# A Golden Age: Conspiracy Theories' Relationship with Misinformation Outlets, News Media, and the Wider Internet

HANS W. A. HANLEY, Stanford University, USA
DEEPAK KUMAR, Stanford University, USA
ZAKIR DURUMERIC, Stanford University, USA

Do we live in a "Golden Age of Conspiracy Theories?" In the last few decades, conspiracy theories have proliferated on the Internet with some having dangerous real-world consequences. A large contingent of those who participated in the January 6th attack on the US Capitol believed fervently in the QAnon conspiracy theory. In this work, we study the relationships between five prominent conspiracy theories (QAnon, COVID, UFO/Aliens, 9/11, and Flat-Earth) and the role that misinformation and political polarization play in spreading them. We find that from 2008 to 2021 the percentage of all external hyperlinks from misinformation websites that went to *conspiracy-oriented* websites went from 10.2% to 22.4%, a 119% relative increase. Using partial Granger-causality, we uncover positive bidirectional relationships between the hyperlinks from misinformation websites and the popularity of conspiracy theory websites, suggesting the deep role that misinformation plays in popularizing many conspiracy theories. Examining the role that political polarization plays in explaining levels of conspiratorial material online, we find that the partisanship levels of news websites have a Pearson correlation of $\rho = 0.521$ with a website's shared domain connections with conspiracy theory websites, increasing to $\rho = 0.621$ when only considering conservative-leaning websites. Combining these results, we conclude by proposing a new metric to understand websites' fringe attitudes, which scores websites based on their penchant towards conspiracy theories and their partisanship level.

CCS Concepts: • **Human-centered computing** → *Collaborative and social computing*; **Empirical studies in collaborative and social computing**; • **Information systems** → **Web Mining**; • **Networks** → *Online social networks*;

Additional Key Words and Phrases: Misinformation, Disinformation, Conspiracy Theories, Social Networks, QAnon, 9/11, COVID-19, UFO, Aliens, Flat-Earth

## 1 INTRODUCTION

In the last few decades, websites have spread non-evidence-based conspiratorial ideas in an unprecedented fashion, resulting in a supposed "Golden Age of Conspiracy" [93, 111]. Today, up to 90% of the American public endorse at least one conspiracy theory [41, 79, 127], including those about 9/11 [113], UFOs/Aliens [33], Flat-Earth [83], COVID-19 [97], and QAnon [11]. According to a poll by the Public Religion Institute and the Interfaith Youth Core in July 2021, approximately 15% of Americans are true "QAnon believers", who believe that "the government, media, and financial worlds in the U.S. are controlled by a group of Satan-worshipping pedophiles who run a global child sex trafficking operation" [110]. Researchers have found that individuals mostly gain exposure to these conspiracy theories through the Internet [115, 124]. Although it is unknown whether this increased exposure to conspiracy theories online has casually led to a broader increase in conspiratorial thinking, prior research has shown that mere exposure to conspiracy theories can unconsciously influence people to believe in them [35]. Furthermore, the belief in just one conspiracy theory can induce paranoia and stimulate belief in other conspiracy theories [14, 26, 49, 58, 139].

Concurrent with the rise in conspiracy theories has been the rise of online misinformation outlets. Misinformation websites like the Gateway Pundit, American Thinker, and InfoWars have played an outsized role in developing toxic, political echo chambers [112]. Furthermore, many misinformation websites have increasingly utilized conspiracy theories for specific political purposes [82, 110, 132]. In spite of this increasing use, the precise relationship between online misinformation and

.

conspiracy theories is largely unknown. In this work, we use web crawling, hyperlink graphs, and partisan bias data to examine the role that conspiracy theories play in misinformation, news media, and the wider Internet's ecosystems.

To do this, we curate a set of 856 conspiracy theory websites, divided into six different categories: QAnon (227 sites), COVID-19 (135), UFO/Aliens (193), 9/11 (104), Flat-Earth (99), and Multiple/Forum (97). Using our own web scrapes and pages historically scraped by Common Crawl,[1] we then document the state of the conspiracy theory ecosystem and its relationship to misinformation, authentic news, and non-news websites. We find that between 2008 and 2021, the percentage of all misinformation websites external hyperlinks to *conspiracy-oriented* websites went from 10.2% to 22.4%, a 119% increase. We then apply partial Granger causality analysis to ascertain whether the behavior of misinformation websites is a *factor* in the popularity of conspiracy theories online. Partial Granger causality is a means of measuring if a given time series is useful for forecasting another while taking into account unmeasured endogenous and exogenous factors. Our results suggest that misinformation sites often enjoy positive feedback loops, where, as they hyperlink to conspiracy theory sites, this, in turn, drives the popularity of conspiracy theory websites. The popularity of the conspiracy theory websites, then, in turn, drives misinformation websites to post more hyperlinks to conspiracy theory websites.

We next seek to understand some of the driving forces behind *why* certain websites promote conspiracy theories. Specifically, we isolate the role of websites' US partisan bias (conservative vs. liberal) in promoting conspiracy theories. We show that as a website's audience becomes more partisan, the website's connectivity to conspiracy theory websites increases—for *both* authentic news and misinformation sites. We find that the partisanship levels of news websites have a Pearson correlation of $\rho = 0.521$ with a website's shared domain connections with conspiracy theory websites, increasing to $\rho = 0.621$ when only considering conservative-leaning websites. This further supports prior work [21, 22, 61, 123] that has shown that high levels of partisanship contribute to and help explain the spread of conspiracy theories.

Lastly, by combining our previous results, we show how conspiracy theories and political polarization can be utilized to understand a website's predilection to post fringe and/or misleading materials. A penchant for conspiratorial ideas or partisan bias alone is an insufficient measure of a website's fringe attitudes. While partisan bias of a website's audience gives some indication of a predilection to fringe beliefs, it is an imperfect measure. Similarly, many conspiratorial misinformation websites targeted at US and Western audiences are not easily categorized on the US political spectrum. For example, Katehon and Global Research, two infamous Russian propaganda and conspiratorial misinformation websites [28, 56, 112, 135], do not have clear biases on the US political spectrum [94]. Using websites' connections to *conspiracy-oriented* domains in combination with the website's partisan bias, we thus develop a method to determine the "fringeness" of websites (0/not-fringe to 1/fringe). We show that this metric can accurately differentiate misinformation from authentic news with a 97.0% accuracy, illustrating its usefulness.

Conspiracy theories have come to play an increasingly large role in world events, as was apparent in the partly QAnon-inspired January 6, 2021, attack on the U.S. capitol [82]. Similarly, as COVID-19 spread throughout the world, conspiracy theories became a major impediment to curbing the pandemic [97]. Our analysis shows the dominant role of misinformation outlets and political polarization in promoting and spreading conspiracy theories throughout the past decade. As conspiracy theories continue to play a larger role on the Internet, we hope that our results shed a light on their spread and show the utility of online structure-based analysis in understanding different their ecosystems.

[1] https://commoncrawl.org/

## 2 BACKGROUND AND RELATED WORK

In this section, we give an overview of how we operationalize our study of non-news, authentic news, misinformation, and conspiracy theories. We further give background on previous studies that have studied these different types of websites.

### 2.1 Terminology and Operalization

We first describe key definitions and concepts that we utilize throughout this work.

***Domain Based Identification:*** As described by Abdali *et al.*, examining misinformation, conspiracies, and news from the website domain level rather than by article level is a more "fruitful" enterprise [3]. It is unlikely that reputable organizations like The New York Times or the Wall Street Journal will start regularly publishing false information. In contrast, websites like The American Thinker or The Conservative Treehouse regularly and consistently publish false information. The consistent publication of misinformation, conspiracy theories, and truthful information is more readily identifiable and can be examined over time [55, 57, 59]. We, therefore, in this work, categorize and understand and reputation on a domain level.

***Misinformation Websites:*** False and misleading information or "*misinformation*" takes a host of different forms including unintentional misreporting, deliberate hoaxes and pranks, political propaganda, and disinformation. Following the 2016 US Presidential election, Jack *et al.* [62] presented an in-depth categorization of different types of misinformation. Since this report, the two criteria for understanding and classifying misinformation that have gained widespread acceptance are *veracity* and *intentionally* [65, 108]. Many works, using these criteria, define *misinformation* as any information that is false or inaccurate regardless of the intention of the author [10, 53, 55, 57, 60, 65, 68, 131]. *Disinformation*, meanwhile, is false and inaccurate information spread with the express and deliberate purpose to mislead. A key aspect of whether news is misinformation and/or disinformation is thus the ability of readers or aggregators to verify its claims [6, 9].

We thus, as in previous works [3, 10, 32, 55, 57, 59, 87, 107, 138], define *misinformation websites* as news websites that regularly publish false information about current events and that do not engage in journalistic norms such as attributing authors and correcting errors. As in these past works, we specifically *choose* to utilize a definition of *misinformation websites* that center around false information *and* disinformation about current events [10, 42, 59, 66]. We do this to ascertain the relationship between prominent websites that regularly publish false information and conspiracy theories.

***Conspiracy Theory Websites:*** A conspiracy theory is "an explanation [of current events] that refers to hidden malevolent forces that seek to advance some nefarious aim" [80]. In prior work, conspiracy theories have been often thought of as a subset of misinformation and disinformation. However, as outlined by others, most markedly by Brotherton *et al.*, conspiracy theories have specific and distinct characteristics that make them different from general misinformation [25, 27, 36]. Due to this particularity, measures taken at correcting and stemming misinformation remain mostly ineffective against conspiracy theories [20, 36, 77, 114]. We thus, as in other works, make the distinction between specific conspiracy theory websites and other more general misinformation websites and platforms [57, 67, 84, 89, 90, 99, 100, 106]. To make this conspiracy theory category concrete, we populate this category with websites that are focused on or dedicated to five different conspiracy theories (QAnon, COVID, 9/11, UFO/Aliens, Flat-Earth) that fall under psychologist Bortherton's *et al.*'s definition of conspiracy theories [25]. We thus consider a website as *misinformation* if it generally regularly publishes false information about current events (without it having a focus on any given conspiracy theory) while considering a website as *conspiracy theory* website if it focuses on our five of our listed conspiracy theories. Establishing this distinction enables us to identify the role that conspiracy theories have in the spread of political news and misinformation as well as to operationalize the study of individual conspiracy theories on the Internet.

***Authentic-News:*** As in previous works, we define authentic news websites as outlets that generally adhere to journalistic norms including attributing authors and correcting errors; altogether publishing mostly true information [57, 59, 138].

***Non-News:*** As in Hounsel *et al.* [59], we define *non-news websites* as those that do not normally traffic in current-event-related topics.

### 2.2 Tracking Misinformation

Our work which examines and measures the impact of specific social phenomena (prominent conspiracy theories and misinformation) builds on others' contributions to identify the long-term influence of toxic social contagions. Several works have attempted to identify misinformation utilizing information gleaned from microblogging messages [48, 63, 92], network infrastructure [55], natural language processing [37, 81, 105], and images [46, 129, 137]. For example, Zannettou *et al.* and Wang *et al.* studied misinformation through shared images on social media [129, 137]. Taking a different image-based approach, Abdali *et al.* identified misinformation websites from screenshots of their articles [3]. Qazvinian *et al.* helped pioneer approaches utilizing text-based linguistic features by identifying particular phrases present in misinformation articles [92]. As shown in Han *et al.*, misinformation websites often contain particular and specialized infrastructure to avoid bans and to monetize their efforts [55]. In contrast to these previous approaches, several other take a more case-study based approach to studying different misinformation campaigns. For example, Wilson and Starbird *et al.* look at the spread of disinformation campaigns related to the Syrian White Helmets [112].

Like our work, many works have exploited the semantic information that exists in mutual hyperlinking amongst semantically similar domains to study online behaviors and misinformation. As shown in Bhatt *et al.*, Hanley *et al.*, Garimella *et al.*, and Miller *et al.*, semantic information is often embedded within hyperlink graphs [23, 47, 57, 70]. Hanley *et al.* showed that QAnon-themed websites often had deeper connections to misinformation compared to authentic news. Garimella *et al.* showed a 0.70 correlation between a given website's polarization level and the average polarization levels of websites linked to it. Starbird *et al.* and Sehgal *et al.* have further shown that mutual hyperlinking persists across misinformation spreaders [103, 112]. In their study, Starbird *et al.* built a domain graph of relatively ignominious misinformation campaigns based on shared hyperlinks. Their approach rests on utilizing the Twitter social graph to draw connections between different domains and campaigns. In contrast, Sehgal *et al.* show that hyperlinks shared between websites and Twitter users can be utilized to identify coordinated efforts to spread disinformation [103]. Their work buttresses this by building a classifier based on mutual link sharing amongst misinformation spreaders on Twitter.

### 2.3 The Spread of Conspiracy Theories

Despite research literature that indicates that the predilection of people to believe in conspiracy theories has remained somewhat constant across time [124, 126], various authors have concluded that large spikes in conspiratorial belief often occur following major societal change [58, 125]. As found by Samory and Mitra [100], conspiracy theories often follow similar *narrative-motifs* that focus on the conspiratorial *agents*, the *actions* these *agents* perform, and the *targets* or *victims* of these actions. While many websites online are known to promote conspiracy theories [115, 124, 133], these beliefs are not new. In the United States, conspiracy theories from those about the death of John F. Kennedy to the Masonry to Watergate have possessed the American mind. According to Gallup, at a high point in the early 2000s, approximately 81% of the American public believed in a conspiracy theory about the JFK assassination [116]. However, Brotherton *et al.* [25], using factor analysis, found that the five key features of modern conspiracy theories are government malfeasance, extraterrestrial cover-up, malevolent global conspiracies, personal wellbeing, and control of information.

While increases or decreases in conspiratorial thinking across time is up for debate, conspiracy theories in the past few decades have increasingly been utilized online as political weapons [43, 82, 110]. Using mainly anecdotes, researchers have found that various conspiracy theories like QAnon have been manipulated for political means [19, 122]. QAnon itself is largely a political conspiracy theory. Prior research has further shown that the Internet is the main vehicle through which people get exposure and share conspiracy theories [115, 124]. As found by Bessi *et al.* [21] found that insular communities and polarized communities often form Facebook groups to discuss and engage with news related to conspiracy theories. Xiao *et al.* [133] after interviewing believers of the "chemtrails" conspiracy theory find individuals often become susceptible to conspiracy theories after reading about them online and finding welcoming communities on social media. To further understand this phenomenon, Samory and Mitra [99] used causal time-series analysis to understand changes in conversation and engagement following dramatic world events on the Reddit subcommunity `r/conspiracy`. They found that after these dramatic events, users on Reddit's `r/conspiracy` subcommunity show signs of emotional shock as well as increased expressions of both certainty and doubtfulness.

## 3 DATASET

Having given background, in this work, we study the relationships between the conspiracy theories, misinformation, authentic news, and other non-news websites as defined in Section 2. Here, we describe the sets of website lists we collect. Each of these website lists is independent and non-overlapping.

### 3.1 Misinformation News Dataset

Much investigation has gone into identifying misinformation websites, and we utilize several previously curated lists utilized in prior research [55, 57, 59]. Specifically, we select websites from lists created by the Columbia Journalism Review,[2] OpenSources,[3] Politifact,[4] Snopes,[5] and Melissa Zimdars.[6] We filter these lists to websites labeled as "fake news", "fake", "unreliable", or that have a factual rating of "low" and "very-low." We group these categories together given our definition of misinformation as news websites that regularly publish false information (Section 2). Combining these filtered lists together, we manually verify that each the *misinformation* websites in the combined list do not focus or are dedicated to our set of conspiracy theories. Finally, after removing duplicates and inactive websites, we arrive at 530 misinformation websites. This list of misinformation websites includes websites that have been widely documented as promoting falsehoods including The American Thinker and The Conservative Treehouse [112].

### 3.2 Conspiracy Theory Websites

As previously stated, a conspiracy theory is "an explanation [of current events] that refers to hidden malevolent forces that seek to advance some nefarious aim" [80]. Our study focuses on five conspiracy theories whose websites are readily identifiable:

[2] https://cjr.org/fake-beta
[3] https://github.com/several27/FakeNewsCorpus
[4] https://www.politifact.com/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/
[5] https://github.com/Aloisius/fake-news
[6] https://library.athenstech.edu/fake

| Conspiracy Theory | Seed Websites | # Websites |
|---|---|---|
| QAnon | 8kun.top, voat.co | 227 |
| UFO/Aliens | ufoabduction.com | 193 |
| COVID | covid-is-fake.blogspot.com | 135 |
| 9/11 | 911truth.org | 104 |
| Flat-Earth | theflatearthsociety.org | 99 |
| Multiple/Forum | – | 98 |

Table 1. **Collecting conspiracy theories**— Using Google searches, we collect a set of seed websites before performing deep crawling and graph analysis to later manually identify conspiracy theory websites in each category, 856 in total.

