Narrative Trends of COVID-19 Misinformation

Thomas Marcoux
txmarcoux@ualr.edu

Nitin Agarwal
nxagarwal@ualr.edu

University of Arkansas at Little Rock, Little Rock AR 72204, USA

Abstract

The COVID-19 crisis has seen the rise of many harmful online narratives. These narratives have seeped into the real world and pose tangible health risks. For this reason, we have leveraged existing techniques and developed tools to further our understanding of online misinformation and its dynamics. This provides policy makers with more tools to sift through otherwise impossibly large data sets. To ensure this, we worked closely with the Arkansas Attorney General.

1 Introduction

One year after the appearance of the novel COVID-19 virus, the U.S. holds the highest number of COVID-19 cases and deaths in the world, behind countries like India with over ten times its population density. While a number of factors can explain this, we believe one unique factor resides in the persistent stream of misinformation spread through multiple online media. While social media is a powerful tool of online interaction and information exchange, it has also seen the rise of forms of deviant behaviors such as spreading fake news, misinformation, and disinformation. For these reasons, we have produced a study of the themes and chronological dynamics of the spreading of misinformation about COVID-19. We leverage topic modeling to identify misinformation themes and introduce a tool to visualize the evolution of these themes [MMA20]. Our main corpus, a collection of unique misinformation stories curated by our team, is augmented with global databases created by independent international organizations and available in the form of an online tracker for COVID-related misinformation[1]. Findings from our study are shared with the Arkansas Office of the Attorney General, that includes reports and misinformation stories available for the public.

2 State of the Art

Studies have shown that the spread of misinformation, even online, constitutes a very real danger to public health and safety [KAJK+20], and makes the identification of such narratives essential to fighting it. That is especially true considering that misinformation spreads in a viral manner and that consumers of misinformation tend to fail to recognize it [PMZ+20]. In order to better understand the misinformation surrounding the pandemic, we look at previous research that has leveraged topic models to understand online discussions surrounding this crisis. Research has shown the benefits of using this technique to understand fluctuating Twitter narratives [SHMB20] over time, and also in understanding the significance of media outlets in health communications [LZZ+20]. To implement topic modeling, we use the Latent Dirichlet Allocation model (LDA). Within the realm of natural language processing (NLP), topic modeling is a statistical technique designed to categorize a set of documents within a number of abstract "topics"[BLS09]. A "topic" is defined as a set of words outlining a general underlying theme. For each document, which in this case, is an individual item of misinformation in our data set, a probability is assigned that designates its "belongingness" to a certain topic. In this study, we use the popular

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Finlayson (eds.): Proceedings of the Text2Story'21 Workshop, Online, 1-April-2021, published at http://ceur-ws.org

[1]Stories and public reports made to the Arkansas Attorney General available at https://cosmos.ualr.edu/covid-19

LDA topic model due to its widespread use and proved performances [BNJ03]. Following recent research claiming that the use of custom stop-words adds little benefits [SMM17], we followed the researchers' recommendation and removed common words **after** the model had been trained. Our model choice has seen use in previous research using LDA for short texts, specifically for short social media texts such as tweets [AAAH+20, CMVM20, ZML17] or hashtags [ARL17] - to provide further context to topic models. We also tested our methodology on a secondary data set using a hierarchical Dirichlet process model (HDP) [TJBB06].

3 Results

For our first set, we used data from a manually curated corpus of 243 unique misinformation narratives spanning from January 2020 to June 2020, and gathered from various news aggregators. Topic modeling revealed various latent narratives. For instance, some topics captured the words "purposeful", "creators", "bill gates", etc., leading us to associate that topic with conspiracy theories suggesting the virus is man-made. As speculations on the origin of the virus dwindled, we saw a rise in various attempts at taking advantage of vulnerable citizens through scams or online identity theft. That topic included the words "scam", "phishing", "giveaways". One other dominant topic that was revealed is included the words "government", "control", "citizens", and "predicted". This highlighted narratives suggesting the virus stems from a government effort. The main takeaway is that misinformation items attempting to spread fear about a potential COVID-19 vaccine and phishing scams remained prominent during June. During the month of July, the main themes of the misinformation items shifted back to attempts to downplay the deadliness of the novel coronavirus. Another prominent theme in July were attempts to convince the public that COVID-19 testing is inflating the results.

