and Mergent charge their commercial clients with a fee for extracting information about the business relationships of targeted firms. Since these agencies mainly employ a manual approach for collecting business intelligence, the information about business relationships is often restricted to a

limited scale and the information may not be up-to-dated to capture evolving relationships among firms. Fortunately, grounded in previous work of automated business network discovery (Lau et al. 2012; Zhang et al. 2012), it becomes technically feasible to automatically or semi-automatically construct large-scale business networks based on financial text corpora such as online financial news articles, investors' comments, and experts' financial reports. In particular, dynamic business networks corresponding to different periods can be built instantly by feeding different financial text corpora into the business network discovery system. Accordingly, a large-scale (e.g., examining most companies of different business sectors) network-based approach for business performance analysis is feasible under the emerging trend of big data analytics in business intelligence research (Chen et al. 2012; Tan 2012).
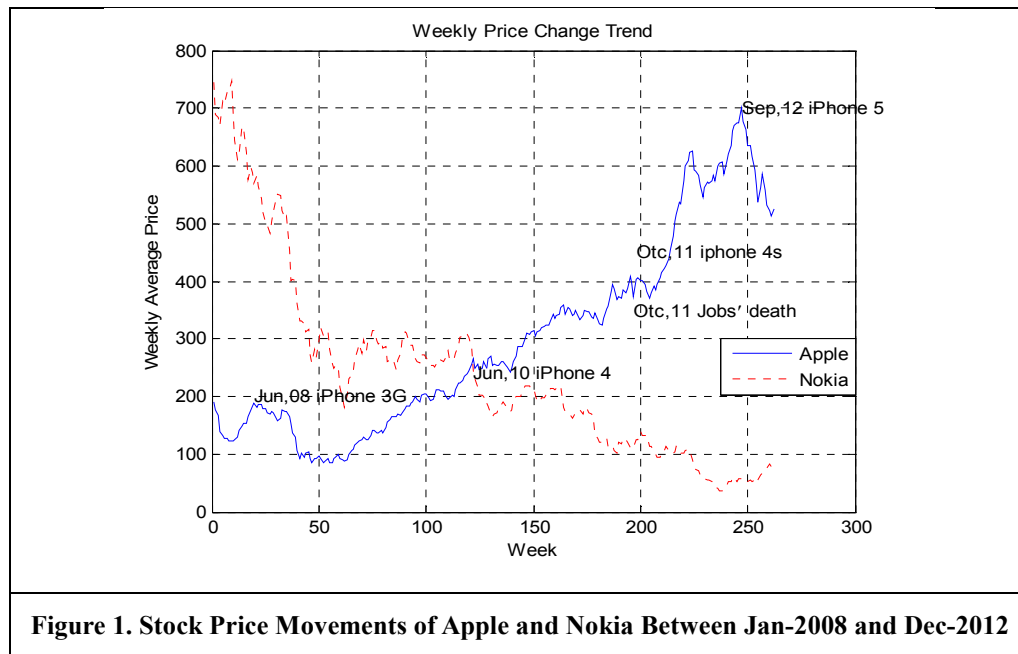
Figure 1. Stock Price Movements of Apple and Nokia Between Jan-2008 and Dec-2012

Guided by the design science research methodology (Hevner et al. 2004; March and Storey 2008; Peffers et al. 2008), the main theoretical contributions of our research work are as follows: (1) designing several novel metrics to measure a firm's stock price movements; (2) designing the novel network-based Energy Cascading Model for the analysis and prediction of firms' business performance using the proxies of stock prices; (3) designing an instantiation (i.e., a prototype service) to analyze the directional stock price movements of firms listed in the S&P 500 index and evaluate the effectiveness of the proposed ECM model. The practical implication of our research is that business managers and financial analysts can apply our design artifacts to more effectively analyze and predict the business performance of targeted firms in a timely manner. As a result, proactive corporate actions can be taken to streamline the business operations of these firms, and therefore enhance their competitive power. Moreover, business consultancy agencies such as Hoover and Mergent can apply our design artifacts to better monitor their targeted firms and extract accurate business intelligence instantly.

Related Work

In the field of economics and finance, researchers found that a firm's stock price often increased as a response to the firm's announcement of a stock repurchase or an increasing dividend (Aharony and Swary 1980; Dann 1981; Loomis 1985; Stewart 1976; Vermaelen 1981). Moreover, stock prices often exhibit a slight "aftermarket" drop following a surge in responding to the announcement of a repurchase (Vermaelen 1981). Ofer and Thakor (1987) built a signaling model to explain such a phenomenon. King (1966) investigated several market and industry factors by examining the historical stock data of 63

securities listed at the New York Stock Exchange (NYSE) from June, 1972 to December, 1960, and found that these factors had significant influences on stock price movements.

Johnson and Magee (1985) analyzed stock price movements in responding to the sudden deaths of top executives of some firms, and they found that stock volatility did occur due to this kind of event; the possible reason of such a short-term stock volatility is that shareholders hold different expectations for the incumbent and replacement of top executives. Smirlock and Starks (1988) examined the relationship between stock price and trading volume. A significant lagged relationship between absolute price change and volume was found, especially in short periods immediately preceding or following quarterly earnings announcement. Jones and Litzenberger (1970) found that stock market did not adjust fully and instantaneously with respect to firms' quarterly earnings reports.

By studying stock price movements with respect to public and private information, Vega (2006) discovered that the abnormal drifts of stock prices had a negative correlation with the information (either public or private) held by investors. Such a finding is consistent with the rational uncertainty theory (Brav and Heaton 2002) in that the arrival rate of informed traders is negatively correlated with structural uncertainty and stock price drifts. On the other hand, Chen et al. (2007) focused on the relationship between private information held by investors and the corporate investment on stock. They found a strong positive correlation between the amount of private information held by investors and the investment-to-price sensitivity.

In the field of information systems, researchers tried to find rules from the historical financial time-series data to predict future stock prices (Al-Qaheri et al. 2008; Chang and Liu 2008; Lu et al. 1998). Moreover, some researchers tried to apply machine leaning algorithms to predict stock prices (Bollen et al. 2011; Ince and Trafalis 2008; Saad et al. 1998). Among the state-of-the-art machine learning algorithms, Artificial Neural Networks (ANNs) have been widely used to predict future stock price movements (e.g. stock price labeled with date) (Charkha 2008; Kohara et al. 1997; Saad et al 1998; Tan and Wittig 1993). Apart from ANNs, other machine learning methods such as Support Vector Machine (SVM) was applied to stock price forecasting (Ince and Trafalis 2004; Ince and Trafalis 2008; Pai and Lin 2005).

Financial texts such as online financial news articles, investor blogs, user-generated stock investment messages, and so on have been shown of great values for analyzing business activities (Bollen et al. 2011; Fama 1965; Schumaker and Chen 2009). Schumaker and Chen (2009, 2010) applied text mining techniques to build the AZFinText System which leveraged breaking financial news articles to predict the directional movements of stock prices. They achieved a prediction accuracy of 57.8% and obtained a financial return of 2.06% in the best case. Bar-Haim et al. (2011) examined the relationships between stock tweets and stock prices, and found that it was beneficial to make stock investment decisions based on experts' micro-blogging messages about stocks (e.g., stock tweets). Such a finding is consistent with the opinion leadership phenomenon (Nair et al. 2010; Schiffman and Gaccione 1974) and the concept of expertise effect (Thomas-Hunt et al. 2003).

Recently, there are some studies about applying financial texts (e.g., news articles, tweets, etc.) to predict stock performance (Bollen et al. 2011; Das and Chen 2007; Schumaker and Chen 2009; Tetlock et al. 2008). One of the most important processes is to mine the sentiments embedded in these text corpora. Previous research has shown the value of applying opinion mining and sentiment analysis techniques to solve practical problems (Liu 2010; Pang and Lillian 2008; Zhang 2008). Some researchers leveraged manually constructed sentiment lexicons and natural language processing (NLP) techniques to analyze the sentiments embedded in user-generated contents (Kouloumpis et al. 2011; Naveed 2011; Saif et al. 2011), while other researchers applied machine learning techniques to mine opinions embedded in these contents (Bermingham and Smeaton 2010; Celikyilmaz et al. 2010; Tsur et al. 2010).

