



Connecting the domains: an investigation of internet domains found in Covid-19 conspiracy tweets

J. D. Moffitt¹ · Catherine King¹ · Kathleen M. Carley¹

Accepted: 15 June 2023
© The Author(s) 2023

Abstract

Conspiracy theories (CTs) have thrived during the COVID-19 pandemic and continue to spread on social media despite attempts at fact-checking. The isolation and fear associated with this pandemic likely contributed to the generation and spread of these theories. Another possible factor is the high rate of Twitter users linking to off-platform alternative news sources through URL sharing (Moffitt et al. 2021). In this paper, we compare URLs and their parent domains linked in CT and non-CT tweets. First, we searched the parent domains of URLs shared in conspiracy theory and non-conspiracy theory classified tweets for the presence of Google tracking codes. We then constructed meta-networks linking domains, tracking codes, and Twitter users to find connections between domains and evidence of an eco-system that may have contributed to the cultivation and spread of conspiracy theories during the pandemic.

Keywords Conspiracy theories · COVID-19 · Network science · Social media analysis

1 Introduction

Soon after the COVID-19 pandemic began in early 2020, the World Health Organization (WHO) acknowledged that they were also fighting an “infodemic” (UN tackles “infodemic”, 2020). An “infodemic” is when too much information, including false and unverified statements, begins spreading rapidly. This “infodemic” of public health mis-/dis-information became a pressing issue because false and misleading stories can

✉ J. D. Moffitt
jdmoffit@cs.cmu.edu

Catherine King
cking2@cs.cmu.edu

Kathleen M. Carley
kathleen.carley@cs.cmu.edu

¹ Carnegie Mellon University, Pittsburgh, PA 15213, USA

spread incredibly quickly and can negatively impact public health behavior (Imhoff and Lamberty 2020). Some of the initial pandemic-related misinformation themes included the ideas that COVID-19 was a hoax, a Chinese bioweapon, a plan to microchip everyone through vaccines, or caused by 5G (Evanega et al. 2020).

During crises, people have a higher tendency to believe in conspiracy theories, as these theories can provide them with an understanding of a complicated and scary problem (Douglas et al. 2019; Oleksy et al. 2020; Sternisko et al. 2020). For example, a Pew Research survey from the summer of 2020 found that just over a third of survey respondents aware of the conspiracy theory that influential people planned the virus responded that they believed that theory to be either probably or definitely true (Basu 2020).

While conspiracy theories help people feel like they are in control, they can negatively impact human behavior. For example, individuals who did not believe the pandemic was genuine were less likely to engage in physical distancing and were less likely to wash their hands frequently (Imhoff and Lamberty 2020). Additionally, conspiratorial beliefs have other dangerous consequences outside of the public health space, such as an association with everyday crime, extremism, and tendency to commit violence (Douglas et al. 2019; Sternisko et al. 2020, 2020). Conspiracy theories can also incite anti-democratic movements (Sternisko et al. 2020).

Social media platforms and the Internet more broadly have increased both the number of people exposed to conspiracy theories and the speed at which these theories can spread (Douglas et al. 2019). Researchers can employ open source intelligence (OSINT) techniques to investigate the “bad actors” using social media platforms to spread conspiracy theories. OSINT is a framework for gathering publicly available information and then collating it for synthesis and insights (Glassman and Kang 2012). In this study, we leverage proven OSINT techniques for finding hidden connections between entities on the Internet to investigate how domains that share COVID-19 conspiracy-related content are connected. Because of the harmful effects on offline public health behavior, studying the spread of conspiracy theories on social media platforms and investigating links to off-platform sources is crucial for the development of effective mitigation strategies.

2 Related work

Conspiracy theories attempt to explain major political or historical events with claims of covert schemes by influential individuals or groups (Aaronovitch 2010; Byford 2011; Dentith and Orr 2018). Prior research on belief in conspiracy theories has come from various disciplines, including history, sociology, and psychology. This previous research has investigated both which individuals believe in conspiracy theories and why, and what real-world effects (if any) these beliefs can have.

2.1 Belief and spread of conspiracy theories

Conspiratorial belief cuts across various demographic groups - no group is above conspiracy theories (Brotherton 2015). For example, most Americans believe that Lee Harvey Oswald did not act alone when assassinating John F. Kennedy. This means that over half of Americans believe in at least one CT (Douglas et al. 2017). Individuals who believe in one theory tend to also place confidence in several others at the same time, even if they are unrelated or even incompatible (Douglas et al. 2017; Georgiou et al. 2020; Goertzel 1994; Wood et al. Nov 2012).

According to Douglas et al. (2017), there are several reasons why individuals believe in conspiracy theories, including wanting to understand and feel in control of a situation or maintain a positive image of their group (Douglas et al. 2017). When information about a confusing event is unavailable, incomplete, or still under investigation, belief in a conspiracy theory that helps explain the event is appealing. Additionally, individuals are substantially more likely to believe an unproven claim about a political opponent over an unproven claim about someone in their own political or identity group (Douglas et al. 2017).

Successfully communicated conspiracy theories with a large following can be modeled by Roger's Diffusion of Innovations model (Radu and Schultz 2017; Rogers 2010). The Diffusion of Innovations model, developed in 1962, seeks to explain how ideas and new technologies can spread throughout a population (Rogers 2010). It is a foundational model in the field of communications (Radu and Schultz 2017) that has been applied to many disciplines and has recently been used in social media applications, including modeling the spread and adoption of new hashtags on Twitter (Chang 2010).

In our study, we use this model as the theoretical basis for how conspiracy theories spread on social media. When applying this model to the spread of conspiracy theories, the innovator in the model is the conspiracy "entrepreneur", or originator of the conspiracy theory. The CT is then distributed by journalistic sources and either adopted or rejected by individuals and groups. Once a CT reaches critical mass, this can lead to socially dangerous behaviors, such as the storming of the US Capitol on January 6th, 2021 (Radu and Schultz 2017). The CT is more likely to be picked up if it is straightforward and resonant with an individual's prior beliefs (Bessi et al. 2015).

The role of journalists is to report the news, and in countries with freedom of the press, to investigate major events. Sometimes in journalistic investigations, reporters uncover misconduct or conspiracies; however, in other cases, their airtime of illogical CTs, even to try to disprove them, ends up boosting these theories (Radu and Schultz 2017). Some previous work shows that attempting to fact-check or disprove unsound CTs can backfire because repeating the theory makes it more salient in people's minds (Radu and Schultz 2017). However, more recent studies have found that this effect may be minor and uncommon (Lewandowsky et al 2020). Either way, some journalistic sources likely are knowingly engaging with conspiracy theories as a way to make money or gain power, as fake news sites can generate significant ad money (Radu and Schultz 2017).

Early on in the COVID-19 pandemic, the COVID-related conspiracy theories spread to large audiences online, often without being countered or removed (Evanega et al. 2020). One study of European social media users found that posts from hostile countries like Russia and Iran received more engagement than posts from regular news sources (Rebello et al. 2020). Another example is the viral *Plan-demic* video that claimed that influential people deployed the virus to profit off of a vaccine. This video received 8 million views across multiple platforms (Basu 2020; Newton 2020).

2.2 Impact and detection of conspiracy theories

Conspiratorial beliefs can often lead to negative offline behavior. While many conspiracy theories do not lead to any real-world actions (ex: JFK assassination, Princess Diana's death, etc.), many others intend to provoke in vs. out-group feelings and demonize the "other", which can encourage violence, extremism, and terrorism (Bartlett and Miller 2010; Douglas et al. 2019). More specifically, recent research on COVID-19 conspiracy theories shows that believers took the pandemic less seriously and did not closely follow public health guidelines on social distancing and hand-washing (Imhoff and Lamberty 2020; Oleksy et al. 2020).

