When the Going Gets Tough, The Tweets Get Going! An Exploratory Analysis of Tweets Sentiments in the Stock Market

Andrey Kretinin
William Paterson University

Jim Samuel
William Paterson University

Rajiv Kashyap
William Paterson University

Information artifacts in social media can have significant impact on domains and subject matter opinions. Asset prices, stock prices and volumes, and metrics are influenced by turbulence in their information ecosystems. Analyses of digital information networks and social media information events have shown information artifacts to affect stock performance without consistent correlation to fundamentals. Thus it becomes important to dispel ambiguity and gain insights into how the sentiments associated with information artifacts in Twitter (due to its extensive usage in relevant signaling), are associated with equity movements. Using an exploratory analysis to study tweet sentiment, we link stock price variations, and explore associated metrics. Our tweets analytics deploying textual analytics, identifies forms in tweet behavior and sentiment shifts.

INTRODUCTION

Information systems, driven by technology, have spawned vast complexities of information overload and complexity. Diversity of formats, access, and networks are all part of the emerging and continually expanding information ecosystem shaped by Google, Facebook, Yahoo, Pinterest, Twitter, Instagram, LinkedIn, and others (Fuchs 2017). This critical and sensitive information ecosystem has expanded explosively and societies are reactively grappling with ways to deal with the fallout (npr.org, 2017). In particular, the increasing influence of social media on extreme behaviors is a cause of great concern among social scientists (Ferrara 2015; Monsted et. al., 2017).

- In 2013, a fake tweet about an explosion in the White House wiped out $130 billion in market value in a matter of seconds, as high frequency traders executed millions of trades per second (‘One fake tweet’, 2013). The heightened response to the expression of negative sentiments in social media has been well researched and documented (e.g. Tsytsarau, Palpanas, and Castellanos 2014).
- In 2014, Cynk Technology Corp’s stock appreciated more than 36,000% as its price stocks surged from less than $0.10 to above $20 a share in a few weeks. This was the result of the publicity

generated via fake discussions among social media bots and spam accounts (Ferrara et al 2015). As this case illustrates, investors also overreact to positive news.

Before the advent of social media, researchers suggested that investors develop heuristics to interpret new information (e.g, Barberis, Shleifer, and Vishny 1998). Specifically, Barberis et al (1998) postulated that investors may categorize information about new company events based upon the most recent event (representativeness bias) or frame of mind (conservatism bias). Accordingly, they might be *too quick* or *too slow* to react to company news. Investors utilizing heuristics based upon representativeness were likely to overreact, while a conservatism bias might cause them to underreact to positive and negative information trends. However, social media has led to a rapid acceleration in investor responses, exacerbating the need to focus on a different set of information attributes, that is, information accuracy and veracity. Two research questions guide our study: How do investors react to news that is not always accurate or easily verifiable? How do they incorporate positive versus negative social media sentiment into their investment decisions?

Start Spreading the News! Social media networks are dynamic and interactive, resulting in a complex system characterized by varying degrees of information accuracy and veracity. Unfortunately, this provides opportunity for benefit as well as deception. Accurate and vital news can be disseminated efficiently, but so can fake news (Ferrara 2015). For instance, social media networks have been used to enable democratization (Conover et al 2013), disaster relief (Sakaki et al 2010), and social awareness of important health (Centola 2011) and political issues (Bond et al. 2012). On the other hand, as noted above, when the information is false, it can lead not only to negative shocks to financial markets but also empower dangerous terrorist and criminal activities (Ferrara et al 2015). The sensitivity of the stock market to information has increased considerably in recent years (Dugast and Foucault 2014). This is evidenced by the increased frequency of mini-flash crashes (www.nanex.net/FlashCrashEquities/), which are followed by reversions to the mean, as investors disconfirm the veracity of information.

At the same time, positive news engenders overreactions from investors. For instance, tweets that imbue positive sentiments are associated with abnormal stock returns and increased trading volumes (Sprenger and Welpe 2010). These effects have been attributed to tweets by influential investors, whose messages get amplified within the investor community. These tweets are accorded more importance and treated as reliable signals due to positive associations with past stock movements. However, as the cost of trading declines, investors are more likely to seek arbitrage opportunities based upon unreliable information (Dugast and Foucault 2014). This stimulates quicker corrections resulting in more efficient markets – accompanied by an increase in volatility.

Twitter's importance in our information ecosystem has been steadily growing due to its immediacy to breaking news. With over 500 million tweets per day and growing, Twitter is a valuable source of raw data, which can be analyzed to gain insights into user sentiment. Interpreting large Twitter datasets is now possible with advances in natural language processing (NLP) techniques and textual analytics algorithms. NLP and textual analytics have been used for a wide variety of sentiment analyses and evaluation of information content, including the classification of inaccuracy in news (e.g., Rubin, V. L., Chen, Y., & Conroy, N. J., 2015).

Experimental evidence using Twitter bots has shown that information is more likely to influence behaviors in social media networks through a *complex contagion* process (Mønsted, Sapieżyński, Ferrara, and Lehmann 2017). The probability that information changes behaviors in a complex contagion process (Centola and Macy 2007) depends upon: 1) the *number of information sources*, and 2) *thresholds of accumulated influence* within the social media network. Specifically, the likelihood of Twitter users retweeting or rephrasing and propagating information is not independent of message reinforcement from multiple sources and external influence (Monsted et al 2017). There is therefore a strong need to better understand how information characteristics (sentiment, curation, credibility) influence investor behaviors. To this end, we use the lens of *dominance theory* to glean new insights into *influential* and *influenced* investors.

