

# The Dual Micro/Macro Informing Role of Social Network Sites: Can Twitter Macro Messages Help Predict Stock Prices?

*Nicholas Evangelopoulos*

*University of North Texas, Denton, TX, USA*

*Michael J. Magro*

*Shenandoah University, Winchester, VA, USA*

*Anna Sidorova*

*University of North Texas, Denton, TX, USA*

[Nick.Evangelopoulos@unt.edu](mailto:Nick.Evangelopoulos@unt.edu); [mmagro@su.edu](mailto:mmagro@su.edu);  
[Anna.Sidorova@unt.edu](mailto:Anna.Sidorova@unt.edu)

## Abstract

This study proposes a framework for understanding the role of Social Network Sites (SNS) in informing clients at individual message (micro) and aggregate (macro) levels. To validate the micro/macro informer framework, we examine if an aggregate of Twitter messages can be used as a predictor of future stock prices of publicly traded companies. Our field study analyzes Twitter posts related to 18 Fortune500 companies using Latent Semantic Analysis to extract the semantic and conceptual content in the form of key themes. Using the factors comprising the themes, we fit a regression model that uses tweet volume and tweet topic strength to predict 8.3% of variability in stock prices beyond what is explained by the fluctuations of the stock market. Our results suggest that Twitter can be viewed as a macro informer for stock markets. This confirms our conjecture that SNS are both platforms for micro informing and macro informers. Our study should stimulate informing science researchers to examine other cases of macro informing and the role of data mining and text mining technologies in message aggregation for macro informing.

**Keywords:** Micro/macro informer framework, Twitter, social networking sites, text mining, latent semantic analysis, stock price modeling

## Introduction

Emergence and proliferation of social media is one of the most significant trends of the 2000s. Starting with the establishment of Friendster in 2003 as the first large-scale social networking site

(Boyd, 2006), a mere 8 years has seen the Internet territory explode with a variety of Social Networking Sites (SNS) which cater to different user categories and are tailored for different purposes. The widespread adoption and popularity of these websites has generated interest from consumers, companies, and researchers alike.

---

Material published as part of this publication, either on-line or in print, is copyrighted by the Informing Science Institute. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact [Publisher@InformingScience.org](mailto:Publisher@InformingScience.org) to request redistribution permission.

Consumers have discovered the value of social network sites and are flocking to them in droves. Social networking sites as a whole almost doubled in unique visitors per month from around 115 million in July of 2007, to just over 200 million in April of 2011 according to one measure (MarketingProfs.com, 2011), and as of June 2012, unique visits to social networking sites have increased by 6% every year (Perez, 2012). Companies are striving to capitalize on the trend of consumer engagement in a variety of ways. Some are using SNS to find new customers (Kulikowski, 2011), to market to existing customers and build their brand (R2Integrated, 2010), and as a new avenue for customer communication for many top U.S. companies (Farfan, 2011). An emerging group of financial traders is closely monitoring SNS for information, crowd pulse, and connection with each other (Harris & Rogers, 2011). The majority of research on SNS is focused on use, adoption, privacy and security issues of SNS (Fogel & Nehmad, 2009; Shen & Khalifa, 2010).

SNS generate large volumes of data which in aggregate contain useful information about views and opinions of SNS users. Several studies have investigated the ways to interpret and utilize the information contained in SNS-generated data (e.g., Naaman, Becker, & Gravano, 2011; Taraszow, Aristodemou, Shitta, Laouris, & Arsoy, 2010). This suggests that SNS play a dual role. On the one hand they provide infrastructure for individual informers to deliver messages to individual clients. On the other hand, due to a largely public nature of the individual message exchange, SNS themselves serve as informers providing aggregate information to clients. The goal of this paper is to introduce and validate a framework for understanding the dual role of SNS as platforms for exchanging micro messages and as macro informers. To validate the micro/macro informing framework, we examine if Twitter can serve as a macro informer to the stock market. By applying text mining techniques to SNS messages for the purpose of stock market prediction, this paper represents transdisciplinary research in the fields of information systems and finance, and thus contributes to the goal of Informing Science transdiscipline to reward and encourage research that crosses disciplinary boundaries (Cohen, 2009).

In the next section we provide a brief overview of key types of SNS and discuss prior work related to mining the information from SNS sites. We then formally introduce the micro/macro informing framework. We continue with our research methodology and present the results of our analysis. We conclude with a discussion of implications for practice and research.

## Overview

The term *social networking site* generally refers to those websites and their derived applications that deliver innovative online communication within and among a person's various social networks. SNS typically share a common set of features which include a profile (representation and/or description) for each user, the means to build and manage a personal relational network (i.e., friends, family, acquaintances, etc.), and access to creative methods to communicate with members of their relational network and the online community (Magro, Ryan, Sharp, & Ryan, 2008).

SNS have grown in popularity at a tremendous rate in the past several years. Facebook, which has emerged as the dominant SNS, reported 900 million active users in April of 2012 (Goldman, 2012). Twitter had an estimated 225 million users as of January 2012, with around 1 billion tweets being sent every 5 day period during 2011 (Pring, 2012). Many people that go online across the globe access SNS on a daily basis. The popularity of SNS has produced specialized networks, such as those which promote professional and occupational social networking (LinkedIn), those centered on food and cooking (i.e., BakeSpace), and even one that brings together people suffering from illness (PatientsLikeMe).

Social networking tools such as Facebook and Twitter have been well researched from both the technological and social perspectives, but as new forms of communication they represent an im-

portant new paradigm in electronic relationship management and information dissemination. Academic research related to SNS focuses on many different facets, including (but not limited to) user behavior (e.g., Bolar, 2009; Ellison, Steinfield, & Lampe, 2007), privacy and trust (e.g., Acquisti & Gross, 2006; Dwyer, Hiltz, & Widmeyer, 2008) and the unique properties of social networks (e.g., Boyd & Ellison, 2008; Richter & Koch, 2008). Recently, studies that are devoted to modeling the textual content generated by SNS users have emerged. Three recent studies that utilize Twitter data for this purpose are discussed below.

Motivated by the need to provide Twitter users with better search and filtering facilities, Ramage, Dumais, and Liebling (2010) used the Latent Dirichlet Allocation (LDA) approach to text modeling for identifying latent topics present in Twitter content. After identifying 200 latent topics in Twitter comments, they classified them into these five categories: substance, status, style, social, and other. The *other* category appeared to include primarily latent topics discussed in languages other than English and was not analyzable. The remaining four categories are referred to as the *4S* categories. *Substance* includes tweets about events, ideas, things, or people. *Status* encompasses personal updates about the author. *Style* refers to tweets focused on language use, and *social* includes tweets related to social communication. Through their analysis the authors were able to characterize and rank individual Twitter users' tweets according to the 4Ss. Additional LDA analysis on tweets led to models that suggest better methods for ranking tweets and for recommending new people to follow on Twitter.

Another recent study examined whether the Dow Jones Industrial Average could be predicted by an analysis of public mood states as derived from a large-scale Twitter feed (Bollen, Mao, & Zeng, 2010). The authors analyzed about nine and a half months' worth of all the posts made on Twitter in 2008. After filtering the tweets and removing those that did not have any indication of author mood, the remaining tweets generally fell into one of the following mood dimensions: *calm*, *alert*, *sure*, *vital*, *kind*, and *happy* (these categories are supplied by the Google Profile of Mood States (GPOMS) tool used in the study). Using Granger causality analysis, the authors found a correlation between the degree of public "calm" (represented by tweets) and the Dow Jones Industrial Average after a lag of three to four days. Other analysis provided in this study suggests that Twitter data may be a sufficient representation of the public sentiment in some respects.

Similar to the above two studies, Asur and Huberman (2010) looked for a way to predict box office movie returns using Twitter. They examined the "tweet-rate" of 24 movies (designated by the number of tweets per hour about the movie) released between November 13, 2009 through February 26, 2010. Using a linear regression model, the authors found a strong correlation between the rate of chatter about the movie on Twitter and the eventual box office revenue of that movie. They also used sentiment analysis to assign a classification of positive, negative, or neutral to a tweet, and found further correlation between positive and negative sentiments about movies and their post-opening revenues.

