“Thought I’d Share First”: An Analysis of COVID-19 Conspiracy Theories and Misinformation Spread on Twitter

Dax Gerts¹, Courtney D. Shelley¹, Nidhi Parikh, Travis Pitts, Chrysm Watson Ross, Geoffrey Fairchild, Nidia Yadria Vaquera Chavez, Ashlynn R. Daughton*

¹These authors contributed equally.
*Corresponding author: adaughton@lanl.gov

Abstract

Background: Misinformation spread through social media is a growing problem, and the emergence of COVID-19 has caused an explosion in new activity and renewed focus on the resulting threat to public health. Given this increased visibility, in-depth analysis of COVID-19 misinformation spread is critical to understanding the evolution of ideas with potential negative public health impact.

Objective: We use Twitter data to explore methods for characterization and classification of major COVID-19 myths and conspiracy theories, and to provide context for the theories’ evolution through the pandemic’s early months.

Methods: Using a curated data set of COVID-19 tweets (N ~ 120 million tweets) spanning late January to early May 2020, we applied methods including regular expression filtering, supervised machine learning, sentiment analysis, geospatial analysis, and dynamic topic modeling to trace the spread of misinformation and to characterize novel features of COVID-19 conspiracy theories.

Results: Random forest models for four major misinformation topics provided mixed results, with narrowly-defined conspiracy theories achieving F1 scores of 0.804 and 0.857, while more broad theories performed measurably worse, with scores of 0.654 and 0.347. Despite this, analysis using model-labeled data was beneficial for increasing the proportion of data matching misinformation indicators. We were able to identify distinct increases in negative sentiment, theory-specific trends in geospatial spread, and the evolution of conspiracy theory topics and subtopics over time.

Conclusions: COVID-19 related conspiracy theories show that history frequently repeats itself, with the same conspiracy theories being recycled for new situations. We use a combination of supervised learning, unsupervised learning, and natural language processing techniques to look at the evolution of theories over the first four months of the COVID-19 outbreak, how these
theories intertwine, and to hypothesize on more effective public health messaging to combat misinformation in online spaces.

Keywords:

COVID-19; coronavirus; social media; misinformation; health communication; twitter; conspiracy theories; vaccine hesitancy; 5g; unsupervised learning; random forest; active learning; supervised learning
1. Introduction

On December 31, 2019 the World Health Organization (WHO) was made aware of a cluster of cases of ‘viral pneumonia’ of unknown origin in Wuhan, Hubei Province, China [1]. The WHO reported this cluster via Twitter on January 4, 2020 [2]. On January 19, the WHO Western Pacific Regional Office tweeted evidence of human-to-human transmission [3]. The first US case was reported the next day. Five days later, on January 26, 2020, GreatGameIndia published the article, “Coronavirus Bioweapon–How China Stole Coronavirus From Canada And Weaponized It,” which claimed that coronavirus was leaked into China from a Canadian lab [4]. The article was reposted the same day on the website ZeroHedge with the title, “Did China Steal the Coronavirus and Weaponize It?” [5]. The story quickly went viral [6].

Misinformation surrounding pandemics is not unique to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease 2019 (COVID-19). At least as far back as the Russian flu of 1889, pandemic spread of misinformation has been concomitant with disease spread [7]. People are susceptible to misinformation, claims of fact that are either demonstrably false based on authoritative sources or unverifiable based on current knowledge [8], when trust in those authoritative sources is low, which can occur when officials provide conflicting information and guidance [9]. Misinformation must also be differentiated from misperception, an incorrect understanding or interpretation of facts, regardless of the truth of those facts [10]. Nyhan and Reifler (2010) differentiate between those who are uninformed and those who are misinformed; the former are open to corrective information because they are aware of their lack of knowledge while the latter may only become more convinced of their misperceptions with corrective information [11].

Misinformation will also include conspiracy theories, which posit explanations of events or circumstances based primarily on a conspiracy, small groups of powerful people acting in secret to their own benefit at the expense of another group [11]. While some conspiracies, such as Watergate or the Tuskegee experiments, may eventually be found to be true, the vast majority of conspiracy theories are not true and their spread can undermine public health efforts [12]. Some conspiracy theories may be better classified as disinformation, false or misleading information that is intentionally passed to a target group [13]. Disinformation is a form of “black propaganda,” in which the true source of information is concealed [14].

The COVID-19 outbreak has left many people isolated within their homes, turning to social media for news and to connect with others. This leaves them especially vulnerable to believing and sharing conspiracy theories [15]. This study examines four oft-repeated and long-lived conspiracy theories surrounding COVID-19: the virus is man-made and was released from a laboratory; Bill Gates or the Gates Foundation created or patented the virus; a COVID-19 vaccine will contain a tracking microchip or be otherwise harmful; and that 5G technology is somehow associated with the disease. None of these conspiracy theories are unique, nor are they entirely distinct. Many theories overlap in specific ideas or end results, and many recycle a verifiable “grain of truth” combined with a previously circulating conspiracy message.

5G Cell Towers Spread Coronavirus. Cellular carriers began limited roll-out of 5G cellular service in 2018, with wider coverage available by the end of 2020 [16]. Coverage
requires the installation of new cellular towers, which are much smaller than existing 4G towers and need to be placed an average of 500 feet apart [17]. These new towers were already the source of a more general conspiracy theory that the signal degrades the human immune system and that these dangers are being “covered up” by “powerful forces in the telecommunications industry” [18]. The theme of wireless technology causing immune damage in humans was previously seen with 2G, 3G, 4G, and WiFi roll-outs [19], and even the 1889 Russian flu was purportedly caused by the then-new technology of the electric light [20]. The COVID-related 5G conspiracy theory began in the first week of January, and may not have evolved past a fringe view into a trending hashtag without being shared by fake news websites with the sole aim of spreading conspiracy theories on Twitter or by people aiming to denounce the theory [19].

**Bill Gates/Gates Foundation.** Conspiracy theories often “are about accusing powerful people of doing terrible things” [J. Uscinski quoted in 20]. The Bill & Melinda Gates Foundation is arguably the biggest philanthropic venture ever attempted, targeting global health through “creative capitalism” solutions [21] that largely bypass existing health systems and local needs [22]. The variety of projects associated with the Bill & Melinda Gates Foundation have proven fertile ground for the development of conspiracy theories, ranging from misinterpretations of a “patent on COVID-19” [23] to incorporation of vaccine-averse concerns. For example, the Foundation funded research at the Massachusetts Institute of Technology to develop injectable invisible ink to serve as permanent records of vaccination for use in developing countries [24,25]. This technology was announced in December 2019, the same month that SARS-CoV-2 emerged in Wuhan, China and has resulted in a conspiracy theory suggesting the COVID vaccine would microchip individuals with the goal of population control [23].

**Laboratory Origins.** In the 1980s the Soviet KGB and the East German Ministry of State Security (“Stasi”) created a disinformation campaign known as “Operation Infektion” to promote the idea that HIV was the result of an escaped virus created by the US Army in Ft. Detrick, Maryland [26]. This conspiracy theory was the result of a real conspiracy between Soviet Bloc countries to tarnish the reputation of the US. This story was purposefully planted throughout the world, which perhaps explains why associations between HIV and other infectious diseases consistently re-emerge, including polio [27], Ebola [28], and COVID-19.

