

From Tweeting About News to Creating News Around Tweets: Characterizing Tweets Embedded in News Articles

Maurício Gruppi¹, Sibel Adali¹, Matthew Salemi¹, and Benjamin D. Horne²

¹ Rensselaer Polytechnic Institute, ² University of Tennessee Knoxville
{gouvem, adalis, salemm2}@rpi.edu, bhorne6@utk.edu

Abstract

Content from Twitter has been largely used as source material in news articles, journalists may use tweets to support a claim or to estimate public opinion. Such content often features in news articles in the form of embedded tweets. Despite the prominence of Twitter in journalism, little work has been done towards characterizing the use of tweets in news articles. We present a study to describe the utilization of tweets by news sources of different credibility levels. Specifically, we show the differences in tweets embedded by reliable and unreliable sources in terms of quantity and quality, the topics they cover, and the individuals they cite. Our results exhibit quantitative differences between the sources of distinct credibility. They also point to discrepancies in the narratives that arise around the same tweet due to the different intentions of each news source.

Introduction

Social media and journalism have had a love-hate relationship over the past decade. On one hand, social media became a key source of reporting, allowing journalist to quickly and easily provide context around developing events, as well as the ability to break news quickly and directly to a target audience (Broersma and Graham 2013; Vis 2013; Papacharissi and de Fatima Oliveira 2012). On the other, changes in journalistic norms due to social media have had negative consequences, including the ease of breaking information before it is fully verified (Hermida 2012), the mixing of opinions and facts on journalist’s social media profiles (Vis 2013), and the rise of untrained, alternative “journalists” who challenge the authority of traditional journalism (Gillmor 2004).

One such new practice brought on by the social media era is the embedding of social media content into news articles. This is most often done with content from the platform Twitter. Specifically, tweets are often embedded into news articles as either sources of reporting, to provide additional context, or to estimate public opinion (McGregor 2019). A study by Broersma and Graham (2013) showed a steep increase in the number of tweets included in newspaper content from 2010 through 2013 in tabloid newspapers from the United Kingdom and the Netherlands. Similarly, Heravi and Harrower (2016) conducted a study to identify the most

common uses of social media in newspapers. Their findings showed that most journalists used Twitter for finding leads and 55% used it for sourcing content, although most of them did not consider social media to be trustworthy.

Despite the significant role of Twitter in journalism and the large body of literature around that role, there has been little work characterizing Twitter’s use in news articles (with the exception of (Broersma and Graham 2013)), and to the best of our knowledge, there have been no works on Twitter’s use in articles from unreliable news outlets. Hence, this short paper aims to begin filling this gap. Namely, we ask four basic questions to characterize tweets embedded in news articles across both reliable and unreliable news outlets:

- Q1. Has the presence of Twitter-based content in news articles increased in recent years?
- Q2. Are tweets embedded into news articles more by reliable or unreliable news outlets?
- Q3. What Twitter users are most frequently embedded in news articles? Do these sets of users differ across the veracity of outlets?
- Q4. What are some examples of how Twitter-based content is used by reliable and unreliable outlets?

To answer these questions, we leverage a unique dataset of news articles and tweets embedded in those news articles over the years 2018, 2019, and 2020. Additionally, we leverage news outlet veracity labels to examine how the practice of embedding tweets differs across news veracity. With this data, we perform a mixed-methods analysis and provide concrete examples of characterized behavior.

We find that while the use of embedded tweets in news articles has increased for both reliable and unreliable outlets, unreliable sources use significantly more Twitter-based content than reliable sources. Furthermore, the two groups often cite different types of Twitter users. However, this is not always the case. In particular, we note a unique case in which the same tweet is used by both types of outlets, but for different purposes.

Data

To answer our proposed research questions, we combine three datasets: the NELA-GT-2018, NELA-GT-2019, and

NELA-GT-2020 datasets (Nørregaard, Horne, and Adali 2019; Gruppi, Horne, and Adali 2020, 2021), and extract articles that contain embedded tweets. The main components of these three datasets are as follows:

- **News articles:** articles published by news sources. It contains title, body text, source, publication date, and may contain multiple embedded tweets.
- **Embedded tweets:** posts from Twitter inserted into the contents of a news article. Its attributes are URL, content, author, and date.
- **Source credibility labels:** annotated information about the credibility of news sources, with sources being labeled as reliable or unreliable.

These combined datasets contain 3,710,005 articles, 300,025 of which have embedded tweets. As these three datasets each contain nearly all articles published by each outlet (see (Gruppi, Horne, and Adali 2021)), we can confidently quantify the use of embedded tweets by each outlet.

