

You are What You Say: The Influence of Company Tweets on Its Stock Performance

Babajide Osatuyi
The Pennsylvania State University
bosatuyi@psu.edu

Behrooz Yoosefi
The Pennsylvania State University
bxy5057@psu.edu

Abstract

This paper investigates the relationship between Twitter metrics and stock price performance of a company. The objective of this research is to contribute to the area of research that seeks to uncover the business value of social media platforms. Building on prior research, this paper identifies two categories of metrics that have been used to examine the relationship between Twitter metrics and stock performance of a company, namely traffic and motivation. While traffic is measured as volume of tweets, motivation is measured from two perspectives; polarity (positive, neutral, and negative) and emotion (positive emotion and negative emotion). Unstructured data from Twitter and Yahoo finance Website about Amazon was gathered to test the study hypothesis. A combination of machine learning techniques for text analytics and hierarchical regression analysis was used to analyze the data. Results indicate that emotional motivation expressed in tweets sent out by a company positively influences the company's stock performance.

1. Introduction

Twitter has become one of the most influential microblogging sites used by people, companies and government entities. Although it is a platform to create situational awareness, individuals, advertisement agencies and businesses in general have ceased the opportunity to use it to draw significant attention to situations of interest in recent times and to improve business processes [1]. Consequently, researchers and businesses are continually interested in developing techniques and approaches to strategically use Twitter to achieve positive outcomes that have business benefits [2].

Findings on the business value of Twitter are mixed at best. While some find a link between Twitter metrics and performance of a company, others are inconclusive

about the link [3]. The focus of the studies in the literature above so far have been on investigating public tweets. This study deviates from the approach in prior studies by focusing on tweets generated by firms. This approach is informed by the *community-building* perspective of the guidelines proposed in prior research [2] on successful implementation of social media strategies for businesses. This research project seeks to contribute to the area of research that is building evidence on the utility of social media platforms for business benefits.

The rest of this paper is organized as follows. Section 2 presents a review of relevant literature and presents the study the hypothesis. Section 3 presents the methodology used to gather and merge data from several sources to conduct analyses. Section 4 describes the results followed by a discussion of the results (Section 5), and conclusions and future work in the final sections.

2. Literature Review

2.1. Stock Prices and Twitter Use

Prior research show that metrics from microblogging sites such as Twitter have better predictive value than traditional industry metrics [4] including the use of historical stock prices. Coyne et al., Coyne, Madiraju [5] developed three models to examine information exchanged on StockTwits, a social media site for investors, to understand and predict future individual stock prices of an organization. The first model found no correlation between postings on StockTwits and stock price. However, a significant relationship was found in the other two models that combine investors' metrics (number of likes, user's followers and how often the user is correct) with their posting sentiments to predict stock price. Over a period of 15 months, Ranco et al., [6] examined Twitter volume and sentiment (positive or negative) of tweets from 30

stock companies that form the Dow Jones Industrial Average index. They found sentiments to be correlated with stock prices during peak Twitter volume.

The literature on the relationship between Twitter use and stock prices revolve around an attempt to understand the motivation behind actors of interest and the amount of traffic on Twitter that may have stock price implications. Motivation assesses the polarity of tweets using basic sentiment analysis of messages (tweets) exchanged on Twitter, whereas traffic assesses the frequency of communication between users on Twitter about a company. Findings on the influence of motivation measured as tweet polarity (i.e., positive, neutral and negative) is contradictory. Some researchers find significant influence [4, 7] while others fail to see significant relationships with target variables of interest [3, 8].

2.2. Stock Prices and Emotions

Research shows that there is a link between publicly expressed emotions about a company and its stock performance [9]. Using Google's sentiment analysis tool, Google-Profile of Mood States (GPOMS) that measures mood across 6 dimensions (calm, alert, sure, vital, kind, and happy), Bollen et al., [10] found that calm mood was the only good predictor of stock prices. The moods measured by GPOMS are predominantly reflective of positive emotions and hence can bias findings on the influence of emotions and stock price performance. There is therefore a need to examine the influence of negative emotions on stock prices of an organization.

Li et al [11] conducted experiments with the effects of news and public moods on stock movements and found that pessimistic public mood had a significantly higher predictive power than optimistic mood. Zhou et al [12] examined the influence of emotions including anger, sadness, joy, disgust and fear on stock prices in China. The study found that joy (a positive emotion) predicted stock price movements better than all the other negative emotions combined.