## Micro-Macro Informing Framework

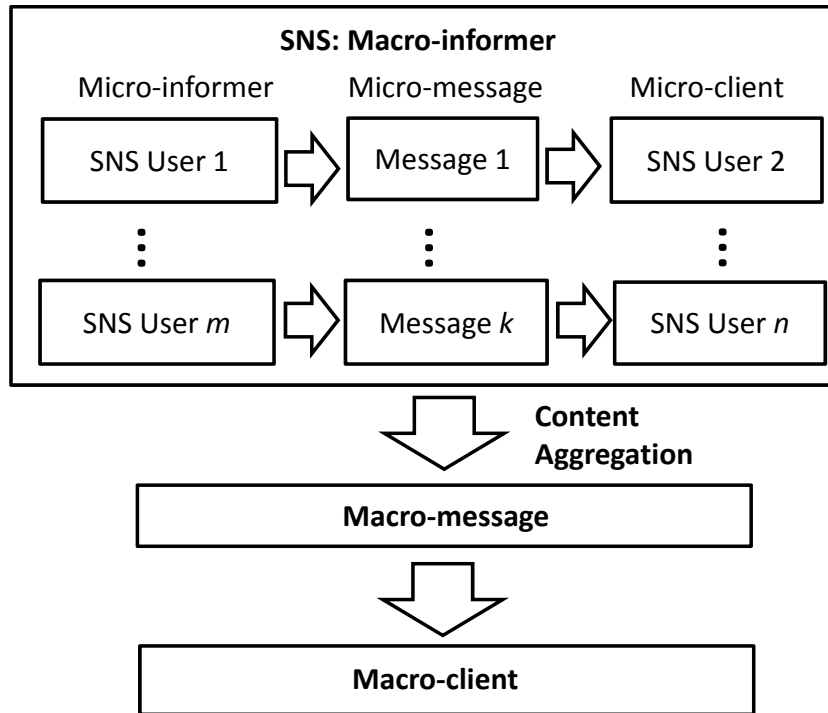
The role of Social Network Sites in information dissemination can be best understood through the lens of the Informing Science framework (Cohen, 2009). According to this framework, an informer has a need to communicate a message to the client (to inform a client). Relaying this message involves its transformation into an appropriate form and its transmission over a medium in the presence of noise. In the context of SNS, SNS users act as both informers and clients, whereas the SNS platform contributes to the context and serves as the medium. As informers, SNS users transform their messages into the format appropriate for a specific SNS. For example, Twitter, a popular micro-blogging SNS, restricts the format of relayed messages to 140 text characters. Users of Facebook can relay their message in richer formats, including longer text and im-

ages. SNS also dictate the degree of control its users have over the clientele of their messages. For instance, Facebook allows its users to limit the clients of their communication to their friends. Google+ allows its users to create several “circles” of clients, such as family, friends, and public. The technologies used by SNS provide a means of dissemination (a channel) of information that serves the needs of the informers and their clients. For example, Twitter technology, based on its lean message format, is designed to maximize its effectiveness through widespread adoption, fast distribution, and concentration of information.

As noted by Cohen (2009), even though in its basic form the Informing Science framework seems to describe a single informer informing a single client through a single channel (micro informing level), senders and clients can also appear in collective form, and messages can be directed to multiple clients. SNS is useful for supporting organizational and individual informers and is particularly instrumental in supporting one-to-many (broadcast style) communication. Regardless of whether informers and clients are collectivities or individuals, the communication is always guided by the goal of the informer. It is however possible, that a message sent by a specific informer to a specific client with a particular purpose can inform an unintended client in a manner that was not originally envisioned by the informer.

The public nature of some SNS communication offers extensive opportunities for such emergent informing, i.e., the informing that goes beyond the goal and the context originally intended by the client. Among widely cited examples of such emergent informing are travel updates posted by individuals with the goal of connecting with their friends and family attracting potential burglars, photos of cute debit cards attracting identity thieves, compromising photos reaching current or potential employers, etc. Yet, there is only a limited room for interpretation of one individual message.

A different level of emergent informing occurs when messages from multiple informers are combined and interpreted together. Such aggregation of individual messages has the potential for informing interested clients about the SNS user community in a variety of ways, many of which are unintended by the individual users. With a high degree of proliferation of SNS, an SNS user community can be representative of certain segment of a society, or of society as a whole. Thus, content generated as a part of individual informing actions within SNS can be used in aggregated form to inform emergent clients. Content analysis technologies, including document summarization, topic extraction, and pattern recognition, can be instrumental in such aggregation. The SNS, which holds all the generated content, can then become a macro informer, which can share the content in raw or aggregated forms with its own clients. An example of such aggregate informing is a trending topic in Twitter, which is a word, phrase, or topic that is tagged by various users at a greater rate than other tags. It's not until statistics on a large number of individual tags are compiled that certain topics are listed as trending, at which point a new informing cycle, involving a new, aggregate message, starts. This highlights the dual informing role of SNS: as a platform for individual informing actions and as a macro informer, informing its clients about their user community and, by extension, by the society at large. The *micro/macro informing framework* presented in Figure 1 illustrates this dual informing role of SNS. Within SNS, individual users send messages to each other, thus acting as micro-informers and micro-clients. The collection of messages can be processed using aggregation technologies to produce a macro-message, which can inform a partially unintended macro-client.



**Figure 1: The micro-macro informing framework.**

While there are privacy-related, legal and technical considerations that may influence the degree and manner in which SNS act as macro-informers, in the next section we present a field study that demonstrates how publicly available content from Twitter can be aggregated and used to inform an unintended client, the stock market.

## Field Study Method

In order to validate our framework presented in the previous section, we conducted a field study that examined the predictive potential of aggregated messages from Twitter for informing the stock market. Twitter is a SNS without a specific focus that encourages users to send and receive short (140 character) text messages between members of their networks. Founded in 2006, Twitter's growth has quickly reached phenomenal rates. In February of 2011 the company reported approximately 200 million Twitter accounts (Johansmeyer, 2011) and a month later experienced a growth spurt of, on average, 460,000 new accounts created *per day* (Smith, 2011). Twitter users must register to post on the site, but non-registrants can typically view any user's public profile and messages. Twitter users operate on a public stage: anyone can "follow" another member's posting history and thus join that member's social network without permission. If privacy is a concern to a member, the onus is on the Twitter user to "opt-out" of people following them, either by blocking specific followers or by protecting all of their messages from the public. When a Twitter user posts a message, it is visible on that user's individual site, and it is also delivered to everyone following that user.

### Data Collection

Twitter data collection was accomplished through a free service offered by TwapperKeeper ([www.twapperkeeper.com](http://www.twapperkeeper.com)). At the time of this research, TwapperKeeper provided online tools through which users could specify search terms that were used to scan the live stream of tweets

and select and save messages that contain the search terms. Archives of tweets containing specific search terms were identified and saved for the time frame associated with our study. The search terms we identified corresponded to 18 publicly-traded companies. The eighteen companies were all in the top 40 of the Fortune 500 and were selected according to whether each had an existing archive being tracked that covered the desired study time frame, as a convenience sample. The time frame for the Twitter portion of the study was November 1, 2010, through January 16, 2011, comprising 53 time periods ending at the close of a corresponding open stock trading day. The companies selected represented the following industries: Retail (online and brick & mortar), manufacturing, energy, finance, and telecommunications.