Results and Discussion

Volume of Twitter-based content in news has increased in recent years The number of embedded Tweets observed in recent years has increased, as shown in Figure 1. As seen in Figure 1, the number of published articles has increased since 2018 and the number of embedded tweets has followed that trend. The ratio of articles published in 2020 with at least one embedded tweet is approximately 22%.

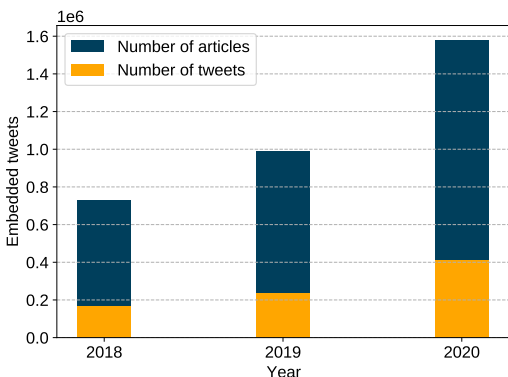


Figure 1: Number of embedded tweets (in millions) in the news articles for the 2018, 2019 and 2020 versions of the NELA dataset.

Unreliable news sources cite considerably more content from Twitter than reliable ones When analyzing the proportion of articles with embedded tweets in reliable and unreliable sources, respectively, a considerable difference can be observed. From 2018 through 2020, the ratio of articles with embedded tweets among reliable sources has dropped from approximately 10% down to 5%, while that of unreliable sources remained more stable with approximately 25% of its news articles containing one or more embedded tweets, as seen in Figure 2.

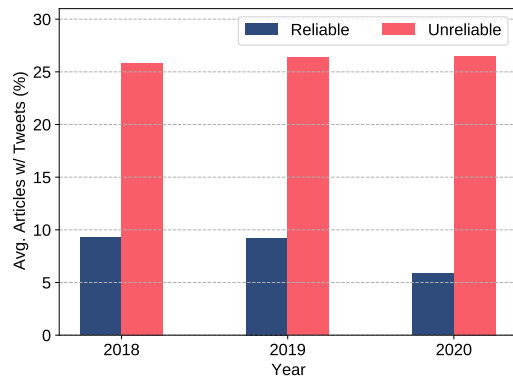


Figure 2: Average percentage of articles containing at least one embedded tweet in the Reliable and Unreliable news source classes. It can be seen that while the unreliable sources maintained a ratio of approximately 25%, the reliable sources decreased this ratio to roughly 5%.

Unreliable news sources cite less credible Twitter accounts more often The fact that a tweet by a questionable user appears in a news article is not exclusively a behavior of unreliable sources. However, our results show that these two source groups exhibit distinct behaviors in the quantity and quality of the tweets they cite. Unreliable sources have a slightly higher chance to cite a tweet by an user with a small number of followers ($p < 0.05$) (see Figure 3).

Moreover, reliable sources cited a total of 13960 users, out of which 529 were banned as of March 2021. Unreliable sources cited a total of 74924 user out of which 5088 were banned as of March 2021. Finally, 41% of the users cited by reliable sources were accounts verified by Twitter, against 14% cited by unreliable. The most cited users by reliable and unreliable sources are shown in Table 1.

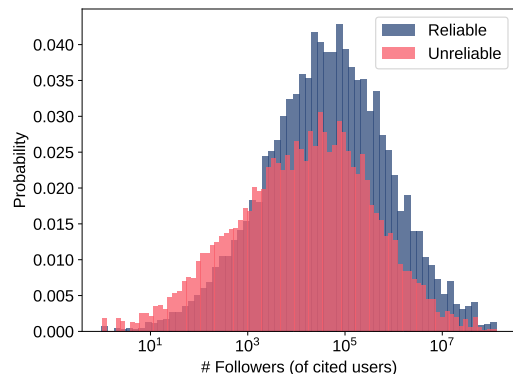


Figure 3: Distributions of the numbers of followers that are cited in news articles by reliable and unreliable sources. This result shows that unreliable news sources are slightly more likely to cite tweets from accounts with lower follower numbers, such as bots and less influential individuals.

