Doctoral Dissertation

Analysis of the COVID-19 Infodemic Using Networks and Machine Learning

XU Wentao

Abstract

During COVID-19 pandemic, an overabundance of information about the coronavirus has been diffused online, making it difficult for the general public to retrieve the legitimate information. This phenomenon is called an “infodemic,” which consists of information plus epidemic. QAnon is a meta-conspiracy theory that emerged during Donald Trump’s presidency, which happened to coincide with the COVID-19 infodemic. In January 2021, QAnon supporters rushed the U.S. Capitol, severely deepening the cleavage of U.S. society. The study of QAnon has used a network approach to analyze the diversity of users and topics. The QAnon meta-conspiracy theory has evolved to become a large conspiracy umbrella during the COVID-19 infodemic. Social bots, a set of computational algorithms, tend to automatically produce or repost misinformation and accelerate the proliferation of online non-credible information sources. The role of bots has been studied in the context of political events, such as the 2016 U.S. presidential election.

In this Ph.D. research, bots that facilitate the diffusion of QAnon misinformation were first identified. Then, the QAnon conspiracy theory umbrella topics were narrowed down to four popular topics, including two conspiracy topics, “5G” and “Bill Gates,” and two misinformation topics, “Trump” and “WHO.” The study found that the bots were segregated in each of the four topics and that the bots’ behaviors were highly correlated to human activity. To be more specific , bots tend to follow humans instead of leading humans in online social networks. We should be alert to the negative role of bots and try to protect less-leaning users from being “infected” by misinformation and conspiracy theories. To maintain the health of the online ecosystem, this research suggested that future work develop more advanced algorithms to identify malicious users.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Computational Social Science

Human society keeps computing. The term “computational social science” (hereafter CSS) comprises two concepts, i.e., “computation” and “social science”. According to a chronicled viewpoint, the computational science of social complexity, one of CSS’s cores, covers the past 12,000-year social history, from ancient human societies of hunter-gatherers to the first socially complex human communities [1]. During the end of the Renaissance and the early Age of Enlightenment, the social sciences began to be applied in naive statistics to explain social phenomena, including decision-making and voting behaviors [2]. For instance, the necessity to make sense of the huge amount of data acquired by population surveys in the nascent European nation states drove the early development of statistics during the Renaissance. The term “statistic” is derived from the Latin word “status” and the Italian word ”stato,” both of which signify a “political state” [3]. The mathematical foundations were greatly enhanced in the Age of Enlightenment by advancements in the theory of probability inspired by gambling. Mathematics and statistics built the methodological foundation, upon which CSS rests.

Later, the invention of computers during the late World War II era and the early Cold War marks the beginning of modern CSS [2]. Computers have played a unique role in the fundamental infrastructure for CSS. It would have been impossible to accelerate CSS to a modern level without computers. Although we are not able to evaluate how much academic research used computers to carry out statistical analysis in the early days, researchers began to use the first business-application computer, Universal Automatic Computer (Univac I) for statistical computations during the early 1950s. Approximately one decade later, the first book of computation-application in social science was published by an American prominent information scientist and psychologist, Harold Borko in 1962 [4].

In addition to the infrastructure role, digital computers have cast fresh light on both traditional and contemporary research topics. Before there were electronic computers, during the 1920s and 1930s, media content analysis became widely popular as a study tool for examining the continually increasing texts, including interview transcripts, clinic conversations, sociological research of narratives, form of films, TV shows, and the editorial and advertising content of newspapers [5].

Harold Lasswell, a political scientist from the United States, was a pioneer in the field of computational content analysis. He elevated the systematic methodology to study mass media [6]. Bernard Berelson, an American behavioral scientist, known for his work in communication and mass media, defined content analysis in his book in 1952 as a "research approach for the objective, systematic, and quantitative description of the manifest content of communication", which demonstrated the early focus on quantitative media text analysis [7].

There was an increasing need for social scientists to use electronic computers to perform statistical data analysis, which was the early term for statistical software. This included SPSS and SAS. SPSS was invented around 1968 and developed by social researchers at the University of Chicago. During the same period, SAS was developed by computational statistics scientists at North Carolina State University.

Computer simulation is another core of CSS. Harold Guetzkow, professor emeritus of political science, psychology and sociology at Northwestern University, developed innovative computer simulation approaches (e.g., InterNation Simulation) to investigate international relations, which are still highly influential today [8–10]. Hayward Alker and Ron Brunner released the first publication on comparative simulation study in 1969 to describe and explain political behaviors, which was a watershed moment in the field [11].

Today, we have entered the era of Big Data. Huge amounts of almost any type of data can be harvested including online shopping, social networking services (SNSs), satellites, smart cars, smart home, administrative data, traffic, public health, geography, marine, space, and industrial sensors. In addition, we have developed much more powerful hardware and service for faster and long-term computations, such as supercomputers, cloud computing, optimized GPU (graphics processing unit, invented by NVIDIA around 1999) for deep learning, and TPU (tensor processing unit, introduced by Google I/O in 2016). Furthermore, worldwide developer communities are maintaining and producing robust machine-leaning tools for CSS. High-performance deep-leaning frameworks (e.g., PyTorch, Tensorflow) have been contributed by related stakeholders.

These developments constitute a healthy CSS research ecosystem. CSS is the study of social systems as information-processing organizations using modern computational systems in an integrated, multidisciplinary manner. However, we have to bear in mind that CSS is not just for Big Data, social complexity, social networks, or social simulations. CSS is not specified in any of these niche areas. Instead, CSS encompasses all of these, as well as other scientific fields, and even new disciplines of the future.

Researchers have been developing diverse methods to study interdisciplinary sciences as well. Gieryn (1983) developed the concept of boundary work in an analysis of the boundaries that separate science from everything else [12]. Fisher (1990) further applied the idea to the boundary-crossing activities involved in interdisciplinary science [13]. This method complements Chubin's conceptualization of knowledge as "core and scatter," in which scatter is the mechanism by which the knowledge domains overlap [14]. Subjects inside the core are equal in importance and interconnected, according to Fisher, but subjects outside the core (scatter) are less integrated knowledge units. Therefore, awareness of multidisciplinary science requires an understanding of the interaction between the core and surrounding domains, which calls for the collection of knowledge of both surrounding and core domains [15]. Since CSS is an interdisciplinary science, it comprises an extensive collection of methodologies, concepts, theories, and even the implementation of programs and simulations. These areas could build on each other and be employed mutually. According to [16], important scientific concepts and the idea of CSS comprise *simulations, human mind, real-world societies, human adaptation, uncertainty, social change, scaling* and *out-of-equilibrium of social distributions.* Here, Cioffi Revilla's framework [2] is introduced. According to his theory, CSS refers to a group of core concepts and research methodologies, including *(1) automated social information extraction, (2) social networks, (3) social complexity* and *(4) social simulation modeling.*

1. *Automated Social Information Extraction*

   The process of parsing and coding materials to extract useful information by computers is known as Automated Social Information Extraction. The primary application of Automated Social Information Extraction is to extract "raw" information from a great amount of data sources. An example would be an investigation to extract information about the background of a target company's profile, including industry analysis, business model analysis, financial strength, management quality, growth analysis, and valuation, through computational content analysis before an unprofessional stock investor makes a final decision.

In 2021, OpenAI has created an artificial intelligence (AI) algorithm that can summarize books of any length. The model, which is a fine-tuned version of GPT-3 [17], works by summarizing sections of a book and those summaries into higher levels, according to a paradigm organized by human feedback with "recursive task decomposition" by OpenAI [18]. The second application is to "visualize" the raw data using a "pivot chart" by constructing a network combined with related information. Imagine that a bank is targeting the money-laundering shell agency by computational content analysis of electronic funds transfer and the corresponding business contracts based on those transfers. For this type of data mining task, we usually need to conduct a knowledge discovery in databases (KDD) [19]. KDD refers to the use of statistical and machine-learning methods to uncover novel relationships in massive relational databases (e.g., Oracle, MySQL, IBM DB2). Recent research also addresses KDD with nonrelational databases (e.g. MongoDB, Redis, Apache HBase) [20].

However, money laundering could be quite complex when transferring funds among tens of thousands of shell accounts, which generates huge networks. To solve this problem, a graph database (eg, Neo4j) has been developed to quickly identify suspicious accounts [21]. The final automated document describes counterparty, time, locations, and text attributes of the contracts associated with suspicious accounts. Ideally, we have a dynamic money-transfer network with preferable attributes at each node. Research of explainable artificial intelligence (XAI) techniques for anti-money laundering has been published to explain non-interpretable models [22]. Other applications include opinion mining and sentiment analysis [23] for customer reviews, named-entity recognition, and extraction of biological relationships in life sciences research by IBM Watson [24].

2. *Social Networks*

   Early sociometric studies can be traced back to the 1950s [25]. In the late 1950s, mathematicians continued to improve graph theory, culminating in the creation of a variety of random network models [26, 27]. These models, in which the edges of the nodes were determined by random processes, have had a tremendous impact on social network analysis because they allow experts to uncover the mechanisms that can cause observed networks to deviate from randomly produced networks [27, 28]. This field has exploded in popularity in recent years, thanks to the rise of social media and Internet sites such as YouTube, Facebook, Twitter, TikTok, Pinterest, Snapchat, etc.

There are two basic forms of graph-theoretic data structures that are used to describe graphs: list and matrix structures. List structures are ideal for storing sparse graph features as less storage space is required. Data sparsity is a widespread issue in social network analysis. This is because there are too many nodes in a huge network, but too few edges to connect these nodes (e.g., Twitter users tend to retweet from influencers other than ordinary ones), because a node's number of relations is restricted in the actual world. In contrast, full matrices are appropriate for representing matrix structures; for example, incidence matrices, adjacency matrices, sociomatrices, Laplacian matrices, and distance matrices [29]. This brings in the computation complexity. The naive multiplication complexity of two square matrices of order n is n^3, although the most recent algorithm deceased it to $n^{2.3728596}$ [30]. As mentioned previously, computers have significantly accelerated the computation of CSS research, particularly modern fast computers, over the last ten to twenty years. As stated above, most huge matrix computations have been taken over by GPUs, which significantly raise the computation efficiency.

3. *Social Complexity*

Sociology adopts complexity principles as a result of its increased focus on complex adaptive systems, and human societies are seen as highly complex, out-of-equilibrium, path-dependent, and self-organizing systems [31].

The studies of social complexity use fundamental principles including non-equilibrium distribution theory, power laws, and information science. The idea behind those principles is that CSS is an interdisciplinary science, and thus, it brings in quantitative concepts and analysis methodologies from thermodynamics.

Pareto distribution is one of the most well-known complexity-theoretic models in CSS [32]. The "80 : 20 rule" is an example of a power-law distribution. J. M. Juran formulated this rule, often known as the Pareto principle, which claims that 20% of causes result in 80% of occurrences [33]. If societies and markets are considered as complex systems, understanding adaptive natures and changing patterns that arise from human behaviors in networks is crucial. This necessitates a greater realism in the behavioral and solid social foundation, upon which social or environmental policies are built.

Furthermore, geography has an affinity for CSS. The association between spatial dynamics and complexity has aided in the integration of geography with CSS and related advances in the physical, biological, and natural sciences, all of which are interested in a number of the same phenomena and processes as geographers [34]. Although there are considerable foundations for complexity-directed social science research, much more of the existing ideas, models, and approaches remain unstudied. In addition, the mix of complexity, related sciences, and simulations enables a hint for future scientific investigation.

4. *Social Simulation Modeling*

Cioffi-Revilla (2021) summarized five types of social simulations including simulation methodology, variable-oriented models, object-oriented models, learning and evolutionary models, and hybrid models [1]. Agent-based modeling (ABM) simulation (belonging to object-oriented models) is one of the important simulation methodologies frequently used by computational social scientists. ABM simulation conceptualizes active components or decision makers as agents, which are modeled and implemented utilizing agent-related ideas and technology. However, even among the small group of computer scientists conducting research on agent models and technologies, the concept of agent remains contentious [35]. Thus, ABM may be defined as a representation of a real-world or reference system imagined as a multi-agent system [36]. A model is usually the theoretic base of a simulation, which is an abstract and simplified depiction of an actual or intended portion of the world. Models are often used to investigate and explain known phenomena, as well as to anticipate future events. Agents and ABMs should not be seen as a technology alone [37, 38], but as a modeling technique that may be used to depict particular system attributes that are not easily stated as properties or functions of individual system components, but rather arise through collective behaviors.

ABM has been addressed to illuminate the link between fundamental behavioral processes, social conditions, and the macroscopic structures of complex networks. Individual aspects interact to produce emergent phenomena. They cannot, by definition, be

> reduced to the system's pieces: the whole is more than the sum of its parts because of the interactions among parts. An emergent phenomenon may have qualities that are independent of the attributes of the part [39]. Broad (2008) gave an example to explain emergent properties [40]. A, B, and C are the characteristics of a system R; one can not simply understand the entire property of object R through the isolated knowledge of A, B, and C, respectively. Usually, new features tend to arise with the interaction of A, B, and C under a certain environment. This knowledge is important to understand complex systems such as social networks.
>
> An example of emergence would be the organs of our body. Can we say we can see the world because the photoreceptor cell in the eyes can see that? The visual process is complex and is facilitated by diverse brain structures and higher cognitive processes. In fact, life is an emergent event. A traffic jam is an emergent event as well. Research has shed light on this problem by examining how ideas of network complexity and emergence theory intertwine. Manley and Cheng (2010) found that, where unlawful maneuvers were detected, there was also a propensity to "follow the leader." When one car is spotted using the hard shoulder to gain an edge over others, a number of others will follow. In many instances, the basic thought of "if they're doing it, I may as well too" was seen [41].

The other important issue that concerns CSS is ethics. One ethical topic could be data privacy. Robust forms of collaboration and data sharing between industry and academia are required to facilitate CSS research, protect customer privacy, and provide firms with liability protection [42]. Appropriate protection of privacy is essential for the academia to reach the latest and most valuable data sustainably. For instance, the General Data Protection Regulation enacted by European Union was adopted in 2016. The regulation allows online users to ask for an explanation of an algorithmic decision that was made about them. On one hand, the EU rule poses significant obstacles for the business sector; on the other hand, it presents an opportunity for scientists to take the lead in building algorithms and evaluation frameworks that minimize discrimination and permit explanation [43]. After all, no off-the-shelf solutions exist for ethical research at the present time. Professional associations, public funders, and private foundations must make an effort to create new guidelines for both the research and industrial areas to guide governments and organizations [44]. CSS has been evolving with both AI technologies and social issues, and now this interdisciplinary subject is facing a global health issue, the COVID-19 pandemic.

1.2 Problems of the Infodemic

COVID-19 broke out in December 2019. Three months later, the outbreak had spread throughout the world, and the World Health Organization (WHO) declared it a pandemic. An overabundance of information about the pandemic emerged and, in addition to credible mainstream news, much of this information consisted of twisted news, rumor, misinformation, and fake news. This misinformation, mixed with legitimate news, confused generic users on social networking services (SNSs) and induced an infodemic. Due to the immense volume of information, the diffusion rate of this mixed news was accelerated by rapid information sharing on SNSs. Biased users tend to share biased information, which is able to affect less-biased users. In addition to human users, social bots (automated controlled accounts) live on SNSs and prey on COVID-19 stakeholders, including influencers, institutional accounts, with the aim of getting exposure, which significantly accelerates the diffusion of misinformation during the pandemic. Although social bots had been studied extensively, their roles in the COVID-19 infodemic is still unknown. In this study, the roles of bots in popular misinformation topics such as “Trump” and “WHO,” and conspiracy theory topics such as “5G” and “Bill Gates” are examined. These four misinformation-related topics associated with a popular American conspiracy theory known as the QAnon movement. QAnon is an emergent conspiracy theory during Trump’s presidency and the COVID-19 pandemic: the conspiracy condemns Democratic elites, accusing them of child sex trafficking, and declaring that political elites in the “deep state” actually control the US. The conspiracy is propagated by Trump’s supporters, who believe that he will save the US from this “political cult.” Although the features of QAnon have been studied [45], the dynamic characteristic of its topic evolution is still unknown.