These findings provide evidence of the predictive ability of emotions expressed on public platforms on the stock performance of a company. This study therefore extends the current research by examining how the expression of positive and negative emotions in tweets by companies affect their stock prices.

2.3. Research Model

People tend to express their feelings through words [13] and therefore to truly understand how an organization truly feels about the state of their

organization, there is need to go beyond simply tagging their conversation as positive, neutral or negative. A deeper understanding of the communication content from a psychological perspective can provide additional insights to produce tangible measures for assessment.

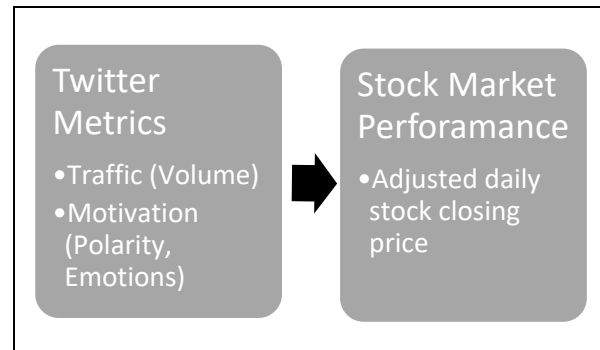


Figure 1. A Twitter Metrics Stock Price Performance Model

Emotions have been studied extensively in the psychology literature and have been found to convey lasting outcomes than mere feelings [14, 15]. The literature demarcates between polarity and emotional states of an individual and considers emotional states to be more stable. Polarity typically measures the predominant valence of a text, whereas classifying emotional states requires coding for both the meaning and context of the text. In the context of social media-mediated communications, emotional states of users expressed as moods in public communication have been found to affect stock prices [16]. Hence, it is expected that coding for the emotional state expressed in conversations about a company will have a stronger influence on stock price than a simple binary sentiment coding.

This paper seeks to extend the current approach to investigating motivational factors in online communication by examining its emotional dimension in addition to the polarity measures to evaluate the (1) individual and (2) combined contribution of those factors to explaining stock price changes.

Study Hypothesis. Therefore, we hypothesize that Twitter metrics related to emotions will have a stronger relationship with stock price market performance of a company than polarity metrics.

3. Method

3.1. Data Collection

Data for this study was gathered from multiple sources. Stock prices for Amazon were downloaded from [Yahoo's](#) finance Website. Tweets created by Amazon were downloaded using NodeXL, an excel plugin for a complete calendar year between June 2017 to May 2018 as it is the common practice in prior research. The data sources were combined in a database using the date as the common matching key. A total of 231 records, which translate to 231 trading days, constituted the final dataset for this study.

3.2. Research Variables

The predictor variables are metrics from Twitter including volume and sentiments, and the target outcome variable is the adjusted closing price from the stock market. These variables and how they are operationalized in this study are illustrated in Table 1.

The choice to use adjusted closing price of the stock rather than the closing price of the stock is that it considers dividends, stock splits and new stock offerings for the day to determine its value. Hence, the adjusted closing price is a more accurate measure of the value of a stock.

Table 1. Research variables and their operationalization.

Variable	Operationalization	Description	How Measured
Traffic	Volume	The total number of tweets per day	Sum of tweets per day
Motivation	Polarity	The polarity of tweets per day	Polarity measured as positive (1), neutral (0) and negative (-1).
Motivation	Emotions	The psychological processes contained in the tweets per day	Average number of positive/negative emotions present in a tweet per day
Stock market performance	Daily stock closing price	The adjusted stock price at the close of each day	Adjusted daily stock price

4. Analysis

Multiple analytical software was used to analyze the data for this study. First, a web crawler was used to gather unstructured data about Amazon from Twitter into a structured database of tweets organized by their date of creation.

Second, sentiment analysis, a machine learning technique to conduct qualitative analysis of textual information (tweets) in order to unravel some of the psychological processes inherent in the tweets. MeaningCloud software was used for the first set of sentiment analysis to identify basic polarity of tweets as either positive, neutral or negative. Furthermore, a second sentiment metric for emotions was computed using a software called Linguistic Inquiry and Word Count (LIWC) [17]. The development of the dictionary in LIWC is grounded in extensive, rigorous psychological and linguistic research that identify affect as positive and negative emotions rather than the basic polarity measurements. Hence, positive or negative emotions as measured by LIWC is the expression of how the message creator feels about the subject of discussion. The latest version of the dictionary in LIWC (i.e., LIWC2015) was used

for analyzing the data in this study. This dictionary contains about 6,400 words, word stems, and select number of emoticons that make it possible to recognize and code phrases, slang words and netspeak language among others. This feature makes the use of the latest LIWC dictionary suitable for analyzing social media texts. Specifically, positive emotion words can include words like love, ☺, sweet whereas negative emotion words can include ugly, ☹, and hurt. Three sub-categories of negative emotions including anxiety, anger and sadness are also coded to further explore negative emotions.