The tweet archives were closed and extracted at the end of the time frame. Once extracted, each archive was examined for spurious information not related to the underlying meaning of the search terms. For example, the tweet archive for the Ford motor company contained many references to people with the last name of Ford. These irrelevant tweets were removed through a combination of filters and manual inspection. Additionally, all tweets made in languages other than English were removed. The results yielded a collection of 3.85M tweets related to the 18 companies. Each tweet contains a date and time stamp from Twitter marking when it was posted, allowing us to make a stock trading day reference. For our text modeling purposes, a stratified sample of 8,464 tweets was compiled (see Table 1). From each company, a proportional number of tweets was included in the sample, with the additional requirement of a 100-tweet minimum. We opted for a proportional sample, rather than a fixed number of tweets from each company, because our intention was to capture the influence of the entire body of tweets on the financial market as a whole, rather than the tweets' separate influence on each individual company. Upon a final examination of the 8,464 sampled tweets, a few remaining non-English language messages were removed. A major SPAMMING event related to Dell was identified, and about 900 tweets offering a free Dell laptop were excluded, since they were considered an artifact that originated outside the community of Twitter users. However, tweets that were complaining about such SPAMMING were included in the sample.

**Table 1:** List of companies included in our study, with tweet sample sizes

F500-Rank	Ticker	Company Name	Total tweets	Sample size
5	BAC	Bank of America Corp.	128,175	275
12	C	Citigroup	20,649	100
25	COST	Costco Wholesale	103,413	221
18	CVS	CVS Caremark	86,498	186
3	CVX	Chevron	48,470	103
38	DELL	Dell	278,137	589
8	F	Ford Motor	388,638	830
39	GS	Goldman Sachs Group	36,721	100
20	IBM	International Business Machines	173,577	371
33	JNJ	Johnson & Johnson	4,240	100
36	MSFT	Microsoft	1,014,088	2,140
22	PG	Procter & Gamble	9,010	100
7	T	AT&T	287,326	614
30	TGT	Target	385,262	814
13	VZ	Verizon Communications	551,601	1,170
32	WAG	Walgreen	73,177	156
1	WMT	Wal-Mart Stores	234,786	495
2	XOM	Exxon Mobil	25,411	100
		<b>Total</b>	<b>3,849,179</b>	<b>8,464</b>

## Topic Extraction

In order to identify key topics that appear in the Twitter messages, we employed Latent Semantic Analysis (LSA), a method that extracts the semantic and conceptual content in textual data. LSA was initially introduced as an information retrieval and query optimization method (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Dumais, 2007), but later was generalized into a family of approaches for analysis of textual data. In IS research, clustering and factor analysis variants of LSA were used in order to identify research communities (Larsen, Monarchi, Hovorka, & Bailey, 2008) or research topics (Sidorova, Evangelopoulos, Valacich, & Ramakrishnan, 2008).

Latent Semantic Analysis starts with the coding of textual data into a tabular form known as the Vector Space Model (Salton, 1975). A term-by-document matrix,  $\mathbf{A}$ , containing the frequency count of term  $i$  in document  $j$ , is first compiled. After appropriate transformations,  $\mathbf{A}$  is subjected to a Singular Value Decomposition (SVD), a matrix operation that produces simultaneous principal components for two sets of variables, the set of  $T$  terms (considering the documents as observations) and the set of  $D$  documents (considering the terms as observations):

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \quad (1)$$

The products of SVD include the  $T \times r$  matrix of term eigenvectors  $\mathbf{U}$ , the  $D \times r$  matrix of document eigenvectors  $\mathbf{V}$ , and the  $r \times r$  diagonal matrix of singular values (square roots of eigenvalues), where  $r \leq \min(T, D)$  is the rank of  $\mathbf{A}$ . After truncating these matrices, keeping the first  $k$  dimensions, and after performing varimax rotations on the term loadings  $\mathbf{U}\mathbf{\Sigma}$  which are then reciprocated on the document loadings  $\mathbf{V}\mathbf{\Sigma}$ , the set of  $k$  rotated factors is associated with  $k$  articulated, socially constructed concepts. Smaller values of  $k$  represent the collection of documents in a space of fewer, more abstract concepts, and higher dimensionalities represent the documents in a space of more specific language patterns.

In our study, LSA was performed on the 8464 tweets included in the sample to identify key themes. Following standard pre-LSA text mining practices, the 8464 tweets were quantified into a term frequency matrix using 537 significant terms. For an overview of pre-LSA text processing steps, see Coussement and Van den Poel (2008). Following an approach similar to that in Sidorova et al. (2008) and taking into account the recommendations in Evangelopoulos, Zhang, and Prybutok (2010), varimax rotations of term loadings were applied to selected numbers of top principal components in order to produce interpretable latent semantic factors. Because of the exploratory nature of this study we did not have a fixed target number of themes; therefore, we conducted an exploratory examination of a number of LSA solutions, ranging between 2 and 30 LSA factors. The results of LSA included (1) a matrix of high-loading description documents, associating all descriptions with the extracted factors, and (2) a set of two sorted loading tables, associating factors with (2a) top-loading terms, and (2b) top-loading documents. The set of tables listing top-loading terms and descriptions was used in labeling the factors. The next section discusses factor interpretations in more detail.

## Influence of Twitter Topics on Stock Returns

There are two distinct goals for modeling stock price performance. One is for the purpose of stock price prediction, and the other is to investigate whether a particular type of event has a positive or negative effect on the value of a firm, where the stock price is considered a representation of the firm value. To achieve the first goal, many methods of forecasting stock prices are used in research and practice, ranging from human “black box” expertise (Soderlind, 2010) to soft computing techniques that number in the hundreds of methods (Atsalakis & Valavanis, 2009). Time series analysis is often a part of these methods and is applied in order to identify trends and analyze patterns (Andersen, Bollerslev, Frederiksen, & Nielsen, 2010; Hwang & Oh, 2010). Recent

studies of note include a time-series forecasting solution using the dividend-price ratio, earnings growth, and price-earnings ratio growth in a sum-of-the-parts method, which yielded better R-squares than standard predictive regressions (Ferreira & Santa-Clara, 2011), and a study on the causal impact of media in financial markets which found that earnings announcements in local media coverage strongly predicted local trading for S&P 500 firms (Engelberg & Parsons, 2011).

The second goal, that of investigating an event's effect on firm value, has been extensively studied. In 1973, Peter Clark theorized that stock prices change when new information violates existing expectations (Clark, 1973). This new information can be in the form of a publicized event. For example, Kaplanski and Levy (2010) found a correlation between aviation disasters and market downturn. They theorized that catastrophic events were likely to broadly affect people's moods and that negative sentiment would result in an irrational investor reaction that would lower market indices temporarily until the investors recovered, which they found typically took two days. Another study suggests that firms recognized for diversity are penalized with a decline in stock price. Specifically, the researchers found that companies receiving Fortune Magazine's "Best Companies for Minorities" award suffered a significant negative share price fluctuation as a consequence of receiving the award (Cook & Glass, 2011).

We posit that Twitter data related to firms inform the public on those firms and, therefore, influence public sentiment regarding those firms. Thus, we acknowledge the potential for reaction to that sentiment in the form of stock price fluctuation. To a certain degree, we expect the public sentiment as reflected in Twitter data to relate to fluctuation in the entire stock market, as examined in Bollen et al. (2010). In this paper we intend to take our level of inquiry one step further and investigate a possible additional influence of public sentiment on the price fluctuation of an individual stock, after accounting for fluctuations of the entire market. The next section discusses our model.

### **Modeling Stock Returns**

The Capital Assets Pricing Model (CAPM), also known as the market model, was introduced by William Sharpe (1964) and has ever since attracted a lot of attention in finance research. According to CAPM, the return  $r_i$  on security  $i$  for some given period is associated with a fixed component,  $\alpha_i$ , and a variable component,  $\beta_{iI} r_I$ , where  $r_I$  is the return on market index  $I$  over the same period, and  $\beta_{iI}$  is known as the beta coefficient of security  $i$  with respect to market index  $I$  over that period. The beta coefficient is a relative measure of sensitivity of an asset's return to changes in the return of the market index. Alpha and beta coefficients can be estimated by fitting the regression model

$$r_i = \alpha_{iI} + \beta_{iI} r_I + \varepsilon_{iI}. \quad (2)$$

Using data from the 300 trading days immediately prior to our main study period, i.e., August 21, 2009, through October 29, 2010, and using the S&P 500 index, adjusted for splits and dividends, as the market index, we fit the market model (2) for each of the 18 companies in our sample. Eighteen pairs of alpha and beta coefficient estimates were thus obtained.

