

Han, Xu; Gu, Xingyu; Peng, Shuai

Article

Analysis of tweet form's effect on users' engagement on Twitter

Cogent Business & Management

Provided in Cooperation with:

Taylor & Francis Group

Suggested Citation: Han, Xu; Gu, Xingyu; Peng, Shuai (2019) : Analysis of tweet form's effect on users' engagement on Twitter, Cogent Business & Management, ISSN 2331-1975, Taylor & Francis, Abingdon, Vol. 6, pp. 1-15,
<https://doi.org/10.1080/23311975.2018.1564168>

This Version is available at:

<https://hdl.handle.net/10419/206149>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Received: 26 November 2018
Accepted: 20 December 2018
First Published: 12 January 2019

*Corresponding author: Xu Han,
Master student in Information
System, Carey Business School, Johns
Hopkins University, Washington, DC,
NW, USA
E-mail: xhan24@jhu.edu

Reviewing editor:
Len Tiu Wright, De Montfort
University Faculty of Business and
Law, UK

Additional information is available at
the end of the article

MARKETING | RESEARCH ARTICLE

Analysis of Tweet Form's effect on users' engagement on Twitter

Xu Han^{1*}, Xingyu Gu¹ and Shuai Peng²

Abstract: This research focuses on the effects on users' engagement of different tweet forms including text length, text sentiment and the usage of hashtag, mention, video or picture URL. In the first part, we analyze the tweets of five companies from the apparel industry and finds out that there is no universal form that can boost user's engagement, but in company scale, the effects of different forms between companies are various due to company attributes. Hence, in our second research, we expand the dataset and analyze the formats of tweets from 70 brands focusing on the attribute of the industry section. The conclusion shows that industries such as luxury and hardware technology are more digital sensitive and benefit more using more hashtag and video or picture URL while industry such as software industry is more digital insensitive. The result could provide evidence and guidance for different categories of companies to design tweets with high customer engagements and serve as a reference for enterprises on other media platforms.

Subjects: Mathematics & Statistics; Economics, Finance, Business & Industry; Information Science

Keywords: social media marketing; Twitter; Tweet Form; user engagement

Nowadays, as the marketing competition becomes fiercer and more intensive in many industries, an increasing number of enterprises are trying to switch their marketing platform from traditional



Xu Han

ABOUT THE AUTHORS

Xu Han, Xingyu Gu, Shuai Peng are master of Marketing from Johns Hopkins University Carey business school. In the time when pursuing the master degree, they learnt data analytics with R and conducted course project about studying the data patterns and predicting heart disease for the Cleveland dataset. After the course, they tried to find interesting and realistic problem and lay their interest on social media marketing. They focused on Twitter and analyzed the tweet form's effect on users' engagement, as their first academic research for publishing. Currently, Xu Han and Xingyu Gu study further at Carey Business school as master student in Information System and dedicate their time to data analysis utilizing more advanced tools. Meanwhile, Shuai Peng works as marketing Specialist in a digital marketing company, applying his knowledge on analysis about social media.

PUBLIC INTEREST STATEMENT

With the prevalence of social media, more brands tend to use their official account on social media such as Twitter to enhance brand loyalty and consumer engagement. This perspective article describes some tips for companies to utilize their Twitter account, based on data gathered via TwitterR, a tool provided by Twitter to obtain public tweets data. It finds that shortening or lengthening the tweet, adding more hashtag or mention, or inserting picture or video will not affect the user's engagement in the same way for different companies. Instead, the effects of different tweets' form are industry-various. Industries such as software technology and luxury are more digital, sensitive, and can benefit more by using more hashtag and video or picture URL. The retail industry is more insensitive in these tweets' form changing. So, a company should identify its industry feature first when managing the Twitter account.

media to social media (Montague, Gazal, Wiedenbeck, & Shepherd, 2016), due to the lower cost and diverse presentation forms for marketing contents (Momany & Alshboul, 2016). Twitter, by virtue of its 68 million monthly active users in the United States (Twitter, 2018), is currently employed by numerous companies as their marketing channel in America (Momany & Alshboul, 2016). In order to establish a deep relationship with users, companies are seeking ways to enhance the users' engagement with tweets and replies, so that the interactions could impress users and thus build stronger brand loyalty (Syrdal & Briggs, 2018) (Burton & Soboleva, 2011).

To figure out what kind of format of tweet can attract more customer engagement, a number of scholars have conducted researches on the qualitative analyses on users' engagement with regards to different advertising formats (Tutaj & Van Reijmersdal, 2012) (Karlo, Sivakumaran, & Marathe, 2017), but there are few studies on quantitative exploration for it and no quantitative findings show how different formats of tweet influence their customer engagements. Besides easily thinking the richer the content of tweet the better of its customer engagement or the simpler the better, there should be some customized strategy for a specific industry of the company. Therefore, based on the previous researches and existing studies, this thesis analyzes the main elements of tweet formation and how these elements affect the users' engagements, respectively, so as to provide advice and direction for companies to devise tweets that could arouse more interaction with users.

1. Theoretical basis and data source

1.1. Theoretical basis and model

Tweets include couple different components not only texts but also URL, hashtags and mentions. Their contents show different sentiment scores, which could be a very important factor that influences customer engagement of this tweet. To build a relationship between formats of tweets and customer engagement, we did some research below to find some theoretical supports.

People today rarely communicate with text only, while pictures and videos are used more and more widely in communication (Burnett, 2015), including communication in social media marketing. According to the statistics, press releases that include a photo or video get 45% more views than those with text only (James, 2012). These media serve as stimuli to influence viewers' perceiving and consequent engagement (Byrum, 2014). In tweets, the URL demonstrates the presence of pictures and videos, so we set the URL number as an independent variable for the engagement measurement.

