The Ebb and Flow of the COVID-19 Misinformation Themes

Thomas Marcoux
txmarcoux@ualr.edu

Esther Mead
elmead@ualr.edu

Nitin Agarwal
nxagarwal@ualr.edu

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

Abstract

The COVID-19 pandemic has seen the emergence of unique misinformation narratives in various outlets, through social media, blogs, etc. This online misinformation has been proven to spread in a viral manner and has a direct impact on public safety. In an effort to improve public understanding, we curated a corpus of 543 misinformation pieces whittled down to 243 unique misinformation narratives along with third party proofs debunking these stories. Building upon previous applications of topic modeling to COVID-19 related material, we developed a tool leveraging topic modeling to create a chronological visualization of these stories. From our corpus of misinformation stories, this tool has shown to accurately represent the ground truth reported by our curator team. This highlights some of the misinformation narratives unique to the COVID-19 pandemic and provides a quick method to monitor and assess misinformation diffusion, enabling policymakers to identify themes to focus on for communication campaigns.

1 Introduction

Following the discovery and subsequent spread of the COVID-19 pandemic, information has become one pivotal entity in determining how each nation responds to the crisis. We have seen a variety of contradictory statements on the national and international scene influencing opinions, in some cases politically polarizing the issue of how to respond to the pandemic. But we have also seen cases of direct, physical - i.e. direct mail scams - attempts at preying on the uninformed or vulnerable such as personal protective equipment (PPE) marketing schemes. In both cases, it is obvious that information has a very real impact on the lives and livelihood of many. As such, we propose a study of the themes and chronological dynamics of the spreading of misinformation about COVID-19. Our corpus is a collection of unique misinformation stories[1] manually curated by our team. To highlight and visualize these misinformation themes, we use topic modeling, and introduce a tool to visualize the evolution of these themes chronologically.

2 Literature Review

The information community has been tackling the issue of misinformation surrounding the COVID-19 pandemic since early in the outbreak. We base the claims found in this paper on the findings that misinformation spreads in a viral fashion and that consumers of misinformation tend to fail at recognizing it as such [Pen+20]. In addition to this, we believe this research is essential as rampant misinformation constitutes a danger to public safety [Kou+20]. We also believe this research is helpful in curbing misinformation since researchers have found that simply recognizing the existence of misinformation and improving our understanding of it can enhance the larger public's ability to recognize misinformation as such [Pen+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

Title of the Proceedings: "Proceedings of the CIKM 2020 Workshops, October 19-20, Galway, Ireland" Editors of the Proceedings: Stefan Conrad, Ilaria Tiddi

[1] Stories can be explored at our official website https://cosmos.ualr.edu/covid-19

benefits of using this technique to understand fluctuating Twitter narratives [Sha+20] over time, and also in understanding the significance of media outlets in health communications [Liu+20].

To implement topic modeling, we use the Latent Dirichlet Allocation model. 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 LDA topic model due to its widespread use and proved performances [BNJ03]. One point of debate within the topic modeling community is the elimination of stop-words: i.e., should analysts filter common words from their corpus before training a model. 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 [ZML17]. Some other social media research using homogeneous social media sources such as tweets or blog posts use associated hashtags to provide further context to topic models [ARL17]. This is a promising lead to expend this research towards big data social media corpora.

In this paper, we propose to leverage topic models to understand the main underlying themes of misinformation and their evolution over time using a manually curated corpus of known fake narratives.

3 Methodology

This study uses a two-step methodology to produce relevant topic streams. First, through a manual curating process, we aggregate different misinformation narratives for later processing. We consider misinformation narratives, any narrative pushed through a variety of outlets (social media, radio, physical mail, etc.) that has been or is later believably disproved by a third party. This corpus constitutes our input data. Secondly, we use this corpus to train an LDA topic model and to generate subsequent topic streams for analysis. We describe these two steps in more details in the next sections.