Above all, the aim of this study is to answer the following research questions (RQs) using computational social science approaches:

RQ1: What kind of conspiracy theories evolved during the COVID-19 infodemic?

RQ2: Did bots play a role in disseminating conspiracy theories as part of the COVID-19 infodemic?

1.3 Proposed Research

To answer these questions, the COVID-19-related Twitter data was collected, and networks and machine-learning techniques were used to analyze the emergent conspiracy theory, QAnon. The first study identified the decreasing features of pro-QAnon users but a large umbrella structure of QAnon topics, which suggests that Twitter's simple removal of malicious users was not enough. The less-leaning users, who leaned less to QAnon, should be better educated to understand the nature of a conspiracy theory; only then will it be possible for them to be pulled toward the anti-QAnon side. The second study then narrows down the QAnon meta conspiracy theory to the fake-news-related topics, such as "WHO," "Trump" and the conspiracy-theory-related topics, "5G" and "Bill Gates" to identify the role of bots during the COVID-19 infodemic. The results demonstrate that the activity of bots is highly correlated with that of humans and that bots have been weaponized for spreading hate speech against Asian people. These results suggest that QAnon has played an important role in the diffusion of misinformation and various conspiracy theories during the COVID-19 infodemic, and that bots have amplified the diffusion of misinformation. Future research could focus on the development of more efficient algorithms for detecting malicious users.

1.4 Outline of the Dissertation

During the COVID-19 infodemic, there is much misinformation and many conspiracy theories circulating online, as listed in [46]. It would cost much time to examine each of them individually. However, it was assumed that there might be a broad and generic conspiracy theory that covers these topics. It was reported that the popular U.S. conspiracy theory umbrella called QAnon has a "complex web of conspiratorial meaning-making narratives" [47]. QAnon's supporters rushed into the U.S. Capitol in January 2021 alerted the assumption that QAnon did not restrict itself in political issues, its supporters might be using this violence to attract more political attention. Then, it was natural to start examining QAnon's dynamics and topics and to find out who were active actors behind the QAnon movement. Furthermore, it was well known that social bots are the major users on social media, and were active in the distribution of misinformation during the 2016 US presidential election [48]. Then, the intuition of this research would be to examine bots and determine their roles in the distribution of misinformation. As mentioned above, it was not possible to examine every single misinformation topic manually. Therefore, we focused on specific cases such as "5G," "Bill Gates," "Trump," and "WHO," all of which were related to QAnon conspiracy theory, to investigate the roles of social bots amid the COVID-19 infodemic. The structure of this research is shown in the diagram in Fig. 1.1

The compendium of this dissertation is as follows. Chapter 2 describes the literature review on related studies. As an outcome of related work and the theoretical background of this dissertation, the following sections have been developed. Chapter 3 presents the first study entitled "QAnon user dynamics and topic diversity during the COVID-19 infodemic," which analyzed the meta conspiracy theory called the QAnon conspiracy theory, including fake news topics, such as "WHO," "Trump," and conspiracy-theory related topics "5G" and "Bill Gates" [49]. Chapter 4 then focuses on these four topics and looks specifically at bot activity, entitled "The roles of bots during the COVID-19 infodemic on Twitter" [50]. Chapter 5 showcases the conclusion and future work.

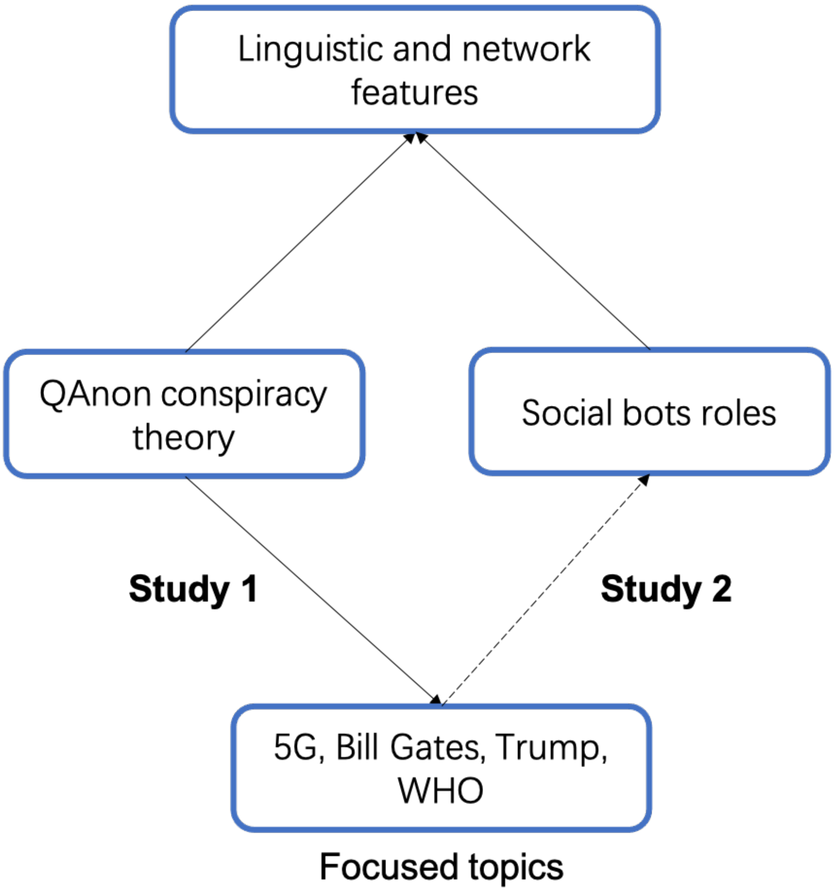


Figure 1.1: The structure of the research. The research started from user dynamics and topic diversity analysis of QAnon conspiracy theory (study 1) and narrowed the candidate topics down to “5G”, “Bill Gates”, “Trump” and “WHO”. These topics were used to study the role of bots during the COVID-19 infodemic (study 2). Finally, the results of the linguistic and network characteristics of both studies were presented.

Chapter 2

Literature Review

2.1 Infodemic

The literal understanding of “Infodemic” consists of “information” plus “epidemic.” Infodemic was defined by Rothkopf in 2003 when SARS broke as “a few facts, mixed, with fear, speculation and rumor, amplified and relayed swiftly worldwide by modern information technologies [that] have affected national and international economies, politics, and even security in ways that are utterly disproportionate with the root realities.”[1] Similarly, the World Health Organization (WHO) declared that the new coronavirus pandemic is accompanied by an “infodemic” of misinformation in February 2022 [51]; one month later WHO announced the epidemic a public health emergency of worldwide concern. WHO defines “An infodemic is an overabundance of information, both online and offline. It includes deliberate attempts to disseminate wrong information to undermine the public health response and advance alternative agendas of groups or individuals.”[2] Infodemic is further considered to be a “second disease” in addition to COVID-19. Asia Center classifies COVID-19 infodemic into four categories, such as “Virus Origin,” “Infections and Deaths,” “Bogus Remedies,” and “Vaccine Efficacy” [52]. “Virus Origin” means the conspiracy theories stating that the origin of the coronavirus was releasing from a Wuhan laboratory. This conspiracy theory originates in the United States, being made for financial gain, or to divert residents’ attention from other political matters.

“Infections and Deaths” indicates unverified or underreported clusters,

[1] https://www.washingtonpost.com/archive/opinions/2003/05/11/when-the-buzz-bites-back/bc8cd84f-cab6-4648-bf58-0277261af6cd/

[2] https://www.who.int/news/item/23-09-2020-managing-the-covid-19-\infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from\-misinformation-and-disinformation

cases, or deaths as a result of the COVID-19 pandemic. “Bogus Remedies” implies unproven viral prevention methods or treatments. The majority of claims include local botanicals, traditional remedies, or even religious or mystical rites. “Vaccine Efficacy” implies misinformation about reactions to various vaccines or their long-term effects. Also included are bogus reports of vaccination given by hospitals and other institutions.

Although experts have discussed the impact of fake news on significant social events such as presidential elections [53, 54] the spread of misinformation has a greater potential to impair public health, particularly during the COVID-19 pandemic. For instance, a series of research found that support for COVID-19 misinformation is strongly related to individuals less inclined to follow public health advice, as well as having less intentions to be vaccinated and to advocate vaccination to others [55–58].

The infodemic ecosystem around COVID-19 is continually evolving, with growing media reports, political repercussions and scientific research, and individual emotions being shared on SNSs. One of the favorite twisted information diffused in infodemic is conspiracy theory. van Prooijen and Douglas (2018) suggested that conspiracy theories are psychological processes, and conspiracy beliefs are extremely responsive to a social environment [59]. The paper further points out that conspiracy theories were conceptualized with four fundamental features; that is, “Consequential,” “Universal,” “Emotional,” and “Social.” It is true that during the COVID-19 pandemic, many cities in the world adopt quarantine, lockdown, and even social distancing to control the pandemic. But these actions induce high-level anxiety [60]. Furthermore, exposure to misinformation is associated with psychological discomfort, including symptoms of anxiety, depression, and post-traumatic stress disorder, in addition to irrational beliefs.

Another important feature of any information is its fast diffusion speed. It is estimated that 211 million users are active on Twitter, and one user has 707 followers on average[3]. Followers can receive updated messages from influencers once influencers tweet anything. Huge mixed information flows on SNSs so fast that the public limits their capacity to verify legitimate information. This feature significantly promotes the propagation of conspiracy theories. In addition, social bots (automated controlled accounts) accelerate the propagation of misinformation and conspiracy theories. It was estimated that 33% of the top shared contents with lower credibility were probably bots [61].

In a recent review [62] of COVID-19 related misinformation topics, the

[3] The number was checked on 30 June 2022. The number can be updated at `https://www.socialpilot.co/blog/social-media-statistics`.

authors summarized several important psychological features of conspiracy theories. The authors stated that the purpose of conspiracy theories is to characterize societal events through the outcome of conspiracies, based on the study of [63]. Furthermore, the authors mentioned that although conspiracy theories are associated with oversimplified and distorted information, conspiracy theories quickly explain puzzling occurrences by attributing their origins to the secret plans of powerful players, according to the research of [63, 64]. Moreover, the review pointed out that certain environmental factors, social incentives, cognitive styles, and tendencies underpin the adherence to views that are not restricted to lunatics and radicals, but belong to common occurrence that crosses political and demographic realms, with significant societal and political effects. This problem impacts a substantial section of people, suggesting the desire to know the society, to connect, to feel cozy, and self-identification in one's social networks. The authors drew the above conclusion based on the study of [65,66]. In addition, the author found the psychological drive for believing conspiracy theories, that is, conspiracy theories lead to consequences: they influence behaviors in a broad area of life such as social relationships, safety, and well-being, as reported by [67]. In addition, most common conspiracy theories promote right-wing extremism, with the objective of highlighting the narratives used to mobilize extreme action and violence and promote a political agenda [68]. The authors further discussed the potential problems of touching misinformation. Right-leaning publications were likely to make false statements about the remedy and origin of COVID-19, and people exposed to much right-leaning news were inclined to express inaccurate opinions, resulting in the tendency of uneducated people to believe that public health professionals overestimated the severity of the pandemic [69]. In this way, conspiracy theories pull innocent people to the pro-conspiracy-theory side, disrupt the society, and entangle people to opposing camps. For example, QAnon supporters rushed into the US Capitol in January 2021.

Although we are still in the middle stage of the COVID-19 infodemic, conspiracy theory during the infodemic has generated chaos. For example, in the years before the pandemic, there had been rumors about nG ($n = 3, 4, 5$) mobile technology (e.g., 3G with "2003 SARS," 4G with "2009 swine flue," 5G with COVID-19). Pertwee et al. (2022) reviewed several popular conspiracy theories [70]. The paper pointed out that the origin of the 5G rumor is associated with COVID-19 in early 2020, after a rumor declared that 5G mobile towers in Wuhan were linked to the outbreak of COVID-19, based on the study of [71]. In addition, people who thought 5G was responsible for the spread of COVID-19 are considered to have vandalized many 5G towers in the UK in early April 2020 [4]. In addition to 5G, anti-vaccination is another popular conspiracy theory. Furthermore, Pertwee et al. (2022) reported that after the readers were exposed to the conspiracy theory that COVID-19 vaccines would change the structure of human DNA and cause infertility, these readers showed a decline in intent to vaccinate. Pertwee et al. (2022) further mentioned that other research has achieved similar results on the impact of online fake news about vaccination, based on the study of [72].

How to fight against the infodemic is a serious problem for all stakeholders in COVID-19 throughout the world. Eysenbach (2020) raised a strategic approach that included four aspects to deal with the infodemic, including providing knowledge to the public in a way they can understand, evaluating information efficiently, building electronic health and science information literacy, and continuous data flow surveillance [73]. WHO established a working flow for daily Internet users to check the legitimacy of online information [5]. These suggestions offer a hint of combating the infodemic at the early stage of the COVID-19 pandemic. However, we have to bear in mind that the pandemic will not end soon. Misinformation and conspiracy theories associated with the COVID-19 pandemic will still be alive online for a considerably long period. Harmful information, which interacts with users, evolves into a global infodemic ecosystem. Stakeholders, including policymakers, politicians, researchers, online users, and SNSs, should work closely to fight against COVID-19 infodemic to keep the online ecosystem healthy.

2.2 Fake News

Fake news is not a new concept and has a long history, dating back centuries to the invention of the earliest writing systems. However, with the advent of social media, the past ten years have witnessed a shift in the way news is disseminated that is quite distinct from traditional media. "Fake news" is usually disguised as legitimate news or interesting stories. Stories are made, told, and sold. Newspaper agencies pay for celebrities, a human interest story, a scandal, or anything else that readers think the good people of the society would want to read about. The interesting stories are well-selling and addicting. A piece of "interesting" and "popular" news usually originates from a story. Of course, being a story does not necessarily mean "fake." Meanwhile, we can not simply consider a professional narrative of

[4] https://www.businessinsider.in/tech/news/77-cell-phone-towers-have-been-set-on-fire-so-far-due-to-a-weird-coronavirus-5g-conspiracy-theory/articleshow/75580457.cms

[5] `https://medium.com/who/lets-flatten-the-infodemic-curve-92e4c840d4bf`

a story by a journalist to be a piece of news. Then what is “news”? Park (1940) defined news as “News, though intimately related to both, is neither history nor politics. It is, nevertheless, the stuff which makes political action, as distinguished from other forms of collective behavior, possible” [74]. This definition suggests that the nature of news serves politics. The common understanding would be “news is what editors say it is” or “News is what sells papers or drives up ratings” and “it is easier to recognize it than define it” [75]. “They know news when they see it” [76]. It seems that news, like a gut feeling [77], has a magical quality that can not be explicitly expressed . A recent research stated ten characteristics of news [78]:

1. The power elite: Stories concerning powerful individuals, organizations, or institutions.
2. Celebrity: Stories concerning people who are already famous.
3. Entertainment: Stories concerning sex, showbusiness, human interest, animals, an unfolding drama, or offering opportunities for humorous treatment, entertaining photographs or witty headlines.
4. Surprise: Stories that have a component of surprise and/or contrast.
5. Bad news: Negatively tinged tales, such as those involving strife or disaster.
6. Good news: Stories with particularly positive overtones, such as rescues and cures.
7. Magnitude: Stories deemed sufficiently substantial, either in terms of the number of individuals involved or their potential effect.
8. Relevance: Readers-relevant narratives about organizations, problems, and countries.
9. Follow-up: Articles about topics that are already in the news.
10. Newspaper agenda: Stories that align with the news and organization’s own agenda.