Because belief in pandemic conspiracy theories can cause distrust in public health guidelines, detecting these stories and understanding how they spread online is crucial in the fight against mis-/dis-information. Previous studies have focused on narrative structures of CTs and detecting mis-/dis-information, but little work has been done on the network side. One prior study on detection focused on the narrative structures of real conspiracies vs. conspiracy theories, including coronavirus conspiracy theories (Shahsavari et al. 2020; Tangherlini et al. 2020). Another study looked at conspiracy discussion in the Reddit community *r/conspiracy* to better understand how online communities detect and spread new conspiracy theories after dramatic events. However, the researchers noted that a lack of network analysis was a limitation in their work (Samory and Mitra 2018). Many studies also tend to focus on detecting mis-/dis-information more generally rather than specifically looking at conspiracy theories (Aphiwongsophon and Chongstitvatana 2018; Cruickshank and Carley 2020). In addition to the substantial research on the belief, spread, and offline impact of conspiracy theories, more network-related research is needed to understand how CTs form and change online and their impact on the overall social media discussion.

2.3 Research questions

Most prior research on conspiracy theories has focused on why people believe conspiracy theories and the detection and impact of those beliefs. While we find some work on detection and spread on social media, little work exists on where the conspiratorial stories originate on the web, organically within conspiracy communities, or as part of coordinated influence operations. Previous investigations using a combination of OSINT and networks analysis highlight the usefulness of these

techniques in uncovering hidden coordinated connections by successfully exposing coordinated web campaigns by Russia in 2015 Alexander (2015) and coordinated “fake news” efforts against legitimate news outlets in 2017 Head (2017). To the best of our knowledge, this approach has not been applied to investigations of COVID-19 conspiracy theories, which presents a unique opportunity due to the critical role URLs play in spreading conspiracy theory-related content (Moffitt et al. 2021).

By examining the domains and the links between domains that host conspiratorial content, we address this gap in the literature. This approach can help us better determine if these conspiracy theories are spread organically in a bottom-up fashion (Wallsten 2007), or in a top-down fashion to either generate revenue or as a part of an influence campaign (Radu and Schultz 2017).

In this paper, we address the following research questions:

1. Are there identifiable connections between domains found in tweets classified as containing CTs?
2. Are there indicators of coordination between domains connected by Google tracking codes?
3. How do CT domain networks differ from non-CT domain networks?

Understanding how these conspiracy theory stories spread and where they originate on the web is critical to help social media platforms develop improved detection and mitigation tools.

3 Data and methodology

3.1 Tweets and URLs

In a previous study, we developed a text classification model to label COVID-19 related tweets as conspiracy or non-conspiracy theory related (Moffitt et al. 2021). We vectorized tweet text using a tuned BERT-large language model (Devlin et al. 2018) and then processed the representations through a multi-layer perceptron model for binary text classification. The model was trained on 8,700 hand-labeled tweets.

We then applied the CT classifier to a collection of 1,508,765 English language tweets collected between February 2020 and May 2020. We focused on these four months to target and capture tweets containing conspiracy-related content about the origin of COVID-19. Our research group collected the tweets through the Twitter live stream API with a set of COVID-19 related collection terms (see Table 1). We then applied additional search terms to find tweets with a high probability of containing conspiracy theories. Our model classified approximately 55% of the tweets as conspiracy theory and 45% as non-conspiracy theory. Note that only tweets supportive of one or more CTs were labeled as CT. If a tweet was dismissive of a CT or merely reporting on it, it was classified as non-CT.

Table 1 This table provides the list of terms used to develop the COVID-19 related conspiracy tweets data set for this study and a previous study (Moffitt et al. 2021)

Terms
5 G, 2019nCoV, bat, bioweapon, bio-weapon, coronavirus, covid-19, covid19, covid 19, lab, NCoV, wuhanvirus, wuhan virus

We then extracted and expanded all URLs found in our data and used them to create a list of associated internet domain names, referred to throughout the paper as ‘domains,’ found in the conspiracy and non-conspiracy tweets. The resulting list contains 5,484 unique domain names. We trimmed this list and only considered domains shared a minimum of ten times in our dataset. The final list contained 2865 domains (1100 found in CT tweets and 1,765 in non-CT tweets). All unique domain names were then matched, where possible, to an accuracy and bias rating. The research group maintains a vast database that contains fact-checking and bias ratings for thousands of websites.

3.2 Identifying links with google tracking codes

Google provides Google AdSense and Google Analytics to internet domain owners and online content creators. Both services work by providing small snippets of code with tracking numbers linked to unique account holders. Google AdSense uses its codes to give domain owners and content creators a way to earn money by providing space in their online presence that companies bid on to project their advertisements ([How Adsense works](#)). Google Analytics works similarly, but the codes are embedded with blocks of Javascript that help facilitate tracking and analytics. The service provides account holders with information such as how long a visitor spends on specific content on their domain and where visitors traveled to next after visiting the account holder’s domain ([Tracking code overview](#)).

To collect Google tracking codes, we utilized the requests python package combined with our curated list of domains found in CT and non-CT tweets to make requests to the host domain and collect its associated HTML code (Alexander 2015; Chandra and Varanasi 2015). Once a domain’s HTML code was acquired, we then used regular expressions to parse out tracking codes if present. From this process, we found that 74% of our input domains contained Google tracking codes, 74.3% for domains found in CT tweets, and 73.9% for domains in non-CT tweets. Table 2 provides a detailed breakdown of Google tracking codes found during this process.

To better understand the potential underlying networks, we employed two-hop snowball sampling (Goodman 1961) to help uncover yet undiscovered nodes. We initiated a single seed snowball query for each tracking code contained in an input domain by querying other domains that also contain that code. We then added newly discovered domains to our list and conducted a second hop for the new domains if not already present. This process was repeated for every input domain and every tracking code.

Table 2 This table provides domain and tracking code information from domains found in conspiracy theory and non-conspiracy theory tweets

	Domains	w/ Codes	AdSense	Analytics
Conspiracy tweets	1100	817	271	652
Non-conspiracy tweets	1765	1305	388	1127

Table 3 provides a final breakdown of the number of domains and tracking codes used in this study.

3.3 Coordination detection

We applied both qualitative and quantitative approaches to identify possible coordination between connected domains. First, we identified tweets containing URLs shared in 24-hour periods by domains linked through google tracking codes. This process helped provide a rich understanding of the content ecosystems (domains) that exist and the similarities between them.

Our first quantitative approach sought to infer semantic coordination through the text similarity of articles shared by domains connected through tracking codes. This approach has proven useful in previous studies of coordination and disinformation (Ng et al. 2021; Pacheco et al. 2020). We sorted out all of the unique URLs for each domain, extracted and cleaned the text describing the content in the link, and then aggregated the text into a single corpus for the domain. We then converted the text to vectors and used cosine similarity to compare text between domains using scikit-learn packages (Pedregosa et al. 2011). Cosine similarity represents the normalized dot product of two vectors. The cosine similarity values range from 0, the two vectors are most dissimilar, to 1, the two vectors are the same. We derived an average cosine similarity value for each group/component of connected domains by summing the cosine similarity for every possible pairwise comparison and dividing by the number of comparisons. The resulting value represents our approximation of connected domain semantic coordination.

Our second quantitative approach sought to infer coordination between Twitter users in our data that share URLs of connected domains. We employed a state-of-the-art coordinated action framework (Magelinski et al. 2021) and extended it by incorporating our connected domain network as an additional view. We formed users' connections if they shared an action type (URL or domain) within a five-minute sliding window. Using a sliding window vs. static time segments helps ensure that the connection

Table 3 This table provides the final results from our two hop snowball sampling of input domains

	Input domains	Discovered domains	Total domains	Final adSense	Final analytics
Conspiracy tweets	817	3427	4244	285	871
Non-conspiracy tweets	1305	5325	6630	388	1127

weighting is non-trivial, thus lowering our chance of creating links between users that were not coordinating (Magelinski et al. 2021).