(1) **QAnon**: This conspiracy holds that the U.S. government is run by a cabal of Satanic pedophiles [11]. This conspiracy theory falls under Brotherton *et. al's* conspiracy features of government malfeasance, global conspiracies, and control of information [25].
(2) **COVID**: A set of theories that promote fake cures for COVID-19, argue that COVID-19 is a fake illness, or spread the idea that COVID-19 vaccines are deadly poisons [97]. This conspiracy theory falls under Brotherton *et. al's* conspiracy features of government malfeasance, malevolent global conspiracies, personal wellbeing, and control of information.
(3) **9/11**: A conspiracy theory that holds that the 9/11 terrorist attack was a false flag event planned and orchestrated by the U.S. government [113]. This conspiracy theory falls under Brotherton *et. al's* conspiracy features of government malfeasance, malevolent global conspiracies, and control of information.
(4) **UFO/Aliens**: A conspiracy theory that holds that the United States government is secretly hiding that intelligent alien life is real and that aliens have repeatedly visited Earth [33]. This conspiracy theory falls under Brotherton *et. al's* conspiracy features of extraterrestrial cover-up, malevolent global conspiracies, and control of information.
(5) **Flat Earth**: A conspiracy that promotes that the Earth is flat, not spherical [83]. Images taken of the earth from space as well as other evidence of the Earth being round are part of an elaborate hoax orchestrated by the world's governments and scientists [86]. This conspiracy theory falls under Brotherton *et. al's*conspiracy features of malevolent global conspiracies and control of information.

To the best of our knowledge, while there exist lists that do label certain websites as promoting conspiracy theories[7], there are no public datasets that provide granular information about which specific conspiracy theory (i.e. QAnon, 9/11) a given website promotes. We thus, to create lists of this specificity, rely on an approach outlined by Hanley *et al.* [57], which leverages seed sets of websites and web crawling to find semantically-related websites. Originally only done for QAnon in Hanley *et al.* [57], we leverage this methodology five separate times, once for each of our conspiracy theories.

To build seed sets of websites per Hanley et al.'s methodology, we rely on Google searches, identifying 1–2 websites for each conspiracy theory (Table 1). After crawling each seed site, we identified additional candidate conspiracy theory websites from their graph-based overlap node similarity [128] (calculated using their set of shared domains connections with the seed websites). This metric helped to create a list of websites with semantically similar content to the original seed set. After identifying candidates, we manually reviewed and confirmed each candidate website was

[7] https://github.com/several27/FakeNewsCorpus

| Website Category | Total Websites |
|---|---|
| Misinformation News | 530 |
| Authentic News | 565 |
| Non-News | 528 |

Table 2. **Curating website lists**— We utilize several previously curated lists including those of the Columbia Journalism Review, OpenSources, Politifact, Snopes, and Melissa Zidmars to create lists of misinformation news and authentic news websites. We utilize semanticweb.org to create a new list of non-news-focused websites.

truly related to the given conspiracy theory in each run of this algorithm. While costly in terms of scraping time, this approach enabled us to build five separate lists of websites dedicated to each of our five conspiracy theories. We crawled, scraped, and built our lists of conspiracy websites throughout June 2021.

In total, we collected 856 conspiracy theory websites: 227 QAnon, 143 COVID, 104 Nine-Eleven, 193 UFO/Aliens, 99 Flat-Earth, and 98 Multiple/Forum websites. These 98 Multiple/Forum websites were found while categorizing QAnon websites and while they most prominently promote QAnon, they also feature multiple other conspiracy theories. While we consider these websites to prominently promote conspiracy theories, particularly QAnon, we thus do not put them in any specific conspiracy theory category. For example, while 8kun.top was useful for finding QAnon-related websites, it also has forums dedicated to COVID, Ebola, and UFOs and thus could not be cleanly categorized.

### 3.3 Authentic News Dataset

For our list of authentic news websites, we again utilize lists curated by OpenSources, Politifact, Snopes, and Melissa Zimdars, combining the sets of websites labeled as “reliable” or simply “biased”. To supplement our list of authentic news websites, we also include a selection of local news websites from lists of U.S. newspapers [8] [9] such as the Orlando Sentinel, the Denver Post, and the Tampa Bay Times. Finally, we expand our list by including the websites of 61 Pulitzer Prize-winning newspapers. Our final list of 565 websites includes sites across the political spectrum, from dailywire.com to cnn.com.

### 3.4 Non-News Dataset

We utilize 528 popular non-news websites as a control group to isolate and understand how conspiracy theories interact with the wider Internet. We include websites from the top 50 lists of 11 different topic category lists as labeled and gathered by semanticweb.com: business and services, community and society, food and drink, health, heavy industry, lifestyle, pets and animals, science and education, sports, travel and tourism, and vehicles.

## 4 METHODOLOGY

Having described the root set of websites that we analyze, we now present our methodology for understanding the relationships between these sites.

[8] https://www.w3newspapers.com/
[9] https://en.wikipedia.org/wiki/List_of_newspapers_in_the_United_States

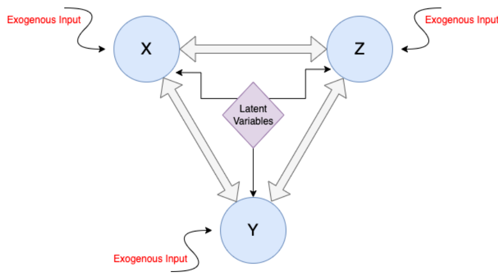



Fig. 1. **Modelling partial Granger causality relationships** — Partial Granger causality models the interaction of multiple time-series while taking into account exogenous environmental inputs and latent endogenous variables.

### 4.1 Common Crawl Page Retrieval and Website Crawling

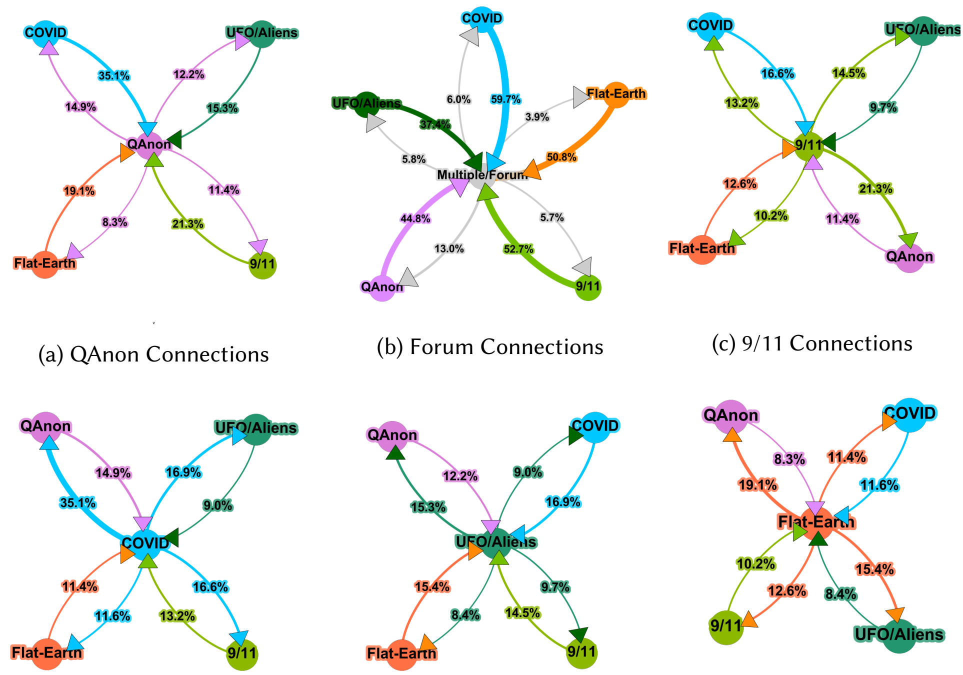
To gather the set of hyperlinks between our websites, we utilize Common Crawl data [109]—widely considered the most complete publicly available source of web crawl data—and our own website crawls. For each website in our dataset, we collect their domain's HTML pages that were indexed by Common Crawl before August 2021. In addition to Common Crawl data, we further utilize our own website scrapes. We utilize our own crawls, in addition to Common Crawl, due to noisiness, missing pages, and missing domains within the Common Crawl dataset [102]. For example, 309 of our conspiracy theory domains were not contained within the Common Crawl dataset. Thus for each website in our dataset, we further gather all the HTML pages 5 hops from each website's homepage (i.e., we collect all URLs linked from the homepage (1st hop), then all URLs linked from the pages that were linked by the homepage (2nd hop), and so forth). For each HTML page from our scrapes and Common Crawl, we parse the HTML and collect hyperlinks to other pages (i.e., HTML <a> tags). Altogether we gather the available Common Crawl pages and scrape the HTML for 855 conspiracy theory, 530 misinformation, 565 authentic news, and 528 non-news websites.

### 4.2 Partial Granger Causality

To ascertain the relationship between authentic news, misinformation, and conspiracy theories, we perform a temporal causal analysis of how these three categories of websites interact. Specifically, after fitting our time-series data to linear vector autoregressive models (VAR), we utilize notions of partial Granger causality [29, 51, 54] to determine if the behavior of websites in one of these groups is predictive of the behavior in another group.

Granger causality is a means of measuring if one time series is useful for forecasting another. In practice, Granger causality is commonly tested utilizing linear vector autoregressive models (VAR) of two different time series $X_t$ and $Y_t$. VAR models forecast future values of a given time series based on past values along with other variables [29]. Granger causality then tests, using VAR models, if knowledge of time-series $Y_t$ increases the predictive power regarding $X_t$ then $X$ Granger causes $Y$ [51, 54]. For a more detailed description of Granger causality see Appendix A.

Partial Granger causality is an extension of Granger causality for more than two time series that also takes into account latent endogenous variables and exogenous environmental variables [54, 136]. More concretely, partial Granger causality models two time-series $X_t$ and $Y_t$ conditioned on a third time-series $Z_t$ while taking into account exogenous environmental inputs $\overrightarrow{\epsilon_t^E}$ as well as endogenous latent variables $\overrightarrow{\epsilon_t^L}$ on the three time-series. We picture our utilized model in Figure 1. partial Granger causality then test against the null hypothesis that the information present in $Y_t$ does not add information when predicting future values of $X_t$. Using this construct, for example, we can thus model the Granger-causal behavior of misinformation websites on the popularity of conspiracy theory websites while taking into account the behavior of authentic news websites as well as




(a) QAnon Connections (b) Forum Connections (c) 9/11 Connections

(d) COVID Connections (e) UFO Connections (f) Flat-Earth Connections

Fig. 2. **Percentage of website domains hyperlinked by each conspiracy theory category that are also hyperlinked by another category**— Each conspiracy theory shares domain connections with each other. As seen QAnon has the most shared domain connections with every other conspiracy theory. 35.1%, 21.3%, 19.1%, and 15.3% of all domains hyperlinked by 9/11, Flat-Earth, and UFO/Aliens websites, respectively, are also hyperlinked by QAnon websites.

unmeasured and unaccounted-for endogenous and exogenous variables. For a detailed overview of partial Granger causality see Appendix B.

### 4.3 Ethical Considerations

Within this work, we largely look at large-scale trends amongst our set of websites. We further utilize only public data and follow ethical guidelines for scraping websites as outlined by others [4, 57].

## 5 THE CONSPIRACY-THEORY/NEWS ECOSYSTEM

Having given background and an overview of our methodology, in this section, we now present the relationships between online conspiracy theories and the news media ecosystem at large.

### 5.1 The Conspiracy Theory Ecosystem

To determine the relationships between different conspiracy theories, in Figure 2, we plot the number of outward domain connections that each conspiracy theory ecosystem shares with one another. For example, we determine the percentage of domains hyperlinked by QAnon websites that are also hyperlinked by COVID conspiracy theory websites. As seen, every conspiracy theory

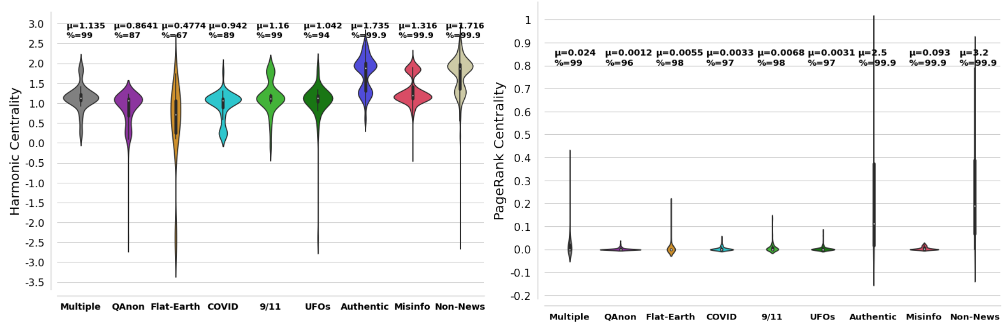



Fig. 3. **Average standardized harmonic and PageRank centrality measures with percentiles**— Utilizing the full web crawl snapshot of Common Crawl from February 2021 to May 2021 [75], we determine each conspiracy theory website group's harmonic and PageRank centralities. For the mainstream, misinformation, and non-news PageRank centralities, we truncate the list to alllow the distribution to fit on the same graph. As seen, each conspiracy theory group is on average more "central" to the web than the average website online by both methods of determining centrality.

has some degree of shared connections, illustrating the depth to which conspiracies interact with one another *and* the same sources.

As in Figure 2b, every individual conspiracy theory's connections are also heavily hyperlinked by the Multiple/Forum category of conspiracy theories. 59.7% of the domains hyperlinked by COVID conspiracy theory are also hyperlinked by the Multiple/Forum conspiracy theory websites. As previously noted, this is due to multiple conspiracy theories being promoted and utilized by these websites. We further see that 13.0% of links from the Multiple/Forum category of websites are also hyperlinked by QAnon websites. QAnon having the strongest connection to these websites was largely expected given that these sets of websites were found while gathering QAnon domains (section 3.2).

Looking at individual conspiracy theories, we observe extensive interconnections between the QAnon and COVID conspiracy theories. 35.1% (Figures 4b and 2d) of all domains hyperlinked by COVID conspiracy theory websites are also hyperlinked by QAnon websites. As reported by The Washington Post and others, following the defeat of Donald Trump in the 2020 election, many QAnon groups began to focus their ire on COVID vaccines and the government lockdowns[72, 121]. This is largely matched by the elevated levels of interconnectivity between these two groups. We further see an elevated connection with QAnon across the rest of the conspiracy theories considered in this work. 21.3%, 19.1%, and 15.3% of all domains hyperlinked by 9/11, Flat-Earth, and UFO/Aliens websites, respectively, are also hyperlinked by QAnon websites; each of them have the most shared connections with QAnon websites as opposed to the other conspiracy theories. This also largely accords with QAnon being a highly active conspiracy theory and news reporting that QAnon has become a "big tent conspiracy theory" that often incorporates the claims of other conspiracy theories [98].