Sensationalism is an important concept of news. Fierce debates have been going on for centuries about sensationalism in journalism and are still going on [79, 80]. The term “sensationalism” refers to the content and form elements of news articles that are capable of arousing the viewer's senses, namely, prompting attention and arousal reactions in viewers [81]. "Sensationalism" is used today to gain readership, ratings, and to make money" [82]. The definition of the word in the dictionary connects to a negative social and emotional implication. Cambridge Dictionary simply defined it as "a way of telling a story that is intended to shock people."[6] The Oxford Advanced Learner's Dictionary described the word as "a way of getting people's interest by using words that are intended to shock you or by presenting facts and events as worse or more shocking than they really are."[7] It was defined as "the presenting of facts or stories in a way that is intended to produce strong feelings of shock, anger, or excitement." by Collins English dictionary [8]. Attracting readers by arousing emotion is a popular methodology in journalism. From the perspective of journalism, sensationalism is defined as information that amuses, titillates, and entertains, "appropriate" news is praised for its presumed potential to improve the audience's political and social awareness by appealing to reason over emotion [80]. Consequently, sensationalism becomes a helpful technique for catching people's attention by using language designed to shock or by portraying facts and events as harsher or more surprising than they really are. Thus, modern economic-driven journalism is considered to have compelled journalists to use sensationalism in order to capture the audience's attention. When sensationalism rules the media landscape, it is reasonable to assume that fake news will draw the attention of some legitimate media, either to avoid missing a story or to generate revenue [83]. Fake news is mainly made up of sensational and divisive headlines, and its emotional language can help spread far and wide [84]. The belief in fake news, as well as the influence and persuasion it has on the public, can be linked to emotion [85]. Content that elicits strong emotions (both positive and negative), such as joy, excitement, or rage, is more likely to be shared [86–88]. Such emotional fake news is exacerbated by the media, which is often guilty of favoring sensationalist reporting over carefully worded scientific messages with a balanced interpretation. The result is a loss of public confidence and a sense of helplessness, which are ideal conditions for the propagation of harmful misinformation and the beginning of a vicious cycle [89].

A well-informed public is essential to develop "news," city legends, and gossips on SNSs during the "self-media" age. The headline is the primary "hook" for the public in a standard print or online news piece. Unfortu-

[6] `dictionary.cambridge.org`

[7] `www.oxfordlearnersdictionaries.com`

[8] `www.collinsdictionary.com`

nately, the public lacks both the professional skills and the time to undertake thorough investigations. Most people are fed sound bites by news and newsletters. People often take what they read as real, particularly if it comes from a respectable source, and do not examine the information, no matter how startling or scary it is. It is all too easy to ignore the saying, "the weight of evidence for an extraordinary claim must be proportioned to its strangeness" [90]. This remark is important to the scientific process and serves as a paradigm for critical thinking, rational reasoning, and skepticism around the world. Of course, it is not a requirement for an ordinary person to dig the truth like a scientist or a professional journalist after reading the news. To make matters worse, people may have a tendency to repeat the most remarkable information, instead of searching for evidence. But what exactly does evidence mean? Although most scientists would disagree, the notion of "extraordinary evidence" aforementioned in science is more a societal agreement than an objective judgment [91, 92].

Since fake news tends to be sensational and could only be verified by professional journalists, what exactly is fake news? We always hear several analog terms for fake news, including satire, yellow journalism, hoax, propaganda, misinformation, disinformation, and rumor. Sometimes, they are interchangeable for fake news. According to Wikipedia [9], fake news is "false or misleading information presented as news." Zhang and Ghorbani (2020) defined fake news as "all kinds of false stories or news that are mainly published and distributed on the Internet, in order to purposely mislead, befool or lure readers for financial, political or other gains;" and they further proposed a fake news map, in which all the related components were presented [93].

According to the Cambridge Dictionary [10], fake news is defined as "False stories that appear to be news, spread on the Internet or using other media, usually created to influence political views or as a joke." Since SNSs are popular nowadays, they are now more 'used to describe false stories spreading on social media" [94].

The meaning of fake news has evolved over time as it is increasingly used to refer to social media rather than traditional media. SNSs have become fertile ground for computational propaganda and the trolling of fake news. Most online fake news is obtrusive and comes in many different forms, styles, and social platforms [95]. According to [96], fake news is "news content published on the internet that aesthetically resembles actual legitimate mainstream news content, but that is fabricated or extremely inaccurate.

[9] en.wikipedia.org

[10] https://dictionary.cambridge.org/

Also referred to as false, junk, or fabricated news." During the COVID-19 infodemic, several conspiracy theories seem popular online [97]. Conspiracy theorists assert that COVID-19 is a bioweapon or coronavirus that has been leaked from a Chinese laboratory in Wuhan; such stories have been consistently debunked [98]. The vaccine conspiracy declares that the COVID-19 pandemic is orchestrated by Bill Gates and the "deep state" (the same phrase in the QAnon conspiracy theory) to implant microchips and further control the population. The 5G conspiracy claims that the 5G technologies are able to alter people's DNA. The conspiracy theories of miraculous remedies suggest that hydroxychloroquine, Bleach, and colloidal silver can cure the virus [99], although hydroxychloroquine has not been shown to be a successful treatment for COVID-19 [100, 101]. QAnon is a meta conspiracy theory, which contains the above-mentioned conspiracy theories.

In addition to fake news and conspiracy theories, the patterns of fake news and misinformation flow during the COVID-19 infodemic have been examined. Cinelli et al. (2020) discovered that the dissemination of information is influenced by the interaction paradigm established by specific social media platforms and/or by the interaction patterns of the engaged users [102]. Huang and Carley (2020) noted that regular users are more likely than news agencies and governments to disseminate tweets containing fake news URLs and stories [103]. Most of the retweets of misinformation websites' contexts are generated by bots. Brennen (2020) determined that 20% of fake news was shared by celebrities, politicians and recognized figures [104]. Meanwhile, fake news is responsible for 69% of the complete engagement of social media (e.g., comments, likes, and shares). On the contrary, fake news, which was spread by generic users, produced much less engagement. These studies suggest that bots and influential users play crucial roles in the dissemination of fake news during the COVID-19 infodemic. Due to the amplifying effect of bots and/or influential users [105], a small proportion of misinformation produced by generic users becomes important or generates large engagement among users. Regrarding the speed of propagation, the spread of fake news on Twitter is much faster than the true news, especially for the political category [84]. Furthermore, according to a report, released in 2019 that carried out research examining a data set with 171 million tweets during the five months before the election day, 25% of the 30 million tweets that featured a link to news sites diffused fraudulent or strongly biased news [54].

The above-mentioned fake news and its diffusing patterns remind us of a question: How do we detect fake news? The problem is usually divided into social context and news content categories (Fig. 2.1, reproduced from [95]).

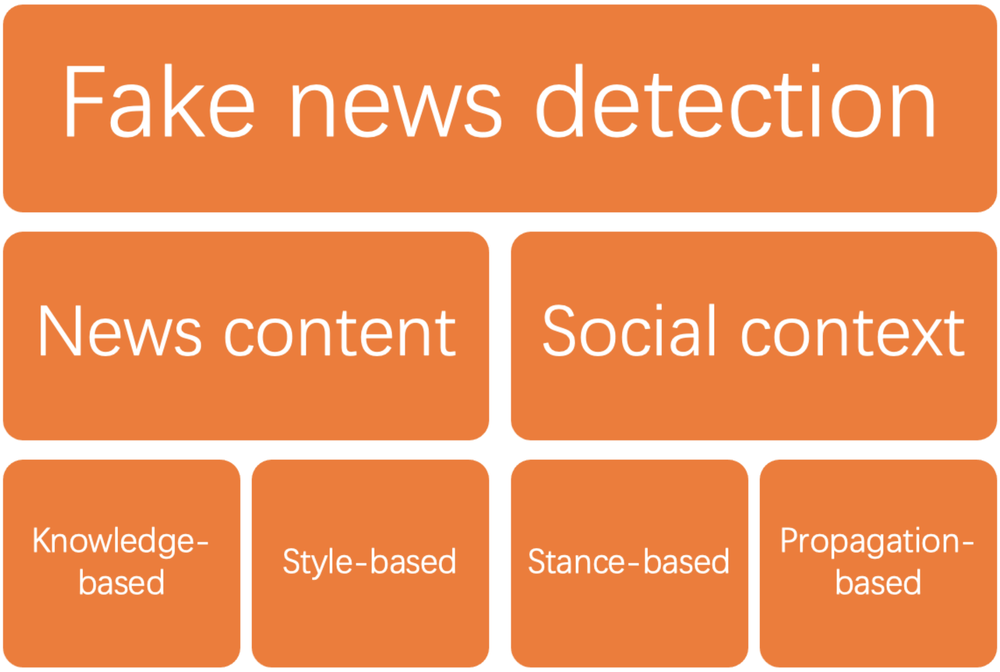


Figure 2.1: Fake news detection pathways (reproduced from [95].)

News content detection comprises knowledge-based and style-based methods. The method of knowledge-based detection is straightforward, which checks the to-be-verified news with ground truth news. This approach includes manual verification, crowdsourcing labeling, and automated models. Experts with professional knowledge, such as journalists, usually take on the task. It is easy to organize and leads to highly accurate results. However, it is expensive and does not ensure scalability when tons of news content need to be checked. The well-known fact-checking websites, including "PolitiFact" and "Snopes", employ experts to prevent the public from fake news. Crowdsourcing relies on a vast population of average annotators participating as fact-checkers. Such a large number of fact-checkers can be found on typical crowdsourcing platforms such as Amazon Mechanical Turk. Compared to experts, crowd workers are more difficult to manage and labeling accuracy could be reduced due to their political bias and insufficient knowledge. However, this method increases the scalability to some extent. To address scalability, automatic fact-checking systems that significantly rely on knowledge graph, machine learning techniques, and natural language processing, as well as network theory, have been developed.

Style-based is the other method in the news content category. Deception-oriented and Objectivity-oriented belong to the category. At the syntax level, part-of-speech (POS) taggers perform shallow syntax tasks to evaluate POS frequencies [106]. Deep syntactic tasks are carried out using Probabilistic Context-Free Grammar (PCFG) parse trees, which allow rewrite of rule frequency analysis [106,107]. Stance models infer the credibility of original news stories based on the opinions expressed by users in relevant post content. In other words, the stance technique extracts the opinions of various users on a news story in order to determine whether it is fraudulent or authentic. The opinions can range from the number of likes a given post has received to the comments made on the piece. The explicit and implicit positions of the users are extracted from their opinions. For example, Jin et al. (2016) established a method for verifying news information based on the analysis of opposing viewpoints [108]. Mohammad et al. (2017) proposed a more refined stance identification technique based on support vector machine (SVM) and input data such as n-grams and word embeddings in another work [109]. The other method contained in the social context belongs to propagation-based analysis. The method consists of news cascade and self-defined graphs. Details of this alternative can be found in [110]. Modern detection methods have used many language models, such as BERT [111] and its derivatives. Innovative information extraction techniques, including the well-known "Transformer" architecture, which employs dot multiplications between "heads" matrix [112]. Most research suggests that a heterogeneous network should incorporate various types of nodes and links into one attention-based architecture network [113].

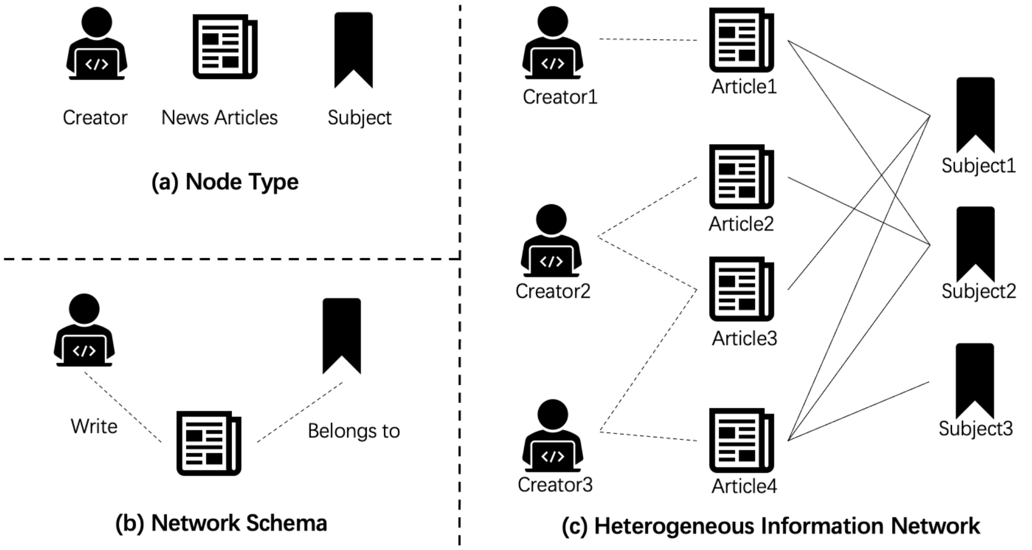


Figure 2.2: An example of a heterogeneous information network. (reproduced from [114]) (a) There are three types of nodes: Creator, News article, and Subject. (b) Network structure. (c) A news production comprises three different types of nodes and two different types of links.

2.3 Bots

Not all communication on social media platforms comes from real people. Bots are referred to as automated accounts that are controlled by a set of algorithms to publish information. Researchers who used agent-based modeling in networks came to the conclusion that in some situations, only 2–4% of bots are sufficient to change the atmosphere of social opinion in social networks and that they can easily "sway public opinion – or the expression thereof" by spiral-of-silence dynamics effects [115].

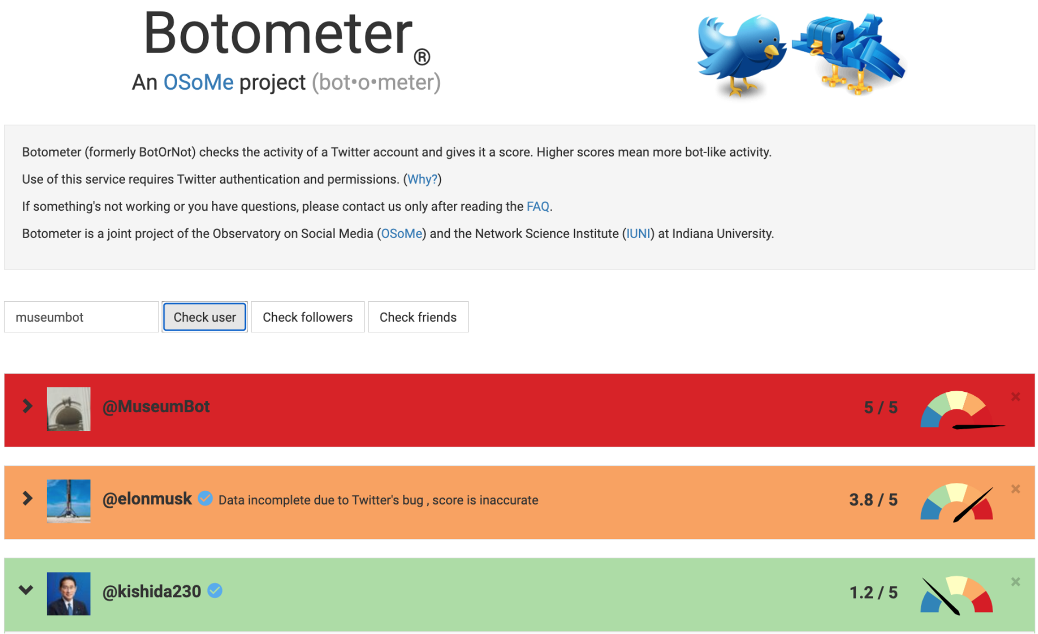
In the late 2000s, early social media bots were developed to handle simple tasks such as automatically retweeting information produced by a set of sources or locating and uploading news from the Web. Bots' capabilities have considerably advanced in recent years: Bots rely on the fast-paced environment. A research paper on the dynamics of bots highlighted the evolution of bots [116]. The paper mentioned that advances in artificial intelligence, particularly in natural language synthesis, take advantage of pre-trained multilingual models such as GPT-2 (invented by OpenAI) [117] to create content that resembles that of a human being. This AI technology enables the development of automated agents or bots that create human like texts on SNS, making it more difficult to differentiate between human and automated accounts [118]. The paper further mentioned that other techniques such as bot-as-a-service (BaaS) significantly decrease the minimum requirement to deploy bots and create huge bot networks: for example, "ChatBots.io" uses BaaS to run bots in cloud computing such as AWS (Amazon Web Services) and Heroku. This kind of cloud computing infrastructure makes it more difficult for bot detection.