For the purposes of the present study, we consider dominance as a personality trait. In this approach, dominance is treated as an inherent human propensity that leads to certain behavioral patterns including specific behaviors associated with electronic communications. We conducted an explorative depth analysis of over 5000 tweets to find patterns or associations between tweet characteristics and stock price and volume fluctuations. To the best of our knowledge, no other study to date has specifically studied tweet characteristics using the lens of dominance theory. Therefore, the findings of our research based on a unique approach to analyzing tweets associated with steep fluctuations using textual analytics applying insights from dominance theory, are expected to serve as a valuable contribution to understanding information behavior in social media networks.

LITERATURE REVIEW

The present literature review provides a summary of our review of past research that has informed our logic and is structured in four parts: Twitter usage and implications, information and stock markets, textual analytics, and dominance theory. With around 330 million users and 500 million tweets a day, tweets are a social phenomenon that are used extensively by individuals, corporations, organizations, politicians, investors, artists, sportspersons and a wide variety of societal leaders (Fuchs 2017). Twitter's far reaching influence is derived from its global network and strong presence of domain leaders, influencers and office bearers (Dubois, and Gaffney 2014; Weller, Bruns, , Burgess, , Mahrt, and Puschmann 2014). Twitter has also been used by eminent personalities in sports, entertainment and media to engage audiences through opinions and near real time updates on events. Additionally, Twitter offers researchers the opportunity to analyze influence on the basis of activity, popularity, and degree of influence (Riquelme and González-Cantergiani, 2016).

Past studies have found that Tweet patterns and volumes before, during, and after events are significantly associated with movements in stock returns and volumes (Sprenger, Tumasjan, Sandner, & Welpe 2014). While demonstrating the causal effects of Tweets and other social media remains a challenge, the association of Tweets with stock price movements is still very valuable. Specifically, associative indicators (based upon influence metrics) that signal potential movements in stock prices and volumes can stimulate new trades. Incorporating multiple and diverse sources of information into an analysis can be challenging. However, it has been shown to improve trading forecasts (Curme, Zhuo, Moat, and Preis, 2017). Spikes in tweets about a particular stock have been associated with changes in volatility, implied volatility and option prices (Wei, Mao, and Wang 2016).

As the world's leading microblogging platform, Twitter is highly representative of information flows on social media networks. Additionally, due to the availability of high quality data, Twitter has become a very useful source for researchers seeking to understand the topology of investor information networks. Much of social media analytics is driven by the analysis of text and unstructured data. Social media communications provide rich unstructured data in vast quantities. Such data contain valuable information which can provide domain insights and enable decision makers to improve the accuracy of their forecasts, disconfirm their understanding of sentiment, and increase the quality of their decisions. However, this requires the use of advanced analytical techniques including natural language processing (NLP) and textual analytics methods to analyze the vast troves of data and discover insights. Past research has demonstrated the importance of understanding the sentiment associated with social media messages and how they interact with news events. Sentiment analysis using textual analytics and NLP techniques can provide insights into the emotional states ('happy' or 'sad' or 'disinterested' or 'positive' or 'negative') and can be filtered by multiple factors such as data, topic, or noun (Tsytsarau, M., Palpanas, T., and Castellanos, M., 2014). Decisions derived from textual analysis based models have been shown to outperform models based on financial information alone (Lee, Surdeanu, MacCartney, and Jurafsky 2014 May). This underscores the importance of developing a deeper understanding of the information contained in social media messages.

The complexity of the social media environment has been increasing due to 'bots' that lead to automated text generation and software driven information propagation. Automation methods have also been used for textual analytics of Tweets to identify trends and the impact of events (Alsaedi, Burnap, and Rana 2016). For instance, automated classification of Tweet textual features such as Twitter tags and emoticons ('smileys') was used to automatically identify Tweet sentiment (Davidov, Tsur, and Rappoport 2010). Such automated and enhanced sentiment analytics are an effective means of identifying self-declared emotions and beliefs contained in Tweets. Such automation also allows for near real-time analysis and quick response, along with vast savings of manual labor and effort. Advanced textual analysis using semantic features and deep neural networks can be used to automatically gauge information content from news text (Chang, Zhang, Teng, Bozanic, and Ke 2016). Research using Thomson Reuters proprietary neural network has shown that predictions based upon textual analysis of news is more effective when news is aggregated over longer periods (Heston and Sinha 2017). Past research clearly indicates the critical importance of using textual analytics to better understand information content and individual behavior.

We introduce a novel basis for understanding Tweets by classifying dominance behavior in Tweets through textual analytics. Dominance theory posits that dominance seeking entities exhibit dominance behavior, with the intent of establishing superiority for the purpose of gaining precedence in perceived physical, material or psychological benefits (Maslow, A., 1937). Dominance has been observed across individuals, groups and organizations (Burgoon, J. K. and Dunbar, N. E., 2006). Various facets of dominance with an emphasis on inherited dominance or system based dominance is contrasted to earned dominance based on intrinsic characteristics and capabilities, and 'interactionist' dominance (Burgoon, J. K., and Dunbar, N. E., 2006). Our primary interest is in dominance behavior expressed towards accomplishing the goal of propagating an information as one or more Tweets. Electronic communication has increased rapidly over the past few decades resulting in information complexity and overload. Face to face communication entailed processing body language, tone and demeanor, in addition to objective words. These characteristics have been replaced by digital features in computer mediated communication, which can be identified using textual analytics and NLP methods to study dominance seeking behavior (Samuel, J., et. al., 2014). Relatively stronger use of Twitter tags, inordinate use of upper case letters, repetition of works, emoticons, alphabets within words, whole or part sentences and special characters are all examples of how dominance behavior is expressed through social media communications. Though much past research on dominance has been focus on anthropological and sociological dimensions, there is a clear need to use dominance concepts to better understand Tweets and Tweets behavior, given the increasing prevalence of social media in society today (Ridgeway, Berger and Smith 1985; Burgoon and Dunbar 2000; Samuel et. al. 2014).