Utilizing Common Crawl network data [75] over the indexed Internet (87.7 million websites), we next determine the *network centrality* of our set of conspiracy-focused websites to understand if each conspiracy theory website category is "core" (regularly utilized on the Internet) or "peripheral". We utilize centralities across Common Crawl's dataset rather than our partial one in order to get a sense of each conspiracy theory's centrality on the *entire* Internet. Various other works have utilized these

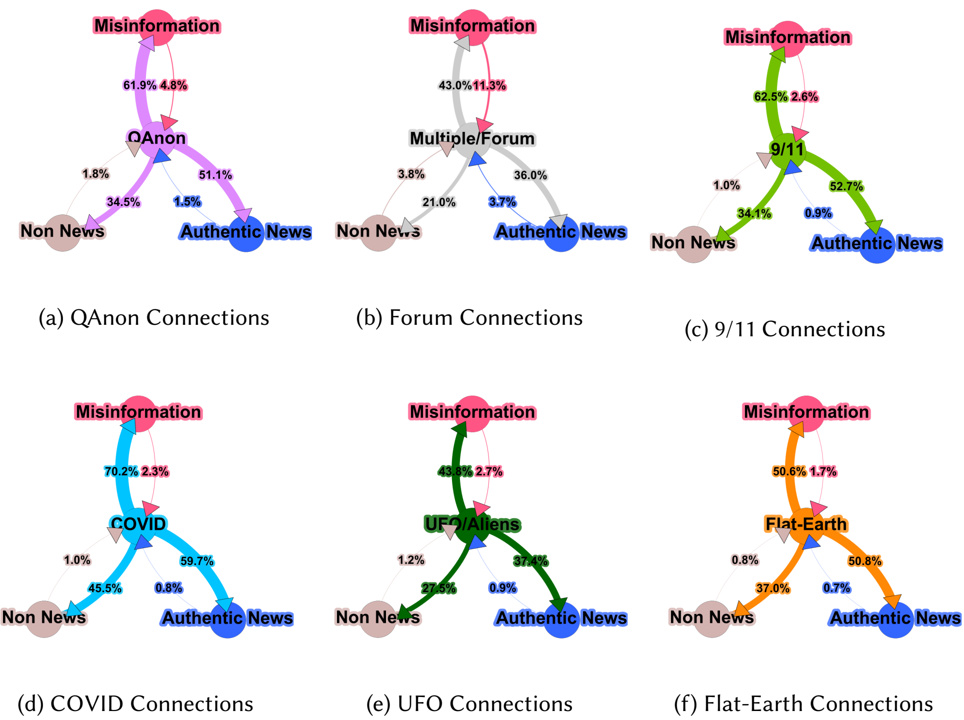
| Category | Domain | Harmonic w/ Perc. | PageRank w/ Perc. | Alexa Rank |
|---|---|---|---|---|
| Multiple/Forum | rense.com | 1.857 (99.98%) | 0.2580 (99.95%) | 59,450 |
| Multiple/Forum | voat.co | 1.272 (99.59%) | 0.2310 (99.95%) | —- |
| QAnon | qactus.fr | 1.230 (99.49%) | 0.0106 (98.49%) | 265,967 |
| QAnon | qmap.pub | 1.210 (99.40%) | 0.0174 (99.32%) | — |
| Flat-Earth | theflatearthpodcast.com | 1.743 (99.86%) | 0.0139 (99.16%) | — |
| Flat-Earth | nasalies.org | 1.712 (99.79%) | 0.0101 (99.82%) | —- |
| COVID | vaccineimpact.com | 1.818 (99.96%) | 0.0508 (99.77%) | — |
| COVID | plandemicseries.com | 1.768 (99.90%) | 0.0163 (99.28%) | 244,082 |
| UFO/Aliens | mufon.com | 1.858 (99.98%) | 0.0798 (99.85%) | — |
| UFO/Aliens | ufosightingsdaily.com | 1.830 (99.97%) | 0.0603 (99.80%) | — |
| 9/11 | ae911truth.org | 1.872 (99.98%) | 0.1330 (99.91%) | — |
| 9/11 | patriotsquestion911.com | 1.820 (99.96%) | 0.0267 (99.56%) | — |

Table 3. **Most central domains per conspiracy-focused category** —The top two most central two websites across the Common Crawled Internet from February 2021 to May 2021 [75], in each conspiracy theory category. The harmonic and PageRank centraliites of each website are reported as z-scores with their respective percentile. The Amazon Alexa rank is given as of March 1, 2021 [7].

network centrality measures to understand how important and utilized different web domains are to given ecosystems [57, 73, 88]. Further, as just seen and seen in Section 3.2, the Internet's hyperlink structure captures large amounts of semantic information and can reveal the role that different websites play. Based on this data, network centralities help determine how core/mainstream a website is on the Internet (*i.e.* websites with high centralities are core/mainstream while those with low centralities are peripheral) [38, 73]. While a widely used and often relied upon centrality measure, PageRank has been found to be susceptible to manipulation and spam [73]. As a result, to capture network centrality, we report a common revision of PageRank, namely the harmonic centrality, a measure that is also utilized by Common Crawl [75, 95] and in other works [73]. We present our analysis utilizing PageRank and harmonic centralities for completeness. As in Morina et al. [73], for interpretability, we present the percentile of these measures and present the raw values as standardized z-scores (*i.e.* a value of 0 represents the average centrality and a value of 1 represents one standard deviation above the mean, *etc.*).

As seen in Figure 3, (from the set of conspiracy theory websites within Common Crawl data) each conspiracy theory website group had an above-average centrality on the Internet between February and May of 2021. Namely, each website group was more central than the average and median website on the Internet. Furhtemore, each group of conspiracy theory websites has domains that are *more* central than many authentic news, misinformation, and nonnews websites. and We note that this is despite many of our conspiracy theory websites not appearing in the Amazon Alexa top million continuously during this period (Table 3) [7]. While as a whole this could be due to our biased selection of websites (Section 3.2), this result *does* illustrate the degree to which several conspiracy theories websites and their associated theories have been utilized and become central to the Internet. Specifically, as seen in Table 3, each conspiracy theory had multiple websites within the top 99% percentile of centralities on the Internet.

Looking at the relative centrality of each conspiracy theory group in Figure 3, we see that the Multiple/Forum group has the highest average harmonic and PageRank centralities. This largely agrees with these websites' central role as discussion places for multiple conspiracy theories. For example, looking at Table 3, we see that voat.co is one of the Multiple/Forum websites with the largest centralities. Now shut down, the former official home for "QAnon research", voat.co was




(a) QAnon Connections (b) Forum Connections (c) 9/11 Connections

(d) COVID Connections (e) UFO Connections (f) Flat-Earth Connections

Fig. 4. **Percentage of website domains hyperlinked by each website category that are also hyperlinked by another category**– Our set of misinformation websites have the highest amount of shared domain connections with conspiracy theory websites. For instance, 4.8% of all domains hyperlinked by misinformation websites are also hyperlinked by QAnon websites. By comparison, only 1.8% and 1.5% of all domains hyperlinked by non-news websites and authentic news websites, respectively, are also hyperlinked by QAnon websites.

known to host a wide array of conspiracy theories, news, and different interest topics, experiencing an influx of users at the height of the PizzaGate conspiracy theory [85].

### 5.2 Interactions between Conspiracies, the News Media, and the Broader Web

Having given an overview of the conspiracy theory ecosystem, we now turn to understand these conspiracy theories' interactions with different parts of the Internet, and in particular, the news media. We do not heavily discuss the Multiple/Forum group of conspiracy theories here as we focus on how each *individual* conspiracy theory interacts with different parts of the Internet.

To understand the degree of interconnection between our conspiracy theory websites and different parts of the Internet, we plot in Figure 4 the percentage of shared domain connections between our conspiracy theory categories and our set of authentic news, misinformation, and non-news-related websites.

*Misinformation, Authentic News, and Non-News' Relationship with Different Conspiracy Theories.* As seen in Figure 4, non-news websites have relatively few connections with the same websites as conspiracy theories. All of the considered conspiracy theories have the least amount of shared connections with non-news websites. At the high end, only 1.8% of the websites collectively

| QAnon | Unique URLs | Multiple/Forum | Unique URLs | COVID | Unique URLs |
|---|---|---|---|---|---|
| zerohedge.com | 22,630 | naturalnews.com | 15,176 | healthimpactnews.com | 2,806 |
| thegatewaypundit.com | 8,751 | zerohedge.com | 11,892 | breitbart.com | 1,565 |
| wikileaks.org | 5,727 | rt.com | 10,825 | bitchute.com | 1,375 |
| theepochtimes.com | 5,502 | thegatewaypundit.com | 9,903 | rt.com | 1,221 |
| breitbart.com | 5,454 | breitbart.com | 8,529 | off-guardian.org | 707 |
| rt.com | 5,452 | bitchute.com | 6,299 | corbettreport.com | 700 |
| veteranstoday.com | 3,511 | activistpost.com | 5,299 | naturalnews.com | 672 |
| bitchute.com | 2,552 | humansarefree.com | 4,964 | zerohedge.com | 672 |
| freebeacon.com | 2,403 | veteranstoday.com | 4,365 | greenmedinfo.com | 566 |
| sputniknews.com | 2,211 | infowars.com | 3,855 | lifesitenews.com | 508 |
| **9/11** | **Unique URLs** | **Flat-Earth** | **Unique URLs** | **UFO/Aliens** | **Unique URLs** |
| globalresearch.ca | 2,082 | breitbart.com | 62 | coasttocoastam.com | 1,402 |
| prisonplanet.com | 1,734 | bitchute.com | 47 | collective-evolution.com | 275 |
| infowars.com | 1,245 | rt.com | 33 | rt.com | 268 |
| rawstory.com | 901 | thegatewaypundit.com | 31 | abovetopsecret.com | 143 |
| antiwar.com | 855 | icr.org | 27 | sputniknews.com | 130 |
| veteranstoday.com | 839 | naturalnews.com | 24 | humansarefree.com | 118 |
| wikileaks.org | 799 | zerohedge.com | 18 | disclose.tv | 102 |
| informationclearinghouse.info | 750 | rawstory.com | 16 | naturalnews.com | 88 |
| whatreallyhappened.com | 684 | globalresearch.ca | 13 | bitchute.com | 81 |
| rt.com | 663 | abovetopsecret.com | 12 | theepochtimes.com | 56 |

Table 4. **Top misinformation websites hyperlinked by each category of conspiracy theory website**—Conspiracy theory websites heavily use misinformation websites. Many of these misinformation websites have been thoroughly described in prior works as promoting toxic ecosystems [57, 112].

hyperlinked by non-news websites are also hyperlinked by QAnon websites. We see similar behavior with authentic news websites. For every conspiracy theory, at most 1.5% (QAnon) of authentic news collective connections are also hyperlinked by a conspiracy theory website category. This illustrates the degree to which authentic news websites have maintained a distance between themselves and conspiracy theories. In contrast, we see our misinformation domains collectively have nearly triple the number of shared connections between themselves and every conspiracy theory website category. For example, while only 1.5% of outward domains connections from authentic news websites are also hyperlinked by QAnon websites, 4.8% of these connections are also hyperlinked by misinformation websites (Figure 4a).

*Conspiracy Theories Relationship with Misinformation, Authentic News, and Non-News Websites.* Conspiracy theories often attempt to interpret current events and use the resources utilized by misinformation and authentic news domains to do so (Figure 4). For both misinformation and authentic news, every conspiracy theory has heightened levels of shared domain connections compared with non-news websites. For every conspiracy category, at least 37.4% of outward connections are hyperlinked by authentic news websites and at least 43.8% of their outward connections are hyperlinked by misinformation websites. This illustrates the high degree to which conspiracy theory websites, while ostensibly fringe, rely on websites and resources also utilized by more "mainstream" and popular websites. Performing a series of Mann-Whitney U-test comparing the relative levels of connections between every conspiracy theory website and mainstream and authentic news websites versus non-news websites, we find (after utilizing the Bonferonni correction

| QAnon | Unique URLs | Multiple/Forum | Unique URLs | COVID | Unique URLs |
|---|---|---|---|---|---|
| cnn.com | 15,606 | nytimes.com | 7,774 | nytimes.com | 1,352 |
| foxnews.com | 7,980 | theguardian.com | 6,360 | theguardian.com | 901 |
| usatoday.com | 6,008 | foxnews.com | 5,765 | cnn.com | 805 |
| nytimes.com | 3,941 | washingtonpost.com | 4,902 | washingtonpost.com | 680 |
| thehill.com | 3,203 | cnn.com | 4,544 | foxnews.com | 600 |
| theguardian.com | 2,747 | huffingtonpost.com | 2,738 | cnbc.com | 447 |
| cbsnews.com | 2,712 | bloomberg.com | 2,736 | bloomberg.com | 436 |
| washingtonpost.com | 2,598 | thehill.com | 2,279 | forbes.com | 384 |
| politico.com | 2,225 | businessinsider.com | 2,053 | thehill.com | 362 |
| washingtonexaminer.com | 1,820 | politico.com | 2,006 | nejm.org | 354 |

| 9/11 | Unique URLs | Flat-Earth | Unique URLs | UFO/Aliens | Unique URLs |
|---|---|---|---|---|---|
| nytimes.com | 3,172 | nytimes.com | 689 | examiner.com | 528 |
| cnn.com | 1,444 | washingtonpost.com | 297 | foxnews.com | 511 |
| washingtonpost.com | 1,299 | cnn.com | 289 | nytimes.com | 470 |
| theguardian.com | 703 | theatlantic.com | 273 | huffingtonpost.com | 414 |
| huffingtonpost.com | 645 | nationalgeographic.com | 249 | heraldtribune.com | 403 |
| salon.com | 465 | npr.org | 238 | livescience.com | 389 |
| examiner.com | 415 | wired.com | 217 | popularmechanics.com | 336 |
| nydailynews.com | 357 | eurekaalert.org | 210 | cnn.com | 328 |
| usatoday.com | 354 | theguardiean.com | 174 | theguardian.com | 317 |
| foxnews.com | 292 | forbes.com | 155 | washingtonpost.com | 309 |

Table 5. **Top authentic news websites hyperlinked by each category of conspiracy theory website**— Conspiracy theory websites make high use of authentic news websites. All conspiracy groups have nytimes.com and theguardian.com within their top 10 hyperlinked authentic news websites.

for multiple statistical [13]) that every conspiracy theory group does indeed have elevated levels of shared domain connections with authentic news and misinformation websites.

Looking at the specific domain-to-domain connections between our website lists, we further note that every conspiracy theory website category in our dataset also thoroughly utilizes our set of misinformation and authentic news websites. Each category collectively hyperlinks to at least 171 authentic news sites and at least 136 misinformation sites. Popular misinformation sites like rt.com, zerohedge.com, bitchute.com, and breitbart.com whose role in promoting extreme beliefs [56, 112] are frequently linked to by conspiracy theory websites (Table 4). Authentic news websites are more popular than misinformation, which may help explain this phenomenon. Using the Amazon Alexa Top List from March 1, 2021, we see that 255 (45.1%) of the websites in our authentic news list are in the top 100K websites while only 93 (17.5%) of the misinformation are in the top 100K [7]. Indeed, nytimes.com, theguardian.com, cnn.com, and washingtonpost.com are some of the most commonly hyperlinked authentic news sites by conspiracy theory websites (Table 5).

### 5.3 Summary

In this section, we documented the complex role conspiracy theories have had with the wider Internet and in particular the news media. First, we showed that QAnon remains a prominent outlet and repository of conspiracy theories generally, not just those specific to Q. This largely accords with news reporting that QAnon has become a "big tent conspiracy theory" that constantly incorporates new claims and theories [98]. We further documented that while our sets of conspiracy theory websites are considered fringe, each of them has websites that occupy a "core" place on the Internet. Each conspiracy theory category that we document has websites that are within the top 99% percentile of central websites on the Internet, according to both harmonic and PageRank centrality measures.

Second, we elucidated the role that the news media plays in the conspiracy theory ecosystem. We found that every conspiracy theory category in our dataset has significant percentages of shared domain connections with mainstream authentic and non-news domains. Thus while promoting fringe ideas, we see that each of the conspiracy theory categories still relies on and hyperlinks to many of the same websites that more "mainstream" websites also utilize. However, as expected, we find that out of the three categories of misinformation, authentic news, and non-news, misinformation websites have by far the most connections with our set of conspiracy theory websites. These misinformation websites have near 3x the percentage of shared domain connections with conspiracy theory websites compared to authentic news and non-news websites. We thus see that our misinformation domains are relatively heavily connected to conspiracy theory websites.

## 6 CONSPIRACY THEORIES AND THE NEWS MEDIA OVER TIME

In this section, we now discuss the interactions between online conspiracy theories and the news media ecosystem, and in particular misinformation, *over time*. Previous social science research has indicated that conspiratorial thinking has remained constant throughout much of human history [124, 126], spiking temporarily only during moments of social unrest. Given that (1) political polarization and social upheaval have increased significantly since 1994 in the United States [1] and (2) having determined that misinformation and authentic news both play a significant role in the conspiracy theory ecosystem, we now measure (1) whether the popularity of conspiracy theories have increased during this period and (2) if news sites have had an increasing effect on the conspiracy-theory ecosystem. Specifically, we analyze the changing popularity of each conspiracy theory, the increased connectivity between conspiracy theories and misinformation, and finally misinformation and authentic news' roles in affecting the popularity of conspiracy theories online.

*Conspiracy-Oriented Websites.* For the rest of this paper, we define *conspiracy-oriented* websites as websites that have more connections from conspiracy theory websites than from authentic news and non-news websites (i.e., the majority of a site's inward links in a domain-based graph are from conspiracy websites [excluding misinformation domains]). We use this definition to understand how misinformation and authentic news interact with conspiracy-related materials generally rather than just our 856 different websites. For this definition, we use our full list of 856 conspiracy theory websites. Given that conspiracy theory websites are smaller and have both fewer pages and hyperlinks to other websites, this is a conservative definition. Indeed, the median conspiracy theory site links to 83.5 other domains, the median non-news site to 176 sites, and the median authentic news website to 2195 sites. In addition, while there are 856 conspiracy theory websites in our dataset, there are a total of 1107 authentic news and non-news websites. Out of the 3.8 million different domains that were hyperlinked in our scrapes, 219K (5.73%) are *conspiracy-oriented.* 324 of our misinformation domains overlap with our *conspiracy-oriented* domains, including websites like zerohedge.com, gulagbound.com, and prisonplanet.tv. All of our conspiracy theory websites are considered *conspiracy-oriented.* We finally note that this definition removed fairly common popular websites like twitter.com, facebook.com, reddit.com, and nytimes.com, ensuring that this set of websites considers only conspiracy-oriented/conspiratorial websites.