In the fields of psychology and behavioral finance, empirical findings show that human decisions are often driven by emotions and moods rather than objective information (Dolan 2002; Kahneman and Tversky 1976; Schwarz 2000). Accordingly, Bollen et al. (2011) applied the Google-Profile of Mood States (GPOMS) to measure a variety of investor moods embedded in tweets. They found that the GPOMS instrument was more effective than traditional sentiment lexicons such as the OpinionFinder for sentiment analysis. Although their experimental method might be controversial, such a finding revealed the potential predictive power of user-generated contents on the Internet. Go et al. (2009) also found that traditional

sentiment lexicons could be extended to effectively perform sentiment analysis based on user-generated comments on the Internet.

To the best of our knowledge, none of the research work reported in existing literature has examined the method of leveraging the business relationships of a firm captured in an inter-firm network to predict its directional stock price movements. In particular, we combine both the sentiments of a targeted firm and the sentiments of related firms captured in an inter-firm network to enhance prediction effectiveness. Our proposed ECM model for the prediction of directional stock price movement is novel, and it is different from the classical machine learning based methods.

The Measures of Directional Stock Price Movements

The historical stock performance data are retrieved through the Yahoo Finance Application Programming Interface (API)[1]. More specifically, we downloaded fields such as "Date of Trading" (t), "Open Price" (O_t), "Highest Price" (H_t), "Lowest Price" (L_t), "Closing Price" (C_t), "Volume of Trading" (V_t), and "Adjusted Closing Price" (AC_t) of each stock under evaluation. To leverage these historical data to identify meaningful stock price movement patterns, we design the following quantitative measures:

Definition 1
Daily Average Price (DAP) *for day t* (Torgo 2010):

$$\text{DAP}_t = \frac{C_t + H_t + L_t}{3} \tag{1}$$

where C_t, H_t, and L_t are the closing, the highest, and the lowest prices of a stock on the trading day t. This measure is used to approximate the daily average stock price.

Definition 2
Weekly Average Price (WAP) *for week i:*

$$WAP_i = \frac{\sum_{t \in T_i} DAP_t}{|T_i|} \tag{2}$$

where T_i = {all the trading days of week i}. Since stock prices may exhibit slight random movements in the short-term, we design the WAP measure to exclude these random movements so as to improve the prediction accuracy of the proposed ECM model.

Definition 3
Change of Weekly Average Price (C-WAP) *for week i:*

$$CWAP_i = WAP_i - WAP_{i-1} \tag{3}$$

where WAP_i and WAP_{i-1} are the weekly average price of a stock in the current period (week) and the immediately preceding period, respectively.

Definition 4
Change Rate of Weekly Average Price (CR-WAP) *for week i:*

$$CRWAP_i = \frac{WAP_i - WAP_{i-1}}{WAP_{i-1}} \tag{4}$$

CR-WAP is applied to normalize the absolute values of weekly price changes of different stocks. For instance, the stock price of Apple Inc. is around $150 per share, whereas Nokia's stock price is about $40. While $10 may only be a small change with reference to the stock price of Apple Inc., the same amount is considered a large stock price movement for Nokia. The proposed *CR-WAP* measure can take into account the relative stock price movements of different stocks which may have large differences in absolute stock values.

 Definition 5

Change of Change Rate of Weekly Average Price (CCR-WAP) *for week i:*

---

[1] http://finance.yahoo.com/

$$CCRWAP_i = CRWAP_i - CRWAP_{i-1} \qquad\qquad (5)$$

We propose *CCR-WAP* to detect the sharp increment or decrement of stock prices. For instance, both Apple Inc. and Samsung's stock prices are rising in 2012 because of the booming smart phone market. However, the rates of increment of these firms vary. With the booming market of Samsung's Android-based smart phones, the rate of increment of Samsung's stock price is rising. In contrast, Apple's rate of increment is falling. The proposed *CCR-WAP* measure is able to capture fierce changes of stock prices.

Definition 6
Abnormal Change Points (ACPs):

The ACPs of a firm's stock price are defined as the time points when its stock price exhibits fierce changes (a sharp increment or a sharp decrement).

We apply the Change of Change Rate of Weekly Average Price (*CCR-WAP*) to identify the Abnormal Change Points (*ACPs*) of a firm. If the *CCR-WAP* of a stock pertaining to a time point is larger than a given threshold, the stock price movement of that time point is considered as an *ACP*. Such a detection threshold is empirically established based on the average *CCR-WAP* of a firm. Given some known ACPs of a stock, we want to predict the occurrence of ACPs of related stocks based on the business relationships of the corresponding firms captured in an inter-firm network.

It should be noted that the series of abnormal change points of two stocks may not demonstrate a one-to-one correspondence even if the corresponding firms are closely related. In other words, not every abnormal change of stock price pertaining to a firm leads to the abnormal change of stock price of a related firm. One possible reason is that an ACP could be triggered by multiple influences from several related firms at the same time. Moreover, an ACP could be caused by other factors rather than business relationships. To address these issues, the novel influence propagation method empowered by the proposed ECM model supports the propagation of multiple influences among firms. In addition, the sentiment analysis and sentiment propagation method of the ECM model aims to capture other factors (e.g., investors' emotion) that may lead to the occurrence of an ACP.

The Design of a Network-based Prediction Model

A basic approach to analyze the impact of firms' relationships on their stock price movements is to apply standard correlation analysis to each pair of firms. For example, the stock volatility of firm X is predicted according to the Pearson correlation coefficient between firms X and Y, and the known stock volatility of firm Y. The problem of such an approach is that only one direct business relationship between two firms is taken into account each time. Consequently, the prediction may be inaccurate due to a partial view of a firm's relationships with other firms.

By way of illustration, two simplified business networks are shown in Figure 2. For the convention adopted in Figure 2, a solid line indicates a collaborative business relationship, whereas a dash line shows a competitive business relationship. In Figure 2-a1, we can see that Google is an important partner (with high collaborative scores) for both Samsung and HTC. On the other hand, HTC and Samsung are fierce competitors to each other. If we only consider the partnership between Google and Samsung, and the partnership between Google and HTC in isolation, the good business performance of Google should positively influence both Samsung and HTC. However, Figure 2-a2 reveals a different reality. Benefited by the release of Android 4.1 in June 2012, which was praised as "the first time as fast and smooth as Apple's ios", both Google and Samsung's stock prices rose. Nevertheless, HTC's stock price slumped sharply in the same period.

On the other hand, Figures 2-b1 and 2-b2 show that the stock price of Clorox and the stock price of Church & Dwight rose following the surge of the stock price of the retailer giant Wal-Mart. However, Clorox and Church & Dwight were rivals to each other. These examples show that it may not be appropriate to evaluate the impact of business relationships on firms' stock prices in isolation. Instead, we should apply a business network as the context to simultaneously consider all the business relationships (direct or indirect) of a targeted firm to infer its stock volatility. For example, the release of Android did positively influence the sales of smart phones manufactured by Samsung and that produced by HTC. However, the increasing market share of Samsung jeopardized HTC's sales because Samsung and HTC