Another unique and important feature of the tweet is the hashtag. It serves as a metadata tag used on Twitter to facilitate a search for a specified topic of interest (Antoine, 2016). For companies, the hashtag can be a supplement to brand image, building community and evoking an efficient response, and companies use the hashtag to launch campaigns, engage directly with consumers (Kleinberg, 2013). Thus, the number of hashtags is reasonable to be perceived to influence consumers' engagement.

Based on the observation of some typical brand tweets, the function of mention ("@") is sometimes used to mention brands' stakeholders, who most of the time are celebrities or influencers (Bao, Hua-Wei, Huang, & Chen, 2013). Similar to celebrity endorsement, the function of mention indicates the public endorsement of the celebrity to the brand, thus enhancing the followers' positive attitude (Tang, Ni, Xiong, & Zhu, 2015).

When Twitter extends its tweet length limits, many researches were conducted to exam the effect of longer tweets. Some literature already indicate that longer text increases the complexity in perception (Arrabal-Sánchez & De-Aguilera-Moyano, 2016), but according to SocialFlow, tweets longer than the old 140-character limit are great for getting attention and engagement (Cho, 2012). Tweets that below 140 characters were retweeted on average 13.71 times and received

29.96 likes. Go above 140 characters and the retweets jump to 26.52 and the likes to 50.28 (Sherman, 2017). So, the tweet length is set as an independent variable.

Twitter offers a unique dataset in the world of brand sentiment as companies can receive sentiment messages directly from consumers. Both the targeted and competing brands can dissect these messages to determine changes in consumer sentiment (Ghiassi, Skinner, & Zimbra, 2013). But the sentiment of tweets posted by the brand can also influence consumer's engagement, as a strong argument with emotion offers a convincing manner (Byrum, 2014). To measure the sentiment of tweets, we use sentiment score as the independent variable.

The concept of customer engagement provides a construct that comprises the total set of behavioral activities toward a firm (Johanna, Veronica, Emil, & Minna, 2012), reflecting the broadest theoretical perspectives of the paradigm of services and of relational marketing and is able to provide enhanced predictive and explanatory power of focal consumer behavior outcomes such as brand loyalty (Linda, Mark, Glynn, & Roderick, 2014).

As for the measurement of users' engagement, a comprehensive algorithm formula (Figure 1.) is referred for the calculation the ratio of weighted engagement on Twitter. (Muñoz-Expósito, Oviedo-García, & Castellanos-Verdugo, 2017)

In this algorithm, the terms mean different actions as detailed below:

Interaction of diffusion (I.diffusion): number of Retweet and Share.

Interaction of conversation (I.conversation): number of Reply.

Interaction of approval (I.approval): number of Likes.

Other interactions: number of Detail Expands, Embedded Media Clicks, Hashtag Clicks, Follows, User Profile Clicks, Link Clicks, Permalick Clicks, Appinstall attempts, App opens, and Leads submitted.

Number of tweets (No. of tweets): number of tweets that posted by the account.

Average impression: total number of views of a tweet divided by the number of tweets considered over the same period.

Average reach: total number of audience of a tweet divided by the number of tweets considered over the same period.

The algorithm covers all the possible actions of the Twitter users that could enable them to interact with the brand and thus reduces the inaccuracy of the result. Moreover, such a formula is also flexible according to the practical situation. Therefore, we choose it as the original algorithm for Twitter engagement calculation.

In this study, we focus on the engagement for every single tweet, so there is no time period for the measurement of all mentioned actions, impressions and reach, as well as other interactions. Besides, only very few tweets contain reply, so we eliminate the effect the reply in the algorithm to reduce further bias. Therefore, only Interaction of diffusion and Interaction of approval are left in terms of user interactions, and their original ratio 3:1.5 could be simplified to 2:1. Furthermore, the number of impressions is almost the same to that of reach, and hence the calculation "impressions/reach" in the algorithm is equal to 1. So, the below-revised algorithm formula will be adopted in the calculation and analysis. (Figure 2)

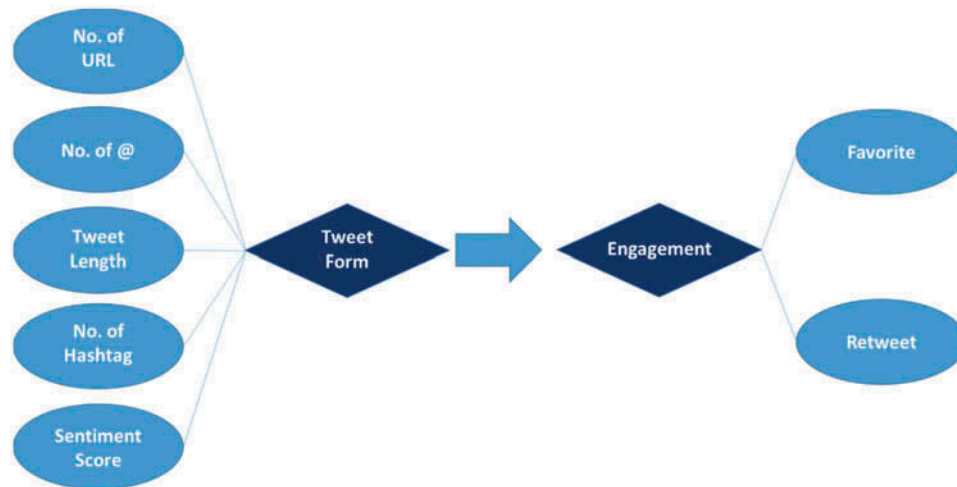
Figure 1. Algorithm formula of engagement on Twitter.

$$\text{Ratio of weighted engagement on Twitter} = \frac{(3 \times (\text{I.diffusion}) + 2 \times (\text{I.conversation}) + 1.5 \times (\text{I.approval}) + \text{other interactions}_{7 \text{ days}}) / (\text{No. of tweets}_{7 \text{ days}})}{\sum \text{Average reach of the tweets}_{7 \text{ days}}} \times 100$$

Figure 2. Revised algorithm formula of engagement on Twitter.

$$\text{Ratio of weighted engagement on Twitter} = \frac{2 \times (\text{I.diffusion}) + (\text{I.approval}) / \text{impressions}}{\text{reach}}$$

Figure 3. Elements of Tweet Form (Independent Variables) and elements of engagements (Dependent variables).