After estimating the alpha and beta coefficients for each stock, we derive stock return anomalies  $RA_i$  as

$$RA_i = r_i - \hat{r}_i = r_i - \left( \hat{\alpha}_{iI} + \hat{\beta}_{iI} r_I \right), \quad (3)$$

where  $\hat{r}_i$  is the return that is expected through an association of the stock return with the market return  $r_I$ . These return anomalies, therefore, represent the component in stock returns that cannot be explained by the movements of the market. Using the 18 pairs of estimated alpha and beta co-



efficients and S&P 500 and individual stock returns from our main study period (the 53 trading days between November 1, 2010, and January 16, 2011), we obtained the  $RA_i$  values in (3) for 18 companies over 53 trading days, for a total of 954 data points.

Can information contained in company-related tweets influence the return anomaly components obtained in (3)? The market efficiency hypothesis (Fama, 1970) proposes that any lag in the response of prices to an event is short-lived. However, a number of event studies have found evidence for long-term effects on stock prices, suggesting market inefficiency. Our study follows a similar argument and proposes the employment of tweet volume  $TV_{ij}$  and up to 6-day lags for 20 different tweet topic factor strengths  $T_{ijd,t-k}$ , over 53 trading days, and across 18 companies, as predictors for stock return anomaly  $RA_{ij}$  obtained from company  $j$  on trading day  $i$ :

$$RA_{ij} = \beta_0 + \theta TV_{ij} + \sum_{k=0}^6 \sum_{d=1}^{20} \beta_{dk} T_{ijd,t-k} + \varepsilon_{ij}, \quad 1 \leq i \leq 53, \quad 1 \leq j \leq 18 \quad (4)$$

We expect, of course, the errors  $\varepsilon_{ij}$  in model (4) to be correlated. However, assuming that the correlation among the errors is not related to the correlation among the independent variables, we can expect the least squares coefficient estimates to be asymptotically unbiased (Greenwald, 1983). We also impose the restriction that the coefficients for the topic strengths  $\beta_{dk}$  in (4) stay stable during the 53-trading day period of our study. The next section presents our results in topic factor extraction, market model coefficients estimation, and fitting of the stock return anomaly model (4).

## Field Study Results

### Key Twitter Topics

After examining several factor solutions we chose to focus on a 20-factor solution for the purpose of further analysis. While increasing the number of factors to 30 resulted in emergence of some interesting unique themes, such as customer service, most other additional factors appeared to be variations on the factors existing in the 20-factor solution. Yet it is important to note that while we focus on the 20-factor solution for this exploratory study, solutions with different numbers of factors may prove to be as, or even more, useful for predicting stock performance.

In order to label the 20 extracted factors, high-loading terms and high-loading tweets for each factor were co-examined. Table 2 lists (stemmed) terms with factor loading equal to 1.0 or higher. Terms with a factor loading equal to 3 or higher are boldfaced. Table 2 also lists a few selected high-loading tweets for each of the 20 factors. Table 3 provides a brief informative description of each factor's content. For easy factor reference in the rest of the paper, both Tables 2 and 3 also provide 20 corresponding short labels.

As evident from Table 3, most of the 20 factors are formed around several key themes. Several factors are formed around types of products and services such as phone (F3), laptop (F19) or car (F16), specific products or services, such as Kinect for Xbox (F1) or Microsoft Office (F14), specific brands and company names, such as Apple (F12). The second group of factors is related to shopping, including gift card (F6), store/sales (F18), Black Friday (F8) or deal (F17).

**Table 2.** High-loading terms and high-loading tweets for the 20-factor solution

F#	Short label	Terms	Sample tweets
F1	Kinect /Xbox	<b>Kinect, Xbox,</b> free	Microsoft Kinect was the Xbox 360 with the New Kinect For 0 cent; Xbox 360 Kinect 250 GB Bundle Supplies from Microsoft Limited; I won a Free Microsoft Kinect + Xbox 360; RT @HarvardBiz: Microsoft's Kinect and the Invisible Handshake
F2	Windows phone	<b>phone, Window,</b> Android, mobil, sale	Dell choisit Windows Phone 7; Bill Gates makes the jump to Windows Phone 7 [TNW Microsoft]; Microsoft ATT Tight-Lipped on Windows Phone 7 Sales; Windows Phone 7 is the Mobile Reset Microsoft So Desperately Needed
F3	New	<b>new,</b> post	National Mortgage News - BofA JPM Resume Foreclosures: BofA has directed its foreclosure ...; Brake Hose Ford E-100 Econoline 75 76 77 78 79 80 NEW; Sweet. RT @CARandDRIVER: A new Ford Bronco? Its a possibility; RT @cbetterley: New Novell podcast: IBM adopts SUSE Linux Enterprise for WebSphere Appliances part II; You Must See This: New RegistryClear - Actually Certified by Microsoft -
F4	Free	<b>Free, xbox,</b> window, ship, bui	Free Microsoft Points XBL NDS PSP IPODS PS3 XBOX 360; Free Microsoft Enterprise 2007; I won a Free Microsoft Kinect + Xbox 360; Verizon Wireless VersaLink Wireless 4 Port Router \$10 Free Shipping:
F5	Get	<b>get,</b> like, look	Get breakfast (@ Citibank Lounge Terminal 2 Soeta); Getting my first flu shot ever! (@ CVS); How-To Geek Gets the Microsoft MVP Award Thanks to You - How-To Geek; Verizon Gets iPhoneonVerizon.com and iPhoneforVerizon.com
F6	Gift card	<b>card,</b> <b>gift,</b> win, give	Win \$200 Target Gift Card and \$25 Starbucks Card coffee thepatio; Congratulations! You have a chance to win a \$1000 Target Gift Card -; \$1000 Gift Card-WalMart-Free card-Must sign up for free card; Hi! I thought you might have interest in a free CVS Gift Card: :)
F7	Job	<b>job,</b> time, sale, manag, hire	Segment Analyst ...: Brussel - CITIBANK job IT; GS: Reinsurance Group Analyst ( NewYork NY) OneWirejobs Finance Jobs Job TweetMyJOBS; 91940 Target jobs in America; MA Jobs - Financial Advisor Trainee - Boston/Andover/Beverly at BofA (Boston MA) Job Hiring MAJobs
F8	Black Friday	<b>black, Friday,</b> ad	RT @ocyrus213: Up here @ cvs gettin this black friday shoppin in :/; Braving the black Friday madness. (@ Target); Lenmar Nickel-Metal Hydride Battery for Motorola -Target Black Friday 2010 @; Black Friday Ads: Launched Walmart Black Friday Ad and Target Black Friday 2010 - Keen Observers
F9	LOL	<b>LOL</b>	lol I got BofA tho :-/; lol RT amy_bby909: crossingq cvs reminded me of when I was drunk :o lol; um I have Verizon lol; RT @Lino0524: That Microsoft phone commercial is dumb funny; Really?!?! Lol<!--you would say that!!! SMDH lol