3.4 Network analysis

To help answer RQ3 and compare the network of CT domains with the network of non-CT domains, we used the data we had to create networks. We created the following five networks for both conspiracy and non-conspiracy domains (including those found in the snowballed sample):

1. *Domain \times AdSense* - In this network, a Domain and an AdSense code have a binary link connecting them if that domain uses that AdSense code.
2. *Domain \times Analytics* - This network is the same as the *Domain \times AdSense*, just with Analytics codes instead.
3. *Domain \times Domain (shared AdSense)* - This network is generated by folding the *Domain \times AdSense* network. In this network, two domains will have a link if they share an AdSense code. That link is weighted by how many AdSense codes they share.
4. *Domain \times Domain (shared Analytics codes)* - This network is the same as the *Domain \times Domain (shared AdSense)* just with Analytics codes instead.
5. *User \times Domain* - In this network, a link exists if a User ever linked to that Domain in the original Twitter data set. The links are weighted by how many times a User linked to that Domain.

The *Shared AdSense* and *Shared Analytics* networks were used for investigating coordination.

We additionally created these five networks for the “Intersection” dataset. This dataset consists of all domains found in both the CT and non-CT datasets. Note that the CT, non-CT, and intersection datasets are not mutually exclusive. Most of the popular and influential domains on both the CT and non-CT side show up in the intersection dataset. We chose not to analyze the conspiracy-only and non-conspiracy-only datasets as they were small and included mostly domains (many of which were low viewership blogs) that were not linked to very often and did not have factual ratings. Note that users only show up in the intersection networks if they had at least one tweet classified as “conspiracy theory” and at least one tweet classified as “non-conspiracy theory”. While the number of domains is high in the intersection networks, the number of users is much lower than in the CT and non-CT domains.

To compare these networks, we loaded all the networks associated with the CT, non-CT, and intersection datasets in ORA (Altman et al. 2020; Carley 2017). ORA is a software tool for network analysis. We ran the “All Measures by Category” report three times, once each for the CT, non-CT, and intersection set of networks. This report calculates node-level centrality measures for the input networks. The *User \times Domain* network was of particular use for RQ3, as the results informed us which domains were most popular in each of the three datasets.

To better understand the off-platform reach and user traffic patterns of the top connected domains in our conspiracy theory data, we processed our domains through Similarweb, a business intelligence website that monitors and analyzes web traffic through Google Analytic codes and other sources ([Website Traffic](#)). Table 4 provides the metrics gathered through this process for the set of domains listed in the previous paragraph. We found that these domains have non-trivial reach beyond what we have discovered in our data. Visitors to these sites are significantly engaged; we considered page per visit numbers greater than two as significant engagement. Looking at direct traffic percentages, we found that half of the domains listed had direct traffic rates greater than 70%, signaling that most visitors to these domains are likely using them as a main or one of their primary sources of information and navigating directly to the website. Analyzing the traffic from social media platforms, we found that connected domains may each be targeting different social media platforms to drive traffic for the collective group. For example, in the [*thegatewaypundit*, *westernjournal*, *wnd*] group, we see that *thegatewaypundit* receives a majority of its social media traffic from Twitter,

Table 4 This table provides domain impact and traffic metrics for domains of interest in this study from September 2021

Domain	Reach		Traffic source (%)			Traffic from social media (%)			
	Total visits	Pages per visit	Direct	Referrals	Social	Twitter	Facebook	YouTube	Reddit
express.co.uk	109,160,000	2.16	29.22	2.77	9.49	14.77	75.24	1.55	3.63
dailystar.co.uk	35,410,000	1.71	22.68	15.91	23.56	6.98	79.81	0.47	12.3
thegatewaypundit	33,570,000	2.29	77.29	10.44	4.21	49.69	26.59	13.55	unk*
westernjournal	22,630,000	1.31	66.99	6.26	19.41	2.46	89.05	2.24	3.98
wnd	3,740,000	2.33	80.17	10.19	5.28	25.96	38.14	6.27	21.08
realclearpolitics	13,180,000	4.00	72.86	4.8	2.38	28.64	21.63	28.84	15.27
thefederalist	5,640,000	1.71	53.57	20.07	11.08	52.81	33.65	0.9	10.61
american-thinker	4,610,000	2.73	71.11	18.03	2.34	23.62	34.2	8.3	20.64
infowars	10,300,000	2.47	82.39	6.31	1.34	29.14	32.8	28.67	4.81
newswars	538,950	2.01	49.11	30.17	4.24	35.96	42.72	17.56	unk*

*<https://www.similarweb.com/> only shows the top 4 referring social media platforms

Total visits is the total number of unique visits to a domain, pages per visit is the average number of sub-pages a user visits once landing on the domain, traffic sources describe how users arrive at the domain (Direct = directly entering a domain into a browser, Referrals = directed from another domain, Social = directed by clicking a link from a social media platform), and Traffic From Social Media provides a break down of social media traffic by major platform

westernjournal from Facebook, and *wnd* has more traffic from Reddit than the other two domains.

When we expanded the connection network through snowball sampling, we found additional connections to domains that we have information for but were not part of our initial collection. Figure 2 provides a view of the Domain by Domain shared tracking codes network where nodes are colored by factual rating. In the figure we see the continued trend that nodes of similar factual ratings are connected most often. When comparing connections between domains with varying factual rating levels, we find that more connections exist between domains with low and mixed factual ratings than with low and high ratings. However, there are three large components in which low, mixed, and high factual rating nodes are connected.

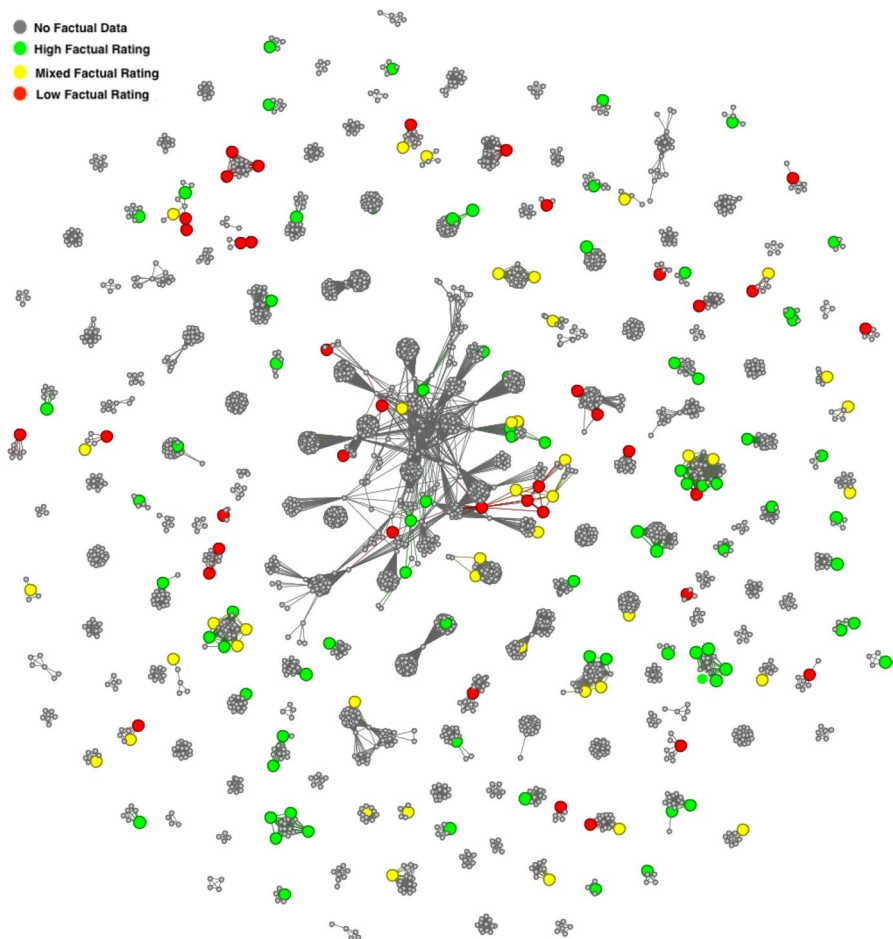


Fig. 2 Illustration of the Domain by Domain shared tracking code (Analytics & AdSense) network formed from the snowball sample of domains discovered in tweets labeled as conspiracy. Nodes are colored by factual rating

4.2 RQ2: Are there indicators of coordination between domains connected by Google tracking codes?

We were able to find several instances of potential coordination between connected domains. In general, domains with a low to mixed factual rating were more likely to show signs of coordination than domains with a high factual rating and high editorial review. Figure 3 provides an example of one possible coordination effort amongst related domains. There are numerous examples of tweets containing URLs from *infowars.com* and *newswars.com*, two domains connected by tracking codes, containing links to articles with the same title. For example, on April 12, 2020, both sites hosted an article titled ‘*global bombshell china admits to developing testing covid19 in wuhan level-4 lab*’. We found similar trends between domains connected by tracking codes such as [*israel365news.com*, *breakingisraelnews.com*], [*latin-times.com*, *sciencetimes.com*, *ibitimes.com*, *newsweek.com*], and [*express.co.uk*, *dailystar.co.uk*].

