Title

Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news.

Journal

Journal of computational social science, 3(2)

ISSN

2432-2717

Authors

Shahsavari, Shadi
Holur, Pavan
Wang, Tianyi
et al.

Publication Date

2020

DOI

10.1007/s42001-020-00086-5

Peer reviewed

Journal of Computational Social Science
https://doi.org/10.1007/s42001-020-00086-5

RESEARCH ARTICLE

Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news

Shadi Shahsavari, et al. ***[full author details at the end of the article]***

Received: 15 July 2020 / Accepted: 10 September 2020

Abstract

Rumors and conspiracy theories thrive in environments of low confidence and low trust. Consequently, it is not surprising that ones related to the COVID-19 pandemic are proliferating given the lack of scientific consensus on the virus's spread and containment, or on the long-term social and economic ramifications of the pandemic. Among the stories currently circulating in US-focused social media forums are ones suggesting that the 5G telecommunication network activates the virus, that the pandemic is a hoax perpetrated by a global cabal, that the virus is a bio-weapon released deliberately by the Chinese, or that Bill Gates is using it as cover to launch a broad vaccination program to facilitate a global surveillance regime. While some may be quick to dismiss these stories as having little impact on real-world behavior, recent events including the destruction of cell phone towers, racially fueled attacks against Asian Americans, demonstrations espousing resistance to public health orders, and wide-scale defiance of scientifically sound public mandates such as those to wear masks and practice social distancing, countermand such conclusions. Inspired by narrative theory, we crawl social media sites and news reports and, through the application of automated machine-learning methods, discover the underlying narrative frameworks supporting the generation of rumors and conspiracy theories. We show how the various narrative frameworks fueling these stories rely on the alignment of otherwise disparate domains of knowledge, and consider how they attach to the broader reporting on the pandemic. These alignments and attachments, which can be monitored in near real time, may be useful for identifying areas in the news that are particularly vulnerable to reinterpretation by conspiracy theorists. Understanding the dynamics of storytelling on social media and the narrative frameworks that provide the generative basis for these stories may also be helpful for devising methods to disrupt their spread.

Keywords COVID-19 · Corona virus · Conspiracy theories · 5G · Bill Gates · China · Bio-weapons · Social media · 4Chan · Reddit · News · Rumor · Narrative · Machine learning · Networks · Data visualization

Published online: 28 October 2020

Introduction

As the COVID-19 pandemic continues its unrelenting global march, stories about its origins, possible cures and vaccines, and appropriate responses are tearing through social media and dominating the news cycle. While many of the stories in the news media are the product of fact-based reporting, many of those circulating on social media are anecdotal and the product of speculation, wishful thinking, or conspiratorial fantasy.

Given the lack of a strong scientific and governmental consensus on how to combat the virus, people are turning to informal information sources such as social media to share their thoughts and experiences, and to discuss possible responses. At the same time, the news media is reporting on the actions individuals and groups are taking across the globe, including ingesting home remedies or defying stay at home orders, and on the information motivating those actions.[1] Consequently, news and social media have become closely intertwined, with informal and potentially misleading stories entangled with fact-based reporting: social media posts back up claims with links to news stories, while the news reports on stories trending on social media. To complicate matters, not all sites purporting to be news media are reputable, while reputable sites have reported unsubstantiated or inaccurate information. Because of the very high volume of information circulating on and across these platforms, and the speed at which new information enters this information ecosystem, fact-checking organizations have been overwhelmed. The chief operating officer of Snopes, for example, has pointed out that, "[there] are rumors and grifts and scams that are causing real catastrophic consequences for people at risk... It's the deadliest information crisis we might ever possibly have," and notes that his group and others like it are "grossly unprepared" [40].

Devising computational methods for disentangling misleading stories from the actual news is a pressing need. Such methods could be used to support fact checking organizations, and help identify and deter the spread of misleading stories. Ultimately, they may help prevent people from making potentially catastrophic decisions, such as resisting efforts at containment that require participation by an entire citizenry or self-medicating with chloroquine phosphate, bleach or alcohol.[2]

As decades of research into folklore has shown, stories such as those circulating on social media, however anecdotal, are not created from whole cloth, but rely on existing stories, story structures, and conceptual frameworks that inform the world view of individuals and their broader cultural groups [49, 62, 68, 71]. Taken

[1] The broad scale use of hydroxychloroquine as a possible treatment for the virus, which first gained a foothold in social media, is only one example of this feedback loop in action [57]. Numerous other claims, despite being repeatedly shown to be false, persist in part because of this close connection between social media and the news [7].

[2] An Arizona man died from using an aquarium additive that contained chloroquine phosphate as a source of chloroquine, a potential "miracle cure" touted by various sources, including the U.S. president[72]. Several poison control centers had to make press releases warning people not to drink or gargle with bleach[12]. The governor of Nairobi included small bottles of cognac in the COVID-19 care kits distributed to citizens, erroneously indicating that WHO considers alcohol a "throat sanitizer" [22].

together, these three features (a shared world view, a reservoir of existing stories, and a shared understanding of story structure) allow people to easily generate stories acceptable to their group, for those stories to gain a foothold in the narrative exchanges of people in those groups, and for individuals to try to convince others to see the world as they do by telling and retelling those stories.

Inspired by the narratological work of Algirdas Greimas [28], and the social discourse work of Joshua Waletzky and William Labov [36], we devise an automated pipeline that determines the frameworks that form the narrative bedrock of diverse knowledge domains, in this case those related to the COVID-19 pandemic [67]. We also borrow from George Boole's famous definition of a domain of discourse, recognizing that in any such domain, there are informal and constantly negotiated limits on what can be said:"In every discourse, whether of the mind conversing with its own thoughts, or of the individual in his intercourse with others, there is an assumed or expressed limit within which the subjects of its operation are confined" [10]. We conceptualize a narrative framework as a network comprising the actants (people, organizations, places and things) and the interactant relationships that are expressed in any storytelling related to the pandemic, be it a journalistic account or an informal anecdote [67, 69]. In our model of story telling, individuals usually activate only a small subset of the available actants and interactant relationships that exist in a discourse domain, thereby recognizing that individual storytelling events are often incomplete. This story incompleteness presupposes knowledge of the broader narrative framework on the part of the storyteller's interlocutors.

Building on folkloric work in rumor and legend, we further recognize that a large number of the stories circulating on and across social networks have a fairly straight forward "threat narrative" structure, comprised of an *orientation* (the who, what, where and when), a *complicating action: threat* (identifying who or what is threatening or disrupting the in-group identified in the orientation), a *complicating action: strategy* (a proposed solution for averting the threat), and a *result* (the outcome of applying that strategy to the threat) [68].

To determine the extent of narrative material available—the actants and their complex, content dependent interactant relationships—we aggregate all the posts or reports from a social media platform or news aggregator site. For social media in particular, we recognize that participants in an online conversation rarely recount a complete story, choosing instead to tell parts of it [38]. Yet even partial stories activate some small group of actants and relationships available in the broader discourse. We conceptualize this as a weighting of a subgraph of the larger narrative framework network.

By applying the narrative framework discovery pipeline to tens of thousands of English-language social media posts and news stories, primarily focused on events in the United States and all centered on conspiracy theories related to the COVID-19 pandemic, we uncover five central phenomena: (i) the attempt by some conspiracy theorists to incorporate the pandemic into well-known conspiracy theories, such as Q-Anon; (ii) the emergence of new conspiracy theories, such as one aligning the domains of telecommunications, public health, and global trade, and suggesting that the 5G cellular network is the root cause of the pandemic; (iii) the alignment of various conspiracy theories to form larger ones, such as one suggesting that Bill Gates is

using the virus as a cover for his desire to create a global surveillance state through the enforcement of a worldwide vaccination program, thereby aligning the conspiracy theory with anti-vaccination conspiracy theories and other conspiracy theories related to global cabals; (iv) the nucleation of potential conspiracy theories, such as #filmyourhospital, that may grow into a larger theory or be subsumed in one of the existing or emerging theories; and (v) the interaction of these conspiracy theories with the news, where certain factual events, such as the setting up of tents in Central Park for a field hospital to treat the overflow of patients, are linked to conspiracy theories. In this particular case, the tents of the field hospital are linked to central aspects of the Pizzagate conspiracy theory, specifically child sex-trafficking, underground tunnels, and the involvement of highly visible public figures [34, 67].

Running the pipeline on a daily basis during the time of this study allows us to capture snapshots of the dynamics of the entanglement of news and social media, revealing ebbs and flows in the overall story graph, while highlighting the parts of the news graph that are susceptible to being linked to rumors and conspiratorial thinking.

Prior work

Conspiracy theories (along with rumors and other stories told as true) circulate rapidly when access to trustworthy information is low, when trust in accessible information and its sources is low, when high-quality information is hard to come by, or a combination of these factors [3, 6, 26, 41, 53, 54, 61]. In these situations, people exchange informal stories about what they believe is happening, and negotiate possible actions and reactions, even as events unfold around them. Research into the circulation of highly believable stories in the context of natural disasters such as Hurricane Katrina [43], man-made crises such as the 9/11 terrorist attacks in 2001 [24] and the Boston Marathon bombings in 2013 [66], or crises with longer time horizons such as disease [4, 30, 32], has confirmed the explanatory role storytelling plays during these events, while underscoring the impact that stories, including incomplete ones, can have on solidifying beliefs and inspiring real-world action [25, 35].

The goal of telling stories in these situations is at least in part to reach group-wide consensus on the causes of the threat or disruption, the possible strategies that are appropriate to counteract that threat, and the likely outcomes of a particular strategy [68]. In an unfolding crisis, stories often provide a likely cause or origin for the threat, and propose possible strategies for counteracting that threat; the implementation of those strategies can move into real-world action, with the strategy and results playing themselves out in the physical world. This pattern has repeated itself many times throughout history, including during recent events such as Edgar Welch's attempt to "free" children allegedly being held in a Washington DC pizza parlor [45], the genocidal eruptions that crippled Rwanda with paroxysms of violence in 1994 [19], and the global anti-vaccination movement that continues to threaten global health [13, 32].

Although the hyperactive transmission of rumors often subsides once credible and authoritative information is brought to the forefront [4, 70], the underlying narrative frameworks that act as a generative reservoir for these stories do not disappear. Even in times of relative calm where people have high trust in their information sources and high confidence in the information being disseminated through those sources, stories based on the underlying narrative frameworks continue to be told, and remain circulating with much lower frequency in and across various social groups. This endemic reservoir of narrative frameworks serves multiple cultural functions. It supports the enculturation of new members, proffering a dynamic environment for the ongoing negotiation of the group's underlying cultural ideology [18]. Also, it provides a ready store of explanatory communal knowledge about potential threats and disruptions, their origins and their particular targets, as well as a repertoire of potentially successful strategies for dealing with those threats [17]. When something does happen that needs explanation—and a possible response—but for which high trust or high confidence information sources do not exist, the story generation mechanism can shift into high gear.

The endemic reservoir of narrative frameworks that exists in any population is not immutable. Indeed, it is the ability of people to change and shape their stories to fit the specific information and explanatory needs of their social groups that makes them particularly fit for rapid and broad circulation in and across groups [68]. While the stability in a story telling tradition suggests that the actants and their relationships are slow to change, their constant activation through the process of storytelling leads to dynamic changes in the weighting of those narrative framework networks; new actants and relationships can be added and, if they become the subject of frequent storytelling, can become increasingly central to the tradition.

Because of their explanatory power, stories can be linked into cycles to form conspiracy theories, often bringing together normally disparate domains of human interaction into a single, explanatory realm [41, 33]. Although a conspiracy theory may not ever be recounted in its entirety, members of the groups in which such a theory circulates internalize, through repeated interactions, the "immanent" narrative that comprises the overall conspiracy theory [14]. In turn, conspiracy theories can be linked to provide a hermetic and totalizing world view redolent of monological thinking [27], and can thereby provide explanations for otherwise seemingly disjoint events while aligning possible strategies for dealing with the event to the story teller's cultural ideology [68].

Summarizing the storytelling of thousands of story tellers and presenting these stories in an organized fashion has been an open problem in folkloristics since the beginning of the field. The representation of narratives as network graphs has been a desiderata in narrative studies at least since the formalist studies of Vladimir Propp [50]. Lehnert, in her work on the representation of complex narratives as graphs, notes that these representations have the ability to "reveal which concepts are central to the story" [39]. In other work specifically focused on informal storytelling, Bearman and Stovel point out that, "By representing complex event sequences as networks, we are easily able to observe and measure structural features of narratives that may otherwise be difficult to see" [5]. Later work on diverse corpora including

national security documents has shown the applicability of the approach to a broad range of data resources [46].

The automatic extraction of these graphs, however, has been elusive given the computational challenges inherent in the problem. In the context of conspiracy theories, preliminary work has successfully shown how simple S–V–O (Subject–Verb–Object) extractions can be matched to a broader topic model of a large corpus of conspiracy theories [55]. Work in our group has shown how the extraction of more complex structures and their concatenation into complex narrative graphs provides a clear overview of, for example, the narrative framework supporting the decision to seek exemptions from vaccination among "antivax" groups posting on parenting blogs [69].