were direct competitors to each other. When the negative influence from Samsung to HTC outweighed the positive influence that HTC might receive from Google, the business performance of HTC as reflected by its stock price suffered. On the other hand, although Clorox and Church & Dwight were main rivals, the competition (i.e., negative influence) between them was relatively small when compared to their strong partnerships (i.e., positive influence) with Wal-Mart. So, the positive influence that Clorox and Church & Dwight received from Wal-Mart outweighed the slight negative influence trigged by the competitive relationship between themselves. As a result, both Clorox and Church & Dwight's stock prices rose following the rising stock price of Wal-Mart.
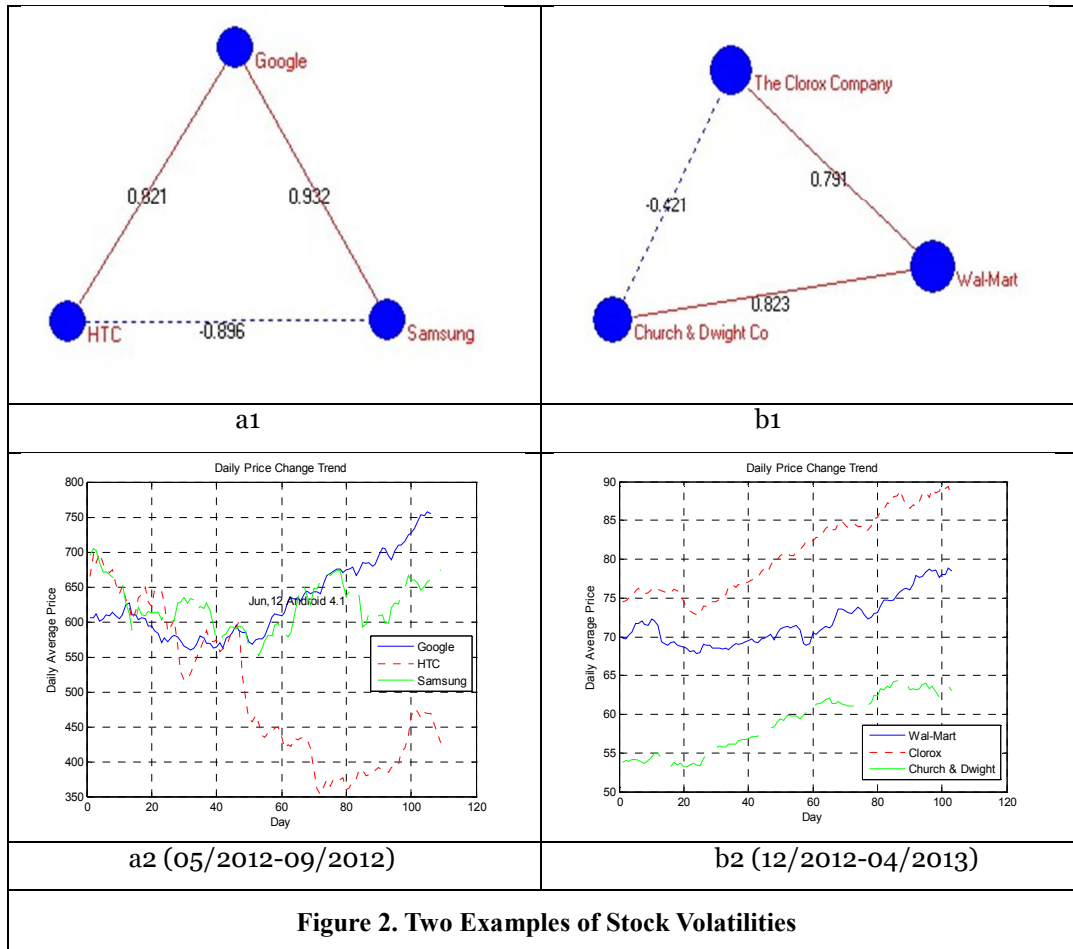

Figure 2. Two Examples of Stock Volatilities

In fact, real-world business markets such as stock markets are very complex. For instance, a firm's stock volatility is affected by both direct and indirect influence of associated firms in an inter-firm network. Given the abnormal change of stock price of one firm triggered by a significant event (e.g., introducing good products to a market), a series of chain reactions occur in a business network. The influence propagation process stops after these reactions have gradually reached equilibrium. It is analogous to the reaction of throwing a stone into a quiet lake. At the beginning, it causes a series of waves. Then, the wavering surface of the lake gradually returns to a peaceful state. Accordingly, we design a novel business performance prediction model that leverages a business network comprising simultaneously interacting agents (firms) to capture complex market dynamics. In fact, a network of interacting agents is considered an effective tool to simulate and study complex economic behavior (Tesfatsion 2002).

One candidate computational model that captures the propagation of influences among firms in an inter-firm network is the spreading activation model (SAM); such a model has been widely used in analyzing semantic networks (Collins and Loftus 1975; Martin et al 1994; Roelofs 1992). A spreading activation process is formalized in that once a source node (e.g., a firm) is activated (e.g., its activation value is greater than a firing threshold f), the influence will be propagated to other nodes through the direct links

captured in a SAM. During the "spreading" process, the activation value decays according to a decay factor $d \in [0,1]$, which is usually established with respect to the weight of a link. For the classical spreading activation model, each node cannot be reactivated once it has been activated. Recently, the independent cascade model (ICM) has been applied to study information diffusion problems (Gruhl et al. 2004; Saito 2008). For ICM, once a node is activated, it has a probability p to trigger its neighboring nodes (p may vary w.r.t. different nodes) until no more reachable inactive nodes are found. Similar to the SAM model, once a node is activated in the ICM model, it cannot be re-activated again. For our application, we are not only interested in inferring whether a node is activated or not but we also want to estimate the degree of influence on a node given the simultaneous influence from multiple direct and indirect nodes. Moreover, we need to consider different types of influences (e.g., positive or negative influence) propagated in a network. Unfortunately, these features are not supported by classical SAM and ICM models. Therefore, we design the novel ECM model to infer the business performance of a targeted firm according to the known business performance of some source firms situated within a business network.

We propose the novel Energy Cascading Model to capture the states of firms and the propagation of business influence (e.g., an ACP) from a source firm to a targeted firm through other intervening firms in a business network. The basic assumption is that a firm's business performance (e.g., stock performance) is influenced by other firms in an inter-firm network (Gnyawali and Madhavan 2001; Ma et al. 2009). The propagation of such an "influence" from a source firm to the targeted firm is simulated by the flow of "energy" from a source node v_i to a sink node v_j in the ECM model. Figure 3 shows a simplified view of an ECM network. Each node of the ECM model has one of three internal states: inactive (e.g., triggered by neutral sentiments of a firm), positively-activated (e.g., triggered by positive sentiments of a firm), and negatively-activated (e.g., triggered by negative sentiments of a firm). A positively-activated node tends to strengthen the positive business influence that passes through it, whereas a negatively-activated node tends to weaken the positive business influence passing through. This characteristic captures our intuition that a performing firm (as reflected by its positive sentiments) tends to play the role as a strong partner or competitor for its associates, whereas an under-performing firm (as reflected by its negative sentiments) tends to be a weak partner or competitor because the firm itself is in trouble. For an inactive node, it simply transmits the external energy to other neighboring nodes. The proposed ECM model differs from classical SAM and ICM models in that every node of the ECM model actively generates and propagates energy to its neighboring nodes, whereas nodes (except the initiating node) of the SAM or the ICM model only passively route signals to their neighboring nodes.
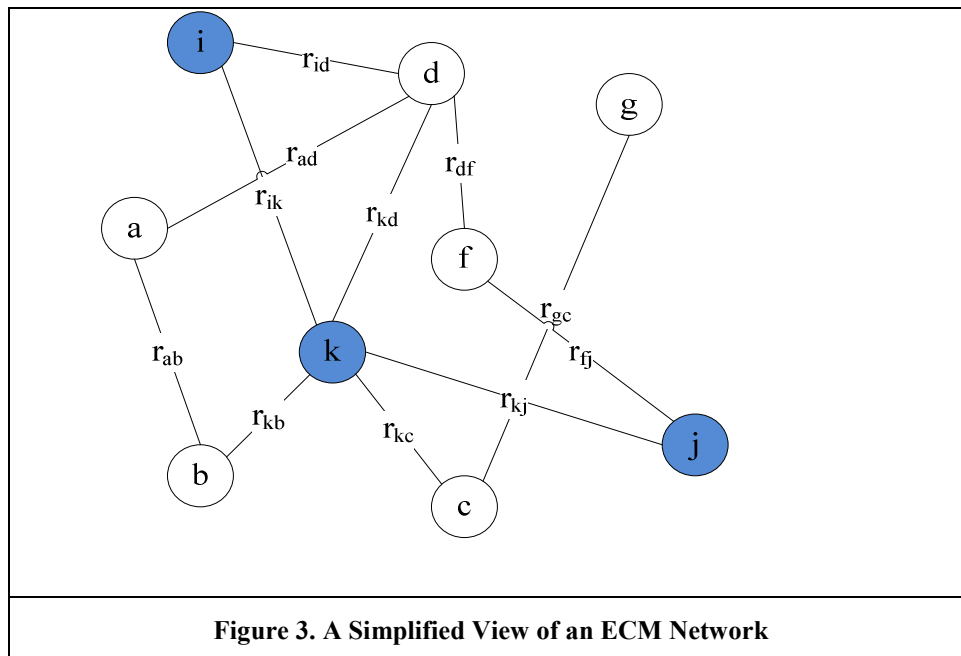

Figure 3. A Simplified View of an ECM Network

There are two types of energies (positive or negative) propagated in an ECM network; positive energy represents an upward ACP or positive sentiments of a firm, whereas negative energy signals a downward ACP or negative sentiments of a firm. Moreover, the type of a link between two nodes (firms) can be either positive (cooperative business relationship) or negative (competitive business relationship). The type of energy as represented by the corresponding sign is "flipped" according to the type of link that it passes through. For a positive link (a cooperative business relationship), the type (sign) of energy that it cascades remains the same. However, the type (sign) of energy is reversed when it passes through a negative link (a competitive business relationship). This feature of ECM aims to capture our intuition that the gain of a firm's competitor tends to be the loss to the firm. When energy is cascaded within the ECM model, it is assumed that energy is propagated from a source node to a sink node through multiple propagation paths. In addition, energy cannot pass through a node twice (i.e., a loop) in each effective propagation path.