Third, this study employs a hierarchical regression modeling (HRM) approach [18] to investigate the relationship between Twitter metrics and a company's stock price performance. The reason for HRM is to unpack the respective contribution of variables in this study.

5. Results

The descriptive results provide some initial understanding of the research variables in this study and the extent to which they are useful for the analyses required to test the study hypotheses. First, Table 2

below, summarizes the descriptive results. It is interesting to note that over the study period, Amazon generated an average of 10 tweets per day. However, the maximum value of 351 tweets in one day from Amazon warrants attention. The plot of daily tweets from Amazon over the study period depicted in Figure 2 indicates that the maximum value of 351 is an anomaly with respect to tweets generated on other days. These tweets were generated on July 11th 2017. The depiction of word frequencies as a word cloud where larger fonts signify high frequencies is illustrated in Figure 3. The word cloud for the tweets on this date reveals that the tweets were about an annual one-day discount event for Amazon Prime members, tagged “Amazon Prime Day.” This tag therefore provides a ground truth for the posts sent out by Amazon on this date. Hence, after removing words that are not related to the central tag (i.e., Amazon Prime Day), about 69.32% of the tweets on that day included the word “prime day” and 28.98% of the

tweets contained “deals.” It follows then that the next most frequent word used in the tweets sent out by Amazon on that day was “deals.”

To gain additional insights into the regular dynamics of tweets outside the annual Amazon prime day event, data from the actual day of the event and the previous day were removed from the dataset leading to the distribution in Figure 4. This new distribution shows a seasonal pattern in the daily tweets where a quarter of high volume tweets is followed by one with low volume tweets. Furthermore, the number of positive emotion tweets are consistently higher than negative emotion tweets.

A shocking finding from the distribution of daily tweets is Amazon’s reaction to accusations from the President of the United States of America on tax issues on March 31, 2018. Only six tweets were generated by Amazon on that day and none of them addressed the accusations.

Table 2. Descriptive statistics of predictor variables				
Variables (per day)	Mean	Minimum	Maximum	Standard Deviation
Volume	10.85	0	351	24.16
Positive	7.10	0	176	12.33
Negative	0.71	0	16	1.51
Neutral	3.04	0	159	10.99
Average positive emotions	8.34	0.00	24.26	4.48
Average negative emotions	0.71	0.00	4.17	0.82
Average sadness	1.40	0.00	89.48	6.48
Average anger	1.31	0.00	23.81	3.36
Average anxiety	1.49	0.00	22.79	3.83
Adjusted closing price	1235.29	938.60	1629.62	236.51

The initial results of the HRM analysis with all the research variables is illustrated in Table 3. The model was built by entering variables into the model one level at a time and the regression algorithm automatically removed variables that are highly correlated. This process led to the automatic elimination of positive polarity tweets from the predictors. In addition, the variance inflation factor values for all but the motivation factors exceed the recommended threshold value of 5 [19]. An alternate model was estimated by replacing average negative emotions with its three sub-categorical variables (anxiety, anger and sadness) from the LIWC2015 dictionary. The estimates were similar to those in Table 3. Additional analysis is conducted in order to develop a valid model that explains how stock prices

of a company can be predicted by the content of its comments on Twitter.

A second HRM model (summarized in Table 4) was built using a stepwise algorithm that ensures that all the predictors entered into the model are significant and are not highly correlated. The results from the final HRM analysis summarized in Table 4 indicate that traffic factors (volume of tweets) does not significantly contribute to the prediction of a company’s daily stock prices. Motivation factors on the other hand are statistically significant predictors of a company’s stock prices. An R^2 change analysis was also reported to evaluate the contribution of adding each predictor into the model. More importantly, the significance of the change in R^2 provides support for the study hypothesis.