Recent work on rumors and conspiracies focuses specifically on the COVID-19 pandemic [1, 23, 74]. An analysis of German Facebook groups whose discussions center on the pandemic uses a similar named entity analysis to our methods, and shows a strong tendency among the Facebook group members to resist the news reported by recognized journalistic sources [9]. A study of the Chinese social media site Weibo revealed a broad range of concerns from disease origin and progression to reactions to public health initiatives [42]. An examination of 4Chan that employs network analysis techniques and entity rankings traces the emergence of Sino-phobic attitudes on social media; these attitudes are echoed in our narrative frameworks [58].

In earlier work, we have shown how conspiracy theories align disparate domains of human knowledge or interaction through the interpretation of various types of information not broadly accessible outside the community of conspiracy theorists [67]. We have also shown that conspiracy theories, like rumors and legends on which they are based, are opportunistic, taking advantage of low information environments to align the conspiracy theory to unexplained events in the actual news [4]. Such an alignment provides an overarching explanation for otherwise inexplicable events, and fits neatly into the world view of the conspiracy theorists. The explanatory power of this storytelling can also entice new members to the group, ultimately getting them to ascribe to the worldview of that group.

Data

Data for this study were derived from two main sources, one a concatenation of social media resources composed largely of forum discussions, and the other a concatenation of COVID-19-related news reports from largely reputable journalistic sources.

We devised a scraper to collect publicly available data from Reddit subreddits and from 4Chan threads related to the pandemic. The subreddits and threads were evaluated for relevance by three independent evaluators, and selected only if there

was consensus. All of the data are available in our Open Science Framework data repository [60].[3]

For 4Chan, we collected ∼ 200 links to threads for the term "coronavirus", resulting in a corpus of 14712 posts. The first post in our corpus was published on March 28, 2020 and the final post was published on April 17, 2020. For Reddit, we accessed ∼ 100 threads on various subreddits with 4377 posts scraped from the top comments. Because these top comments are not necessarily sorted by time but rather by the process of up-voting, we did not include these timestamps in our analysis. Specifically, we targeted r/coronavirus and r/covid19, along with threads from r/conspiracy concentrating on Corona virus. We removed images, URLs, advertisements, and non-English text strings from both sources to create our research corpus. After running our pipeline, we were able to extract 87079 relationships from these social media posts.

For news reports, we relied on the GDELT project, an Open Source platform that scrapes web news (in addition to print and broadcast) from around the world (https://www.gdeltproject.org/).[4] Our search constraints through this dynamic corpus of news reports included a first-order search for conspiracy theories. The corpus was subsequently filtered to only include articles written in English (GDELT built-in feature) from U.S. news sources. The top 100 news articles (as sorted by the GDELT engine) were aggregated daily from January 1, 2020 to April 14, 2020 (prior to filtering), and the body of each filtered news report was scraped with Newspaper3K. These articles were then cleaned and staged for our pipeline to extract sentence-level relationships between key actors. We extracted ∼ 60 relationships from each report, ∼ 50 filtered news reports per day, and 324510 relationships.

Methods

The methods used here are a refinement of those developed for an earlier study of conspiracies and conspiracy theories [67]. We estimate narrative networks that represent the underlying structure of conspiracy theories in a large social media corpus (4Chan, Reddit) where they are most likely to originate, and the corresponding reporting about them in the news (GDELT). This approach allows us to analyze the interplay between the two corpora and to track the time-correlation and pervasive flow of information from one corpus to the other. The latent structure of the social media networks also provides features which enable the identification of key actants (people, places and things) in conspiracies and conspiracy theories, and the detection of threat elements in these narratives. The following subsections introduce the graphical narrative model for conspiracy theories in social media as well as the

[3] We ensured that our data were free from personal identifying information (PII), and that our data collection was allowed by the terms of service of the two sites. To the best of our knowledge, neither our corpus nor the news data contain data from private discussions, private chat rooms, or any other sources with restrictions on access for public use.

[4] Research use of the platform is explicitly permitted on the GDELT "about" pages.

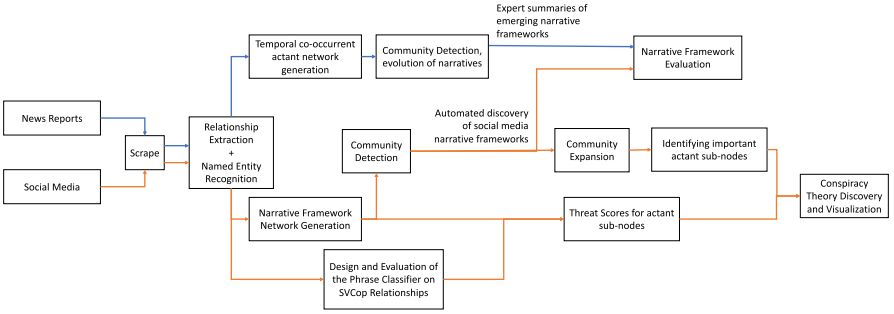


Fig. 1 Automated pipeline of processing data and discovering narrative networks in social media and news reports

pipeline for processing news reports. The end-to-end automated pipeline is summarized in Fig. 1.

A graphical narrative model and its estimation

We model narratives as generated by an underlying graphical model [67]. The narrative graph is characterized by a set of n nodes representing the actants, a set of r relationships $R = \{R_1, R_2, \ldots, R_r\}$ defining the edges, and k contexts $C = \{C_1, C_2, \ldots, C_k\}$ providing a hierarchical structure to the network. These parameters are either given *a priori* or estimated from the data. A context C_i is a hidden parameter, or the 'phase', of the underlying system that defines the particular environment in which the actants operate. It expresses itself in the distributions of the relationships among the actants, and is captured by a labeled and weighted network $G_{C_i}(V_{C_i}, E_{C_i})$. Here, $V_{C_i} = \{A_1, A_2, \ldots, A_n\}$, where each A_j is an actant. The edge set E_{C_i} consists of m_{C_i} ordered pairs $e_{C_i,j} = (A_{j_1}, A_{j_2})$, where each such pair is labeled with a distribution over the relationship set R.

Each post to a thread describes relationships among a subset of actants. A user picks a context C_i and then, from the network, draws a set of actants and inter-actant edges $G_{C_i}(V_{C_i}, E_{C_i})$. The forum posts constitute the output of this generative process. From a machine learning perspective, given a text corpus, we need to estimate all the hidden parameters of the model: the actants, the contexts, the set of relationships, and the edges and their labels. In other words, it is necessary to jointly estimate all the parameters of the different layers of the model.[5]

First, each sentence in our corpus is processed to extract various syntax relationship tuples. Some tuples are described as (arg_1, rel, arg_2) where each arg_i is a noun phrase and *rel* is a verb or other type of relational phrase. Others include Subject–Verb–*copula* (SVCop) relationships that connect a subject to a *copula* or subject complement. The relationship extraction framework combines dependency parse tree and Semantic Role Labeling (SRL) tools used in Natural Language

[5] The code for the pipeline is available on a github repository at https://tinyurl.com/ydftuez6

Table 1 Examples of contextual groups

Contextual group	Frequency
Virus, viruses, corona	2966
China	7865
Chinese, chinese, government	3622
Wuhan, lab	2691
Trumps, trump, donald	1913
Gate, gates, bill	1158
Jews, jew	572
USA, military	525
Doctor, doctors, nurse	520

Processing (NLP) [15]. We design relationship patterns that are frequently found in these narratives, and then extract tuples matching these patterns from the parsed corpus.

The phrases arg_1 and arg_2, along with entity mentions obtained using a Named Entity Recognition tool [2], provide information about the actant nodes and the contexts in which they appear. Noun phrases are aggregated across the entire corpus into contextually relevant groups referred to as super-nodes or Contextual Groups (CGs). Several different methods of clustering noun phrases and entity mentions into semantically meaningful groups have been proposed in the literature [51]. We accomplish the aggregation presented here by generating a list of names across the corpus, which then acts as a seed list to start the process of creating CGs from the noun phrases. We group nouns in the NER list using co-occurrence information from the noun phrases. For example, "corona" and "virus" tend to appear together in phrases in our corpus (see Table 1). We leverage recent work on word embedding that allows for the capture of both word and phrase semantics. In particular, the noun phrases in the same contextual group are clustered using an unsupervised k-Means algorithm on their BERT embeddings [20]. While the process of merging and deleting the subnodes detailed in prior work [67] offers flexibility over the choice of k we chose $k = 40$ as a conservative parameter to preserve the resolution of cluster members.

In prior work [67], we have shown that the final network comprising sub-nodes and their relationship edges is not sensitive to the exact size and content of the contextual groups derived at this stage. This CG grouping, which we undertake prior to applying k-Means clustering on word embedding, enables us to distill the noun phrases into groups of phrases that have a semantic bias. This distillation mitigates the inherent noise issues with word embeddings when they are directly used to cluster heterogeneous sets of words over a large corpus [56]. The CGs can also be viewed as defining macro-contexts in the underlying generative narrative framework. It is worth noting that the nodes are not disjoint in their contexts: a particular sub-node might have phrases that are also associated with other sub-nodes. These sub-nodes, when viewed as groups, act as different micro-contexts of the super-nodes in the overall model.

The final automatically derived narrative framework graph is composed of phrase-cluster nodes or, "sub-nodes", which are representations of the underlying actants. We automatically label these sub-nodes based on the TF-IDF scores of the words in each cluster resulting from the k-Means clustering of the contextual groups.

The edges between the sub-nodes are obtained by aggregating the relationships among all pairs of noun phrases. Thus, each edge has a set of relationship phrases associated with it, and the number of relationships can serve as the weight on that edge. The relationship set defining an edge can be further partitioned into semantically homogeneous groups by clustering their BERT embeddings [20, 67].

Narrative network communities and their relationship to conspiracy theories

Because conspiracy theories connect preexisting domains of human activity through creative speculation, often presented as being based on a theorist's access to "hidden knowledge", we expect that the narrative frameworks that we construct will have clusters of nodes and edges corresponding to the different domains. Since these clusters are densely connected within themselves, with a sparser set of edges connecting them to other clusters, we can apply community detection algorithms to discover them. For example, the domain of *"public health"* will have dense connections between sub-nodes such as *"doctors"* and *"hospitals"*, with relatively few connections to the domain of *"telecommunications"*, which will in turn have dense connections between sub-nodes such as *"5G"* and *"cell towers"*. Traversing these different communities mimics the conspiracy theorist's cross-domain exploration in the attempt to create a conspiracy theory.

Given the unsettled nature of discussions on social media concerning the COVID-19 pandemic, it seems likely that there are multiple, competing conspiracy theories in the corpus. Therefore, one would expect to find a large number of communities in the overall network, some isolated from the rest and others with a limited number of shared sub-nodes. One would also expect that this network would have a hierarchical structure.

To capture any such hierarchical structure, we compute overlapping network communities, where each community is defined by (i) a core set of nodes that constitute its backbone, and (ii) a set of peripheral nodes of varying significance that are shared with other communities. Currently, to determine the communities in our network, we run the Louvain greedy community detection algorithm multiple (~ 1000) times using the default resolution parameter in NetworkX [8]. We define *two nodes as belonging to the same core if they co-occur* in the same community for almost all of the runs; here we use a threshold of 850. As in [67], the threshold is aligned with the precipitous drop in the size of the Giant Connected Component (GCC).

Next, a core that defines a community is a set of nodes that is closed under the co-occurrence transitive relationship: If nodes A and B belong to the same core, and nodes B and C also belong to that same core then, by transitivity, we say nodes (A, B, C) are all in the same core. The resulting disjoint sets of core nodes (i.e., equivalence classes under the co-occurrence transitive relationship), along with their edges in their original network, define non-overlapping communities that form

the multitude of narrative frameworks in the corpus.[6] Overlapping nodes are then brought into the communities by relaxing the co-occurrence threshold [67].[7] These interactions among core communities, and hence, the respective narrative frameworks, capture the alignments among multiple knowledge domains that often underlie conspiracy theories.

Threats and special nodes in communities

Taken as a whole, the narrative framework comprising networks of actants and their inter-actant relationships (along with other metadata) reveals aspects of conspiracy theories including the threatening sub-nodes identified by the conspiracy theorists, and the possible strategies that they suggest for dealing with those threats. For instance, the network community consisting of sub-node [tower, 5g, danger] along with its associated SVCop relationships "[5g] is deadly", "[tower]s should be burned", imply a threat to human well-being posed by 5G, and a strategy for dealing with that threat: burn the cell towers (strategy) to protect people from the deadly 5G (threat). Because threats are often followed by strategies, we prioritize the classification of threats.

To classify threats, we look for sub-nodes in the network communities that, given their associated descriptions, might be considered threatening. For example, a descriptive reference to a sub-node "vaccines" that suggests that they "can kill", would allow us to code "vaccines" as a possible threat. We repeat this process for all the sub-nodes in the network communities, and find that strong negative opinions are associated with some subset of sub-nodes, which we identify as candidate threats. By applying a semi-supervised classification method to these candidate sub-nodes, we can confirm or reject our suspicions about their threatening nature.

The threat classifier is trained on the relationships extracted from social media posts. In particular, SVCop relationships (described in Section 4.1) play a special role in providing information about a particular sub-node: these relationships provide important information about the first argument and are generally descriptive in nature. In such relationships, the second argument is most often a descriptive phrase with an associated to-be verb phrase. For example (5g,is,dangerous/a threat/harmful) are SVCop relationships describing the [5g] argument. We consider these relationships as self-loops for their first arguments, which are aggregated into sub-nodes.