The Energy Cascading Model

Definition 7 (Business Network) *A business network is a weighted undirected graph $G = (V, E)$ that comprises a finite set of nodes V and edges E. Each node $v_i \in V$ represents a firm com_i. An edge $e_{ij} \in E$ is an unordered pair of nodes v_i and v_j. In particular, e_{ij} indicates that company com_i is associated with another company com_j through the business relationship with type $type(e_{ij}) \in R$, where $R = \{collaboration, competition\}$. The weight of an edge denoted $r_{ij} \in [-1,0) \cup (0,1]$ represents the strength of a specific type of business relationship and it satisfies the conditions: $type(e_{ij}) = collaboration => r_{ij} \in (0,1]$ and $type(e_{ij}) = competition => r_{ij} \in [-1,0)$.*

Definition 8 (Effective Propagation Path) *Let P_{ij} be the set of all possible effective propagation paths from a source node v_i to a sink node v_j, then an effective propagation path $p_{ij} \in P_{ij}$ is a directed acyclic path that comprises a sequence of pair of nodes such as $(v_i, v_{i+1}), (v_{i+1}, v_k), \cdots, (v_{k+1}, v_j)$ satisfying the conditions: (1) the source node v_i and sink node v_j only appearing once in the sequence; (2) $\forall v_k \in P_{ij}, v_k \neq v_i, v_k \neq v_j$, node v_k only appearing twice in the sequence with the first appearance as the ending node in a pair and the second appearance as the starting node of the immediately following pair. The length of an effective propagation path denoted $len(p_{ij})$ is defined by $len(p_{ij}) = |\{p_{ij}\}|$.*

In ECM, the external energy $en_{ext} \in [-1,1]$ triggered by a downward or upward ACP is propagated from the source node v_i to the sink node v_j through the set of all possible effective propagation paths P_{ij}. The net energy received by the sink node v_j through one of the effective propagation paths $p_{ij} \in P_{ij}$ is derived according to the following formula.

$$en_{p_{ij}}(v_i, v_j, en_{ext}) = \left(\left(\left((en_{ext} \cdot tp(v_i) + en(v_i)) \cdot r_{i,k} \right) \cdot tp(v_k) + en(v_k) \right) \cdot r_{k,k+1} \cdots \cdots \right) \cdot tp(v_j) + en(v_j) \quad (6)$$

where $en(v_k)$ is the energy generated from node v_k according to the sentiments of the corresponding firm. According to previous studies in economics (Beck 2005; Liu 1995), smaller firms which have limited access to resources are much easier to be influenced by external forces. In contrast, larger firms which can access to more diverse sources of resources tend to be less influenced by external forces. Accordingly, we use the adjustment function $tp(v_k) = \frac{avgsize(sector_i)}{size(com_k)}$ to strengthen or weaken the external energy when it is absorbed by different types (large or small) of firms. The size of a firm denoted $size(com_k)$ is determined according to its market capitalization, and the average size of a business sector $avgsize(sector_i)$ is derived from the average market capitalization of a specific business sector. More specifically, if the size of the firm com_k is relatively large, $tp(v_k) < 1$ is returned; otherwise, $tp(v_k) \geq 1$ is established.

The total energy $En(v_i, v_j, en_{ext}) \in [-1,1]$ that is cascaded from the source node v_i to the sink node v_j through the set of all the effective propagation paths P_{ij} is defined by:

$$En(v_i, v_j, en_{ext}) = \frac{\sum_{p_{ij} \in P_{ij}} en_{p_{ij}}(v_i, v_j, en_{ext})}{\sum_{p_{ij} \in P_{ij}} |en_{p_{ij}}(v_i, v_j, en_{ext})|} \quad (7)$$

In addition, given a set V_{set} of initiating nodes (firms), the total net energy $TNEn(V_{set}, v_j, EN_{ext}) \in [-1,1]$ received by a targeted node v_j is then defined by:

$$TNEn(V_{set}, v_j, EN_{ext}) = \frac{\sum_{v_i \in V_{set}, en_{ext} \in EN_{ext}} \sum_{p_{ij} \in P_{ij}} en_{p_{ij}}(v_i, v_j, en_{ext})}{\sum_{v_i \in V_{set}, en_{ext} \in EN_{ext}} |\sum_{p_{ij} \in P_{ij}} en_{p_{ij}}(v_i, v_j, en_{ext})|} \qquad (8)$$

The internal energy en(v_k) of a node v_k is estimated according to the sentiment score derived from the corresponding firm com_k. In particular, we analyze the sentiment of each firm extracted from Twitter via the publicly available API provided by Topsy[2]. OpinionFinder[3], a publicly available sentiment lexicon, is applied to identify the opinion phrases *op* in each stock tweet and predict their sentiment polarities. Let OP_k^+ and OP_k^- be the set of positive and the set of negative opinion phrases for the firm com_k, respectively. The term strength(op) $\in \{1,2\}$ represents the strength of an opinion phrase *op*. In particular, an opinion phrase is assigned the score of 1 if it contains a "weak" sentiment indicator defined according to OpinionFinder. If an opinion phrase *op* contains a "strong" sentiment indicator, its strength is strength(op) = 2.The sentiment score (i.e. the internal energy) of the firm com_k is derived according to the following formula.

$$en(v_k) \approx sent(com_k) = \frac{\sum_{op \in OP_k^+} strength(op) - \sum_{op \in OP_k^-} strength(op)}{\sum_{op \in OP_k^+} strength(op) + \sum_{op \in OP_k^-} strength(op)} \qquad (9)$$

In ECM, since the external energy en_{ext} propagated through different paths may be countervailed, the total net energy $TNEn(V_{set}, v_j, EN_{ext})$ received by the sink node v_j is applied to predict the ACP of the corresponding firm com$_k$. If $TNEn(V_{set}, v_j, EN_{ext}) > \xi^+$ is established, node v_j is likely to be positively activated which results in an upward ACP. On the other hand, if $En(v_i, v_j, en_{ext}) < \xi^-$, node v_j is negatively activated and results in an downward ACP. Otherwise, the energy cascaded from different paths is countervailed, and so an ACP is unlikely to be observed for node v_j. ξ^+ and ξ^- are empirically established ACP prediction thresholds. The computational complexity of predicting the ACP of an arbitrary node v_j given an observed ACP (external energy) of a node v_i in ECM is characterized by $O(|V|^2)$ in the worst case, where V is the set of nodes of the ECM network G.

Experiments and Analysis

To conduct empirical experiments for the proposed ECM Model, we retrieved historical stock data, financial news articles, and Twitter postings for firms that were classified under four different business sectors, Information Technology (IT), Energy (EN), Financials (FN), and Consumer Staples (CS) according to the S&P 500 Index. More specifically, five-year historical stock data corresponding to the period from Jan 2008 to Dec 2012 were retrieved through the Yahoo Finance API (a total of 292,880 stock records). The basic descriptive statistics of our data set is depicted in Table 1. The stock tweets about the chosen S&P 500 firms of the corresponding period were also collected using the API provided by Topsy. Financial news articles of the corresponding period were downloaded from Reuters Finance[4]; these news articles were used as the basis to semi-automatically construct the business networks for the four business sectors. For each evaluation year, a separate set of business networks was constructed using our latent text mining based business network discovery method (Zhang et al. 2012). These sequential sets of business networks were applied to capture evolving business relationships. In addition, JGraphT[5], a free Java graph library that provides mathematical graph-theory objects and algorithms, was applied to analyze and compute the initial statistics pertaining to each business network. Another publicly available programming package called Pajek Java library[6] (Batagelj and Mrvar 1998) was applied to visualize the mined business networks. The proposed ECM model can also operate based on manually constructed business networks if the format of these networks is the same as the one defined in this paper.