Since social bots are becoming more sophisticated, detecting them is not an easy task. For common users, it is even more difficult for them to distinguish between humans and bots [119], because around $9 - 15\%$ of active Twitter users are likely to be bots [120]. Bots' ability to bias discussion about politicians has been identified as well. For example, bots have successful experience in influencing elections and political debates [121, 122]. Social bots are also applied to distract public attention from hot topics, spamming, and exaggerating trends [123, 124]. In addition, social bots may be capable of working synergistically to modify the influence scores of several centrality metrics [125, 126]. Twitter has previously taken precautions against bots in order to reduce the impact of bot destructive acts, and thus targeted users that were actively used to amplify and propagate news from suspicious sources [127].

To reduce the impact of bots, many researchers have addressed bot detection. The majority of bot detection methods rely on supervised learning with manually-labeling data. Rout et al. (2020) proposed a malicious social bot detection system to identify members of trustworthy Twitter networks by combining a trust computation model with URL-based features [128]. Zhang et al. (2020) identified bots and gender in two unique languages (English and Spanish) [129]. For English accounts, AdaBoost was used to solve the bot detection, while an SVM model was used for Spanish accounts. Botornot [130] (Botometer's former name), was proposed for detecting social media bots. Fig. 2.3 shows examples of "@MuseumBot," (probably a bot) "@elonmusk," (probably a bot) and "@kishida230" (probably a human). [11] According to reference [131], Botometer considers six types of features, including dating, user, network, content, time, and emotion. Botometer further extracted more than 1,000 characteristics to determine whether a user was a social bot or an ordinary user. Botometer compared logistic regression, AdaBoost, random forest and decision tree and found that random forest achieved the best classification performance, with a 95% accuracy rate. The Botometer has been used in a series of research to quantify bot online behavior [84, 105]. Botometer computes a "Complete Automation Probability" (CAP) for each user that falls between $[0, 1]$. The greater the value, the more likely the user

[11] https://botometer.osome.iu.edu/

is a bot.



Botometer (formerly BotOrNot) checks the activity of a Twitter account and gives it a score. Higher scores mean more bot-like activity.

Use of this service requires Twitter authentication and permissions. (Why?)

If something's not working or you have questions, please contact us only after reading the FAQ.

Botometer is a joint project of the Observatory on Social Media (OSoMe) and the Network Science Institute (IUNI) at Indiana University.

museumbot

@MuseumBot 5 / 5

@elonmusk Data incomplete due to Twitter's bug , score is inaccurate 3.8 / 5

@kishida230 1.2 / 5

Figure 2.3: Botometer examples (https://botometer.osome.iu.edu/).

Kudugunta and Ferrara (2018) proposed an LSTM-based deep-learning infrastructure that used content and account metadata to detect bots on a tweet-by-tweet basis [132]. This method extracted content features from user profiles as additional input features. In addition, social network structure-based analysis [133], an unsupervised clustering-based detection scheme [134], and a crowd-sourcing-based detection method [135] are also available.

Bots are believed to diffuse conspiracy theories during the COVID-19 infodemic [61,136], but the characteristics of the role of bots is still unknown. With unprecedented levels of communication and the emergent COVID-19 infodemic, it is necessary to identify the features of social networks that help the propagation of misinformation with the help of bots.

2.4 Social networks

A network is a set of connections that represent relationships of objects in the network. More formally, a network contains a set of nodes (in mathematics, vertex) and a mapping or description (edge) between nodes. Networks can be categorized as directed and undirected. In directed networks, each line with an arrow was utilized as an edge to indicate a special direction relationship, whereas in undirected networks, a line without an arrow was used to signify a relation without any directional property between two nodes.

A "sociogram" is an early representation of a social network. Fig. 2.4 shows a "sociogram," raised up by Jacob Moreno (1934) [137], who is considered a key founder of modern network studies. The sociogram is a diagram in the tradition of spatial geometry, in which persons are represented as nodes, and their connections are depicted as lines linking the nodes. The sociogram gives a visual depiction of the investigated social structure and reveals specific characteristics of the interactions that comprise it. When relationships were shown as sociograms, people could almost immediately see what was going on in small networks that were not too complicated. Adding graph theory to the tools for understanding networks made it possible to understand and control networks that were much bigger and more complicated. A triad is a simple group of three people that are related to each other. This simple network turns out to be the basis for more complicated social relationships.

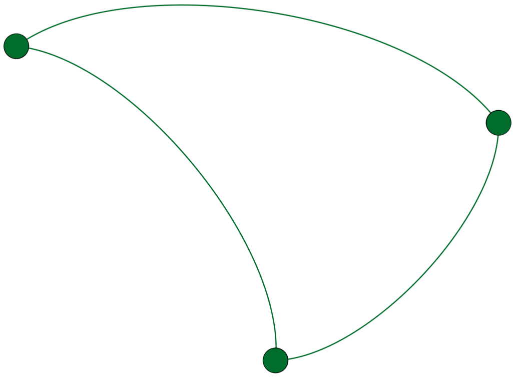


Figure 2.4: Sociogram example of three people. Each node is a person, and each edge represents a relationship between two people.

Around 10 years after the Internet was invented, the phrase "surfing the internet" was becoming popularized in 1992. In the same year, the first web image, which is for a band "Les Horribles Cernettes," was posted on the web. Geocities, the first social networking website, was launched in 1994. Geocities enabled users to construct and personalize their own websites, categorizing them into different "cities" depending on the content of the site. The next year, TheGlobe.com was made available to the public, allowing people to post their own material and to engage with others who shared their hobbies and interests. Due to the ever-growing popularity of the World Wide Web and the Internet, an expanding variety of services are accessible via computer networks. People who use these services have established a new type of virtual society commonly referred to as “online social networks” [138–141], or “web-based social networks,” [142] “computer-supported social networks,” [143] and “virtual communities” [144].

However, a social network is more than merely a collection of people; it is the total of the relationship that bind these people together [145]. Twitter, founded in 2006, is a well-known SNS. The basic relation netween users on Twitter is the follower-follower relationship, which generates a friendship network. In addition to that, a “retweet network” can be generated by examining the “retweeted_status” object of a tweet object. In each “retweeted_status” object, one can find a retweeting relationship between two users. Through studying all the retweeting relationships, it is possible to produce a retweet network of all the users. The social network shares an important feature: clustered. A cluster signifies that there is a high probability that two neighbors of a given node are also direct neighbors themselves. Clustering of nodes is also a popular research topic for SNSs.

Clustering algorithms attempt to capture the intuitive concept that nodes should be linked to many nodes within the same community (intra-cluster density) but to few nodes within different communities (inter-cluster sparsity).

The Louvain algorithm [146] is among the first scalable approaches based on Newman-Girvan modularity maximization [144]. It is an agglomerative hierarchical technique with a greedy approach to local optimization. Two stages make up the algorithm. In the first stage, the method iterates over the graph nodes and assigns each node to a community if doing so would boost modularity. In the second stage, the method constructs node clusters from the clusters discovered in the previous step. Iteratively, the increased value (can be negative) in modularity is always calculated using the base graph. Even though the underlying computational problem is NP-hard, the Louvain algorithm uses an efficient and effective heuristic that strikes a balance between the quality of the solution and the amount of work it takes to solve the problem, which scales roughly linearly with the number of edges. Eq. (2.1) demonstrates the computation of the modularity measure, where A_{ij} denotes the edge weight between node i and node j, $k_i = \sum_j A_{ij}$ represents the accumulation of all edge weights of vertex i, c_i denotes the community to which vertex i is assigned, the δ function $\delta(u,v)$ is 1 if $u = v$ and 0 otherwise and $m = \frac{1}{2}\sum_{ij} A_{ij}$.

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta\left(c_i, c_j\right) \tag{2.1}$$

The Louvain method initializes the partition with each node in its own cluster. The modularity gain for relocating each node to nearby clusters is then computed for each node. The highest positive increased value is selected, and then the node is relocated. This is repeated until no nodes are moved throughout a complete loop. Every cluster is then integrated into one node; intra-cluster edges are accumulated with loops, while inter-cluster edges are merged into one edge, and the weights of edges are accumulated. This intermediate graph is utilized in the following iterations until no more advantages can be obtained by moving another node. The modularity range is located in $[-1, 1]$. A score of 1 indicates ideal clustering, in which there are no edges between clusters and all clusters are strongly linked. A lower value implies a worse outcome. A negative value indicates an extremely poor clustering. To calculate the increased value for Q of bringing a node i into the cluster C, rather than computing the new modularity of the entire cluster, it is more efficient to compute a local value that reflects the change in the modularity value, as indicated in Eq. (2.2) [147].

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m}\right)^2\right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2\right], \tag{2.2}$$

where $\sum_{in}$ is the sum of the link weights within the cluster C, $k_{i,in}$ is the sum of the link weights from i the other nodes within C, $\sum_{tot}$ is the cluster degree, and k_i is the node i's degree.

Recall that a tweet should be able to connect or include hashtags that don't have a clear relationship with each other but are used a lot together with hashtags it has already used. Thus, the number of times a hashtag is used in a tweet indirectly shows a wider semantic relationship between tweets, which is able to show the topic diversity of tweets. With the help of the Louvain algorithm, it is possible to suggest topics, in addition to advanced topic modeling methods, such as Latent Dirichlet Allocation (LDA) [148].

In addition, researchers use other popular algorithms for network clustering, including Local Moving [149], Informap [150], and Label Propagation [151]. Compared to other methods, Louvain is more popular. Louvain has many benefits, such as being easy to use, fast, and able to handle large and weighted networks. More importantly, compared to a series of clustering methods [152], the Louvain algorithm has shown that it gives better results for detecting communities in real data. For example, the Louvain method has been used in COVID-19 infodemic-related research [153–155]. During the COVID-19 infodemic, conspiracy theories and misinformation have penetrated SNSs, such as Twitter [136]. Since the diffusion of non-credible information has been an online social network, it is necessary to find out their network features to generate corresponding counter-solutions.

2.5 Machine learning

Nowadays, machine learning and deep learning are popular tools both for research and for media. Machine learning as we know it today can be traced back to a psychologist at Cornell University named Frank Rosenblatt. He led a group that built a machine that could recognize letters of the alphabet based on ideas about how the human nervous system works [156–158]. Frank Rosenblatt called the machine "perceptron." It used both continuous and discrete signals and had a "threshold" part that turned continuous signals into discrete ones. It became the model for modern artificial neural networks, and the way it learned was similar to how animals and people learn. Rosenblatt is the first scientist to study the perceptron mathematically [157]. But the Novikoff theorem [159], which specifies the conditions for a perceptron learning algorithm to converge, was more popular in the scientific field. The well-known "Backpropagation" was introduced in 1974 [160]. The first practical deep neural network, LeNet, was developed by Yann LeCun in 1990 [161]. In 2007, Deep Belief Networks and layerwise pretraining techniques [162] opened the modern deep learning era.

Natural Language Processing (NLP) has computers to understand, interpret, and work with human languages (natural languages). Usually, humans and computers communicate through a programming language. When it comes to using natural language to interact with a machine, it is not easy to directly carry out like that because natural language is vague, with jargon words, linguistic ambiguity, and social contexts. When we take into account the culture of countries, local traditions, and even different accents, the NLP interpretation task becomes much more difficult.

NLP can comprise five major components: morphological analysis, syntactic analysis, semantic analysis, discourse analysis, and pragmatic analysis [163]. Syntactic and semantic analysis are two of the most important NLP tasks. The purpose of syntax analysis is to rearrange the words of a phrase so that it begins to make grammatical sense. It helps NLP in determining the meaning of a sentence based on grammatical principles. The purpose of semantic analysis is to determine the meaning of words and their usage in a phrase. NLP uses it to comprehend the structure and meaning of a phrase. Due to the availability of a large number of texts (data) for natural languages, researchers in NLP have recently paid significant attention to language models with deep learning architecture such as BERT [164] and GPT series [17, 117].

In addition to deep learning methods, word importance can be measured by "Term Frequency-Inverse Document Frequency" (TF-IDF). IDF [165] is one of the fundamental and widely-applied notions in NLP. A highly successful approach to word weighting has been used in information retrieval systems [166]. TF-IDF determines the relative frequency of words in a specific document relative to the inverse proportion of that word across the entire corpus of documents. This calculation determines the relevance of a given word within a given document. Common words in a single document or a small group of documents tends to have higher TF-IDF values than common words such as articles and prepositions [167]. The formal procedure for implementing TF-IDF varies slightly across its various applications, but the overall strategy is as follows: Given a document collection D, a single document $d \in D$, and a token w, the TF-IDF value of w is calculated as follows:

$$TF - IDF = f_{w,d} \cdot \log\left(|D|/f_{w,D}\right), \tag{2.3}$$

where $f_{w,d}$ is the number of times w occurs in document d. $|D|$ is the corpus size and f_w, D is the number of documents in which w appears in corpus $|D|$. TF-IDF has the following advantages: 1) It is easy to calculate, 2) the most descriptive terms in a document can be an identifier with basic metrics; 3) since TF-IDF weights words depending on their significance, this approach may be used to discover which words are the most significant. This may be used to more effectively summarize articles or simply identify keywords (or even tags) for a text. However, TF-IDF does not take the positional information into account, and therefore, the associations between words cannot be calculated and the linguistic ambiguity cannot be measured either. Additionally, sentiment words cannot be captured by the TF-IDF approach.

Word2vec is able to compensate for this disadvantage. According to the original paper [168], this model employs a two-layer neural network to produce a vector for each word. Both syntactic and semantic features of words should be implied in the word vectors, carried out by the continuous bag of words (CBOW) technique and skip-gram of Word2Vec. For a more accurate representation of words, it is advised to train the model using a large corpus. Word2Vec has proven to be effective for several NLP-related tasks [169]. Word2vec was created to make embedding training more meaningful, and it has subsequently become the standard for producing pre-trained word representation. Word2Vec predicts using one of two neural network models, such as the CBOW and Skip-gram models, based on the context. Two algorithms are proposed to address the problem of learning the final vectors. The negative sampling technique used for Word2Vec training ensures that only a portion of the vectors is updated based on a noise distribution. This technique helps to accelerate the training speed. The other technique is hierarchical softmax, which is based on the Huffman tree. Huffman tree is a binary tree that provides all possible words based on their counts. Huffman tree ensures that a more frequent word has a shorter code, which is a useful data compression method. Word2vec is commonly used to compare the similarity of word meaning.

Chapter 3

QAnon user dynamics and topic diversity during the COVID-19 infodemic

3.1 Introduction

3.1.1 A brief history of QAnon

In recent years, global populism has become more and more popular. Conspiracy theories and populism are mutually relevant. Typically, they contain two roles: the powerful elites who control social resources and privilege, and the commoners who are portrayed as vulnerable victims [170].

One of the most popular conspiracy theories is QAnon. QAnon is a conspiracy theory umbrella that encompasses various individuals, including Trump supporters, COVID-19 deniers, and anti-vaccination activists. In 2017, an anonymous government official known as “Q” appeared on 4chan (an anonymous English-language forum), claiming that a cabal of upper hierarchy elites controlled the United States and abused children covertly (#pizzagate); People are encouraged to support Donald Trump (The QAnon conspiracy theory emerged during his presidency) because they believe that he will arrest all members of the “deep state,” including Hillary Clinton and Barack Obama, and bring the cabal to justice [171–173]. Despite the fact that QAnon is not an extremist organization, extremists existed within the QAnon movement. On 6 January 2021, an organized group of QAnon and Trump protesters stormed the U.S. Capitol building. This well-known violence implied that far-right extremists are present in QAnon supporters.