MODEL

FIGURE 1
GENERAL ANALYTICAL FRAMEWORK FOR TWEETS AND STOCK

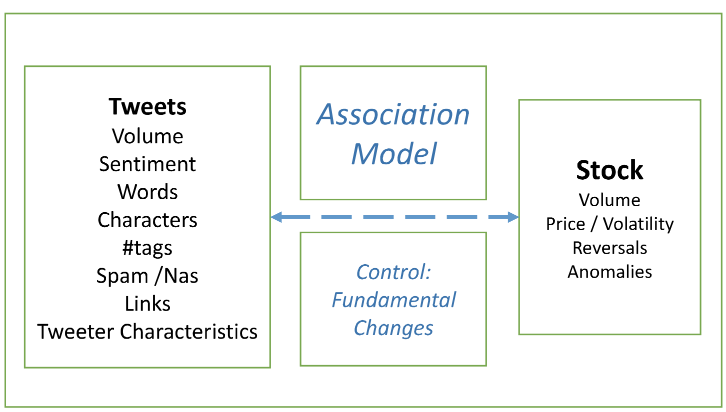


Figure 1 represents a general analytical framework for analyzing Tweets and their association with variation in stock metrics include volume, sentiment, word counts, use of special character and '#' tags, links, multimedia, character repetition and emoticons. Within the scope of dominance theory, our focus is on extraordinary patterns in tweets and tweet behavior which can be associated with dominance seeking behavior. We are therefore particularly interested in repetition of characters, use of forceful words such as 'must', volume of tweets and volume of words, characters and emoticons per tweet and such other electronic dominance seeking behavior (Samuel, J., et. al., 2014).

METHOD

Data Collection: We prescreened over 100 stock based on their Beta, Market Capitalization, price-earnings ratio (PE) and brand value using stock screening tools and stock market data providers including finviz.com, Bloomberg, Yahoo Finance and news articles. We then downloaded 2017 annual data for these stock from Yahoo Finance and shortlisted stock which we found interesting based on relatively steep fluctuations in prices or volumes or both, or relative stability for comparative purposes. Twitter Data - downloaded tweets dataset: We then wrote code in R to collect data from Twitter, filtering tweets which mentioned the selected stock. We ran the code iteratively to collect tweets associated with the day before the fluctuation, the day of the fluctuation and the day after the fluctuation. We then cleaned the data to weed out spam posts with links, posts with multiple stock symbols or names and other row items which were not applicable.

TABLE 1
ADSK

Stock	Date	Tweets	Open	High	Low	Close	Adj Close	Volume
ADSK	1/16/2018	95	116.24	117.08	111.72	111.98	111.98	2682700
ADSK	1/15/2018	70	MLK Holiday					
ADSK	1/14/2018	52	Sunday					
ADSK	1/13/2018	45	Saturday					
ADSK	1/12/2018	77	113.57	116.09	113.07	115.91	115.91	2553000
ADSK	1/11/2018	80	111.67	113.96	111.47	113.26	113.26	2716000
ADSK	1/10/2018	70	111.71	112.61	110.55	111.47	111.47	1645000
ADSK	1/9/2018	61	111.57	112.31	110.65	112.11	112.11	1848900
ADSK	1/8/2018	41	110.42	111.74	109.04	111.42	111.42	1782100

Understand Data, Study Data/textual Characteristics: we then used textual analytics –including standard packages in R software and our own proprietary textual analytics frameworks to study the Tweets and understand the data. We also performed a manual verification of the results of the "SentimentR" (SentimentR, 2018) package in R software as NLP and textual analytics tools can be contextually subjective and there is a need to validate their performance on a case to case basis.

We found reasonable validity in the ability of SentimentR to evaluate positive and negative sentiment in Tweets. We also explored the semantics associated with the Tweets using a random selection so as to have a face value understanding of the nature of Tweets related to the stock – this served as a preliminary manual textual analytics mechanism which enabled us to gain familiarity with nature of Tweets associated with the stock.

TABLE 2
INTC

Stock	Date	Tweets	Open	High	Low	Close	Adj Close	Volume
INTC	1/16/2018	330	43.55	43.79	42.89	43.14	43.14	37892700
INTC	1/15/2018	243	MLK Holiday					
INTC	1/14/2018	165	Sunday					
INTC	1/13/2018	256	Saturday					
INTC	1/12/2018	585	43.45	43.6	43.01	43.24	43.24	29973600
INTC	1/11/2018	582	42.8	43.58	42.45	43.41	43.41	35371500
INTC	1/10/2018	606	43.33	43.6	42.44	42.5	42.5	45735000
INTC	1/9/2018	800	44.7	44.84	43.49	43.62	43.62	44282300
INTC	1/8/2018	528	44.27	44.84	43.96	44.74	44.74	33733800