---

[2] http://topsy.com/

[3] http://mpqa.cs.pitt.edu/opinionfinder/

[4] http://www.reuters.com/finance

[5] http://jgrapht.org/

[6] http://download.cnet.com/Pajek/3000-2076_4-10662544.html

Table 1. Basic Descriptive Statistics											
Sector	#Firms	Daily Average Stock Price Records					Weekly Average Stock Price Records				
		Record.	Max	Min	Mean	Std.	Record.	Max	Min	Mean	Std.
IT	70	85,559	769.0	1.8	47.5	81.8	17,805	762.6	1.9	47.6	81.9
EN	45	51,661	146.3	6.2	50.6	25.4	10,756	144.5	7.6	50.6	25.4
FN	83	104,497	4786	0.35	60.6	254.5	21,746	4738.5	0.4	60.6	254.2
CS	43	51,163	139.3	0.2	43.7	22.1	10,648	138.8	0.2	43.7	22.1

To highlight the potential of predicting stock price movements based on business relationships among firms, the Pearson correlation coefficients of the daily stock prices of some example firms are shown in Table 2. It is easy to find that the stock volatility of primary competitors such as Apple Inc. and Nokia, Apple Inc. and RIM, and so on are negatively correlated. In contrast, the stock volatility of business partners such as Google and Motorola is positively correlated. However, although some firms are competitors to each other (e.g. Apple Inc. and Google, Apple Inc. and Microsoft), their stock volatility seems positively correlated according to the Pearson correlation analysis. As explained in the previous section, simply studying the isolated relationship between each pair of firms may not be able to fully evaluate all the influences that affect a firm's stock price. Nevertheless, the correlation analysis reported in Table 2 does reveal that business relationships among firms may influence their stock volatility. What we need is a more robust and sophisticated model that can leverage both business relationships and sentiments of firms to more accurately predict stock volatility.

Table 2. Correlations of Daily Average Prices of Some Firms (* $p < .05$)									
	Mean	Std.	1	2	3	4	5	6	7
1 Amazon	142.39	62.43	1						
2 Apple	297.09	166.85	0.92*	1					
3 Google	530.19	101.73	0.79*	0.78*	1				
4 Microsoft	26.51	3.62	-0.18*	0.54*	0.78*	1			
5 Motorola	21.11	20.80	0.47*	0.88*	**0.65***	0.37*	1		
6 Nokia	12.22	8.35	-0.75*	**-0.71***	-0.35*	0.02	**-0.66***	1	
7 RIM	54.20	34.05	-0.73*	**-0.74***	-0.33*	-0.03	-0.72*	0.89*	1

To examine the impact of business relationships on stock price movements, the abnormal change points (ACPs) of each evaluating firm were first identified based on the measure Change of the Change Rate of Weekly Average Price (CCR-WAP). According to King's (1966) study, market and industrial factors may have an impact on stock volatility as well. Therefore, we first identified the ACPs (stock volatility) that were mainly caused by market factors and excluded them from our experiments. Based on the Dow Jones Industrial Average (DJIA) and the S&P 500 Index, we selected top 5 leading firms from each targeted business sector as the basis to estimate the market-led stock volatility of the corresponding sector. These leading firms were usually the hubs (i.e., connecting to many firms) in a sector-based business network (Zhang et al. 2012). If all the chosen leading firms showed the same directional stock price movement in a particular week, the ACPs of that week were considered mainly influenced by market-led factors, and so these ACPs would be excluded from our experiments.

To evaluate the experimental system ECM, we first used the historical stock data in the first six months of each evaluation year to identify the top ϑ most influential firms for each testing firm of a business sector. For the experiments reported in this paper, we set the selection parameter $\vartheta = 5$. For instance, a testing firm was first taken as a source node in a business network. Then, the top ϑ firms (i.e., the sink nodes) that received the largest positive or negative energy in this period were taken as the most influential firms. For the remaining six months of each evaluation year, we applied the ACPs of these top ϑ most influential firms as the source nodes to evaluate the energy received by the testing firm, and hence to predict its ACPs. Finally, the ACPs of each testing firm were compared with the ground truth. The widely used performance

measures such as Precision, Recall, F-measure, and Accuracy were applied to assess the performance of the experimental and the baseline systems. The prediction performance of each system across four different business sectors is reported in Table 3. For our experiments, the absolute values of the positive and the negative thresholds (i.e., ξ^+ and ξ^-) are the same.

For the CORR baseline system, Pearson correlation coefficient was applied to identify the top 5 closely related firms for each testing firm based on the weekly average stock prices in each training period (i.e., the first six months of each evaluation year). The ACPs of a testing firm and the CCR-WAPs of its closely related firms were then taken as training examples to train a SVM classifier. Afterwards, the SVM classifier took the CCR-WAPs of these closely related firms as inputs to predict the testing firm's ACPs in each test period (i.e., the second half of each evaluation year). LIBSVM[7] (with RBF kernel), one of the most popular classification tools, was applied to implement the CORR baseline system. Investors' sentiments for firms have shown to be useful to predict stock price movements in the finance literature (Das and Chen 2007; Tetlock et al. 2008). Accordingly, our second baseline system called SENT utilized the Twitter sentiments of a firm to predict its ACPs. More specifically, a firm's sentiment score pertaining to each period is computed according to Eq.9. For instance, if the weekly sentiment score of a firm is greater (lower) than an empirically established positive (negative) threshold, an upward (downward) ACP is predicted for the firm.

Table 3 shows that the ECM system achieves an average accuracy of 62.8%. It represents an improvement of 34.6% when compared to the result of the best baseline system CORR. When compared to other state-of-the-art prediction models such as the AZFinText System which achieves a directional accuracy of around 57% (Schumaker and Chen 2009, 2010), our proposed ECM model demonstrates a considerable improvement for predicting directional stock price movement given the merits of a business network-based approach. Surprisingly, the SENT baseline system performs poorly despite the promising results reported in the previous literature (Das and Chen 2007; Tetlock et al. 2008). One possible reason is that our sentiment scores were computed based on stock tweets which were relatively short and noisy when compared to formal financial reports. As a whole, our empirical experiments confirm the effectiveness of the proposed ECM model.

Table 3. Comparative Performance of ECM and Baseline Methods					
		IT	EN	FN	CS
ECM	threshold	0.030	0.150	0.150	0.030
	Precision	0.678	0.546	0.691	0.598
	Recall	0.718	0.831	0.776	0.638
	F-measure	0.698	0.659	0.731	0.617
	Accuracy	0.672	0.553	0.687	0.603
CORR	threshold	0.050	0.150	0.150	0.050
	Precision	0.556	0.440	0.394	0.455
	Recall	0.715	0.837	0.774	0.647
	F-measure	0.625	0.577	0.522	0.534
	Accuracy	0.560	0.448	0.399	0.461
SENT	threshold	0.050	0.050	0.050	0.050
	Precision	0.519	0.471	0.372	0.436
	Recall	0.754	0.733	0.785	0.644
	F-measure	0.615	0.541	0.515	0.527
	Accuracy	0.524	0.419	0.387	0.406

---

[7] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

According to previous studies in economics (Beck 2005; Liu 1995), large firms (e.g. sector leaders) have more influential power than small firms do. In addition, the business performance of small firms is easier to be influenced by outside factors (e.g. market or industrial factor). Accordingly, we performed an empirical study to compare the prediction accuracy for different kinds of firms. More specifically, we selected 10 large firms and 10 small firms based on their market capitalization from each of the four S&P 500 business sectors. Similar to the first experiment, we first identified the top ϑ most influential firms for each testing firm in each training period. Then, given the known ACPs of these most influential firms, the ACPs of each testing firm in each group (large or small) were predicted for each test period. The average prediction performance pertaining to these two groups is reported in Table 4. It seems that the ACPs of large firms are relatively easier to be predicted. In particular, the average prediction accuracy for large firms is 14.8% higher than that for small firms. The possible explanation is that the business performance of small firms is sensitive to many outside factors. As a result, it is more difficult to predict the business performance of small firms solely based on business relationships and firms' sentiments. Our empirical result suggests that the impact of business relationships on stock price movements may not be the same for different kinds of firms.