1.2. Hypothesis generation

Therefore, to figure out the influence on customer engagement from the contents of tweets, we make several hypotheses to test the influence contents of tweet and customer engagement. Each type of tweet forms may influence the engagement which can be demonstrated by two attributes, Favorite and Retweet behavior (Figure 3).

H1: the number of URL in a tweet will influence the engagement.

H2: the number of hashtag in a tweet will influence the engagement.

H3: the number of at in a tweet will influence the engagement.

H4: the length of a tweet will influence the engagement..

H5: the sentiment of a tweet will influence the engagement.

1.3. Data collection

To extract information such as tweets and followers from each account, we use R Language and the *Twitter* package. By requiring and utilizing APIs from Twitter (authentication from twitter developer website, apps.twitter.com), we obtain 3200 tweets at maximum for each account including text content, favorite count, created time, retweet count and so on. To set-up a data set for further analysis, different pre-handling steps are needed (See Figure A1 in Appendix).

2. Data collection and empirical analysis

2.1. Data sample selection

In deciding which companies' Twitter accounts should be chosen, we exam several criteria to guarantee the quality of rudimental data for further analysis. First of all, the Twitter accounts of these companies must be active, with abundant tweets daily or weekly. It also demonstrates that these companies indeed spend marketing efforts on social media. Besides, all selected companies should be in the same industry to eliminate the potential risk of heterogeneity in different industries (Sevin, 2013). Also, the companies should be the leading companies in the industry so that the result can be more instructive to a certain industry. For the measurement of this criterion, we focus on the followers in each account, as more followers mean greater brand awareness and higher possibility of quality communication.

Enlightened by Sevin (2013), we exam Twitter accounts in several industries such as airlines, smartphone manufacturers, supermarkets and apparel. And ultimately, we decide to collect data from apparel industry and specifically, from Forever 21 (@Forever21), Gap (@Gap), H&M (@hm), Abercrombie & Fitch (@Abercrombie) and GUESS (@GUESS). Their Twitter accounts all enjoy hundred-thousands of followers and thousands of "favorite".

2.2. Data pre-analysis processing

First, tweets returned by TwitteR inquiry contain both original tweets post by the brand and retweets sent to other users. As retweets by the brand are hard to be observed by its followers and usually irrelevant with its own marketing attempt, we exclude from the rudimental dataset about retweets with "mention" as initial in text attribute.

Before data cleaning, we count the number of hashtags and URLs as they are variables that we need in SPSS analysis. Then, as the process of sentiment analysis, we practice data cleaning to the text, eliminating irrelevant information in the corpus, including URLs, punctuation, number and stop words (Feinerer, Hornik, & Software, 2017). After the data cleaning, some rows in the data frame are vacant in the text, hence we have to delete these rows too for sentiment analysis (Pak & Paroubek, 2010). By Sentiment140 package, we categorize the dataset into three classes: positive sentiments, negative sentiments and a set of objective texts (no sentiments), assigned 1, -1, 0, respectively (See Figure A2 in Appendix).

2.3. Data description

After the data collection and preliminary data process, we finally get a dataset including 316 rows of data. The columns of the table include (1) relevant information: post-data-cleaning text, brand name; (2) independent variables: URL number, hashtag number, mention number, tweet's length, sentimental score (set positive and negative separately as dummy variable), and (3) dependent variables: favorite count, retweet count, reach, impression and engagement ratio.

From the description of data (Table 1), we can see that the average number of favorite count of a tweet is 141.5 and the average number of retweet count of a tweet is 17.75. In our case, the influence on engagement caused by the number of favorites is higher than the number of retweets. Besides, most tweets have no mention and hashtag but have more than one URL. We can see that the median of the favorite count, Retweet count and engagement score are lower than the means, which means there are some extremely big number in these dimensions.

2.4. Empirical analysis

According to the previous hypotheses and meanwhile considering the law of diminishing marginal utility, we use log-linear regression to test these hypotheses. To avoid interaction influence among companies, we create dummy variables for different companies to decrease the influence caused by companies themselves. Finally, we got the following result.

Table 1. Descriptive analysis of the data sample

Descriptive Statistics

| | | Favoritecount | Retweetcount | URLnum | Hashtagnum | Atnum | Length | Engagement. score | Positive | Negative |
|----------------|---------|---------------|--------------|---------|------------|---------|---------|----------------------|----------|----------|
| N | Valid | 316.000 | 316.000 | 316.000 | 316.000 | 316.000 | 316.000 | 316.000 | 316.000 | 316.000 |
| | Missing | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Mean | | 141.500 | 17.753 | 1.595 | 0.491 | 0.472 | 11.465 | 177.382 | 0.054 | 0.013 |
| Median | | 115.000 | 11.000 | 2.000 | 0.000 | 0.000 | 11.000 | 144.000 | 0.000 | 0.000 |
| Std. Deviation | | 145.619 | 33.420 | 0.541 | 0.830 | 0.732 | 5.669 | 204.049 | 0.226 | 0.112 |
| Percentiles | 25 | 51.000 | 5.000 | 1.000 | 0.000 | 0.000 | 6.000 | 58.250 | 0.000 | 0.000 |
| | 50 | 115.000 | 11.000 | 2.000 | 0.000 | 0.000 | 11.000 | 144.000 | 0.000 | 0.000 |
| | 75 | 183.000 | 19.000 | 2.000 | 1.000 | 1.000 | 16.000 | 226.750 | 0.000 | 0.000 |

From the result, we use 95% confidence interval and compare it to a p-value of all independent variables from three different regressions. So, we can have a conclusion to our hypothesis based on the statistical significance.