Although a variety of misinformation themes were identified, particular dominant themes stood out. These themes were considered as dominant based on a simple sum of their frequency of occurrence in our data set. During the month of March, the prominent misinformation theme was the promotion of remedies and techniques to supposedly prevent, treat, or kill the novel coronavirus. During the month of April, the prominent themes still included the promotion of remedies and techniques, but additional prominent themes began to stand out. For example, several misinformation stories attempted to downplay the deadliness of the novel coronavirus. Others discussed the anti-malaria drug hydroxychloroquine. Others promoted the idea that the virus was a hoax meant to defeat President Donald Trump. Others consisted of various attempts to attribute false claims to high-profile people, such as politicians and representatives of health organizations. Also in April, although first signs of these were seen in March, the idea that 5G caused the novel coronavirus began to become more prevalent. During the month of May, the prominent themes shifted to predominantly false claims made by high-profile people, followed by attempts to convince citizens that face masks are either more harmful than not wearing one, or are ineffective at preventing COVID-19, and how to avoid rules that required their use. The number and variety of identity theft phishing scams also increased during May. Misinformation items attempting to attribute false claims to high-profile people continued throughout May. Also becoming prominent in May were misinformation items attempting to spread fear about a potential COVID-19 vaccine, and items promoting the use of hydroxychloroquine. During the month of June, the prominent theme shifted significantly to attempts to convince citizens that face masks are either more harmful than not wearing one, and how to avoid rules that required their use. Phishing scams also remained prominent during June. During the month of July, the dominant themes of the misinformation items shifted back to attempts to downplay the deadliness of the novel coronavirus. Another prominent theme in July were attempts to convince the public that COVID-19 testing is inflating the results.

For our second data set; after seeing promising results while using small samples of data, we experimented with YouTube data: the video titles and associated comments of videos coming up when searching YouTube for relevant keywords: "covid", "outbreak", "virus", etc. Since the data was not as finely curated for YouTube videos, the resulting topics were unsurprisingly not as detailed. However we obtained some promising results, especially when studying YouTube comments. With a much larger set (652,120 comments), we found that a HDP model outperform LDA models insofar as it was able to identify a probable topic for misinformation. When applied to our comments set, our LDA model mostly found general terms while also successfully isolating non-English comments. Figure 1 shows two topics of interests detected by our LDA model. One is Topic "7", characterized by the keywords "china", "virus", and "made". While discussion of China has been on a downward trend since the start of the pandemic, the mention of the term "virus" along with "china" suggests toxic behavior. The other Topic, "17", includes toxic language and keywords that could be used in a hostile way or communicate further sinophobic sentiments - for example, "dumb", and "'bats".

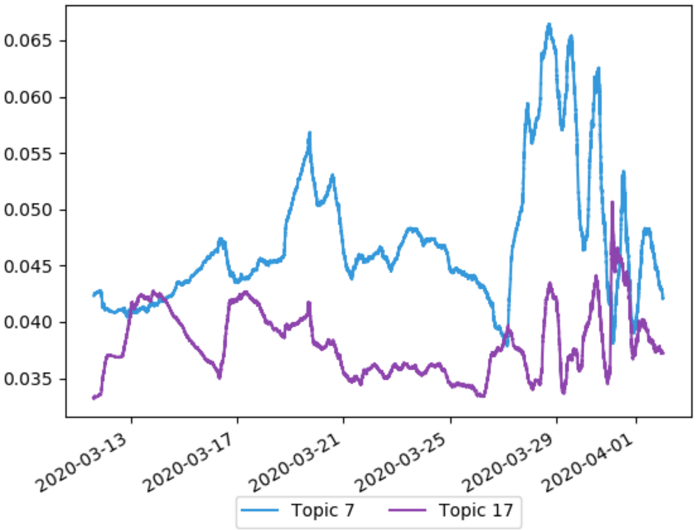

Figure 1: LDA Model Topic Streams - Daily Chance for Document to be Part of Topic

Our LDA model, however, behaved as expected and was able to identify major topics, mostly news videos, as well as what we suspect to be a vehicle of misinformation. On this very large set, our HDP model somewhat outperformed LDA for our purposes as it was able to identify a probable topic for misinformation. When applied to our comments set, our LDA model mostly found general terms while also successfully isolating non-English comments. The model did identify a topic with some toxic language and some that could be used in a hostile way or communicate sinophobic sentiments. While discussion of China has so far been on a downward trend since the start of the pandemic, the mention of the term "virus" along with "china" suggests toxic behavior.