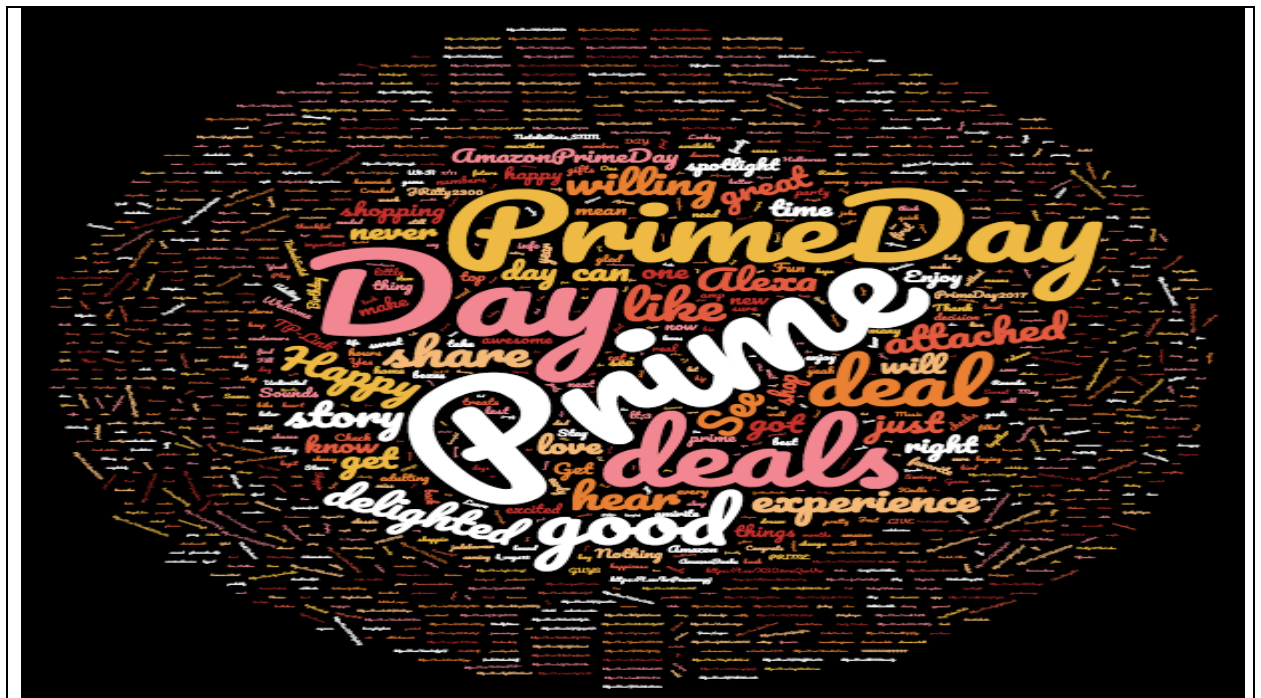
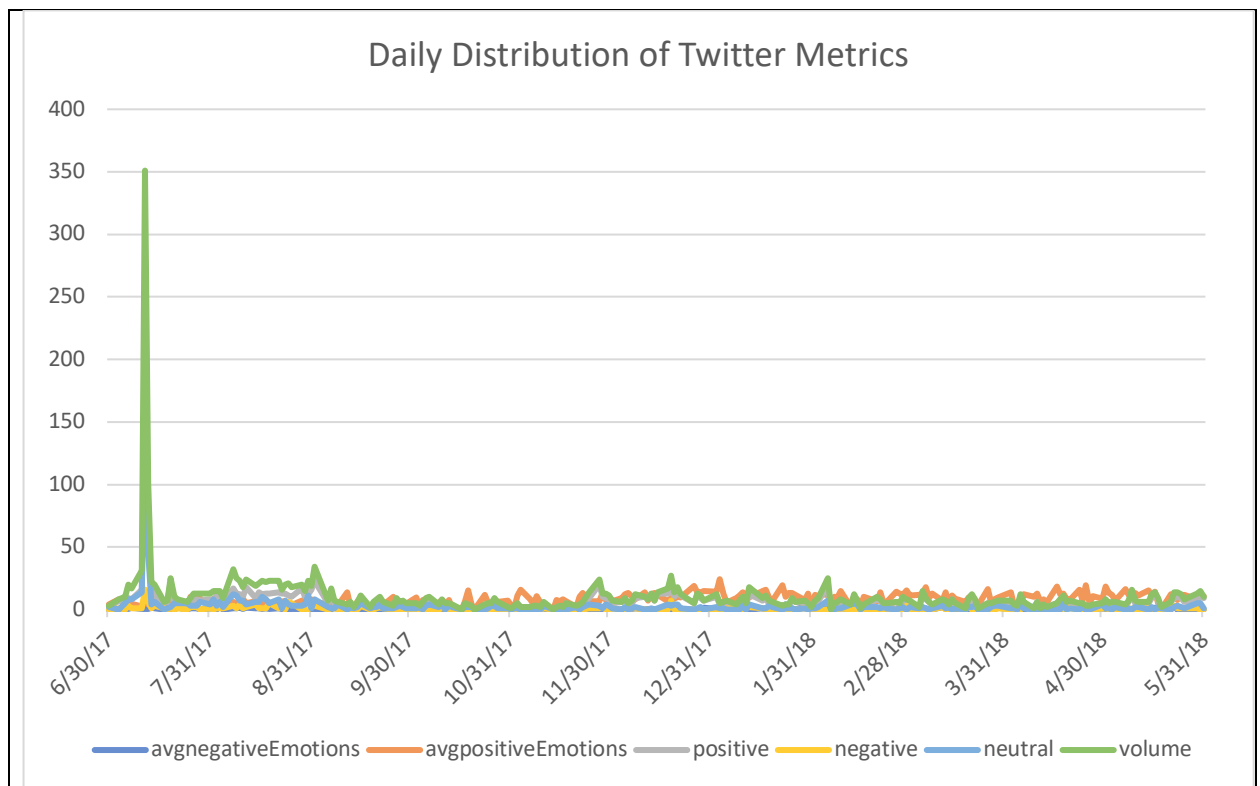