The most discussed sub-nodes tend to have a high number of such self-loop relationships, and the descriptive phrases often carry powerful characterizations of these entities. Sub-node-specific aggregation of these relationships can inform us about the role of a particular actant in its community. For example, we find ~ 350 SVCop relationships describing the node "virus" as "harmful", "deadly", "dangerous", and "not real".

[6] See [67] for details on how to select an optimal co-occurrence threshold and how to efficiently determine the core community networks.

[7] Future work will focus on testing other network community detection methods.

We aggregate the entire corpus of SVCop relationships ($\sim$ 10000) and then label them in a hierarchical fashion as follows: First, each such SVCop phrase is encoded using a 768 dimensional BERT embedding from a model fine-tuned for entailment detection between phrases [20]. Next, the vectors are clustered with HDBSCAN [44], resulting in a set of $\sim$ 1000 density-based clusters C, with an average cluster membership size of 7. Approximately, 3000 of the phrase encoding vectors are grouped in a cluster labeled as -1, indicating that they are not close to others and are best left as singletons. For the rest, each cluster represents a semantically similar group, and can be assigned a group semantic label. Thus, the task of meaningfully labeling $\sim$ 10000 phrases as 'threat' or 'non-threat' is reduced by almost a factor of 10.

We define a binary label for each cluster. A *threat* is a phrase that is universally recognized as threatening: [5g] is *dangerous*, [a tower] is *a bioweapon*. Here, the phrases *dangerous* and *bioweapon* are clearly indicative of threats. The remaining phrases are labeled as *neutral/vague* comments.

For every cluster $c \in C$, we assign a label l_c to c such that every descriptive phrase $d \in c$ is also assigned label l_c. Clearly, label quality is contingent on the manual labeling of the clusters and the semantic similarity of descriptive phrases as aggregated by the BERT and density-based clustering. This is ensured by three independent reviewers labeling each cluster and, in the case of disagreement, choosing the label receiving the majority vote (REPORT KRIPPENDORF ALPHA HERE).

We measure the inter-rater reliability with respect to the majority vote by the three different raters. Our results for a sample size of 100, are 0.745, 0.87 and 0.829.

The semantic similarity in each cluster is verified by a qualitative analysis of the clusters undertaken by domain experts. For example, most of the clusters have exact phrase matches such as

- **Cluster 1**: [the ongoing trade war, the trade war]
- **Cluster 2**: [radiation, radiation, more EM radiation, a result of radiation, electromagnetic radiation, also radiation]

that support high-fidelity hierarchical labeling. Other clusters validate the usage of BERT embedding as a means for clustering semantically similar phrases. For example,

- **Cluster 1**: [SLAVERY, members of race enslaved, a slave]
- **Cluster 2**: [a liberal hoax, a liberal lie designed]
- **Cluster 3**: [rabid supporters of SCIENCE, rabid supporters of SCIENCE, scientists f***]

capture semantic similarity in addition to exact matches of phrases.

Since our BERT model is fine-tuned to detect entailment, the clustering is sensitive to negation in particular, which is important in classifying phrases as threats. For example, the following clusters are distinct and complete:

- **Cluster 1**: [not convenient, not beneficial, not fun, not helpful]

– **Cluster 2**: [useful, helpful]

These labeled phrases are used to train a k-nearest neighbor (kNN)-based phrase classifier to identify threatening descriptions. Once again we use the fine-tuned BERT embedding.

Many competing kNN models provide useful classification results for phrases. We found that setting $k = 4$ results in a model that reasonably classifies most phrases. The kNN classifier is *binary*: 0 represents the class of *non-threat* and 1 represents the class of "threat". The cross-validation part is carried out at the level of the clusters: that is, when designing the training sets (for kNN, the set of phrases used in performing the kNN classification of a given phrase) and validation sets, we partition the phrases based on their cluster assignments. All phrases belonging to the same cluster are assigned to the same set and are not split across the training and validation sets. Because the labeled phrases have duplicate second arguments and repeated phrases occur in the same cluster, this approach to cross-validation ensures against repeating phrases in both the training and validation set, which is achieved by partitioning data at the cluster level.

The primary purpose of designing the phrase classifier is to identify threatening sub-nodes, which appear as core nodes in the narrative framework communities. Aggregated second arguments of SVCop relationships corresponding to a particular sub-node are classified with the kNN phrase classifier. Based on a majority vote on these second arguments, we can classify a sub-node as a potential threat. An outline of this algorithm is provided in Algorithm 1.

Algorithm 1 Threat classification of sub-nodes using majority vote

Input: A sub-node p, descriptive phrases in SVCop relationships of sub-node p, D_p and the $(k = 4)$ kNN model f
Output: s_p: A score between 0 and 1 assigned to p as an proportional indicator of threat

```
count ← 0
samples ← |D_p|
for each descriptive phrase arg_2 in D_p do
    label ← f(arg_2)
    count ← count + label
end for
s_p ← count / samples
```

Conspiracy theory narrative framework discovery

A narrative framework for a conspiracy theory, which may initially take shape as a series of loosely connected statements, rumors and innuendo, is composed from a selection of subnodes from one or more of these communities and their intra- and inter-community relationships. Each community represents a group of closely connected actant sub-nodes with those connections based on the context-dependent inter-actant relationships. Traversing paths based on these inter-actant relationships within and across communities highlights how members posting to the forums understand the overall discussion space, and provide insight into the negotiation process concerning main actants and inter-actant relationships.

This search across communities is guided by the extended overlapping communities (which connect the core communities), as described in 4.1, taking into consideration the sub-nodes that are classified as threat nodes. The inter-actant relationship paths connecting the dominant threat nodes, both within and across communities, are then pieced together to create the various conspiracy theories.

Searching conspiracy theories from social media in the news

Many conspiracy theories detected in social media are addressed in news reports. By temporally aligning the communities discovered from social media with the evolving communities emerging from news collected daily, we can situate the 4Chan commentary alongside mass media discussions in the news. Such a parallelism facilitates the analysis of information flow from smaller community threads in social media to the national news and from the news back to social media.

Extraction of inter-actant communities in the news

To aggregate the published news, we consider (1-day time-shifted) intervals of 5 days. This sliding window builds $s = 101$ segments from January 1, 2020 to April 14, 2020. We have discovered that a longer interval, such as the one chosen here, provides a richer backdrop of actants and their interactions than shorter intervals. In addition, narratives on consecutive days retain much of the larger context, highlighting the context-dependent emergence of new theories and key actants.

We use the major actants and their mentions discovered in the social media data to filter the named entities that occur in the news corpus. A co-occurrence network of key actants in news reports (conditioned on those discovered from social media), provides a day-to-day dynamic view of the emergence of various conspiracy theories over time. In addition, we model the flow of information between social media and news reports by monitoring the frequency of occurrence of social media communities (as captured by a group of representative actants in each community) in the text of news reports (see Evaluation). With minimal supervision, a few actant mentions are grouped together including, [trump, donald]: **donald trump**, [coronavirus, covid19, virus]: **coronavirus** and [alex, jones]: **alex jones**. While such groupings are not strictly required and could be done more systematically (see [59]), this actant-grouping enhances the co-occurrence graph by reducing the sparsity of the adjacency matrix representing subject–object interaction.

Co-occurrence actant network generation

For each 5-day segment of aggregated news reports, the corpus of extracted relationships R_i and the associated set of entities E_i are parsed with Algorithm 2 to yield a co-occurrence actant network. Day-to-day networks reveal the inter-actant dynamics in the news. While many metrics can be used for summarizing the dynamics within these networks, we considered the Number of Common Neighbors (NCN) between them. If the adjacent vertices of a_1 are S_{a_1} and of a_2 are S_{a_2}, the NCN score is defined as:

$$n_{a_1,a_2} = |S_{a_1} \cap S_{a_2}|. \tag{1}$$

Algorithm 2 Co-occurrence Actant Network Generation for a Segment $i < s$ of News

Input: R_i relationship tuples, E_i entities
Output: $G_i(R_i, E_i)$
  $M \leftarrow []$
  for $(arg_1, rel, arg_2) \in R_i$ **do**
    $s \leftarrow H(arg_1)$ $\{H(arg)$ is the headword of $arg\}$
    $o \leftarrow H(arg_2)$
    $r \leftarrow H(rel)$
    if $(s, o \in E_i)$ AND $(s \neq o)$ AND (r NOT stop word) **then**
      $M[s, o] \leftarrow M[s, o] + 1$
      $M[o, s] \leftarrow M[o, s] + 1$
    end if
  end for
  $M_{norm} =$ `normalize`(M) {along each row}
  $G_i(R_i, E_i) \leftarrow M$ {Color-coded based on the labels of actants decided by the Entity Extractor}

Evaluation

Temporal alignment of communities derived from news reports and social media

We temporally align the conspiracy theories discussed in social media and in news reports by first capturing a group of representative actants in each social media community. Let the set of keywords representing a particular community be V_i. The timestamps present in 4Chan and GDELT data make these corpora suitable for temporal analysis with respect to V_i (our Reddit corpus does not contain dates). To facilitate a comparison between the two corpora conditioned on V_i, let C_j denote the raw 4Chan data and D_j denote the raw GDELT news data in time-segment j. The time segments are 5-day intervals between March 28,2020 and April 14, 2020, which is the intersection of date ranges for which we have temporal 4Chan and GDELT data. We define a Coverage Score (m) that captures the presence of actants from V_i in C_j and D_j.

$$m_C(V_i, j) = \frac{\sum_{w_V \in V_i} \sum_{w_C \in C_j} \mathbb{1}(w_V = w_C)}{|V_i||C_j|}, \tag{2a}$$

$$m_D(V_i, j) = \frac{\sum_{w_V \in V_i} \sum_{w_D \in D_j} \mathbb{1}(w_V = w_D)}{|V_i||D_j|}. \tag{2b}$$

To normalize the coverage scores to a baseline, we compute a Relative Coverage Score (r), where V^* is a random set of actants (of size 500) as:

$$r_C(V_i,j) = \frac{m_C(V_i,j)}{m_C(V^*,j)}, \quad r_D(V_i,j) = \frac{m_D(V_i,j)}{m_D(V^*,j)}. \tag{3}$$

Computed across all time-segments, $r_C(V_i)$ and $r_D(V_i)$ represent a series of relative coverage scores for 4Chan and GDELT, respectively, with one sample for every time segment. This metric now provides a normalized measure for coverage of a community derived from social media in the temporal corpora of 4Chan and GDELT data.

The cross-correlation function of these relative coverage scores $\mathcal{R}_{C,D}(\tau) = E[r_C(V_i,t)r_C(V_i,t+\tau)]$ can provide interesting insight into the co-existence of conspiracy theory communities in the two corpora where τ is the number of offset days between the news and 4Chan data (see Fig. 5). This cross-correlation score peaks for the number of offset days that results in the maximum overlap of relative coverage scores. For example a τ of 10 days would imply that information about a specific set of representative actants occurred in the news and 4Chan data roughly 10 days apart. τ captures the latency or *periodicity* lag between communities mentioned in the news and in 4Chan data. The error bars are generated over 20 random communities used for normalizing the coverage scores before cross-correlation.

Other standard metrics to compare communities derived from the news and social media

We present standard metrics to further compare communities of actants derived from temporal news reports and social media. Our metrics are standard measurements used for clustering evaluations based on ground truth class labels [73]. Algorithm 3 describes this evaluation process.

Algorithm 3 Unsupervised evaluation of communities

Input: $C_{i,t}$ News community indexed i at time t, K_j Social media community indexed j
Output: Pr_t Percentage of coverage for time t, h_t Homogeneity at time t, c_t Completeness at time t, v_t V-Measure at time t

```
Y_gr ← []
Y_pred ← []
count ← 0
for each C_{i,t} do
    for each actant a in C_{i,t} and K_j do
        if a in K_j then
            count ← count + 1
            Y_gr[a] ← j
            Y_pred[a] ← i
        end if
    end for
end for
Pr_t ← count/|K_j|
h_t ← Homogeneity(Y_gr, Y_pred)
c_t ← Completeness(Y_gr, Y_pred)
v_t ← V-Measure(Y_gr, Y_pred)
```

Evaluation of phrase-based threat detection

We use average recall and average accuracy to evaluate the performance of the phrase-based threat classifier. The average is computed across the fivefold group-shuffled cross-validation of phrases. Here, recall and accuracy are defined as,

$$\mathtt{Recall} = \frac{\mathtt{DetectedThreats}}{\mathtt{Groundtruththreats}},$$
$$\mathtt{Accuracy} = \frac{\mathtt{Detectedthreats} + \mathtt{Detectednon} - \mathtt{threats}}{\mathtt{Sizeofthevalidationset}}.$$

Limitations

There are limitations with our approach, including those related to data collection, the estimation of the narrative frameworks, the labeling of threats, the validation of the extracted narrative graphs, and the use of the pipeline to support real-time analytics.

Data derived from social media sources tend to be very noisy, with considerable amounts of spam, extraneous and off-topic conversations, as well as numerous links and images interspersed with meaningful textual data. Even with cleaning, a large number of text extractions are marred by spelling, grammatical and punctuation errors, and poor syntax. While these problems are largely addressed by our NLP modules, they produce less accurate entity and relationship extractions for the social media corpus than for the news corpus. Also, unlike news articles which tend to be well archived, social media posts, particularly on sites such as 4Chan, are unstable, with users frequently deleting or hiding posts. Consequently, re-crawling a site can lead to the creation of substantively different target data sets. To address this particular challenge, we provide all of our data as an OSF repository [60].