3. Conclusion

Through in-depth evaluation and comprehensive analysis, we find that all these five independent variables (number of URLs, number of hashtags, number of mentions, length of the tweets, and sentiment of the tweets) do not pose direct influence on users' engagement of the tweet, which is different to public common sense. Namely, there is a general recommendation for these five companies to spend efforts on specific tweets' format as the result show that no independent variables will affect the engagement (Table 3). But from the regression result, we find that the *companies* factors (dummy variables in Table 2) are statistically significant, which means that different companies have different Twitter performance according to their own format of tweets, their brand image and other individual factors, such as industry section. Rather than analyzing the relationship between the contents of individual tweets and their engagement performance, it is more necessary to explore the tweets difference between companies. So, instead, based on the individual tweets, we conduct a further analysis based on individual companies, to examine if there is any pattern in groups of companies that have similar Twitter performance.

4. Effect analysis on specific industries

4.1. Data sample selection

In this section, we expand our selection of companies and their Twitter accounts. Because the number of followers can indicate the activity of companies' social media marketing to some extent

Table 2. R language result of different Tweet Form's effect on engagement score

| Regression result | | | | |
|-------------------|----------|------------|---------|----------|
| | Estimate | Std. Error | t value | Pr(> t) |
| (Intercept) | 114.638 | 75.497 | 1.518 | 0.13 |
| URLnum | -30.189 | 24.204 | -1.247 | 0.213 |
| Hashtagnum | 5.563 | 13.979 | 0.398 | 0.691 |
| Atnum | 13.562 | 17.381 | 0.78 | 0.436 |
| Length | -3.199 | 2.918 | -1.097 | 0.274 |
| Positive | -32.674 | 48.475 | -0.674 | 0.501 |
| Negative | -92.75 | 95.673 | -0.969 | 0.333 |
| GUESS | 155.576 | 45.092 | 3.45 | 0.001 |
| Forever21 | 196.264 | 48.167 | 4.075 | 0 |
| Abercrombie | 22.521 | 42.935 | 0.525 | 0.6 |
| hm | 223.835 | 46.221 | 4.843 | 0 |

Table 3. Result of hypothesis

| Hypothesis | Influence of contents of tweet on customer engagement | Result |
|------------|---|-------------|
| H1: | The number of URL in a tweet will influence the engagement. | No support* |
| H2: | The number of hashtag in a tweet will influence the engagement. | No support |
| H3: | The number of @ in a tweet will influence the engagement. | No support |
| H4: | The length of a tweet will influence the engagement. | No support |
| H5: | The sentiment of a tweet will influence the engagement. | No support |

*No support means not statistically significant.

and influence the sample size (more followers can have a significant difference in the number of retweets and favorites on different tweets), we choose 100 companies with the top number of followers on Twitter. To eliminate the heterogeneity caused by the followers' geography, we exclude the companies and accounts that are local-based and finally obtain 70 global companies with a high number of followers. Among them, there are 22 technology companies, 17 luxury brands, 7 game companies 7 retail companies, 5 FMCG companies, 4 auto brand, 2 news companies and 6 others.

4.2. Data description

We collect data from 70 different companies and filter 46,822 pieces' tweets as our dataset, which have all the factors we need to analyze. We summarize all data in Table 4, and you can see all tweets have diversified components, which give us more chance to analyze the influence of format of tweet on its engagement. As we conclude above, for different companies, their Twitter performance (tweets engagement, the number of retweets and number of favorites) are determined by different factors statistical significantly. For example, some companies' Twitters performance may be significantly influenced by URL numbers when others may be not by it but hashtag number. Also, the influence can be positive or negative to different companies.

So, we summarize all the factors that are both positive significant and negative significant for each company and use 1 as the positive coefficient, -1 as the negative coefficient and 0 we present in the table shows that this independent variable is not statistically significant to the specific company. Then, a company with result 1 should enrich its use on this variable. If a factor is significant, we then call the company is sensitive towards the factors and the proportion of these companies in the same industry is the sensitive rate of a factor for this industry.

General speaking, for the influence of tweet contents on engagement, almost 1/3 companies are sensitive on the URL number and hashtag number, reaching 31.43% and 34.29%, respectively. Among these companies, 2/3 of their engagement performance are positively influenced. Mention number has an equal rate in positive sensitive and negative sensitive while the length is mainly a negative influence on engagement with 7.14% positive sensitive ratio and 17.14% negative sensitive rate (Table 5). As for sentiment score, it affects the engagement score of companies only in a positive way.

4.3. Data empirical analysis

From Table 6, we can see that the retail industry is a good category of companies to design different messages to increase companies' customer engagement. The number of URL, hashtag and sentiment score of tweets have a positive influence on 14.29%, 28.57% and 42.86% of the companies' customer engagement, respectively, and no negative influence on any company's customer engagements. Besides, the number of mention and the length of a tweet have a negative influence on 14.29% and 14.29% of the companies' customer engagement and no positive influence on any company's customer engagement. All in all, retail companies should make their tweets content including more URL, hashtag and positive words by decreasing mentions and length at the same time to increase customer engagement.

Besides, the sentiment score has no negative influence on any company's customer engagement and has a positive influence on 18.57% of the companies' customer engagement. It positively influences 42.86% of the retail companies' customer engagement, who are influenced by sentiments score most severely. Generally, people prefer positive content, and companies should take care of their tweets design by making their contents more positive.