TABLE 3
WMT

Stock	Date	Tweets	Open	High	Low	Close	Adj Close	Volume
WMT	1/16/2018	461	101.34	101.91	100.34	100.69	100.69	6533800
WMT	1/15/2018	329	MLK Holiday					
WMT	1/14/2018	384	Sunday					
WMT	1/13/2018	299	Saturday					
WMT	1/12/2018	908	100.39	101.44	100.3	100.87	100.87	6951200
WMT	1/11/2018	1611	99.7	100.45	98.78	100.02	100.02	6537700
WMT	1/10/2018	357	99.75	99.89	98.53	99.67	99.67	7930400
WMT	1/9/2018	365	102.03	102.35	100.31	100.39	100.39	7312700
WMT	1/8/2018	346	100.25	101.75	100.21	101.61	101.61	8843900

For the purposes of the present white paper we have limited our analysis to exploring Tweets for four stock through early January of 2018, using R and R-Studio, along with relevant R packages, primarily “TwitteR”. We chose ‘Autodesk, Inc.’ (ADSK) stock and downloaded 628 tweets from 7th January to 16th January and combined the summary of the Twitter dataset with ADSK financial data from Yahoo Finance. Similarly we also downloaded 5328 tweets for Walmart (WMT), 817 tweets for Eastman Kodak Company (KODK) and 4470 tweets for Intel (INTC) and combined the Tweet volumes with financial metrics in tables 1-4.

TABLE 4
KODK

Stock	Date	Tweets	Open	High	Low	Close	Adj Close	Volume
KODK	1/16/2018	408	8.85	9.1	8	8.5	8.5	8524000
KODK	1/15/2018	118	MLK Holiday					
KODK	1/14/2018	139	Sunday					
KODK	1/13/2018	130	Saturday					
KODK	1/12/2018	0	8.45	9.45	7.95	9.2	9.2	13264300
KODK	1/11/2018	0	9.7	9.9	8.05	8.4	8.4	23936500
KODK	1/10/2018	0	12.5	13.28	10.1	10.7	10.7	1.08E+08
KODK	1/9/2018	0	3.1	7.65	3.05	6.8	6.8	71427400
KODK	1/8/2018	0	3.15	3.2	3.05	3.1	3.1	255700

We then collected the text of the tweets for each of the stock and performed Sentiment Analysis using the SentimentR tool and classified the Tweets into positive, negative neutral and negative sentiment. Given the explorative approach taken, we chose to focus on the positive and negative sentiment, as a vast majority of the Tweets which were classified as neutral (0 score) were a result of links and textual processing characters being treated as independent tweets, and hence such scores were removed from our plots and summaries.

FIGURE 2
KODK TWEET SENTIMENT DISTRIBUTION

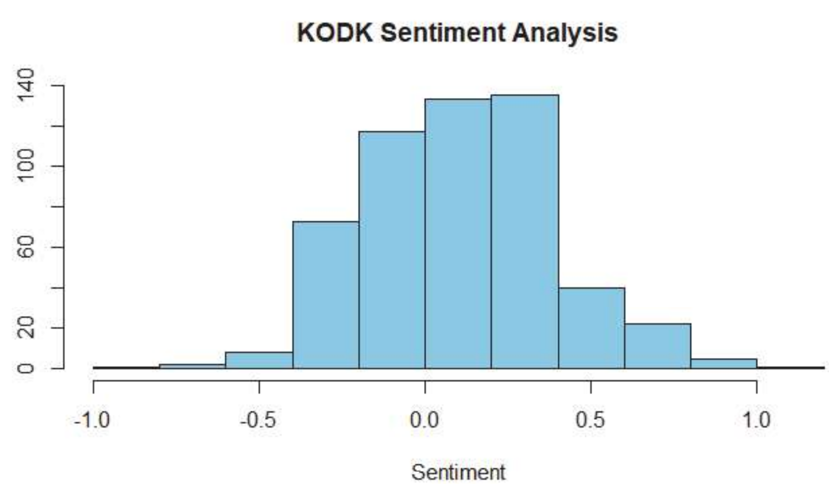
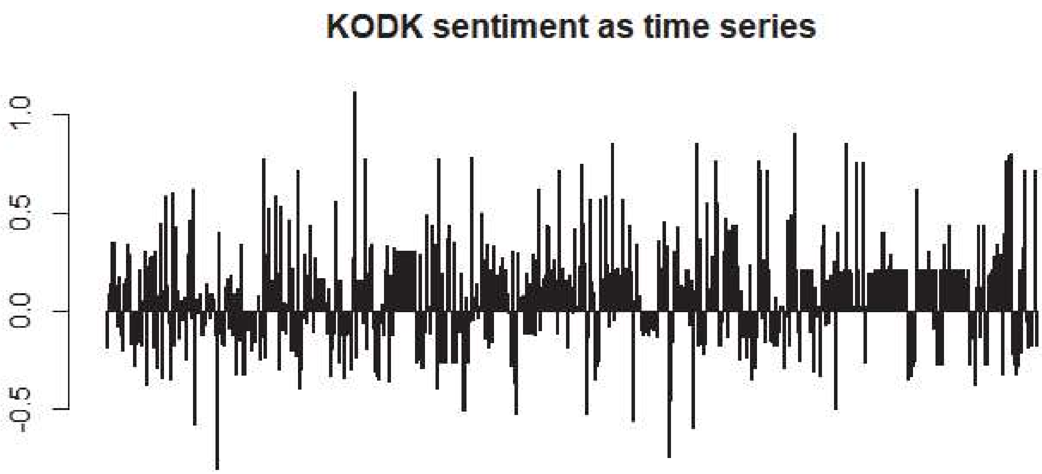


FIGURE 3
TIME-SEQUENCED BARPLOT OF SENTIMENT SCORES FOR KODK TWEETS



The sentiment analysis method assigns a sentiment score – in this case along a continuous scale of positive sentiment (positive values) to negative sentiment (negative values) and the scores are unbounded. However, the unboundedness can create outlier effects and so for the present study we boundary sentiment scores between -2.5 & 2.5. For the period considered of the second week of January 2018, the histogram (Figure 2) for sentiment analysis of KODK shows a slightly positive trend, but we do not see extremely positive sentiment. Similarly the time series bar plot shows an overall positive sentiment(Figure 3), with a positive push towards the end.