Figure 3. Word cloud for Tweets from Amazon on July 11 2017

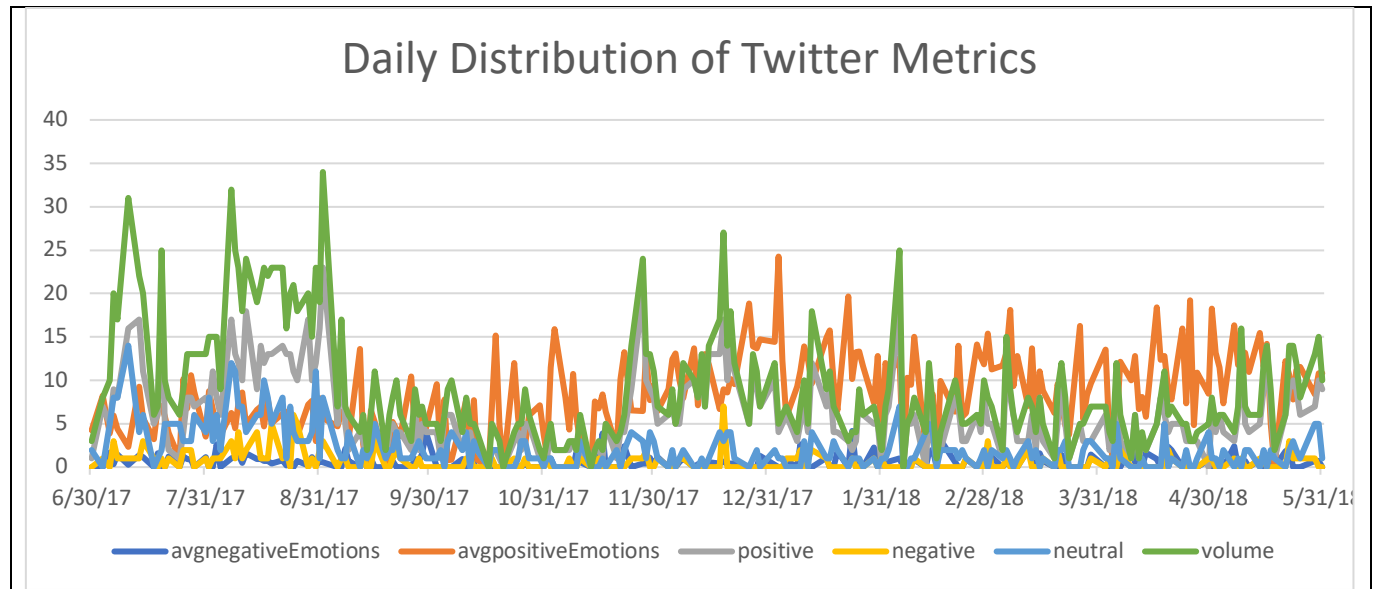


Figure 4. Daily Tweet Metrics Distribution without Outliers

Table 3. Hierarchical regression modeling results					
Model	Variables	Coefficient	Variance Inflation Factor (vif)	R ²	R ² Change
1	Intercept	1251.54***		0.019	0.020***
	Volume	-1.50*	1.000		
2	Intercept	1241.19***		0.037	0.047***
	Volume	4.89	39.333		
	Negative	-41.01*	2.776		
	Neutral	-9.79	33.649		
3	Intercept	1063.56***		0.206	0.000***
	Volume	-4.95	45.943		
	Negative	-18.10	2.987		
	Neutral	10.15	39.054		
	Negative Emotions	16.53	1.012		
	Positive Emotions	23.68***	1.176		
	N	231			

NOTE:***significant at 0.001; **significant at 0.01; *significant at 0.05

Table 4. Revised hierarchical regression modeling results					
Model	Variables	Coefficient	Variance Inflation Factor (vif)	R ²	R ² Change
1	Intercept	1258.38***	1.000	0.039	0.002***
	Negative	-32.52***			
2	Intercept	1069.98***	1.012	0.214	0.000***
	Negative	-25.30***			
	Positive Emotions	21.97***			
	N	231			

NOTE:***significant at 0.001; **significant at 0.01; *significant at 0.05

Overall, results from the modeling analyses indicate that tweet volume per day, negative polarity of tweets per day and tweets with positive emotions per day are significant predictors of Amazon's daily stock prices. However, a closer look at the results reveal the following insights. First, Model 1 in Table 3 includes only tweet volume in the model and the model is significant ($F=5.487$, $p=0.02$). The negative relationship between tweet volume and stock price performance suggests that a company might not need to say too much daily. Rather, communication may need to be fewer on microblogging platforms like Twitter. Based on recommendations for model fitness [19], Models 2 and 3 are inaccurate leading to the results in Table 4. The revised analysis indicates that negative tweets and positive emotions are the only variables that are significant predictors of the daily stock prices of a company. Model 2 in Table 4 shows that a unit increase in negative tweets lead to \$25.30 loss in the stock price of Amazon. However, a unit increase in positive emotions expressed in tweets by Amazon increases their stock price by \$21.97.

The explained variance of daily stock prices for the significant polarity factor (i.e., negative tweets) is only 3.9%, which is significantly lower (R-square difference=0.171, $p=0.000$) compared to when positive emotion tweets are included in the model leading to a combined explained variance of 21.4%. The low variance inflation factor values in the revised model indicate that the model satisfies the recommended model fit indices.