Technology companies who based on hardware business have different influences from tweets format design compare with those who based on the software business. Different

| Table 4. Descriptive analysis of the data sample | | | | | | | | | |
|--|---------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Descriptive Statistics | | | | | | | | | |
| N | Valid | FavoriteCount | RetweetCount | URLnum | Hashtagnum | Atnum | Length | Score | Engagement. score |
| | Missing | 46,820.000 2.000 | 46,822.000 0.000 | 46,822.000 0.000 | 46,820.000 2.000 | 46,820.000 2.000 | 46,820.000 2.000 | 46,820.000 2.000 | 46,820.000 2.000 |
| Mean | | 424.120 | 169.720 | 1.270 | 0.930 | 0.330 | 14.000 | 0.030 | 817.730 |
| Median | | 188.000 | 54.000 | 1.000 | 1.000 | 0.000 | 14.000 | 0.000 | 303.000 |
| Mode | | 28.000 | 21.000 | 1.000 | 1.000 | 0.000 | 17.000 | 0.000 | 57.000 |
| Percentiles | 25 | 88.000 | 26.000 | 1.000 | 0.000 | 0.000 | 11.000 | 0.000 | 153.000 |
| | 50 | 188.000 | 54.000 | 1.000 | 1.000 | 0.000 | 17.000 | 0.000 | 303.000 |
| | 75 | 386.000 | 116.000 | 2.000 | 1.000 | 1.000 | 17.000 | 17.000 | 0.000 |

Table 5. Further R Language Result of Different Tweet Form's Effect on Engagement Score

| Engagement-Tweets format relationship results | | | | | | | | | | | |
|---|--------|-----------|-------|--------|-------|-------------------------|--------|-----------|--------|--------|--------|
| Company | URLnum | hashagnum | athum | length | score | Company | URLnum | hashagnum | athum | length | score |
| 1 CHANEL | 1 | 1 | 0 | 0 | 0 | 39 NORDCURRENT | 1 | 0 | 0 | 0 | 0 |
| 2 XBOX | 1 | 1 | 0 | 0 | 0 | 40 RAZER | 1 | 0 | 0 | 0 | 0 |
| 3 SAMSUNGMOBILE | 1 | 1 | 0 | 0 | 0 | 41 PEPSI | 1 | 0 | 0 | 0 | 0 |
| 4 ANDROID | 1 | 1 | 0 | 0 | 0 | 42 LOUBOUTIN | 1 | 0 | 0 | 0 | 0 |
| 5 APPLE MUSIC | 1 | 1 | 0 | 0 | 0 | 43 WORLDBANK | 1 | 0 | 0 | 0 | 0 |
| 6 ROCKSTARGAMES | 1 | 1 | 0 | 0 | 0 | 44 FIREFOX | 1 | 0 | 0 | 0 | 0 |
| 7 MARCJACOBS | 1 | 1 | 0 | 0 | 0 | 45 BENZ | 1 | 0 | 0 | 0 | 0 |
| 8 HM | 1 | 1 | 0 | 0 | 0 | 46 SPOTIFY | 1 | 0 | 0 | 0 | 0 |
| 9 BURBERRY | 1 | 1 | 0 | 0 | 0 | 47 HTC | 1 | 0 | 0 | 0 | 0 |
| 10 MICROSOFT | 1 | 1 | 0 | 0 | 0 | 48 SEPHORA | 1 | 0 | 0 | 0 | 0 |
| 11 DIOR | 1 | 1 | 0 | 0 | 0 | 49 SOUNDCLLOUD | 1 | 0 | 0 | 0 | 0 |
| 12 HOOTSUITE | 1 | 1 | 0 | 0 | 0 | 50 RALPHLAUREN | 1 | 0 | 0 | 0 | 0 |
| 13 NIKE | 1 | 1 | 0 | 0 | 0 | 51 GOPRO | 1 | 0 | 0 | 0 | 0 |
| 14 LOUISVUITTON | 1 | 1 | 0 | 0 | 0 | 52 NOKIA | 1 | 0 | 0 | 0 | 0 |
| 15 UBISOFT | 1 | 1 | 0 | 0 | 0 | 53 REDBULL | 1 | 0 | 0 | 0 | 0 |
| 16 GOOGLEPLAY | 1 | 1 | 0 | 0 | 0 | 54 ZAIN | 1 | 0 | 0 | 0 | 0 |
| 17 WINDOWS | 1 | 1 | 0 | 0 | 0 | 55 LUMIA | 1 | 0 | 0 | 0 | 0 |
| 18 GUCCI | 1 | 1 | 0 | 0 | 0 | 56 VALENTINO | 1 | 0 | 0 | 0 | 0 |
| 19 SKYPE | 1 | 1 | 0 | 0 | 0 | 57 LIGHTROOM | 1 | 0 | 0 | 0 | 0 |
| 20 DOLCEGABBANA | 1 | 1 | 0 | 0 | 0 | 58 MCQUEEN | 1 | 0 | 0 | 0 | 0 |
| 21 EA | 1 | 1 | 0 | 0 | 0 | 59 FACOBELL | 1 | 0 | 0 | 0 | 0 |
| 22 TEAM_GAMES | 1 | 1 | 0 | 0 | 0 | 60 WORDPRESS | 1 | 0 | 0 | 0 | 0 |
| 23 INTEL | 1 | 1 | 0 | 0 | 0 | 61 SEGA | 1 | 0 | 0 | 0 | 0 |
| 24 APPSTORE | 1 | 1 | 0 | 0 | 0 | 62 LAMBORGHINI | 1 | 0 | 0 | 0 | 0 |
| 25 DROPBOX | 1 | 1 | 0 | 0 | 0 | 63 FORSCHE | 1 | 0 | 0 | 0 | 0 |
| 26 VERSACE | 1 | 1 | 0 | 0 | 0 | 64 AWS | 1 | 0 | 0 | 0 | 0 |
| 27 BLACKBERRY | 1 | 1 | 0 | 0 | 0 | 65 BMW | 1 | 0 | 0 | 0 | 0 |
| 28 YSL | 1 | 1 | 0 | 0 | 0 | 66 PUMA | 1 | 0 | 0 | 0 | 0 |
| 29 ADIDASORIGINALS | 1 | 1 | 0 | 0 | 0 | 67 NYSE | 1 | 0 | 0 | 0 | 0 |
| 30 TRIPADVISOR | 1 | 1 | 0 | 0 | 0 | 68 ROBERTO | 1 | 0 | 0 | 0 | 0 |
| 31 CALVINKLEIN | 1 | 1 | 0 | 0 | 0 | 69 TOMMY | 1 | 0 | 0 | 0 | 0 |
| 32 JUMPMAN23 | 1 | 1 | 0 | 0 | 0 | 70 JIMMYCHOO | 1 | 0 | 0 | 0 | 0 |
| 33 GOOGLEMAPS | 1 | 1 | 0 | 0 | 0 | Number of positive | | | | | |
| 34 GIORGIOARMANI | 1 | 1 | 0 | 0 | 0 | Number of negative | | | | | |
| 35 COCACOLA | 1 | 1 | 0 | 0 | 0 | Positive sensitive rate | | | | | |
| 36 PHOTOSHOP | 1 | 1 | 0 | 0 | 0 | Negative sensitive rate | | | | | |
| 37 OPTIC | 1 | 1 | 0 | 0 | 0 | Over all sensitive rate | | | | | |
| 38 MONSTER | 1 | 1 | 0 | 0 | 0 | Sum | | | | | |
| | | | | | | 20 | 16 | 5 | 8 | 18 | 67 |
| | | | | | | 10 | 8 | 14 | 13 | 4 | 49 |
| | | | | | | 28.57% | 22.86% | 7.14% | 11.43% | 25.71% | 19.14% |
| | | | | | | 14.29% | 11.43% | 20.00% | 18.57% | 5.71% | 14.00% |
| | | | | | | 42.86% | 34.29% | 27.14% | 30.00% | 31.43% | 33.14% |