The lack of consistent time stamping across and within social media sites makes determining the dynamics of the narrative frameworks undergirding social media posts difficult. In contrast to the news data harvested from the GDELT project, the social media data are marked by a coarse, and potentially inaccurate, time frame due to inconsistent time stamps or no time stamps whatsoever. Comparing a crawl from one day to the next to determine change in the social media forums may help attenuate this problem. Given the potential for significant changes due to the deletion of earlier posts, or the move of entire conversations to different platforms, the effectiveness of this type of strategy is greatly reduced. Because of the limited availability of consistently time-stamped data, our current comparison between the social media conspiracy theory narrative frameworks, and those appearing in the news, is limited to a three-week window.

There appears to be a fairly active interaction between the “Twittersphere” and other parts of the social media landscape, particularly Facebook. Many tweets, for instance, point to discussions on social media and, in particular, on Facebook. Yet,

because of restrictions on access to Facebook data for research purposes, we are unable to consider this phenomenon. Future work will incorporate tweets that link to rumors and other conspiracy theories in our target social media arena. As part of this integration, we also plan to include considerations of the trustworthiness of various Twitter nodes, and the amplification role that "bots" can play in the spread of these stories [16, 23].

As with a great deal of work on social media, there is no clear ground truth against which to evaluate or validate. This problem is particularly apparent in the context of folkloric genres such as rumor, legend and conspiracy theories, as there is no canonical version of any particular story. Indeed, since folklore is always a dynamically negotiated process, and predicated on the concept of variation, it is not clear what the ground truth of any of these narratives might be. To address this problem, we consider the narrative frameworks emerging from social media and compare them to those arising in the news media. The validation of our results confirms that our social media graphs are accurate when compared to those derived from news media.

Currently, our pipeline only works with English language materials. The modular nature of the pipeline, however, allows for the inclusion of language-specific NLP tools, for parsing of languages such as Italian or Korean, both areas hard hit by the pandemic, and likely to harbor their own rumors and conspiracy theories.

In addition, we believe that our semi-supervised approach to threat detection would require less human effort if we had more accurate semantic embeddings.

Finally, we must note that the social media threads, particularly those on 4Chan, are replete with derogatory terms and abhorrent language. While we have not deleted these terms from the corpus, we have, wherever possible, masked those terms in our tables and visualizations, with obvious swears replaced by asterisks, and derogatory terms replaced by "dt" for derogatory term, or "rdt" for racially derogatory term and a qualifier identifying the target group.

Results and evaluation

After running the pipeline and community detection, we find a total of two hundred and twenty-nine communities constituting the various knowledge domains in the social media corpus from which actants and interactant relationships are drawn to create narrative frameworks. Many of these communities consist of a very small number of nodes. It is worth noting that several of the communities are "meta-narrative" communities, and focus on aspects of communication in social media (e.g., communities 11 and 74), or platform specific discussions (e.g., communities 44 and 46 that focus on Facebook and 181 focusing on YouTube and Twitter). Other communities are "background" communities and focus on news coverage of the pandemic (e.g., communities 7 and 62), the background for the discussion itself (e.g., community 30 that connects the pandemic to death, and community 35 that focuses on hospitals, doctors, and medical equipment such as ventilators), or discussions of conspiracy theories in general (e.g., communities 108 and 109).

We find that these "meta-narrative" and "background" communities, after thresholding, tend to be quite small, with an average of 3.9 sub-nodes per community. Nevertheless, several of them include sub-nodes with very high NER scores, such as community 155, with only four nodes: "use", "microwave", "hybrid protein" and "cov", all with high NER scores. This community is likely to be included as part of more elaborated conspiracy theory narrative frameworks such as those related to 5G radiation.

The five largest communities, in contrast, range in size from 66 to 172 nodes. These five communities, along with several other large communities, form the main reservoir of actants and inter-actant relationships for the creation of conspiracy theory narrative frameworks. We find thirty communities with a node count ≥ 14. (See Fig. 2). Table 2 shows the temporary labels for these communities, which are based on an aggregation of the labels of the three nodes with the highest NER scores and node(s) with the highest-degree.

The relationship between the discussions occurring in social media and the reporting on conspiracy theories in the media changed over the course of our study period. In mid to late January, when the Corona virus outbreak appeared to be limited to the central Chinese city of Wuhan, and of little threat to the United States, news media reporting on conspiracy theories had very little connection to reporting on the Corona virus outbreak. As the outbreak continued through March 2020, the reporting on conspiracy theories gradually moved closer to the reporting on the broader outbreak. By the middle of April, reporting on the conspiracy theories being discussed in social media, such as those in our research corpus, had moved to a central position (see Fig. 3).

The connection between these two central concepts in the news—"coronavirus" and "conspiracy theory"—can also be seen in the rapid increase in the shared neighbors of these sub-nodes (defined in Eq. (1)) in the overall news graph during the period of study (see Fig. 4).

Since our dataset contains dated 4Chan *and* GDELT data from March 28, 2020 to April 14, 2020, communities from the social media corpus were explored within the subset of news media between the same dates using Relative Coverage Scores defined in Eq. (3). The cross-correlation of the ratio of coverage scores for different fixed communities to a random community is provided in Fig. 5.

The higher average scores for the "5G" community including words such as {"5g", "waves", "antenna", "radio", "towers", "radiation"}, suggests that this community was matched more frequently than other communities compared to a baseline random community. A peak at zero days offset within the time period from March 28, 2020 to April 14, 2020 implies that the news reports are correlated in time to 4Chan thread activity. In addition, these plots suggest that few communities dominate conspiracy theories more than others. The viability of other communities such as {"army", "us", "bioweapon"} and {"lab", "science", "wuhan"} suggests the lack of a single dominant conspiracy theory *consensus* narrative. Instead, it appears that numerous conspiracy theories may be vying for attention.

We examine "Bill Gates", a key actor frequently found in the common neighbors set between "coronavirus" and "conspiracy theory". Key relationships extracted by

Table 2 The largest thirty communities in the social media corpus in descending order of size

ID	Core size	Community label	High degree nodes
0	172	China, Government, End World	Bioweapon
1	89	Chinese, lab, research	Truth, animal(s)
5	88	Virus, 5G, cell	Bill Gates, vaccine
6	72	Coronavirus, flu, test	SARS
35	66	Chloroquine, doctor, patient	Hospital
41	53	Medium, fact, video	Chinese
21	39	Question, trump, impeachment	Fauci
30	39	Death, Connecticut, pandemic	CDC
51	32	Bacterial, post, economy	YouTube, bot
32	27	Medical, misinforming, life	Quarantine, Italy
56	25	Virus, 5G, vaccine	(Radio) frequency
58	24	Med, CIA, commie	Evidence
74	24	Physician, source, official	CNN
57	23	Journal, conference, quantum tattoo	Bill Gates
40	22	Chinese, lab, bat	Wet market
75	22	American, cognitive dissonance, question	Lost cause
7	19	Diagnosis, Fauci, wireless	Coronavirus
18	19	Wuhan, advancement opportunity, medical company	Flu
59	18	Guy, repugnant organization, backstab	Asymptomatic
82	17	Virus, cell, vaccine	Thermodynamic load
11	16	Financial, community, coronavirus	Biolab
42	16	Nation, consequence, dedicated worker	Attack
62	16	News, cancer, american	(Un)reliable
73	16	Accepted narrative, scientist, hand	Cell level
8	15	Corona, chan, 5g dumba**	Friend
15	14	Coronavirus, test, accurate	Spain
24	14	China, government, weapon	CCP, bioweapon
45	14	Coronavirus, SARS, flu	Disease
52	14	Medium, fact, video	Agenda
80	14	Shill, epidemic, defining factor	Political, blame

The labels are derived from the sub-node labels for the semantically meaningful nodes with the highest NER scores in each community (racially derogatory terms and swears have been skipped). The label of the highest degree node(s) not included in the community label is listed in the third column. Nodes with a threat score ≥ 0.5 are underlined

our pipeline on the news reports provide a qualitative overview of the emergence of "Bill Gates" as a key actor (see Table 5).

Finally, the evaluations based on Algorithm 3 are shown in Fig. 6. The plots indicate the saturation of completeness and homogeneity scores at $\sim 92\%$ and $\sim 82\%$ respectively across time. Similarly, the V-measure saturates at $\sim 86\%$. These scores per time sample, represent the fidelity of the process of cluster matching.

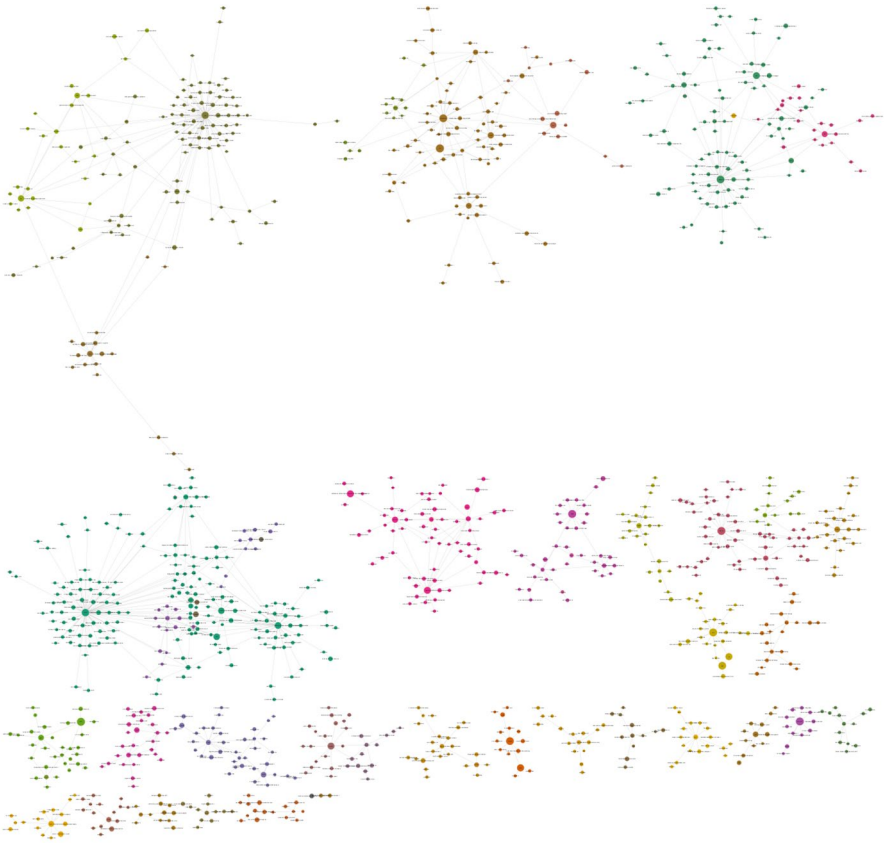


Fig. 2 Overview graph of the largest thirty communities in the social media corpus. Nodes are colored by community, and sized by NER score. Narrative frameworks are drawn from these communities, each of which describes a knowledge domain in the conversation. Nodes with multiple community assignments are colored according to their highest ranked community. An overarching narrative framework for a conspiracy theory often aligns subnodes from numerous domains

Descriptive phrase classification of threats from SVCop relationships

The phrase classifier described in the methods was cross-validated and the recall and accuracy across the validation sets are provided in Table 3. Recall is used as the primary performance measure in the detection of threats, as the sensitivity to threatening phrases is the most important feature of the classifier.