Table 4. Comparative Prediction Performance for Large and Small Firms					
		IT	EN	FN	CS
Large Firms	threshold	0.030	0.150	0.150	0.030
	Precision	0.717	0.547	0.696	0.677
	Recall	0.718	0.833	0.782	0.642
	F-measure	0.718	0.661	0.736	0.659
	Accuracy	0.720	0.554	0.699	0.663
Small Firms	threshold	0.050	0.150	0.150	0.050
	Precision	0.595	0.537	0.672	0.498
	Recall	0.675	0.730	0.744	0.621
	F-measure	0.633	0.619	0.706	0.552
	Accuracy	0.589	0.541	0.667	0.498

Our third empirical study aimed to examine the stock volatility of different kinds of firms. Figure 4 depicts an automatically constructed business network for the IT sector by invoking our business network mining method (Zhang et al. 2012). We can observe that some firms (e.g. Google and Microsoft) have more collaborators (solid lines) than competitors (dash lines), whereas other firms (e.g. Apple Inc. and Samsung) have more competitors than collaborators. To distinguish between competitive firms (firms with many business competitors) and collaborative firms (firms with many business partners), a relationship type measure $RT(com_i) \in [-1,1]$ is developed. The relationship type of a firm com_i is derived according to the following formula:

$$RT(com_i) = \frac{|Coll(com_i)| - |Comp(com_i)|}{|Coll(com_i)| + |Comp(com_i)|}$$

(10)

where functions $Coll(com_i)$ and $Comp(com_i)$ return the set of collaborators and competitors of firm com_i. Thresholds for competitive and collaborative firms were empirically established according to the typical scores of competitive and collaborative firms of a business sector. To examine the stock volatility of different kinds of firms, weekly average price (WAP) was chosen as the instrument to exclude the short term fluctuations of stock prices. In particular, change rate of weekly average price (CR-WAP) and the change of change rate of weekly average price (CCR-WAP) were applied to identify the ACPs for each kind of firms. The number of ACPs and its standard deviation for each kind of firm were then analyzed. In addition, we applied standard deviations of CR-WAP and CCR-WAP to estimate the range of dispersion of these average prices pertaining to individual firms of each group.
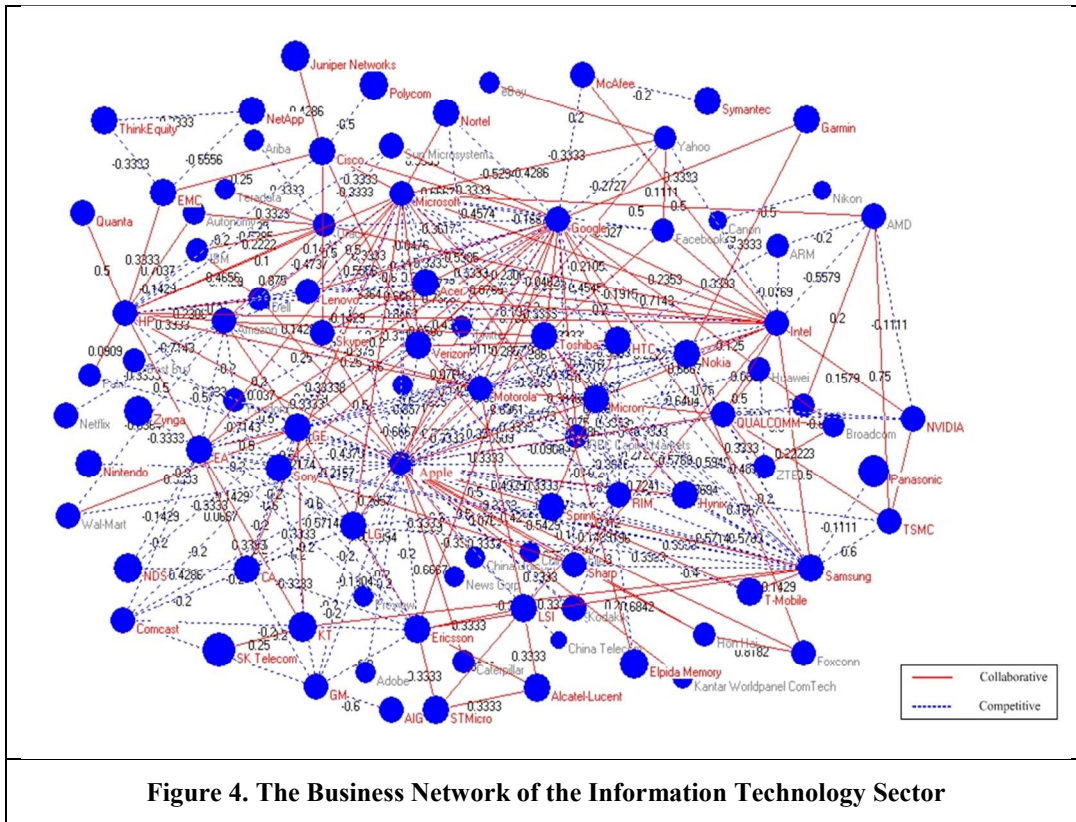
Figure 4. The Business Network of the Information Technology Sector

Table 5. The Stock Volatility of Competitive and Collaborative Firms							
			IT	EN	FN	CS	All
# Firms			68	41	81	41	191
Competitive	threshold		0.05	0.30	0.30	0.10	--
	# Firms		9	18	10	11	48
	Avg No. of ACPs	CR	112.83	105.50	97.86	89.00	101.50
		CCR	109.50	103.72	100.57	88.67	100.70
	Std.	CR	4.81	3.71	18.56	2.97	6.48
		CCR	4.62	6.55	19.20	3.42	8.10
Collaborative	threshold		0.05	0.10	0.30	0.10	--
	# Firms		16	8	30	9	63
	Avg No. of ACPs	CR	90.43	97.15	98.75	85.40	94.53
		CCR	92.69	105.07	98.81	88.60	96.23
	Std.	CR	3.79	4.32	5.40	2.85	4.49
		CCR	4.48	4.63	6.01	3.39	5.13

The result of our empirical study is shown in Table 5. We can observe that both the number of ACPs and the standard deviations of CR-WAPs and CCR-WAPs of collaborative firms are smaller than that of competitive firms. Our empirical finding reveals that stock volatility of collaborative firms is smaller than that of competitive firms in general. Such a result may be explained by collaborative firms' better access to essential resources and expertise to carry out diversified business strategies and operations

(Brandenburger and Nalebuff 1996). In general, a collaborative firm can obtain the necessary support from its business partners if the firm is in trouble. Thus, it is much easier for the firm to solve its problems. Furthermore, if one of its collaborative firms is in trouble, the negative influence from that partner may be alleviated through the positive support from other business partners. As a result, the stock price of a collaborative firm fluctuates less. In contrast, when a competitive firm is in trouble, it is difficult for that firm to obtain assistance given a limited number of collaborators. Instead, its competitors may even launch hostile activities given such an opportunity. These hostile activities tend to make the stock price of the competitive firm fluctuate a lot.

Conclusions and Future Work

While a large body of research has been devoted to the analysis of social networks, very few studies about mining and analyzing business networks are reported in the literature. The main theoretical contributions of our research include the design of several metrics to effectively measure stock price movements, and the design of the novel Energy Cascading Model for the prediction of firms' business performance through the proxy of stock prices. In particular, a firm's directional stock price movement is predicted based on its relationships with other firms and the sentiments of these firms. Our experimental results show that the proposed ECM model can effectively predict directional stock price movements; it achieves an average accuracy of 62.8% which is considerably higher than that achieved by other state-of-the-art prediction methods. Our empirical studies also show that it is relatively easier to predict stock price movements of large firms and the stock volatility of collaborative firms is smaller than that of competitive firms. The business implication of our research is that business managers and financial analysts can apply our design artifacts to more effectively analyze and predict the business performance of targeted firms based on automatically mined or manually constructed business networks. Accordingly, they can take proactive business strategies to streamline the operations of these firms.