FIGURE 4
INTC TWEET SENTIMENT DISTRIBUTION

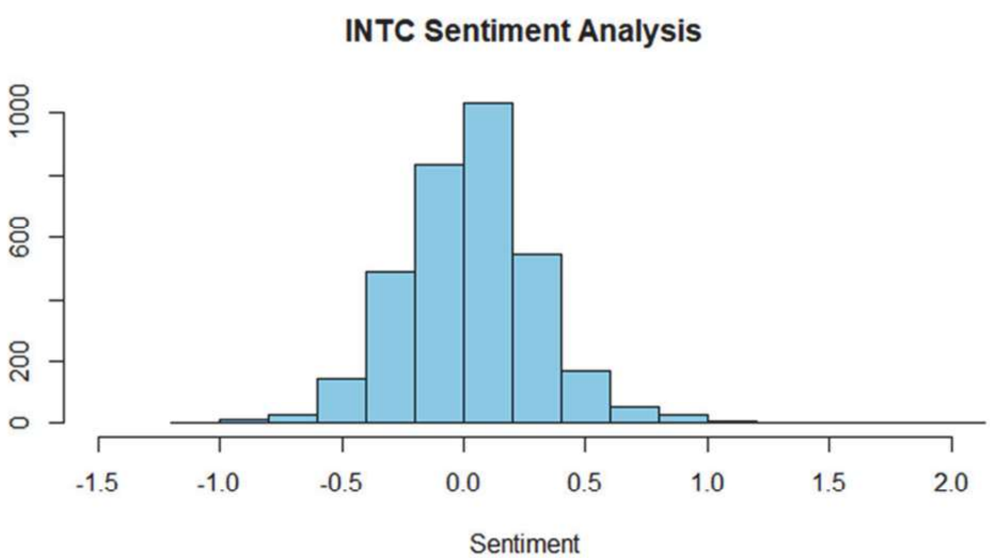
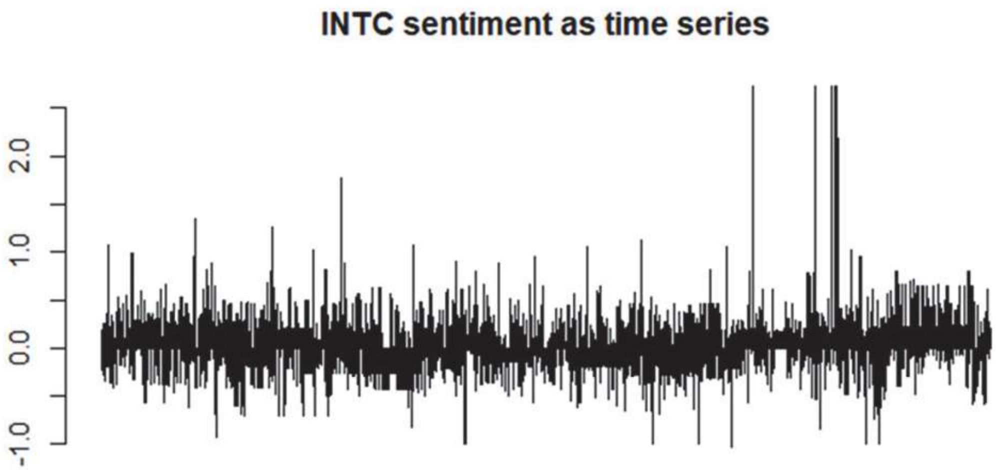


FIGURE 5
TIME-SEQUENCED BARPLOT OF SENTIMENT SCORES FOR INTC TWEETS



EXPLORATORY ANALYSIS

The value used for representing sentiment evaluation is called the 'polarity score' and this value depends on the polarity dictionary used. We use the SentimentR package which use the Jockers (2017) polarity dictionary with adaptation. Another good option that has been extensively used is Hu, M., & Liu, B. (2004) – the choice of an appropriate polarity benchmark depends on a number of factors, including the subject matter being analyzed. Very often, the sentiment analysis is repeated with multiple polarity references to identify discrepancies, validation and additional insights. For the period considered of the second week of January 2018, the histogram (Figure 4) for sentiment analysis of INTC shows an almost flat trend, but we do see it skewed with a positive sentiment. Also, the time series bar plot shows a flat sentiment trend, with positive spikes towards the end (Figure 5).