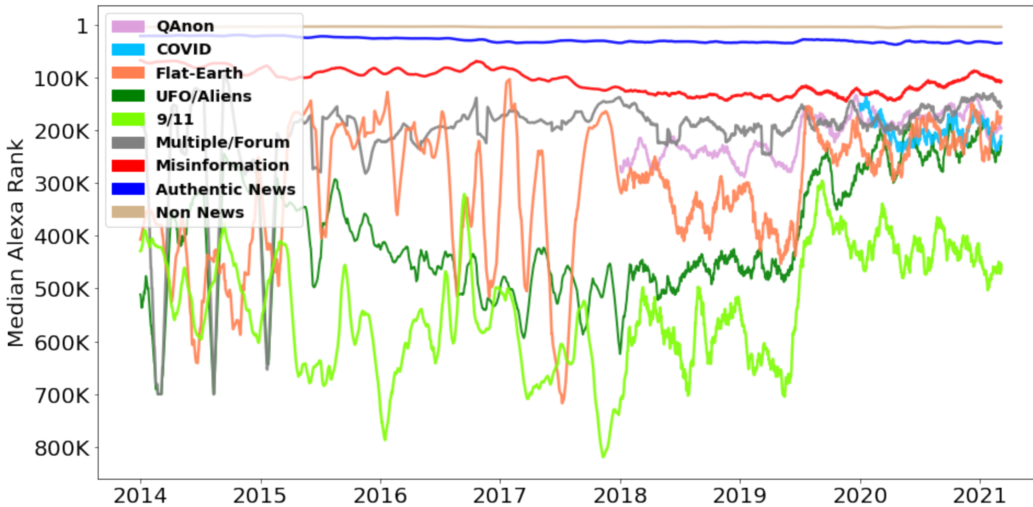



Fig. 5. **30-Day moving average of the median alexa rank of non-news, authentic news, misinformation, and each category of conspiracy theory website**— We note a large increase in the median popularity of all conspiracy theories website categories around August 2019 (we ensured that this trend is specific to these groups of websites by plotting the median Alexa rank of 500 random websites, validating that the rank of the set of websites was stable and did not exhibit the increase seen for the conspiracy theory websites). This increase in popularity preceded and was concurrent with both the Dayton and El Paso mass shootings in the USA. In both cases, the shooters posted their manifestos on the QAnon-associated website 8chan.net.

### 6.1 The Web's Changing Relationship with Different Conspiracy Theories

To begin, we first investigate the relative popularity of authentic news, misinformation, non-news, and each conspiracy theory over time. To do this, we consider each group's median rank in the Alexa Top Million list [7].

Between 2014–2021, the median Alexa rank of authentic news, misinformation, and non-news websites have all remained relatively stable, while conspiracy theory websites' popularity has varied widely (Figure 5). However, all conspiracy theories experienced a dramatic increase in popularity between July and August 2019. We ensured that this trend is specific to these groups of websites by plotting the median Alexa rank of 500 random websites. validating that the rank of the set of websites was stable and did not exhibit the increase seen for the conspiracy theory websites. While we cannot show definitive causality, we note that this coincides with the El Paso, Texas, and Dayton, Ohio, mass shootings that took place on August 3rd and 4th, respectively. In both cases, the shooters posted manifestos on the QAnon-associated website 8chan.net. As previously noted, QAnon material and QAnon websites like 8chan often host and share material for other conspiracy theories; the large simultaneous increase in popularity among each conspiracy website group largely follows from QAnon acting as a means by which people access other conspiracy theories [98]. We finally note that following these events, 8chan.net shut down, subsequently rebranding as 8kun.top. Since this increase in popularity, the median popularity of each conspiracy theory website group has remained high, with the exception of the 9/11 conspiracy theory (Figure 5).

Next looking at the interaction over time between news websites and conspiracy theory websites we see that misinformation websites have increasingly hyperlinked to conspiracy theory and *conspiracy-oriented* websites, concurrent with the increase in popularity of conspiracy theory

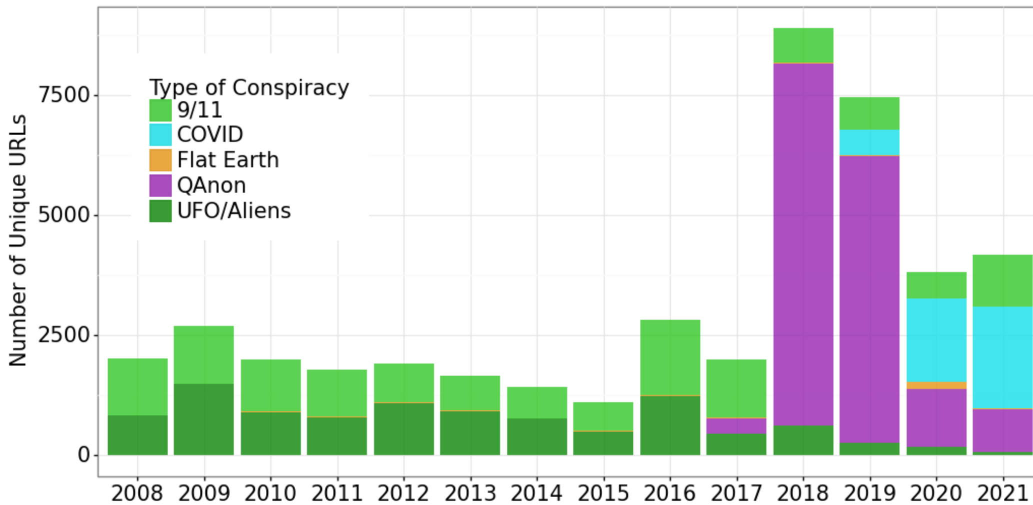



Fig. 6. **Hyperlinks from misinformation websites to individual conspiracy theory website groups**—(Note: Hyperlinks were only collected to Aug 2021) The vast majority of links from news websites to conspiracy theory websites are from misinformation sites. After the appearance of QAnon, there was a massive increase in links to conspiracy theory websites. The decline of QAnon was followed by a large increase in COVID conspiracy theory website links.

websites (Figures 5, 6, and 7). We specifically observe a massive jump in the number of hyperlinks to conspiracy theories due to the arrival of QAnon in 2017–2018 (Figure 6). This was later supplemented by COVID hyperlinks at the start of the COVID pandemic in late 2019 and early 2020. We see relatively few Flat-Earth links; this may suggest its somewhat separate role within the conspiracy theory ecosystem. This lack of links further matches the lower centrality of Flat-Earth websites that we observed in Section 5.1 and in Tables 5 and 4.

Between 2008 and 2021, looking at the average percentage of *conspiracy-oriented* external hyperlinks per source domain (especially compared to authentic new websites) misinformation websites have hyperlinked to more and more *conspiracy-oriented* domains. This peaked in 2020 at 20.9% from a low of 13.9% in 2008, a 50.4% relative increase (Figure 7). Looking at of all external hyperlinks from misinformation websites, instead of an average per domain, we see that the percentage that go to *conspiracy-oriented* starting from a low of 10.2% in 2008 goes to a high of 22.4% in 2021, a 119% relative increase. Looking at some of the top *conspiracy-oriented domains* hyperlinked, Blazetv.com is one of the most commonly linked. Blazetv.com is a pay television network founded by conservative commentator Glenn Beck [5]. The television network has long been known to spread misinformation. For example, the network spread debunked network 1000 mail-in ballots were found in a dumpster in California on September 2020 [16]. Some of other most commonly linked *conspriacy-oriented* websites are Gab, Parler, and MeWe (Table 6). All three sites are known to spread disinformation, conspiracy theories, and hate [8, 69, 74]. Parler was even utilized as a communication forum by rioters who attacked the United States Capitol on January 6, 2021 [18].

Even considering the larger set of 219K conspiracy-oriented domains, authentic news domains have largely avoided linking their audience to conspiratorial material, with the notable exception of some Russian websites. As seein in Table 6, The *conspiracy-oriented* domains that authentic news websites hyperlink to are primarily Russian websites. ok.ru is a Russian social media second

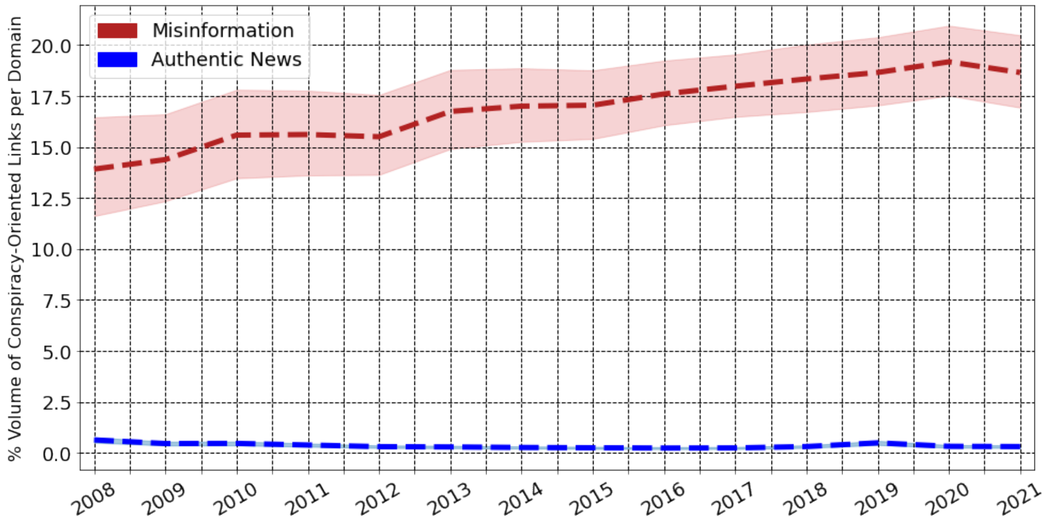



Fig. 7. **Average percentage of hyperlinks to *conspiracy-oriented* websites from misinformation and authentic news websites**— Plotted is the average percentage per source domain with 95% bootstrap confidence intervals of hyperlinks per to conspiracy-oriented websites from misinformation and authentic news websites. Misinformation websites have increasingly linked their users to conspiracy-oriented domains since 2008. Authentic news websites have refrained from directing their users to websites that promote conspiracy theories.

| Domain | Unique Hyperlinks | Domain | Unique Hyperlinks |
|---|---|---|---|
| blazetv.com | 3,308,119 | ok.ru | 59,803 |
| theabovenetwork.com | 2,581,659 | 9nl.us | 8,432 |
| mitocopper.com | 2,093,941 | rt.com | 8,169 |
| russian-faith.com | 1,445,839 | zerohedge.com | 1,559 |
| parler.com | 1,170.825 | universalpressrelease.com | 1,312 |
| banned.video | 931,672 | thegatewaypundit.com | 933 |
| zerohedge.com | 808,906 | paulcraigroberts.org | 920 |
| gab.com | 671,900 | christchurch.org.nz | 908 |
| gab.ai | 659,738 | antiwar.com | 783 |
| mewe.com | 622,483 | sputniknews.com | 768 |

Table 6. **Top *conspiracy-oriented* websites hyperlinked by misinformation websites (left) and authentic news websites (right) between January 2008 and August 2021**— Misinformation websites hyperlink to a large number of websites normally considered to promote conspiracy theories and misinformation, most prominently Gab, Parler, and MeWe [8, 69, 74]. Mainstream misinformation websites largely do not direct users to conspiracy-related domains but do occasionally link to Russian-operated websites that are known to spread disinformation.

in popularity in Russia, behind Russian social media website VK/VKontakte, and just ahead of Facebook [118]. rt.com is the website of the Russian Television (RT) network, known to spread disinformation [112]. 199 different authentic news websites hyperlink to articles from rt.com including nytimes.com and washingtonpost.com. As documented by Yablokov *et al.* [135], RT and the Russian government have consistently engaged with conspiracy theories as a form of public

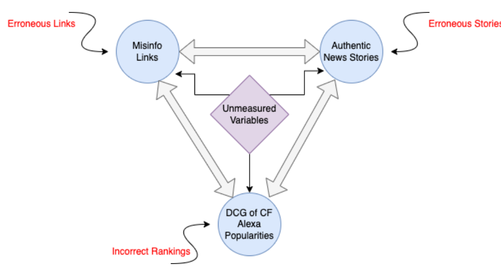



Fig. 8. **Model to determine partial Granger causality relationships between misinformation hyperlinks, authentic news stories, and the popularity of conspiracy-focused websites**— We model the influence of misinformation website hyperlinks and authentic news stories about conspiracy theories on the popularity of our set of conspiracy focused websites while taking into account erroneously documented stories, incorrect hyperlinks, and incorrect rankings from the Amazon Alexa top million list [7].

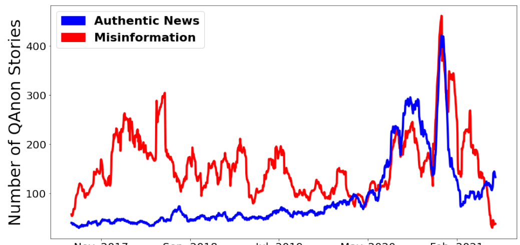
diplomacy. While, even including these websites, the percentage of *conspiracy-oriented* hyperlinks posted by authentic news sites is negligible (Figure 7), Russian websites appear as one of the few main vestiges of *conspiracy-oriented* ideas to which authentic news outlets hyperlink.

### 6.2 The Role of Misinformation and Authentic News in the Popularity of Conspiracy Theories

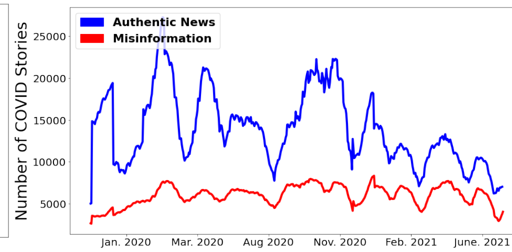
In the last section, we showed that misinformation websites frequently hyperlink to conspiracy theories, whereas authentic news websites rarely do. We now analyze whether the behavior of misinformation and authentic news have actively encouraged/discouraged the popularity of conspiracy theories. Specifically, we determine how changes in the volume of hyperlinks to conspiracy theory websites from misinformation websites affect the popularity of our conspiracy theory domains. Similarly, (given the relative lack of hyperlinks to our conspiracy theory websites from authentic news domains), we determine the relationship between the popularity of conspiracy theory websites and authentic news websites' frequency of mentioning different conspiracies.

*Experimental Setup.* To examine these relationships, we utilize the notion of *partial Granger causality* outlined in Section 4.2. For full details on partial Granger causality see Appendix B.[10] Utilizing this construct, we determine whether the number of hyperlinks to specific category of conspiracy theory websites from misinformation websites has a Granger-causal influence on the popularity of this same category of conspiracy theory websites. This is while also taking into account the influence of authentic news websites writing about the considered conspiracy theory as well as the influence of environmental exogenous inputs and endogenous latent variables. Conversely, we also determine whether the number of stories about a given conspiracy theory from authentic news websites has a partial Granger-causal influence on the popularity of this same conspiracy theory while also taking into account the influence of misinformation websites and environmental exogenous input and endogenous latent variables. For our experiment, we note that

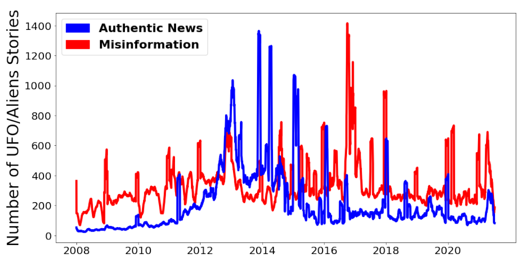
[10]To perform tests for Granger-causality, analyzed time series must be *stationary*. Stationarity implies that statistics (i.e. mean, variance) of the time series data must not change over time. Our untransformed data is not stationary given the large changes it makes over time (Figures 6, 5 and 9). We perform *difference transformation*, performing cross-checks against the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test to ensure stationarity. For all time-series data that we analyze, each series rejects the null hypothesis in the ADF test (implying stationarity) and accepts the null hypothesis in the KPSS test (also implying stationarity). To pick the appropriate lag (i.e., the number of past values to consider in the autoregressive model), we minimize the Aikake Information Criterion (AIC). We further note that to perform tests for Granger-causality, after fitting the autoregressive model, there must not be serial correlations among the residuals after fitting. Serial correlations imply that there are leftover patterns that are not explained by the model. For each model that we fit, we apply the Durbin-Watson test to test for serial correlations. For each model we fit, we find, utilizing this test, no evidence of serial correlations.



