

# Does Bad News Spread Faster?

Anna Fang

Department of Computer Science  
Cornell University  
Ithaca, New York 14853

Zina Ben-Miled

Department of Electrical and Computer Engineering  
Indiana University Purdue University Indianapolis  
Indianapolis, IN 46202

**Abstract**—Bad news travels fast. Although this concept may be intuitively accepted, there has been little evidence to confirm that the propagation of bad news differs from that of good news. In this paper, we examine the effect of user perspective on his or her sharing of a controversial news story. Social media not only offers insight into human behavior but has also developed as a source of news. In this paper, we define the spreading of news by tracking selected tweets in Twitter as they are shared over time to create models of user sharing behavior.

Many news events can be viewed as positive or negative. In this paper, we compare and contrast tweets about these news events among general users, while monitoring the tweet frequency for each event over time to ensure that news events are comparable with respect to user interest. In addition, we track the tweets of a controversial event between two different groups of users (i.e., those who view the event as positive and those who view it as negative). As a result, we are able to make assessments based on a single event from two different perspectives.

**Index Terms**—Data analytics, data mining, news, sentiment analysis, social media, Twitter

## I. INTRODUCTION

What kind of news do we read and why? While the media is often criticized for its emphasis on negativity, does media cover bad news because of its higher level of public interest? In this paper, we aim to answer this question by investigating the differences in sharing behavior based on the user sentiment of a news story. To the best of our knowledge, this work is the first to confirm the intuition that news spread is affected by public sentiment.

In modern society, social networks have proven to be vital in the spread of information and have become a representation of human behavior [1]. The rising role of social networks in information spread is especially prevalent with news sharing. Social media users are able to mobilize, discuss, and influence one another easily, which produces a rich social environment yielding dynamic information flow and propagation. Past work has confirmed, though, that information spread differs based on various factors. One such factor is the exposure of a user to this information, depending on the sharing patterns of his or her network neighbors [2]. In particular, the microblogging site Twitter, which we used for our experiment, is often used in studies to track how the spread of stories die down over time. For Twitter, the influences that affect the proliferation of news can be

categorized as follower, reply, or retweet influence [3]. However, the impact of a user's opinion of a particular topic on the message spread has not been factored in previous studies.

Twitter is a primary social media platform used as a news source by many users. On Twitter, users may "tweet" short messages of up to 140 characters that can then be broadcasted to their followers. These tweets can contain text, photos, links, and other media. Often, tweets include a "hashtag", which is a short phrase or word preceded by a hash sign (#) that usually represents the subject of the tweet. Users can also "retweet" other users' tweets, which often contain links to news stories and the user's own commentaries. After retweeting, these tweets are rebroadcasted to a user's followers in addition to being posted on the creator's own profile page. Users therefore discover tweets through who they are following as well as through searching Twitter for tweets containing certain keywords or hashtags. Users may follow other users (although not necessarily in a reciprocated fashion) and view their activity on their homepage feed. Another activity on Twitter is "liking" (formerly known as "favoriting") a tweet, which places the tweet on a user's own profile but does not actively share it to the user's followers. As opposed to retweeting, a user does not have the option of adding his or her own comment to a liked tweet and the tweet itself is not shared again on Twitter. As such, retweeting is an active sharing behavior whereas liking is not [4].

The goal of our paper is to explore whether the propagation of news on social media, namely Twitter, depends on public sentiment. To conduct our study, we used Twitter as our data source and tracked Twitter user sharing behavior. After extracting tweets pertaining to various news stories, we analyzed and modeled the differences in user behavior based on the user sentiment towards the news story. In aggregate, we use this information to assess the claim that bad news travels faster than good news. Our results beg the question of, if "bad news sells", is this affinity to negativity an indication of an inherent human interest in tragedy?

## II. METHOD

There are two possible approaches to the analysis of news spread based on user sentiment. The first involves the selection of a controversial topic that cannot be definitively classified as negative or positive. For this topic, users are divided into those who support and those who oppose the event. In essence, users are segregated according to their *perception* of the target news story. The second analysis involves multiple topics, each of which is generally viewed by the public as either positive or negative. After splitting the topics into two groups, namely positive or negative, we can observe general user sharing behavior with respect to these news stories. In this paper we focus on the first method. The second method is out of scope of this paper and will be considered in our future work.