Using quantitative text comparison techniques allowed us to speed up and scale the process of identifying possible indicators of coordination between connected domains. Table 5 provides the top *average cosine similarity* scores for connected domains. The cosine similarity scores are calculated based on the text in the URL, since most URLs include the title of the article. In this way, we are comparing the headlines shared by connected domains. The mean similarity score for all connected domains is 0.383, the upper quartile score is 0.587, and the standard deviation is 0.27. We found similar results to our qualitative approach; connected domains with low to mixed factual ratings tended to have higher *average cosine similarity scores* than connected domains with high factual ratings. The one exception in our table that we found is [*smh.com.au*, *theage.com.au*]. We added a [*newyorker.com*, *vanityfair.com*] to the list to highlight an example of connected domains that we would not expect to see coordinated behavior from. Condé Nast is the parent company for both of these domains; the content distributed by these domains typically caters to

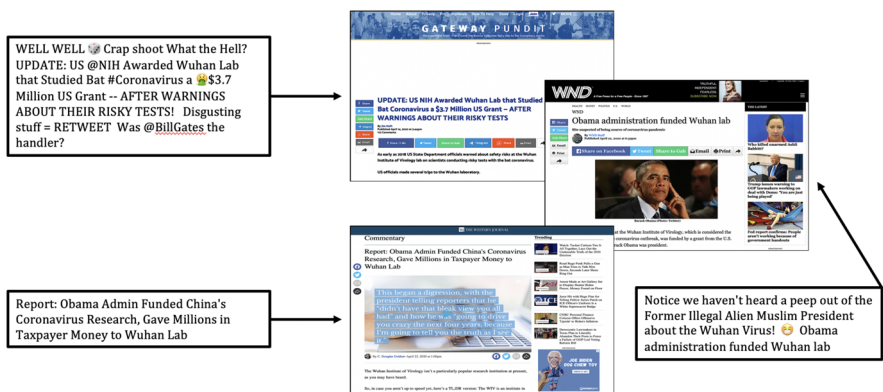


Fig. 3 An illustrative example of tweets from 24 April 2020 containing URLs to three connected domains pushing a common narrative

Table 5 This table provides the top average cosine similarity scores for domains found in conspiracy tweets connected by tracking codes.

Connected domains	Avg cosine similarity
infowars.com, newswars.com	0.93
ibtimes.com, sciencetimes.com, latintimes.com, newsweek.com	0.89
breakingisraelnews.com, israel365news.com	0.78
smh.com.au, theage.com.au	0.75
express.co.uk, dailystar.co.uk	0.69
thegatewaypundit.com, wnd.com, westernjournal.com	0.65
waynedupree.com, lifezette.com	0.65
*newyorker.com, vanityfair.com	0.32

newyorker.com, *vanityfair.com* were added for reference to demonstrate scores we would expect to see if connected domains were publishing markedly different content

different audiences, and we expect article titles to vary as evidenced by the low *average cosine similarity* score.

Next, we compared the *cosine similarity* pairwise between each group of connected domains resulting in 465 comparisons. Table 6 provides the top pairwise comparisons in terms of *cosine similarity*. The mean *cosine similarity* score for all comparisons is 0.349, the upper quartile score is 0.483, and the standard deviation is 0.178. The top pairwise comparisons are significant in terms of similarity as most comparisons have minimal similarity. While not officially connected via tracking codes, these results could further aid the analyst in defining the ecosystem that people who believe in conspiracy theories and disinformation use to consume ‘news.’

Table 6 This table provides the top cosine similarity scores comparing clusters of connected domains

Connected domains	Connected domains	Cosine similarity
thegatewaypundit.com wnd.com westernjournal.com	infowars.com newswars.com	0.82
express.co.uk dailystar.co.uk	independent.co.uk standard.co.uk	0.78
chemistryworld.com asiatimes.com malaymail.com newspunch.com rebelnews.com	infowars.com newswars.com	0.75
express.co.uk dailystar.co.uk	mirror.co.uk birminghammail.co.uk	0.75
ibtimes.com sciencetimes.com latintimes.com newsweek.com	infowars.com newswars.com	0.74

Using the coordinated action framework, we set a five-minute sliding window to identify users in our data who share similar URLs from connected domains. Once identified, edges were formed connecting the users. We used this data to form a User by User shared coordination action network. We added the newly formed network as an additional layer to our existing Domain by Domain shared tracking code network; see Fig. 4. We found that most of the possible coordination centers around the [*thegatewaypundit*, *westernjournal*, *wnd*] and [*infowars.com* and *newswars.com*] clusters of connected domains. We also found ten components of connected domains that do not exhibit any evidence of coordination.

4.3 RQ3: How do conspiracy domain networks differ from non-conspiracy domain networks

We split up the domains into those present in conspiracy theory-labeled tweets and those present in the non-conspiracy theory-labeled tweets. Table 7 shows the number of domains, tracking codes, and users in the intersection of these two data sets. As might be expected, a small but sizeable minority of domains were present in both data sets. Additionally, a small number of users were in both data sets. This means that some of their tweets discussed conspiracy theories, but others of their tweets did not. These non-conspiracy theory-related tweets were sometimes factual tweets, but

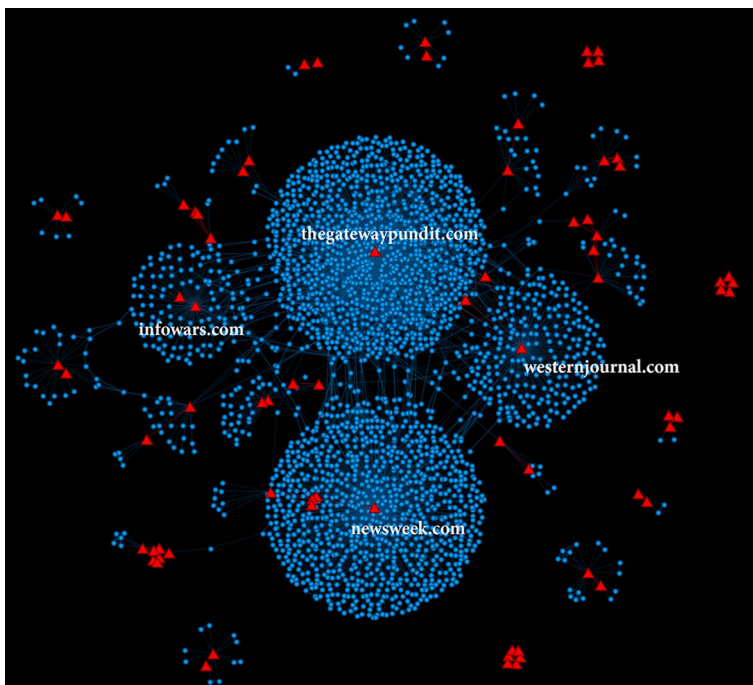


Fig. 4 Conspiracy Coordination Meta Network with Domain by Domain shared tracking code, User by Domain, and User by User shared coordination layers

Table 7 This table provides domain and tracking code information from domains found in conspiracy theory, non-conspiracy theory, and intersection tweets

	Domains	AdSense	Analytics	Users
CT Tweets	4244	285	871	133,693
Non-CT Tweets	6630	388	1127	147,617
Intersection	2481	150	479	5,844

many also discussed mis-/dis-information or were dismissive of certain conspiracy theories. In our analysis, we compared the top domains, domain attributes, and the network structure of the User x Domain network in the conspiracy, non-conspiracy, and intersection data sets.