FIGURE 6
WMT TWEET SENTIMENT DISTRIBUTION

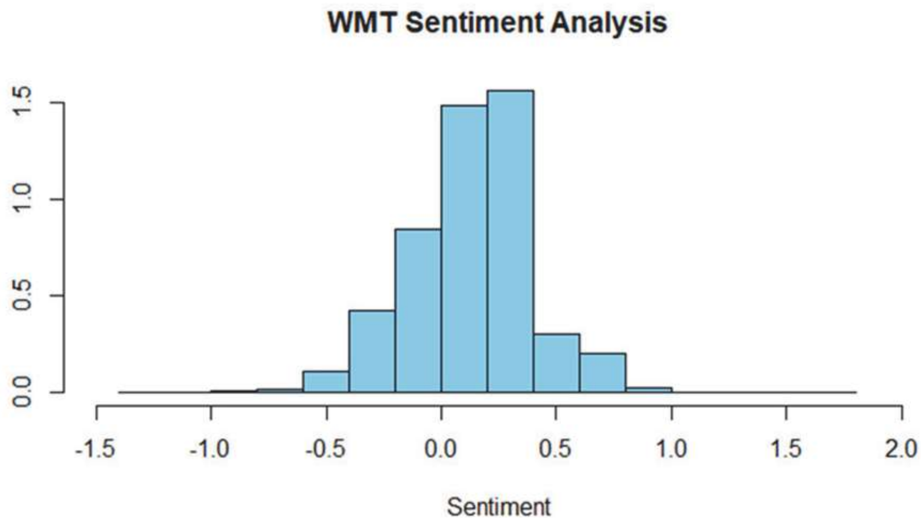
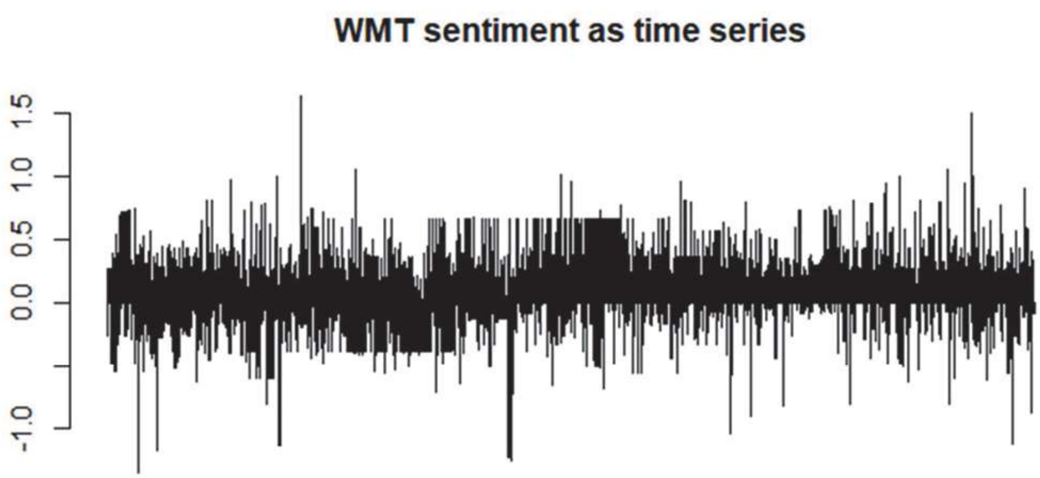


FIGURE 7
TIME-SEQUENCED BARPLOT OF SENTIMENT SCORES FOR WMT TWEETS



There was a strong interest in Walmart on Twitter in terms of Tweet volumes but that did not necessarily translate into a clearly polarized sentiment on unusual price movements in the stock. For the period of the second week of January 2018, the histogram (Figure 6) for sentiment analysis of WMT, like INTC, shows a near flat trend, but it shows a stronger positive sentiment than INTC. Interestingly, the time series bar plot shows more variance in the early period and a flat sentiment trend, with a positive trend in the second half of the period(Figure 7).

FIGURE 8
ADSK TWEET SENTIMENT DISTRIBUTION

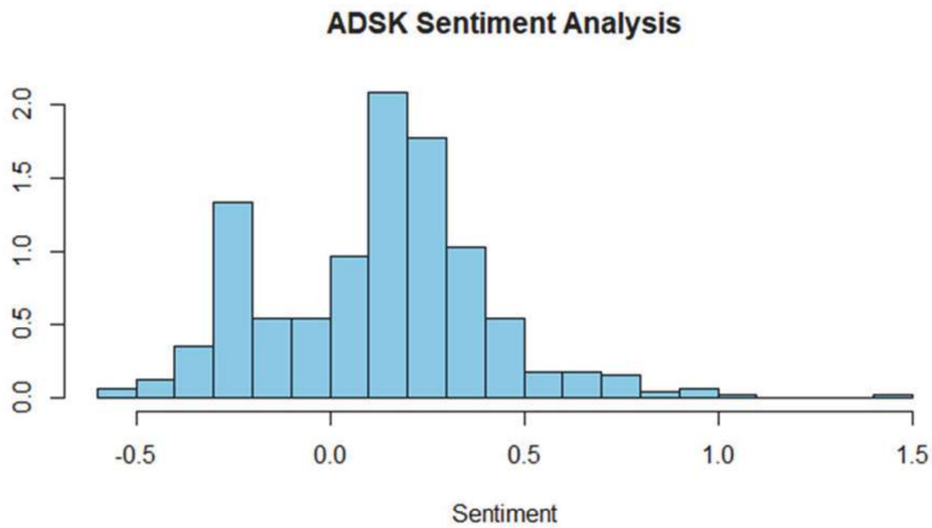
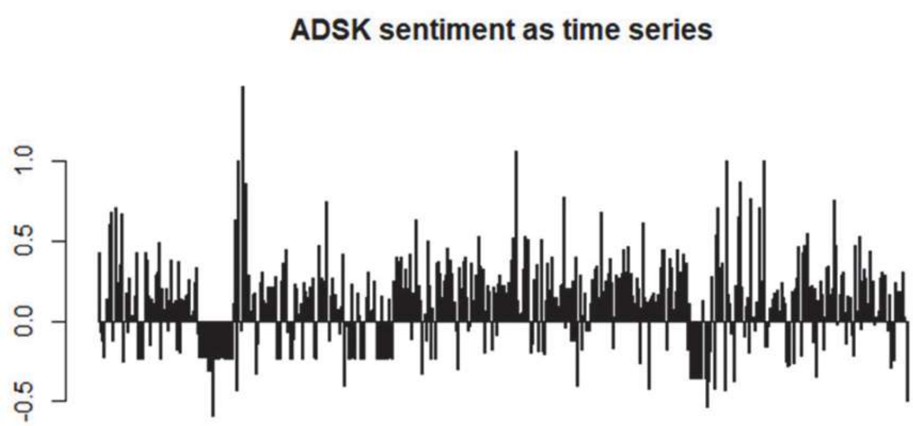