The first method is used in conjunction with a highly controversial topic: Britain leaving the EU in June 2016 (also known as Brexit). Opinions on the referendum were quite polarized, and the story unleashed much support as well as backlash on social media. The divide between users was interestingly also observed in relation to the users' sentiment towards Donald Trump. Users who tweet about Brexit were identified and then split into Trump supporters and opponents. This experiment showed a strong correlation between public support of Brexit and support for Donald Trump's campaign for President of the United States in 2016. In other words, we identified users who tweet about Brexit and then split these users into Trump supporters and opponents to compare the behavior of these two user groups.

## III. DATA COLLECTION

Although news reports are present on many social media sites, most are unable to provide adequate information for our experiment. An important characteristic of Twitter data is its transparency, as tweets are provided along with many details including the tweet creation timestamp and user information [1].

To collect the needed raw data, the statistical programming language R was used [5]. By using the *twitteR* package of R and the Twitter web API, a keyword search is able to return hundreds of tweets at a time. Along with keywords, it is possible to search Twitter using additional criteria like geolocation, time period, and user handle. The search results includes information such as the tweet's number of retweets and likes, the timestamp, the user handle and ID, links, and other additional information. The results of the search includes the most recent public tweets that match the search criteria as well as the associated retweets. For the purpose of this study, we only focused on tracking original tweets and therefore filtered out the retweets.

In addition to finding recent tweets through the keyword search, we also performed a manual search on

Twitter's search engine using commonly used topic keywords. We chose only tweets that had a clear bias in sentiment, and added these popular tweets to our collection. Thus, our raw data set consisted of both recent and highly circulating tweets associated with both positive and negative sentiments. Ultimately, the top tweets found using Twitter's search engine have higher influence as they spread further and faster than the average tweets found from the keyword search, and appropriately will contribute the most to our calculations.

In summary, in order to extract the target tweets for the Brexit topic, we searched for tweets that include one or more popular hashtags or keywords related to Brexit, including "#brexit", "britain", "EU", "no2eu", "yes2eu", and "#euref".

## IV. TRACKING

We recorded the selected tweets in a table along with all the related metadata mentioned above. We then narrowed down our collection by choosing the tweets that had sufficient details, clear stances, were recently created, and with non-negligent circulation traffic. The collected tweets were created anywhere from June 25, 2016 to June 30, 2016. The total number of retweets of our collected tweets ranged from 24 to 10,360 while the total number of likes ranged from 14 to 29,233. We removed the tweets that were deleted and the tweets that had no activity during the tracking period. Because of the lack of activities in these tweets, this deletion did not affect our results. The final collection consisted of 19 tweets for pro-Brexit and 19 tweets for anti-Brexit for a total collection of 38 tweets. We labeled the anti-Brexit tweets as "Remain" and the pro-Brexit tweets as "Leave".

For 6 consecutive days, we manually recorded each tweet's number of retweets and likes. The vast majority of tweets stopped being retweeted or liked after 3 to 4 days of their creation.

## V. NEWS SPREAD

In order to validate our first hypothesis (i.e., does bad news travel faster), we analyzed the behavior associated with the "Leave" and "Remain" groups. The difference in retweets and likes per tweet were measured for each day. We refer to these as "New Retweets" and "New Likes", respectively. In other words, only new activity related to a given tweet is recorded each day. This measure represents the *growth* in a tweet from one day to the next.

$$\frac{NewRetweets}{NewRetweets + NewLikes} \quad (1)$$

As previously mentioned, we view a retweet as an act of sharing whereas a like is not. Our goal is to investigate whether Remain users (anti-Brexit) have a

Date	27-Jun-16	28-Jun-16	29-Jun-16	30-Jun-16	1-Jul-16	2-Jul-16
<i>Remain: Bad News, Anti-Brexit</i>	<b>0.4182161</b>	0.3922594	<b>0.3856209</b>	<b>0.5037037</b>	<b>0.46168</b>	<b>0.31373</b>
<i>Leave: Good News, Pro-Brexit</i>	0.2799078	<b>0.4284479</b>	0.2971698	0.4323944	0.42331	0.22807

TABLE I: New Retweet Ratio for Brexit

higher proportion of retweet activity than that of the Leave user group (pro-Brexit). If this hypothesis is true, we can conclude that there is a more active spread of bad news and users who perceive news as negative are more inclined to share it with followers.

