Issues in Information Systems
Volume 23, Issue 4, pp. 62-69, 2022

DOI: https://doi.org/10.48009/4_iis_2022_106

The impact of environmental news stories on twitter content

Joshua Maxwell, *Missouri University of Science and Technology, jnmx59@mst.edu*
Prosper Ayawah, *Missouri University of Science and Technology, peadp3@mst.edu*
Lauren Shaffer, *Missouri University of Science and Technology, lshc4@mst.edu*
Wen-Bin Yu, *Missouri University of Science and Technology, yuwen@mst.edu*
Bih-Ru Lea, *Missouri University of Science and Technology, leabi@mst.edu*

Abstract

There has been a lot of 'talk' about environmental issues and climate change lately in traditional media. Some of these issues are also thoroughly discussed on social media outlets. This study, therefore, hypothesized that there is a correlation between the traditional media news on environmental issues and Twitter discourse on the same. Text mining was used to extract keywords from news articles text posted on CNN and BBC within a certain period. These keywords were used to mine Twitter posts within the same period the articles were posted on the news outlets. It was observed that the densities of the identified topics from the news were, with varying strengths, correlated with the densities of tweets. Sometimes, a positive or negative lag was observed. The positive lag suggests that Twitter possibly spearheaded those topics before traditional media caught up, while the negative lag implies that Twitter is responding to the issues on the main media.

Keywords: Text mining, Twitter news, Environment, BBC news, CNN news

Introduction

Online news sites, such as CNN and BBC, are beginning to advocate the consequences of climate change. The Amazon rainforest is burning and the Arctic glaciers are melting, oceans are becoming more acidic and their levels are rising, and the mass extinction of plants and wildlife is looming. Consumer awareness is slowly encouraging a shift from fossil fuels to cleaner energy sources and more sustainable products and packaging. Protests are encouraging political leaders to endorse environmental policy changes and stricter Carbon emission regulations for a cleaner tomorrow.

Since the industrial revolution, humans have had a drastic impact on the health of the planet.. Environmentalists warn of nearing a 'tipping point.' Many plants and animal species are in danger of extinction. Animals are being killed by plastic consumption, a deer in Thailand consumed 15lbs of plastics, a whale consumed 80 plastic bags, and a turtle had 104 pieces of plastic found in an autopsy (Jacobo, 2019; Katz, 2018; Thongtep & Yongstar, 2019). In Zimbabwe, 55 elephants recently died after an extended drought. Some died while searching for water while others were killed by humans while searching for water in neighboring communities (Adebayo & Mavhunga, 2019).

Consumer awareness is driving many industries to produce more sustainable products and reduce or recycle single-use plastic. Maritime and aviation emissions account for approximately 6% of global carbon dioxide output (Gerretsen, 2019). Companies in both industries are working to lower emissions. British Airways has pledged to offset domestic flight carbon emissions by investing in solar energy and planting trees in South America, Africa, and Asia (Riley, 2019).

Renewable energy sources are projected to grow by 50% according to the International Energy Agency (Egan, 2019a). Coal is still the most prominent source of power through 2024, but solar, wind, and natural gas are gaining market share (Egan, 2019b).Carlsberg is working on producing paper-fiber beer bottles, which are more environmentally friendly than glass and aluminum because they are sustainably sourced. Absolut, Coca-Cola, and L'Oréal have also begun the development of paper packaging (Wiener-Bronner, 2019).

"The older generations are failing us, and the political leaders are failing us, but we will be watching and holding them accountable," Greta Thunberg announced to a Denver crowd (Tabachnik, 2019). She began skipping school in August 2018 to stand outside of the Swedish Parliament; this initiated a worldwide movement where students began walking out of schools in over 100 countries (Tabachnik, 2019). As Greta Thunberg says, 'whether you like it or not, change is coming.' Industry leaders are recognizing the growing concern for the planet and initiating changes before regulations force them to do so. From airlines and shipping to reducing plastic production and a Swedish teenager rallying students across the globe to strike for the climate, change is indeed coming.

News Impact on Social Media Output

Twitter and other channels of social media have begun to play an important role in defining trends, consumer behavior, and even tracking health epidemics. As print media evolves, more individuals are turning to social media as their source for breaking headline news. Twitter can act as a place for individuals to voice their opinions, raise awareness and make newsworthy posts. Most Tweets are not individually rich in information; however, an aggregate of Tweets can provide insight into a population (Paul & Dredze, 2011).