One limitation of our current work is that business relationships are modeled at the coarse-grained level. However, in the real world, a firm can have both collaborative and competitive relationships with another firm at the same time. For instance, Samsung is the largest rival of Apple Inc. for smartphones and tablet PCs. At the same time, Samsung is the largest supplier of Apple Inc. for display and battery devices. Our future work will extend the ECM network by capturing entity-based (e.g., products, events, or locations) relationships among firms. The prediction of ACPs can then be conducted based on a chosen set of entities captured in an inter-firm network. Since business relationships among firms tend to evolve over time, future work will examine the optimal lag time for using a current or historical business network to predict future business performance. Moreover, the ACP prediction thresholds will be fine-tuned using a dynamic threshold learning method. We will apply the structural properties (e.g., centrality, in-out degree, etc.) widely used in social network analysis to enhance the current ECM model. Larger scale of experiments with more baseline models and business sectors will be conducted in the future. Finally, we will apply the ECM model to analyze firms' business performance based on other indicators such as turnovers, net profits, and number of patents generated.

References

Aharony, J. and Swary I. 1980. "Quarterly Dividend and Earnings Announcements and Stockholder's Returns: An Empirical Analysis," Journal of Finance (35), pp. 1-12.

Al-Qaheri, H., Hassanien, A. E. and Abraham, A. 2008. "Discovering Stock Price Prediction Rules using Rough Sets," Neural Network World (18:3), pp. 181-189.

Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M. and Goldstein, G. 2011. "Identifying and Following Expert Investors in Stock Microblogs," Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, July 27-31, pp. 1310-1319.

Batagelj, V. and Mrvar, A. 1998. "Pajek - Program for Large Network Analysis," Connections (21:2), pp. 47-57.

Beck, T., DEMIRGÜÇ - KUNT, A. S. L. I., and Maksimovic, V. 2005. "Financial and legal Constraints to Growth: Does Firm Size Matter?" The Journal of Finance (60:1), pp.137-177.

Bermingham, A., and Smeaton, A. F. 2010. "Classifying Sentiment in Microblogs: Is Brevity an Advantage?" In Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 1833-1836.

Bernstein, A., Clearwater, S., and Provost, F. 2003. "The Relational Vector-space Model and Industry Classification," In Proceedings of Workshop on Learning Statistical Models from Relational Data.

Bollen, J., Mao, H. and Zeng, X. "Twitter Mood Predicts the Stock Market," Journal of Computational Science (2), pp. 1-8.

Brandenburger, A. M., and Nalebuff, B. J. 1997. "Co-Opetition," 1st Edition, New York: Doubleday.

Brav, A. and Heaton, J.B. 2002. "Competing Theories of Financial Anomalies," Review of Financial Studies (15), pp. 575–606

Cameron, M. A., Power, R., Robinson, B., and Yin, J. 2012. "Emergency Situation Awareness from Twitter for Crisis Management," In Proceedings of the 21st international conference companion on World Wide Web, pp. 695-698.

Case, K.E., Quigley, J.M. and Shiller, R.J. 2005. "Comparing Wealth Effects: The Stock Market versus the Housing Market," Advances in Macroeconomics (5:1), Article 1, pp. 1-32.

Celikyilmaz, A., Hakkani-Tur, D., and Feng, J. 2010. "Probabilistic Model-based Sentiment Analysis of Twitter Messages," In Spoken Language Technology Workshop (SLT), IEEE, pp. 79-84.

Chakrabarti, D., and Punera, K. 2011. "Event Summarization using Tweets," In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp. 66-73.

Chang, P.C. and Liu, C.H. 2008. "A TSK Type Fuzzy Rule based System for Stock Price Prediction," Expert System with Application (34), pp. 135-144.

Charkha, P.R. 2008. "Stock Price Prediction and Trend Prediction using Neural Networks," First International Conference on Emerging Trends in Engineering and Technology, IEEE, pp. 592-594.

Chaudhuri, K., and Wu, Y. 2003. "Random Walk versus Breaking Trend in Stock Prices: Evidence from Emerging Markets," Journal of Banking & Finance (27:4), pp. 575-592.

Chen, H., Chiang, R.H.L., Storey, V.C. 2012. "Business Intelligence and Analytics from Big Data to Big Impact," MIS Quarterly, (36:4), pp. 1165-1188.

Chen, Q., Goldstein, I. and Jiang, W. 2007. "Price Informativeness and Investment Sensitivity to Stock Price," The Review of Financial Studies (20:3), pp. 620-650.

Collins, A. M., and Loftus, E. F. 1975. "A Spreading-activation Theory of Semantic Processing," Psychological review (82:6), pp. 407-428.

Cootner, P.H. 1964. "The Random Character of Stock Market Prices," MIT.

Dann, L. 1981. "Agency Stock Repurchases: An Analysis of Returns to Bondholders and Stockholders," Journal of Financial Economics (9), pp. 113-138.

Das, S. and Chen, M. 2007. "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," Management Science (53:9), pp. 1375–1388.

Dolan, R.J. 2002. "Emotion, Cognition, and Behavior," Science (298), pp. 1191-1194.

Elton, E. J., Gruber, M. J., & Gultekin, M. 1981. "Expectations and Share Prices," Management Science (27:9), pp. 975-987.

Fama, E.F. 1965. "The Behavior of Stock-market Prices," The Journal of Business (38:1), pp. 34-105.

Fama, E. F. 1991. "Efficient Capital Markets: II," The journal of finance (46:5), pp. 1575-1617.

Feldman, S. J., Soyka, P. A., and Ameer, P. G. 1997. "Does Improving a Firm's Environmental Management System and Environmental Performance Result in a Higher Stock Price?" The Journal of Investing (6:4), pp. 87-97.

Gnyawali, D. R., and Madhavan, R. 2001. "Cooperative Networks and Competitive Dynamics: A Structural Embeddedness Perspective," Academy of Management Review (26:3), pp. 431-446.

Go, A., Bhayani, R., and Huang, L. 2009. "Twitter Sentiment Classification using Distant Supervision," CS224N Project Report, Stanford, pp.1-12.

Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A. 2004. "Information Diffusion through Blogspace," In WWW '04: Proceedings of the 13th international conference on World Wide Web, ACM, New York, NY, USA, pp. 491-501.

Hevner, A., March, S., Park, J., Ram, S. 2004. "Design Science in Information Systems Research," MIS Quarterly (28:1), pp. 75-105.

Ince, H. and Trafalis, T.B. 2004. "Kernel Principal Component Analysis and Support Vector Machines for Stock Price Prediction," IEEE International Joint Conference on Neural Networks, pp. 2053-2058.

Ince, H. and Trafalis, T.B. 2008. "Short Term Forecasting with Support Vector Machines and Application to Stock Price Prediction," International Journal of General Systems (37:6), pp. 677-687.

Johnson, W.B. and Magee, R.P. 1985. "An Analysis of the Stock Price Reaction to Sudden Executive Deaths—Implications for the Managerial Labor Market," Journal of Accounting and Economics (7), pp. 151-174.

Jones, C. P. and Litzenberger, R.H. 1970. "Quarterly Earnings Reports and Intermediate Stock Price Trends," The Journal of Finance (25:1), pp. 143-148.

Kahneman, D. and Tversky, A. 1976. "Prospect Theory: An Analysis of Decision under Risk," Econometrica (47:2), pp. 263-291.

Knack, S., & Keefer, P. 1995. "Institutions and Economic Performance: Cross-country Tests using Alternative Institutional Measures," Economics & Politics (7:3), pp. 207-227.

Kimoto, T. and Asakawa, K. 1990. "Stock Market Prediction System with Modular Neural Network," IEEE International Joint Conference on Neural Network (1), pp. 1-6.

King, B.F. 1966. "Market and Industry Factors in Stock Price Behavior," The Journal of Business (39:1), pp. 139-190.

Kohara, K., Ishikawa, T. Fukuhara, Y. and Nakamura, Y. 1997. "Stock Price Prediction using Prior Knowledge and Neural Networks," Intelligent Systems In Accounting, Finance and Management (6), pp. 11-22.

Kouloumpis, E., Wilson, T., and Moore, J. 2011. "Twitter Sentiment Analysis: The Good the Bad and the Omg," In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, pp. 538-541.

Levine, R. and Zervos, S. 1996. "Stock Market Development and Long-run Growth," World Bank Economic Review (10:2), pp. 323-39.