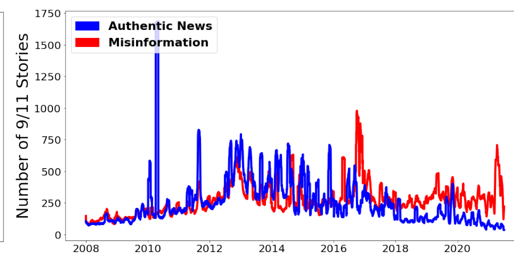

(a) QAnon



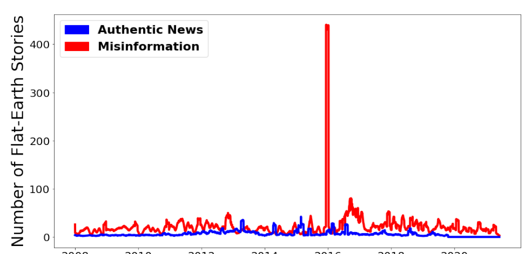

(b) COVID




(c) UFO/Aliens




(d) 9/11




(e) Flat Earth

Fig. 9. **News stories mentioning each conspiracy theory**— We show the number of misinformation and authentic news story that mention each conspiracy theory in our study.

the exogenous variables modeled would represent potentially erroneously documented links and stories, as well as ranks that were incorrect within the Amazon Alexa top list. The endogenous variables would represent variables that we did not measure for in this experiment but that had an underlying influence on our three time series. This would, for example, include the effect of social media on the number of misinformation hyperlinks, authentic news stores, and the popularity of conspiracy theory websites. We picture this model in Figure 8.

As a proxy for the popularity of each conspiracy theory, we utilize a binary version of the discounted cumulative gain (DCG) of the daily Amazon Alexa ranks. DCG is a measure of the quality of rankings [30, 64]. The binary version DCG is calculated as follows:

$$DCG(rankings) = \sum_{rank} \frac{1}{log_2(rank+1)}$$

DCG is useful for modeling the popularities of each conspiracy theory group because (1) we wish to model that movement between rankings near the top of the Amazon Alexa list is more important than near the bottom (i.e. moving from a rank of 100,000 to 10 is more significant than moving from a rank of 900,000 to 800,0010) and (2) DCG accounts for all the rankings of each conspiracy theory list. This is opposed to the median or mean, which do not have these two desired properties. We further note that to prevent us from comparing intra-conspiracy DCGs with different numbers of rankings (i.e. a given website can go in and out of the Alexa rankings without it meaning much for its actual popularity), as recommended [64], we enforce a small minimum value (i.e. rank of 1,000,000+) for websites without ranks (we find that this adjustment did not change our results). We finally note that Amazon Alexa changed its methodology for calculating rankings on January 30, 2018 [101, 117]. Previously, Alexa rankings were averaged over three months of data from URLs visited by users with a given browser extension installed. While the exact change was not announced by Amazon, it was found that the rankings after January 30, 20218, were not averaged over multiple days [117]. Given this change in methodology, for this analysis, we only utilize rankings and data from February 1, 2018, onward.

For the frequency of hyperlinks to conspiracy theory websites, we utilize the total daily number of links from all misinformation websites in our dataset to each category of conspiracy theory websites. To ascertain the date when each hyperlink was published we utilize the Python package `htmldate` [15]. As previously noted, due to the change in the methodology of Amazon Alexa ranking, we only measure the effect of hyperlinks on the popularity of websites from February 1, 2018, onward, filtering out hyperlinks published before then.

Given the paucity of links from authentic news websites to conspiracy theory websites, we rely on the daily number of authentic news articles mentioning each conspiracy theory topic ("QAnon", "COVID", "Flat-Earth", "UFO", "9/11") as a proxy for connection frequency. For example, we count the daily number of authentic news articles that mention “QAnon” and measure its effect on the popularity of QAnon websites. Given that we do not have a particular key word for the Multiple/Forum group, we exclude this conspiracy theory group from this analysis. In our scrapes of authentic news websites, there are 1.08M pages mentioning UFOs, 29K mentioning Flat-Earth, 1.23M mentioning 9/11, 134K mentioning QAnon, and 8.23M mentioning COVID.[11] How often authentic news sites have written about each of these conspiracy topics over time is shown in Figure 9. We again utilize the Python package `htmldate` to extract the publication date for each misinformation and authentic news page that we scrape [15]. Again, due to the change in the methodology of Amazon Alexa ranking, we only measure the effect of stories on the popularity of conspiracy theory websites from February 1, 2018, onward.

We finally note that after fitting our models, to gain a sense of the directionality of the Granger-causal relationships (*i.e.* if an increase in hyperlinks from misinformation websites to conspiracy theory websites leads to an increase or decrease in the popularity of these conspiracy theory websites), we examine the coefficients of our fit VAR models [44]. Specifically, if the fit coefficients of the causal variable are on average positive, then we consider the relationship positive (otherwise negative). This is such that if our model specifies, for example, that more hyperlinks from misinformation websites to conspiracy theory websites would have led to an increase in popularity in these same conspiracy websites, then we consider that relationship to be positive. Finally, we note that given that we test across our five different conspiracy theories, we utilize Benjamini-Hochberg correction procedure [17] for multiple hypothesis testing with a false discovery rate (FDR) of 0.05 to infer partial Granger causality. For details on the Benjamini-Hochberg procedure, see Appendix C.

*Misinformation Websites: Propping up Conspiracy Theories.* In Figures 10–14, we see that for the QAnon, 9/11, Flat-Earth, and COVID conspiracy theories, the DCG of the popularity of each

[11] In our scrapes our misinformation websites, there are 1.59M pages mentioning UFOs, 108K mentioning Flat-Earth, 1.30M mentioning 9/11, 205K mentioning QAnon, and 3.03M mentioning COVID.

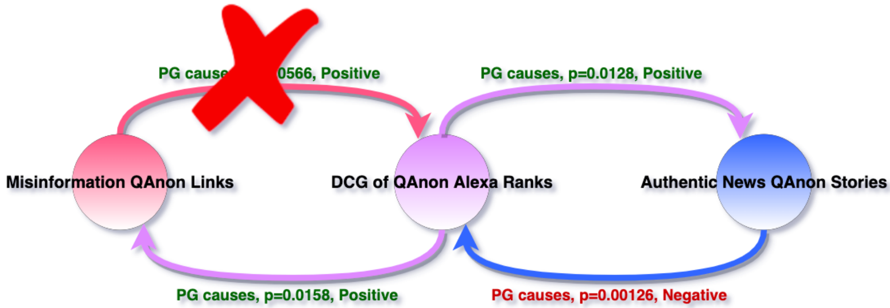



Fig. 10. **QAnon Granger causality loops**— (Non Granger-causal relationships are Xed) As the overall popularity of QAnon conspiracy theory websites increases, misinformation websites hyperlink to QAnon-focused conspiracy theory websites. As the popularity of QAnon websites increases, authentic news websites write about QAnon more; however, as authentic news sites write about QAnon more, the overall popularity of QAnon conspiracy tends to decrease.

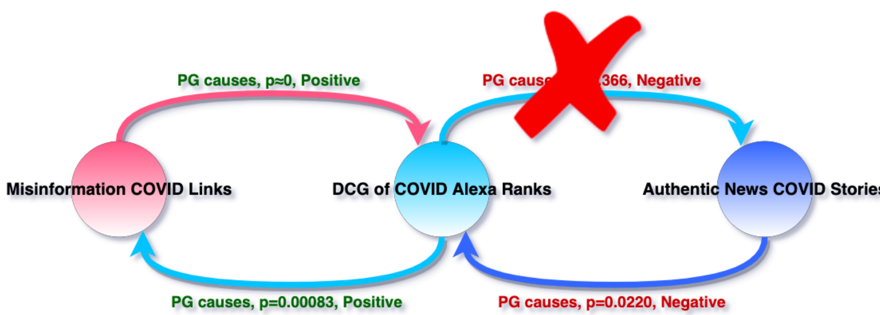



Fig. 11. **COVID Granger causality loops**— (Non Granger-causal relationships are Xed) As the number of hyperlinks from misinformation websites to COVID conspiracy theory websites increases this in turn increases the popularity of COVID conspiracy theory websites. As the popularity of these COVID conspiracy theory websites goes up, this in turn increases the number of hyperlinks that misinformation websites utilize. However, as authentic news websites write about COVID more, the overall popularity of COVID conspiracy theory websites decreases.

conspiracy theory group had a positive partial Granger causality with the number of hyperlinks to these same websites. Simply put, the popularity of conspiracy theory websites positively affects the number of hyperlinks from the misinformation outlets to these websites. We thus see, that for these four conspiracy theories that in addition to helping to spread extreme beliefs, misinformation websites take advantage of periods when these conspiracies are popular to spread them further. We further see that for the COVID, UFO, 9/11, and Flat-Earth conspiracy theories that as misinformation websites hyperlinked to them more, the corresponding websites increased in popularity. For these four conspiracy theories, misinformation websites thus play a role in helping popularizing the conspiracy theories. We suspect that for the QAnon conspiracy theory that social media played a larger and more dominating role in popularizing it across the Internet and for this reason we were unable to identify a definitive relationship between misinformation hyperlinks and its popularity [57, 85, 85]. Altogether, these results suggest that misinformation websites play a prominent role in popularizing and spreading conspiracy theories online.

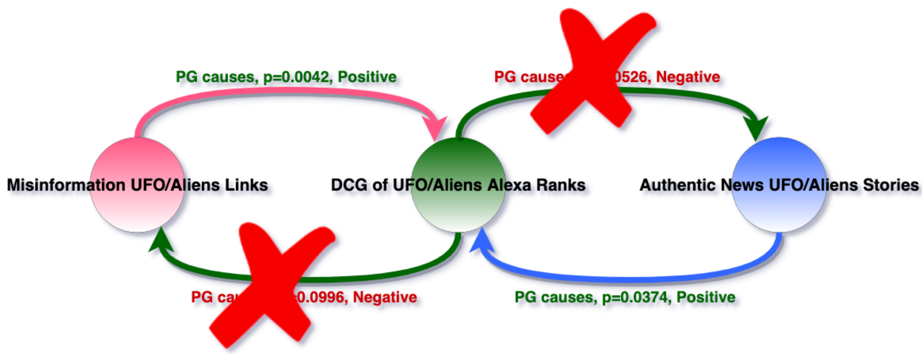



Fig. 12. **UFO/Aliens Granger causality loops**— (Non Granger-causal relationships are Xed) As the number of hyperlinks from misinformation websites to UFO/Aliens conspiracy theory websites increases this, in turn, increases the popularity of UFO/Aliens websites. Similarly, as authentic news sites write about UFO/Aliens more, the overall popularity of conspiracy theory sites increases.

*Authentic News Websites: Complex Relationships.* In contrast to hyperlinks from misinformation websites, the effect of stories from authentic news websites on the popularity of these conspiracy theories is more complex. For instance, we observe a positive Granger-causal relationship between the popularity of QAnon websites and the number of authentic news QAnon stories (Figure 9). Thus as QAnon websites became more popular, authentic news websites wrote more about QAnon. This result concurs with the fact that authentic news websites only began to write about the conspiracy theory after it began to have a marked effect on U.S. politics (Figure 9). In contrast, however, we see that the more authentic news websites wrote about QAnon, the more the popularity of QAnon websites decreased. We suspect that as prominent outlets began to write about the conspiracy theory, this led to heightened pressure on platforms to reign in QAnon's spread, decreasing the conspiracy theory's popularity. For example, Reddit [119], Twitter [120], and Facebook [78] all cracked down on the QAnon conspiracy theory following intense media scrutiny of its spread on their platforms [52, 57, 104]. We see a similar relationship for Flat-Earth websites and COVID conspiracy theories. Flat-Earth conspiracy theories have been repeatedly excoriated in mainstream news media [34, 71] and this severe scrutiny appears to have been correlated with a decrease in the popularity of several Flat-Earth websites. Similarly, many news websites have taken a highly proactive role in combating COVID-19 conspiracy theories [12, 40, 45], and this appears to have been correlated with a decrease in websites advocating these conspiracy theories. NewsGuard, for instance, even published a list of specific websites that were spreading COVID conspiracy theories [76] that was used extensively by news organizations and researchers.[12]

Besides these relationships, we further see a positive Granger-casual feedback relationship between the popularity of 9/11 conspiracy theory websites and the number of stories that authentic news publish concerning 9/11 (Figure 14). We suspect that as 9/11 annually returns as a topic of interest in the news, 9/11 conspiracy theory websites receive more traffic and authentic news websites publish more about the event. Finally, for the UFO/Aliens conspiracy theory, we see in Figure 12 that as authentic news websites write about UFOs and Aliens more frequently UFO/Alien conspiracy theory websites become more popular. This suggests that as media sites write about UFOs, Internet users *do* look up and visit UFO/Aliens websites more often.

[12]We did not use this list in our work due to the $6,000 licensing fee.

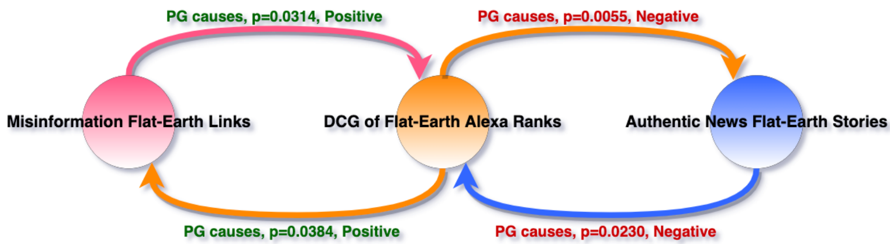



Fig. 13. **Flat-Earth Granger causality loops**— As the amount of hyperlinks from misinformation websites to Flat-Earth conspiracy theory websites increase, this in turn increases the popularity of Flat-Earth conspiracy theory websites. As the popularity of these Flat-Earth conspiracy theory websites goes up, the amount of hyperlinks that misinformation websites use also increases. In contrast, as the popularity of Flat-Earth websites increases, authentic news websites wrote about Flat-Earth less. Furthermore when authentic news websites wrote more about the Flat-Earth conspiracy theory, the popularity of Flat-Earth websites decreased.


Fig. 14. **9/11 Granger causality loops**— As the number of hyperlinks from misinformation websites to 9/11 conspiracy theory websites increases this, in turn, increases the popularity of 9/11 conspiracy theory websites; similarly as the popularity of these websites increases, the number of hyperlinks from misinformation to these 9/11 conspiracy theory websites increases as well. As the popularity of these 9/11 conspiracy theory websites goes up, the number of stories that authentic news stories write about 9/11 increases; and finally when authentic news websites write about 9/11 the popularity of 9/11 conspiracy theory websites increases.

### 6.3 Summary

In this section, we documented the intensifying role that misinformation sites have played in promoting conspiracy theories. Indeed, from 2008 to 2020, the average percentage of *conspiracy-oriented* hyperlinks from misinformation websites per domain increased from 13.9% to 20.4%, a 50.4% relative increase. Looking at the raw percentage of all external hyperlinks from misinformation websites that went to *conspiracy-oriented* websites, instead of a per-domain average, this percentage went from 10.2% to 22.4%, a near 119% relative increase. Then using partial Granger causality analysis, we demonstrated how misinformation websites hyperlinks have contributed to the popularity of conspiracy theory websites and illustrated the more complex relationship authentic news websites have with conspiracy theory websites.

## 7 EXPLAINING AND MEASURING AND MISINFORMATION WEBSITES' FRINGE AND CONSPIRATORIAL ATTITUDES

In Sections 5 and 6, we explored the interconnectivity between the news media and conspiracy theory ecosystems and showed how misinformation websites promote the growth of conspiracy theories. We now investigate a driving force behind conspiratorial and fringe material online—political polarization. Prior work has suggested that political polarization can help explain conspiratorial

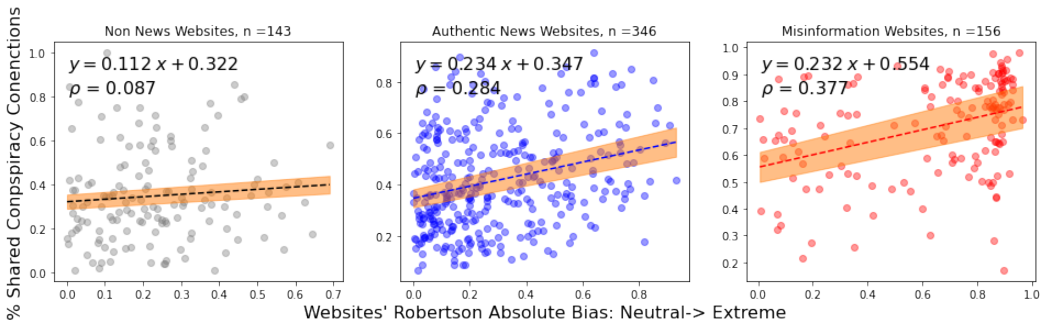



Fig. 15. **Partisan bias vs. conspiracy theory similarity**—For news-related domains (i.e, misinformation and authentic news websites), as partisan bias levels increase, websites have a higher percentage of shared domain connections with conspiracy theory websites.

beliefs [21, 22, 123], and according to the Pew Research Center, partisan bias has increased significantly since 1994 [1]. Furthermore, this increase in partisanship largely corresponds with the rise in conspiracy theories online (Figure 7). In this section, we thus investigate political polarization's role in the evolution and growth of conspiracy theories.