3.1 Collection of Misinformation Stories

Initially, the misinformation stories in our data set were obtained from a publicly available database created by EUvsDisinfo in March of 2020 [EUv20]. EUvsDisinfo's database, however, was primarily focused on "pro-Kremlin disinformation efforts on the novel coronavirus". Most of these items represented false narratives that were communicating political, military, and healthcare conspiracy theories in an attempt to sow confusion, distrust, and public discord. Subsequently, misinformation stories were continually gleaned from publicly available aggregators, such as POLITIFACT[2], Truth or Fiction[3], FactCheck.org[4], POLYGRAPH.info[5], Snopes[6], Full Fact[7], AP Fact Check[8], Poynter[9], and Hoax-Slayer[10]. The following data points were collected or each misinformation item: title, summary, debunking date, debunking source, misinformation source(s), theme, and dissemination platform(s). The time period of our data set is from January 22, 2020 to July 22, 2020. The data set is comprised of 548 unique misinformation items. For many of the items, multiple platforms were used to spread the misinformation. For example, oftentimes a misinformation item will be posted on Facebook, Twitter, YouTube, and as an article on a website. For our data set, the top-used platforms used for spreading misinformation were websites, Facebook, Twitter, YouTube, and Instagram, respectively.

3.2 Topic Modeling

In order to derive lexical meaning from this corpus, we built a pipeline executing the following steps. First, we processed each document in our text corpus. All that is needed is a text field identified by a date. Because in most cases of word of mouth or social media it is impossible to pinpoint the exact date the idea first emerged, we use the date of publication of the corresponding third party "debunk piece". We trained our LDA model using the Python tool Gensim[11] using the methodology and pre-processing best practices as described by its author [ŘS10] as well as best stop words practices as described earlier [SMM17]. In this study, we found that generating 20 different topics best matched the ground truth as reported by the researchers curating the misinformation stories.

Once the model was trained, we ordered the documents by date and created a numpy matrix where each document is given a score for each topic produced by the model. This score describes the probability that

[2] https://www.politifact.com/coronavirus/
[3] https://www.truthorfiction.com/
[4] https://www.factcheck.org/
[5] https://www.polygraph.info/
[6] https://www.snopes.com/fact-check/
[7] https://fullfact.org/health/coronavirus/#coronavirus
[8] https://apnews.com/APFactCheck
[9] https://www.poynter.org/ifcn-covid-19-misinformation
[10] https://www.hoax-slayer.net/category/covid-19/
[11] https://radimrehurek.com/gensim/

the given document is categorized as being part of a topic, i.e. if a score is high enough (here, a 10% probability), the document is considered part of the topic. This allowed us to leverage the Python Pandas[12] library to plot a chronological graph for each individual topic. We averaged topic distribution per day and used a moving average window size of 20. This helped in highlighting the overarching patterns of the different narratives. The tool is publicly available and can be found in the footnotes[13].

4 Results

In this section, we discuss the thoughts of our data collection team and the ground truth as they were observed, and compare these with the results obtained through our topic modeling visualization tool.

4.1 Prominent Misinformation Themes Over Time

Although a variety of misinformation themes were identified, particularly dominant themes stood out, changing over time. 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.

4.2 Topic Streams

After using the tool described in 3.2, we generated the graphs and tables described and discussed in this section. Our data contains 243 unique misinformation narratives spanning from January 2020 to June 2020. The data was curated by our research team through the process described in the methodology. Each entry contains, among other fields, a "date" used as a chronological identifier, a "title" describing the general idea the misinformation is attempting to convey, and a "theme" field putting the story in a concisely described category. For example, a story given the title *"US Department of Defense has a secret biological laboratory in Georgia"* is categorized in the following theme: *"Western countries are likely to be purposeful creators of the new virus."* Each topic was represented by an identification number up to 20 and a set of 10 words. We picked the three most relevant words that best represented the general idea of each topic. Notably, obvious words such as *covid* or *coronavirus* were removed from the topic descriptions since they are common for every topic.

In Tables 1 and 2, we described some of the twenty topics found by each of our LDA models. These topics were chosen because they each described a precise narrative and have a low topic distribution (or proportion within the corpus). A low proportion is desirable because this indicates the detection of a unique narrative within the corpus; as opposed to an overarching topic including general words such as "world", "outbreak", or "pandemic". Do note that topic inclusiveness is not exclusive and documents can be part of multiple topics.

This becomes apparent in the tables below: from our topic model, we found a dominant topic encompassing 68% of narratives. It includes words such as "Trump", "outbreak", "president", etc. Some other narratives also included words such as "flu", "news", or "fake". Because the evolution of these narratives are consistent across the corpus and show little temporal fluctuation, we chose not to report on them further.

[12]https://pandas.pydata.org/
[13]https://github.com/thomas-marcoux/TopicStreamsTools

For these reasons, the narratives we focused on below show a low percentage of distribution.

Table 1: Most frequent dominant topics from titles.