To gain more exposure during the COVID-19 pandemic, QAnon has been using social topics that are controversial and popular. For example, QAnon conspiracy theories blamed China for the long-term cover-up of the coronavirus; spread the idea that mandated quarantine helped protect Joe Biden during the election; questioned the travel ban and advocated the use of hydroxychloroquine; arbitrarily linked COVID-19 to the presidential election and China so that the coronavirus was just a media-hyped tool to secure the Democrats' victory in the election; and even introduced a discord element such as "Black Lives Matter" to the 2020 U.S. presidential election [45].

Meanwhile, to spread its beliefs, QAnon arbitrarily linked COVID-19 to the U.S. presidential election and China [45]. According to surveys on the QAnon conspiracy theory, most U.S. citizens who have heard of QAnon believe the conspiracy theory is harmful to the country [174]. There are, however, many people who fall somewhere in the middle (referred to as "less-leaning users") who consider QAnon as neither harmful nor helpful; they must not be overlooked because they have the potential to become pro-QAnon in the long run.

QAnon has been on popular social network services (SNSs) for a long time before Facebook, Twitter, and YouTube realized that the QAnon conspiracy theory could cause the U.S. Capitol violence in 2021, as mentioned above. Followers of QAnon tend to talk about violence on Twitter [175]. In 2020, thousands of QAnon accounts were deleted from these platforms. Faced with this reality, people who believed in QAnon started looking for new spiritual homes on SNSs like Parler and Telegram. Parler is an American alt-tech microblogging SNS known for the discussions among Trump supporters. On Telegram, there are active QAnon channels where people from many different countries can talk about QAnon. Although QAnon is still a mystery, it is clear that the COVID-19 infodemic has helped the conspiracy theory spread throughout the world.

3.1.2 Related works

The connection between information regarding COVID-19 and the pandemic has thrown light on epidemiology policy and the attitude of the local community toward the advice of experts source [176]. The COVID-19 infodemic is a scenario in which the proliferation of COVID-19-related mis/disinformation on SNSs makes it impossible for the general public to get reliable information regarding the pandemic.

Several studies have analyzed the linguistic characteristics of QAnon. Aliapoulios et al. (2021) compiled a dataset of 4,949 "Q drops" and determined that they were not produced by a single individual, suggesting the presence of apocrypha among those drops [171]. Phadke et al. (2021) examined

483 language features and developed a computer framework to analyze self-disclosures of dissonance and calculate changes in user participation in the presence of dissonance [177]. Hoseini et al. (2021) examined QAnon in several languages using a BERT-based topic model and found that the German language is popular in QAnon groups and channels on Telegram [178]. Anwar et al. (2021) used VADER to analyze QAnon-related users' attitudes toward Trump and Biden, and a BERT model was used to define user profiles [179]. They discovered that the majority of QAnon users were Donald Trump followers, with Twitter profiles that accommodate "MAGA," "God," "Patriot," and "WWG1WGA." Miller (2021) examined QAnon comments on YouTube and discovered significant worldwide conversations regarding China, Russia, and Israel [180]. These language results show that the QAnon conspiracy is widespread online and that QAnon has established a global presence.

Other studies have used networks to investigate semantic components of the QAnon conspiracy theory. Papasavva et al. (2021) discovered QAnon-relevant word graphs in the Voat community using a word embedding [181]. Hanley et al. (2021) built a QAnon-related domain network and trained a random forest classifier to distinguish between misleading and legitimate news sites [182].

In addition, the challenge of SNSs detecting QAnon groups and banning malicious members is growing increasingly difficult. Twitter's rules and procedures did not get widespread public notice until January 2021. It was stated that 355K Twitter users participating in the 2020 U.S. Presidential Election scandal had been deleted [183]. Chowdhury et al. (2020) found that more than 60% of the purged users lived for more than two years before being deleted by Twitter, raising concerns about the efficacy of the purge [184]. Meanwhile, whether or not removing disruptive accounts adds to a healthy online social community remains debatable, particularly among QAnon users.

3.1.3 Research questions

QAnon appears to be using the plethora of COVID-19 mis/disinformation to achieve political impact. It distributes misinformation and instills negative emotions, both of which are damaging to "less-leaning users," those who do not have a strong preference for QAnon but have the potential to become pro-QAnon in the long term. Although some features of QAnon have been explored, there is a dearth of research on how QAnon developed in terms of user dynamics and topic diversity during the COVID-19 infodemic. The RQs of the study 1 are therefore raised here, and investigated using a network-based approach:

RQ1-1: What are the pro- and anti-QAnon user dynamics during the COVID-19 infodemic?

RQ1-2: What topics do QAnon users spread during the COVID-19 infodemic?

3.2 Data and Methods

The dataset and the techniques used to study the dynamics of QAnon during the COVID-19 infodemic with a network-based approach will be described in this section.

3.2.1 Data

Between February 20, 2020 and March 1, 2021, 880,278,195 posts from 58,519,206 unique users (including tweets and retweets) were collected using the Twitter Search API and the following COVID-19-related keywords: “corona virus,” “coronavirus,” “covid19,” “2019-nCoV,” “SARS-CoV-2,” and “wuhanpneumonia.” This dataset is called **base dataset**. In addition, English tweets containing at least one of the phrases “QAnon,” “#QAnon,” or “deep state,” was extracted, which produced 308,631 tweets from 135,740 unique accounts [1]. This subset is called **QAnon dataset**. This study included both datasets.

3.2.2 Identification of pro-/anti-QAnon users and their leaning

Because QAnon is a conspiracy theory that has sparked debate both for and against its assertions, it is anticipated to find a characteristic retweet (RT) network with segregated pro- and anti-users. Using the QAnon dataset, an RT network was constructed and the k-core decomposition [185] ($k = 2$) to applied to identify pro- and anti-QAnon users, with each node representing a user and directed edges between nodes representing retweeting relationships. This resulted in an RT network with two large clusters, as expected. By manually evaluating larger-indegree users in each cluster (who were retweeted many) in terms of their tweets and profile descriptions, which cluster corresponded to the pro- or anti-QAnon group can be determined.

[1] https://github.com/myrainbowandsky/A-network-based-approach-to-QAnon-user-dynamics-during-COVID-19-infodemic

A manual verification was carried out to ensure that the categorization of pro- and anti-QAnon users was reliable enough. All users were divided into two groups for the manual verification. Two annotators participated in this exercise and categorized 60 randomly chosen users, with 30 labelled as pro-QAnon and the other 30 designated as anti-QAnon. The annotators were provided with the user names and asked to read their profiles and tweets. The annotators finally categorized them as pro- or anti-QAnon. The consistency of their classifications was then evaluated using Cohen's kappa. The final Kappa was 0.76, which signified strong agreement and verified the user classification result as statistically valid [2].

In addition, "QAnon-leaning" was defined as follows and three types of users were identified including: "pro-leaning users," "anti-leaning users," and "less-leaning users."

$$L = \frac{P - A}{P + A}, L \in [-1, 1], \tag{3.1}$$

where P is the number of retweets from pro-QAnon users and A is the number of retweets from anti-QAnon users. L compares the leaning of a user between pro-QAnon and anti-QAnon based on retweet tendencies. If a user has more than 70% probability of retweeting from the pro-QAnon side, this user is considered pro-leaning in this study, and vice versa. Thus, $-0.4 \leq L \leq 0.4$ indicates that the user is less-leaning; $L > 0.4$ indicates that the user is pro-leaning; $L < -0.4$ indicates that the user is anti-leaning.

Please keep in mind that QAnon-leaning is measured by L (Eq. 3.1), while pro- and anti-QAnon classifications are based on retweet network clustering, the results of which were validated as previously stated.

3.2.3 Classification of Bots and humans

The Botometer (described in Sec. 2.3) API V4 was used in this study to classify users into bots and humans. The major difference between Botometer API V4 and V3 is that V4 retained that model with new training data and applied an architecture called " Ensemble of Specialized Classifiers", which is able to predict a bot on a bot style basis [187].

3.2.4 Hashtag co-occurrence networks

Latent Dirichlet Allocation (LDA) is a typical method for modeling topics from a given text. However, due to the short text length of tweets, LDA

[2] Note that according to [186], Cohen's kappa value is interpreted as follows: 0.0-0.2 for slight agreement; 0.2-0.4 for fair agreement; 0.4-0.6 for moderate agreement; 0.6-0.8 for substantial agreement; and 0.8-1.0 for near-perfect agreement.

frequently fails to extract definite topics. In this study, LDA modeling was applied to obtain topics from retweets using the library pyLDAvis [188], but the method did not obtain informative topics (see Table A.2 in Appendix).

Therefore, hashtag co-occurrence network was used to examine the topic diversity of QAnon conspiracy theory. The network is simpler yet useful for examining intricate links among topics, which LDA topics cannot achieve. In order to comprehend the topic richness of the QAnon conspiracy theory, co-occurrence networks for both the base and QAnon datasets were generated. Each node in the hashtag co-occurrence network represents a hashtag, and the undirected edges between nodes indicate the co-occurrence of two hashtags. A hashtag co-occurrence network was produced with the base dataset, and the k-core decomposition ($k = 10$) was applied to the network. Finally, all direct neighbors of "#QAnon" and itself were retrieved. Recall that the first dataset contains various languages (not only English). For a clear visibility, a 1000-core hashtag co-occurrence network with 336 unique hashtags (nodes) was generated. Similarly, from the QAnon dataset, a 10-core hashtag co-occurrence network with 323 unique hashtags was produced as well.

The Louvain algorithm [146], a modularity-based community recognition method, was applied to the hashtag co-occurrence network using the Gephi [189] [3] in the hashtag community detection. In the end, each node of the hashtag co-occurrence network was assigned a unique modularity class ID for subsequent analysis purposes.

3.2.5 Hashtag semantic map

To understand the similarity of topics, word2vec [168], a well-known word embedding technique, was applied to visualize a semantic map of QAnon hashtags. In this analysis, only the top 50 degree hashtags were considered, except the most QAnon-related hashtags, including: "#QANON," "#QANONAS," "#Q," "#QANON2020," "#THESTORM," "#WWG1GWA," and "#WWG1WGA" because these hashtags could be linked to any semantic cluster in the QAnon dataset and finally form a dense and large semantic cluster. Word2Vec model was finally trained using the Gensim topic modeling library [4] by exploiting tweet texts and hashtags. Then, the HDBSCAN[5] [190] was used to cluster the hashtag embeddings with Word2Vec. After that, UMAP was used to reduce the dimensionality of the embeddings ($d = 2$) and visualized them [6] [191].

[3] https://gephi.org/

[4] https://github.com/RaRe-Technologies/gensim

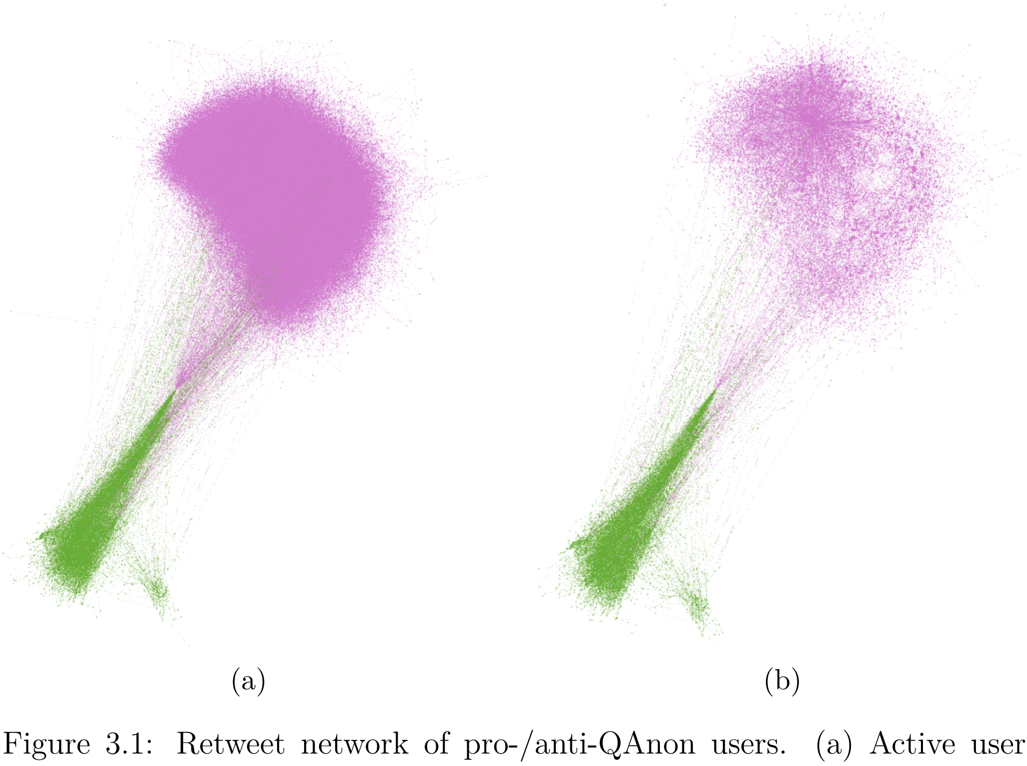
[5] https://github.com/scikit-learn-contrib/hdbscan

[6] https://github.com/lmcinnes/umap

3.3 Results

3.3.1 QAnon's user dynamics

The retweet (RT) network (2-core) that was constructed from the QAnon dataset between February 2020 and March 2021 is shown in Fig. 3.1. The RT network demonstrates that pro- and anti-QAnon clusters are topologically separated from each other. The pro-QAnon cluster occupied a much larger size ($n = 40,512$), compared to the anti-QAnon cluster ($n = 5,480$) (See Table 3.1). Recall that the pro- and anti-QAnon classification was manually validated, as described in Sec. 3.2.2. The activity of each in the network was examined again in August 2021 in order to estimate how many of them were suspended by Twitter. From Fig. 3.1a to Fig. 3.1b, more than 50% (25,318) of the users in the pro-QAnon cluster had their accounts closed or were suspended, whereas in the anti-QAnon cluster, only 653 users were suspended (Table 3.1).


(a)

(b)

Figure 3.1: Retweet network of pro-/anti-QAnon users. (a) Active users from February 2020 to March 2021; (b) Active users in August 2021, in which magenta denotes pro-QAnon and green denotes anti-QAnon.

The dynamics of the users was then investigated through "QAnon-leaning"

(L). Fig. 3.2 is a user scatter plot generated from the QAnon dataset, demonstrating the connection between the number of retweets from pro-QAnon users and those from anti-QAnon users. The figure shows that there were not only users who retweeted the most from pro-QAnon users (that is, "pro-leaning") but also users of "anti-leaning" and "less-leaning" (see Fig. A.2a in the Appendix).