In order to aggregate all tweets within a group, we sum all New Retweets within a given user group (i.e., Remain or Leave) and divide this number by the sum of total number of activities in the same group as shown in Equation 1. This equation computes the *proportion* associated with sharing compared to the total number of activity related to the group. We believe that this metric is a valid indication of the spread of news in each group.

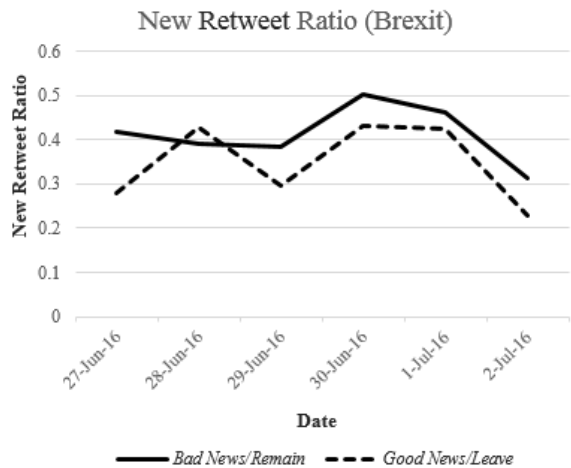


Fig. 1: Graph of Bad News Retweet Ratio (Black) and Good News Retweet Ratio (Red)

## VI. RESULTS

The retweet ratio for both groups of users is shown in Table 1. The numbers in Table 1 correspond to the New Retweet ratio per user group for every day as defined by Equation 1. Bolded entries highlight the higher ratios for each day.

As seen in Table 1, for every day except June 28, bad news have consistently higher New Retweet ratio. The outlier of June 28 is the result of one or two tweets that had a sudden peak of activity. Furthermore, this outlier shows a difference of .036, which is significantly less than the 0.084 average difference on other days between the proportion spread of bad news and good news. The results of Table 1 indicate that bad news are subject to more active sharing than good news. Figure 1 depicts the same data and highlights that tweet activities reduce considerably 5 days after the date of creation. Over time, other studies have shown that the levels of retweets and

likes will stabilize to a generally fixed level for the rest of time [1].

Based on the above analysis, we can determine that bad news is more to be shared and spread on Twitter than good news in the context of the Brexit topic. Whether or not we can generalize this finding other topics or general human behavior is the subject of our future work.

## VII. USER CLASSIFICATION

The second hypothesis that we wanted to verify was whether there was an alignment between Trump and pro-Brexit supporters. Towards this goal, we searched through each tweet creator's account to identify whether he or she is a Trump supporter or not. Specifically, for each candidate tweet in our collection, we searched the corresponding creator's account for the keyword "Trump" over the last 5 months. Then, we manually identified whether or not a user is a Trump supporter. Candidate tweets associated with creators that cannot be classified as either Trump supporters or non-supporters are discarded. In order to verify the alignment of Trump supporter with Brexit supporter and Trump non-supporter with Brexit non-supporter, we scanned each tweet and its creator manually. The analysis confirmed that each tweet classified as Leave was created by a Trump supporter and each tweet labeled as Remain was not created by a Trump supporter. Although outliers are likely to exist, for all tweets and the corresponding creators we identified, no mismatches were found.

## VIII. FUTURE WORK

Much future work remains to be done on this topic in order to ensure that our results are not biased by demographic, cultural, or topic of interest.

In this paper, we have explored user sharing behavior for one topic, Brexit, between two user groups. We plan to analyze other topics in the same way to further strengthen our understanding of the subject.

### A. Topic Grouping

As Brexit cannot be universally classified as positive or negative, our current work has split users into two different groups. However, users may react to many headlines unanimously. For example, space exploration could be universally viewed as positive whereas mass shootings will trigger negative reactions. Therefore, we have no need to split users into categories. Instead, by selecting these highly biased topics, we can group topics

into positive or negative categories, then contrast general user reaction and sharing behavior with respect to these general categories.

### B. Sentiment Analysis

Our work touches upon sentiment analysis but we hope to further investigate the effect of sentiment on spread specifically. Since our tweet pool was of a reasonable size, we were able to classify a tweet as positive or negative by simply reading it and classifying it manually. In the future, we hope to rank how positive or negative a tweet is using automated natural language processing. Previously, proposed methods include sorting keywords into positive and negative lists, then summing up the score of a tweet based on keyword content [6]. Future work may investigate how the degree of a tweet's negativity rating affects its spread.