ID	Word 1	Word 2	Word 3	Proportion
10	china	chinese	spread	2%
12	scam	hydroxy...	health	2%
17	state	donald	trump	2%
18	vaccine	gates	bill	5%

Table 2: Most frequent dominant topics from themes.

ID	Word 1	Word 2	Word 3	Proportion
3	fear	spread	western	2%
9	predicted	pandemic	vaccine	2%
16	phishing	hydroxy...	vaccine	2%

4.2.1 Using narrative titles as a corpus

The general narratives described by the topics were thus:

- Topic 10 described the narratives related to the Chinese government and its responsibility in the spread of the virus. These stories represented an estimated 2% of the 243 stories collected.
- Topic 12 described the narratives related to personal health and scams or misinformation such as the benefits of hydroxychloroquine. These stories represented an estimated 2% of the 243 stories collected.
- Topic 17 described the narratives related to the response of Donald Trump and his administration. These stories represented an estimated 2% of the 243 stories collected.
- Topic 18 described the narratives related to the involvement of Bill Gates in various conspiracies, mostly linked to vaccines. These stories represented an estimated 4% of the 243 stories collected.

Figure 1 shows the evolution of Topic 10, the topic describing China-related narratives. It shows that these narratives were already in full force from the beginning of our corpus and slowly came to a near halt during the month of April. We notice a short spike again towards the end of the corpus during the month of June. This is consistent with the ground truth of

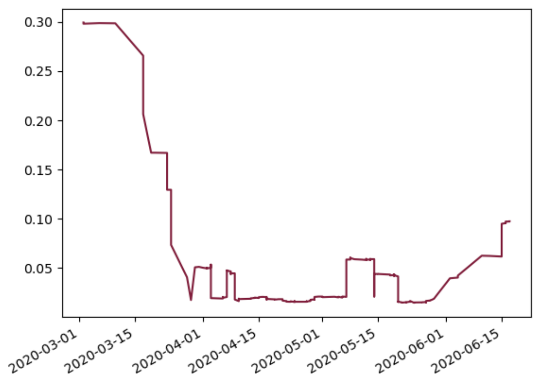

Figure 1: Topic distribution of titles for topic 10 (keywords: china, chinese, spread)

online narratives that focused on the provenance of the virus during the early stages.

Figure 2 shows the evolution of Topic 12, the topic describing narratives related to health, home remedies, and general hoaxes and scams stemming from the panic. We can see it was consistent with the rise of cases in the United States and panic increased as with the spread of the virus. It is interesting to note that this figure roughly coincides with the daily number of confirmed cases for this time period [Rit+20].

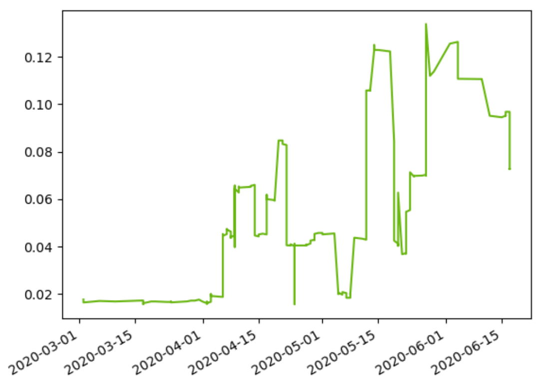

Figure 2: Topic distribution of titles for topic 12 (keywords: hydroxychloroquine, health, scam)

Figure 3 shows the evolution of Topic 17. This topic described stories related to Donald Trump and his administration. These stories generally referred to claims that the virus was manufactured as a political strategy, or claims that various public figures were speaking out against the response of the Trump administration.

Figure 4 shows the evolution of Topic 18. This

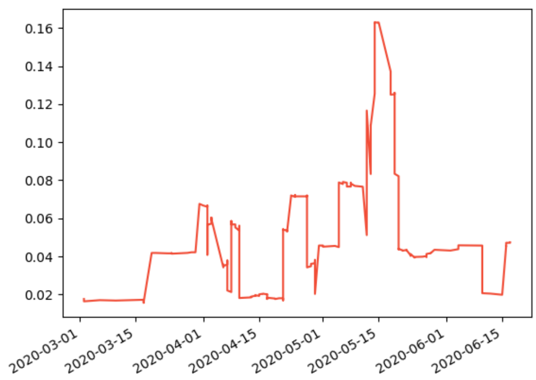

Figure 3: Topic distribution of titles for topic 17 (keywords: donald, trump, state)

topic described stories such as Bill Gates and his perceived involvement with an hypothetical vaccine, and other theories describing the virus' appearance and spread as an orchestrated effort. As with Figure 1, these narratives were especially strong early on (albeit this narrative remained active for a slightly longer time), before coming to a near halt.