6. Discussion

This paper sought to investigate the relationship between Twitter metrics and a company's stock market performance. The results highlight some interesting trends that have implications for future research and management. Following the approach in prior research, this study categorized Twitter metrics into traffic (volume of daily tweets) and motivational coding of the tweets which is further grouped into polarity (negative, neutral and positive tweets) and emotions (positive emotions and negative emotions). Rather than examining correlations between Twitter metrics and stock price as commonly done in prior research, this study investigates the relationships using hierarchical regression model to understand the individual and combined percent change and contribution of the predictors on a company's stock price.

This study finds positive emotions, a motivational factor to be a stronger predictor of daily stock price of a company than tweet volume, a traffic factor. This result provides empirical support for the need for companies to engage in meaningful conversations that foster community-building [20], rather than posting generic information about services on social media platforms. An examination of the tweets generated by Amazon shows that the company is well aware of the need to engage with each customer more than applying a one-size-fits-all approach to their communication.

The negative relationship between tweet volume and stock price suggests that tweeting too much can be harmful to a company's economic performance. These results point to intentional strategic communication guideline that should be in place to build a healthy community with customers and other stakeholders of the company. For instance, Amazon's decision to only

focus on communicating positive emotions about their business rather than responding to accusations that may have led to the use of some harsh words indicate the company's intentional decision to focus on staying positive in their communication.

To address the study hypothesis, this study found that emotional factors are stronger predictors than polarity measures. Emotional factors explain more of the variance in stock prices of a company compared to simple polarity factors. This result highlights the value of using a sentiment classifier that seeks to unveil affect expressed in texts rather than the commonly used polarity classification.