4 Conclusion

This study has highlighted some of the narratives that surfaced during the COVID-19 pandemic. From January 2020 to July 2020, we collected 243 unique misinformation narratives and proposed a tool to observe their evolution. We have shown the potential of using topic modeling visualization to get a bird's eye view of the fluctuating narratives and an ability to quickly gain a better understanding of the evolution of individual stories. We have seen that the tool is efficient to chronologically represent actual narratives pushed to various outlets, as confirmed by the ground truth observed by our misinformation curating team and independent international organizations. Working with the Arkansas Office of the Attorney General, this study illustrates a relatively quick technique for allowing policy makers to monitor and assess the diffusion of misinformation on online social networks in real-time, which will enable them to take a proactive approach in crafting important theme-based communication campaigns to their respective citizen constituents. We have made most of our findings available online to support this effort. We then scaled up our data and repeated our methodology on social media data: YouTube video titles and comments. Over concerns of our LDA topic model becoming difficult to scale, we experimented with a HDP (Hierarchical Dirichlet Process) model, which attempts to infer the number of topics. We found promising but unsurprisingly less precise results. We notice that HDP was able to isolate a probable subset of polarizing comments. One possible way forward would be to use HDP to identify these subsets, filter out irrelevant comments, then apply the LDA model. This may reveal various narratives, some of them spreading misinformation, and further automate the process of identifying online misinformation in uncontrolled spaces.

4.0.1 Acknowledgements

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121), U.S. Air Force Research Lab, U.S. Army

Research Office (W911NF-17-S-0002, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

[AAAH+20] Alaa Abd-Alrazaq, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, and Zubair Shah. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *J Med Internet Res*, 22(4):e19016, April 2020.

[ARL17] Md. Hijbul Alam, Woo-Jong Ryu, and SangKeun Lee. Hashtag-based topic evolution in social media. *World Wide Web*, 20(6):1527–1549, November 2017.

[BLS09] David M. Blei, John D. Lafferty, and Ashok N. Srivastava. *Text Mining: Classification, Clustering, and Applications*. CRC Press, 2009.

[BNJ03] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[CMVM20] Ranganathan Chandrasekaran, Vikalp Mehta, Tejali Valkunde, and Evangelos Moustakas. Topics, trends, and sentiments of tweets about the covid-19 pandemic: Temporal infoveillance study. *J Med Internet Res*, 22(10):e22624, Oct 2020.

[KAJK+20] Ramez Kouzy, Joseph Abi Jaoude, Afif Kraitem, Molly B El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W Akl, and Khalil Baddour. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus*, 12(3):e7255–e7255, March 2020. Publisher: Cureus.

[LZZ+20] Qian Liu, Zequan Zheng, Jiabin Zheng, Qiuyi Chen, Guan Liu, Sihan Chen, Bojia Chu, Hongyu Zhu, Babatunde Akinwunmi, Jian Huang, Casper J P Zhang, and Wai-Kit Ming. Health communication through news media during the early stage of the covid-19 outbreak in china: Digital topic modeling approach. *J Med Internet Res*, 22(4):e19118, 4 2020.

[MMA20] Thomas Marcoux, Esther Mead, and Nitin Agarwal. The ebb and flow of the covid-19 misinformation themes. 5th International Workshop on Mining Actionable Insights from Social Networks Special Edition on Dis/Misinformation Mining from Social Media (MAISON 2020), 10 2020.

[PMZ+20] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G. Lu, and David G. Rand. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, 31(7):770–780, 2020. _eprint: https://doi.org/10.1177/0956797620939054.

[SHMB20] Hao Sha, Mohammad Al Hasan, George Mohler, and P. Jeffrey Brantingham. Dynamic topic modeling of the covid-19 twitter narrative among u.s. governors and cabinet executives, 2020.

[SMM17] A. Schofield, M. Magnusson, and D. Mimno. Pulling out the stops: Rethinking stopword removal for topic models. In *15th Conference of the European Chapter of the Association for Computational Linguistics*, volume 2, page 432–436. Association for Computational Linguistics, 2017.

[TJBB06] Yee Whye Teh, Michael I. Jordan, Matthew J. Beal, and David M. Blei. Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006. Publisher: Taylor & Francis.

[ZML17] Y. Zhang, W. Mao, and J. Lin. Modeling topic evolution in social media short texts. In *2017 IEEE International Conference on Big Knowledge (ICBK)*, pages 315–319, 2017.