We notice that as theories about the origins of the virus slowed down, hoaxes and scams on personal protection increased as shown on Figure 2.

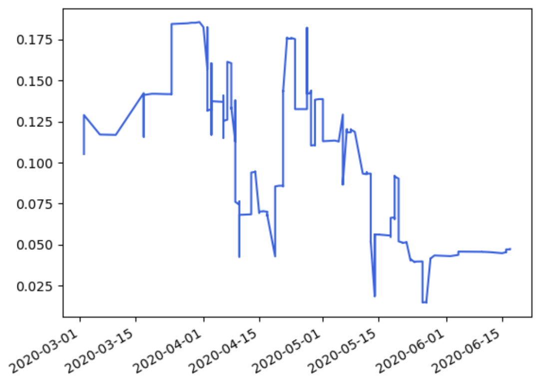

Figure 4: Topic distribution of titles for topic 18 (keywords: bill, gates, vaccine)

4.2.2 Using narrative themes as a corpus

For this section, we inputted narrative themes as the corpus. Note that the topic IDs are independent from the previous set of topics using titles. Similarly to section 4.2.1, we found a dominant topic encompassing 68% of narratives as well. This time including words such as "attempt", "countries", and "purposeful". As for section 4.2.1, we chose not to report on that topic as well as other smaller but general topics showing little fluctuation. Therefore, the narratives we focused on below show a low percentage of distribution. The general narratives described by the topics are thus:

- Topic 3 described the narratives related to the speculations on the spread of the virus, especially in an international relations context. These stories represented an estimated 2% of the 243 stories collected.
- Topic 9 described the narratives related to stories claiming the creation and propagation of the virus were either designed or predicted, along with voices claiming a vaccine already exists. These stories represented an estimated 3% of the 243 stories collected.
- Topic 16 described the narratives related to personal health and scams or misinformation such as the benefits of hydroxychloroquine. These stories represented an estimated 2% of the 243 stories collected.

Figure 5 shows the evolution of Topic 3. It is linked to early fear of the virus and presented narratives as opposing the western block with the East, notably China. It matched closely with Figure 1 and its China-related narratives. In both cases, we see an early dominance of the topic followed by a near halt as the virus touched the United States.

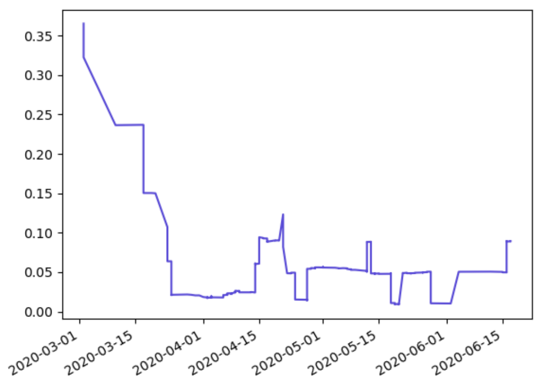

Figure 5: Topic distribution of themes for topic 3 (keywords: fear, spread, western)

Figure 6 describes the evolution of narratives claiming the virus was predicted or even designed. This figure is consistent with the results shown by Figure

4 which shows claims regarding Bill Gates, early vaccines, etc. They both showed stories of early knowledge of the virus and peaked early, appearing more or less sporadically as time goes on and as cases increased.

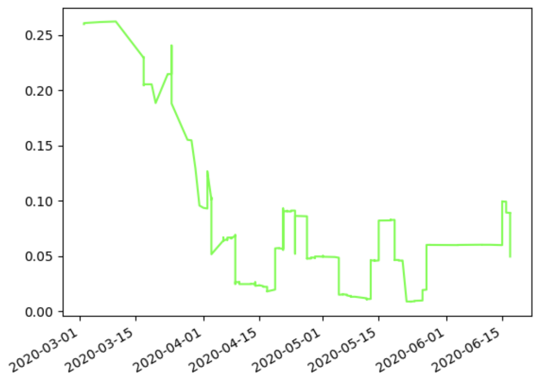

Figure 6: Topic distribution of themes for topic 9 (keywords: predicted, pandemic, vaccine)

Figure 7 is parallel to Figure 2. Both showed hoax stories promoting scams and health-related misinformation. We noticed an early rise on Figure 7, most likely due to the inclusion of the keyword "vaccines" in the topic, which caused some overlap with Topic 9 as shown in Figure 6.