**Table 2 (continued).** High-loading terms and high-loading tweets for the 20-factor solution

F#	Short label	Terms	Sample tweets
F10	Not	<b>not</b>	the BofA-GSE settlement: Were we robbed or not? Hw; Microsoft website not opening; To go to Target or to not go to Target. That is the question.....; Four reasons to not get the Verizon iPhone
F11	All	<b>all</b> , do, us, mobile, know, like, get, work, come, no, good, time, ther, service, more, custom, not	Converting and burning Microsoft PowerPoint (PPT) presentations in nearly all formats (including ppt pptx pptm pp...; Target is where all the hipster adults hang out; Off to Lowes Sam's and Spec's... and probably WalMart since I shot up all my ammo :) Ok- so what do you all think of Ford Escape or Jeep Liberty / GCherokee? Anything from Chevy? GMC? What makes/models have "stigmas"?
F12	Apple	<b>Apple</b> , ipad, google	Apple iPhone to go corporate with BofA Citigroup via @MobileBurn; RT @glynmoody: Will Visa MasterCard Paypal BofA & Apple Terminate Relationships With the NYTimes For Revealing Military Secrets? - ht ...; Apple and Verizon: the bumpy road to love; Ooh...Apple-ey.... in response to the new Microsoft Powerpoint
F13	Like	<b>like, get, store</b> , myworldsacoust, look, album, justinbieb, don't	Apple - Verizon partnership is finally hashed out "If you cant get what you like you have to like what you can get"; ATT gots me out here in the dark like a dummy; the Target lady looks like Be thenny Frankel!!!; RT @justinbieber: glad you guys out there are liking MY-WORLDSACOUSTIC ...if you dont have it get to WalMart and grab that cd mang!
F14	Office	<b>Office</b> , google	Microsoft Office 2010 Portable (12-17-2010); Microsoft Office For Mac Standard 2011-XiSO: Microsoft Office For Mac Standard 2011-XiSO; Anybody have Microsoft Office 2010 and Adobe Suite; Trinidad Guardian Facebooks Sexy Microsoft Affair Continues With Office App Integration ... guardian.co.tt
F15	Mobile	<b>mobile</b> , window, more, announce, report, servic	Why is mobile banking for BofA so shitty! Microsoft Wireless Mobile Mouse 3000 for \$10 + \$6 s&h -; Mobile phones: WITH Wednesday's announcement by Verizon Wireless the largest of America's four big mobile phone ..; Daily Mobile News: Verizon ATT and T-Mobile debuting a joint mobile-payment project smartphone
F16	Car	<b>car, sale</b> , focus, price, mustang, video	car sale cincinnati 2007 Ford Focus S 47000mi (Western Hills) \$6800 (price); 1932 Ford V-8 Bonnie And Clyde Diecast Car; the Ford Raptor may have surpassed the Edge as my favorite car. the raptor is legit; For sale: ATT Car Charger w/ USB Port for Apple iPhone 3GS 3G & iPod

**Table 2 (continued).** High-loading terms and high-loading tweets for the 20-factor solution

F#	Short label	Terms	Sample tweets
F17	Deal	<b>deal,</b> dai, unwrap, dail, season, surpris	Dell Deals of the Day;; Ford E-150 Econoline Accessory Belt Tensioner Dayco - Cheap Deals; Target: new LEGO deals from 16 Dec including 7208 Fire Station for \$89; software deals: Hidden Mysteries: Vampire Secrets
F18	Store/Sales	<b>store,</b> <b>sale,</b> time, like, mobil, percent	\$20 OFF Wireless Desktop 3000 at the Microsoft Store! Limited time only. Microsoft; Costco Sustains Sales Momentum - Analyst Blog: After a 6 percent increase in October 2010 Costco's comparable-store sa...; espana Costco November same-store sales up 9 percent economia em-presas; BofA Resumes Foreclosure Sales
F19	Windows laptop	<b>window,</b> <b>laptop,</b> xp, more, android	[H N?] Laptop dell inspiron 1564 - vga r?i windows; Laptop Choice - Dell Mac or Sony; Dells new Convertible Windows 7 tablet Dell Inspiron Duo thetechcheck dell laptop Windows7 RT @vickinebetienne: @shelbyfero Dell Laptop Your New Dell XPS M1530 Laptop is here
F20	Buy	<b>bui,</b> google, alla	Someday Google will buy Microsoft Rocky Termanini @clarkcollge; turn around go back to town buy a ford raptor; Buying stoges (@ Walgreens); Thank you so much Google for donating \$100k 4 Bpark to buy the Turing papers today!!! @IBM @Microsoft can you help??

Another group of factors is formed around actions such as get (F5) or buy (F20), or expressions of emotions such as LOL (“laughing out loud”) (F10). Two factors seem to stand on their own and do not clearly belong to any of the aforementioned groups. These include the job factor (F7) and the new factor (F4). In addition, two factors are formed around individual words such as like (F13) and all (F11) and do not appear to represent any particular theme. It is important to note that the themes associated with these factors are not representative of the overall population of Twitter content, since we analyzed only tweets that mention names or ticker symbols of specific companies. Rather they are indicative of the community discourse related to specific publically-traded companies of large capitalization.

Table 4 shows the number of high-loading tweets by factor and by company. As evident from the table, different companies account for different numbers of overall tweets. Microsoft (MSFT) accounts for the largest number of tweets, perhaps reflecting the release of two major consumer products (Kinect and Windows Phone 7) during or shortly before the observation period, as well as the tech-savvy customer base and consumer focus (as opposed to IBM). AT&T (T) and Verizon (VZ) also had a large number of tweets associated with them, perhaps also reflecting consumer focus, combined with tech-savvy customers. Surprisingly, both consumer products giants Johnson&Johnson (JNJ) and Proctor&Gamble (PG) had relatively few tweets in the sample, and even fewer among the tweets loading on the top 20 factors. Banking and investment companies Citigroup (C), Goldman Sachs (GS), and Bank of America (BAC), as well as oil and gas companies Chevron (CVX) and ExxonMobil (XOM), also had relatively few tweets.

**Table 3.** List of factors, labels and description

<b>Factor</b>	<b>Short label</b>	<b>Description</b>
F1	Kinect /Xbox	Tweets about Microsoft Kinect for Xbox 360
F2	Windows phone	Tweets about Windows Phone 7 and other mobile phones as well as Windows OS
F3	New	Tweets about news, new products, policies, as well as new items for sale
F4	Free	Advertisements and reminders of promotions involving free products & services
F5	Get	Tweets on a variety of topics that include the word "get" , but often refer to getting a product or service
F6	Gift card	Advertisements of free gift cards from major retailers, or chances to win such gift cards
F7	Job	Mostly job posts, but also includes references to news about job creation, loss, outsourcing, etc.
F8	Black Fri- day	Announcements of Black Friday deals, references to news about Black Friday shopping, status posts involving Black Friday
F9	LOL	Tweets on a variety of topics that include the acronym "lol". May indicate positive mood.
F10	Not	Tweets on a variety of topics including the word "not". May indicate negativity.
F11	All	Tweets on a variety of topics including the word "all". Does not appear to have a common theme.
F12	Apple	News involving Apple and its products. Apple is usually mentioned as a collaborator or competitor of one of the companies involved in the study.
F13	Like	Tweets on a variety of topics including the word "all". Does not appear to have a common theme.
F14	Office	Tweets about Microsoft Office applications.
F15	Mobile	Tweets about mobile phones and mobile technologies and their use. Also includes references to companies T-Mobile (usually as a competitor of AT&T or Verizon) an Exxon Mobil.
F16	Car	"Car for sale" offerings, mostly of various Ford models
F17	Deal	New of corporate deals and announcements of consumer deals, usually at retail stores or websites
F18	Store/Sales	Announcements of same store sales and store openings by retailers and arrival of new products to stores, announcements of sales at retail locations, store jobs announcements, other store and sale related tweets.
F19	Windows laptop	Tweets about models of laptops, often with references to Windows OS
F20	Buy	Tweets about personal purchases (or purchase intentions), advises to buy or not to buy a product, news of corporate stock/share purchases

Not surprisingly, the majority of tweets loading highly on factors representing types of products or specific products are associated with companies that produce or carry these products or competing products. For example, almost all Kinect/Xbox tweets (F1) and Office tweets (F14) were associated with Microsoft, Windows Phone tweets (F2) were mostly associated with Microsoft (the producer), as well as AT&T and Verizon (the distributors). Windows/laptop tweets split between Dell and Microsoft, and car tweets were mostly related to Ford. The job factor was relatively highly represented in the tweets associated with banking companies, as well as in the technology sector (Microsoft, IBM, AT&T). Other factors seemed to be rather evenly represented in tweets associated with different companies.