Reliable			Unreliable		
Username	Articles	Followers	Username	Articles	Followers
realDonaldTrump	1822	<i>Suspended</i>	realDonaldTrump	8238	<i>Suspended</i>
atrupar	486	657909	SecretNews	6986	1419
mims	271	98390	MrAndyNgo	2143	762481
TPMLiveWire	177	9431	BreitbartLondon	1260	86929
TPM	160	280127	stillgray	965	241328
JoeBiden	136	29139233	atrupar	903	657909
elonmusk	93	48523294	dauidicke	849	<i>Suspended</i>

Reliable			Unreliable		
Username	Sources (%)	Followers	Username	Sources (%)	Followers
realDonaldTrump	10	<i>Suspended</i>	realDonaldTrump	26	<i>Suspended</i>
JoeBiden	6	29139233	MrAndyNgo	19	762481
NYGovCuomo	4	2630017	DailyCaller	17	742341
thehill	4	4141371	ABC	17	16559317
atrupar	4	657909	Breaking911	17	817097
SpeakerPelosi	4	7067914	thehill	17	4141371
senatemajldr	4	<i>Suspended</i>	CBSNews	17	8073550

Table 1: Twitter users cited in most articles (top row) and users cited by most sources (bottom row) and their number of followers on Twitter as of March 2021. Users suspended before March 2021 are not shown with the exception of *realDonaldTrump*.

Examples of Embedded Tweet Use To better understand the use of embedded tweets in news and to highlight paths for future research. We qualitatively examined articles related to Covid-19. To extract Covid-19 articles from the NELA-2020 dataset, we filtered using a set of Covid-19 related keywords. We started with a list of keywords used by the COR-19 Dataset (Wang et al. 2020) and then expanded this list by reviewing a random sample of articles returned by the initial keyword list. In total, this list contained 16 keywords and returned 40,932 articles that had embedded tweets.

Our data set contains many articles that are collections of tweets used to summarize public or expert opinion on an event or news. There are also many tweets that are embedded in many news articles over time. In this section, we introduce two case studies of such highly referenced tweets.

Case Study 1: Alternative Narratives Around Tweets

The quantitative analysis shows substantial differences in the tweet citation patterns from reliable and unreliable news sources. Beyond these quantitative aspects, we show, in a qualitative analysis, how narratives are constructed around tweets in news articles from each of the source classes.

In a tweet published by the official Twitter account of the World Health Organization (WHO) on January 14, 2020¹, they state:

“Preliminary investigations conducted by the Chinese authorities have found no clear evidence of human-to-human transmission of the novel #coronavirus (2019-nCoV) identified in #Wuhan, #ChinaFlag of China”.

This tweet was embedded into 116 news articles being used to source various narratives across news outlets.

¹<https://twitter.com/WHO/status/1217043229427761152>

Narratives used by reliable sources include criticism of controversial statements given by the WHO, such as the one in an article by The Washington Post published on June 10, 2020, titled “THE HEALTH 202: WHO HAS STUMBLER REPEATEDLY IN COMMUNICATING ABOUT THE CORONAVIRUS”. On July 8, 2020, the British Broadcast Corporation (BBC), published an article describing the charges then-president Donald Trump had against the WHO, justifying that the organization had failed in obtaining and sharing information in a timely manner (“CORONAVIRUS: WHAT ARE PRESIDENT TRUMP’S CHARGES AGAINST THE WHO?”).

The narratives shared by unreliable sources tend to imply that the WHO deliberately attempted to promote Chinese Government Propaganda, as seen in articles from two unreliable sources Breitbart (“WHO SPREAD FALSE CHINESE GOVERNMENT PROPAGANDA: CORONAVIRUS NOT CONTAGIOUS AMONG HUMANS”) and The Gateway Pundit (“PRESIDENT TRUMP HAS GOOD REASON NOT TO TRUST THE COVID-19 ‘EXPERTS’ – AMERICA WAS LIED TO ABOUT THE CHINA CORONAVIRUS AND WE STILL ARE”).

These examples show how news producers may use content from Twitter as source material to their articles while constructing narratives that best suit their agenda. Tweets can be directly inserted into news articles in an attempt to directly cite people of interest using narratives to convey a distorted message.

Case Study 2: Combating Disinformation Campaigns

Our second case study involves a tweet by a Chinese government official from March 12, 2020, introducing disinformation about the origins of Covid-19 and trying to attribute it to US. Note that state-affiliated media account labels were not visible at the time of this tweet. This tweet has been liked 14K times and retweeted 6,888 times, cited by 40 articles from 25 different sources in our dataset. Twitter currently

has a warning sign for this tweet and links to a news item with title: “WHO says evidence suggests COVID-19 originated in animals and was not produced in a lab”.