Social media allows users to link to others that are similar to their ideas and beliefs or to someone whose ideas they find interesting (Lerman & Ghosh, 2010). The dynamics of information sharing on these platforms may be different than in the general population or using a larger, more conventional news outlet. Tweets consist of short messages that may contain links to news stories, or retweets of stories with a comment from the users. A known limitation for any research using the Twitter platform is that the maximum number of characters a user is permitted to type is 140 (Barbosa & Feng, 2010). Lerman & Ghosh (2010) discovered that news on Twitter travels slower but farther into the social network and continues to do so at this same rate as the story ages, in comparison with Digg, where the news story may generate a larger audience response. In 2015, a study was conducted using Twitter to track and gauge the perceived risk of Ebola in the United States, concluding that Twitter was a valuable tool for public health communication (Lazard et al., 2015).

From this research, it seems that traditional news should impact Twitter content. We hypothesize, for a given topic,
1. There exists some correlation between News activity and Twitter activity.
2. There exists a characteristic lag between the initial occurrence of the News Activity and Twitter activity.

Methodology

The data collection and analysis were executed using R. The analysis was divided our analysis into three phases,
• Extract Keywords that uniquely identify a specific topic from a corpus of News documents.
• Use extracted keywords to mine Twitter for relevant tweets.
• Compare and contrast the frequency at which the News and Twitter discussed the topic.

News Collection, Topic Identification, and Keyword Extraction

We first manually collected a corpus of news articles from two popular News providers' websites, CNN and BBC. We chose BBC for its increased propensity to cover global news stories and we chose CNN for its large US viewership. Our selection criteria were that each article had to be environmentally related and published within the last seven days. We collected a total of 102 distinct CNN articles and 51 distinct BBC articles. Next, we cleaned and processed the text data from the News documents. To preprocess the data, we removed numbers, removed stop-words, and converted all text to lower case (Uysal & Gunal, 2014). To avoid the risk of over-stemming or under-stemming, we refrained from stemming (Paice, 1994). The lack of stemming did not seem to impact the quality of our topic identification or keyword extraction. We then used a Latent Dirichlet Allocation (LDA) algorithm (Lu, Ott, Cardie, & Tsou, 2011; Phan, Nguyen, & Horiguchi, 2008) to cluster the News documents into topics and identify keywords. However, before we could use this algorithm, we needed to identify the number of topics present in each corpus. We used several different metrics to identify the number of topics in each collection. We found the Arun et al. (2010) and Cao et al. (2009) metrics were the most instructive. Using these metrics, we determined that the number of topics in the CNN and BBC collections was six and eight, respectively. Now, with the number of topics determined, the documents sorted into those topics, and keywords within each topic identified, we moved on to mining Twitter.

Twitter Mining

The LDA procedure for identifying keywords yields an ordered list of terms for each topic. Each term has an associated probability that measures the likelihood that the topic generated the term. To search Twitter, we had to select the terms that we would use to query Twitter. To select the search terms, we had to establish a threshold integer n, by which we could select the n most likely terms for each topic and search for tweets that contained all n terms. A large value of n would be much too strict and diminish the number of tweets the query would return, while a small n would be too general and return tweets off-topic. Through trial and error, we found a value for n of three to be suitable. Our Twitter search for tweets with each of the 3 keywords previously identified yielded 3188 tweets.

Data Analysis

Once we had the timestamped News and Twitter data sorted by topic, we began analyzing the data. We chose to overlay kernel density estimations (KDE) to compare the data sets. We chose KDE to account for two characteristics in the data. First, the News data timestamps were accurate up to the day, while Twitter data timestamps were in seconds. Being that it was not the case that the News agencies wrote and published all their articles at the same time every day, we needed a method that smoothed out the data for a fair comparison. Second, because the query returned fewer tweets than the maximum limit, we believe that our query of tweets was complete; however, we cannot be sure. Therefore, we wanted to smooth out the data to account for the possibility of missing data and maintain a fair caparison. To improve the accuracy of the KDE, we dropped topics that had fewer than three observations of either News data or Twitter data. The omission of the sparse topics was not much of a burden, as most topics had greater than five observations in the News data and many more in the Twitter data.

Results

Initially, we hypothesized that the traditional news outlets, CNN and BBC would have a characteristic lag with respect to Twitter. A pronounced lag occurred on several topics, but it was not a consistent finding throughout the evaluation. For the CNN data, Twitter responses for energy/emissions (1) and plastics/waste

(4) did lag News responses; this was in line with our hypothesis. There was an unexpected lack of lag between Twitter and CNN activity on the water/Venice (6) topic (Figure 1).