Specifically, utilizing our lists of conspiracy theory websites and external political polarization data, we show that partisan bias has a $\rho = 0.521$ Pearson correlation with websites' graph similarity to conspiracy theory websites. We conclude this section by proposing a new metric to measure websites' fringe attitudes, which objectively measures a website's fringe attitudes by taking into account, partisanship and levels of *conspiracy-oriented* hyperlinking.

### 7.1 Political Partisanship and Similarity to Conspiracy

Although not a causative analysis and limited to US partisanship, we now look to establish whether a website's partisanship level corresponds with its similarity to conspiracy theory websites or its likelihood to post conspiratorial content. For this analysis, we determine whether a website's orientation to conspiracy theories is a function of the website's audience's partisanship, as defined by Robertson *et al.*'s partisanship bias dataset [94]. Robertson *et al.*'s dataset presents partisanship data on a scale of -1 (liberal/Democratic-leaning) to +1 (conservative/Republican-leaning) for 19K websites. Cross-referencing this dataset with our curated lists of websites, we find that 192 authentic, 186 misinformation, and 143 non-news websites have scores in Robertson *et al.*'s dataset [94] and that had at least 100 domain connections. For each website in this combined dataset, we define the website's similarity to conspiracy theory websites as the percentage of external domains that it links to that are *also* linked to by our group of 856 conspiracy theory websites. This metric is essentially the overlap node similarity metric [50, 134], giving us an approximation of how similar a given website is to conspiracy theory websites generally. We utilize all domains rather than just *conspiracy-oriented* domains to fully capture how the behavior of a website emulates that of conspiracy theory websites. Given our use of connections as a factor, we finally note that filter our dataset to contain only the set of websites having connections to at least 100 other domains, leaving 346 authentic, 156 misinformation, and 143 non-news websites.

Echoing Section 5, in Figure 15, we see that individual misinformation websites have the highest average level (55.4%) of shared domain connections with conspiracy theory websites. Authentic news websites have the second-highest average (34.7%), marginally higher than non-news websites at (32.2%). We further observe that for both authentic news and misinformation websites, as partisan bias levels increase, the similarity to conspiratorial websites increases. For authentic news and misinformation, the Pearson correlation values are 0.284 and 0.377 respectively, suggesting a moderate level of correspondence for both groups. For non-news-related sites, partisanship levels have nearly no correlation/slight-correlation with a website's similarity to conspiracy theory websites.

Combining the 156 misinformation and 346 authentic news websites, we observe an overall 0.521 Pearson correlation between political partisanship and similarity to conspiracy theories. We note that for solely conservative/Republican-leaning news websites, this correlation increases to 0.621. For solely liberal/Democratic-leaning news websites, this correlation decreases to 0.284. This indicates that for liberal-leaning news websites, partisanship may not be as prominent a factor in the promotion of conspiratorial material as for conservative-leaning news websites. These results largely agree with concurrent work based on survey data that found across 26 different countries that as polarization increases on both ends of the political spectrum (left and right), conspiratorial thinking increases [61].

### 7.2 Measuring Websites' Fringe Attitudes

US partisanship on the Republican/Democratic spectrum can explain part of how fringe or a website is, but it is not the only factor. Indeed, Robertson *et al.* [94] found that several Russian misinformation websites have little to no US right/left partisan bias. For example, Russian misinformation websites like katehon.com and globalresearch.ca do not have a clear place on the US spectrum, and yet both often spread conspiracy theories much more readily than news sites like cnn.com and the dailywire.com that have liberal and conservative biases. We thus propose to combine these two concepts (political polarization and conspiracy-orientation) to fully understand how websites' fringe attitudes by proposing a new metric that takes into account both of these factors. Given that similarity to *conspiracy-oriented* websites and political polarization levels (as calculated by Robertson *et al.* [94]) can be objectively and automatically determined, this new metric can be utilized at scale to identify fringe websites.

For this section, we define *conspiracy-oriented* websites as we did in Section 6.1 (websites that have more connections to conspiracy theory websites than to non-news or authentic news websites). We use *conspiracy-oriented* websites in this metric to ensure that only connections to more fringe material are considered (the similarity to conspiracy theory metric from Section 7.1 considers all connections that conspiracy theory websites have, including to benign websites). Given our use of connections as a factor, we again filter this dataset to contain only the set of domains having connections to at least 100 other domains (N=502).

Before building our metric, we observe in Figure 16 that using simple thresholding on partisanship levels and each website's percentage of *conspiracy-oriented* connections, we can easily differentiate between misinformation and authentic news. Creating a simple threshold with just these two inputs using an 80% training subset of our N=502 dataset, we achieve a 97.0% accuracy on the remaining 20% test set (using a Support Vector Machine (SVM) to decide on the threshold boundary). Across both all websites considered, this simple thresholoding approach achieved a 93.1% precision and a false positive rate of 2.6% in labeling the 20% test set. Given that misinformation websites generally are more fringe (see Section 7.1), this suggests that these two inputs are highly discriminatory in identify fringe attitudes. We again note that both of these inputs can be automatically determined for a given website using their hyperlinks and utilizing Robertson *et al.* [94] approach.

Utilizing the drawn boundary for classification, we now define our metric for understanding a website's "fringe attitudes." We define this metric as a function of the distance of each website to the SVM boundary dividing misinformation from authentic news. Namely, we calculate this metric in the same way SVMs calculate class probabilities. SVM probabilities in the binary case

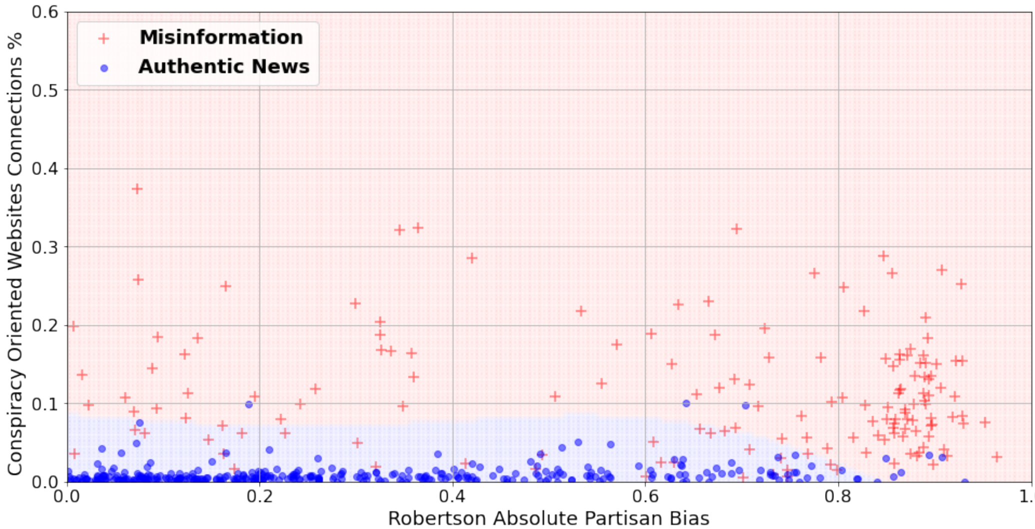



Fig. 16. **Utilizing partisan bias and conspiracy orientation to understand websites' "fringe attitudes"**— Using a simple threshold on the partisan bias of a website and the number of conspiracy domains that it links to, one can differentiate between authentic news and misinformation. This illustrates the usefulness of these two criteria in understanding websites' fringe attitudes.

are normally calculated utilizing Platt scaling: $P(class|input) = P(y = 1|x) = p(x) = \frac{1}{1+exp(-f(x))}$ where f(x) is a function learned from the training data. For more details on SVMs and Platt scaling see Platt *et al.* [91]. Thus, in our case, the score for a given website is the probability it is grouped with misinformation websites given its partisanship level and percentage of *conspiracy-oriented* hyperlinks. Derived as such, values vary from 0 (not fringe) to 1 (fringe).

Looking at the scores for the different groups of sites shown in Table 7, we see that among the authentic news sites that were misclassified, nearly all were hyper-partisan websites. Thus even among websites that were misclassified as misinformation, most had fringe elements, evidencing our new metric's usefulness. For several, it is questionable whether they are authentic news websites. For example, the most politically extreme and conspiratorial mainstream website, humanevents.com is a known far-right hyper-conservative site [24]. On the day of writing (October 1, 2021), the top story on the site is "THE DUCHESS IS A VICTIM AND FOOTBALL IS GAY" (emphasis not added). We further see that the metric was able to correctly identify several misinformation websites that do not fall on US political spectrum. For example, even though globalresearch.ca is not particularly partisan on the US political spectrum, it was still given a high fringe score (Table 7).

## 8 DISCUSSION

In this work, we utilize readily identifiable conspiracy theory website groups to understand how conspiracy theories have influenced and affected the Internet, particularly the news media ecosystem over the past decade. From initial analysis, we see that conspiracy theory websites characterize and share a similar ecosystem with misinformation websites. The self-reinforcing relationship between misinformation and conspiracy theory websites, we found to have positive partial Granger-causal feedback loops. We further found that political partisanship levels largely correlate with websites' shared domain connections with conspiracy theory websites. Using these initial findings, we presented a new metric for identifying a given website's penchant for politically polarizing

| Domain | Fringe Score | Domain | Fringe Score |
|---|---|---|---|
| healthnutnews.com | 1.0 | humanevents.com | 0.953 |
| disclose.tv | 1.0 | freedomworks.org | 0.948 |
| americanjournalreview.com | 1.0 | kenyonreview.org | 0.912 |
| teaparty.org | 1.0 | ronpaulinstitute.org | 0.897 |
| oathkeepers.org | 1.0 | dailywire.com | 0.855 |
| usdefensewatch.com | 1.0 | consortiumnews.com | 0.851 |
| thecommonsenseshow.com | 1.0 | davidstockmanscontracorner.com | 0.836 |
| newstarget.com | 1.0 | heritage.org | 0.790 |
| guccifer2.wordpress.com | 1.0 | dissentmagazine.com | 0.719 |
| globalresearch.ca | 1.0 | amsterdamnews.com | 0.556 |

Table 7. **Most Fringe (*i.e* partisan and conspiratorial) misinformation (left) and authentic news websites (right)**—Based on the distance of a given website from the boundary dividing misinformation and authentic news, we devise a metric to measure a website's politically extreme and conspiratorial nature. As seen the most fringe mainstream sites are hyper-partisan websites.

material and conspiratorial thinking. Now, in this section, we give an overview of the implications of these results.

### 8.1 Conspiracies on the Internet

The past decade has been one of significant political upheaval with some arguing that we live in a "Golden Age of Conspiracies" [111]. In this work, we document the role of six different conspiracy theory categories on the Internet. We unsurprisingly find that QAnon is one of the largest and most prolific conspiracy theories on the Internet. Given QAnon has become a repository for a lot of different conspiratorial beliefs, we argue that understanding exactly how this conspiracy theory works and de-radicalizing its adherents can help de-radicalize those influenced by other conspiracy theories.

While some have argued that conspiratorial thinking has remained constant throughout human history, we also see that several conspiracy theories have become a larger and larger part of the hyperlinks on misinformation websites during the past decade. We indeed see that since 2008, the percentage of conspiracy-oriented links by misinformation websites has increased dramatically. Besides misinformation websites, however, we find that other types of websites have largely not promoted conspiracy theory websites or conspiratorial material. This suggests somewhat of a divergence of the types of websites online: those prone to conspiracy and those that are not. While it is uncertain whether the true believers of conspiracies have increased since 2008, misinformation websites' audiences *have* been exposed to more conspiratorial content.

We further find that misinformation websites' role goes further than merely hyperlinking to conspiracy theory websites. We find that as the misinformation publish more hyperlinks to conspiracy theory websites that this in turn leads to the increased popularity of these websites; similarly as the popularity of these conspiracy theory websites increase, misinformation websites often hyperlink to them more in turn. This suggests that as misinformation outlets gain more prominence, conspiracy theories will as well.

If platforms and researchers want to tackle conspiracy theories and their potentially detrimental effects, we argue that we should focus our efforts on curtailing the role of large misinformation outlets. Given this uptick in exposure to conspiracy theories from misinformation websites, online platforms must take action against the most egregious actors. With our metric that accounts for both conspiratorial and political partisanship, platforms can filter and identify websites with the most fringe elements.

### 8.2 Polarization and Conspiracy

Prior research suggests that polarization can help explain conspiratorial beliefs [21, 22, 123]. For example, researchers have shown that Americans with higher political polarization levels are more likely to believe in birtherism (the conspiracy theory that former President Barack Obama was not born in the United States) [39]. In this work, we find that the partisanship levels of news websites have a Pearson correlation of $\rho = 0.521$ with a website's shared external domain connections with conspiracy theory websites, increasing to $\rho = 0.621$ when only considering conservative-leaning websites. This indicates, reinforcing prior work [61] published in Nature, that as societies, websites, and the Internet as a whole become more politically polarized, conspiratorial ideas will increase in tandem.

### 8.3 Tracking Misinformation

Labeling a website as a peddler of misinformation or as fringe is a difficult thing to do. It requires rigorous fact-checking and objective measures of truth, which are expensive and difficult to obtain. Furthermore, the decision to categorize a website as misinformation or fringe has become intensely political [2]. For example, the site mediabiasfactcsheck.com labels the Washington Post as "mostly factual" due to failing two fact checks and using loaded words, despite many considering it a very trusted source [130].

We argue that given that conspiracy theories are readily identifiable and that the number of shared connections is an objective measure, our metric, which is based on websites' penchants to promote *conspriacy-orietned* materials and their partisanship, can help ameliorate issues surrounding the labeling of websites. When building our presented metric, we note that it is much easier to label large conspiratorial movements than individual news sites that may get some things right and other things wrong. Identifying flat-earth websites is largely uncontroversial. Similarly, for obtaining partisanship scores, Robertson *et al.* used how often Republicans and Democrats linked to given websites on Twitter, another objective measures [94]. For those websites not already within their dataset, partisanship levels can be easily obtained using their methodology. As such, using our metric can help introduce a level of objectivity into grading given sites as conspiratorial or as promoting fringe materials. If researchers can use objective metrics, building on previous work in misinformation labeling, they will be able to more effectively identify and further stem the flow of misinformation and fringe material on the Internet.

## 9 CONCLUSION

In this work, we studied the relationships between conspiracy theories, misinformation, and political polarization on the Internet by detailing the inner workings and interactions of 856 websites dedicated to conspiracy theories and prominent news outlets. We found that QAnon remains the most prominent conspiracy theory on the Internet, feeding people into other online conspiratorial movements. We then detailed how misinformation websites in particular connect people to conspiracy materials in contrast to authentic news websites and non-news websites. Using the notion of partial Granger-causality, we detailed the time-correlative relationships that exist between the number of hyperlinks on misinformation websites and the popularity of these conspiracy theory websites, also describing the more complex relationship that authentic news websites have with conspiracy theory websites. Examining the driving forces behind these phenomena, we discuss how political polarization levels heavily correlate ($\rho = 0.521$) with news websites' shared connections with conspiracy theory websites. We finally combined these two features into a single metric to measure the political polarization and conspiratorial nature of a given website, finishing by discussing the implications for future research into online conspiracy theories.

## REFERENCES


[1] 2017. The Partisan Divide on Political Values Grows Even Wider | Pew Research Center. https://www.pewresearch.org/politics/2017/10/05/the-partisan-divide-on-political-values-grows-even-wider/

[2] 2020. Fake news is fooling more conservatives than liberals. Why? | The Economist. https://www.economist.com/international/2020/06/03/fake-news-is-fooling-more-conservatives-than-liberals-why

[3] Sara Abdali, Rutuja Gurav, Siddharth Menon, Daniel Fonseca, Negin Entezari, Neil Shah, and Evangelos E Papalexakis. 2021. Identifying Misinformation from Website Screenshots. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 2–13.

[4] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *ACM SIGSAC Conference on Computer and Communications Security*.

[5] Rizwan Ahmad. 2022. Analysis of Media Bias—Glenn Beck TV Shows: A Content Analysis. *Journal of Creative Communications* 17, 1 (2022), 67–87.

[6] Syeda Zainab Akbar, Anmol Panda, Divyanshu Kukreti, Azhagu Meena, and Joyojeet Pal. 2021. Misinformation as a Window into Prejudice: COVID-19 and the Information Environment in India. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–28.