**Table 4.** Tweets by factor by company

F#	Short label	BAC	C	COST	CVS	CVX	DELL	F	GS	IBM	JNJ	MSFT	PG	T	TGT	VZ	WAG	WMT	XOM	Total
F1	Kinect /Xbox	0	0	0	0	0	0	0	0	0	0	354	0	2	10	0	0	1	0	367
F2	Windows phone	6	2	1	1	0	38	2	0	8	0	266	0	107	3	175	4	3	0	616
F3	New	17	3	11	6	6	49	78	11	45	7	195	10	39	32	110	3	22	2	646
F4	Free	0	4	7	15	3	33	11	2	5	0	250	1	25	31	34	13	26	1	461
F5	Get	15	5	32	30	5	22	44	2	18	4	109	2	55	75	126	17	73	4	638
F6	Gift card	15	14	7	18	13	19	11	1	2	0	19	0	8	92	26	14	23	2	284
F7	Job	15	8	2	2	2	9	14	6	25	10	49	4	59	14	16	4	2	2	243
F8	Black Friday	5	1	4	5	1	20	10	1	5	0	17	2	8	115	16	4	39	1	254
F9	LOL	6	1	7	16	3	12	16	0	4	2	22	0	19	31	20	9	29	2	199
F10	Not	9	4	13	6	2	9	14	2	10	3	72	5	31	27	45	4	13	0	269
F11	All	35	7	33	31	4	37	60	7	28	4	173	4	95	93	144	24	52	14	845
F12	Apple	4	0	5	1	0	11	0	1	7	1	112	3	28	8	84	0	4	3	272
F13	Like	9	5	28	19	7	13	29	1	10	2	108	2	45	71	89	16	114	4	572
F14	Office	2	0	0	0	0	6	1	2	3	0	159	1	2	4	6	0	1	1	188
F15	Mobile	8	1	5	0	2	25	12	5	7	3	91	6	92	14	119	0	7	30	427
F16	Car	7	2	10	6	5	8	310	1	17	6	37	4	25	31	23	8	17	2	519
F17	Deal	10	0	10	23	3	24	11	4	6	1	44	2	19	65	77	7	13	1	320
F18	Store/Sales	10	4	23	14	8	17	62	3	11	5	94	3	71	71	58	19	57	4	534
F19	Windows laptop	3	0	3	1	1	196	4	1	11	0	205	1	6	5	44	0	3	0	484
F20	Buy	9	1	11	11	6	19	19	0	12	2	141	4	14	24	28	3	18	4	326
	<b>All 20</b>	<b>185</b>	<b>62</b>	<b>212</b>	<b>205</b>	<b>71</b>	<b>567</b>	<b>708</b>	<b>50</b>	<b>234</b>	<b>50</b>	<b>2517</b>	<b>54</b>	<b>750</b>	<b>816</b>	<b>1240</b>	<b>149</b>	<b>517</b>	<b>77</b>	
	<b>Sample size</b>	<b>275</b>	<b>100</b>	<b>221</b>	<b>186</b>	<b>103</b>	<b>589</b>	<b>830</b>	<b>100</b>	<b>371</b>	<b>100</b>	<b>2140</b>	<b>100</b>	<b>614</b>	<b>814</b>	<b>1170</b>	<b>156</b>	<b>495</b>	<b>100</b>	

In the next section we discuss how the content of tweets is related to the stock performance of individual companies.

### **Stock Price Predictors**

In order to establish the macro informing value of Twitter as an example SNS, we examine whether the content of tweets can be a significant predictor of stock price performance for the companies mentioned in the tweets. We started by fitting the 18 market models in (2). Table 5 presents 18 pairs of alpha and beta estimates. They were all estimated using 300 returns (obtained from 301 trading days) from the period that immediately precedes our main study period, i.e., immediately prior to November 1, 2010. Beta coefficients vary from 0.375 (for Wal-Mart) to 1.614 (for Bank of America).

Using (3), return anomalies were computed for each of the 18 companies over the 53-trading day period. These return anomalies were used as the dependent variable in fitting the regression model in (4). Results of the regression analysis indicate that a number of the topic factors are significant leading indicators in predicting stock price change among the companies included in our study. Table 6 presents significant predictors and the lag time under which each term becomes effective. The regression model explains 8.3% of the variability in return anomalies (based on the regression model's  $R^2$  coefficient). The overall regression model F-statistic is equal to 4.69, which is significant at the 0.001 level.

**Table 5.** Estimated alpha and beta coefficients using the Market Model

Stock Symbol	Est. alpha	Est. beta	F-stat	Significance	R-sq (%)
BAC	-0.001993	1.614	485.24	< 0.001	62.0
C	-0.000833	1.505	188.58	< 0.001	38.8
COST	0.000660	0.600	138.16	< 0.001	31.7
CVS	-0.000722	0.733	67.59	< 0.001	18.5
CVX	0.000249	0.963	832.24	< 0.001	73.6
DELL	-0.000417	1.150	185.06	< 0.001	38.3
F	0.001518	1.482	249.40	< 0.001	45.6
GS	-0.000371	1.009	159.93	< 0.001	34.9
IBM	0.000336	0.728	346.26	< 0.001	53.7
JNJ	0.000038	0.466	238.27	< 0.001	44.4
MSFT	0.000014	0.816	241.33	< 0.001	44.7
PG	0.000467	0.507	176.12	< 0.001	37.1
T	0.000311	0.572	214.29	< 0.001	41.8
TGT	0.000160	0.750	217.04	< 0.001	42.1
VZ	0.000165	0.498	116.13	< 0.001	28.0
WAG	0.000002	0.800	116.62	< 0.001	28.1
WMT	0.000090	0.375	87.45	< 0.001	22.7
XOM	-0.000450	0.810	497.90	< 0.001	62.6

**Table 6.** Significant predictors for stock return anomalies and their coefficients

<b>Predictor</b>	<b>Time lag</b>	<b>Coefficient</b>	<b>T statistic</b>	<b>significance</b>
Constant		0.0003659	0.69	0.491
Tweet Volume	$t$	-0.00006303	-2.43	0.015
Get (F5)	$t - 1$	0.0198	2.5	0.013
Gift Card (F6)	$t - 1$	-0.0174	-2.93	0.003
Black Friday (F8)	$t - 1$	0.0227	1.89	0.06
Apple (F12)	$t - 1$	0.0185	2.04	0.041
Deal (F17)	$t - 1$	-0.0233	-2.23	0.026
Store/Sales (F18)	$t - 1$	-0.0253	-2.69	0.007
Buy (F20)	$t - 1$	0.0136	1.79	0.074
Deal (F17)	$t - 2$	0.0210	2.05	0.041
Store/Sales (F18)	$t - 2$	-0.0224	-2.25	0.025
Get (F5)	$t - 3$	-0.0161	-1.83	0.068
Black Friday (F8)	$t - 3$	0.0323	2.71	0.007
All (F11)	$t - 4$	0.0340	3.12	0.002
Car (F16)	$t - 5$	-0.0326	-2.7	0.007
Store/Sales (F18)	$t - 5$	0.0219	2.17	0.03
Car (F16)	$t - 6$	0.0246	2.1	0.036

The results suggest that both the total number of tweets mentioning the company and the number of LSA factors appear to be predictive of stock performance. Specifically, the model suggests that the total number of tweets about a company is negatively related to the stock performance during the same trading day. The number of tweets loading on LSA factors such as get (F5), Black Friday (F8), Apple (F12) and buy (F20) has a positive effect on the company's stock performance during the next trading day (with the lag of 1 day). The number of tweets about gift cards, deals and stores/sales has a negative effect on the stock performance during the next trading day. Some of these effects are reinforced and some are corrected during the following trading days. For example, Black Friday tweets have a positive effect with lag 1 and with lag 3, whereas tweets about store/sales have a negative effect with lag 1 and lag 2, and a positive effect with lag 5. Similarly, the positive effect of the get factor (F5) with lag 1 is partially corrected with two days later (lag 3). Similar correction can be observed for the deal factor (F17) and the car factor (F16).

Overall, our results suggest that tweet topics can explain over 8% of variance in stock performance above that explained by the market. While previous research (Asur & Huberman, 2010; Bollen et al., 2010; Ramage et al., 2010, etc.) suggested that Twitter data can, in some manner, serve as an indicator of public sentiment, attitude, or intent and even predict the movement of the stock market, this study suggests that Twitter data carries important information that can be used to predict the performance of individual stocks.