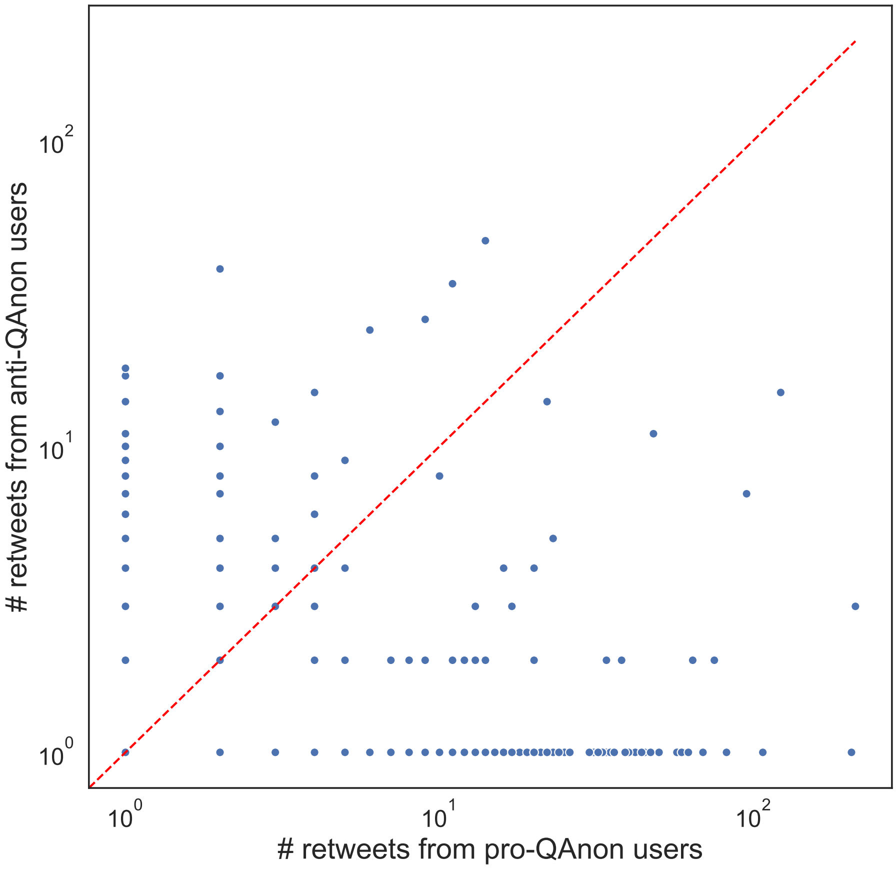


Figure 3.2: User scatter plot with the number of retweets from pro-QAnon users and the number of retweets from anti-QAnon users (logarithmic scale) based on Twitter data between February 20, 2020 and March 1, 2021.

Fig. 3.3 shows temporal dynamics of QAnon-leaning (L) distributions for less-leaning users. With the exception of a bimodal peak (about 0.3) in

June 2020, the majority of less-leaning users are constantly centered on 0.0 over months. However, the two opposing groups of users continued to evolve over time. In February 2020, all users were pro-QAnon, but in subsequent months, anti-QAnon users outnumbered pro-QAnon users.

The same plots for pro- and anti-leaning users are illustrated in Fig. A.2b and c in the Appendix. Unlike less-leaning users, the distributions of QAnon-leaning users were more stable, demonstrating that both pro- and anti-leaning users were consistent in the contents they retweeted over time. This finding implies that Twitter's policy in removing malicious users may have prevented less-leaning users from adopting a pro-QAnon stance. Even though less-leaning users are in the minority, an SNS like Twitter needs to figure out how to protect them from a large number of pro-QAnon users.

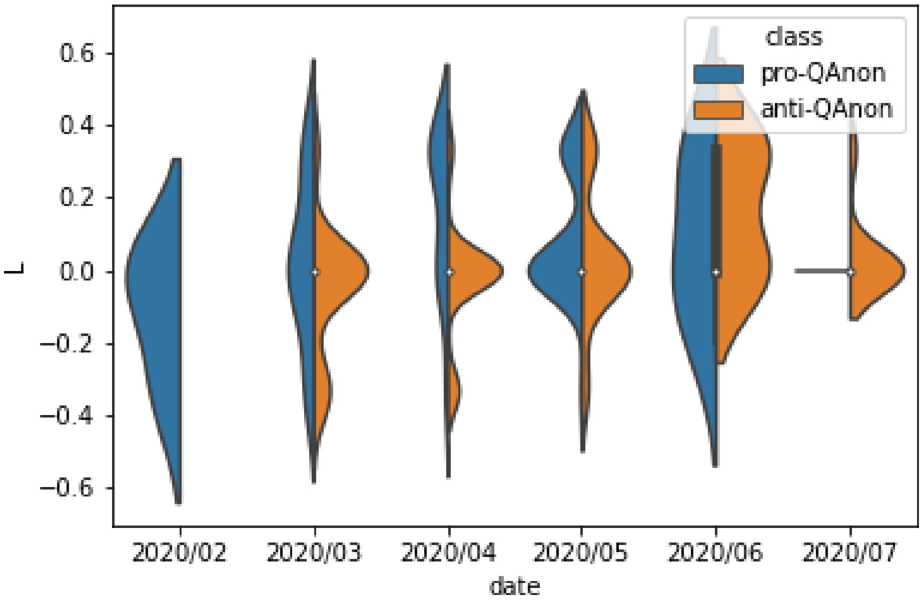


Figure 3.3: QAnon-leaning (L) distributions for less-leaning users. Note that pro- and anti-QAnon classifications are based on the retweet network clustering.

Next, the monthly change of active users—pro-leaning, anti-leaning, less-leaning, and total was measured Fig. 3.4). The overall number of active users peaked in March 2020, then began to decline. However, one month later, the numbers of pro-leaning, anti-leaning, and less-leaning people all peaked. After that, the number of pro-leaning and less-leaning users kept decreasing,

Table 3.1: Summary of pro- and anti-QAnon users (February 2020 to March 2021; suspended or closed accounts as of August 2021).

	#pro-QAnon	#anti-QAnon
All users	40,512	5,480
Suspended users	25,318	653
Bots	8,239	2,861
Humans	6,016	2,592

while the number of anti-leaning users rose again in July 2020. Similar patterns were observed in the way these users retweeted (see Fig. A.3 in the Appendix). All of these statistics suggest that Twitter's removal of malicious users may have led to a decrease in pro-QAnon users and an increase in anti-QAnon users.

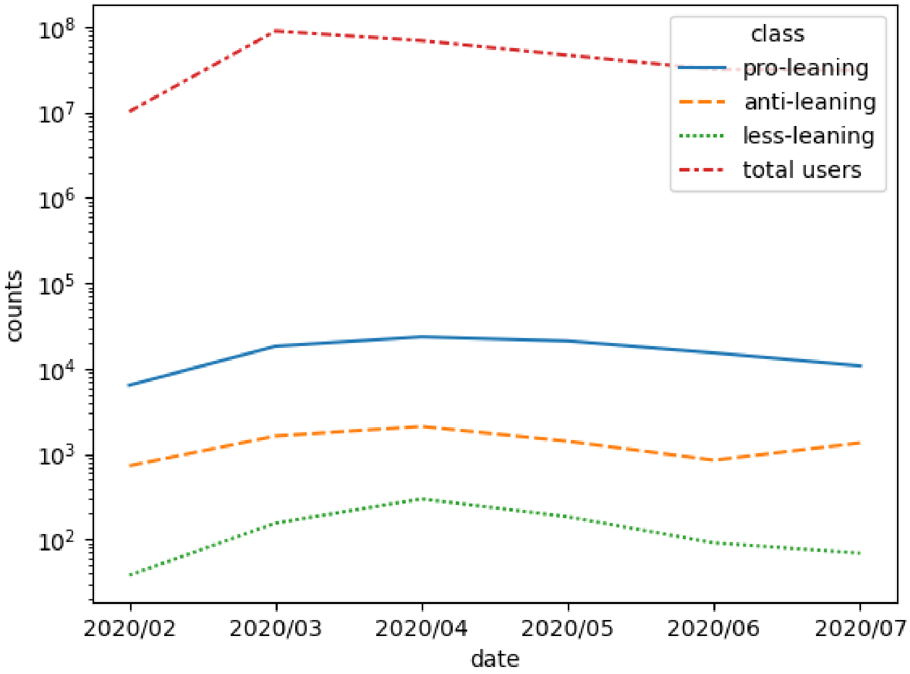


Figure 3.4: User activity: the numbers of active users who retweeted at least once a month, including pro-leaning, anti-leaning, and less-leaning users, and total.

3.3.2 Prevalence of bots in QAnon clusters

In addition, how many bots were involved in pro- and anti-QAnon users was examined. In the pro-QAnon cluster, there were 8,239 bots and 6,016 humans; while in the anti-QAnon cluster, there were 2,861 bots and 1,592 humans (as seen in Table 3.1). Bots are, therefore, prevalent not only in the pro-QAnon cluster but also in the anti-QAnon cluster, playing a major role in the dissemination of QAnon conspiracy topics, on the one hand, and passing on contents debunking them, on the other hand. Note that it was not possible to obtain all bot scores owing to Twitter's policy or their inaccessibility due to private settings; hence, the number of bots and humans indicated here may be lower than the real amount.

3.3.3 Hashtag co-occurrence network as a conspiracy theory umbrella

Using the base dataset, the global hashtag co-occurrence network (1000-core) was created. The resulting network is depicted in Fig. 3.5 (n = 336), which clearly shows a topic landscape for the QAnon conspiracy theory during the COVID-19 infodemic, given that the base dataset contains several languages and various COVID-19 topics.

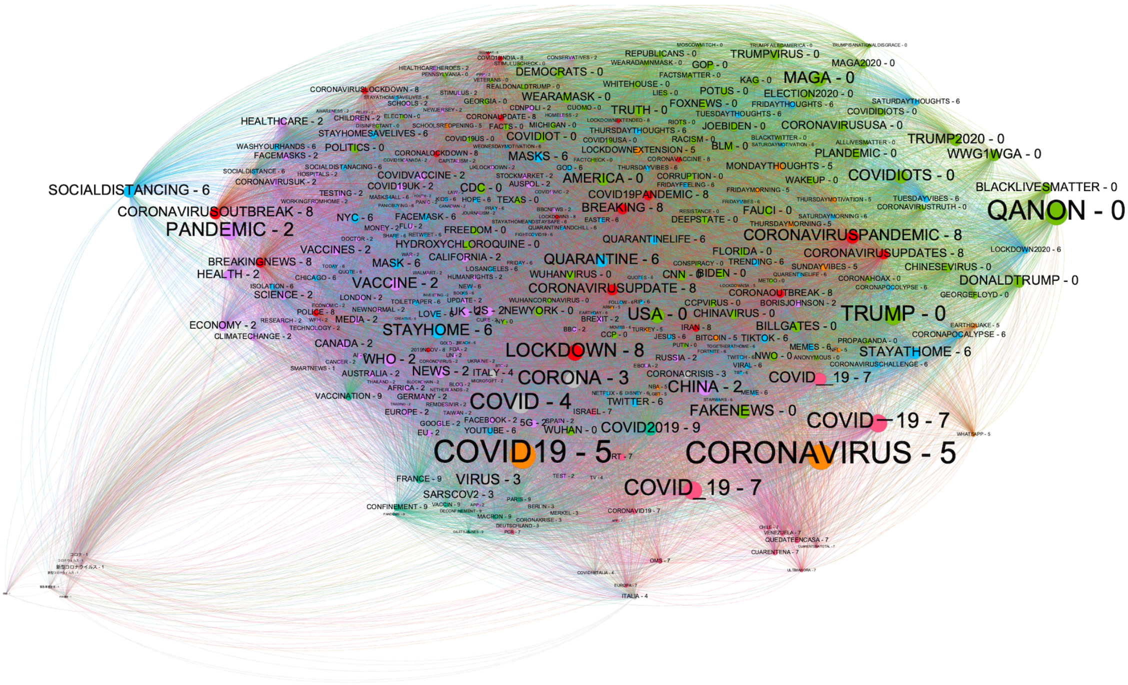


Figure 3.5: Global hashtag co-occurrence network. Numbers denote hashtag classes. “#QANON” is in class 0 (green). The degree is represented along with each hashtag. The label size of a node is proportional to its degree.

The top three topics are “U.S. politics”, “News,” and “Daily life.” Additionally, #QAnon has co-occurred with human rights hashtags such as “#LGBT,” ($k = 1,418$) “#METOO,” ($k = 1,073$) and “#BLACKLIVESMATTER,” ($k = 6,390$) which is consistent with [45]. Note that k represents the frequency or degree of occurrence. The co-occurrence pattern of popular hashtags reflected the topical richness of the QAnon conspiracy theory and suggested that the QAnon co-occurrence network increased users’ exposure during the pandemic.

Moreover, there is an isolated cluster (class 1) of Japanese hashtags in the bottom left corner of Fig. 3.5, which corresponds to J-Anon, the Japanese counterpart of QAnon. Users of J-Anon concur that (former) President Trump is a hero in the struggle against the “deep state.” Furthermore, the topic links of QAnon, France (French language tweets, class 9), Spain (Spanish language tweets, class 7) and Italy (Italian language tweets, class 4), demonstrate that QAnon is becoming a global conspiracy theory movement, particularly in western countries. This study lends credence to the notion that the QAnon conspiracy theory developed in local specialized communities, such as 4chan and 8chan, before migrating to become a worldwide popular conspiracy theory [171]. Moreover, the religious hashtags relevant to the "apocalypse" that Trump supporters believe in were connected to #QAnon. They believed that Trump was sent by God [192]. In fact, there is a tweet mentioning "Armor of God ! ! #qanon #wearethenewsnow #factsmatter #wwg1wga #wakeupamerica #covid19 #unitednot."

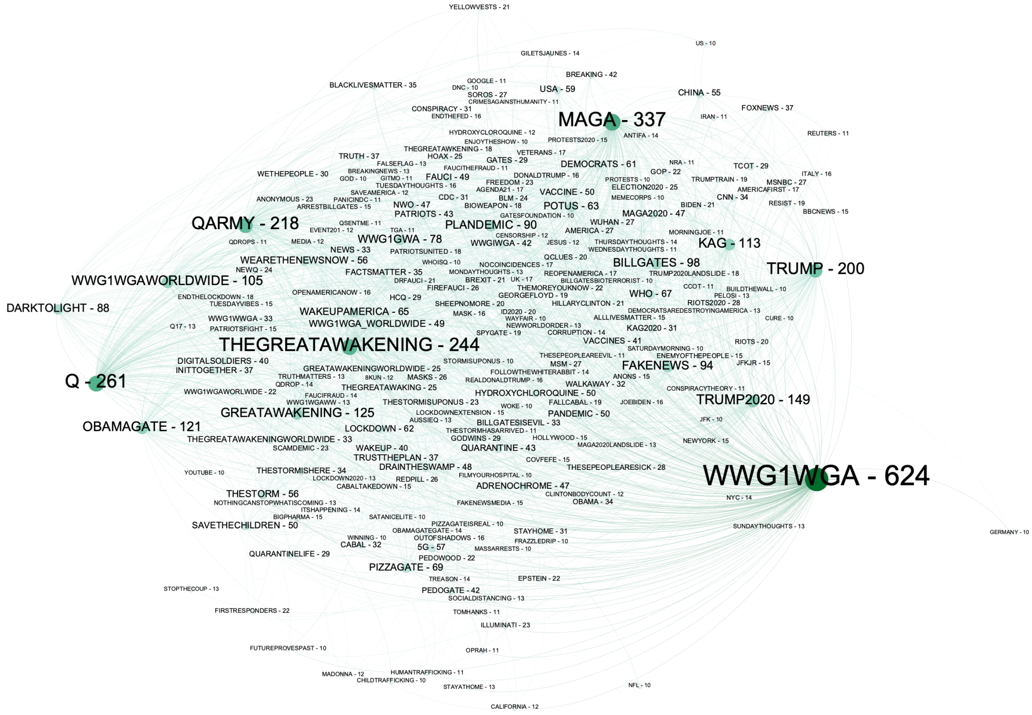


Figure 3.6: English QAnon hashtag co-occurrence network.

In addition, the well-known QAnon hashtags were observed, such as "#WWG1WGA," ($k = 624$) "#MAGA," ($k = 337$) "#THEGREATAWAKENING ($k = 244$); it seems that QAnon debunking information was also present in the network, for instance, "#FAKENEWS," ($k = 94$) "#FAKENEWSMEDIA," ($k = 15$) and "#CONSPIRACY" ($k = 31$) were identified in the network. Since "#FAKENEWS" is identified in both global and English hashtag co-occurrence networks, it is considered that there are at least two voices about QAnon: one is pro-QAnon and the other is against QAnon, which is in line with what the visualizations of QAnon users show (Fig. 3.1). Furthermore, "#FAKENEWS" and its 64 neighbors existed, which shows that there was a voice debunking the truth about QAnon.