First common narrative in the news articles reports this tweet as misinformation. Many unreliable and conspiracy sources report that this tweet is incorrect and provide information discrediting the author (“US Army Behind Covid-19 in Wuhan’: China’s Foreign Ministry Levels Bombastic Charge”). Even though the tone of these stories are more sensationalist than reliable sources, they provide even more detail regarding the official response to this tweet (“US summons China’s ambassador over comments on coronavirus conspiracy”) and also criticize China’s initial missteps in handling the pandemic and cite multiple other sources including CNN. One highly unreliable source cites scientific arguments debunking the claim using an article copied from a different source outside of our collection. Only two of the sources that cover this tweet are reliable. It is possible that other sources have covered this information without linking to this tweet. There are 15 reliable sources reporting to the author of this tweet in the same month without linking to this tweet. This may be the result of a deliberate policy not to link to misinformation to avoid its amplification.

The second main narrative involves criticism of Twitter for not flagging this tweet as false while fact-checking others (“Twitter Says China Claiming Coronavirus Started in U.S. Not a Violation of Rules”) or flagging it too late (“Twitter Retroactively Fact-Checks Chinese Official’s Tweet Months Later”). This tweet is used in further criticism of Twitter’s uneven enforcement of their rules (“Twitter Fact Checks President Trump - But Not Communist China”).

Additionally, one conspiracy source, Veteran’s Today covers this information as truth, even fact checking Donald Trump who debunks the information in this tweet. It is the only source in our collection to do so. The story is also used in narratives against other targets, for example, to attack credibility and independence of World Health Organization (“Is the World Health Organization Merely a Bunch of Incompetent Goobers or Is It Wholly Owned by China?”). It is used to attack people critical of anti-Asian statements by President Trump or to claim that Democrats believe this claim (“Donald Trump Jr. Slams Democrats & Media Who Believe China On Coronavirus: ‘Get Your Heads Examined’”) or against Chinese Government in general in some conspiracy sources (“China Purges Major U.S. Media Outlets From Country Amid Coronavirus Information War”).

Conclusions and Future Work

We presented an analysis of embedded tweets in the news landscape, comparing the presence of Twitter content in news articles published by reliable and unreliable news sources. We use the data from the 2018, 2019, and 2020 versions of the NELA-GT dataset. Our first set of results showed that the presence of Twitter content in news has increased in recent years. Additionally, we showed that reliable and unreliable sources exhibit significant differences with respect to the volume of tweets embedded in their articles, with unreliable publishing tweets in about 25% of its articles against less than 10% in reliable sources. We also

demonstrated how unreliable sources tend to include tweets by less credible individuals from Twitter. Finally, we showed a qualitative description of the approaches news sources may take to introduce new narratives using tweets as source material. The same tweet may be used by a source to propagate legitimate information, and by a malicious source who attempts to spread disinformation by distorting the narratives around statements made on Twitter by politicians, government officials and other influential people.

In future work, we would like to study and develop a principled method to identify and characterize multiple narratives around a tweet. This may require complete comprehension of the context in which the tweet appears, as well as an understanding of the news ecosystem as a whole.

References

- Broersma, M.; and Graham, T. 2013. Twitter as a news source: How Dutch and British newspapers used tweets in their news coverage, 2007–2011. *Journalism practice* 7(4): 446–464.
- Gillmor, D. 2004. We the media: The rise of citizen journalists. *National Civic Review* 93(3): 58–63.
- Gruppi, M.; Horne, B. D.; and Adalı, S. 2020. NELA-GT-2019: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. *arXiv e-prints* arXiv:2003.
- Gruppi, M.; Horne, B. D.; and Adalı, S. 2021. NELA-GT-2020: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. *arXiv preprint arXiv:2102.04567*.
- Heravi, B. R.; and Harrower, N. 2016. Twitter journalism in Ireland: sourcing and trust in the age of social media. *Information, communication & society* 19(9): 1194–1213.
- Hermida, A. 2012. Tweets and truth: Journalism as a discipline of collaborative verification. *Journalism Practice* 6(5-6): 659–668.
- McGregor, S. C. 2019. Social media as public opinion: How journalists use social media to represent public opinion. *Journalism* 20(8): 1070–1086.
- Nørregaard, J.; Horne, B. D.; and Adalı, S. 2019. Nela-gt-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 630–638.
- Papacharissi, Z.; and de Fatima Oliveira, M. 2012. Affective news and networked publics: The rhythms of news storytelling on# Egypt. *Journal of communication* 62(2): 266–282.
- Vis, F. 2013. Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots. *Digital journalism* 1(1): 27–47.
- Wang, L. L.; Lo, K.; Chandrasekhar, Y.; Reas, R.; Yang, J.; Eide, D.; Funk, K.; Kinney, R.; Liu, Z.; Merrill, W.; et al. 2020. Cord-19: The covid-19 open research dataset. *ArXiv*.