Because of its demonstrative nature, the field study presented here has several important limitations. First, the study assumes a rather simplified model of stock performance and also assumes that market reaction to information takes an entire day. Interested researchers are encouraged to examine the predictive value of Twitter data using more sophisticated models for stock prediction and also try to see if shorter reaction time (less than one day) yields better predictive results. Secondly, different alternative LSA factor solutions, perhaps with as many as 100 or 200 factors



should be examined. This would allow examining the predictive value of rare topics which may be more relevant for investment decisions. Finally, interested researchers are encouraged to replicate our study by expanding its scope to include more companies and perhaps to incorporate other information into the predictive model.

## Discussion of the Results

The results of this study have important implications for informing science researchers and practitioners alike. First and foremost, they highlight the fact that content generated through SNS communication can, in aggregate form, be used to inform clients beyond the original informing goals set by senders of individual messages. While concerns may be raised over the potential change in informers' behavior once they become aware of their macro-informing role, considering the public nature of Twitter, such changes are expected to be minimal.

Our study results may appear to some readers, to be somewhat intuitive: How could Twitter *not* have reflected what market players are doing? However, we point out that our results find certain Twitter topics to be leading indicators of stock market activity, rather than a reflection on it. Specifically, our results suggest that tweeting about “buying” or “getting” something is positively reflected in the future stock performance of the companies mentioned in those tweets, even though some of the effects are delayed by several days. While both of these factors are rather general in nature, they potentially indicate the perceived value of goods and services (or even stock) of the companies mentioned in those tweets and thus are leading indicators of future stock appreciation.

Interestingly, tweets about store/sales, deals, and gift cards of companies appear to be an indicator of future negative stock performance. A possible explanation is that these factors represent promotions by companies designed to overcome negative circumstances and may be indicative of, for example, weaker demand for their goods and services, thereby either providing a tip to investors about, or serving as a reflection of, potentially poorer financial performance. Marketing studies have generally found that investor reaction mirrors consumer reaction to price promotion (Blattberg, Briesch, & Fox, 1995; Pauwels, Silva-Risso, Srinivasan, & Hanssens, 2004; Srinivasan, Pauwels, Hanssens, & Dekimpe, 2004), and that reaction to sales and deals is generally strong, immediate, and positive (Srinivasan & Hanssens, 2008). Therefore, another explanation is that consumer reaction to the particular deals from the companies studied was poor, resulting in a similar reaction from investors.

Another interesting finding is the positive effect of the Apple factor, which suggests that being associated with Apple leads to improved stock performance of companies. Interestingly, none of the factors associated with specific products were found to be predictors of stock performance. One possibility is that market reaction to the initial release of these products occurred before the beginning of the observation period, and the product factors failed to differentiate between positive and negative consumer sentiment regarding those products.

Do the factors captured by our Twitter analysis represent components of rational decision making, or do they reflect the realm of “animal spirits”? Animal spirits refer to a motivational force within individuals that moves them toward being restless or inconsistent in the face of uncertain economic factors (Akerlof & Schiller, 2009). The theory behind animal spirits attempts to answer economic questions that defy explanation by purely economic motivational factors pursued rationally. As the theory's name suggests, these motivational factors are somewhat mysterious. We believe that our results capture both decision making dimensions, including some factors that are mainly rational and some others that are primarily emotional. Unlike more formal communication channels, such as market analyst reports, Twitter, in its role as a macro-informer, offers the opportunity to capture the animal spirit perspective.

Given the subject and scope of this research, an obvious question is what causal role, if any, tweets might exert on stock prices. While we do not imply causality in this paper, we certainly do not rule it out. One potential causal role for Twitter messages is to bolster the animal spirit confidence in some investors who were already considering buying a stock, so that upon seeing positive mention of the company or stock, their feeling of confidence increases and they move to invest. This may be reflected in the positive effect of the “Buy” (F20) and “Get” (F5) factors on next-day stock prices (see Table 6). Another causal mechanism may simply be the informing nature of Twitter. Since Twitter enables rapid, selective exchange of information in a push rather than pull format, specific and timely information can be received by investors faster than in the past. This could have the dual effect of speeding up the pace of the market and informing more clients than before, resulting in increased market activity.

From an informing science point-of-view, this exploratory study demonstrates the macro-informing potential of SNS. While this research study focuses on stock prices, potential areas of macro-informing by SNS can range from marketing research to political analysis, to sociological and anthropological studies. Macro-informing phenomena in such areas present fruitful directions for future research in informing science. One such direction could be the design and evaluation of real-time macro informing systems for different macro audiences. Twitter’s *trending* feature is an example of such a macro informing system. An evaluation of use and effects of this system might be an interesting direction for future study.

It is important to note that the emergence of macro-informing is closely related to the proliferation of technologies that enable efficient and inexpensive information storage, retrieval, and analytics. Data mining and text mining methods are particularly instrumental in distilling targeted macro messages from the multitude of spontaneous micro messages.

The results of this study can help identify some new questions. For example, how do tweets compare to other communication by the participants? This question could be the basis for future research, where multiple communication channels including Twitter are monitored and compared. Obviously Twitter’s forced brevity of 140 characters and the public nature of Twitter communication is bound to have an effect on what and how people say things. This could be a direction for a future study.

As we try to correctly interpret our findings, we should also add a reader’s caveat. Because our analysis is frequency-based, it may have overlooked certain Twitter messages that are specifically focused on stock prices, targeting the investor audience. Our study was designed to capture the general public discourse and evaluate its influence on stock prices; therefore its insights should be considered complementary to insights afforded by specialized informing systems designed for investors.

## Conclusion

In this paper we introduce the micro/macro informing framework for SNS-based communication. The framework suggests that SNS allow for informing at two levels: an individual message (micro) level, and an aggregate (macro) level. A field study conducted to validate our framework, suggests that Twitter, a popular social networking site, plays a dual micro/macro informing role: At a micro level, it operates as a platform for connecting individual informers (those that post, or “tweet” information) with their clients (those that receive that information). At a macro level, Twitter community-generated content can be analyzed using text mining technologies to distill a macro message which can inform potential unintended clients about future economic activity, such as stock price performance. The micro/macro informing framework is useful in explaining how content from the growing number of social networking sites can be mined to inform inter-

ested parties about diverse societal and economic trends. Such information can be used to support decision making in the era of Big Data.

## References

- Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. In P. Golle & G. Danezis (Eds.), *Proceedings of 6th Workshop on Privacy Enhancing Technologies* (pp. 36-58). Cambridge, UK: Robinson College.
- Akerlof, G. A., & Schiller, R. J. (2009). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton, NJ: Princeton University Press.
- Andersen, T. G., Bollerslev, T., Frederiksen, P., & Nielsen, M. (2010). Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics*, 25, 233-261.
- Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. *arXiv:1003.5699v1* [cs.CY], March 29, 2010. Retrieved from <http://arxiv.org/abs/1003.5699v1>
- Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, 36(3), 5932-5941.
- Blattberg, R. C., Briesch, R., & Fox, E. (1995). How promotions work. *Marketing Science*, 14(3), 122-132.
- Bolar, K. P. (2009). Motives behind the use of social networking sites: An empirical study. *ICFAI Journal of Management Research*, 8(1), 75-84.
- Bollen, J., Mao, H., & Zeng, X. (2010). Twitter mood predicts the stock market. *arXiv:1010.3003v1* [cs.CE], October 14, 2010. Retrieved from <http://arxiv.org/abs/1010.3003v1>
- Boyd, D. (2006). Friends, friendsters, and Top 8: Writing community into being on social network sites. *First Monday*, 11(12), December 4, 2006. Retrieved from <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1418/1336>
- Boyd, D., & Ellison, N. B. (2008). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13, 210-230.
- Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1), 135-155.
- Cohen, E. B. (2009). A philosophy of informing science. *Informing Science: the International Journal of an Emerging Transdiscipline*, 12, 1-15. Retrieved from <http://www.inform.nu/Articles/Vol12/ISJv12p001-015Cohen399.pdf>
- Cook, A., & Glass, C. (2011). Leadership change and shareholder value: How markets react to the appointments of women. *Human Resource Management*, 50(4), 501-519.
- Coussement, K., & Van Den Poel, D. (2008). Improving customer complaint management by automatic email classification using linguistic style features as predictors. *Decision Support Systems*, 44(4), 870-882.
- Deerwester S., Dumais, S., Furnas, G., Landauer, T., & Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.
- Dwyer, C., Hiltz, S. R., & Widmeyer, G. (2008). Understanding development and usage of social networking sites: The social software performance model. *Proceedings of the 41st Hawaii International Conference on System Sciences (HICSS)*, Waikoloa, Big Island, HI, January 7-10, 2008.
- Dumais, S. (2007). LSA and information retrieval: Getting back to basics. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 293-322). Mahwah, NJ: Lawrence Erlbaum Associates.