Table 6. R language result of different Tweet Form's effect on engagement score for different categories of companies

| Engagement-Tweets format relationship summary | | | | | | |
|---|--------|------------|---------|---------|--------|--------|
| | URLnum | hashtagnum | atnum | length | score | sum |
| Positive sensitive rate(average) | 21.43% | 27.14% | 10.00% | 7.14% | 18.57% | 16.86% |
| Negative sensitive rate(average) | 10.00% | 7.14% | 11.43% | 17.14% | 0.00% | 9.14% |
| Overall sensitive rate(average) | 31.43% | 34.29% | 21.43% | 24.29% | 18.57% | 26.00% |
| Overall relative sensitive rate(average) | 11.43% | 20.00% | -1.43% | -10.00% | 18.57% | 7.71% |
| Positive sensitive rate(retail) | 14.29% | 28.57% | 0.00% | 0.00% | 42.86% | 17.14% |
| Negative sensitive rate(retail) | 0.00% | 0.00% | 14.29% | 14.29% | 0.00% | 5.71% |
| Overall sensitive rate(retail) | 14.29% | 28.57% | 14.29% | 14.29% | 42.86% | 22.86% |
| Overall relative sensitive rate(retail) | 14.29% | 28.57% | -14.29% | -14.29% | 42.86% | 11.43% |
| Positive sensitive rate(luxury) | 29.41% | 41.18% | 23.53% | 11.76% | 17.65% | 24.71% |
| Negative sensitive rate(luxury) | 11.76% | 5.88% | 11.76% | 11.76% | 0.00% | 8.24% |
| Overall sensitive rate(luxury) | 41.18% | 47.06% | 35.29% | 23.53% | 17.65% | 32.94% |
| Overall relative sensitive rate(luxury) | 17.65% | 35.29% | 11.76% | 0.00% | 17.65% | 16.47% |
| Positive sensitive rate(game) | 28.57% | 42.86% | 0.00% | 14.29% | 14.29% | 20.00% |
| Negative sensitive rate(game) | 28.57% | 0.00% | 28.57% | 28.57% | 0.00% | 17.14% |
| Overall sensitive rate(game) | 57.14% | 42.86% | 28.57% | 42.86% | 14.29% | 37.14% |
| Overall relative sensitive rate(game) | 0.00% | 42.86% | -28.57% | -14.29% | 14.29% | 2.86% |
| Positive sensitive rate(tech-hard) | 42.86% | 57.14% | 14.29% | 0.00% | 28.57% | 28.57% |
| Negative sensitive rate(tech-hard) | 14.29% | 14.29% | 14.29% | 14.29% | 0.00% | 11.43% |
| Overall sensitive rate(tech-hard) | 57.14% | 71.43% | 28.57% | 14.29% | 28.57% | 40.00% |
| Overall relative sensitive rate(tech-hard) | 28.57% | 42.86% | 0.00% | -14.29% | 28.57% | 17.14% |
| Positive sensitive rate(tech-soft) | 6.67% | 13.33% | 6.67% | 6.67% | 0.00% | 6.67% |
| Negative sensitive rate(tech-soft) | 0.00% | 6.67% | 13.33% | 20.00% | 0.00% | 8.00% |
| Overall sensitive rate(tech-soft) | 6.67% | 20.00% | 20.00% | 26.67% | 0.00% | 14.67% |
| Overall relative sensitive rate(tech-soft) | 6.67% | 6.67% | -6.67% | -13.33% | 0.00% | -1.33% |

formats have some influence on 40% of the companies' customer engagement who are technology companies based on hardware business, but only 14.67% of those technology companies based on the software business. The format design of tweets matters more for technology companies based on hardware business compare with those based on the software business.

Finally, if game companies use more hashtags and fewer mentions in their tweets and technology companies use fewer words in their tweets content, their tweets' formats will help them to increase customer engagement.

5. Summary

5.1. Conclusion

Based on the comprehensive analysis on tweets from above-mentioned brands and companies, we come up with the conclusion that company marketing teams should design the tweet with considerations toward which category their company lies in, since different forms of tweets will contribute to the different results of engagement score in terms of different categories of

companies. For managers, when considering the utilization of Twitter as part of social media marketing strategy, it is better to analyze the Twitter performance of companies similar patterns, in our study, same industry, before generating own tweets content. In that way, managers can recognize whether the company is necessary to manage the Twitter account in terms of social media sensitivity and which tweet format will improve the Twitter engagement at most.