FIGURE 9
TIME-SEQUENCED BARPLOT OF SENTIMENT SCORES FOR ADSK TWEETS



ADSK histogram shows the presence of some strong negative sentiment along with cumulatively positive sentiment, indicating mixed views and positions in ADSK tweets. The time series bar plot (Figure 9) confirms this, demonstrating string spikes in the early half of the period.

From an exploratory analysis, for all the four companies, KODK, INTC, WMT & ADSK, we see a reasonable alignment between their respective tweets sentiment trends and their price movements during the period under consideration. We also see a pattern of certain Twitter users creating or retweeting relatively higher volumes of tweets as an expression of a form of dominance behavior. Overall, based on Samuel, Holowczak, Benbunan-Fich, & Levine (2014) we observe the following patterns of dominance behavior in the tweets we analyzed:

- Repetition of tweets by the same Twitter user
- Use of upper case words and sentences
- Continued use of relatively longer tweets and higher number of words

- Use of strong and emphatic language
- Counter posts with aggression, indicating submission seeking behavior.

Our analysis is encouraging as it provides early support to the stream of research we have initiated involving the study of social media posts – the findings validate our expectation of tweets sentiment being associated to stock metrics and the presence of dominance behavior in tweets. However, we bear in mind that extensive additional empirical analysis would be required for validation of the exploratory findings.

DISCUSSION

This explorative study is strong in the novel theoretical framework it uses to address microblogging behavior as expressed in tweets by Twitter users. The study provides an insightful approach to analyzing tweets and the use of textual analytics.

The study also has numerous weaknesses which we plan to address as we develop the stream of research. Twitter data obtained through APIs have a limited time range availability, in most cases from around 7 days prior to the date of query. This limited the study from seeking the most volatile periods of 2017, which was the initial intent. We also did not, deliberately, weed out some of the tweets which could have been spam posts and double posts. While other research strategies may omit such data, we chose to retain it, as such data is reflective of the real we live in today. Very often, unaided users are unable to distinguish between actual humans posting tweets and bots posting tweets and dealing with such data strengthens the external validity of the findings, given the explorative nature of the findings. There is also a need to conduct in-depth textual analytics using manual and automated processes to better identify dominance behavior expressed in tweets (Samuel, J., et. al., 2014).

The uniqueness of this study is primarily based on the novel approach to framing the analysis from an information systems perspective using dominance theory. An improved understanding of tweet sentiment and patterns will from information perception, information science and dominance. Our continued and future research is expected to provide fresh and unique insights into the use of words, phrases and character patterns which have not yet been identified. The identification of unique features will be actionable in stock investment decision making. Even the smallest of cues and improvement of accuracy in evaluating Tweets can lead to significant advantages including risk reduction.

CONCLUSION

Practitioners have sought to understand how tweets propagate – the present research addresses this dimension theoretically from a dominance theory perspective. Future research will involve the development of propositions for theoretical modeling to conceptually validate the exploratory and inductive findings of the present study. The present study serves as a beginning and we see it as the initiation of a stream of research into the textual behavior of social media posts, under the lens of dominance theory. Thus, the research possibilities that this initiation creates will be a valuable contribution to researchers and those interested in studying social media from a textual analytics and information behavior perspectives. The exploratory insight from the present study will also support future research in behavioral machine learning which focuses on explaining and classifying information artifact driven performance (Samuel, J., 2017).

Furthermore, the present research has also identified the presence of dominance behavior in tweets – this has interesting implications for researchers and practitioners as it opens up avenues for identifying micro-influencers in the various virtual groups and communities. Dominance behavior can have positive or negative consequences subject to context, subject matter and audience. An improved understanding of dominance behavior can be very useful for targeted policy creation to serve at least two objectives – create guidelines for what kind of textual expressions would constitute positive dominance and this enhance the likelihood of desirable results and similarly policy guidelines can also be developed to identify, manage and mitigate the undesirable results of negative dominance behavior expressed through tweets and social media posts. It is also possible to extend this research to create automated dominance identification mechanisms. It is our intention to further pursue this stream to research and develop rigorous and actionable frameworks to meet the aforesaid goals.

DISCLOSURE

This paper is a significantly updated and expanded version of Samuel, J., Kashyap, R., & Kretinin, A., (2018). ‚Going Where The Tweets Get Moving! An Explorative Analysis Of Tweets Sentiments In The Stock Market. 45th NBEA Annual conference proceedings, with overlapping content.