Classification of sub-nodes as threats

The phrase classifiers applied to descriptive phrases of a particular sub-node provide insight into the context of the sub-node. For the phrase classifier, Fig. 8 describes a histogram of the number of sub-nodes across the percentage of associated phrases

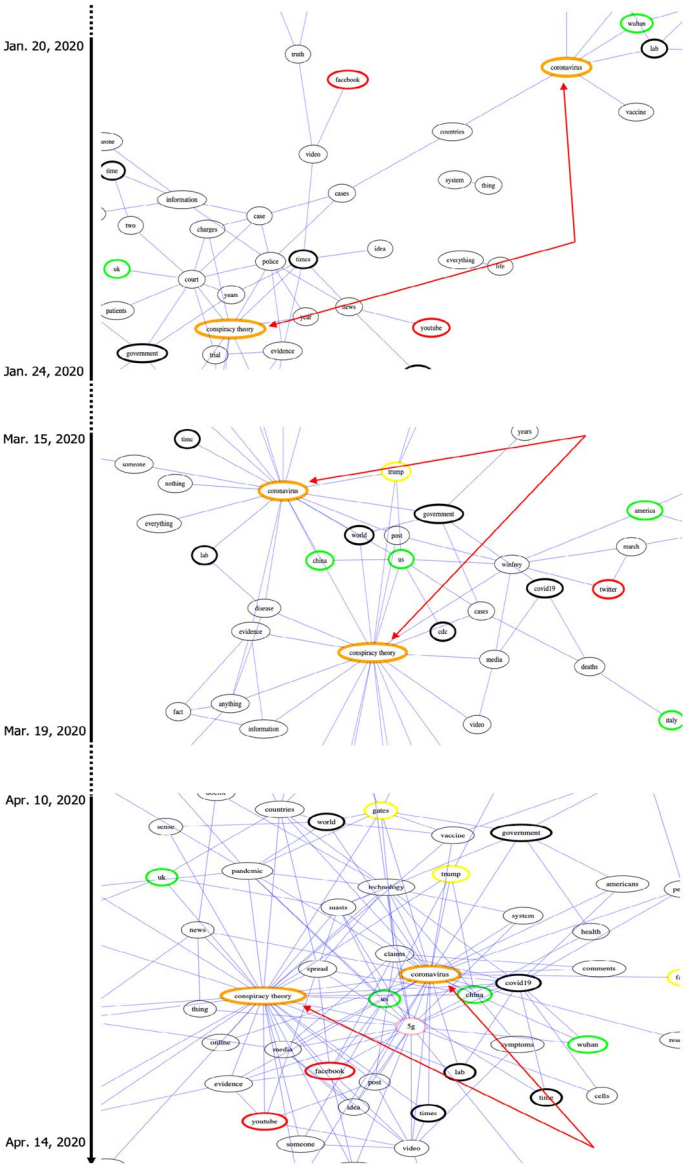


Fig. 3 Progressive attachment of "coronavirus" to "conspiracy theory" in the co-occurrence network of news reports conditioned on entities found in social media: The orange-outlined nodes represent the two concepts, as they gravitate toward one another over time and form new simple paths. From top to bottom, 5-day intervals starting on January 20, 2020, March 15, 2020, and April 10, 2020. Nodes are colored as follows: Celebrities in yellow, media outlets in red, important actants in pink (manually colored), places in green and corporations/entities in black

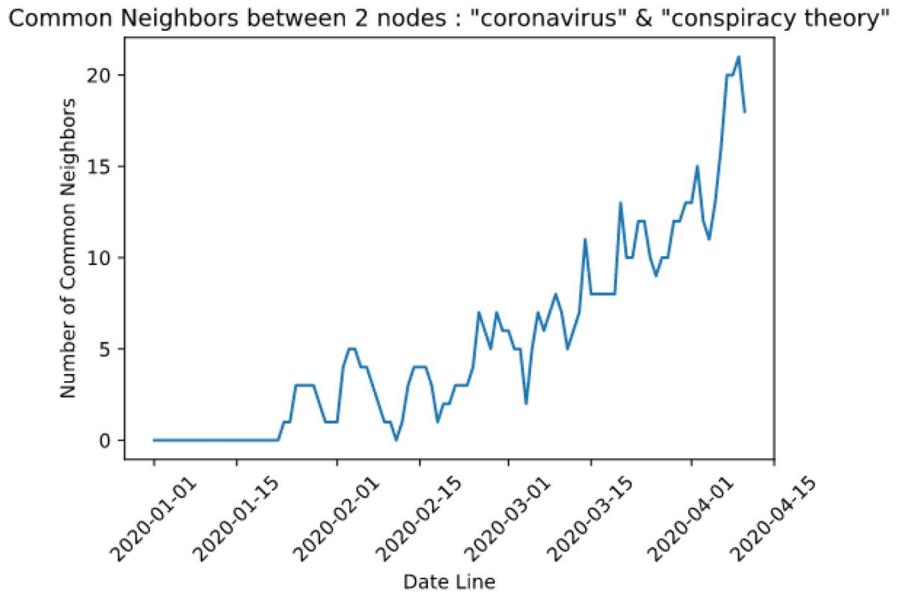
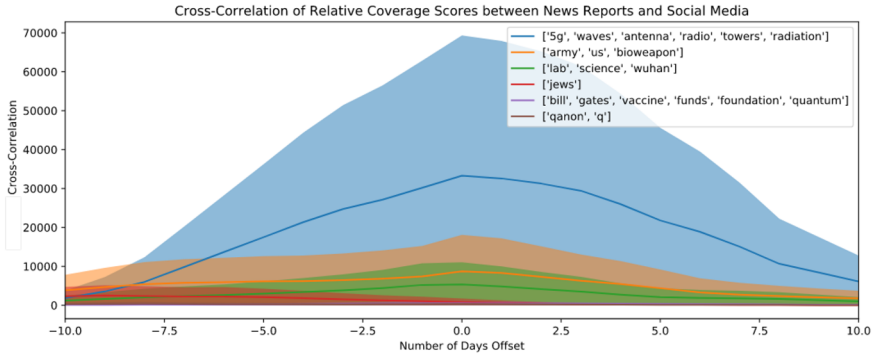


Fig. 4 Number of common neighbors between "coronavirus" and "conspiracy theory" over time in the news reports: Across all 101 segments of 5-day intervals, the number of simple paths empirically increases rapidly, suggesting the closer ties between the two entities across time

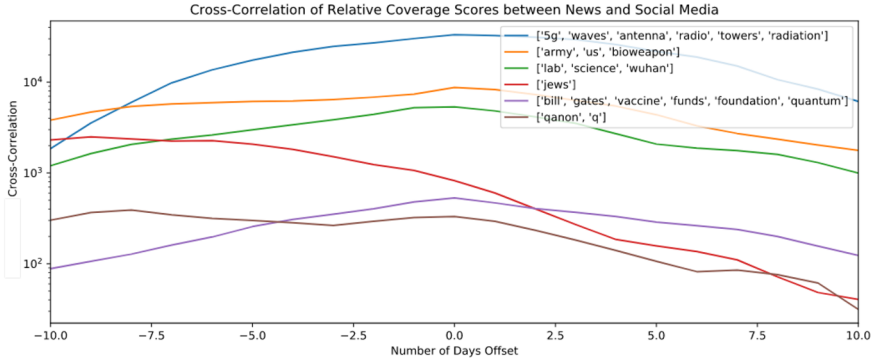
classified as threats. Table 4 provides a sample set of sub-nodes with their respective threat scores based on the majority vote. A sample sub-node "CCP" has 53% of its associated descriptive phrases classified as threats. The end-to-end classification pipeline, along with sample nearest neighbors during the phrase classification task, is shown in Fig. 9.

Discussion

The lack of authoritative information about the COVID-19 pandemic has allowed people to provide numerous, varied explanations for its provenance, its pathology, and both medical and social responses to it. These conversations do not occur in isolation. They not only circulate on and across various social media platforms but also interact with news reporting on the pandemic as it unfolds. Similarly, journalists are keenly aware of the discussions occurring in social media, thereby creating a feedback loop between the two. The interlocking computational methods described above facilitate the discovery of a series of important features of the (i) narrative frameworks that bolster conspiracy theories and their constituent rumors circulating



(a) Words in a community are matched to words present in the news reports and social media. Both the news reports and social media are smoothed for 5-day intervals. The mean and standard deviation are computed per time stamp and marked.



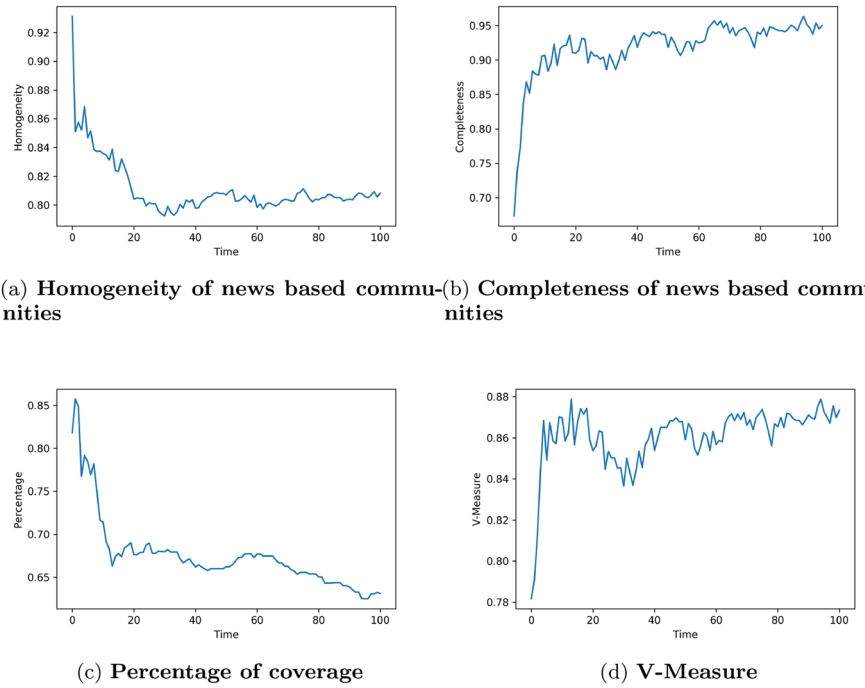
(b) Mean trajectories show the relative differentiation of each community. Note that the scaling on the y-axis is logarithmic.

Fig. 5 Cross-Correlation of Relative Coverage Score for Word-Level Community Hits in social media against the news reports: The mean and standard deviation of the relative coverage score are computed per time stamp across 20 trials with 500 community members each. The peak at 0 days offset suggests that social media and the news are intertwined in a very responsive manner

on and across social media, and (ii) the interaction between social media and the news.

Conspiracy theories in social media

The main communities and their interconnections in the aggregated social media corpus reveal the centrality of several significant conspiracy theory narrative frameworks. In particular, groupings of large communities form expansive frameworks and may well represent the dominant conspiracy theory frameworks in the corpus.



(a) **Homogeneity of news based communities**

(b) **Completeness of news based communities**

(c) **Percentage of coverage**

(d) **V-Measure**

Fig. 6 Scores of (**a**) Homogeneity, (**b**) Completeness, (**c**) Coverage and (**d**) V-Measure are provided to compare News based communities with Social Media communities. Here we used Y_{pred} and Y_{gr} derived in algorithm 3 as our cluster label and classes. Completeness measures how members of a given class are assigned to the same cluster, while homogeneity measures how each cluster contains only members of a single class. Their harmonic mean is the V-Measure [73]. Coverage percentage is the fraction of actants in news report communities that also are found in social media network communities

In other cases, coherent narrative frameworks can be discovered within a single community. These communities may have some connections or overlap with communities describing the contours of the pandemic, as well as to other small communities that provide support for aspects of the narrative framework.

We find four large community groupings which present easy-to-interpret conspiracy theory frameworks. The first of these groupings is comprised of nodes from communities 5, 56, and 82 (see Fig. 7). The narrative framework suggests that the Corona virus is closely linked to the 5G cellular network, and Bill Gates's associations with both faulty research and wide-scale vaccination programs. Eager to expand a global vaccination program to help prevent the explosion of the world's population, Gates has contributed to the design of the Corona virus, which can be characterized as a bio-weapon. Potentially activated by 5G signals (a technology that is also the result of faulty research), the virus is intended to eradicate various populations throughout the world.

Certain key sub-nodes play key roles in connecting these communities to create the conspiracy theory narrative. For example, the sub-node "facility, faulty, practice,

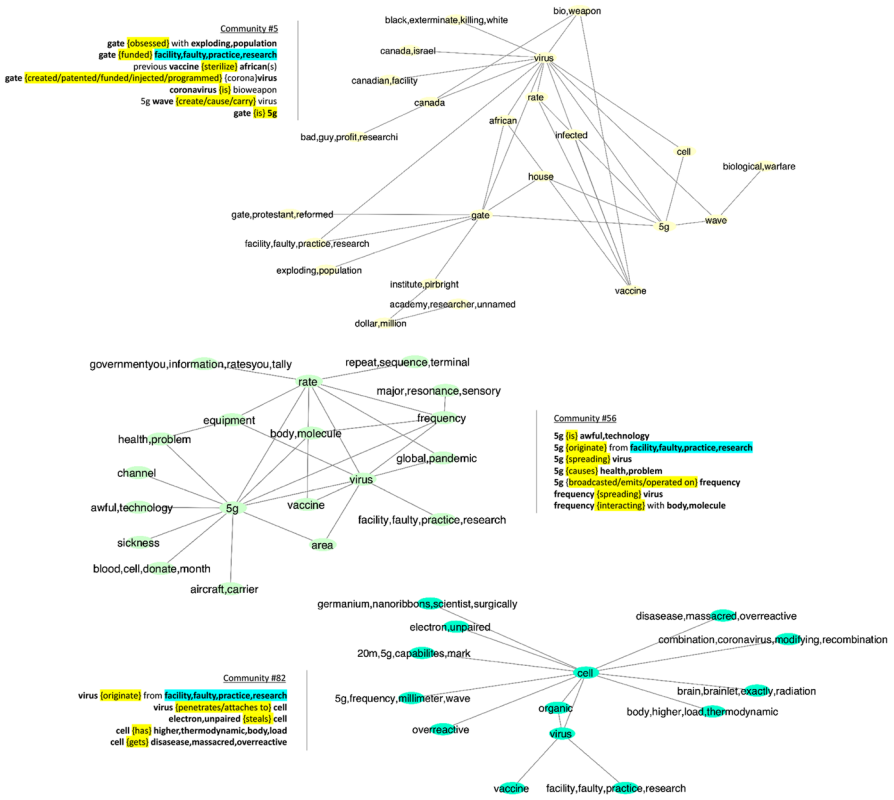


Fig. 7 Communities with index 5, 56 and 82 sequentially describe the conspiracy theory surrounding "Bill Gates" and "5g". The words in bold are the sub-nodes present in the narrative network and the yellow-highlighted phrases are automatically extracted relationships between the sub-nodes. The blue-highlighted sub-node is a key actant that exists in all 3 communities and is one of the connecting components between "Bill Gates" and the conspiracy theory around "5g". Community 5 describes Gates's supposed *obsession* with population control along with his funding of faulty research. The same research is alleged to have created "5g" as a means of spreading the "virus" which is allegedly intended as a "bioweapon". Community 56 takes it a step further tying "5g" to its carrier frequency and the associated interactions of this frequency with the human body. Community 82 concludes the origin story of the virus (back to the "faulty" research conducted by "Gates") and mentions the cell-level interaction between the virus and the body