The COVID-related HIV conspiracy theory began on January 31, 2020 with the preprint publication of “Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag” [29], which was quickly retweeted by Anand Ranganathan, a molecular biologist with over 200,000 followers on Twitter, legitimizing the conspiracy theory with his scientific reputation by tweeting, “Oh my god. Indian scientists have just found HIV (AIDS) virus-like insertions in the 2019-nCoV virus that are not found in any other coronavirus. They hint at the possibility that this Chinese virus was designed...” [30]. Within two hours Ross Douthat, a prominent New York Times opinion columnist, retweeted Ranganathan to his 140,000+ followers [31], further legitimizing the theory through a reputable news outlet and greatly furthering the story’s reach outside the scientific community [32]. Three days after the pre-print’s initial release, the original paper was retracted [29] after many scientists discredited the team’s overly simplistic methodology. Laboratory origins theories have since become

---

1 Retracted paper.
something of a political weapon, with US President Donald Trump claiming to have evidence of a Chinese laboratory origin \[^{33}\], prompting a Twitter response from a Chinese government account \[^{34}\] that was flagged by Twitter as misinformation \[^{35}\]. Both of these events signal a transformation from misinformation to disinformation, opening up the potential for government-sponsored propaganda.

Additional laboratory-related conspiracy theories quickly emerged, such as that the virus was created to achieve global population reduction, or to impose quarantines, travel bans, and martial law, all of which were previously seen during the 2014 Ebola outbreak \[^{29}\] and the 2015-2016 Zika outbreak \[^{36,37}\].

**Vaccines.** In addition to conspiracy theories surrounding HIV similarities and laboratory origins, the 2014 Ebola outbreak also spawned conspiracy theories surrounding the CDC’s supposed patent of the virus and ability to profit off of vaccines while conspiring with American pharmaceutical companies \[^{29}\]. People have feared vaccines since the smallpox vaccine was first introduced in England in 1796, and conspiracy theories quickly followed \[^{38}\]. In addition to the genuine conspiracy to disseminate HIV disinformation described above, conspiracy theories surrounding HIV included the ineffectiveness of antiretroviral treatments \[^{39}\] and that a cure existed but was being withheld by government and pharmaceutical companies \[^{40}\].

Vaccine-related social media articles in 2019 still consist of debunked associations with autism and general mistrust of government and the pharmaceutical industry, with most of those who share articles being relatively knowledge-deficient and vaccine-averse compared to non-sharers \[^{41}\]. Despite representing fewer users, anti-vaccine Facebook pages are more numerous than pro-vaccine pages; are more varied in discussion interests; provide more opportunities for engagement; and are more geographically diverse, ranging from local to global focus \[^{42}\]. Johnson et al. (2020) predicts that users expressing vaccine-averse opinion on Facebook will overtake pro-vaccine opinion within the next decade due to having many more interlinked pages with rapidly growing follower bases \[^{42}\].

With the above framing in mind, this paper offers the following contributions to the literature. First, we use a combination of regular expression-based filtering and supervised learning to identify tweets associated with the four conspiracy theories described above to separate misinformation from other COVID-related tweets. We additionally explore sentiment of tweets and geospatial properties of misinformation tweets. We use an unsupervised learning approach of dynamic topic modeling to explore the changes in word importance among topics within each theory. All of this is done using a large corpus of Twitter data (120 million initial tweets, and 1.8 million that met our initial regular expression filtering step). Lastly, we contextualize these findings within the rich, interdisciplinary literature on misinformation, and health-related content on social media.

Health officials too often fail in crafting effective messaging campaigns because they target what they want to promote rather than addressing the recipients’ existing misperceptions \[^{43}\]. Misinformation can spread rapidly and without clear direction, as evidenced by one tweet (anonymized for privacy) we uncovered while conducting this research, which said “[Article on 5g/coronavirus connection]... Haven’t vetted, thought I’d share first.” An understanding of the appearance, transmission, and evolution of COVID-19 conspiracy theories can allow public health officials to better craft outreach messaging, and adjust those messages if public
perceptions measurably shift. This study aims to demonstrate that identifying and characterizing the most common and long-lived COVID-related conspiracy theories using Twitter data is possible, even when these messages shift in content and tone over time.

Related Work

In the past year, substantial work has emerged investigating the onslaught of misinformation related to COVID-19. Multiple studies have found that misinformation is common, both on social media [44–46], as well as in the web pages returned for common COVID-19 queries at the beginning of the pandemic [47]. One survey-based study even found that scientists and clinicians think scientific journals without rigorous enough review processes have contributed to misinformation [48].

Though sparse, there have been a few studies focusing on social media data. In general, work until this point has focused on quantifying the amount of misinformation present in online spaces. One study found that original tweets more often presented false information than evidence-based information, but evidence-based information was more often retweeted [49]. Similarly, Singh et al. found that during the first three months of the outbreak the volume of misinformation tweets was small compared to the overall conversation [49]. Broniatowski et al. conducted a much larger analysis of Twitter data and found that the amount of data related to COVID-19 dwarfed that of other health-related content but found that proportionally more of the data was from credible websites compared to other health-related datasets, with the caveat that it is difficult to assess website credibility [50].

Separately, some researchers have attempted to characterize how many people believe misinformation and the characteristics of those likely to believe misinformation. Some evidence suggests the proportion of the general population that believe COVID-19 conspiracy theories could be quite high. One nationally representative study in the US found that some myths (e.g., that the virus was created or spread on purpose) were believed by over 30% of respondents [50]. Another study across several countries found that those that believe misinformation are more likely to get information from social media or have a self-perceived minority status [51]. At the same time, the study found characteristics like “trusting scientists” and getting information from the WHO had a negative relationship with belief in misinformation (see [51] for a full list of covariates considered).

This work builds on prior work by delving into specific conspiracy theories on Twitter from January until May 2020. We use a combination of supervised learning (to identify relevant tweets and analyze sentiment) and unsupervised learning (to characterize the evolution of themes within theories) to understand changes in conspiracy theories over time in a novel fashion. To our knowledge this is the first paper to do a deep analysis of these specific conspiracy theories, or to do so using these methods.
2. Data

2.1 Twitter Data

The Twitter data used for this study is derived from Chen et al. 2020 [52], who constructed and made publicly available a COVID-19 Twitter dataset by filtering for known COVID-19 keywords and significant Twitter accounts. Due to limitations in the Twitter API, this data is a 1% sample of tweets that include the keywords, which we rehydrated for use in this study. While the data repository constructed by Chen et al. continues to grow based on lists of known COVID-19 keywords and significant Twitter accounts, the analysis for this paper includes approximately 120 million tweets from January 21, 2020 to May 8, 2020 (see Figure 1). Tweets collected were primarily in English, but also include a variety of other languages [52].

Figure 1. Volume of Twitter data & important events. Twitter data were collected from January 21, 2020 - May 8, 2020. On average, we were able to rehydrate approximately one million tweets per day. Important events in the outbreak are noted for context.