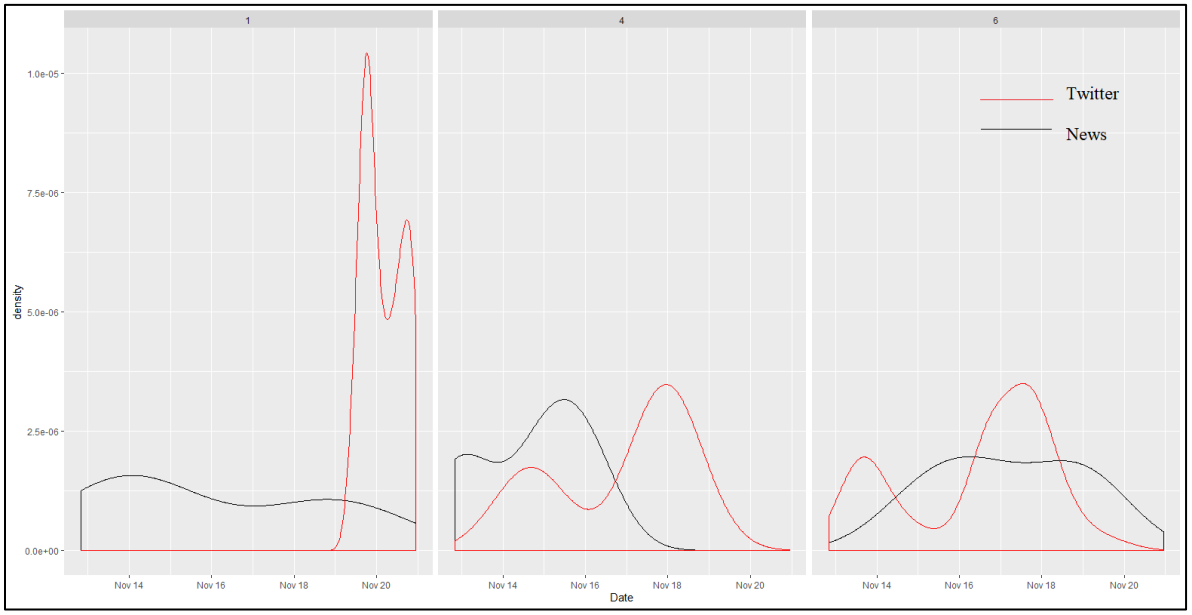

Figure 1: Kernel density estimation plot of CNN & Twitter correlation

The BBC-Twitter relationship yielded similar results. The topics of Wales/green (4), bears/population(5), and company/climate/emissions (8) all had a clear correlation between Twitter activity and News output; however, only topic (8) had the expected lag. The coal/China (1) and climate/change (7) topics unexpectedly had a lack of correlation (Figure 2). Our data, with some exceptions, seem to support our first hypothesis.