To understand the topics in Fig. 3.5 in detail, the top-50 degree hashtags were examined in relation to pro- and anti-QAnon users. (See the statistical summary in Table 3.2.) The three most popular topics are the same as the ones described above: US politics (class 5), COVID-19 (class 0) and News (class 2). These two networks indicate that QAnon has been evolving into a much larger conspiracy umbrella worldwide, which may potentially attract less-leaning users, who are neutral to pro- and anti-QAnon clusters.

Table 3.2: Top 10 hashtags preferred by pro- and anti-QAnon users.

Topic class	%Pro	%Anti	%Pro/%Anti
U.S. politics	80	20	4.0
J-Anon	32	68	0.5
News	70	30	2.3
Lockdown	67	33	2.0
Italy	67	33	2.0
COVID-19	61	39	1.6
Daily life	73	27	2.7
Spain	72	28	2.6
India	70	30	2.3
France	78	22	3.5

To determine hashtag preference, the ratio ($\%Pro/\%Anti$) of pro-users' %hashtags to anti-users' %hashtags was computed (Table 3.2). Here, a higher ratio means a tendency to lean toward the anti-QAnon side. If $\%Pro/\%Anti$ >1, the users are holding pro-QAnon tendency in that hashtag topic; if $\%Pro/\%Anti$ <1, the users are holding an anti-QAnon tendency in the topic; and if $\%Pro/\%Anti = 1$, the users are holding balanced or neutral views toward the topic. Except for J-Anon, most of the hashtags showed a pro-QAnon tendency. J-Anon hashtag community is located in the bottom left of the Fig. 3.5). The detailed is enlarged in the Fig. 3.7. It seems that these hashtags (e.g., corona, and declaration of a state of emergency) are probably from the mainstream media. No suspicious hashtags, which were closely related to fake news, were identified. Thus, the J-Anon hashtag community is more likely to be related to mainstream news and therefore anti-QAnon. Other topics, in particular, users associated with U.S. politics and French tweets were more pro-QAnon.

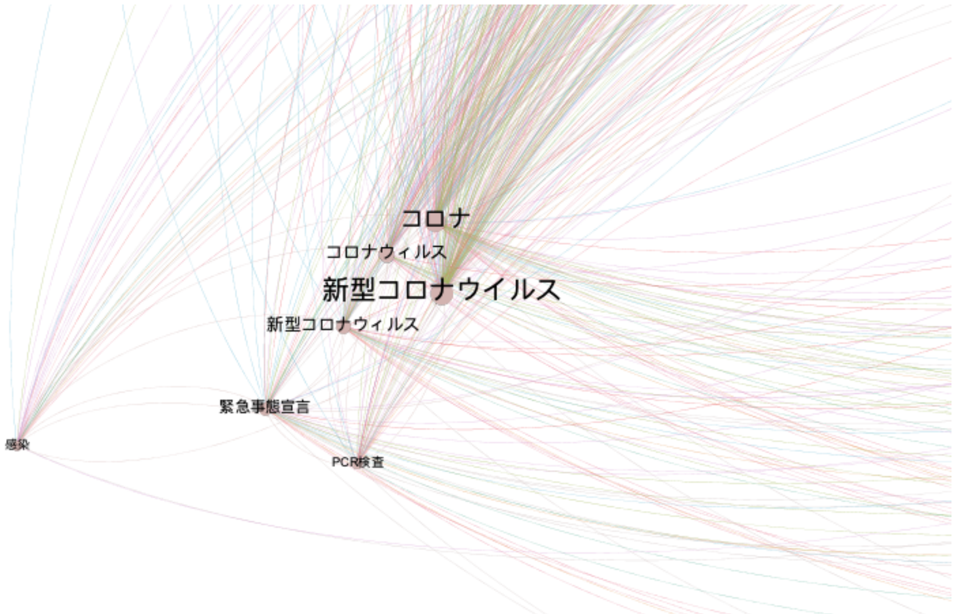


Figure 3.7: J-Anon hashtag community in the QAnon hashtag co-occurrence network.

3.3.4 Hashtag semantics and dynamics

The hashtag co-occurrence network shows topic diversity, but lacks the ability of measuring the similarity between hashtags. Therefore, a semantic map of the top 50 popular hashtags in the global hashtag co-occurrence network (Fig. 3.8) was generated. Semantically similar hashtags are clustered on the map: the conservative cluster (cluster 0: e.g., #trump, #maga), the QAnon cluster (cluster 1: e.g., #plandemic, #5g), the vaccine cluster (cluster 2: e.g., #vaccine, #fauci) and outliers (cluster -1: e.g., #china, #fakenews). The map includes diverse QAnon topics, such as #plandemic, #5g, #pizzagate, and #obamagate as well. The lexical resemblance of the largest cluster 3 could be explained by the fact that "#plandemic" and QAnon were associated in the context of community victimization [193].

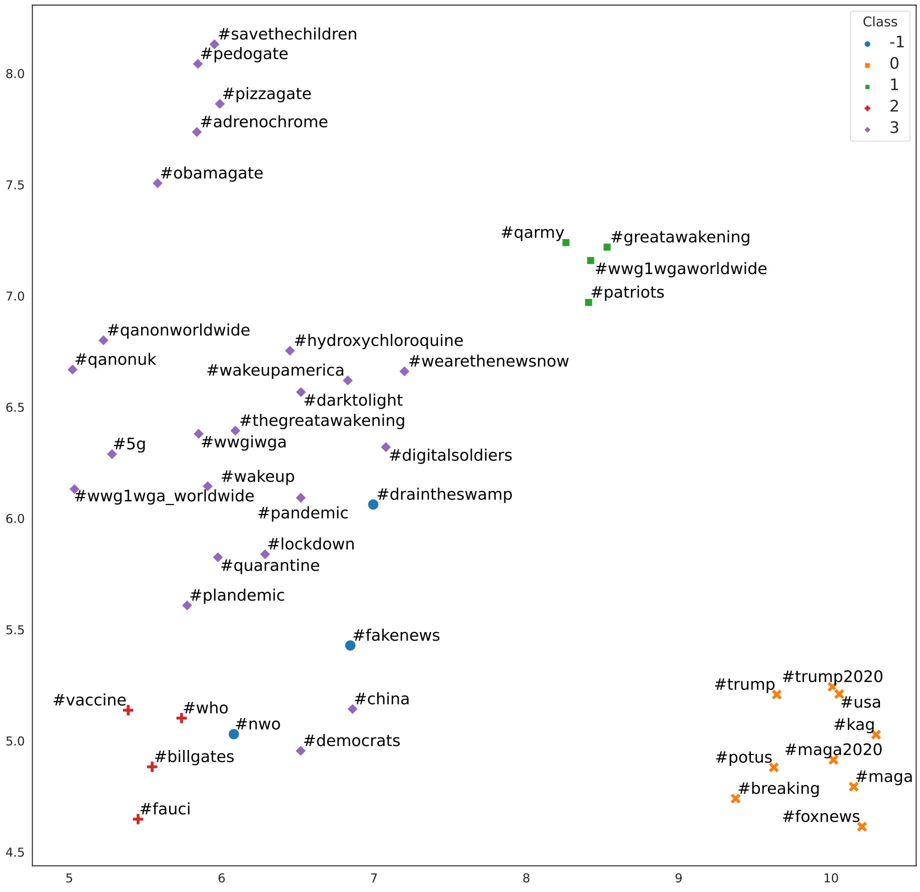


Figure 3.8: Semantic map of top 50 popular hashtags.

Finally, temporal changes of hashtags occurrences was examined (Fig. 3.9). QAnon representative hashtags, including "#WWG1WGA," "#Q," "#QARMY" and "#THEGREATAWAKENIN," appeared together in sync. It turns out that these hashtags were involved in the gigantic component of the global hashtag co-occurrence network. The degrees of these hashtags reached their peaks between April and May 2020, during which QAnon topics flourished.

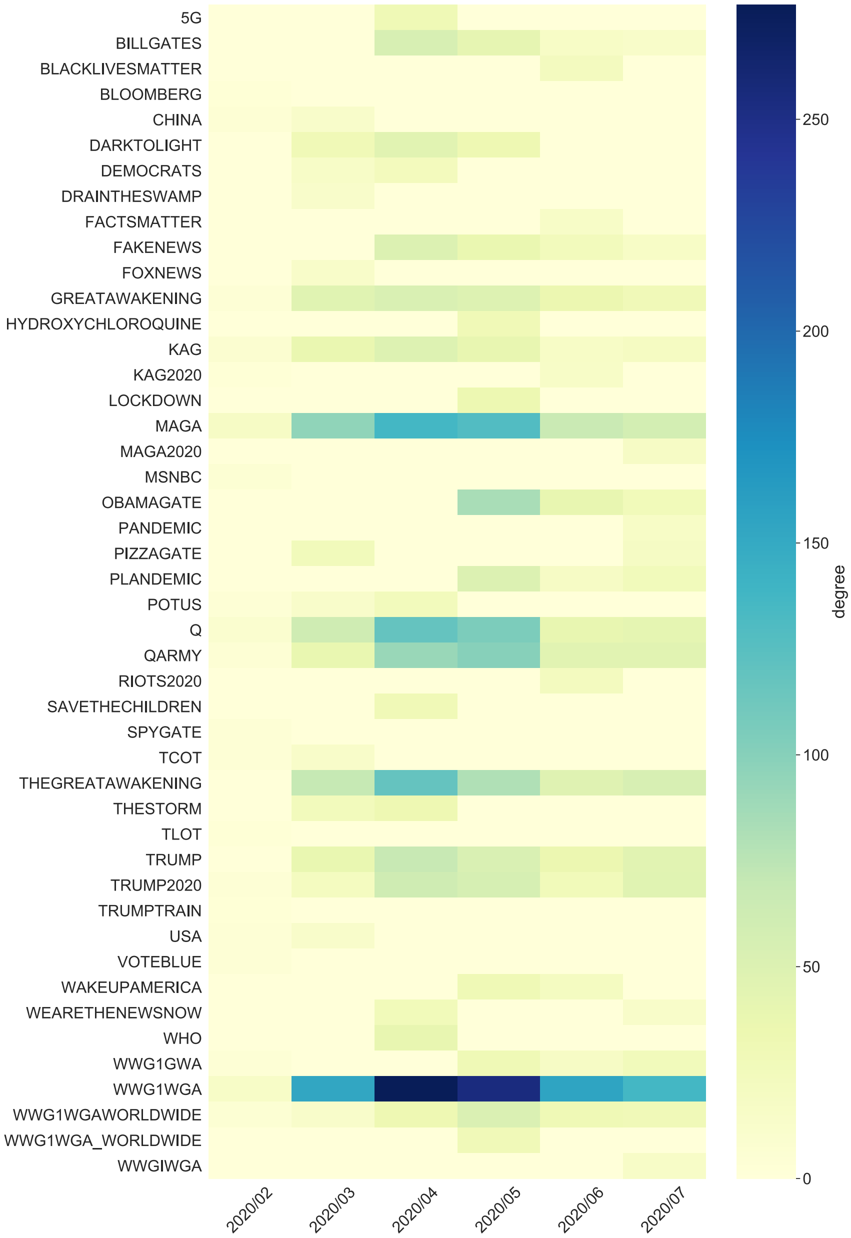


Figure 3.9: Temporal changes of top 20 popular hashtags. A darker hashtag indicates a higher degree.

3.4 Discussion

The findings of this study are based on a simple network-based method and explain implications for countering the QAnon movement. To answer RQ1-1, the pro-QAnon cluster is much larger than the anti-QAnon cluster, despite the fact that more than 50% of pro-QAnon users were suspended. A notable finding is that the numbers of pro- and anti-leaning users were both peaked in April 2020, but then pro-leaning users monotonically decreased whereas anti-leaning users increased again in July 2020. In addition, late less-leaning users were mostly anti-QAnon users. The phenomenon suggested that Twitter's policy might contribute to curbing the QAnon movement.

However, just removing malicious users is not enough to prevent pro-QAnon users and keep other users from attracting various pro-QAnon content. For instance, QAnon-related topics (hashtags) still existed in the Twitter community and the topics have evolved to a larger conspiracy theory umbrella. It is often difficult to identify "malicious users." For example, a person who is anti-QAnon might retweet or share a pro-QAnon post to show that they don't agree with it, and content-based algorithms might mistakenly label them as non-credible, if only the retweeted contents are considered. An alternate strategy would be educating anti- and less-leaning users by displaying credible information sources at the appropriate time to promote their participation, while simultaneously suspending excessively pro-leaning users. If we can better communicate with a comparable emotional tone and objective perspective, less-leaning users are more likely to convert their positions to the anti-QAnon side.

Regarding RQ1-2, QAnon has evolved into a broad and diversified conspiracy theory umbrella. Compared to other extremist groups, QAnon lacks a defined organizational structure and a concentration of interpretative functions, according to previous research [194]. However, during the COVID-19 infodemic, QAnon became a prevalent conspiracy theory. QAnon has been extended to countries such as France, Spain, Italy, and Japan, in addition to the United States. In addition, topics linked to human rights, such as "#LGBT" and "#BLACKLIVESMATTER," as well as COVID-19-related issues, such as "#STAYHOME" and "#SOCIALDISTANCING" were identified as well. These findings showed that QAnon has been expanding in a semantic network, forming a larger conspiracy theory umbrella.

It is suggested here that neutral users, such as less-leaning users and "a slight majority," may play a significant role in the formation of the QAnon conspiracy theory. SNS platforms must safeguard themselves against an overwhelming number of pro-QAnon groups. In addition, current popular SNS recommender algorithms are popping up similar voice and topics to a user based on the user's preference. This kind of SNS recommender algorithm might make the echo-chamber effect more severe. It is necessary to reconsider an experimental SNS recommender algorithm from scratch, which is able to inform malicious users and less-leaning users of the truth of a conspiracy theory.

From the perspective of psychology, a "backfire effect" happens, when people are increasingly confident in some misinformation after an effort is made to correct their beliefs [195]. In order to minimize the "backfire effect" of neutral users entering the pro-QAnon cluster, it is necessary to better educate them about the nature of QAnon. As shown in this study, although the proportion of pro-leaning Twitter users has been declining, they remained the majority at a late stage. In addition, some of them may have migrated to other social media platforms and are looking for a second opportunity, while improving topic diversity to attract more users with a broader range of viewpoints.

This research employs a network-based method to capture the social and topic evolution of the QAnon movement using a simple yet effective approach. Journalists, fact-checkers, and platforms who want to construct effective counters to the QAnon movement may find the facts and insights gleaned from this helpful method.

Recall that QAnon conspiracy theory umbrella consists of "WHO," "Trump," "5G," and "Bill Gates," in which the first two are mis/disinformation-related topics and the other two are conspiracy theory-associated. In addition, Bots took a major portion of all the users related to QAnon conspiracy theory. To figure out the role of bots. The study then narrows down from the QAnon conspiracy theory umbrella down to the "WHO," "Trump," "5G," and "Bill Gates" topics.