2.2 NewsGuard

Theories listed in NewsGuard’s “Special Report: COVID-19 Myths” [23] were investigated and used to identify the four major conspiracy theories listed in the introduction. NewsGuard provides thorough evaluations of thousands of websites based on criteria including funding transparency, journalistic integrity, and editorial track record. Since the emergence of COVID-19, NewsGuard has also provided a summary of major myths and conspiracy theories associated with the pandemic, earliest documented claims, major events that caused significant spread, and detailed reports of major sources of COVID-19 misinformation.
3. Methods

3.1 Filtering & Supervised Classification

We first filtered the data into four datasets using regular expressions (see Figure 2) to increase the number of relevant tweets in the category of interest [53–57]. The four datasets are hereafter referred to using the following terms:

- **5G**: 5G technology is somehow associated with COVID-19.
- **Gates**: Bill Gates or the Gates Foundation funded, patented, or otherwise economically benefited from the virus.
- **Lab**: The COVID-19 virus is man-made or bioengineered and was released (intentionally or accidentally) from a laboratory.
- **Vax**: A COVID-19 vaccine would be harmful in a way not supported by science (e.g., it could contain a microchip).

**COVID-19 Tweets (Filter 1)**: Curated tweets collected using COVID-19 keywords and relevant Twitter accounts

**Topic Tweets (Filter 2)**: Regular expressions applied to identify major misinformation theories

**Misinformation Theory Patterns**:
- **5G**: (^|[^a-z0-9])*(?:[a-z0-9][a-z0-9][a-z0-9]$)?
- **Gates**: (^|[^a-z0-9]*|bill|melind)[gates]
- **Lab**: (^|lab|detr|oraroi7r|)
- **Vax**: (^|\|\|\|\[#|\]{0,25}[vacc|vax])

**Topic Misinformation Tweets (Filter 3)**: Machine learning used to classify tweets as misinformation or not

**Machine Learning Classification**:
- ~750 labeled tweets in each topic
- Random Forest model (10 rounds active learning applied)

**Figure 2. Tweet filtering flow**. Three filtering stages were applied to increase the proportion of COVID-19 misinformation tweets in the data: (1) Chen et al. [52] dataset used keywords and known accounts to provide a large sample of COVID-19 Twitter data, (2) Regular expressions were used to create four conspiracy theory datasets (“5G”, “Gates”, “Lab”, and “Vax”), and (3) machine learning classifiers were built for each dataset to automatically identify tweets within each category of COVID misinformation.

Within each keyword-filtered conspiracy theory dataset, we randomly sampled 750 tweets to use for training. Two authors coded each set of tweets and established agreement by jointly coding a subset of tweets (see Table 1). Any tweet promoting or engaging with misinformation, even to refute it, was labeled COVID misinformation. This was done with the rationale that tweeting about misinformation, even in the context of a correction, increases the audience exposed to that misinformation. Inter-rater analysis found relatively high agreement and reasonable Kappa scores (mean 0.759, Table 1). However, the effort demonstrated the
difficulty of reliably identifying misinformation; there were many cases where oblique references and jokes fell in a gray area which raters labeled “Uncertain” (~6.1% of coded tweets). To account for this without removing data (and thus shrinking the available training data), we collapsed “uncertain” labels into COVID misinformation or not COVID misinformation and identified which grouping generated higher inter-rater agreement. Results are shown in Table 1.

Table 1. Training data creation inter-rater results.

| Theory | Unique Tweets Labeled | Tweets Labeled By Multiple Authors | Agreement | Cohen’s Kappa |
|--------|------------------------|-------------------------------------|-----------|---------------|
| 5G     | 725                    | 146                                 | 0.852     | 0.708         |
| Gates  | 711                    | 143                                 | 0.893     | 0.782         |
| Lab    | 735                    | 146                                 | 0.901     | 0.796         |
| Vax    | 775                    | 199                                 | 0.915     | 0.751         |

Tweets were tokenized and both URLs and stop words were removed. Unigrams and bigrams were used as features in a document-term matrix and the most sparse (< 0.05% populated) terms were removed. Additionally, we added boolean features describing relationships to domains identified by NewsGuard as sources of misinformation. Features included (1) a tweet originating from a misinformation-identified domain, (2) a tweet replying to an originating tweet, (3) a tweet retweeting an originating tweet, or (4) a tweet that was otherwise linked (e.g., replying to a retweet of a tweet from a misinformation source). As noted elsewhere, only English tweets were used in this analysis.

The data were partitioned into a two-thirds/one-third training-test split. Data were sampled such that the training data had an equal sample distribution (50% misinformation, 50% non-misinformation). The testing data used the remaining available data and were thus an uneven sample distribution.

Classifiers were built using random forest models with 150 trees, up to 25 terminal nodes, and 25 variables randomly sampled at each split. We used an active learning approach such that after each run of the random forest classifier, the calculated posterior entropy was used to select the three unlabeled tweets that caused the most uncertainty in the model. These were then hand-labeled by an author and applied to the next run of the model. Additionally, for each hand-labeled tweet, highly similar tweets (string similarity >= 0.95) were identified and given the same label. The models which performed the best (measured by F1 score) were used to assign labels to the full dataset.
3.2 Analysis

Sentiment Analysis
Two well-documented sentiment dictionaries were used to label tokenized tweets. The first, AFINN [59], provided an integer score ranging from -5 (negative sentiment) to +5 (positive sentiment) for each word in the dictionary. The second dictionary, NRC [59], was used to tag words with categories of emotion, providing labels for eight emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust in addition to an overall “positive” or “negative” sentiment. We then compared the sentiment for each classified dataset over time. For each tweet, aggregate sentiment metrics were calculated, including the sum of integer scores and the counts for each emotion label.

Geospatial Analysis
The OpenStreetMaps geocoding service Nominatim [60] was used to match Twitter’s non-standardized, user-provided location data with geographic coordinates where possible, and was able to locate approximately 50 percent of tweets. A graduated symbol map approach [61] was used to identify geospatial clusters by grouping located tweets into ordered quantitative classes, with a maximum cluster radius commensurate with the average size of a major metropolitan area. Finally, each class was assigned a graduated symbol to visualize clusters from smallest to largest.

3.3 Dynamic Topic Modeling
Dynamic topic modeling was used to characterize themes and analyze temporal changes in word importance [62] by dividing tweets into weekly time slices based on the time they were generated. The set of topics at each time slice is then assumed to evolve from the set of topics at the previous time slice using a state space model. The result is an evolving probability distribution of words for each topic that shows how certain words become more or less important over time for the same topic. Traditional topic models, such Latent Dirichlet Allocation [63], assume that all of the documents (which are here equivalent to tweets) are drawn exchangeably from the same topic distribution, irrespective of the time when they were generated. However, a set of documents generated at different times may reflect evolving topics.

Dynamic topic models were trained for each conspiracy theory, with the number of topics ranging from two to five. Small numbers of topics were chosen because these tweets were already classified to be relevant for individual misinformation topics, and because our goal was to identify potential sub-topics that evolved over time. The optimal number of topics was assessed qualitatively by reviewing topic modeling results.
4. Results

4.1 Filtering & Supervised Classification

Filtering using regular expressions reduced our initial collection of tweets from approximately 120 million to roughly 1.8 million unique tweets across all four conspiracy theories. The relative volume of tweets in each dataset is shown in Figure 3, with the number of tweets appearing in multiple datasets visualized by edge thickness. All datasets showed some degree of overlapping tweets between categories, defined as the proportion of tweets appearing in more than one category, with “Gates” showing the most overlap and “5G” showing the least. “5G” additionally had a low volume of tweets compared to other theories.