Table 3 Cross-validation (fivefold) result of the phrase classifier

Hyperparameters	Recall	Accuracy
$k = 4$	$73.1\% \pm 3.8\%$	$84.9\% \pm 1.3\%$

research" interacts with "Bill Gates" and his supposed obsession with exploding populations and vaccination efforts, the "virus"' origin story, and the emerging "5g" technology, thereby offering one potential route traversed by conspiracy theorists. This traversal aligns three distinct communities as the conspiracy theorists create

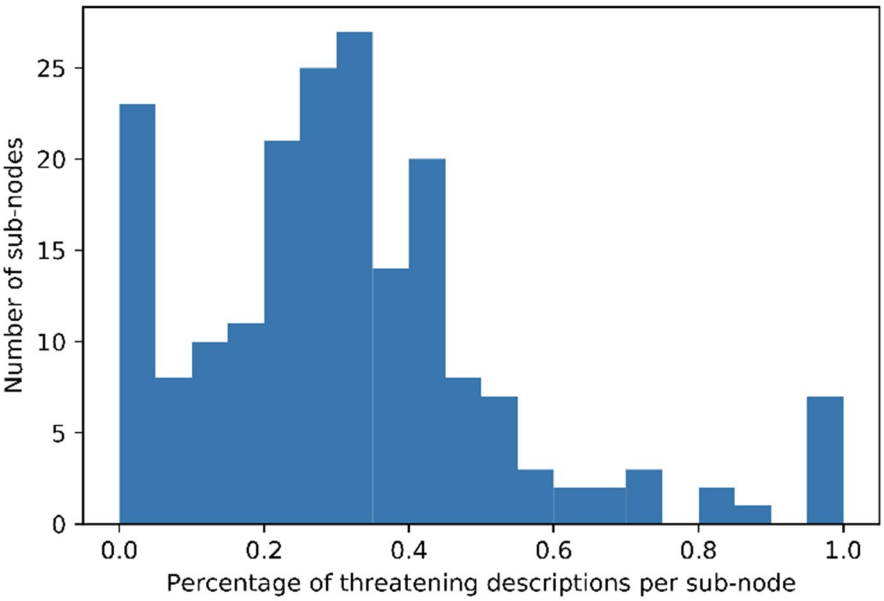


Fig. 8 The histogram of threat scores across the sub-nodes from the phrase classifier. The bi-modality encourages binary classification thresholds around 0.2. In our networks, we use 0.25 which is at the 57th percentile of sub-nodes classified as threats

Table 4 Sample threat scores: Note the increasing threat score from the sub-nodes "china" to "chinese" to "chinese, government", which reflects the threat carried by more specific "china" contextualized actants

Sub-node	Score
China	0.50
Chinese	0.53
China,government	0.57
CCP	0.53
5g	0.52
Cia	0.64
Shill	0.20
Result	0.17
Year	0.12

a unifying theory. None of these key nodes are innocuous, but rather have all been classified as threats (See Fig. 10).

A second group is comprised of nodes from communities 1, 40, and 65. In this narrative framework, the limited information about the virus released by the Chinese Communist Party is coupled to the virus's origin either in Chinese wet markets selling pangolins, presumably for human consumption, or labs studying bats (or potentially both). The narrative framework is informed by bigoted discussions of Chinese food practices coupled to an ongoing critique of the truthfulness of Chinese

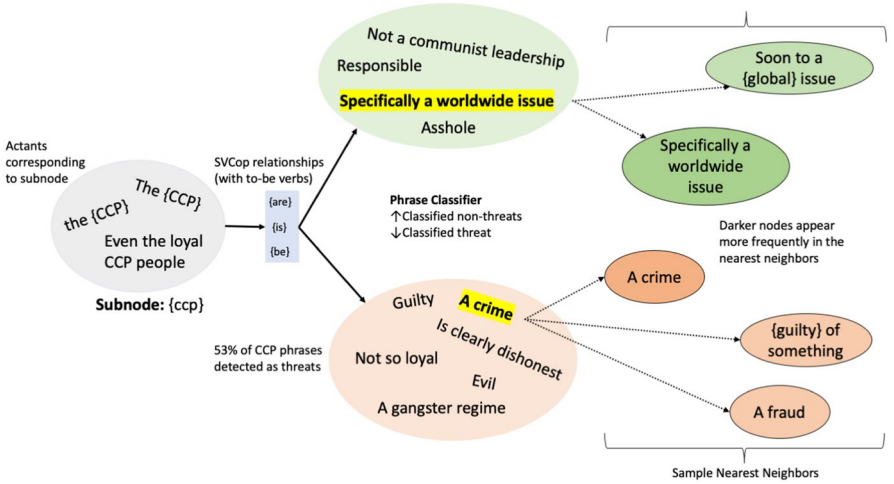


Fig. 9 The sub-node "*CCP*" has associated noun phrases shown in the gray box. The noun phrases have descriptive SVCop relationships, whose descriptive phrases are sampled in the **light red** and **green** blobs. The phrases in the **red blob** are classified as *threats* by our majority classifier and the phrases in the **green blob** are classified as *non-threats*. The highlighted and bold descriptive phrases are sample phrases for which the nearest neighbors are shown. *The kNN classifier reasonably clusters phrases that are syntactically different but semantically similar using the BERT embedding.* Darker nearest neighbors occur more frequently

researchers. Several intriguing elements of the narrative framework are the "fluoroquinolone" sub-node, an antibiotic which is also a favored medication in other narrative frameworks, and the inclusion of a Bill Gates sub-node. Both of these suggest clear points of potential attachment with other conspiracy theory frameworks, such as the 5G one described above, and another one focused on information cover-ups and the virus-as-hoax (See Fig. 11).

A third group, comprised of communities 0, 23, 24, 121 and 150, presents an expansive narrative framework. Here, the virus is presented as an engineered bioweapon, either deliberately or accidentally released from a lab. Confirmation of the engineered nature of the virus can be provided by scientists (pulmonologists) or members of the military (researcher, soldier). The subnodes in the graph set up a clear dichotomy between western governments and the Chinese government, and the controlling Chinese Communist Party (CCP), all of which are classified as threats. It is worth noting that the CCP abbreviation is used by some social media contributors as a reference to the "Chinese Communist Plague", a racially derogatory term for the virus analogous to Trump's reference to the virus as the "Kung flu" [47]. Aspects of the framework also support discussions of the economic impact of the pandemic, as well as the role of "globalists" in promoting the danger of the virus through inaccurate reporting and inflated counts of victims across the world, including Europe (See Fig. 12).

A fourth grouping comprised of communities 18, 21 and 75, constitutes a narrative framework proposing that the pandemic is a hoax on the same level as the

global warming "hoax". This framework includes actants such as Trump, the American news commentator Sean Hannity, the right-wing podcaster Nick Fuentes and Republicans writ large who are fighting against globalists, Democrats, scientists such as Anthony Fauci and, in keeping with the long history of anti-Semitism in conspiracy theorizing, the "Jews", all of whom have conspired to perpetrate this hoax, which is wreaking havoc on the economy. The British conspiracy theorist, David Ickes appears with a direct link to a node representing the "Jewish globalists". Interestingly, albeit perhaps not surprisingly, Bill Gates appears once again in this framework, now more closely related to the Mueller inquiry and Democrats such as Obama. While the goal of the hoax is not made explicit, the framework bolster the erroneous suggestion that the virus presents with mild symptoms, and is no more dangerous than the flu (See Fig. 13). The belief that the pandemic is a hoax inspires the "#filmyourhospital" movement as a means for publicizing the "discovery" that the virus poses no meaningful threat other than the economic threat of stay-at-home orders [31].

Several related narrative frameworks intersect with the main "hoax" framework in interesting ways. For example, a grouping of communities 6, 15 and 45, reveals a discussion of the disease, its relation to SARS and the flu, the testing regimen, the accuracy of the tests and the efficacy of masks. It also includes an apparent critique of media figures who often endorse conspiracy theories (See Fig. 14). While aspects of this narrative framework can be deployed as part of the more elaborated hoax framework—which seems to be the case particularly given the "threat" coding of "masks"—it can also be activated in the service of a counter-hoax narrative, given the inclusion of anti-conspiratorial content. In that regard, this particular grouping of communities captures the ongoing negotiation of the framework and the activation of parts of a framework as individuals come together (or move apart) to construct a totalizing narrative.

There are numerous other nucleations of narrative frameworks in the overall space that are worth noting. A particularly interesting community is 51 which has strong connections to the well-known "Pizzagate" conspiracy theory, as well as connections to the much broader QAnon conspiracy theory [67]. The intrusion of QAnon, and the alignment of the pandemic with the broader narrative of a ring of pedophile human traffickers gained strong support in certain conversations associated with the pandemic-as-hoax frameworks. It also aligns well with the belief, noted above, that tents erected in Central Park were part of an operation to save children trafficked through underground tunnels, a key feature of the "Pizzagate" conspiracy theory [67].

These smaller frameworks suggest that there is a lively, ongoing negotiation of community beliefs about the pandemic. As the conversations progress, many of these smaller narratives are likely to become more closely connected with larger groupings, while others are likely to fade away. Community 42, for instance, describes the pandemic as a deliberate attack on the nation perpetrated by the Democrats; despite the impact on the global economy, the virus is no worse than a bad flu. Such a community could be easily subsumed in the broader virus-as-hoax narrative framework. Two additional examples of much smaller nucleations would be community 171 which consists of three sub-nodes: "cell phone", "ear"

Table 5 A qualitative overview of key relationships that refer to "Bill Gates" in social media and the news reports

Date	News reports	4Chan threads
04/04	[Bill {Gates}] → [{predicted}] → [the {outbreak} and the China biolabs]	[{5g}] → [{causes}] → [{coronavirus}], [regular {people}] → [{go}] → [{untested}]
04/07	[conspiracy theorist David {Icke}] → [{added}] →[that Bill {Gates}, who is helping fund vaccine research, should be jailed]	[{Gates}] → [{saying}] → [[...]we all [...] accept his discount mark of the beast]
04/09	[Bill {Gates}] →[{invented}] →[{5G} to depopulate the world]	[the satanic {cabal}] → [to {leverage}] → [crisis into a forced vaccination /I D {program}]

These relationships describe the role that the "Bill Gates" node plays in connecting the Corona virus to conspiracy theories

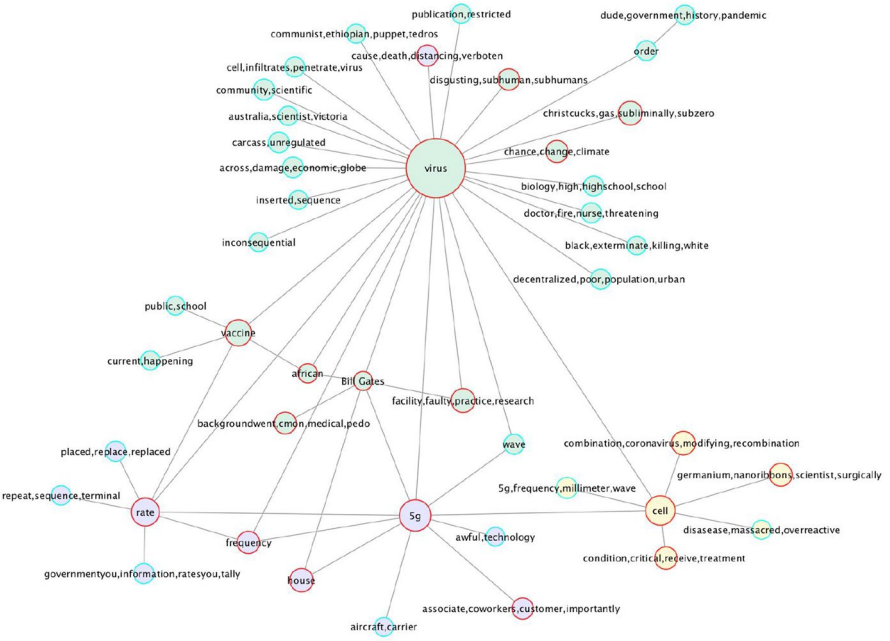


Fig. 10 A conspiracy theory narrative framework that links the virus to 5G, Bill Gates, and vaccination. Nodes have been scaled by NER mentions; those with fewer than 250 mentions have been filtered for the sake of clarity. Nodes are colored by community, and outlined with red if they represent a threat

and "state surveillance", and community 123 with six semantically meaningful sub-nodes: "cancer", "cell phone", "cell tower", "microwave", "human cell", and "substance". One would expect, as discussions continue, that these communities would move closer to the 5G conspiracy theory narrative framework.