## The Dual Micro/Macro Informing Role of Twitter

- Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "Friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4), 1143-1168.
- Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, 66(1), 67-97.
- Evangelopoulos, N., Zhang, X., & Prybutok, V. (2012). Latent Semantic Analysis: Five methodological recommendations. *European Journal of Information Systems*, 21(1), 70-86.
- Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25, 383-417.
- Farfan, B. (2011). Complete and updated list of U.S. retail industry companies using Facebook. About.com. 100(3). Retrieved from [http://retailindustry.about.com/od/retailbestpractices/a/us\\_retail\\_companies\\_on\\_facebook\\_complete\\_list.htm](http://retailindustry.about.com/od/retailbestpractices/a/us_retail_companies_on_facebook_complete_list.htm)
- Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, 100(3), 514-537.
- Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior*, 25, 153-160.
- Goldman, D. (2012). Facebook tops 900 million users. *CNN/Money*. Retrieved from <http://money.cnn.com/2012/04/23/technology/facebook-q1/index.htm> .
- Greenwald, B. C. (1983). General analysis of bias in the estimated standard errors of least squares coefficients. *Journal of Econometrics*, 22(3), 323-338.
- Harris, J., & Rogers, J. (2011). INSIDER – Mining Twitter for trading profits. Reuters.com. Retrieved from <http://www.reuters.com/article/2011/06/15/twitter-idUSN159366620110615>
- Hwang, H., & Oh, J. (2010). Fuzzy models for predicting time series stock price index. *International Journal of Control, Automation, and Systems*, 8(3), 702-706.
- Johansmeyer, T. (2011, February 3). 200 million Twitter accounts... But how many are active? *Social Times*. Retrieved from [http://socialtimes.com/200-million-twitter-accounts-but-how-many-are-active\\_b36952](http://socialtimes.com/200-million-twitter-accounts-but-how-many-are-active_b36952)
- Kaplanski, G., & Levy, H. (2010). Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics*, 95(2), 174-201.
- Kulikowski, L. (2011). Survey: Social media is a must for business. TheStreet.com. Retrieved from <http://www.thestreet.com/story/11149978/1/survey-social-media-is-a-must-for-business.html>
- Larsen, K., Monarchi, D., Hovorka, D., & Bailey, C. (2008). Analyzing unstructured text data: Using latent categorization to identify intellectual communities in information systems. *Decision Support Systems*, 45, 884-896.
- Magro, M. J., Ryan, S. D., Sharp, J. H., & Ryan K. A. (2009). Using social networking for educational and cultural adaptation: An exploratory study. *Proceedings of the Fifteenth Americas Conference on Information Systems*, San Francisco, CA, August 6-9, 2009.
- MarketingProfs.com. (2011). Facebook growth outpacing other social media. MarketingProfs.com. Retrieved from <http://www.marketingprofs.com/charts/2011/5212/facebook-growth-outpacing-other-social-media>
- Pauwels, K., Silva-Risso, J., Srinivasan, S., & Hanssens D. M. (2004). New products, sales promotions and firm value, with application to the automobile industry. *Journal of Marketing*, 68(4), 142-156.
- Naaman, M., Becker, H., & Gravano, L. (2011). Hip and trendy: Characterizing emerging trends on Twitter. *Journal of The American Society For Information Science and Technology*, 62(5), 902-918.

- Perez, S. (2012). ComScore U.S. Internet report. Techcrunch.com. Retrieved from <http://techcrunch.com/2012/06/14/comscore-us-internet-report-yoy-pinterest-up-4000-amazon-up-30-android-top-smartphone-more/>
- Pring, C. (2012). 100 social media statistics for 2012. Thesocialskinny.com . Retrieved from <http://thesocialskinny.com/100-social-media-statistics-for-2012/>
- R2Integrated. (2010). Social media market survey. R2Integrated.com. Retrieved from [http://www.r2integrated.com/Portals/30/pdfs/surveys/SMM\\_12\\_10.pdf](http://www.r2integrated.com/Portals/30/pdfs/surveys/SMM_12_10.pdf)
- Ramage, D., Dumais, S., & Liebling, D. (2010). Characterizing microblogs with topic models. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington, D.C., May 23-26, 2010.
- Richter, A., & Koch, M. (2008). Functions of social networking services. *Proceedings of the Eighth International Conference on the Design of Cooperative Systems (COOP '08)*, Carry-le-Rouet, Provence, France, May 20-23, 2008.
- Salton, G. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620.
- Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3), 425-442.
- Shen, K. N., & Khalifa, M. (2010). A research framework on social networking sites usage: Critical review and theoretical extension. *IFIP Advances in Information and Communication Technology*, 341, 173-181.
- Sidorova, A., Evangelopoulos, N., Valacich, J., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly*, 32(3), 467-482 & A1-A20.
- Soderlind, P. (2010). Predicting stock price movements: Regressions versus economists. *Applied Economics Letters*, 17, 869-874.
- Smith, C. (2011, March 3). Twitter user statistics show stunning growth. *The Huffington Post*. Retrieved from [http://www.huffingtonpost.com/2011/03/14/twitter-user-statistics\\_n\\_835581.html](http://www.huffingtonpost.com/2011/03/14/twitter-user-statistics_n_835581.html)
- Srinivasan, S., & Hanssens, D. M. (2008). Marketing and firm value: Metrics, methods, findings, and future directions. *Boston University School of Management Research Paper Series*, No. 2009-6. Retrieved from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1136332](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1136332)
- Srinivasan, S., Pauwels, K., Hanssens, D., & Dekimpe, M. (2004). Do promotions benefit manufacturers, retailers or both? *Management Science*, 50(5), 617-629.
- Taraszow, T., Aristodemou, E., Shitta, G., Laouris, Y., & Arsoy, A. (2010). Disclosure of personal and contact information by young people in social networking sites: An analysis using Facebook profiles as an example. *International Journal of Media and Cultural Politics*, 6(1), 81-101.

## Biographies



**Nicholas Evangelopoulos** is an Associate Professor of Decision Sciences at the University of North Texas and a Fellow of the Texas Center for Digital Knowledge. His research interests include Statistics and Text Mining. He received his Ph.D. in Decision Sciences from Washington State University and his M.S. in Computer Science from the University of Kansas. His research interests include statistics and text mining. His publications include articles appearing in *MIS Quarterly*, *Communications of the ACM*, *Communications in Statistics-Simulation & Computation*, and *Computational Statistics & Data Analysis*.



**Michael J. Magro** is an Assistant Professor of Information Systems and Computing Technology at Shenandoah University in Virginia. He received his Ph.D. in Business Computer Information Systems from the University of North Texas and an MS in Information Technology from the University of Redlands. Prior to pursuing his Ph.D., he worked in industry as a software developer, project manager, and IT solutions provider. His research interests are social media and social networking applications, hedonic information systems, IT and IS education, and system design.



**Anna Sidorova** is an Assistant Professor at the University of North Texas. She received her Ph.D. in MIS from Washington State University. Her research and professional interests include IT-enabled organizational transformation, business process management, and open source software development. Her publications include articles appearing in *MIS Quarterly*, *Journal of MIS*, *Journal of the AIS*, and *Decision Support Systems*.