5.2. Limitation

In the study, we focus on the engagement for tweets as our main dependent variable. And we generally indicate that the engagement is the only goal for the company using Twitter. But sometimes the company will use Twitter for other reasons such as announcement which do not care much about engagement. Also, we do not consider the time period for the measurement of the actions, which is hard to obtain from Twitter. Further study can include these considerations.

Based on our first data analysis, we find the further research necessity on the exam the relationship between tweets' engagement performance regarding different formats and companies' different attributes. In this thesis, we only focus on industry section as one feature of companies. Further study can explore other factors that differ the results on tweet performance, to provide other aspects for companies considering applying Twitter.

Besides, the number of companies we select is limited and not applicable to log-linear analysis. So our conclusion about how a company should design the format of its tweet is based on the descriptive analysis. We will expand our database with more companies from different industries to conduct our research and find out statistical significance to approve our conclusion.

Acknowledgements

This paper is totally conducted by us without other financial supports or direct guides. There is no conflict of interest regarding the publication of this paper. Someone helped us with some useful suggestions are Chen Tao, Minghong Xu and Ling Hu. This paper uses data from Twitter API, which is an open source data analysis access operated by Twitter. We are also grateful for this platform provided by Twitter.

Funding

The authors received no direct funding for this research.

Author details

Xu Han¹
E-mail: xhan24@jhu.edu
Xingyu Gu¹
E-mail: xingyu_gu@jhu.edu
Shuai Peng²
E-mail: speng14@jhu.edu

¹ Master student in Information System, Carey Business School, Johns Hopkins University, Washington, DC, NW, USA.

² Marketing Specialist, WarmStone Co. Ltd., Johns Hopkins University, Changsha, Hunan Province, China.

Supplementary material

Supplementary material for this article can be accessed [here](#).

Citation information

Cite this article as: Analysis of Tweet Form's effect on users' engagement on Twitter, Xu Han, Xingyu Gu & Shuai Peng, *Cogent Business & Management* (2019), 6: 1564168.

References

Antoine, H. A. (2016). Hashtagging: What you need to know about hashtags as trademarks, hashtag litigation, the FTC, viral campaigns, and more. *Computer & Internet Lawyer*, 33(11), 6–10. Retrieved from <http://search.ebscohost.com.proxy1.library.jhu.edu/login.aspx?direct=true&db=bsu&AN=119547933&site=ehost-live&scope=site>

Arrabal-Sánchez, G., & De-Aguilera-Moyano, M. (2016). Communicating in 140 characters. How Journalists in Spain use Twitter. *Comunicar*, 24(46), 9–17. doi:10.3916/C46-2016-01

Bao, P., Hua-Wei, S., Huang, J., & Chen, H. (2013, March). *Mention effect in information diffusion on a micro-blogging network*. San Francisco: Public Library of Science.

Burnett, R. (2015, July 6). *Images and videos - media for communication*. London: MENA Report.

Burton, S., & Soboleva, A. (2011). Interactive or reactive? Marketing with Twitter. *Journal of Consumer Marketing*, 28, 491–499. doi:10.1108/07363761111181473

Byrum, K. L. (2014). *A comparison of the source, media format, and sentiment in generating source credibility, information credibility, corporate brand reputation, purchase intention, and social media engagement in a corporate social responsibility campaign presented via soci*. Ann Arbor: ProQuest Dissertations Publishing.

Cho, C. (2012, December). Domains/SocialFlow. *PRweek*, p. 1.

Feinerer, I., Hornik, K., & Software, A. (2017, December 6). Text Mining Package “tm”. Retrieved from R-Project: <https://cran.r-project.org/web/packages/tm/tm.pdf>

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40(16), 6266–6282. doi:10.1016/j.eswa.2013.05.057

James, M. (2012, April). The power of pictures in marketing. *LIMRA's Marke Facts Quarter*, p. 19.

Johanna, G., Veronica, L., Emil, W., & Minna, P. (2012). Customer engagement in a facebook brand community. *Management Research Review*, 35(9), 857–877. Retrieved from <https://doi-org.proxy1.library.jhu.edu/10.1108/01409171211256578>