This tendency toward monological thinking already appears to be at work in the alignment of the 5G conspiracy theory with the biological weapons conspiracy theory, with both of those frameworks sharing close connections with the narrative framework describing the pandemic as a whole. Other alignments seem possible, with the 5G conspiracy and the hoax conspiracy potentially aligning through community 32, which in general focuses on Italy and quarantine measures across Europe. The inclusion of two peripheral sub-nodes, one labeled "5G, chemtrailz" and another one that labels the quarantines "ridiculous", not only provide an opportunity to challenge the meaningfulness of quarantine measures (thus providing a potential alignment with the hoax narrative), but also provide a connection between 5G and the longstanding "chemtrailz" conspiracy theory [75].

In earlier work on conspiracy theories, we discovered that conspiracy theorists, as part of their theorizing, tend to collaboratively negotiate a single explanatory

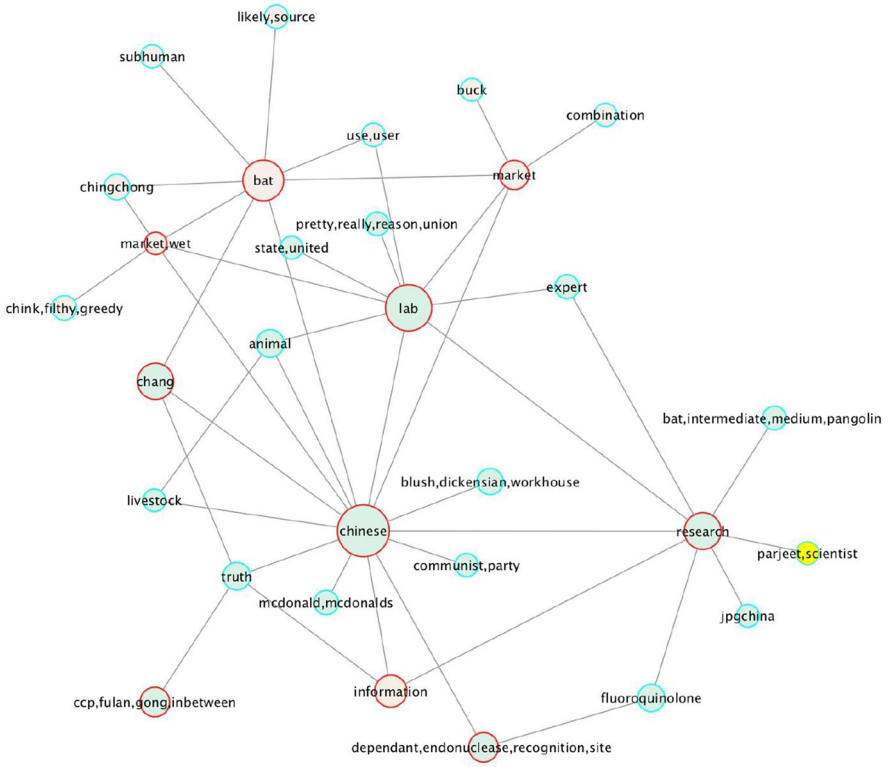


Fig. 11 Communities comprising the narrative framework suggesting that the virus is a result of Chinese wet markets and deliberate information cover-ups. The narrative framework focuses heavily on markets, exotic animals such as pangolins, and the role of Chinese Communist Party in hiding information about the initial outbreak. Nodes are colored by community, and outlined with red if they represent a threat. The graph is filtered to show nodes with degree *geq* 2

narrative framework, often composed of a pastiche of smaller narratives, aligning otherwise unaligned domains of human interaction as they develop a totalizing narrative [27, 55, 67]. In many conspiracy theories, this coalescence of disparate stories into a single explanatory conspiracy theory relies on the conspiracy theorists' self-reported access to hidden, secret, or otherwise inaccessible information. They then use this information to generate "authoritative" links between disparate domains, engaging in what Goertzel has labeled "monological thinking" [27].

For the current pandemic, however, a single unifying corpus of special or secret knowledge does not yet exist—there are no "smoking guns" to which the conspiracy theorists can point, such as the Wikileaks emails on which Pizzagate conspiracy theorists relied [67]. Consequently, the social media space is crowded by a series of conspiracy theories. In the various forums we considered, proponents of different narratives fight for attention, while also trying to align the disparate sets of actants and interactant relationships in a manner that allows for a single narrative framework

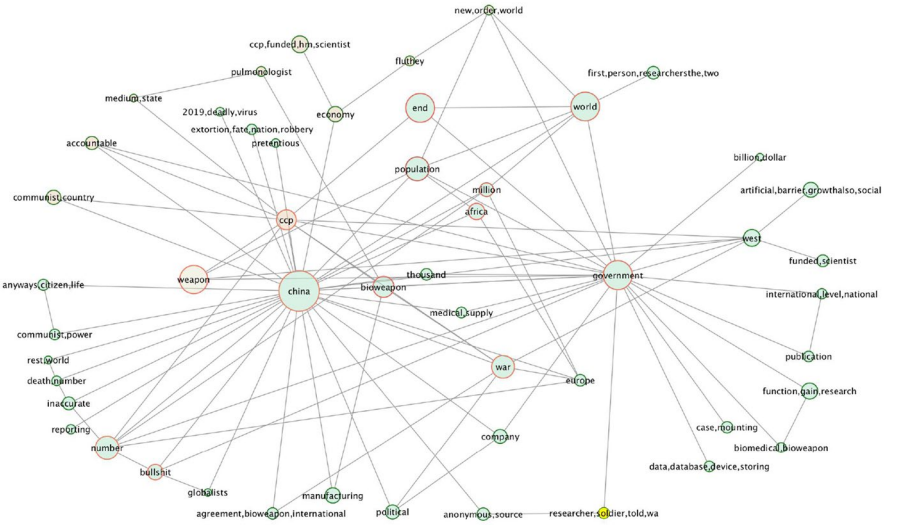


Fig. 12 Communities comprising the COVID-19-as-bioweapon narrative framework. The narrative framework focuses heavily on laboratories and the potential role of the virus as a weapon. Nodes are colored by community, and outlined with red if they represent a threat. The graph is filtered to show nodes with degree *geq* 2

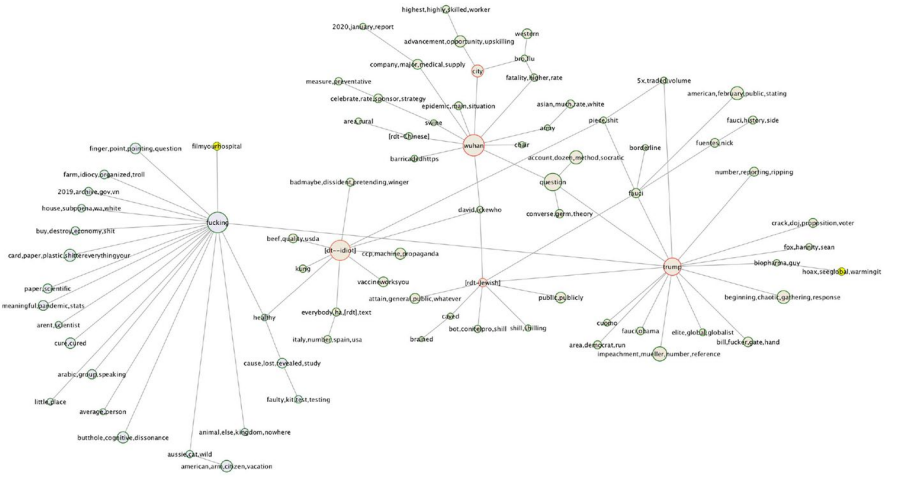


Fig. 13 The communities comprising the globalist hoax narrative framework: Here, a globalist cabal has conspired to foist the hoax of the Corona virus on the world, with the virus presenting with mild flu-like symptoms. Trump and his allies are fighting against the Democrats and their surrogates to stave off the economic impact of the hoax. Nodes are colored by community, and outlined with red if they represent a threat. The graph is filtered to show nodes with degree *geq* 2. Two nodes, "filmyourhospital" and "hoax,see global warming" have been highlighted in yellow

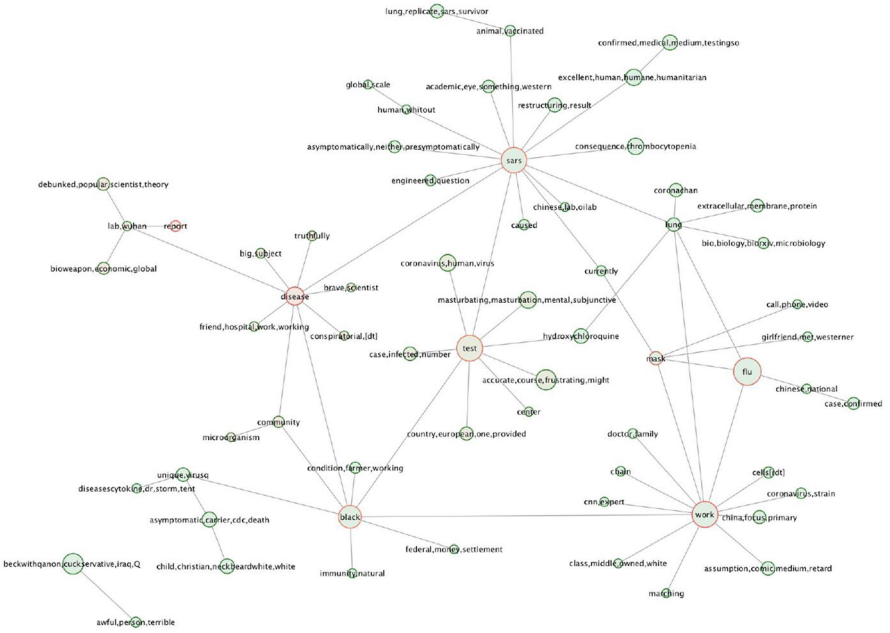


Fig. 14 A narrative framework that can be deployed by multiple groups. The framework focuses on the relationship between the virus, SARS, the flu and the testing regimen. Also included are nodes representing research on the virus and questions of immunity. A small disconnected component on the filtered graph provides a critique of QAnon and Glenn Beck. Nodes are colored by community, and outlined with red if they represent a threat. The graph is filtered to show nodes with degree *geq* 2

to dominate and, by extension, to provide the "winning" theorists with the bragging rights of having uncovered "what is really going on." This type of jockeying for position is also reflected in the news.

Social media and the news

Importantly, there is a lively interaction between the news media and the discussions about the pandemic on social media. Consequently, while the news media reports on the conspiracy theories that are evolving on social media, the social media groups point back toward reporting on the pandemic in the news media. The interaction is not, however, one of simple endorsement. Rather, the conversations on social media frequently contest, poke holes in, and otherwise challenge the narratives presented in the news media. In turn, the news media explores not only the content but also the veracity (or lack thereof) of the social media discussions. Unlike normal fact checking, however, the rejection in the news media of a particular social media position may be fuel for the conspiracy theorists, given their frequent suspicion of the news media.

This interaction between social media and the news, modeled by the cross-correlation of relative coverage scores in Fig. 5, indicates that the information flow

between the two corpora is swift: the correlation-maximizing offset of days was 0 or nearly 0 for all considered actant groups. Since the data is smoothed over five days, this finding implies that the major actants appearing in narrative frameworks get aligned within days of appearing in either channel.

A qualitative example expanding upon this dynamic of knowledge synchronization between the news and social media is observed in Table 5 where "Bill Gates" was earlier highlighted as an important actant. News reports on April 4th actively mentioned Gates's prediction of the COVID-19 outbreak. At the same time, 4Chan threads were embroiled in the discussion of "5g" causing the "Coronavirus". Perhaps the shock of such an accurate prediction—-and Bill Gates's continued investment in pandemic prevention and vaccine research—-helped motivate David Icke, an influential conspiracy theorist, to proclaim on April 7th that "Bill Gates belongs in jail", echoing comments of a Florida pastor, Adam Fannin, who believes Gates is involved in a global effort to depopulate the world. In the ensuing days after Ickes's comments, 4Chan threads began denigrating "Gates", alleging him to be a part of a satanic cabal (thereby creating a direct link to "Pizzagate"), labeling him the anti-Christ, and accusing him of being an opportunist forcing the world into a crisis to further his *alleged* forced vaccination campaign. News reports, seemingly in response, summarized the conspiracy theories circulating on 4Chan communities with headlines such as, "The Dangerous Coronavirus Conspiracy Theories Targeting 5G Technology, Bill Gates, and a World of Fear" [65].

Conclusion

As the global COVID-19 pandemic continues to challenge societies across the globe, and as access to accurate information both about the virus itself and what lies in store for our communities continues to be limited, the generation of rumors and conspiracy theories will continue unabated. Although news media have paid considerable attention to the well-known Q-Anon conspiracy theory (perhaps the most capacious of conspiracy theories of the Trump presidency), social media conversations have focused on four main conspiracy theories: (i) the virus as related to the 5G network, and Bill Gates's role in a global vaccination project aimed at limiting population growth; (ii) a cover-up perpetrated by the Chinese Communist Party after the virus leaped to human populations based largely on Chinese culinary practices; (iii) the release, either accidental or deliberate of the virus from, alternately, a Chinese laboratory or an unspecified military laboratory, and its role as a bio-weapon; and (iv) the perpetration of a hoax by a globalist cabal in which the virus is no more dangerous than a mild flu or the common cold. As the conversations evolve, these conspiracy theories appear to be connecting to one another, and may eventually form a single coherent conspiracy theory that encompasses all of these actants and their relationships. At the same time, smaller nucleations of emerging conspiracy theories can be seen in the overall social media narrative framework graph.