4.3.1 Top Domains

First, Table 8 shows the top ten domains in the User x Domain network in both the CT, non-CT, and intersection data sets. Note that the intersection table shows the top 11 because one of the top 10 is a blog that is unrated. The top ten is based on two metrics: normalized in-degree centrality (how many users are linking into the domain) and normalized authority centrality (the extent to which a domain's in-links are from users that link to many other domains). A domain acts as an authority when they are linked to by many users, each of whom also links a large number of other domains.

The majority of the top ten domains in the conspiracy theory tweets are websites that range from very low to mixed accuracy. Several are listed as completely “fake news”, like *www.americanthinker.com* and *www.thegatewaypundit.com*. Only two domains are listed as having high accuracy: The Washington Times and The Washington Post. Even if individuals are discussing low credibility conspiracy theories, it is not unusual for them to link to legitimate news sources to debunk or dismiss those real news sources.

In contrast, most of the top ten domains in the non-conspiracy theory tweets are websites rated as having “high” accuracy. Only two are “mixed” rather than “high”: CNN and The New York Post. It is reasonable that individuals not discussing conspiracy theories would tend to link to more credible news sources on average. As one may expect, the average accuracy of the intersection domains is in the middle between the CT and non-CT tweets.

4.3.2 Domain attributes

We then looked at the domain attributes in the various data sets. We analyzed the ratings of the domains in the conspiracy, non-conspiracy, and intersection data sets. Most domains were unrated; they were random blog posts, unknown domains linked to known domains through Google codes, domains in other languages, etc. Table 9 shows the percent of rated domains in each rating ranging from very low to very high accuracy. The non-conspiracy data set has the lowest fraction of its domains in the very low/low/mixed ratings and the highest fraction in the high/very high ratings.

Table 8 This table lists the top domains, found in both the conspiracy and non-conspiracy User x Domain networks

Domains	Network type	Bias	Factual rating	In degree	Authority
americanthinker.com	Conspiracy	fake news	very low	1,667	0.19
thegatewaypundit.com	Conspiracy	fake news	very low	31,944	1.22e-4
dailymail.co.uk	Conspiracy	fake news	very low	23,655	3.73e-5
foxnews.com	Conspiracy	right	mixed	17,999	3.30e-5
newsweek.com	Conspiracy	left	mixed	13,846	9.91e-6
breitbart.com	Conspiracy	fake news	very low	10,029	2.21e-5
nypost.com	Conspiracy	right-center	mixed	6,485	1.42e-5
washingtontimes.com	Conspiracy	right-center	high	6,424	9.04e-6
washingtonpost.com	Conspiracy	left-center	high	5,342	2.90e-6
dailycaller.com	Conspiracy	right	mixed	4,491	1.33e-5
newsweek.com	Intersection	left	mixed	430	0.06
dailymail.co.uk	Intersection	fake news	very low	2532	6.04e-5
ponderwall.com	Intersection	–	–		0.027
nypost.com	Intersection	right-center	mixed	1404	5.28e-5
breitbart.com	Intersection	fake news	very low	1142	2.88e-5
axios.com	Intersection	left-center	high	1071	6.58e-6
theguardian.com	Intersection	left-center	high	811	3.38e-5
thegatewaypundit.com	Intersection	fake news	very low	681	1.18e-5
cnn.com	Intersection	left	mixed	667	3.95e-5
dailycaller.com	Intersection	right	mixed	596	1.17e-5
foxnews.com	Intersection	right	mixed	430	1.67e-5
cnn.com	Non-conspiracy	left	mixed	14,014	0.174
theguardian.com	Non-conspiracy	left-center	high	7,089	0.044
nypost.com	Non-conspiracy	right-center	mixed	10,036	0.011
telegraph.co.uk	Non-conspiracy	right	high	8,647	0.018
latimes.com	Non-conspiracy	left-center	high	8,095	0.016
axios.com	Non-conspiracy	left-center	high	6,117	0.022
smh.com.au	Non-conspiracy	left-center	high	5,739	0.020
nytimes.com	Non-conspiracy	left-center	high	4,651	0.020
nationalgeographic.com	Non-conspiracy	pro-science	high	4,908	0.017
reuters.com	Non-conspiracy	center	very high	4,780	0.015

The in-degree and authority metrics are unscaled

The conspiracy dataset is reversed. It has the highest fraction of its domains in the very low/low/mixed ratings, and lowest in high/very high. As one might expect, the intersection data is in the middle.

Table 9 additionally shows the fraction of users that are classified as a bot or not at the 0.75 bot probability threshold in each of the three datasets. The fraction of bots was roughly identical in the conspiracy and non-conspiracy data sets, at 31%. However, users in the intersection, who posted at least one tweet classified as conspiracy and one tweet as non-conspiracy, were substantially more likely to be bots

Table 9 This table provides the fraction of rated domains in each factual rating category. It also shows the percent of users in each network that are classified as bots

	Conspiracy (%)	Intersection (%)	Non-conspiracy (%)
Ratings			
Very low	12.95	10.79	5.04
Low	8.56	6.12	1.68
Mixed	26.41	26.24	21.08
High	45.48	48.98	64.18
Very High	6.60	7.87	8.02
Bot Prediction			
Bot	31	55	31
Not Bot	68	44	68

- at 55%. There are substantially fewer users in the intersection than in either the conspiracy theory or non-conspiracy theory group. The users in the intersection mean those users posted tweets that were classified as both conspiracy theory and not. Perhaps some bots just retweet or spread content in general, no matter what side it is on.

4.3.3 Domain traffic

We then compared the traffic patterns of the top ten domains in the conspiracy, non-conspiracy, and intersection data sets. These top domains are shown in Table 8. We used the same business intelligence website, Similarweb, as in our RQ2 analysis ([Website Traffic](#)). The averages of the main traffic metrics for each of the three groups of domains is shown in Table 10.

We see that on average the conspiracy-theory linked domains had lower traffic than the non-conspiracy domains. The intersection domains had the highest, likely because both users promoting and users debunking conspiracy theories were linking to those domains.

We also found that the conspiracy theory domains were engaging their viewers more. They had a higher average number of pages visited and a higher average visit duration than the intersection and the non-conspiracy theory domains. We also see

Table 10 This table provides the average monthly traffic metrics of the top ten domains in each data set using data from September-October 2021

Statistics	Conspiracy	Intersection	Non-Conspiracy
Avg total vsits	94.3M	165.8M	147.3M
Avg pages per visit	2.37	2.18	2.04
Avg visit duration	3:55	3:09	2:43
Direct traffic	54.22%	53.00%	42.79%
Referral traffic	9.24%	7.55%	5.10%
Search traffic	24.99%	27.51%	40.90%
Social traffic	9.86%	10.24%	9.66%

that the percentage of traffic coming directly (users directly inputting the domain into their browser) and through referrals from other websites is higher for both the conspiracy and intersection domains than for the non-conspiracy theory domains. Non-conspiracy theory domains got substantially more traffic through search engines. Search engines may be prioritizing more reliable sources.

Finally, we see that social media traffic accounts for roughly 10% of all traffic to these domains in all three data sets. This is a small but meaningful percentage, and it gives us an idea of how the spread of misinformation on Twitter can drive some of the traffic on these websites.

4.3.4 User x domain network structure

We then compared the network structures between the different data sets. Table 11 shows the density and component information in the User x Domain network in the conspiracy, non-conspiracy, and intersection data sets. In the User x Domain network, a link exists between a user and a domain if the user linked to a page on that domain in a tweet. The links are weighted. If a user has multiple tweets that link to the same domain, the link value increases.