6.1. Limitations and Future Research

It is useful to mention some limitations that may influence generalizability of the findings of this study. First, the analysis only considered data for one company rather than multiple companies. This approach was taken to fully investigate how social media metrics relate to a company's stock performance. Second, the data examined only one year since the stock market follows a seasonal trend. Third, with reference to the negative relationship between tweet volume and stock prices, future analysis will examine the optimal threshold for tweet volume to better manage stock price loss. Future studies can build on the limitations of this study by including additional companies in the analysis and examining stock performance beyond one calendar year. In addition, future studies seek to examine the effect of communication content of other entities including customers, competition and media corporations on a company's stock performance.

7. Conclusion

This paper reports some of the preliminary findings from a larger study design to understand how the use of social media platforms by companies can influence their economic performance. These initial results support the hypothesis that the content of information exchanged by companies have an influence on their stock performance. In other words, a company [and its performance] is what it says about itself. Research is underway to build upon these findings to examine how the cognitive, emotional, and structural components of the communication by companies on social media platforms influence their stock movement and performance.

8. References

- [1] S. Wigley and B.K. Lewis, *Rules of engagement: Practice what you tweet*. Public Relations Review, 2012. 38(1): pp. 165-167.
- [2] M.J. Culnan, P.J. McHugh, and J.I. Zubillaga, *How large US companies can use Twitter and other social media to gain business value*. MIS Quarterly Executive, 2010. 9(4): pp. 243-259.
- [3] E.J. Ruiz, et al. *Correlating financial time series with micro-blogging activity*. in *Proceedings of the fifth ACM international conference on Web search and data mining*. 2012. ACM.
- [4] L. Liu, et al., *A social-media-based approach to predicting stock comovement*. Expert Systems with Applications, 2015. 42(8): pp. 3893-3901.
- [5] S. Coyne, P. Madiraju, and J. Coelho. *Forecasting Stock Prices Using Social Media Analysis*. in *Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence & Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 2017 IEEE 15th Intl*. 2017. IEEE.
- [6] G. Ranco, et al., *The effects of Twitter sentiment on stock price returns*. PloS one, 2015. 10(9): pp. e0138441.
- [7] T.H. Nguyen, K. Shirai, and J. Velcin, *Sentiment analysis on social media for stock movement prediction*. Expert Systems with Applications, 2015. 42(24): pp. 9603-9611.
- [8] N. Oliveira, P. Cortez, and N. Areal. *On the predictability of stock market behavior using stocktwits sentiment and posting volume*. in *Portuguese Conference on Artificial Intelligence*. 2013. Springer, Berlin, Heidelberg.
- [9] M. Baker and J. Wurgler, *Investor sentiment in the stock market*. Journal of Economic Perspectives, 2007. 21(2): pp. 129-152.

- [10] J. Bollen, H. Mao, and X. Zeng, *Twitter mood predicts the stock market*. Journal of Computational Science, 2011. 2(1): pp. 1-8.
- [11] Q. Li, et al., *The effect of news and public mood on stock movements*. Information Sciences, 2014. 278: pp. 826-840.
- [12] Z. Zhou, J. Zhao, and K. Xu. *Can online emotions predict the stock market in China?* in *International Conference on Web Information Systems Engineering*. 2016. Springer.
- [13] M.D. Lieberman, et al., *Putting feelings into words*. Psychological Science, 2007. 18(5): pp. 421-428.
- [14] W.B. Cannon, *The James-Lange theory of emotions: A critical examination and an alternative theory*. The American Journal of Psychology, 1927. 39(1/4): pp. 106-124.
- [15] A. Moors, et al., *Appraisal theories of emotion: State of the art and future development*. Emotion Review, 2013. 5(2): pp. 119-124.
- [16] Q. Li, et al., *The effect of news and public mood on stock movements*. Information Sciences, 2014. 278(1): pp. 826-840.
- [17] J.W. Pennebaker, et al., *The development and psychometric properties of LIWC2015*. 2015.
- [18] K.J. Klein and S.W. Kozlowski, *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. 2000: Jossey-Bass.
- [19] C.M. Ringle, S. Wende, and J.-M. Becker *SmartPLS 3. Bönningstedt: SmartPLS*. 2015.
- [20] M.J. Culnan, P.J. McHugh, and J.I. Zubillaga, *How large US companies can use Twitter and other social media to gain business value*. MIS Quarterly Executive, 2010. 9(4): pp. 243-259.