[7] Alexa Internet, Inc. 2020. Top 1,000,000 Sites. http://s3.amazonaws.com/alexa-static/top-1m.csv.zip.

[8] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. 2021. An early look at the parler online social network. *arXiv preprint arXiv:2101.03820* (2021).

[9] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.

[10] Hunt Allcott, Matthew Gentzkow, and Chuan Yu. 2019. Trends in the diffusion of misinformation on social media. *Research & Politics* 6, 2 (2019), 2053168019848554.

[11] Amarnath Amarasingam and Marc-André Argentino. 2020. The QAnon conspiracy theory: A security threat in the making. *CTC Sentinel* 13, 7 (2020), 37–44.

[12] James T. Areddy. 2020. Coronavirus Conspiracy Theory Claims It Began in the U.S.—and Beijing Is Buying It - WSJ. https://www.wsj.com/articles/canadian-writer-fuels-china-u-s-tiff-over-coronaviruss-origins-11585232018

[13] Richard A Armstrong. 2014. When to use the Bonferroni correction. *Ophthalmic and Physiological Optics* 34, 5 (2014), 502–508.

[14] Jeffrey M Bale. 2007. Political paranoia v. political realism: On distinguishing between bogus conspiracy theories and genuine conspiratorial politics. *Patterns of prejudice* 41, 1 (2007), 45–60.

[15] Adrien Barbaresi. 2020. htmldate: A Python package to extract publication dates from web pages. *Journal of Open Source Software* 5, 51 (2020), 2439.

[16] Samia Benaissa Pedriza. 2021. Sources, Channels and Strategies of Disinformation in the 2020 US Election: Social Networks, Traditional Media and Political Candidates. *Journalism and Media* 2, 4 (2021), 605–624.

[17] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.

[18] Katie Benner. 2021. Parler says it sent the F.B.I. posts about threats to the Capitol ahead of Jan. 6. - The New York Times. https://www.nytimes.com/2021/03/25/us/parler-fbi-capitol-attack.html

[19] Eirikur Bergmann. 2018. *Conspiracy & populism: The politics of misinformation.* Springer.

[20] Alessandro Bessi, Guido Caldarelli, Michela Del Vicario, Antonio Scala, and Walter Quattrociocchi. 2014. Social determinants of content selection in the age of (mis) information. In *International conference on social informatics*. Springer, 259–268.

[21] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one* 10, 2 (2015), e0118093.

[22] Alessandro Bessi, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Viral misinformation: The role of homophily and polarization. In *Proceedings of the 24th international conference on World Wide Web*. 355–356.

[23] Shweta Bhatt, Sagar Joglekar, Shehar Bano, and Nishanth Sastry. 2018. Illuminating an ecosystem of partisan websites. In *Companion Proceedings of the The Web Conference 2018*. 545–554.

[24] Jake Bittle. 2021. Ronald Reagan's Favorite Magazine Has Gone Full MAGA | The New Republic. https://newrepublic.com/article/164396/human-events-trump-ronald-reagan-magazine-maga

[25] Robert Brotherton, Christopher C French, and Alan D Pickering. 2013. Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in psychology* (2013), 279.

[26] Susanne Bruckmüller, Peter Hegarty, Karl Halvor Teigen, Gisela Böhm, and Olivier Luminet. 2017. When do past events require explanation? Insights from social psychology. *Memory Studies* 10, 3 (2017), 261–273.

[27] Michael Butter and Peter Knight. 2020. *Routledge handbook of conspiracy theories*. Routledge.

[28] Global Engagement Center. 2020. Pillars of Russia's disinformation and propaganda ecosystem.

[29] Kung Sik Chan and Howell Tong. 1986. On estimating thresholds in autoregressive models. *Journal of time series analysis* 7, 3 (1986), 179–190.

[30] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. 2018. Investigating the impact of gender on rank in resume search engines. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–14.

[31] Yonghong Chen, Govindan Rangarajan, Jianfeng Feng, and Mingzhou Ding. 2004. Analyzing multiple nonlinear time series with extended Granger causality. *Physics letters A* 324, 1 (2004), 26–35.

[32] Lu Cheng, Ruocheng Guo, Kai Shu, and Huan Liu. 2021. Causal understanding of fake news dissemination on social media. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 148–157.

[33] Jodi Dean. 1998. *Aliens in America: Conspiracy cultures from outerspace to cyberspace*. Cornell University Press.

[34] EJ Dickson. 2022. Flat Earth: TikTok, YouTube Launched 19th Cent. Conspiracy Theory - Rolling Stone. https://www.rollingstone.com/culture/culture-features/flat-earth-conspiracy-theory-podcast-1311262/

[35] Karen M Douglas and Robbie M Sutton. 2008. The hidden impact of conspiracy theories: Perceived and actual influence of theories surrounding the death of Princess Diana. *The Journal of social psychology* 148, 2 (2008), 210–222.

[36] Karen M Douglas, Joseph E Uscinski, Robbie M Sutton, Aleksandra Cichocka, Turkay Nefes, Chee Siang Ang, and Farzin Deravi. 2019. Understanding conspiracy theories. *Political Psychology* 40 (2019), 3–35.

[37] Lauren Dyson and Alden Golab. 2017. Fake News Detection Exploring the Application of NLP Methods to Machine Identification of Misleading News Sources. *CAPP 30255 Adv. Mach. Learn. Public Policy* (2017).

[38] David Easley and Jon Kleinberg. 2010. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge university press.

[39] Adam M. Enders and Steven M. Smallpage. 2018. On the measurement of conspiracy beliefs. *Research and Politics* 5 (1 2018). Issue 1. https://doi.org/10.1177/2053168018763596

[40] Adam M. Enders and Jospeh E. Uscinski. 2020. Coronavirus conspiracy theories are how people try to feel more in control - The Washington Post. https://www.washingtonpost.com/outlook/2020/05/05/coronavirus-conspiracy-theories-pandemic/

[41] Adam M. Enders, Joseph E. Uscinski, Casey A. Klofstad, Michelle I. Seelig, Stefan Wuchty, Manohar N. Murthi, Kamal Premaratne, and John R. Funchion. 2021. Do Conspiracy Beliefs Form a Belief System? Examining the Structure and Organization of Conspiracy Beliefs. *Journal of Social and Political Psychology* 9, 1 (Jun. 2021), 255–271. https://doi.org/10.5964/jspp.5649

[42] Dan Evon. 2019. Did a Georgia Lawmaker Claim a Chick-fil-A Employee Told Her to Go Back to Her Country? | Snopes.com. https://www.snopes.com/fact-check/georgia-lawmaker-go-back-claim/

[43] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[44] Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and Agenda-setting in Russian News: a Computational Analysis of Intricate Political Strategies. In *2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[45] Max Fisher. [n. d.]. Why Coronavirus Conspiracy Theories Flourish. And Why It Matters. - The New York Times. https://www.nytimes.com/2020/04/08/world/europe/coronavirus-conspiracy-theories.html

[46] Kiran Garimella and Dean Eckles. 2020. Images and misinformation in political groups: Evidence from WhatsApp in India. *Harvard Kennedy School Misinformation Review* (2020).

[47] Kiran Garimella, Tim Smith, Rebecca Weiss, and Robert West. 2021. Political Polarization in Online News Consumption. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 152–162.

[48] Amira Ghenai and Yelena Mejova. 2017. Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 518–518.

[49] Ted Goertzel. 1994. Belief in conspiracy theories. *Political psychology* (1994), 731–742.

[50] Mark K Goldberg, Mykola Hayvanovych, and Malik Magdon-Ismail. 2010. Measuring similarity between sets of overlapping clusters. In *2010 IEEE Second International Conference on Social Computing*. IEEE, 303–308.

[51] Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424–438.

[52] Rachel Greenspan. 2021. QAnon Bans: How Tech Companies Are Responding to the Conspiracy Theory. https://www.businessinsider.com/how-every-tech-company-social-media-platform-handling-qanon-2020-

10#youtube-announced-a-bigger-crackdown-on-qanon-but-stopped-short-of-an-explicit-ban-3
[53] Andrew Guess, Brendan Nyhan, and Jason Reifler. 2018. Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. *European Research Council* 9, 3 (2018), 4.
[54] Shuixia Guo, Anil K Seth, Keith M Kendrick, Cong Zhou, and Jianfeng Feng. 2008. Partial Granger causality—eliminating exogenous inputs and latent variables. *Journal of neuroscience methods* 172, 1 (2008), 79–93.
[55] Catherine Han, Deepak Kumar, and Zakir Durumeric. 2022. On the Infrastructure Providers That Support Misinformation Websites. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 287–298.
[56] Hans WA Hanley, Deepak Kumar, and Zakir Durumeric. 2022. Happenstance: Utilizing Semantic Search to Track Russian State Media Narratives about the Russo-Ukrainian War On Reddit. (2022).
[57] Hans WA Hanley, Deepak Kumar, and Zakir Durumeric. 2022. No Calm in the Storm: Investigating QAnon Website Relationships. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 299–310.
[58] Richard Hofstadter. 2012. *The paranoid style in American politics.* Vintage.
[59] Austin Hounsel, Jordan Holland, Ben Kaiser, Kevin Borgolte, Nick Feamster, and Jonathan Mayer. 2020. Identifying Disinformation Websites Using Infrastructure Features. In *USENIX Workshop on Free and Open Communications on the Internet.*
[60] Y Linlin Huang, Kate Starbird, Mania Orand, Stephanie A Stanek, and Heather T Pedersen. 2015. Connected through crisis: Emotional proximity and the spread of misinformation online. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 969–980.
[61] Roland Imhoff, Felix Zimmer, Olivier Klein, João HC António, Maria Babinska, Adrian Bangerter, Michal Bilewicz, Nebojša Blanuša, Kosta Bovan, Rumena Bužarovska, et al. 2022. Conspiracy mentality and political orientation across 26 countries. *Nature human behaviour* 6, 3 (2022), 392–403.
[62] Caroline Jack. 2017. Lexicon of lies: Terms for problematic information. *Data & Society* 3, 22 (2017), 1094–1096.
[63] Suchita Jain, Vanya Sharma, and Rishabh Kaushal. 2016. Towards automated real-time detection of misinformation on Twitter. In *2016 International conference on advances in computing, communications and informatics (ICACCI).* IEEE, 2015–2020.
[64] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* 20, 4 (2002), 422–446.
[65] Shan Jiang and Christo Wilson. 2018. Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.
[66] Alex Kasprak. 2019. Did Colin Kaepernick Lobby to Remove the National Anthem from Football? | Snopes.com. https://www.snopes.com/fact-check/kaepernick-remove-national-anthem/
[67] Quyu Kong, Emily Booth, Francesco Bailo, Amelia Johns, and Marian-Andrei Rizoiu. 2022. Slipping to the Extreme: A Mixed Method to Explain How Extreme Opinions Infiltrate Online Discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 524–535.
[68] Stephan Lewandowsky, Ullrich KH Ecker, Colleen M Seifert, Norbert Schwarz, and John Cook. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest* 13, 3 (2012), 106–131.
[69] Binny Mathew, Ritam Dutt, Pawan Goyal, and Animesh Mukherjee. 2019. Spread of hate speech in online social media. In *Proceedings of the 10th ACM conference on web science.* 173–182.
[70] Mahalia Miller, Conal Sathi, Daniel Wiesenthal, Jure Leskovec, and Christopher Potts. 2011. Sentiment flow through hyperlink networks. In *AAAI Conference on Web and Social Media.*
[71] Shaheed N Mohammed. 2019. Conspiracy theories and flat-earth videos on YouTube. *The Journal of Social Media in Society* 8, 2 (2019), 84–102.
[72] Jeremiah Morelock and Felipe Ziotti Narita. 2022. The Nexus of QAnon and COVID-19: Legitimation Crisis and Epistemic Crisis. *Critical Sociology* (2022), 08969205211069614.
[73] Durim Morina and Michael S Bernstein. 2022. A Web-Scale Analysis of the Community Origins of Image Memes. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–25.
[74] Mary E Myers. 2021. Propaganda, Fake News, and Deepfaking. In *Understanding Media Psychology.* Routledge, 161–181.
[75] Sebastian Nagel. 2021. Host- and Domain-Level Web Graphs February/March, April and May 2021 – Common Crawl. https://commoncrawl.org/2021/05/host-and-domain-level-web-graphs-feb-apr-may-2021/
[76] NewsGuard. 2021. COVID-19 Misinformation Resources - NewsGuard. https://www.newsguardtech.com/covid-19-resources/
[77] Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.
[78] Ciaran O'Conner, Cooper Gatewood, Kendrick McDonald, and Sarah Brandt. 2020. The Boom Before the Ban: QAnon and Facebook - ISD. https://www.isdglobal.org/isd-publications/the-boom-before-the-ban-qanon-and-facebook/

[79] J Eric Oliver and Thomas Wood. 2014. Medical conspiracy theories and health behaviors in the United States. *JAMA internal medicine* 174, 5 (2014), 817–818.

[80] J. Eric Oliver and Thomas J. Wood. 2014. Conspiracy Theories and the Paranoid Style(s) of Mass Opinion. *American Journal of Political Science* 58 (10 2014), 952–966. Issue 4. https://doi.org/10.1111/AJPS.12084

[81] Ray Oshikawa, Jing Qian, and William Yang Wang. 2020. A Survey on Natural Language Processing for Fake News Detection. In *Proceedings of the 12th Language Resources and Evaluation Conference*. 6086–6093.

[82] Shira Ovide. 2021. Trump. Twitter. QAnon. Who's to Blame for DC Capitol Riot? https://www.nytimes.com/2021/01/07/technology/trump-twitter-qanon.html.

[83] John C Paolillo. 2018. The flat earth phenomenon on YouTube. *First Monday* (2018).

[84] Antonis Papasavva, Max Aliapoulios, Cameron Ballard, Emiliano De Cristofaro, Gianluca Stringhini, Savvas Zannettou, and Jeremy Blackburn. 2022. The gospel according to Q: Understanding the QAnon conspiracy from the perspective of canonical information. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 735–746.

[85] Antonis Papasavva, Jeremy Blackburn, Gianluca Stringhini, Savvas Zannettou, and Emiliano De Cristofaro. 2021. "Is it a Qoincidence?": An Exploratory Study of QAnon on Voat. In *Proceedings of the Web Conference 2021*. 460–471.

[86] Stephanie Pappas. 2021. Flat Earth 'theory' — why do some people think the Earth is flat? | Live Science. https://www.livescience.com/24310-flat-earth-belief.html

[87] Marius Paraschiv, Nikos Salamanos, Costas Iordanou, Nikolaos Laoutaris, and Michael Sirivianos. 2022. A Unified Graph-Based Approach to Disinformation Detection using Contextual and Semantic Relations. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 747–758.

[88] Han Woo Park. 2003. Hyperlink network analysis: A new method for the study of social structure on the web. *Connections* (2003).

[89] Shruti Phadke, Mattia Samory, and Tanushree Mitra. 2021. Characterizing social imaginaries and self-disclosures of dissonance in online conspiracy discussion communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–35.

[90] Shruti Phadke, Mattia Samory, and Tanushree Mitra. 2022. Pathways through Conspiracy: The Evolution of Conspiracy Radicalization through Engagement in Online Conspiracy Discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 770–781.

[91] John Platt et al. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers* 10, 3 (1999), 61–74.

[92] Vahed Qazvinian, Emily Rosengren, Dragomir Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 1589–1599.

[93] Sonia Rao. 2022. Trevor Noah's best jokes at the White House correspondents' dinner - The Washington Post. https://www.washingtonpost.com/arts-entertainment/2022/04/30/trevor-noah-white-house-correspondents-dinner/

[94] Ronald E Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. 2018. Auditing partisan audience bias within google search. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–22.

[95] Yannick Rochat. 2009. *Closeness centrality extended to unconnected graphs: The harmonic centrality index*. Technical Report.

[96] Bjorn Roelstraete and Yves Rosseel. 2012. Does partial Granger causality really eliminate the influence of exogenous inputs and latent variables? *Journal of neuroscience methods* 206, 1 (2012), 73–77.

[97] Daniel Romer and Kathleen Hall Jamieson. 2020. Conspiracy theories as barriers to controlling the spread of COVID-19 in the US. *Social science & medicine* 263 (2020), 113356.

[98] Kevin Roose. 2021. What Is QAnon, the Viral Pro-Trump Conspiracy Theory? - The New York Times. https://www.nytimes.com/article/what-is-qanon.html

[99] Mattia Samory and Tanushree Mitra. 2018. Conspiracies online: User discussions in a conspiracy community following dramatic events. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[100] Mattia Samory and Tanushree Mitra. 2018. 'The Government Spies Using Our Webcams' The Language of Conspiracy Theories in Online Discussions. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–24.