Table 2. Results of regular expression filtering step. Number tweets per conspiracy theory and overlap between multiple categories.

| Theory | All       | 5G        | Gates     | Lab       | Vax       |
|--------|-----------|-----------|-----------|-----------|-----------|
| # Tweets | 1,796,164 | 127,209   | 278,130   | 526,115   | 969,654   |
| # Tweets in other theories | 6,300      | 69,566     | 44,198     | 82,380     |

Model performance metrics for each theory are presented in Table 3. For the theories “5G”, “Gates”, and “Lab” the balance of misinformation to non-misinformation was roughly equal in the training data. For the vaccine-averse dataset (“Vax”), the split was unbalanced, with only ~18% of labeled training data identified as COVID misinformation. These models were then used on the regular-expression filtered datasets to improve our confidence in the downstream analyses. Figure 3. shows the breakdown of each theory by model-assigned label and what proportion of tweets was composed of COVID Misinformation.

Table 3. Random forest model training results.

| Theory | Training Data (COVID Misinfo / Not COVID Misinfo) | Metrics | Random Forest | Random Forest (with Active Learning) | Delta |
|--------|---------------------------------------------------|---------|---------------|-------------------------------------|-------|
| 5G     | 367 / 356 (50.8%)                                  | Accuracy | 0.779         | 0.783                               | 0.004 |
|        |                                                   | Recall  | 0.908         | 0.872                               | -0.036|
|        |                                                   | Precision | 0.728       | 0.744                               | 0.016 |
| Theory | Counts | Accuracy | Recall | Precision | F1 Score |
|--------|--------|----------|--------|-----------|---------|
| Gates  | 354 / 356 (49.9%) | 0.622 | 0.675 | 0.608 | 0.64 |
|        |         | 0.5819 | 0.793 | 0.556 | 0.654 |
|        |         | -0.04  | 0.118 | -0.052 | 0.014 |
| Lab    | 407 / 327 (55.4%) | 0.782 | 0.699 | 0.9 | 0.787 |
|        |         | 0.84 | 0.833 | 0.883 | 0.857 |
|        |         | 0.058 | 0.134 | -0.017 | 0.070 |
| Vax    | 142 / 632 (18.3%) | 0.507 | 0.653 | 0.170 | 0.270 |
|        |         | 0.751 | 0.474 | 0.274 | 0.347 |
|        |         | 0.244 | -0.1786 | 0.104 | 0.077 |

Dataset Volume and Overlap by Theory.

COVID Misinformation
Not COVID Misinformation

Gates
Lab
Vax
Figure 3. Dataset volume and overlap by theory. Node size indicates total number of tweets discussing each conspiracy theory, while edge thickness corresponds to the number of tweets discussing any pair of conspiracy theories simultaneously. Color represents the proportion of tweets falling under the associated label.

4.2 Analysis

Sentiment Analysis

Figure 4 shows “Gates”-related tweets by net sentiment score over time. The range in sentiment is significantly greater for COVID misinformation, with tweets more consistently at net -10 sentiment or less, especially in April and May. See appendix for additional figures related to other conspiracy theories.

Figure 4. Sentiment comparison for “Gates” data by label. Color indicates the number of tweets found on that date with corresponding net sentiment score.
Figure 5. Sentiment comparison for each theory by classification. Red lines correspond to COVID misinformation and blue lines correspond with tweets labeled not COVID misinformation.

Figure 5 shows the daily average sentiment of tweets in each conspiracy theory subset across eight emotions and general “negative” or “positive” sentiment. While tweets related to “5G” conspiracies are similar between misinformation and non-misinformation, there are clear differences in the other four conspiracy theories. In general, tweets classified as misinformation tend to rate higher on negative sentiment, “fear”, “anger” and “disgust” compared to tweets classified as not misinformation.
Geospatial Analysis

Figures 6 and 7 show examples of geospatial analyses for different misinformation theories, with replies and retweets removed. See appendix for additional geospatial renderings of remaining conspiracy topics.

A large percentage of tweets corresponded with major population centers including London, Los Angeles, New York City, Paris, and Washington DC, as expected. However, additional patterns indicate locations with higher engagement with the conspiracy theory. For example, Figure 6 shows discussion of the “5G” conspiracy theory was widespread throughout Europe where this theory was especially prominent. Of note, because this study used English regular expression patterns, results are skewed to English-speaking locations.

Figure 6. Geospatial rendering of tweets about 5G COVID-19 conspiracy theory.
Figure 7. Geospatial rendering of tweets about Gates mandatory vaccine conspiracy theory.

4.3 Dynamic Topic Modeling
Figure 8. Topic distribution (number of tweets) over time for 2-topic dynamic topic model of tweets related to Bill and Melinda Gates. Number of tweets (y-axis) assigned to each topic in the 2-topic dynamic topic model, over time (x-axis). Tweets belonging to topic 1 are more common in the conversation in January, while topic 2 becomes more prominent in the spring. Additionally, there are distinct peaks showing the popularity of tweets related to this conspiracy theory overall.
Figure 9. Word cloud and topic evolution for Topics 1 and 2 for tweets related to Bill and Melinda Gates. Top Panel: Word evolution, the change in word importance over time, for tweets related to Bill and Melinda Gates. The x-axis represents time while the y-axis shows important words. Color represents the importance of words, with darker color denoting higher importance. Bottom Panel: Word clouds for each topic. Word size corresponds to word weight (higher weighted words appear larger).
For each conspiracy theory dataset, dynamic topic modeling was used to identify two to five potential subtopics and understand their evolution over time. The optimal model was assessed qualitatively by reviewing the results. Models with two topics led to the best results (qualitatively coherent topics with the least overlap) for the Gates, 5G, and laboratory origin theories, while the model with three topics was qualitatively the best for the vaccine theory. Results for the Bill and Melinda Gates theory are discussed here, with the remaining theories visualized in the multimedia appendix.

The Bill and Melinda Gates theory was composed of two topics, distinguishable both in relative volume of subtopic tweets over time (Figure 8) and in evolution of important words over time (Figure 9). Both topics showed peaks of increased Twitter discussion in mid-January to mid-February, with reemergence in April. The initial peaks in Topic 1 corresponded to high weighting of the words “predicted”, “kill_65m”, “event”, and “simulation”, while the later spike in April showed higher weights in words like “fauci” and “buttar”. The model identified a second topic that referred to several conspiracy theories about Bill Gates, coronavirus, and vaccines. This second topic initially focused on theories about the origins of the virus with highly weighted words like “pirbright”, and “patent”. In late April, higher weight is given to the words “kennedy”, “jr”, and “fauci.”

The vaccine theory showed high weighting for the word “Bakker” in Topic 1 and a brief increase in the word “microchip” in early April within Topic 2. A linguistic shift in referring to the virus was also observable within the vaccine theory, with “coronavirus” highly weighted until mid-March, when “COVID” became more frequently used.

In the laboratory origin theory, words like “biosafety”, “biowarfare”, “warned”, and “laboratory” were more highly weighted early in the outbreak. Words like “escaped”, “evidence”, and “originated” became more strongly weighted as the theory evolved over time. Overlap was seen between the laboratory theory and the Gates theory, with words such as “kill”, “kill_65m”, “kill_forget”. In addition, we observed terms related to other, more distant theories, such as “ebola” in Topic 2 in mid-January and terms related to Jeffrey Epstein and associated conspiracy theories surrounding his death (“epstein”, “forget_epstein”) [64].

The 5G theory showed high weighting of words and phrases like “conspiracy” and “conspiracy theory” in Topic 1 from late March through early April.