We found that while there were fewer components of size greater than three in the conspiracy network, on average these components were substantially larger in size than those in the non-conspiracy network. In the conspiracy network, only 0.47% of components were of size larger than three, and the mean component size was 8404 for those in larger components. In the non-conspiracy network, 2.41% of the components were of size larger than three, but the mean component size was much lower at 1114. Additionally, the density of the User x Domain network is higher in the conspiracy and intersection networks than in the non-conspiracy network. This information indicates that the conversation is more widely dispersed in the non-conspiracy theory network compared with the conspiracy theory network.

Figures 5, 6, and 7 show a visualization of the User x Domain networks in the CT, non-CT, and intersection data sets. These visualizations show how the User x Domain network in the conspiracy data set is more condensed and organized around just a handful of domains when compared with the more widely dispersed network in the non-conspiracy data set. The green means those nodes were predicted not to

Table 11 This table provides density and component information for the User x Domain networks in the conspiracy, non-conspiracy, and intersection data sets

Statistics	Conspiracy	Intersection	Non-Conspiracy
Network density	3.80e ⁻⁴	4.56e ⁻⁴	1.98e ⁻⁴
Isolates	97.87%	91.14%	93.62%
Dyads	1.29%	4.67%	2.90%
Triads	0.38%	1.38%	1.06%
Larger components	0.47%	2.81%	2.42%
Mean size	8404	91.8	1144

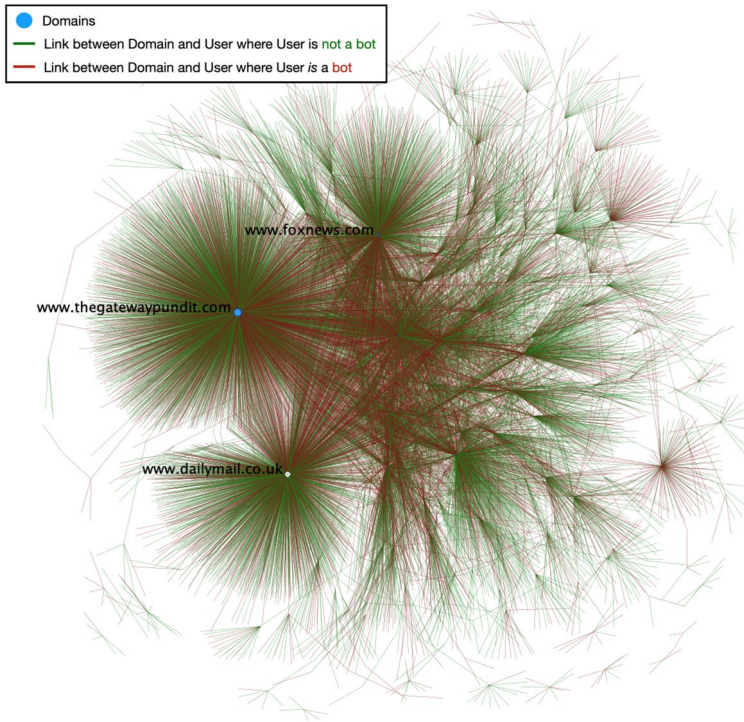


Fig. 5 The User \times Domain network in the conspiracy theory data set. The blue domain nodes are sized by degree centrality

be a bot, while the red means they were predicted to be a bot at the 0.75 bot probability thresholds.

5 Discussion

We investigated the network of domains found in tweets labeled as “conspiracy theory,” those labeled “not conspiracy theory,” and the intersection of the two. We discovered many connections through shared AdSense and Analytics codes in the CT and non-CT data sets. In this dataset, connected domains tended to share similar factual ratings and media bias. There were also many instances of domains with mixed factual ratings connected to domains with high factual ratings, and of domains with mixed factual ratings connected to domains with low factual ratings. Connections between domains of varying factual ratings may serve as opportunities for conspiracy theories to transition from alternative news sources to more mainstream news sources.

We next analyzed whether there was evidence of coordination among domains that shared Google tracking codes using a variety of techniques. We were

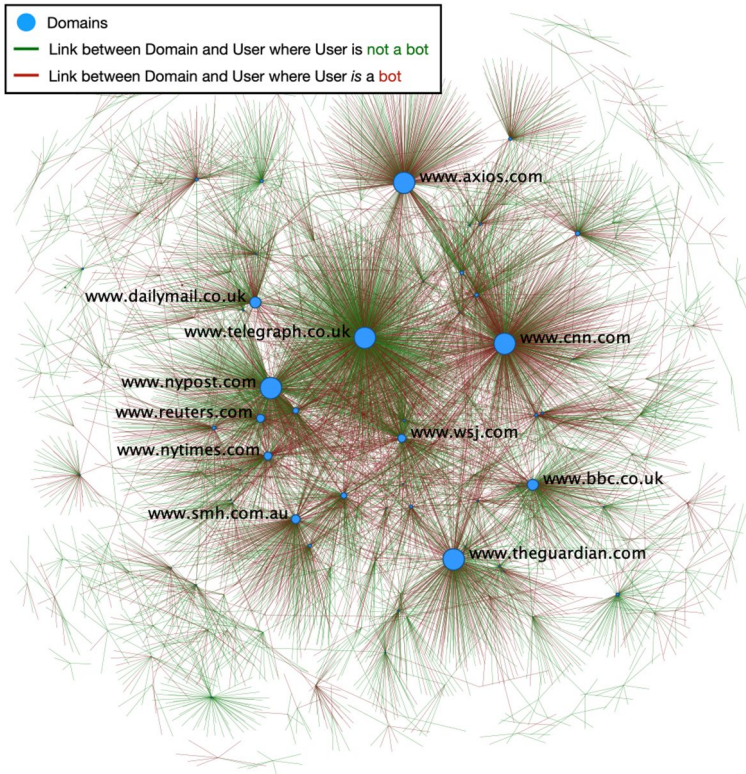


Fig. 6 The User \times Domain network in the non-conspiracy theory data set. The blue domain nodes are sized by degree centrality

qualitatively able to find some evidence of possible coordination among a few connected CT domains. For example, several URLs linking to both *infowars.com* and *newswars.com*, which share tracking codes, linked to articles with similar or even the same headline. We quantitatively compared domains in the same group using the average cosine similarity score of all of their articles. Connected domains with low or mixed factual ratings tended to have higher average cosine similarity scores within their connected groups than connected domains with high factual ratings. Finally, we used the coordinated action framework (Magelinski et al. 2021), which identifies coordinated URL sharing within a five-minute sliding window, and found evidence of possible coordination in the same groups of CT domains identified by the other techniques. Overall, there appears to be consistent evidence of coordination in the connected domains in the conspiracy theory dataset.

We found more evidence of possible organization in conspiracy theory domains when we compared the structure of the User \times Domain networks in the CT and non-CT datasets. We found that users tended to connect to the same domains over and over again in the CT data set. The non-CT data set had more components of size greater than three, which indicates that users were engaging in a variety of

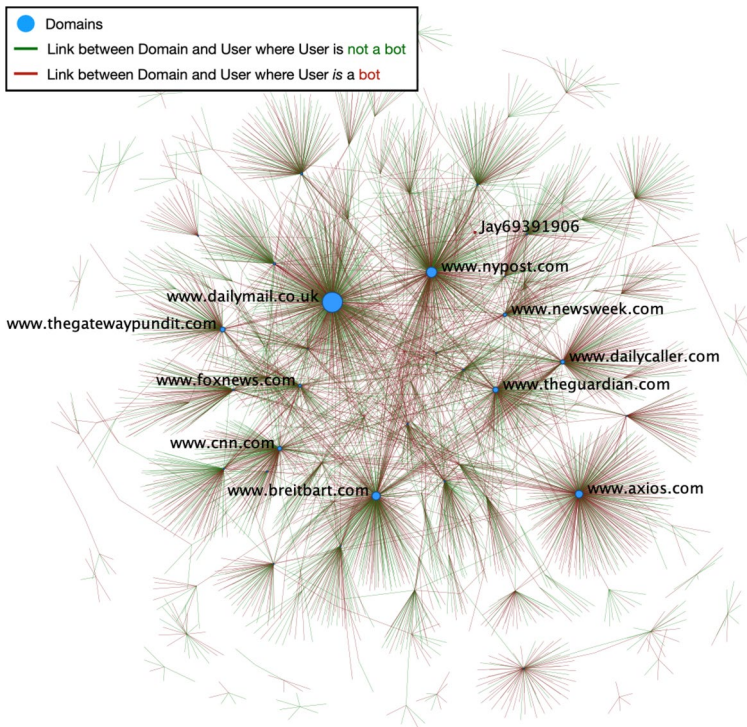


Fig. 7 The User \times Domain network in the intersect data set. The blue domain nodes are sized by degree centrality

sources and were more fragmented when compared to users promoting CTs. Previous research on Twitter conversations surrounding #ReOpen campaigns in the spring of 2020 and climate change show a similar pattern where the side that is forcing controversy (pro-reopening (Babcock and Carley 2021) and anti-climate change (Tyagi et al. 2020)) is more well-organized and well-connected.