[101] Quirin Scheitle, Oliver Hohlfeld, Julien Gamba, Jonas Jelten, Torsten Zimmermann, Stephen D Strowes, and Narseo Vallina-Rodriguez. 2018. A long way to the top: Significance, structure, and stability of internet top lists. In *Proceedings of the Internet Measurement Conference 2018*. 478–493.

[102] Djamé Seddah, Farah Essaidi, Amal Fethi, Matthieu Futeral, Benjamin Muller, Pedro Javier Ortiz Suárez, Benoît Sagot, and Abhishek Srivastava. 2020. Building a user-generated content North-African Arabizi treebank: Tackling hell. In *ACL 2020-58th Annual Meeting of the Association for Computational Linguistics*.

[103] Vibhor Sehgal, Ankit Peshin, Sadia Afroz, and Hany Farid. 2021. Mutual Hyperlinking Among Misinformation Peddlers. *arXiv preprint arXiv:2104.11694* (2021).

[104] Ari Sen and Brandy Zadrozny. 2020. QAnon groups have millions of members on Facebook, documents show. https://www.nbcnews.com/tech/tech-news/qanon-groups-have-millions-members-facebook-documents-show-n1236317

[105] Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. 2020. NLP-based feature extraction for the detection of COVID-19 misinformation videos on YouTube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.

[106] Karishma Sharma, Emilio Ferrara, and Yan Liu. 2022. Characterizing online engagement with disinformation and conspiracies in the 2020 US presidential election. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 908–919.

[107] Karishma Sharma, Emilio Ferrara, and Yan Liu. 2022. Construction of Large-Scale Misinformation Labeled Datasets from Social Media Discourse using Label Refinement. In *Proceedings of the ACM Web Conference 2022*. 3755–3764.

[108] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.* 19, 1 (sep 2017), 22–36. https://doi.org/10.1145/3137597.3137600

[109] Jason Smith, Herve Saint-Amand, Magdalena Plamadă, Philipp Koehn, Chris Callison-Burch, and Adam Lopez. 2013. Dirt cheap web-scale parallel text from the common crawl. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1374–1383.

[110] PRRI Staff. [n. d.]. Understanding QAnon's Connection to American Politics, Religion, and Media Consumption. https://www.prri.org/research/qanon-conspiracy-american-politics-report/

[111] Zack Stanton. 2020. You're Living in the Golden Age of Conspiracy Theories - POLITICO. https://www.politico.com/news/magazine/2020/06/17/conspiracy-theories-pandemic-trump-2020-election-coronavirus-326530

[112] Kate Starbird, Ahmer Arif, Tom Wilson, Katherine Van Koevering, Katya Yefimova, and Daniel Scarnecchia. 2018. Ecosystem or echo-system? Exploring content sharing across alternative media domains. In *Proceedings of the International AAAI Conference on Web and Social Media*.

[113] Carl Stempel, Thomas Hargrove, and Guido H Stempel III. 2007. Media use, social structure, and belief in 9/11 conspiracy theories. *Journalism & Mass Communication Quarterly* 84, 2 (2007), 353–372.

[114] Ana Stojanov et al. 2015. Reducing conspiracy theory beliefs. *Psihologija* 48, 3 (2015), 251–266.

[115] Cass R Sunstein. 2018. *# Republic: Divided democracy in the age of social media*. Princeton University Press.

[116] Art Swift. 2013. Majority in US still believe JFK killed in a conspiracy. *Gallup. com* 15 (2013).

[117] Samaneh Tajalizadehkhoob and Victor Le Pochat. 2019. The Tale of Website Popularity Rankings: An Extensive Analysis | RIPE Labs. https://labs.ripe.net/author/samaneh_tajalizadehkhoob_1/the-tale-of-website-popularity-rankings-an-extensive-analysis/

[118] Sprinklr Team. 2017. Sprinklr Spotlight: OK.ru, Russia's Oldest Social Network. https://www.sprinklr.com/blog/okru-russia-oldest-social-network/

[119] Kaitlyn Tiffany. 2020. Why QAnon Left Reddit. https://www.theatlantic.com/technology/archive/2020/09/reddit-qanon-ban-evasion-policy-moderation-facebook/616442/.

[120] Craig Timber. 2020. QAnon still rampant on Twitter despite July takedown. https://www.washingtonpost.com/technology/2020/10/03/twitter-banished-worst-qanon-accounts-more-than-93000-remain-site-research-shows/.

[121] Craig Timberg and Elizabeth Dwoskin. 2021. QAnon groups on Telegram seethe with covid denialism and vaccine misinformation - The Washington Post. https://www.washingtonpost.com/technology/2021/03/11/with-trump-gone-qanon-groups-focus-fury-attacking-covid-vaccines/

[122] Joseph E Uscinski, Adam M Enders, Casey Klofstad, Michelle Seelig, John Funchion, Caleb Everett, Stefan Wuchty, Kamal Premaratne, and Manohar Murthi. 2020. Why do people believe COVID-19 conspiracy theories? *Harvard Kennedy School Misinformation Review* 1, 3 (2020).

[123] Joseph E Uscinski and Santiago Olivella. 2017. The conditional effect of conspiracy thinking on attitudes toward climate change. *Research & Politics* 4, 4 (2017), 2053168017743105.

[124] Joseph E Uscinski and Joseph M Parent. 2014. *American conspiracy theories*. Oxford University Press.

[125] Jan-Willem Van Prooijen and Karen M Douglas. 2017. Conspiracy theories as part of history: The role of societal crisis situations. *Memory studies* 10, 3 (2017), 323–333.

[126] Jan-Willem Van Prooijen and Paul AM Van Lange. 2014. *Power, politics, and paranoia: Why people are suspicious of their leaders*. Cambridge University Press.

[127] Cornelius Adrian Vermeule and Cass Robert Sunstein. 2009. Conspiracy theories: causes and cures. *Journal of Political Philosophy* (2009).

[128] MK Vijaymeena and K Kavitha. 2016. A survey on similarity measures in text mining. *Machine Learning and Applications: An International Journal* 3, 2 (2016), 19–28.

[129] Yuping Wang, Fatemeh Tamahsbi, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, David Magerman, Savvas Zannettou, and Gianluca Stringhini. 2020. Understanding the Use of Fauxtography on Social Media. *arXiv preprint arXiv:2009.11792* (2020).

[130] Amy Watson. 2021. • Washington Post credibility in the U.S. 2021 | Statista. https://www.statista.com/statistics/739442/washington-post-credibility-usa/

[131] Brian E Weeks. 2015. Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of communication* 65, 4 (2015), 699–719.

[132] Elizabeth Williamson and Emily Steel. 2018. Conspiracy Theories Made Alex Jones Very Rich. They May Bring Him Down. https://www.nytimes.com/2018/09/07/us/politics/alex-jones-business-infowars-conspiracy.html.

[133] Sijia Xiao, Coye Cheshire, and Amy Bruckman. 2021. Sensemaking and the Chemtrail Conspiracy on the Internet: Insights from Believers and Ex-believers. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–28.

[134] Jierui Xie, Stephen Kelley, and Boleslaw K Szymanski. 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *Acm computing surveys (csur)* 45, 4 (2013), 1–35.

[135] Ilya Yablokov and Precious N Chatterje-Doody. 2021. *Russia Today and Conspiracy Theories: People, Power and Politics on RT.* Routledge.

[136] Funda Yurdakul and Erdogan Cevher. 2015. Determinants of current account deficit in Turkey: the conditional and partial Granger causality approach. *Procedia Economics and Finance* 26 (2015), 92–100.

[137] Savvas Zannettou, Tristan Caulfield, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. 2020. Characterizing the use of images in state-sponsored information warfare operations by russian trolls on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media.*

[138] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtelris, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2017. The web centipede: understanding how web communities influence each other through the lens of mainstream and alternative news sources. In *Proceedings of the 2017 internet measurement conference.* 405–417.

[139] Marvin Zonis and Craig M Joseph. 1994. Conspiracy thinking in the Middle East. *Political Psychology* (1994), 443–459.


## A GRANGER CAUSAILITY

Granger causality is a means of measuring if one time series is useful for forecasting another. Granger causality compares the efficacy of predicting a stochastic process $Y$ using *all the information in the universe* $U$ with the efficacy of using all information except for some other stochastic process $X$, denoted $U \backslash X$. If $X$ reduces the predictive power of $U$ regarding $Y$, then $X$ Granger causes $Y$. More formally:

(1) Let $X$ and $Y$ be stationary stochastic processes.
(2) Denote $U_i$ as all the information in the universe $U$ up to time $i$ ($U_{i-1}, ..., U_{i-\infty}$), $Y_i$ as all the information in $Y$ up to time $i$ ($Y_{i-1}, ..., Y_{i-\infty}$), and $X_i$ as all the information in $X$ up to time $i$ ($X_{i-1}, ..., X_{i-\infty}$).
(3) Denote $\sigma^2(Y_i|U_i)$ as the variance of the residual from the prediction of $Y_i$ using $U_i$ at time $i$.
(4) Denote $\sigma^2(Y_i|U_i \backslash X_i)$ as the variance of the residual from the prediction using all the information in $U_i$ at time $i$ except for $X_i$.

if $\sigma^2(Y_i|U_i) < \sigma^2(Y_i|U_i \backslash X_i)$ then $X$ Granger-causes $Y$ or $X \rightarrow Y$. If $Y$ also Granger-causes $X$ then we also say, $X \leftrightarrow Y$ and *feedback* is occurring [51].

## B PARTIAL GRANGER CAUSALITY

Previous works have found that repeatedly applying the original notions of Granger causality to identify causal relationships can lead to misleading results [31, 54]. An extended form of Granger causality, conditional Granger (CG) causality attempts to correct for the interaction of multiple times series while identifying relationships. CG causality thus determines whether one time-series Granger causes a second time-series conditional on a third or fourth time series. However, CG causality assumes that all relevant variables have been included in the model. Given that we cannot fully model *all* interactions amongst news, misinformation, and conspiracy theory websites, we instead utilize an extended form of CG causality namely *Partial Granger Causality* (PGC) to determine and understand these interactions.

Time-domain Partial Granger Causality specifically models exogenous environmental variables and unmeasured endogenous variables as well as conditioned interactions to mitigate confounding effects [54, 136]. Specifically, to determine the Granger-causal impact of a time-series $Y_t$ on another time series $X_t$ given the influence of a third time-series $Z_t$ as well as exogenous inputs and unmeasured latent models, PGC models the three time series together in a restricted and an unrestricted set of linear vector autoregressive (VAR) models. The restricted model assumes that the two time series $X_t$ and $Z_t$ are linearly dependent on past values of themselves, as well as time-dependent exogenous inputs, endogenous latent variables, and noise terms.

$$\begin{aligned} X_t &= \Sigma_{i=1}^{p} a_{1i} X_{t-i} + \Sigma_{i=1}^{p} b_{1i} Z_{t-i} + \overrightarrow{\epsilon_{1t}} + \overrightarrow{\epsilon_{1t}^{E}} + \overrightarrow{B_1(L)\epsilon_{1t}^{L}} \\ Z_t &= \Sigma_{i=1}^{p} c_{1i} X_{t-i} + \Sigma_{i=1}^{p} d_{1i} Z_{t-i} + \overrightarrow{\epsilon_{2t}} + \overrightarrow{\epsilon_{2t}^{E}} + \overrightarrow{B_2(L)\epsilon_{2t}^{L}} \end{aligned}$$

with noise covariance matrix $S$ and noise terms $e_{it}$,

$$S = \begin{bmatrix} var(e_{1t}), cov(e_{1t}, e_{2t}) \\ cov(e_{2t}, e_{2t}), var(e_{2t}) \end{bmatrix}$$

$$e_{it} = \overrightarrow{\epsilon_{it}} + \overrightarrow{\epsilon_{it}^{E}} + \overrightarrow{B_i(L)\epsilon_{it}^{L}}$$

where p is lag/the number of past terms that affect the current values of the time-series, matrices $A_1$, $B_1$, $C_1$, and $D_1$ contain coefficients that model these autogressive effects, $\overrightarrow{\epsilon_i}$ model Gaussian white noise processes, $\overrightarrow{\epsilon_i^{E}}$ model zero-mean exogenous Gaussian variables, and $B_1(L)\overrightarrow{\epsilon_i^{L}}$ model latent variables. Specifically $B_1(L)\overrightarrow{\epsilon_{it}^{L}}$ incorporate latent variables by assuming that the $t$th network element receives unmeasured inputs of the form $\Sigma_j^N x_{tj}/N$ where here each $x_{tj}$ is a stationary time series and $j$ is the latent index.

In contrast, the unrestricted model assumes $X_t$ and $Z_t$, as well as $Y_t$, are all linearly dependent on the past values of these three time series as well as time-dependent exogenous inputs, endogenous latent variables, and noise terms.

$$\begin{aligned} X_t &= \Sigma_{i=1}^{p} a_{2i} X_{t-i} + \Sigma_{i=1}^{p} b_{2i} Y_{t-i} + \Sigma_{i=1}^{p} c_{2i} Z_{t-i} + \overrightarrow{\epsilon_{3t}} + \overrightarrow{\epsilon_{3t}^{E}} + \overrightarrow{B_3(L)\epsilon_{3t}^{L}} \\ Y_t &= \Sigma_{i=1}^{p} d_{2i} X_{t-i} + \Sigma_{i=1}^{p} f_{2i} Y_{t-i} + \Sigma_{i=1}^{p} g_{2i} Z_{t-i} + \overrightarrow{\epsilon_{4t}} + \overrightarrow{\epsilon_{4t}^{E}} + \overrightarrow{B_4(L)\epsilon_{4t}^{L}} \\ Z_t &= \Sigma_{i=1}^{p} h_{2i} X_{t-i} + \Sigma_{i=1}^{p} k_{2i} Y_{t-i} + \Sigma_{i=1}^{p} m_{2i} Z_{t-i} + \overrightarrow{\epsilon_{5t}} + \overrightarrow{\epsilon_{5t}^{E}} + \overrightarrow{B_5(L)\epsilon_{5t}^{L}} \end{aligned}$$

with noise covariance matrix $\Sigma$ and noise terms $e_{it}$,

$$\Sigma = \begin{bmatrix} var(e_{3t}), cov(e_{3t}, e_{4t}), cov(e_{3t}, e_{5t}) \\ cov(e_{4t}, e_{3t}), var(e_{4t}), cov(e_{4t}, e_{5t}) \\ cov(e_{5t}, e_{3t}), cov(e_{5t}, e_{4t}), var(e_{5t}) \end{bmatrix}$$

$$e_{it} = \overrightarrow{\epsilon_{it}} + \overrightarrow{\epsilon_{it}^{E}} + \overrightarrow{B_i(L)\epsilon_{it}^{L}}$$

Taking this all into account, PGC calculates the ratio of (1) a measure based on the noise covariance matrix $S$ in the first model of the accuracy of predicting the present value of X using the history of X, conditioned on Z and eliminating the influence of the exogenous variables $\overrightarrow{\epsilon_i^{E}}$ and the latent variables $\overrightarrow{B_i(L)\epsilon_i^{L}}$, and (2) a measure based on the noise covariance matrix $\Sigma$ in the second model of the accuracy of predicting the present value of X using on the history of both X *and* $Y$ conditioned on Z and eliminating the influence of the exogenous variables $\overrightarrow{\epsilon_i^E}$.

$$F_1 = ln\left(\frac{S_{11} - S_{12}^{-1}S_{21}}{\Sigma_{11} - \Sigma_{12}^{-1}\Sigma_{21}}\right)$$

As specified [54, 96] the distribution of $F_1$ can be used to determine whether $Y_t$ Granger-causes $X_t$. Using the same noise covariance matrices, whether $Y_t$ Granger-causes $X_t$ can be determined as well as $Z_t$ Granger-causing $Y_t$, etc. Thus to test for Granger causality between two time-series $X_t$ and $Y_t$ conditioned on a third time-series $Z_t$ and taking into account exogenous environmental variables as well as endogenous latent variables we test against the null hypothesis that the second model does not add information in predicting future values of $X_t$.

## C BEJAMINI-HOCHBERG PROCEDURE

The Benjamini-Hochberg procedure is means of controlling the false discovery rate when performing multiple statistical tests. The procedure, after performing all tests and acquiring their individual p-values is as follows:

(1) Sort the p-values in ascending order.
(2) Assign ranks to the p-values of each test.
(3) Calculate each individual p-value's Benjamini-Hochberg critical value, using the formula:

$$(i/m)Q$$

where $i$ is the individual p-value's rank, $m$ is the total number of tests, and $Q$ is the false discovery rate (FDR).

(4) Compare the original p-values to the critical Benjamini-Hochberg values from Step 3. If the original p-value is smaller than its critical values, then reject the null hypothesis for that given statistical test; otherwise, accept the null hypothesis.