5. Discussion

5.1 Principle Findings

The best performing random forest models were for the “5G” and “Lab” theories, with F1 scores of 0.804 and 0.857, respectively. While results for the “Gates” theory were weaker, “Vax” was the most problematic, with a F1 score of 0.347. This is likely due to the imbalanced nature of the dataset; many of tweets we labeled were associated with scientifically valid news. In most cases, the addition of active learning produced at least minor improvements over the base models. Despite the variance in model performance, this approach was intended to increase the
proportion of tweets pertaining to each conspiracy theory, not to provide perfect classification, so results were deemed adequate for downstream analysis.

Within our labeled data, a slight majority of tweets were COVID-19 misinformation in all conspiracy theories aside from the vaccine-averse dataset. However, when looking at overlap between the vaccin-eaverse dataset and the other datasets (see Figure 3), misinformation made up at least half of the overlap in each case, with significant connections between the “Lab”, “Gates”, and “Vax” conspiracy theories.

Across the conspiracy theories analyzed, this study showed a variety of factors to be consistent. When considering sentiment, the upper bounds (positive sentiment) were similar between misinformation and non-misinformation, but the data consistently showed that misinformation corresponded to a greater range in negative sentiment. We observe a similar trend when considering sentiment categorically. Tweets classified as misinformation consistently averaged higher scores in the categories of “negative”, “fear”, “anger”, and “sadness”. These differences were the smallest when comparing misinformation and non-misinformation in the 5G data, despite the fact that this classifier was quite high performing. This could be a result of intense political polarization surrounding the rollout of 5G in Europe, even when discussed outside the context of COVID-19. Geospatial features proved useful in characterizing misinformation, not only in verifying intuitive trends in the data (high volume in population centers) but also in identifying spread in unexpected communities (conspiracy theories gaining adoption in Africa).

Dynamic topic modeling often identified specific conspiracy theories as they went from fringe discussion to larger topics. Early Gates terms of “predicted”, “kill_65m”, “event”, and “simulation” all refer to the simulation of a novel zoonotic coronavirus outbreak at Event 201, a global pandemic exercise co-hosted by the Johns Hopkins Center for Health Security, the World Economic Forum, and the Bill and Melinda Gates Foundation [65]. The simulation predicted a disease outbreak would spread to multiple countries and result in 65 million deaths. These terms were also observed within the first vaccine subtopic. The later Gates spike in April identified the words “fauci” and “buttar”, which correspond to news coverage that Dr. Rashid Buttar stated that coronavirus was manufactured to hurt the economy and that Dr. Fauci and Gates were using the pandemic to drive their hidden agendas [66]. Topic 2, which identified early terms of “pirbright”, and “patent”, corresponds to theories that the Gates Foundation funded and/or patented the virus through the Pirbright Institute, a UK based company. Later terms identified were “kennedy”, “jr”, and “fauci,” corresponding to claims from Robert Kennedy Jr. that a coronavirus vaccine would personally benefit Dr. Fauci and/or Bill Gates.

The vaccine theory identified “bakker” in Topic 1. This refers to the tele-evangelist Jim Bakker who promoted theories about colloidal silver as a cure for COVID-19 on his show and who was given a cease-and-desist letter from the federal government as a result [67]. A brief increase in the word “microchip” occurs a few weeks after the Bill Gates “Ask Me Anything” (AMA) on Reddit [68], which increased the popularity of theories that the Gates Foundation was attempting to use vaccines to microchip individuals for population control [24].

Laboratory origin words such as “biosafety”, “biowarfare”, “warned”, and “laboratory” suggest that people were discussing a malicious laboratory release [69], whereas later words of “escaped”, “evidence”, and “originated” correspond to theories tending to focus on an accidental
release of the virus from a laboratory [7] (Multimedia Appendix figure 11, topic 2). The laboratory theory also saw a high weight placed on the word “ebola,” another disease outbreak that sparked conspiracy theories surrounding bioengineering and laboratory origins. [9]. The laboratory origins topic also captured belief in unrelated conspiracy theories surrounding Jeffrey Epstein. This is consistent with studies which have shown that those who believe in a conspiracy theory are more likely to also believe in others, or are prone to conspiratorial thinking broadly [70,71].

The 5G theory identified “conspiracy” and “conspiracy theory” in Topic 1, suggesting an increase in people tweeting corrections. This is consistent with Ahmed 2020, which found that approximately 32% of tweets analyzed from March 27, 2020 to May 4, 2020 were criticizing the theory [24].

Within the dynamic topic modeling results, similar patterns repeat across theories. First, we see consistent evidence of overlap in conspiracy theories. Gates-related terms were seen outside the Gates theory, namely within laboratory origin theories and vaccine theories. The Gates theories are in many ways the quintessential conspiracy theory, identifying a very small group of powerful and wealthy people acting for their own benefit at the expense of the larger population [11].

Second, we repeatedly see the evolution of conspiracy theories evolving with reference to real-world events as they occur. This is especially evident in the 5G theory, where related hashtags trended on Twitter in the United Kingdom and resulted in suspected arson attacks on 5G towers [72]. We see this pattern in other theories as well. In the Gates conspiracy theory, we were able to tie the overlap between Gates and the vaccine-averse community to his March “Ask Me Anything” on Reddit, which highlighted Gates-funded research at the Massachusetts Institute of Technology to develop injectable invisible ink to serve as permanent records of vaccination for use in developing countries [24,25]. This association morphed into a conspiracy theory suggesting the COVID-19 vaccine would secretly microchip individuals for population control [24]. Further, within the laboratory origin theory, we see evolution from a theory that focuses on the malicious release of COVID [69] to a more passive and accidental release, and the corresponding connections to political leaders in the United States. Terms like “Trump” and “Pompeo” (the US Secretary of State) gained importance in early May when Pompeo suggested there was evidence linking the release of COVID to a laboratory in Wuhan [73].

5.2 Limitations

One of the difficulties of this work was the active nature of the subject matter. Not only has COVID-19 misinformation continued to spread past our May data cutoff, but emergent conspiracy theories and topics are still being traced back to the early months of the year. For instance, our research into claims of a “laboratory origin” for the virus focused on popular conspiracy theories around a Chinese lab in Wuhan, a Canadian lab, and Fort Detrick in the U.S. However, at the time of writing, two additional theories were identified which have gained traction, one that the virus originated from the French Pasteur Institute and another that it originated in a lab at the University of North Carolina [23].
Further, our exclusive use of Twitter data fails to capture the entirety of the spread of misinformation. Social media platforms have faced significant challenges in identifying and containing the spread of misinformation throughout the course of the COVID-19 pandemic. While Twitter data can be used to identify links from other platforms, without a means to analyze the source content, the ability to determine when misinformation crosses platforms to and from Twitter is limited.

Twitter users are also known to be a demographically biased sample of the US population. In 2012, Twitter users represented 16% of internet users, and were more likely to be aged 18-29 years, African-American, and urban. By 2015, Twitter users represented 23% of all internet users, including 30% of internet users under age 50. Minority groups are over-represented in some geographic regions and under-represented in others, and geographic regions are also differentially represented, with much of the Midwest significantly under-represented. As technological barriers to internet access decline, user demographics have shifted, with users becoming increasingly more educated and wealthy over time.

Lastly, while we are aware that our study population consisting of a 1% Twitter sampling subset of coronavirus-related keywords and hashtags is not generalizable to the English-speaking U.S. population as a whole, the goal of this study is not to achieve generalizability, but rather to achieve internal validity by accurately categorizing sentiment and describing misinformation patterns within the Twitter population, defined by active Tweeters from January 21, 2020 to May 8, 2020.