REFERENCES

Alsaedi, N., Burnap, P., & Rana, O. F. (2016). Automatic summarization of real world events using twitter.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A Model of Investor Sentiment. *Journal of Financial Economics,* 49(3), 307-343.

Bond, R. M., et al. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature,* 489(7415), 295–298.

Bovet, A., Morone, F., & Makse, H. A. (2016). Validation of Twitter opinion trends with national polling aggregates: Hillary Clinton vs Donald Trump. *arXiv preprint arXiv:1610.01587*.

Burgoon, J. K., & Dunbar, N. E. (n.d.). Nonverbal Expressions of Dominance and Power in Human Relationships. In V. Manusov and M. L. Patterson (Eds.), The Sage Handbook of Nonverbal Communication. Thousand Oaks, CA: Sage, 2006, pp. 279-297.

Burgoon, J. K., & Dunbar, N. E. (2000). An interactionist perspective on dominance submission: Interpersonal dominance as a dynamic, situationally contingent social skill. *Communication Monographs*, 67, 96-121.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American journal of Sociology,* 113, 702–734.

Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060), 1269–1272.

Chang, C. Y., Zhang, Y., Teng, Z., Bozanic, Z., & Ke, B. (2016). Measuring the Information Content of Financial News. In *COLING* (pp. 3216-3225).

Conover, M. D., Davis, C., Ferrara, E., Mckelvey, K., Menczer, F., & Flammini, A. (2013). The geospatial characteristics of a social movement communication network. *PloS one,* 8(3), e55957.

Curme, C., Zhuo, Y. D., Moat, H. S., & Preis, T. (2017). Quantifying the diversity of news around stock market moves.

Davidov, D., Tsur, O., & Rappoport, A. (2010, August). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd international conference on computational linguistics: posters* (pp. 241-249). Association for Computational Linguistics.

Dillard, J. P., Solomon, D. H., & Samp, J. A. (1996). Framing social reality: The relevance of relational judgments. *Communication Research*, 23, 703–723.

Dubois, E., & Gaffney, D. (2014). The multiple facets of influence: identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist*, 58(10), 1260-1277.

Ferrara, E. (2015). Manipulation and abuse on social media by emilio ferrara with ching-man au yeung as coordinator. ACM SIGWEB Newsletter, (Spring), 4.

Fuchs, C. (2017). *Social media: A critical introduction*. Sage.

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. National Conference on Artificial In telligence.

Jockers, M. L. (2017). Syuzhet: Extract sentiment and plot arcs from text. Retrieved from https://github.com/mjockers/syuzhet

Heston, S. L., & Sinha, N. R. (2017). News vs. Sentiment: Predicting Stock Returns from News Stories. *Financial Analysts Journal*, 73(3), 67-83. Retrieved from https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf

Lee, H., Surdeanu, M., MacCartney, B., & Jurafsky, D. (2014, May). On the Importance of Text Analysis for Stock Price Prediction. In *LREC* (pp. 1170-1175).

Maslow, A. (n.d.). Dominance-feeling, Behavior and Status. *Psychological Review*, 44(1937), 404-429.

Mønsted, B., Sapieżyński, P., Ferrara, E., & Lehmann, S. (2017). Evidence of Complex Contagion of Information in Social Media: An Experiment Using Twitter Bots. arXiv preprint arXiv:1703.06027.

Ridgeway, C. L., Berger, J., & Smith, L. (1985). Nonverbal cues and status: An expectation states approach. *American Journal of Sociology*, 90(5), 955-978.

Riquelme, F., & González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information processing & management*, 52(5), 949-975.

Rubin, V. L., Chen, Y., & Conroy, N. J. (2015). Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1), 1-4.

SentimentR - https://cran.r-roject.org/web/packages/sentimentr/sentimentr.pdf https://www.npr.org/sections/thetwo-way/2017/12/26/573609499/library-of-congress-will-no-longer-archive-every-tweet

http://business.time.com/2013/04/24/how-does-one-fake-tweet-cause-a-stock-market-crash/ http://www.nanex.net/FlashCrashEquities/FlashCrashAnalysisEquities.html

Sakaki, T., Okazaki, M., And Matsuo, Y. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In 19th International Conference on World Wide Web. ACM, 851–860.Wei, W., Mao, Y., & Wang, B. (2016). Twitter volume spikes and stock options pricing. *Computer Communications*, 73, 271-281.

Samuel, J., Holowczak, R., Benbunan-Fich, R., & Levine, I. (2014, January). Automating Discovery of Dominance in Synchronous Computer-Mediated Communication. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 1804-1812). IEEE.

Samuel, J. (2017). Information Token Driven Machine Learning For Electronic Markets: Performance Effects In Behavioral Financial Big Data Analytics. *Journal of Information Systems and Technology Management*, 14(3).

Samuel, J., Kashyap, R., & Kretinin, A., (2018). Going Where The Tweets Get Moving! An Explorative Analysis Of Tweets Sentiments In The Stock Market. 45th NBEA Annual conference proceedings.

SIGWEB Newsletter Spring 2015, The Curious Case of Cynk, an Abandoned Tech Company Now Worth $5 Billion — mashable.com/2014/

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5), 926-957.

Tsytsarau, M., Palpanas, T., & Castellanos, M. (2014, August). Dynamics of news events and social media reaction. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 901-910). ACM.

Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (2014). Twitter and society (Vol. 89, p. 447). P. Lang.