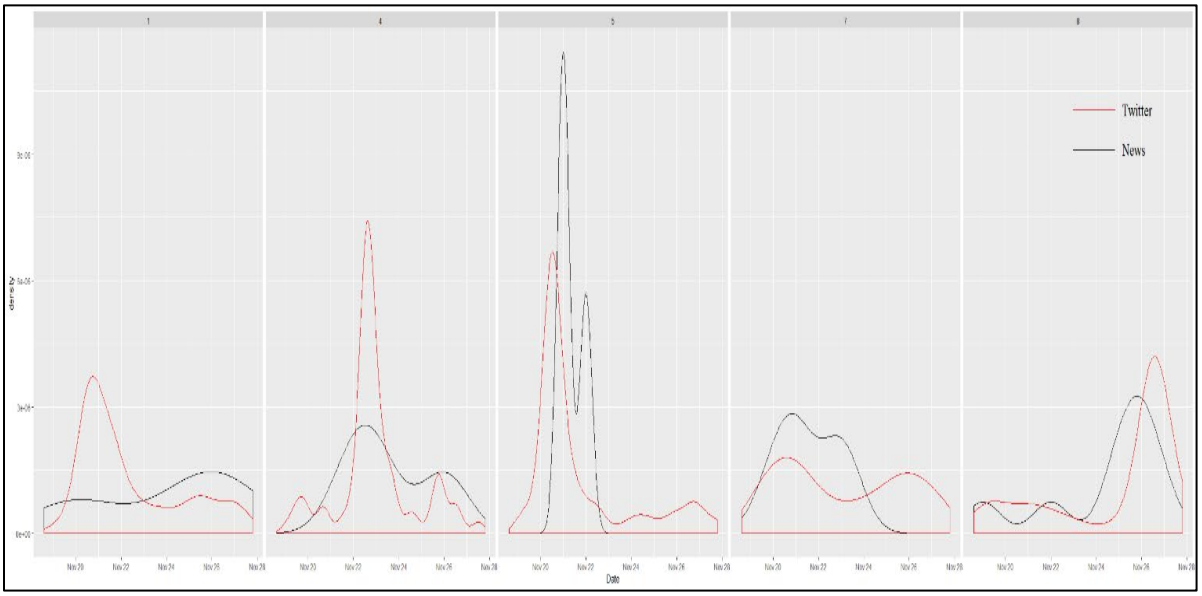

Figure 2: Kernel density estimation plot of BBC & Twitter correlation

It has been observed that false news spreads much faster and further on social media than true news (Soroush Vosoughi et al, 2018). The increased virality of false news on social media persists even when bot activity is accounted for, Soroush et al attribute this increased transmission to the enhanced novelty of false news stories compared to true news stories. Differences in novelty and sensationalism can help explain why some of the topics have a pronounced lag and some do not. For example, the topic Wales/green (4) has increased novelty due to the relative obscurity of Wales and the topic bears/population (5) concerns bear attacks which are inherently sensational. Another factor to consider is the influence of other news agencies, if another news agency broke the story significantly earlier than BBC or CNN then our window of Twitter activity may have missed the conversation.

Discussions and Conclusions

Popularity does not always equate to quality on social media platforms (Salganik, Dodds, & Watts, 2006). Certain users or influences may have a large following but not put out quality content especially regarding news reporting. Depending on an individual's network, the news is also filtered depending on their interests or who they follow. This creates a filter that may limit exposure to certain categories of news and interests altogether.

"Big Media has lost its monopoly on the news. Now that it is possible to publish in real-time to a worldwide audience, a new breed of grassroots journalists are taking the news into their own hands" (Newman, 2011).

The lines between professional journalists and citizens journalism/bloggers have become increasingly blurred. Professional journalists have adapted to the trends by beginning blogs, joining social media and interacting with the online communities. An underlying, increasingly popular trend is for individuals to let social media filter the news that 'comes to me' (Newman, 2011). Depending on who you follow, and how you create your social media network, the topics that come into your newsfeed may be limited in scope.

In 2009, after a failed attempt to blow up a United States commercial flight, Twitter was the first source to release upgraded security standards for international flights (Lerman & Ghosh, 2010). This shows that news channels such as CNN and BBC can be beaten by Twitter to post a breaking story. Twitter and other social media may also lead to a 'burst' bias where a tweet or post affects others. As a result, a few tweets about a topic may lead to a burst of tweets on that same topic (Aramaki, Maskawa, & Morita, 2011).

Additional research may be necessary to better define the relationship between traditional news channels and social media. Most users rarely tweet, but the most prolific 10% create 80% of tweets from adult U.S. users (Wojcik & Hughes, 2019). Increasing the size and diversity of the News providers, News topics, time period, and Twitter sample size would help define the cases when the expected causal relationship holds and when it fails. This study used a manual collection of News articles, web scraping would greatly enhance the sample size. As this was a purely observational study, even with a larger sample size proving a bona fide causal relationship is difficult. However, future studies could do a more controlled study by creating the news article first, releasing it into social media, and then using keywords to track the resulting activity. This method would be more conducive to A/B testing.

References

Adebayo, B., & Mavhunga, C. S. (2019). Zimbabwe says 55 elephants have died in two months following severe drought - CNN. Retrieved October 24, 2019, from https://www.cnn.com/2019/10/21/africa/zimbabwe-elephant-drought-starvation/index.html

Aramaki, E., Maskawa, S., & Morita, M. (2011). Twitter Catches The Flu : Detecting Influenza Epidemics using Twitter The University of Tokyo The University of Tokyo National Institute of. In Computational Linguistics (Vol. 2011, pp. 1568–1576). Retrieved from http://www.aclweb.org/anthology/D11-1145

Arun, R., Suresh, V., Madhavan, C. E. V., & Murty, M. N. (2010). On finding the natural number of topics with Latent Dirichlet Allocation: Some observations. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 6118 LNAI, pp. 391–402). https://doi.org/10.1007/978-3-642-13657-3_43

Barbosa, L., & Feng, J. (2010). Robust Sentiment Detection on Twitter from Biased and Noisy Data. Retrieved from http://twittersentiment.appspot.com/