5.3 Conclusions

We found that an approach combining regular expression filtering and machine learning classification can be used to isolate tweets about four known conspiracy theories. Using this approach, we increased the proportion of tweets relevant to each conspiracy theory, which allowed for downstream sentiment analysis, unsupervised learning, and geospatial analysis with minimal manual components. Combining these methods allowed for a rich description of the evolution of each conspiracy theory over the course of the COVID-19 outbreak from January to May 2020. We show common patterns of theory evolution and conspiracy theory intersection over this time period, and describe how real world events impact the trajectory of conspiracy theories in online spaces.

This approach is valuable for characterizing misinformation that poses concerns to public health and the approach taken here is generalizable to types of misinformation spread in the health domain and beyond. An ability to assess conspiracy theories before they become widespread would allow public health professionals to craft effective messaging to preempt misperceptions rather than to combat established false beliefs.

5.4 Future Work

The difficulty in identifying new conspiracy theories, the time sensitivity of their development, and the intensely manual nature of creating labeled data, have shown the limitations imposed by relying on a traditional approach to identification using surveys or hand-labeled training data sets. Future efforts would benefit from an increased focus on
unsupervised methods for identification of new misinformation topics. While verification of any new approach would likely still be dependent on hand-labeled data, the relatively short half-life of "expert rated" data when applied to misinformation would suggest the need for unsupervised validation methods as well.

In addition, the methodology described in this paper did not distinguish between tweets promoting misinformation and those refuting it in some way. Prior work has identified both rumor-correcting and rumor-promoting tweets during crises using Twitter data [78]. To promote a more complete view of how misinformation spreads and evolves, future work would benefit from adopting this approach.

**Acknowledgements**

DG, NP, ARD, CWR, GF, NYVC collected and analyzed data. DG and ARD labeled data for supervised learning classifiers. NP ran dynamic topic models which NP, DG, CDS and ARD analyzed. TP performed geospatial analyses. DG ran sentiment analysis and built the supervised models. CDS and DG wrote the initial manuscript. All authors contributed to critical revisions of the manuscript. ARD led the project.

We thank NewsGuard for licensing the data. Research support provided by the UC Office of the President through its UC National Laboratory Fees Research Program (Award # LFR-18-547591), the Laboratory Directed Research and Development program of Los Alamos National Laboratory (project no. 20200721ER), and the US Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (Contract No. 89233218CNA000001). LA-UR-20-28305

**Conflicts of Interest**

None declared.

**Abbreviations**

AMA - “Ask Me Anything”
COVID-19 - coronavirus disease 2019
LDA - Latent Dirichlet Allocation
SARS-CoV-2 - severe acute respiratory syndrome coronavirus 2
WHO - World Health Organization

**6. References**

1. PRO/AH/EDR. Undiagnosed pneumonia - China (HU): Request For Information. Pro-MED. Published December 30, 2019. https://promedmail.org/promed-post/?id=6864153
2. @WHO. #China has reported to WHO a cluster of #pneumonia cases —with no deaths— in
Wuhan, Hubei Province. Investigations are underway to identify the cause of this illness. Published online January 4, 2020. https://twitter.com/WHO/status/1213523866703814656?s=20.

3. @WHOWPRO. According to the latest information received and @WHO analysis, there is evidence of limited human-to-human transmission of #nCOV. This is in line with experience with other respiratory illnesses and in particular with other coronavirus outbreaks. Published online January 18, 2020. https://twitter.com/WHOWPRO/status/1218741294291308545.

4. Coronavirus Bioweapon – How China Stole Coronavirus From Canada And Weaponized It. GreatGameIndia. Published January 26, 2020. Accessed August 18, 2020. https://web.archive.org/web/20200313192627/https:/greatgameindia.com/coronavirus-bioweapon/

5. Durden T. Did China Steal the Coronavirus and Weaponize It? ZeroHedge.

6. Deutz G. How One Particular Coronavirus Myth Went Viral. Published March 19, 2020. Accessed August 18, 2020. https://www.wired.com/story/opinion-how-one-particular-coronavirus-myth-went-viral/

7. Knapp A. The original pandemic: Unmasking the eerily familiar conspiracy theories behind the Russian flu of 1889. Forbes. Published 2020. https://www.forbes.com/sites/alexknapp/2020/05/15/the-original-plandemic-unmasking-the-eerily-parallel-conspiracy-theories-behind-the-russian-flu-of-1889/#3c4e54ff50d5

8. Kouzy R, Abi Jaoude J, Kraitem A, et al. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. Cureus. 2020;12(3):e7255. doi:10.7759/cureus.7255

9. Ognyanova K, Lazer D, Robertson RE, Wilson C. Misinformation in action: fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. Misinformation Rev. 2020;1(4):1-19. doi:10.37016/mr-2020-024

10. Nyhan B, Reifler J. When Corrections Fail: The Persistence of Political Misperceptions. Polit Behav. 2010;32(2):303-330. doi:10.1007/s11109-010-9112-2

11. Uscinski JE. Conspiracy Theories: A Primer. Rowman & Littlefield; 2020.

12. Turse N. How This Pastor of a Megachurch Is Fueling Ebola Conspiracy Theories. Published October 18, 2019. Accessed September 9, 2020. https://time.com/5703662/ebola-conspiracy-theories-congo/

13. Shultz RH, Godson R. Dezinformatsia: Active Measures in Soviet Strategy. Pergamon-Brassey’s; 1984.

14. Jowett G, O’Donnell V. Propaganda & Persuasion. Seventh edition. SAGE; 2019.

15. Gorman SE, Gorman JM. Denying to the Grave: Why We Ignore the Facts That Will Save Us. Oxford University Press; 2017.

16. Oswald E, de Looper C. Verizon 5G rollout: Everything you need to know. Digital Trends. Published May 28, 2020. Accessed September 9, 2020. https://www.digitaltrends.com/mobile/verizon-5g-rollout/

17. Holmes A. 5G Cell Service Is Coming. Who Decides Where It Goes? New York Times. Published March 2, 2018. Accessed September 9, 2020. https://www.nytimes.com/2018/03/02/technology/5g-cellular-service.html

18. Morgan A. What Is The Truth Behind the 5G Coronavirus Conspiracy Theory? Euronews. Published 2020. Accessed June 22, 2020. https://www.euronews.com/2020/05/15/what-is-the-truth-behind-the-5g-coronavirus-conspiracy-theory-culture-clash.

19. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res. 2020;22(5):e19458. doi:10.2196/19458
20. Wakefield J. How Bill Gates became the voodoo doll of Covid conspiracies. BBC News. Published June 5, 2020. Accessed September 9, 2020. https://www.bbc.com/news/technology-52833706

21. Wilhelm I. Bill Gates Promotes “Creative Capitalism” to Aid the World’s Poor. The Chronicle of Philanthropy. Published January 24, 2008. Accessed September 9, 2020. https://www.philanthropy.com/article/Bill-Gates-Promotes-Creative/163247

22. Bowman A. The flip side to Bill Gates’ charity billions. New Internationalist. Published April 1, 2012. Accessed September 9, 2020. https://newint.org/features/2012/04/01/bill-gates-charitable-giving-ethics/

23. Gregory J, McDonald K. Trail of Deceit: The Most Popular COVID-19 Myths and How They Emerged. NewsGuard. Published June 2020. Accessed September 2, 2020. https://www.newsguardtech.com/covid-19-myths/

24. Goodman J, Carmichael F. Coronavirus: Bill Gates ‘microchip’ conspiracy theory and other vaccine claims fact-checked. BBC. Published May 30, 2020. Accessed September 2, 2020. https://www.bbc.com/news/52847648

25. Weintraub K. Invisible Ink Could Reveal whether Kids Have Been Vaccinated. Scientific American. Published December 18, 2019. Accessed September 9, 2020. https://www.scientificamerican.com/article/invisible-ink-could-reveal-whether-kids-have-been-vaccinated/

26. Jeppsson A. How East Germany Fabricated the Myth of HIV Being Man-Made. J Int Assoc Provid AIDS Care JIAPAC. 2017;16(6):519-522. doi:10.1177/2325957417724203

27. Blancou P, Vartanian J-P, Christopherson C. Polio Vaccine Samples Not Linked to Aids. 2001;Nature(410):1045-1046.

28. Feuer A. The Ebola Conspiracy Theories. The New York Times. Published October 18, 2014. Accessed September 2, 2020. https://www.nytimes.com/2014/10/19/sunday-review/the-ebola-conspiracy-theories.html

29. Pradhan P, Pandey AK, Mishra A, et al. Uncanny Similarity of Unique Inserts in the 2019-NCoV Spike Protein to HIV-1 Gp120 and Gag. Evolutionary Biology; 2020. doi:10.1101/2020.01.30.927871

30. @ARanganathan72. Oh my god. Indian scientists have just found HIV (AIDS) virus-like insertions in the 2019-nCoV that are not found in any other coronavirus. They hint at the possibility that this Chinese virus was designed [“not fortuitous’]. Scary if true. Published online January 31, 2020.

31. @DouthatNYT. Ummmmmmmmmm … [Link to @ARanganathan72]. Published online January 31, 2020.

32. Samorodnitsky D. Don’t believe the conspiracy theories you hear about coronavirus and HIV: Especially if you work for the New York Times. Massive Science. Published January 31, 2020. Accessed August 7, 2020. https://massivesci.com/notes/wuhan-coronavirus-ncov-sars-mers-hiv-human-immunodeficiency-virus/

33. Singh M, Davidson H, Borger J. Trump claims to have evidence coronavirus started in Chinese lab but offers no details. The Guardian. Published April 30, 2020. Accessed September 2, 2020. https://www.theguardian.com/us-news/2020/apr/30/donald-trump-coronavirus-chinese-lab-claim

34. Wallbank D, Bloomberg. Twitter applies another fact check—this time to China spokesman’s tweets about virus origins. Fortune. Published May 28, 2020. Accessed September 9, 2020. https://fortune.com/2020/05/28/twitter-fact-check-zhao-lijian-coronavirus-origin/
35. Galbraith A. Twitter flags China spokesmen’s tweet on COVID-19. Reuters. Published April 30, 2020. https://www.reuters.com/article/us-twitter-china-factcheck/twitter-flags-china-spokesmans-tweet-on-covid-19-idUSKBN235061.

36. Pruss D, Fujinuma Y, Paul MJ, et al. Zika discourse in the Americas: a multilingual topic analysis of Twitter. *PLOS ONE*. 2019;14(5):e0216922. doi:10.1371/journal.pone.0216922

37. Klofstad CA, Uscinski JE, Connolly JM, West JP. What drives people to believe in Zika conspiracy theories? *Palgrave Commun*. 2019;5(1):36. doi:10.1057/s41599-019-0243-8

38. Stewart-Peters E, Kevin C. A short history of vaccine objection, vaccine cults and conspiracy theories. The Conversation. Published July 9, 2017. https://theconversation.com/a-short-history-of-vaccine-objection-vaccine-cults-and-conspiracy-theories-78842

39. Bogart LM, Wagner G, Galvan FH, Banks D. Conspiracy beliefs about HIV are related to antiretroviral treatment nonadherence among african american men with HIV. *J Acquir Immune Defic Syndr*. 1999;10(5):648-655. doi:10.1097/00042014-199910000-00026

40. Jaiswal J, Singer SN, Siegel K, Lekas H-M. HIV-related “conspiracy beliefs”: lived experiences of racism and socio-economic exclusion among people living with HIV in New York City. *Cult Health Sex*. 2019;21(4):373-386. doi:10.1080/13691058.2018.1470674

41. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Soc Sci Med*. 2019;240:112552. doi:10.1016/j.socscimed.2019.112552

42. Johnson NF, Velásquez N, Restrepo NJ, et al. The online competition between pro- and anti-vaccination views. *Nature*. 2020;582(7811):230-233. doi:10.1038/s41586-020-2281-1

43. Larson HJ. The biggest pandemic risk? Viral misinformation. *Nature*. 2018;562(7727):309-309. doi:10.1038/d41586-018-07034-4

44. Memon SA, Carley KM. Characterizing COVID-19 Misinformation Communities Using a Novel Twitter Dataset. *ArXiv200800791 Cs*. Published online September 19, 2020. Accessed December 8, 2020. http://arxiv.org/abs/2008.00791

45. Broniatowski DA, Kerchner D, Farooq F, et al. The COVID-19 Social Media Infodemic Reflects Uncertainty and State-Sponsored Propaganda. *ArXiv200709682 Phys*. Published online July 19, 2020. Accessed December 8, 2020. http://arxiv.org/abs/2007.09682

46. Pulido CM, Villarejo-Carballido B, Redondo-Sama G, Gómez A. COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. *Int Sociol*. 2020;35(4):377-392. doi:10.1177/0268580920914755

47. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR Public Health Surveill*. 2020;6(2):e18444. doi:10.2196/18444

48. Gupta L, Gasparyan AY, Misra DP, Agarwal V, Zimba O, Yessirkepov M. Information and Misinformation on COVID-19: a Cross-Sectional Survey Study. *J Korean Med Sci*. 2020;35(27):e256. doi:10.3346/jkms.2020.35.e256

49. Singh L, Bansal S, Bode L, et al. A first look at COVID-19 information and misinformation sharing on Twitter. *ArXiv200313907 Cs*. Published online March 30, 2020. Accessed December 8, 2020. http://arxiv.org/abs/2003.13907

50. Uscinski JE, Enders AM, Klofstad C, et al. Why do people believe COVID-19 conspiracy theories? *Harv Kennedy Sch HKS Misinformation Rev*. Published online 2020. doi:https://doi.org/10.37016/mr-2020-015

51. Roozenbeek J, Schneider CR, Dryhurst S, et al. Susceptibility to misinformation about
COVID-19 around the world. *R Soc Open Sci.* 2020;7(10):201199. doi:10.1098/rsos.201199

52. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health Surveill.* 2020;6(2):e19273. doi:10.2196/19273

53. ElSherief M, others. Hate Lingo: A Target-based Linguistic Analysis of Hate Speech in Social Media. In: *Proc. 12th International AAAI Conference on Weblogs and Social Media*. AAAI; 2018. https://arxiv.org/pdf/1804.04257.pdf

54. Tay Y, Tuan LA, Hui SC. COUPLENET: Paying Attention to Couples with Coupled Attention for Relationship Recommendation. In: *Proc. 12th AAAI Int. Conf. on Weblogs and Social Media*. AAAI; 2018. https://arxiv.org/pdf/1805.11535.pdf