Lau, R.Y.K., Liao, S., Wong, K.F., and Chiu, D. 2012. "Web 2.0 Environmental Scanning and Adaptive Decision Support for Business Merger and Acquisition," MIS Quarterly, (36:4), pp.1239-1268.

Liu, B. 2010. "Sentiment analysis: a multifaceted problem," IEEE Intelligent Systems 25(3), pp. 76-80.

Liu, H. 1995. "Market orientation and firm size: an empirical examination in UK firms," European Journal of Marketing (29:1), pp. 57-71.

Lo, A. W. and MacKinlay, A. C. 2011. "A Non-random Walk Down Wall Street," Princeton University Press.

Loomis, C. 1985. "Beating the Market by Buying Back Stock," Fortune, pp. 42-48.

Lu, H., Han, J. and Feng, L. 1998. "Stock Movement and n-Dimensional Inter-Transaction Association Rules," Proc. 1998 SIGMOD Workshop Research Issues on Data Mining and Knowledge Discovery (DMKD '98), Article 12, pp. 1-7.

Ludvigson, S. and Charles, S. 1999. "How Important is the Stock Market Effect on Consumption?" Economic Policy Review, (5:2), pp. 29-52.

Ma, Z., Sheng, O., and Pant, G. 2009. "Discovering Company Revenue Relations from News: A Network Approach," Decision Support Systems (4:47), pp. 408-414.

March, S. T. and Storey, V.C. 2008. "Design Science in the Information Systems Discipline," MIS Quarterly (32:4), pp. 725-730.

Martin, N., Dell, G. S., Saffran, E. M., and Schwartz, M. F. 1994. "Origins of Paraphasias in Deep Dysphasia: Testing the Consequences of a Decay Impairment to an Interactive Spreading Activation Model of Lexical Retrieval," Brain and language (47:4), pp. 609-660.

Nair, H. S., Manchanda, P., and Bhatia, T. 2010. "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders," Journal of Marketing Research, (47:5), pp. 883-895.

Naveed, N., Gottron, T., Kunegis, J., and Alhadi, A. C. 2011. "Bad News Travel Fast: A Content-based Analysis of Interestingness on Twitter." Available at:
http://www.websci2011.org/fileadmin/websci/Papers/2050_paper.pdf

Ofer, A. R. and Thakor, A.V. 1987. "A Theory of Stock Price Responses to Alternative Corporate Cash Disbursement Methods: Stock Repurchases and Dividends," The Journal of Finance, (42:2), pp. 365-394.

Pai, P.F. and Lin, C.S. 2005. "A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting," The International Journal of Management Science, Omega 33, pp. 497-505.

Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," Foundations and trends in information retrieval 2, no. 1-2, pp. 1-135.

Peffers, K., Tuunanen, T, Rothenberger, M., Chatterjee, S. 2008. "A Design Science Research Methodology for Information Systems Research," Journal of Management Information Systems (24:3), pp. 45-77.

Petram, L.O. 2011. "The World's First Stock Exchange: How the Amsterdam Market for Dutch East India Company Shares Became a Modern Securities Market, 1602-1700," Dissertation, University of Amsterdam.

Porter, M. E. 1980. "Competitive strategy: Techniques for Analyzing Industries and Competitors," New York: Free Press.

Poterba, J.M. 2000. "Stock Market Wealth and Consumption," Journal of Economic Perspectives (14:2), pp. 99-118.

Roelofs, A. 1992. "A Spreading-activation Theory of Lemma Retrieval in Speaking," Cognition (42:1), pp. 107-142.

Saad, E.W., Prokhorov, D.V. and Wunsch, II, D.C. 1998. "Comparative Study of Stock Trend Prediction Using Time Delay, Recurrent and Probabilistic Neural Networks," IEEE Transactions on Neural Networks (9:6), pp. 1456-1470.

Saif, H., He, Y., and Alani, H. 2011. "Semantic Smoothing for Twitter Sentiment Analysis," In Proceeding of the 10th International Semantic Web Conference (ISWC).

Saito, K., Nakano, R., and Kimura, M. 2008. "Prediction of Information Diffusion Probabilities for Independent Cascade Model," In Knowledge-Based Intelligent Information and Engineering Systems, Springer Berlin Heidelberg, pp. 67-75.

Sakaki, T., Okazaki, M., and Matsuo, Y. 2010. "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors," In Proceedings of the 19th international conference on World wide web, pp. 851-860.

Schiffman, L. G., and Gaccione, V. 1974. "Opinion Leaders in Institutional Markets," Journal of Marketing (38:2), pp. 49-53.

Schumaker, R.P. and Chen, H. 2009. "Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System," ACM Transactions on Information Systems (27:2), Article 12.

Schumaker, R.P. and Chen, H. 2010. "A discrete Stock Price Prediction Engine Based On Financial News," IEEE Computer (43:1), pp. 51-56.

Schwarz, N. 2000. "Emotion, Cognition, and Decision making," Cognition and Emotion (14:4), pp. 433-440.

Smirlock, M. and Starks, L. 1988. "An Empirical Analysis of the Stock Price-Volume Relationship," Journal of Banking and Finance (12), pp. 31-41.

Stewart, S. 1976. "Should A Corporation Repurchase Its Own Stock," Journal of Finance, pp. 911-921.

Summers, L.H. 1986. "Does the Stock Market Rationally Reflect Fundamental Values?" The Journal of Finance (41:3), pp. 591-601.

Tan, H., W. Luo, and Ni, L.M. 2012. "CloST: A Hadoop-based Storage System for Big Spatio-Temporal Data Analytics," Proceedings of the 2012 ACM International Conference on Information and Knowledge Management, pp. 2139-2143.

Tan, C. N. W. and Wittig, G.E. 1993. "A Study of the Parameters of a Back Propagation Stock Price Prediction Model," In Proceedings of first New Zealand international two-stream conference on artificial neural networks and expert systems, pp. 288-291.

Tesfatsion, L. 2002. "Agent-based Computational Economics: Growing Economies from the Bottom Up," Artificial Life (8:1), pp. 55–82.

Tetlock, P., Saar-Tsechansky, P.M., and Macskassay, S. 2008. "More than Words: Quantifying Language to Measure Firms' Fundamentals," Journal of Finance (63:3), pp. 1437–1467.

Thomas-Hunt, M.C., Ogden, T.Y. and Neale, M.A. 2003. "Who's Really Sharing? Effects of Social and Expert Status on Knowledge Exchange Within Groups," Management Science (49:4), pp. 464-477.

Torgo, L. 2010. "Data Mining with R-Learning with Case Studies," Chapman & Hall/CRC, Taylor & Francis Group, pp. 108-111.

Trafimow, D., and Sniezek, J. A. 1994. "Perceived Expertise and Its Effect on Confidence," Organizational Behavior and Human Decision Processes (57), pp. 290-302.

Tsur, O., Davidov, D., and Rappoport, A. 2010. "ICWSM–A Great Catchy Name: Semi-supervised Recognition of Sarcastic Sentences in Online Product Reviews," In Proceedings of the fourth international AAAI conference on weblogs and social media, pp. 162-169.

Vega, C. 2006. "Stock Price Reaction to Public and Private Information," Journal of Financial Economics (82), pp. 103-133.

Vermaelen, T. 1981. "Common Stock Repurchases and Market Signaling," Journal of Financial Economics (9), pp. 139-183.

Wang, J.H and Leu, J.Y. 1996. "Stock Market Trend Prediction using ARIMA-based Neural Networks," IEEE International Conference on Neural Networks (4:6), pp. 2160-2165.

Wilder, D. A. 1984. "Intergroup Contact: The Typical Member and the Exception to the Rule," Journal of Experimental Social Psychology (20:2), pp. 177-194.

Zhang, W., Lau, R., Liao, S. S., and Kwok, R. C. W. 2012. "A Probabilistic Generative Model for Latent Business Networks Mining," In Proceedings of the Thirtieth-third International Conference on Information Systems, Completed Research Paper 138, Orlando, Florida, USA, December 16-19, 2012.

Zhang, Z. 2008. "Weighing Stars: Aggregating Online Product Reviews for Intelligent Ecommerce Applications," IEEE Intelligent Systems 23(5), pp. 42-49.