### C. Deceleration Rate

Our calculations of the New Retweet Ratio is essentially the velocity of spread. Another quantitative approach can be the *deceleration* of growth.

To do this, we can utilize the final number of retweets and likes for each user group. We first add all retweets and likes for every tweet within a user group at the end of the tracking period. After calculating this final total sum, we subtract the sum of initial retweets and initial likes that were already present prior to the beginning of the tracking period. This calculation will yield the total of retweets and likes that occurred from the start until the end of the tracking period. The proportion of New Retweets per day can then be given by the following equation.

$$\frac{NewRetweets}{(FinalRetweets + Likes) - (InitialRetweets + Likes)} \quad (2)$$

This equation is the ratio of activity on tweets that occur on each day. It can be used to evaluate whether or not spread of bad news decelerates slower than the spread of good news.

### D. Other Sources

A valid question may be whether or not our findings hold true for the general population. Studies have shown that the Twitter population with respect to the US population is diverse, but not necessarily representative [7]. Generally, minorities such as Black and Hispanic ethnic groups are over-represented while the White ethnic group is approaching an accurate proportion to its share in the United States [7]. Users are younger than the general average US population, but age groups in the range of 18-29 years old are on the decline [7].

However, an important note is that the distribution of topic discussion differs between age groups. For example, topics like immigration are overrepresented in discussion between age groups of 30-49 and 65+ years old compared to their proportion of total users on Twitter [7]. As such, general data demographic is

not necessarily the same as Twitter's demographic, and may be comparatively more representative of the actual population who are discussing the targeted topics.

Still, we hope to collect data from other social media and news sources to debunk any concerns about potential demographic-specific bias. However, the lack of data transparency may make this task difficult.

## IX. CONCLUSION

The importance of social media in the spread of news is becoming increasingly obvious, and allows us to explore user sharing behavior easily. Prior work has shown the effect of social influences and the degree of these influences with respect to various topics [8] as well as defined relationships between these influences on Twitter [3]. Yet, no work has shown the effect of positive or negative sentiment on social media spread or quantified the difference in sharing of news stories depending on their positive or negative classification.

In this study, we investigated the widely accepted claim that bad news spreads faster than good news. We showed that sentiment towards news does affect its rate of spread on social media. We hope to extend our investigation to different topics on other social media in order to strengthen our conclusion. More work is needed in order to conclusively ascertain that bad news travels faster than good news.

Our results do however suggest that sentiment does affect the spread of news stories on social media, but an overarching question that may be asked is *why* people like to read and share bad news. A guiding motivation to our study is to wonder if this liking to negativity is an inherent human quality or a result of constant exposure to bad news.

## ACKNOWLEDGMENT

This research was made possible with the support of the Indiana University-Purdue University Indianapolis Department of Computer Information and Information Science, as well as through funding from the National Science Foundation and the United States Department of Defense.

## REFERENCES

- [1] Lerman, Kristina, and Rumi Ghosh. "Information contagion: An empirical study of the spread of news on Digg and Twitter social networks." ICWSM 10 (2010): 90-97.
- [2] Romero, Daniel M., Brendan Meeder, and Jon Kleinberg. "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter." Proceedings of the 20th international conference on World wide web. ACM, 2011.
- [3] Ye, Shaozhi, and S. Felix Wu. "Measuring message propagation and social influence on Twitter. com." International Conference on Social Informatics. Springer Berlin Heidelberg, 2010.

- [4] Boyd, Danah, Scott Golder, and Gilad Lotan. "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter." System Sciences (HICSS), 2010 43rd Hawaii International Conference on. IEEE, 2010.
- [5] Team, R. Core. "R: A language and environment for statistical computing." (2013): 409.
- [6] Go, Alec, Richa Bhayani, and Lei Huang. "Twitter sentiment classification using distant supervision." CS224N Project Report, Stanford 1 (2009): 12.
- [7] Oktay, Huseyin, Aykut Firat, and Zeynep Ertem. "Demographic breakdown of twitter users: An analysis based on names." (2014).
- [8] Tang, Jie, et al. "Social influence analysis in large-scale networks." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009.