Finally, we compared the domain attributes between the CT and non-CT URLs and found stark differences in user engagement. We found that on average, domains found in the intersection and non-conspiracy theory data sets had higher traffic, and a noticeably higher percentage of their traffic came from search engines. However, those in the conspiracy theory data set had a higher average pages per visit, average visit duration, and percentage of direct traffic. Looking at the origin of each domain's traffic, it became clear that some of the groups of connected conspiracy theory domains may be targeting different social media platforms as a way to expand their reach. This type of activity may be indicative of information operations, and coordinated inauthentic behavior as identified by Facebook in a recent report (Gleicher et al. 2021).

Our results show that there is both qualitative and quantitative evidence of possible coordination between conspiracy theory-pushing web domains in this

dataset. Additionally, the domains linked to conspiracy theories are highly engaging their viewers, even if their current average traffic tends to be lower than domains not linked to conspiracy theories.

6 Limitations and future work

Our data for this study were collected from February to May of 2020, and therefore primarily focuses on conspiracy theories that were circulating at the beginning of the pandemic about the origins of the virus. We focused on the beginning of the pandemic because in the early stages of the pandemic there was less public information known, and lack of information is known to foster fear and conspiracy theories (Douglas et al. 2017). Additionally, these conspiracy theories were almost all about the origins of the virus and were therefore comparable with each other. Future work could consider using later COVID-19 data to see if the patterns we found hold for different types of conspiracy theories. Later data may show more conspiracy theories related to the vaccines, and it is possible the domains and types of coordination involved in those conspiracy theories may differ in some way. Additionally, future work could focus on conspiracy theories not related to COVID-19.

For this study, we used the limited domain attribute data we had available to us. The research group has a vast database that collects fact-checking ratings for thousands of websites. However, when we snowball sampled domains, we ended up having a dataset where the vast majority of domains were unrated in terms of accuracy. Updating this data is a daunting task that would be impossible to do manually, as new websites and blogs pop up all the time. Access to complete and accurate media bias and ratings is a limitation in this field. However, we noticed many of the rated domains that share codes tended to be of the same or similar accuracy rating. Future work could investigate the potential accuracy and usefulness of propagating an average rating from rated domains to unrated domains that share a code. This work could lead to a much more robust domain rating database.

7 Conclusion

In this work, we have highlighted the utility of using Google tracking codes to uncover connections between domains and aid in the study of mis-/dis-information on social media. We found that there were identifiable connections between domains found in tweets classified as conspiracy theory, and that there was evidence of coordination based on shared Google tracking codes. Websites of similar media bias and factual ratings are more often connected together than different levels of media bias and factual ratings. There exists an ecosystem of alternative news sites connected by both google tracking codes and by regular user traffic (identified by tracking codes). This system appears to replace traditional news sources for a large number of users.

We additionally found noticeable differences in the accuracy, the network structure, and the web traffic patterns of the top domains in the conspiracy theory vs. non-conspiracy theory tweets. The domains in the conspiracy theory-labelled tweets

were on average of lower factual rating. The condensed network structure of the User x Domain network of the conspiracy theory domains indicates a higher level of coordination and echo chamber-like behavior. Finally, the web traffic data shows that while the conspiracy domains on average have less traffic, their audience is more engaged and spends more time on those websites.

This work shows that conspiratorial domains are designed to keep people on the page and highly engaged. They are often connected via Google tracking codes to other conspiratorial domains that sometimes target a different social media platform to attain a larger audience. Similar article titles and content are used, and users on social media often post links to the CT URL domains within five minutes of each other, giving further evidence of coordination. While currently, these domains get less web traffic than mainstream news sources, this level of high engagement and this evidence of intentional coordination is worrisome as it is likely contributing to the proliferation of COVID-19 conspiracy theories.

Acknowledgements The research for this paper was supported in part by the Office of Naval Research (ONR) under Grant N00014182106, the Knight Foundation, the United States Army, and by the center for Informed Democracy and Social-cybersecurity (IDeaS). The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, the ONR, the United States Army, or the US Government

Funding Open Access funding provided by Carnegie Mellon University.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aaronovitch D (2010) *Voodoo histories: The role of the conspiracy theory in shaping modern history*. Riverhead Books, New York, NY
- Alexander L (2015) Unveiling hidden connections with google analytics ids. <https://www.bellingcat.com/resources/how-tos/2015/07/23/unveiling-hidden-connections-with-google-analytics-ids/>
- Alexander L (2015) Open-source information reveals pro-kremlin web campaign. <https://globalvoices.org/2015/07/13/open-source-information-reveals-pro-kremlin-web-campaign/>
- Altman N, Carley KM, Reminga J (2020) *Ora user's guide 2020*. Carnegie Mellon University, Tech. rep.
- Aphiwongsophon S, Chongstitvatana P (2018) Detecting fake news with machine learning method. *Int Conf Electrical Eng/Electron Comput Telecommun Inform Technol*. <https://doi.org/10.1109/ECTIC on.2018.8620051>
- Babcock M, Carley KM (2021) Operation gridlock: opposite sides, opposite strategies. *J Computat Social Sci*. <https://doi.org/10.1007/s42001-021-00133-9>
- Bartlett J, Miller C (2010) *The power of unreason: conspiracy theories, extremism, and counter-terrorism*. Demos, London, United Kingdom
- Basu T (2020) How to talk to conspiracy theorists-and still be kind (2020), <https://www.technologyreview.com/2020/07/15/1004950/how-to-talk-to-conspiracy-theorists-and-still-be-kind/>