Because the news cycle appears to chase social media conversations, before feeding back into it, there is a pressing need for systems that can help monitor the emergence of conspiracy theories as well as rumors that might presage real-world action.

We have already seen people damage 5G infrastructure, assault people of Asian heritage, deliberately violate public health directives, and ingest home remedies, all in reaction to the various rumors and conspiracy theories active in social media and the news. We have shown that a pipeline of interlocking computational methods, based on sound narrative theory, can provide a clear overview of the underlying generative frameworks for these narratives. Recognizing the structure of these narratives as they emerge on social media can assist not only in fact checking but also in averting potentially catastrophic actions. Deployed properly, these methods may also be able to help counteract various dangerously fictitious narratives from gaining a foothold in social media and the news. At the very least, our methods can help to identify the emergence and connection of these complex, totalizing narratives that have, in the past, led to profoundly destructive actions.

References

1. Ahmed, W., Vidal-Alaball, J., Downing, J., & Seguí, F. L. (2020). COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *Journal of Medical Internet Research*, *22*(5), e19458.
2. Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., Vollgraf, R. (2019). FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations); 2019. pp. 54–59.
3. Allport, G.W., Postman, L. (1947). The psychology of rumor.
4. Bandari, R., Zhou, Z., Qian, H., Tangherlini, T. R., & Roychowdhury, V. P. (2017). A resistant strain: revealing the online grassroots rise of the antivaccination movement. *Computer*, *50*(11), 60–67.
5. Bearman, P. S., & Stovel, K. (2000). Becoming a Nazi: a model for narrative networks. *Poetics*, *27*(2–3), 69–90.
6. Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs conspiracy: collective narratives in the age of misinformation. *PloS One*, *10*(2), e0118093.
7. Blake, A. (2020). Six zombie claims about the coronavirus that just won't go away. Washington Post April 15, 2020.
8. Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment.*, *2008*(10), P10008.
9. Boberg, S., Thorsten, Q., Schatto-Eckrodt, T., Lena, F. (2020). Pandemic populism: facebook pages of alternative news media and the corona crisis–a computational content analysis. arXiv preprint arXiv:2004.02566.
10. Boole, G. (1854). *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities*. London: Walton and Maberly.
11. Börner, K., Oyvind, E., Tamara, M., Malte, R., Gerik, S. (2019) Network Visualization in the Humanities (Dagstuhl Seminar 18482). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.
12. Capatides, C. (2020) Coronavirus cannot be cured by drinking bleach or snorting cocaine, despite social media rumors. CBS News. March 9, 2020.
13. Capurro, G., Greenberg, J., Eve Dubé, S., & Driedger, Michelle. (2018). Measles, moral regulation and the social construction of risk: media narratives of "anti-vaxxers" and the 2015 Disneyland outbreak. *Canadian Journal of Sociology*, *43*(1), 25–48.
14. Clover, C.J. (1986). The long prose form.
15. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*, 2493–2537.

16. Davis, C.A., Onur Varol, E.F., Alessandro, F., Filippo, M. (2016). Botornot: a system to evaluate social bots. In Proceedings of the 25th international conference companion on world wide web, pp. 273–274.
17. De Michel, C. (1984). *The practice of everyday life*. Berkeley: U California Press.
18. Dégh, L., Andrew, V. (1978). The crack on the red goblet or truth and modern legend. In, Folklore in the Modern World, pp. 253–272.
19. Des Forges, A.L. (1999) Leave none to tell the story: genocide in Rwanda. 3169 189, 1999 Human Rights Watch New York.
20. Devlin, J., Chang, M.W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:181004805.
21. Editorial board. Stop the coronavirus stigma now. Nature. https://www.nature.com/articles/d41586-020-01009-0
22. Feleke, B. (2020). Kenya governor under fire after putting Hennessy bottles in coronavirus care packages. CNN News. April 17, 2020.
23. Ferrara, E. (2020). # COVID-19 on Twitter: bots, conspiracies, and social media activism. arXiv preprint arXiv:2004.09531.
24. Fine, G. A., & Bill, E. (2013). *The global grapevine: Why rumors of terrorism, immigration, and trade matter*. Oxford: Oxford University Press.
25. Fine, G. A., & Patricia, A. T. (2001). *Whispers on the color line: rumor and race in America*. Berkeley: University of California Press.
26. Fine, G.A. (2005). *Veronique Campion Vincent and Chip Heath (eds)*. Rumor mills: The social impact of rumor and legend. Transaction Aldine.
27. Goertzel, T. (1994). Belief in conspiracy theories. *Political Psychology*, pp. 731–742.
28. Greimas, Algirdas Julien. (1966). Éléments pour une théorie de l'interprétation du récit mythique. *Communications*, *8*(1), 28–59.
29. Hathaway, R. V. (2005). 'Life in the TV': The visual nature of 9/11 lore and its impact on vernacular response. *Journal of Folklore Research*, *42*(1), 33–56.
30. Heller, Jacob. (2015). Rumors and realities: making sense of HIV/AIDS conspiracy narratives and contemporary legends. *American Journal of Public Health*, *105*(1), e43–e50.
31. Izahi, E., Sarah, E. (2020). Americans want to see what's happening in hospitals now. But it's hard for journalists to get inside. The Washington Post, April 3, 2020.
32. Kitta, A. (2012). *Vaccinations and public concern in history: legend, rumor, and risk perception*. New York: Routledge.
33. Knight, P., ed. Conspiracy theories in American history: An encyclopedia. Vol. 1. Abc-clio, 2003.
34. Knight, P., ed. Conspiracy theories in American history: An encyclopedia. Vol. 1. Abc-clio, 2003.
35. Kuperman, A. J. (2004). Provoking genocide: a revised history of the Rwandan Patriotic Front. *Journal of Genocide Research*, *6*(1), 61–84.
36. Labov, W.W., Joshua, W. Narrative analysis: oral versions of personal experience. Essays on the Verbal and Visual Arts, Seattle 12 (1967): 44.
37. Langlois, J.L. (2005). 'Celebrating Arabs': Tracing Legend and Rumor Labyrinths in Post-9/11 Detroit. Journal of American Folklore, pp. 219–236.
38. Laudun, J. (2001). Talk about the Past in a Midwestern Town:"It Was There At That Time.". Midwestern Folklore. 27(2):41–54.
39. Lehnert, W.G. (1980). Narrative text summarization. In AAAI, pp. 337–339.
40. Leskin, P. (2020). One of the internet's oldest fact-checking organizations is overwhelmed by coronavirus misinformation - and it could have deadly consequences. Business Insider, March 31, 2020.
41. Lewandowski, S., & Cook, John. (2020). *The Conspiracy Theory Handbook*. Fairfax: Center for Climate Change Communication. George Mason University.
42. Li, J., Qing, X., Cuomo, R., Purushothaman, V., & Mackey, T. (2020). Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. *JMIR Public Health and Surveillance*, *6*(2), e18700.
43. Lindahl, C. (2012). Legends of hurricane Katrina: The right to be wrong, survivor-to-survivor storytelling, and healing. *The Journal of American Folklore*, *125*(496), 139–176.
44. McInnes, L., Healy, J., & Astels, Steve. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, *2*(11), 205.
45. Metaxas, P., Finn, S.T. (2017). The infamous# Pizzagate conspiracy theory: Insight from a TwitterTrails investigation. Wellesley College Faculty Research and Scholarship.

46. Mohr, J. W., Wagner-Pacifici, R., Ronald, L. B., & Petko, B. (2013). Graphing the grammar of motives in National Security Strategies: cultural interpretation, automated text analysis and the drama of global politics. *Poetics*, *41*(6), 670–700.
47. Nakamura, D. (2020) With 'kung flu,' Trump sparks backlash over racist language - and a rallying cry for supporters. Washington Post. June 24.
48. Oliver, J. E., & Wood, T. J. (2014). Conspiracy theories and the paranoid style (s) of mass opinion. *American Journal of Political Science*, *58*(4), 952–966.
49. Pentikainen, J. (1978). Oral repertoire and world view. An anthropological study of Marina Takalo's life history. *FF Communications Turku*, *93*(219), 1–366.
50. Propp, V. (1928). The morphology of the folktale.
51. Ren, X., El-Kishky, A., Chi, W., Fangbo, T., Clare, R.V., Jiawei, H. (2015). Clustype: Effective entity recognition and typing by relation phrase-based clustering. In, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 995–1004.
52. Rosenberg, A., Hirschberg, J. (2007). V-Measure: a conditional entropy-based external cluster evaluation measure. In, Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL), pp. 410–420.
53. Rosnow, R. (1976). *L, and Gary A Fine. Rumor and gossip: The social psychology of hearsay*. Amsterdam: Elsevier.
54. Rosnow, Ralph. (1980). Psychology of rumor reconsidered. *Psychological Bulletin*, *87*(3), 578–591.
55. Samory, M., Tanushree, M. (2018). Conspiracies online: User discussions in a conspiracy community following dramatic events. In Twelfth International AAAI Conference on Web and Social Media. 2018.
56. Samory, M., Tanushree M. (2018). 'The Government Spies Using Our Webcams' The Language of Conspiracy Theories in Online Discussions. Proceedings of the ACM on Human-Computer Interaction 2, no. CSCW (2018): 1–24.
57. Samuels, E., Meg, K. (2020). How false hope spread about hydroxychloroquine to treat COVID-19 - and the consequences that followed. Washington Post, April 13, 2020.
58. Schild, L., Chen, L., Jeremy, B., Gianluca, S., Yang, Z., Savvas, Z. (2020). 'Go eat a bat, Chang!': An Early Look on the Emergence of Sinophobic Behavior on Web Communities in the Face of COVID-19. arXiv preprint arXiv:2004.04046.
59. Shahsavari, S., Ehsan, E., Behnam, S., Misagh, F., Pavan, H., Roja, B., Timothy, R.T., Vwani, R. (2020). An automated pipeline for character and relationship extraction from readers literary book reviews on Goodreads.com. In, WebSci '20, July 07–10, 2020, Southampton, UK.
60. Shahsavari, S., Pavan, H., Timothy, R.T., Vwani, R. (2020)."Covid-19_conspiracy-Theories." OSF. April 26, 2020. osf.io/j5tg6.
61. Shibutani, T. (1966). Improvised news: a sociological study of rumor. Ardent Media, 1966.
62. Siikala, A.-L. (1990). Interpreting Oral Narrative. Folklore Fellows'. *Communications*, *245*
63. Skomorovsky, M., Mintz, M. (2019). Oligrapher; 2019. https://github.com/public-accountability/oligrapher.
64. Sommer, W. (2020). Naturally, we now have a cottage industry of coronavirus truther assholes. The Daily Beast, March 30, 2020.
65. Sorkin, A.D. (2020). The dangerous coronavirus conspiracy theories targeting 5G technology, Bill Gates, and a World of Fear. The New Yorker, April 24, 2020.
66. Starbird, K., Jim, M., Mania, O., Peg, A., Robert, M.M. (2014). Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. IConference 2014 Proceedings, 2014 iSchools
67. Tangherlini, T. R., Shadi, S., Behnam, S., Ehsan, E., & Vwani, R. (2020). An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate, Pizzagate and storytelling on the web. *Plos One*, *15*(6), e0233879.
68. Tangherlini, T. R. (2018). Toward a generative model of legend: Pizzas, bridges, vaccines, and witches. *Humanities*, *7*(1), 1.
69. Tangherlini, T. R., Vwani, R., Beth, G., Catherine, M. C., Roja, B., Akshay, W., et al. (2016). "Mommy Blogs" and the vaccination exemption narrative: results from a machine-learning approach for story aggregation on parenting social media sites. *JMIR Public Health and Surveillance*, *2*(2), e166.
70. Victor, J.S. (1993). Satanic panic: The creation of a contemporary legend. Open Court Publishing, 1993.

71. Sydow, V. (1948). *Carl W* (pp. 11–43). On the spread of tradition. CW von Sydow: Selected papers in folklore.
72. Waldrop, T., Dave, A., Eliott, CM. (2020). Fearing coronavirus, Arizona man dies after taking a form of chloroquine used to treat aquariums. CNN. March 25, 2020.
73. https://scikit-learn.org/stable/modules/clustering.html#homogeneity-completeness
74. Yang, K.-C., Torres-Lugo, C., Filippo, M. (2020). Prevalence of low-credibility information on twitter during the COVID-19 outbreak. arXiv preprint arXiv:2004.14484.
75. Oliver, J., & Thomas, J. W. (2014). Conspiracy theories and the paranoid style (s) of mass opinion. *American Journal of Political Science*, *58*(4), 952–966.

Affiliations

Shadi Shahsavari[1] · Pavan Holur[1] · Tianyi Wang[1] · Timothy R. Tangherlini[2] · Vwani Roychowdhury[1]

✉ Shadi Shahsavari
shadihpp@g.ucla.edu

Pavan Holur
pholur@g.ucla.edu

Tianyi Wang
tianyiw@g.ucla.edu

Timothy R. Tangherlini
tango@berkeley.edu

Vwani Roychowdhury
vwani@g.ucla.edu

[1] Electrical and Computer Engineering, UCLA, Los Angeles, CA, USA

[2] Department of Scandinavian, University of California, Berkeley, CA, USA