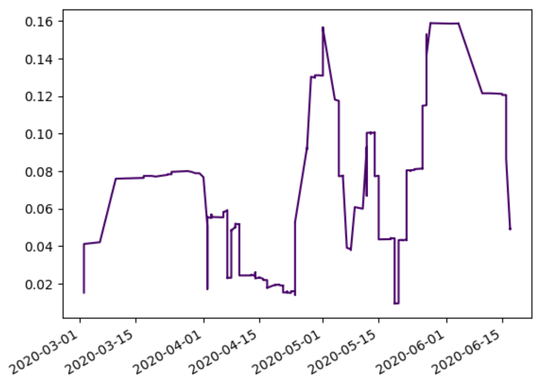

Figure 7: Topic distribution of themes for topic 16 (keywords: hydroxychloroquine, vaccine, phishing)

5 Conclusion

This study has highlighted some of the narratives that surfaced during the COVID-19 pandemic. We collected 243 unique misinformation narratives over six months 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. This work 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 also seen in this study that using carefully curated "themes" - which offer a lexical value close to the abstract topics provided by the LDA model - yields similar results to using misinformation narratives "title". This paves the way for scaling this method with much larger corpora such as a set of news headlines, blog titles, or social media posts.

LDA is generally viewed as more reliable due to the control one can have over the number of topics. Finding an optimal level of granularity through trial and error tends to perform well when tailored to the use-case. Because the LDA topic model may become difficult to scale, however, we consider using the HDP (Hierarchical Dirichlet Process) model for future works involving multiple larger corpora. This model attempts to infer the number of topics computationally, which may become more scalable on large sets of documents with an unknown number of topics.

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), 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

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

[BLS09] David M. Blei, John D. Lafferty, and Ashok N. Srivastava. *Text Mining: Classification, Clustering, and Applications*. CRC Press, 2009, pp. 71–88.

[ŘS10] Radim Řehůřek and Petr Sojka. "Software Framework for Topic Modelling with Large Corpora". In: May 2010, pp. 45–50. DOI: 10.13140/2.1.2393.1847.

[ARL17] Md. Hijbul Alam, Woo-Jong Ryu, and SangKeun Lee. "Hashtag-based topic evolution in social media". In: *World Wide Web* 20.6 (Nov. 2017), pp. 1527–1549. ISSN: 1573-1413. DOI: 10.1007/s11280-017-0451-3. URL: https://doi.org/10.1007/s11280-017-0451-3.

[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*. Vol. 2. Association for Computational Linguistics. 2017, pp. 432–436.

[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)*. 2017, pp. 315–319.

[EUv20] EUvsDisinfo. *EUvsDisinfo. March 16, 2020. The Kremlin and Disinformation About Coronavirus.* 2020. URL: https://euvsdisinfo.eu/the-kremlin-and-disinformation-about-coronavirus/. (accessed: 07.16.2020).

[Kou+20] Ramez Kouzy et al. "Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter". eng. In: *Cureus* 12.3 (Mar. 2020). Publisher: Cureus, e7255–e7255. ISSN: 2168-8184. DOI: 10.7759/cureus.7255. URL: https://pubmed.ncbi.nlm.nih.gov/32292669.

[Liu+20] Qian Liu et al. "Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach". In: *J Med Internet Res* 22.4 (Apr. 2020), e19118. ISSN: 1438-8871. DOI: 10.2196/19118. URL: http://www.ncbi.nlm.nih.gov/pubmed/32302966.

[Pen+20] Gordon Pennycook et al. "Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention". In: *Psychological Science* 31.7 (2020). _eprint: https://doi.org/10.1177/0956797620939054, pp. 770–780. DOI: 10.1177/0956797620939054. URL: https://doi.org/10.1177/0956797620939054.

[Rit+20] Hannah Ritchie et al. *United States: Coronavirus Pandemic - Our World in Data.* 2020. URL: https://ourworldindata.org/coronavirus/country/united-states?country=~USA. (accessed: 07.29.2020).

[Sha+20] Hao Sha et al. *Dynamic topic modeling of the COVID-19 Twitter narrative among U.S. governors and cabinet executives.* 2020. arXiv: 2004.11692 [cs.SI].