- Bessi A, Petroni F, Del Vicario M, Zollo F, Anagnostopoulos A, Scala A, Caldarelli G, Quattrociocchi W (2015) Viral Misinformation: The Role of Homophily and Polarization. In: Proceedings of the 24th International Conference on World Wide Web. pp. 355–356. WWW '15 Companion, Association for Computing Machinery, New York, NY, USA (2015). <https://doi.org/10.1145/2740908.2745939>
- Brotherton R (2015) Suspicious minds: Why we believe conspiracy theories. Bloomsbury Publishing, London
- Byford J (2011) Conspiracy theories: A critical introduction. Palgrave MacMillan, Basingstoke, United Kingdom
- Carley KM (2017) Ora: a toolkit for dynamic network analysis and visualization. In: Alhajj R, Rokne J (eds) Encyclopedia of social network analysis and mining. Springer. https://doi.org/10.1007/978-1-4614-6170-8_309
- Chandra RV, Varanasi BS (2015) Python requests essentials. Packt Publishing, Birmingham
- Chang HC (2010) A new perspective on Twitter hashtag use: diffusion of innovation theory. *Proceed Am Soci Inform Sci Technol* 471:1–4. <https://doi.org/10.1002/meet.14504701295>
- Cruickshank IJ, Carley KM (2020) Characterizing communities of hashtag usage on twitter during the 2020 covid-19 pandemic by multi-view clustering. *Appl Netw Sci* 5(1):1–40
- Dentith MRX, Orr M (2018) Secrecy and conspiracy. *Episteme* 15(4):433–450. <https://doi.org/10.1017/epi.2017.9> (www.cambridge.org/core/journals/episteme/article/abs/secrecy-and-conspiracy/4A9A8CB9AD858835E532956380E9E0B4)
- Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv preprint. <https://doi.org/10.48550/arXiv.1810.04805>
- Douglas KM, Sutton RM, Cichocka A (2017) The psychology of conspiracy theories. *Cur Dir Psychol Sci* 26:538–542
- Douglas KM, Uscinski JE, Sutton RM, Cichocka A, Nefes T, Ang CS, Deravi F (2019) Understanding conspiracy theories. *Polit Psychol* 40(S1):3–35. <https://doi.org/10.1111/pops.12568>
- Evanega S, Lynas M, Adams J, Smolenyak K, Insights CG (2020) Coronavirus misinformation: quantifying sources and themes in the covid-19 infodemic. *JMIR Preprints* 19:1
- Georgiou N, Delfabbro P, Balzan R (2020) Covid-19-related conspiracy beliefs and their relationship with perceived stress and pre-existing conspiracy beliefs. *Pers Individ Differ* 166:110201
- Glassman M, Kang MJ (2012) Intelligence in the internet age: the emergence and evolution of open source intelligence (osint). *Comput Hum Behav* 28(2):673–682
- Gleicher N, Franklin M, Agranovich D, Nimmo B, Belogolova O, Torrey M (2021) Threat Report: The State of Influence Operations 2017–2020. Facebook
- Goertzel T (1994) Belief in conspiracy theories. *Polit Psychol* 15(4):731–742. <https://doi.org/10.2307/3791630> (www.jstor.org/stable/3791630)
- Goodman LA (1961) Snowball sampling. *Annals of Math Stat* 32:148–170
- How adsense works - adsense help. <https://support.google.com/adsense/answer/6242051?hl=en>
- Head T (2017) Surprise, surprise: Multiple websites pushing 'white monopoly capital' agenda traced back to india. <https://www.thesouthafrican.com/news/surprise-surprise-multiple-websites-pushing-white-monopoly-capital-agenda-traced-back-to-india/>
- Imhoff R, Lamberty P (2020) A bioweapon or a hoax? The link between distinct conspiracy beliefs about the coronavirus disease (COVID-19) outbreak and pandemic behavior. *Soci Psychol Pers Sci* 11(8):1110–1118
- Lewandowsky S, Cook J, Lombardi D (2020) Debunking Handbook 2020. Doi: <https://doi.org/10.17910/B7.1182>
- Magelinski T, Ng LHX, Carley KM (2021) A synchronized action framework for responsible detection of coordination on social media. ArXiv preprint. <https://doi.org/10.48550/arXiv.2105.07454>
- Moffitt J, King C, Carley KM (2021) Hunting conspiracy theories during the covid-19 pandemic. *Soci Media Soci* 7(3). <https://doi.org/10.1177/20563051211043212>
- Netanomics: Netmapper (version 1.0.0.68) [computer software & data, retrieved from <https://netanomics.com>]
- Newton C (2020) How the 'Plandemic' video hoax went viral , <https://www.theverge.com/2020/5/12/21254184/how-plandemic-went-viral-facebook-youtube>
- Ng LHX, Cruickshank I, Carley KM (2021) Coordinating narratives and the capitol riots on parler. ArXiv preprint. <https://doi.org/10.1007/s10588-022-09371-2>
- Oleksy T, Wnuk A, Maison D, Łys A (2020) Content matters. Different predictors and social consequences of general and government-related conspiracy theories on COVID-19. *Pers Individ Diff*. <https://doi.org/10.1016/j.paid.2020.110289>

- Pacheco D, Flammini A, Menczer F (2020) Unveiling coordinated groups behind white helmets disinformation. Companion Proceed of the Web Conf 10(1145/3366424):3385775
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
- Radu RN, Schultz T (2017) Conspiracy Theories and (the) Media (Studies). SSRN Scholarly Paper ID 3089178, Social Science Research Network, Rochester, NY (2017). Doi: <https://doi.org/10.2139/ssrn.3089178>
- Rebello K, Schwieter C, Schliebs M, Joynes-Burgess K, Elswah M, Bright J, Howard PN (2020) Covid-19 news and information from state-backed outlets targeting french, german and spanish-speaking social media users. Project on Computational Propaganda, Oxford
- Rogers EM (2010) Diffusion of Innovations, 4th edn. The Free Press, New York, New York
- Samory M, Mitra T (2018) Conspiracies Online: User Discussions in a Conspiracy Community Following Dramatic Events. Proceedings of the International AAAI Conference on Web and Social Media 12(1) (2018), <https://ojs.aaai.org/index.php/ICWSM/article/view/15039>, number: 1
- Shahsavari S, Holur P, Wang T, Tangherlini TR, Roychowdhury V (2020) Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. *J Comput Social Sci* 3(2):279–317. <https://doi.org/10.1007/s42001-020-00086-5>
- Tracking code overview, <https://developers.google.com/analytics/resources/concepts/gaConceptsTrackingOverview>
- Website traffic - check and analyze any website, <https://www.similarweb.com/>
- Sternisko A, Cichocka A, Cislak A, Bavel JJV (2020) Collective narcissism predicts the belief and dissemination of conspiracy theories during the COVID-19 pandemic. *PsyArXiv*. <https://doi.org/10.31234/osf.io/4c6av>
- Sternisko A, Cichocka A, Van Bavel JJ (2020) The dark side of social movements: social identity, non-conformity, and the lure of conspiracy theories. *Curr Opin Psychol* 35:1–6
- Tangherlini TR, Shahsavari S, Shabbazi B, Ebrahimzadeh E, Roychowdhury V (2020) An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate Pizzagate and storytelling on the web. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0233879>
- Tyagi A, Babcock M, Carley KM, Sicker DC (2020) Polarizing tweets on climate change. In: Thomson R, Bisgin H, Dancy C, Hyder A, Hussain M (eds) Social Cultural and Behavioral Modeling Lecture Notes in Computer Science. Springer International Publishing, Cham, pp 107–117
- UN tackles ‘infodemic’ of misinformation and cybercrime in COVID-19 crisis (March 2020), <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-‘infodemic’-misinformation-and-cyber-crime-covid-19>
- Wallsten K (2007) How Conspiracies Rise, Spread and Fall: The Case of Voter Fraud, the Blogosphere and the 2004 Election. Working paper, UC Berkeley
- Wood MJ, Douglas KM, Sutton RM (2012) Dead and alive: beliefs in contradictory conspiracy theories. *Social Psychol Pers Sci* 3(6):767–773. <https://doi.org/10.1177/1948550611434786>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

J.D. Moffitt is a Societal Computing Ph.D student in the Software and Societal Systems Department at Carnegie Mellon University. At the Naval Postgraduate School, he earned his M.S. in Operations Research. His research interests include applications of Natural Language Processing to identify and understand disinformation and applications of dynamic network analysis and machine learning to identify and mitigate influence operation

Catherine King is a Societal Computing Ph.D. student in the Software and Societal Systems Department at Carnegie Mellon University. At the College of William & Mary, she earned both her M.S. in Computational Operations Research and her B.S. in Mathematics with a minor in Computer Science. Her research focuses on the societal impact of misinformation and polarization, including their impact on elections and public policy

Kathleen M. Carley (Ph.D. Harvard, H.D. University of Zurich) is a Professor of Computer Science in the Software and Societal Systems Department, IEEE Fellow, Director of the Center for Computational Analysis of Social and Organizational Systems (CASOS) and Director of the center for Informed DEMocracy And Social-cybersecurity (IDeaS) at Carnegie Mellon University, and CEO of Netanomics. She is the recipient of the USGA Academic Award at GEOINT 2018 for her work on geo-spatially enabled dynamic network analytics, the Allen Newell award for research excellence, the Lifetime Achievement Award from the Sociology and Computers Section of the ASA (2001), and the Simmel Award for advances in social networks from INSNA (2011). Her research combines cognitive science, sociology, and computer science to address complex social and organizational issues. Her pioneering research led to the areas of computational social science, dynamic network analysis, and social cybersecurity. She has over 400 publications and has served on multiple National Academies panels.