- Kalro, A. D., Sivakumaran, B., & Marathe, R. R. (2017). The ad format-strategy effect on comparative advertising effectiveness. *European Journal of Marketing*, 51, 99–122. doi:10.1108/EJM-11-2015-0764
- Kleinberg, S. (2013, May 20). To hashtag or not to hashtag?: That is the question - how to use them correctly without going overboard. *Chicago Tribune; Chicago, Ill*, pp. 2–5. doi:10.1016/j.jhazmat.2013.03.057
- Linda, H., Glynn, M., & Roderick, B. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, 28(2), 149–165. ISBN 1094-9968. Retrieved from doi:10.1016/j.intmar.2013.12.002
- Momany, M., & Alshboul, A. (2016). Social media marketing: Utilizing social media to advance brand awareness and increase online sales. *International Journal of Business, Marketing, & Decision Science*, 9(1), 33–54.
- Montague, I., Gazal, K. A., Wiedenbeck, J., & Shepherd, J.-G. (2016). Forest products industry in a digital age: A look at E-Commerce and social media. *Forest Products Journal*, 66, 49–57. doi:10.13073/FPJ-D-14-00104
- Muñoz-Expósito, M., Oviedo-García, M., & Castellanos-Verdugo, M. (2017). How to measure engagement in Twitter: Advancing a metric. *Internet Research*, 27, 1122–1148. doi:10.1108/IntR-06-2016-0170
- Pak, A., & Paroubek, P. (2010). *Twitter as a corpus for sentiment analysis and opinion mining*. Paris: European Languages Resources Association.
- Sevin, E. (2013). Places going viral: Twitter usage patterns in destination marketing and place branding. *Journal of Place Management and Development*, 6, 227–239. doi:10.1108/JPM-10-2012-0037
- Sherman, E. (2017, December). A genius way to get more attention for tweets also makes life easier. Retrieved from <https://www.inc.com/erik-sherman/a-genius-way-to-get-more-attention-for-tweets-also-makes-life-easier.html>
- Sherman, E. (2018, July 16). A genius way to get more attention for tweets also makes life easier. *Inc.: A genius way to get more attention for tweets also makes life easier*.
- Syrdal, H. A., & Briggs, E. (2018, January 1). Engagement with social media content: A qualitative exploration. *Journal of Marketing Theory & Practice*, 26, 4–22. doi:10.1080/10696679.2017.1389243
- Tang, L., Ni, Z., Xiong, H., & Zhu, H. (2015, July). Locating targets through mention in Twitter. *World Wide Web*, 18, 1019–1049. doi:10.1007/s11280-014-0299-8
- Tutaj, K., & Van Reijmersdal, E. A. (2012). Effects of online advertising format and persuasion knowledge on audience reactions. *Journal of Marketing Communications*, 18, 5. doi:10.1080/13527266.2011.620765
- Twitter. (2018). *Number of monthly active Twitter users in the United States from 1st quarter 2010 to 4th quarter 2017 (in millions)*. United States: Author.
- Twitter Developer. (n.d.). Twitter API reference. Retrieved from <https://developer.twitter.com/en/docs/api-reference-index>
- Unknown. (n.d.). Sentiment140. Retrieved from <http://help.sentiment140.com/home>

Appendix A.

Figures below show R codes this study used to catch, clean and reorganize data into formats that can be done further analysis. These codes can be replicable used to address further or related conclusions of this study.

Figure A1. R language code of TwitterR to catch data from Twitter API.

```
> library(twitterR)
> library(ROAuth)
> consumer_key<-"TRbkGY43*****1DIW0fK"
> consumer_secret="4GdrHLgo*****KUYHn0dVPfjWbObRGt"
> access_token="810736*****Uw4cDQnlsb3"
> access_secret="j0tCOPQ*****sSrytYd9hhj"
> setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)

[1] "Using direct authentication"
> HM=userTimeline("hm",n=3200,excludeReplies = TRUE)
> HM.tweets.df <- twListToDF(HM)
```

Figure A2. R language code for data cleaning and reorganization.

```
> removeURL <- function(x) gsub("http[^\s:]*", "", x)
> removeNumPunct <- function(x) gsub("[^\s:]*", "", x)
> ##SAMPLE
> SAMPLE.tweets.df <- twListToDF(userTimeline("SAMPLE",n=3200,excludeReplies = TRUE))###66
> SAMPLE.tweets.df$picturenum=str_count(SAMPLE.tweets.df$text, "t.co")
> SAMPLE.tweets.df$hashtagnum=str_count(SAMPLE.tweets.df$text, "#")
> SAMPLE.tweets.df$atnum=str_count(SAMPLE.tweets.df$text, "@")
> ###Count number of URL, hashtag and at###
> SAMPLE.myCorpus <- Corpus(VectorSource(SAMPLE.tweets.df$text))
> SAMPLE.myCorpus <- tm_map(SAMPLE.myCorpus, content_transformer(removePunctuation))
> SAMPLE.myCorpus <- tm_map(SAMPLE.myCorpus, content_transformer(removeURL))
> SAMPLE.tweets.newtwxt1 <- data.frame(text=apply(SAMPLE.myCorpus, identity,
stringsAsFactors=F)
> SAMPLE.tweets.df$length=lengths(gregexpr("\\W+", SAMPLE.tweets.newtwxt1$text))
> SAMPLE.tweets.df$length=ifelse(SAMPLE.tweets.df$length<=2,SAMPLE.tweets.df$length-
1,SAMPLE.tweets.df$length)
> ###Count the length of each tweet###Data clean for length###
> SAMPLE.myCorpus <- tm_map(SAMPLE.myCorpus, content_transformer(removeNumPunct))
> SAMPLE.myCorpus <- tm_map(SAMPLE.myCorpus, removeWords, stopwords("english"))
> SAMPLE.myCorpus <- tm_map(SAMPLE.myCorpus, tolower)
> SAMPLE.myCorpus <- tm_map(SAMPLE.myCorpus, stripWhitespace)
> ###Other data clean###
> SAMPLE.tweets.newtwxt2 <- data.frame(text=apply(SAMPLE.myCorpus, identity,
stringsAsFactors=F)
> SAMPLE.tweets.df$text=SAMPLE.tweets.newtwxt2$text
> ###transfer into data frame###
> SAMPLE.sentiments <- sentiment(SAMPLE.tweets.df$text)
> # sentiment plot
> SAMPLE.sentiments$score <- 0
> SAMPLE.sentiments$score[SAMPLE.sentiments$polarity == "positive"] <- 1
> SAMPLE.sentiments$score[SAMPLE.sentiments$polarity == "negative"] <- -1
> SAMPLE=cbind(SAMPLE.tweets.df,SAMPLE.sentiments)
> ###Sentiment analysis###
> SAMPLE$engagementscore=2*SAMPLE$retweetCount+SAMPLE$favoriteCount
> #engagement score
> SAMPLE.3sd <- mean(SAMPLE$favoriteCount)+ 3*sd(SAMPLE$favoriteCount)
> SAMPLE$favoriteCount <- ifelse(SAMPLE$favoriteCount
SAMPLE.3sd,SAMPLE.3sd,SAMPLE$favoriteCount)
```




© 2019 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format. Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

***Cogent Business & Management* (ISSN: 2331-1975) is published by Cogent OA, part of Taylor & Francis Group.**

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

