
A SYSTEMATIC APPROACH TO ANALYZING SYNDICATION OF VIRAL NEWS STORIES

A PREPRINT

Gaurav Deshpande
Institute for Software Research
Carnegie Mellon University
Pittsburgh, PA 15213
gdeshpan@andrew.cmu.edu

Sam E. Teplov
Institute for Software Research
Carnegie Mellon University
Pittsburgh, PA 15213
steplov@andrew.cmu.edu

Alon Peer
Department of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
apeer@andrew.cmu.edu

Dr. Nicolas Christin
Department of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
nicolasc@andrew.cmu.edu

December 4, 2019

ABSTRACT

In this work, we present a novel, semiautonomous methodology for analyzing syndicated news articles. We first detail our manual process of news syndication analysis from which we observe a number of unsettling patterns. From there, we go on to categorize and define two different forms of syndication: top-down and bottom-up. We then transform our manual process into a semiautomated one which combines web scraping and machine learning techniques to create and then analyze a dataset of viral news stories. From the analysis, we are able to make some inferences about the nature of syndicated news articles, as well as trace back a viral news story to its original publisher and article. By presenting this work, we hope to minimize the effects that blind news syndication has on the spread of misinformation.

Keywords Misinformation · News syndication · Fake news

1 Introduction

News are everywhere- television, newspapers, social media. We consume it whether we like it or not. But the question is, how do we know which news are true, and which are false?

Fake news has been a real issue since the spark of social media, especially in the recent years, with its spike coming during the 2016 U.S. presidential election. In fact, a study from Ohio State University shows that fake news had a significant impact on the results of the elections. The researchers found a correlation between the amount of Obama voters who believed fake news about Clinton and the amount of them who chose to vote for Trump. At that time, the media was swarmed with fake news, and one could hardly blame people who couldn't distinguish them from real news.

But what exactly is fake news? According to Wikipedia, "fake news is a form of news consisting of deliberate disinformation or hoaxes spread via traditional news media (print and broadcast) or online social media". In addition to this definition, fake news may also include clickbait stories, propaganda, satire, biased news, or even sometimes just bad journalism.

These days, anyone can publish content on the internet, whether it is on a website, a blog, or a social media profile, and potentially reach large audiences. The fact that people nowadays rely on the internet to consume news doesn't make this problem any easier to solve. Fake news is very profitable for the publisher, as viral stories generate large

amounts of traffic, thus increasing advertisement revenue. For that reason, fake news are widely common around the internet. Another motive to publish fake news is to promote a political agenda. Fake political news stories are meant to persuade consumers to accept a biased or false beliefs (like the fake news regarding Hillary Clinton during the 2016 U.S. presidential election). This is the main reason detecting fake news is such a challenge. These false stories were created to sound true and interesting.

A different variation of fake news is syndication. A website can take a real news story, and change it a bit, so it will fit that website's beliefs. A news story can be syndicated in various ways- changing the title, adding interpretations that aren't necessarily true, and much more. A popular form of syndication is news stories that are posted by local newspapers and websites, and as they get viral, these stories get picked up by bigger and more popular news websites. On the way from the local news source to the national (or international) news website, the story can get syndicated. Sometimes, it's not necessarily on purpose, but it happens due to the lack of due diligence by the mainstream media.

One can try and detect fake news by checking the source of the story, read beyond the title (which is intended to attract users and thus contains more interesting and controversial topics), try to trace the article back to the original source, or even make sure the article is not a satire by any chance. Even though it's possible, doing so is not an easy task, and that is why there has been a lot of work put into automating this process. A step towards this, is detecting and analyzing the syndication of viral news stories. That is, how stories change through time and the difference between articles regarding the same stories but posted in different websites.

2 Background

3 Related Work

[1]

4 Methodology

5 Analysis

6 Contributions

7 Future Work

8 Conclusion

References

- [1] Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. On the origins of memes by means of fringe web communities. In *Proceedings of the Internet Measurement Conference 2018*, pages 188–202. ACM, 2018.