55. Daughton AR, Paul MJ. Identifying Protective Health Behaviors on Twitter: Observational Study of Travel Advisories and Zika Virus. *J Med Internet Res.* 2019;21(5). doi:10.2196/13090

56. Saha K, Weber I, De Choudhury M. A Social Media Based Examination of the Effects of Counseling Recommendations After Student Deaths on College Campuses. In: *Proc. 12th International AAAI Conference on Weblogs and Social Media*. AAAI; 2018. Accessed July 16, 2018. http://koustuv.com/papers/ICWSM18_CounselingRecommendation.pdf

57. Rizoiu M-A, others. DEBATENIGHT: The Role and Influence of Socialbots on Twitter During the First 2016 U.S. Presidential Debate. In: *Proc. 12th International AAAI Conference on Weblogs and Social Media*. AAAI; 2018. https://arxiv.org/pdf/1802.09808.pdf

58. AFINN.; 2011. http://www2.imm.dtu.dk/pubdb/pubs/6010-full.html

59. Mohammad SM. *NRC Word-Emotion Association Lexicon (Aka EmoLex).*; 2010. Accessed September 1, 2020. http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

60. Nominatim. Accessed September 1, 2020. https://nominatim.openstreetmap.org/ui/search.html

61. Dot Distribution vs Graduated Symbols vs Proportional Symbol Maps. GIS Geography. Published July 16, 2018. Accessed September 2, 2020. https://gisgeography.com/dot-distribution-graduated-symbols-proportional-symbol-maps/

62. Blei DM, Lafferty JD. Dynamic Topic Models. In: *Proceedings of the 23rd International Conference on Machine Learning.* ; 2006.

63. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. *J Mach Learn Res.* 2003;3:993-1022.

64. Chaffin J. Epstein’s death proves feeding ground for conspiracy theories. Financial Times. Published November 22, 2019. Accessed August 31, 2020. https://www.ft.com/content/

65. The Event 201 Scenario. Event 201. Accessed September 1, 2020. https://www.centerforhealthsecurity.org/event201/scenario.html

66. Dr Buttar Accuses Fauci, Gates & The Media For Using COVID-19 To Drive Hidden Agenda. Youtube. Accessed September 1, 2020. https://www.youtube.com/watch?v=mbf9wccdxE

67. Ferré-Sadurní L, McKinley J. Alex Jones Is Told to Stop Selling Sham Anti-Coronavirus Toothpaste. New York Times. Published March 13, 2020. Accessed September 1, 2020. https://www.nytimes.com/2020/03/13/nyregion/alex-jones-coronavirus-cure.html

68. Gates B. I’m Bill Gates, co-chair of the Bill & Melinda Gates Foundation. AMA about COVID-19. Reddit. Published September 8, 2020. https://www.reddit.com/r/Coronavirus/comments/fksnbf/im_bill_gates_cochair_of_the_bill_melinda_gates/

69. Stanway D. China lab rejects COVID-19 conspiracy claims, but virus origins still a mystery. Reuters. Published April 28, 2020. Accessed September 2, 2020.
https://www.reuters.com/article/us-health-coronavirus-china-lab/china-lab-rejects-covid-19-conspiracy-claims-but-virus-origins-still-a-mystery-idUSKCN22A0MM
70. Lewandowsky S, Cook J. *The Conspiracy Theory Handbook.*; 2020. http://sks.to/conspiracy
71. Swami V, Voracek M, Stieger S, Tran US, Furnham A. Analytic thinking reduces belief in conspiracy theories. *Cognition.* 2014;133(3):572-585. doi:10.1016/j.cognition.2014.08.006
72. Brewis H. Nightingale hospital phone mast attacked as 5G conspiracy theory rages on. Evening Standard. Published April 14, 2020. https://www.standard.co.uk/news/uk/nhs-nightingale-phone-mast-arson-attack-5g-conspiracy-a4414351.html
73. Brewster J. Pompeo: ‘Enormous Evidence’ Linking Wuhan Lab To Covid Outbreak. Forbes. Published May 4, 2020. Accessed September 2, 2020. https://www.forbes.com/sites/jackbrewster/2020/05/03/pompeo-enormous-evidence-linking-wuhan-lab-to-covid-outbreak/#57d51a5f1432
74. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social Media Use in the United States: Implications for Health Communication. *J Med Internet Res.* 2009;11(4):e48. doi:10.2196/jmir.1249
75. Mislove A, Lehmann S, Ahn Y-Y, Onnela J-P, Rosenquist JN. Understanding the Demographics of Twitter Users. In: *ICWSM.* ; 2011. Accessed June 26, 2018. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2816/3234
76. Duggan M, others. *Social Media Update 2014.* Pew Research Center; 2015. Accessed August 2, 2016. http://www.pewinternet.org/2015/01/09/demographics-of-key-social-networking-platforms-2/
77. Duggan M, Smith A. 6% of Online Adults Are Reddit Users. Pew Research Center; 2013. Accessed August 4, 2016. http://www.pewinternet.org/2013/07/03/6-of-online-adults-are-reddit-users/
78. Arif A, Robinson JJ, Stanek SA, et al. A Closer Look at the Self-Correcting Crowd: Examining Corrections in Online Rumors. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing.* ACM; 2017:155-168. doi:10.1145/2998181.2998294

7. Multimedia Appendix

**Figure 1.** *Sentiment comparison for “5G” data by label.* Color indicates the number of tweets found on that date with corresponding net sentiment score.
Figure 2. Sentiment comparison for “Lab” data by label. Color indicates the number of tweets found on that date with corresponding net sentiment score.

Figure 3. Sentiment comparison for “Vax” data by label. Color indicates the number of tweets found on that date with corresponding net sentiment score.
Figure 4. Geospatial rendering of tweets about possible engineered origin of COVID-19.

**COVID-19 Contains HIV-like Insertions Suggesting Engineering**

1,296 (original & geolocated) of 2,591 total original tweets - 1/31/20 – 3/30/20

Figure 5. Geospatial rendering of tweets alleging Gates funding tied to COVID-19.
Figure 6. Geospatial rendering of tweets about COVID-19 being a manmade bioweapon.
Figure 7. Geospatial rendering of tweets about theory COVID-19 was predicted in simulation.
Figure 8. Geospatial rendering of tweets alleging COVID-19 was stolen from a Canadian lab.
Figure 9: Word cloud and topic evolution for topics of tweets related to 5G technology. **Top Panel**: Word evolution (change in word importance) over time for tweets. Here, the x-axis represents time, the y-axis shows important words, and the color represents importance of words with darker color denoting higher importance. **Bottom Panel**: Word clouds for each topic. Word size corresponds to word weight (larger words have a higher weight).
Figure 10: Word cloud and topic evolution for topics of tweets related to laboratory origins of the virus. **Top Panel:** Word evolution (change in word importance) over time for tweets. Here, the x-axis represents time, the y-axis shows important words, and the color represents importance of words with darker color denoting higher importance. **Bottom Panel:** Word clouds for each topic. Word size corresponds to word weight (larger words have a higher weight).
Figure 11. Word cloud and topic evolution for topics of tweets related vaccine discussion. **Top Panel:** Word evolution (change in word importance) over time for tweets. Here, the x-axis represents time, the y-axis shows important words, and the color represents importance of words with darker color denoting higher importance. **Bottom Panel:** Word clouds for each topic. Word size corresponds to word weight (larger words have a higher weight).