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. Neurocomputing, 72(7–9), 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Egan, M. (2019a). Renewable energy is booming. But it's not growing fast enough to fight climate change. Retrieved October 24, 2019, from https://www.cnn.com/2019/10/21/business/renewable-energy-solar-iea/index.html

Egan, M. (2019b). Renewables set for "meteoric" growth. But that's not enough to fight climate change - CNN. Retrieved October 24, 2019, from https://www.cnn.com/2019/10/21/business/renewable-energy-solar-iea/index.html

Gerretsen, I. (2019). Shipping is one of the dirtiest industries. Now it's trying to clean up its act. Retrieved October 24, 2019, from https://www.cnn.com/2019/10/03/business/global-shipping-climate-crisis-intl/index.html

Jacobo, J. (2019). 104 pieces of plastic found in baby sea turtle that washed ashore in Florida - ABC News. BBC News. Retrieved from https://abcnews.go.com/US/104-pieces-plastic-found-baby-sea-turtle-washed/story?id=66096881

Katz, B. (2018). Whale Dies in Thailand With 80 Plastic Bags in Its Stomach | Smart News | Smithsonian. Retrieved December 11, 2019, from https://www.smithsonianmag.com/smart-news/whale-dies-thailand-80-plastic-bags-its-stomach-180969232/

Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., & Suran, M. (2015). Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. American Journal of Infection Control, 43(10), 1109–1111. https://doi.org/10.1016/j.ajic.2015.05.025

Lerman, K., & Ghosh, R. (2010). Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks. Retrieved from https://networkchallenge.darpa.mil

Lu, B., Ott, M., Cardie, C., & Tsou, B. K. (2011). Multi-aspect sentiment analysis with topic models. In Proceedings - IEEE International Conference on Data Mining, ICDM (pp. 81–88). https://doi.org/10.1109/ICDMW.2011.125

Mackintosh, E. (2019). London becomes first city to ban Extinction Rebellion protests - CNN. Retrieved October 24, 2019, from https://edition.cnn.com/2019/10/15/uk/extinction-rebellion-london-ban-gbr-intl/index.html

Newman, N. (2011). Mainstream media and the distribution of news in the age of social media. Retrieved from https://ora.ox.ac.uk/objects/uuid:94164da6-9150-4938-8996-badfdef6b507

Newton, P. (2019). Buckle up -- Canada's election will be a cliffhanger. Retrieved October 24, 2019, from https://www.cnn.com/2019/10/20/world/canada-election-october-21-intl/index.html

Paice, C. D. (1994). An evaluation method for stemming algorithms. In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1994 (pp. 42–50). Association for Computing Machinery, Inc. https://doi.org/10.1007/978-1-4471-2099-5_5

Paul, M. J., & Dredze, M. (2011). You Are What You Tweet: Analyzing Twitter for Public Health. Retrieved from www.aaai.org

Phan, X. H., Nguyen, L. M., & Horiguchi, S. (2008). Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08 (pp. 91–99). https://doi.org/10.1145/1367497.1367510

Riley, C. (2019). Airlines have a climate problem. Now one company is promising carbon-neutral flying by 2050. Retrieved October 24, 2019, from https://www.cnn.com/2019/10/10/business/british-airways-iag-climate-change/index.html

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. Science, 311(5762), 854–856. https://doi.org/10.1126/science.1121066

Tabachnik, S. (2019). Greta Thunberg in Denver: Teen climate activist leads rally in Civic Center Park. The Denver Post. Retrieved from https://www.denverpost.com/2019/10/11/greta-thunberg-denver-climate-strike/

Thongtep, W., & Yongstar, S. (2019). Thai deer found dead with 7kg of "underwear, plastic bags" in stomach - BBC News. BBC News. Retrieved from https://www.bbc.com/news/world-asia-50554911

Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. Information Processing and Management, 50(1), 104–112. https://doi.org/10.1016/j.ipm.2013.08.006

Wiener-Bronner, D. (2019). Carlsberg is working on beer bottles made of paper - CNN. Retrieved October 24, 2019, from https://www.cnn.com/2019/10/10/business/carlsberg-paper-beer-bottles/index.html

Wojcik, S., & Hughes, A. (2019). Sizing Up Twitter Users FOR MEDIA OR OTHER INQUIRIES. Retrieved from www.pewresearch.org.

Soroush Vosoughi, Deb Roy & Sinan Aral (2018) The spread of true and false news online. Science Vol 359, issue 6380. https://doi.org/10.1126/science.aap9559