Chapter 4

The roles of bots during the COVID-19 infodemic

4.1 Introduction

A flood of disinformation about the COVID-19 has been circulating on prominent social networking services (SNSs), which has played a significant role in misinformation propaganda. According to [104], top-down disinformation from politicians, celebrities, and other public figures accounted for 69% of overall social media activity. Furthermore, approximately 60% of COVID-19-related content on Twitter was distorted, repurposed, and rewritten, while 38% of disinformation was entirely fabricated. Politifact.com, a nonprofit fact-checking project, also mentioned that trustful and mostly trustful news about the coronavirus represented only around 10% of the overall distributed information. SNSs users tend to connect with other like-minded users, a phenomenon known as “birds of a feather” or homophily [196, 197]. Besides human users, earlier research has shown that bots (i.e., non-human users controlled by algorithms) play a critical role in disinformation propagation. Messias et al. [125] described how bots interacted with people to obtain exposure. In politics, a well-known example is the 2016 U.S. presidential election, during which bots were utilized to extensively disseminate disinformation. According to Pew Research Center, 66% of all tweeted links to famous websites were shared by bots. In addition, bots have been proposed to be essential to promote viral transmission of material from sources of low credibility and may be able to amplify disinformation [105]. While investigating the frequency and behaviors of Twitter followers from seven German parties before the 2017 political campaigns, Keller and Linger (2019) discovered that social bots increased from 7.1% to 9.9% [198]. Furthermore, bots were used to propagate and promote fake news about vaccination [199]. Gallotti, et al (2020), carried out the research on the COVID-19 infodemic and revealed the information flow paths between humans and bots [200]. A study of social media manipulation during the 2020 U.S. presidential election identified distinctions in the behavior of right-leaning and left-leaning bots [201]. Moreover, bot activity is associated with hate toxicity in more densely populated and isolated local areas in the United States and the Philippines, according to [202].

Based on this background, an essential research topic is how bots operated during the COVID-19 infodemic to spread misinformation. A retweet is a kind of information-sharing behavior by which any user can instantly share messages with their followers. A retweet may be both a useful communication tool and a self-serving action by attention seekers [203]. According to [204], a popular tweet has a fascinating context or is retweeted by an influencer. Based on the research of [205], the popularity of a user does not always suggest a high degree of influence, and vice versa, demonstrating that the popularity and impact of an influencer are loosely associated. In addition, it is believed that a user's contextual information (e.g., social network topology, tweet content, URLs) influences retweeting behavior [206–208]. Retweets employed in this study sheds light on how COVID-19 disinformation is spread in an information environment populated by bots.

Misinformation is divided into numerous forms, one of which is conspiracy theory [209]. A conspiracy theory has the negative consequence of eliciting emotions such as avoidance, anxiety, wrath, and hostility, which may lead to illogical conduct [209]. For example, a local Belgian newspaper reported on the 5G conspiracy theory on January 22, claiming that a local doctor suggested that 5G might be related to the coronavirus [71, 210]. Due to this conspiracy belief, 5G cell phone towers in the United Kingdom have been set on fire [211]. According to another version of this conspiracy theory, 5G modifies people's immune systems and changes DNA architecture, making them more susceptible to the coronavirus [212, 213]. Another prominent conspiracy theory focused on Bill Gates, the co-founder of Microsoft Corporation. It alleged that Gates supported the implantation of monitoring chips in individuals under the guise of forced coronavirus vaccination [214, 215]. U.S. political groups were found to have a considerable partisan bias about this conspiracy [216]; the right-wing was more inclined to believe in this conspiracy than the left-wing.

The above-mentioned conspiracy theories about 5G and Bill Gates were concentrated to analyze the propagation of misinformation during the COVID-19 infodemic. Other topics such as "WHO" and "Trump" (the 45^{th} U.S. president) were analyzed for comparison as well. These keywords were chosen because disinformation on health and politics grew during the COVID-19 infodemic. Furthermore, a recent study [217] found that Trump was the main source of fake news amid the COVID-19 pandemic.

In this study, the categorization of credible and non-credible bots in retweet networks are based on four topics. The retweet activity, as well as other attributes, are then compared among the four themes. This research helps to understand how bots contributed to the COVID-19 infodemic and provides insights into a mitigation approach.

4.2 Data and Methods

4.2.1 Data

To characterize the COVID-19 infodemic, Twitter was used as the data source. Case-insensitive COVID-19 relevant keywords, including "corona virus," "coronavirus," "covid19," "2019-nCoV," "SARS-CoV-2," and "wuhanpneumonia" were prepared in advance for data collection. Using Twitter's standard search API, 279,538,960 English posts were collected from February 20 to May 31, 2020. Of those posts, 23,1515,441 (82.8%) were retweets. As stated previously, the four fake news-related topics that included "WHO," "Trump," "Bill Gates," and "5G" were concentrated in this study. Table 4.1 shows the breakdown of this dataset. Some accounts in the dataset might be deemed harmful and were suspended based on Twitter's spam policy between the dates the data was collected the tweets and the date corresponding bot scores were calculated; these users were therefore excluded from this study.

4.2.2 Methods

According to a list of non-credible websites posted on MisinfoMe [1], and a list of non-credible news website domains issued in [218], 893 responsive domains from a total of 1,143 domains were collected and used as the non-credible domain list. A list of credible media domains given by [219] was examined and 30 (all responding) credible media sites were identified (The dataset was achieved in our published paper [50]). Furthermore, two prominent scientific publications *Nature* [2] and *Science* [3] were included to the list of credible websites. Totally, 32 credible domains were determined. Based on

[1] https://misinfo.me

[2] https://www.nature.com

[3] https://www.sciencemag.org/

Table 4.1: Overview of COVID-19 tweets by topic.

	Unique users (U)	Unique Users with Bot score (US)	Percentage (US/S)	# Tweets	# Retweets
WHO	88,719	73,704	83.1	128,016	46,650
Trump	1,125,251	947,694	84.2	5,631,459	2,322,036
5G	67,523	55,315	81.9	97,638	31,814
Bill Gates	94,584	77,896	82.3	138,042	75,885

the credible and non-credible domain lists, each tweet was labeled as "credible" if it had a URL from the credible domain list, and "non-credible" if it contained a URL from the non-credible domain list. Then, given a topic, each user was labeled as "credible" if they exclusively retweeted credible tweets and "non-credible" if they exclusively retweeted non-credible tweets. In other words, non-credible users are those who retweeted, at least one time, a URL from the non-credible domain list but have never retweeted a URL from the credible domain list.

Similarly, credible users were defined as well. Note that a user's label may vary from topic to topic. For example, in the 5G topic, an account is labeled "credible" if the user retweets credible domains only in that topic, even if the user retweets non-credible domains in other topics.

In addition, Botometer (as described in Sec. 2.3) API V3 was used to classify users into bots and humans. 0.54 was determined as the criterion in this study after several trials. This threshold means that if a user's CAP value was greater than or equal to 0.54, the user was deemed a bot; otherwise, the user was regarded as human. Users were classified into five categories for analysis based on the preceding discriminative requirements for credible/non-credible and bot/human: credible humans (CH), non-credible humans (NH), credible bots (CB), non-credible bots (NB), and other.

4.2.3 User verification by human annotators

The user categorization (i.e., CH, CB, NH, NB) mentioned in the previous section has an impact on the results of the analysis. To confirm the reliability of the classification, human annotators carried out manual verification. First, 200 accounts were randomly sampled for bot classification, with 100 labeled as bots and 100 labeled as humans. The accounts were then assigned to two human annotators. They were asked to examine each user's profile and tweets independently before labeling a user as a bot or human. The categorization consistency was then tested by Cohen's kappa (κ). $\kappa = 0.68$ suggests significant agreement between the two annotators, showing that the bot score threshold used is sufficiently reliable. Cohen's κ is interpreted as follows: 0.0-0.2 for mild agreement; 0.2-0.4 for reasonable agreement; 0.4-0.6 for moderate agreement; 0.6-0.8 for significant agreement; and 0.8-1.0 for near perfect agreement [186].

With two additional annotators, a similar verification procedure for credible / non-credible categorization was carried out. They evaluated 100 randomly selected accounts, with 50 labeled as credible and 50 labeled as non-credible, by reading their tweets and profiles, and then classified them as credible or non-credible users. Following that, Cohen's was calculated, yielding $\kappa = 0.70$, suggesting a significant agreement. With the help of manual verification, the credible / non-credible categorization based on the aforementioned criteria is also adequately reliable.

4.2.4 Retweeted behavior analysis

To investigate information spreading patterns by topic, a retweet network corresponding to each topic was generated, with nodes representing users, and a directed edge formed between the source and the target user whenever a target user is retweeted by a source user. The retweet network was then used to quantify structural properties. The retweet network, for the purpose of visibility, was visualized with bots and chosen accounts (as anchors) by using the graph layout algorithm ForceAtlas2 [189] embedded in Gephi [220]. The selected users with a large indegree were highlighted. Those users include famous politicians, well-known mainstream media, and right-wing media from the top 40 indegree users. Moreover, temporal patterns of retweet activities by topic among four types of users (i.e., CH, CB, NH, NB) were compared. Furthermore, the way bots interacted with media and celebrities using the aggregated retweet network ($n = 211$) based on the bots that appeared across the topics was quantified as well. For this, 19 credible media and celebrity accounts (CM and CC), and 12 non-credible counterparts (NM and NC) were first identified through examining their user profiles, their Wikipedia articles, their official websites, and Media Bias / Fact Check, [4] etc. Then, media and celebrity accounts were classified as: (a) those retweeted by non-credible bots only; (b) those retweeted by both credible and non-credible bots; and (c) those retweeted by credible bots only. The retweet network visualization process was the same as above.

[4] https://mediabiasfactcheck.com

4.2.5 Retweeted contents analysis

In addition to the retweeting behaviors, the retweeted content was also examined in each topic. Due to the insufficient amount of text content in retweets of the dataset, retweeted Internet publications were investigated by visiting corresponding hyperlinks or URLs. Then, the articles retweeted by credible/non-credible humans and bots were collected, respectively. The retweeted articles were characterized by their noun phrases, with their importance measured by the TF-IDF (as described in Sec. 2.5) score. For this study, the top 30 nouns from credible users and the top 30 nouns from non-credible users were sorted, and then, the nouns were combined into a single list without duplicates for each topic.

To compare relevant phrases used in articles retweeted by credible and non-credible users, TF-IDF results were summarized by another form, the laterality index (LI) [221], which is defined as follows (similar with Sec. 3.2.2):

$$LI = \frac{C - NC}{C + NC}, LI \in [-1, 1], \tag{4.1}$$

where C represents the TF-IDF score for words used in articles retweeted by credible users and NC represents phrases used in articles retweeted by non-credible users. LI compares the significance of a phrase between credible and non-credible sites. A negative LI implies that the word is associated with non-credible sites; a positive LI shows that the word is associated with credible sites; and $LI = 0$ indicates that the word is equally relevant in both sites.

4.3 Results

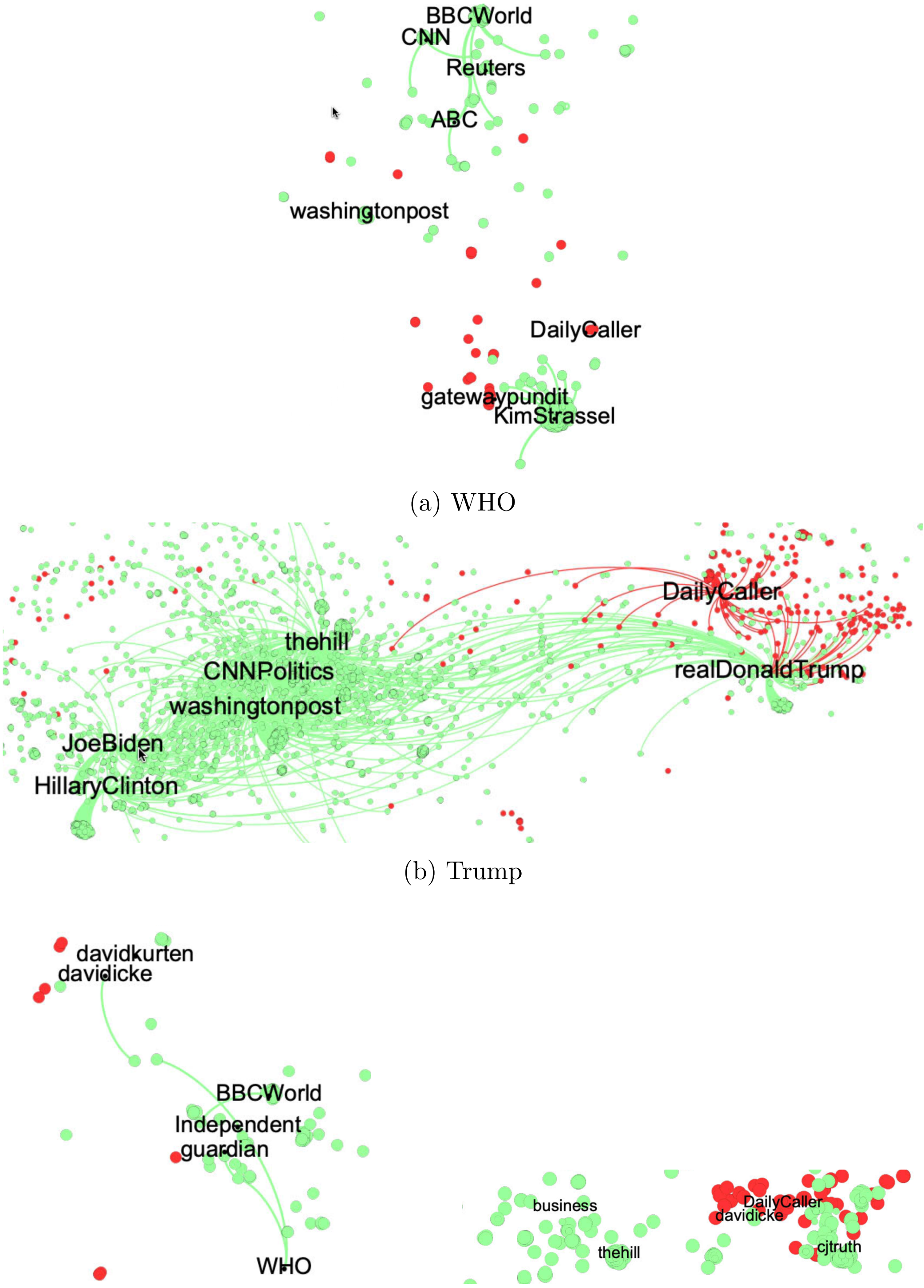
4.3.1 Retweet network structure

Using pre-processed COVID-19 retweets, the retweet interactions between humans and bots for each topic were investigated. Fig. 4.1 shows the resulting retweet networks. It is worth noting that segregated patterns occurred across all topics studied, with dense connections inside and sparse connections in between.

In the “WHO” network ($n = 88,719$), 3.8% of branded users are non-credible, whereas 19.7% are credible bots. The credible group included official media accounts such as “@washingtonpost,” “@ABC,” “@Reuters,” “@CNN,” and “@BBCWorld,” and was separated from the non-credible group, which included “@DailyCaller,” “@gatewaypundit,” and “@KimStrassel” (Fig.

4.1a). It is discovered that non-credible bots were emerging around the U.S. conservative columnist "@KimStrassel" (Kimberley Strassel), as well as "@DailyCaller" (an American right-wing misinformation website) and "@gatewaypundit" (an American right-wing misinformation website) (an American far-right website publishing misleading news). A similar result was reported in [222]. These findings suggest that the non-credible bots may be aiming to connect with politically right-leaning users in order to enhance their exposure to negative content. Although WHO itself is a neutral topic, partisan asymmetry was evident during the COVID-19 infodemic.

According to [223], the retweet network during the 2010 U.S. midterm election displayed typical "left" and "right" separated groups. To investigate whether "Trump" under the context of COVID-19 infodemic was holding comparable characteristics, the retweet network of "Trump" was generated. Fig. 4.1b depicts the Trump network ($n = 1,125,251$) with 3.2% of the labeled users being non-credible bots and 23.5% being credible bots. The number of non-credible bots and credible bots are 694 and 5,400, respectively. Here, "@HillaryClinton" (Hillary Clinton) and "@JoeBiden" (Joe Biden), representing the progressives, were separated from the conservative cluster, which included "@realDonaldTrump" (Donald Trump). In the context of the COVID-19 infodemic in 2020, the political echo chamber was detected again. A notable discovery was that "@realDonaldTrump" was largely retweeted by non-credible bots (shown in red), whereas "@HillaryClinton" and "@JoeBiden" were less so.



(a) WHO

(b) Trump

(c) 5G

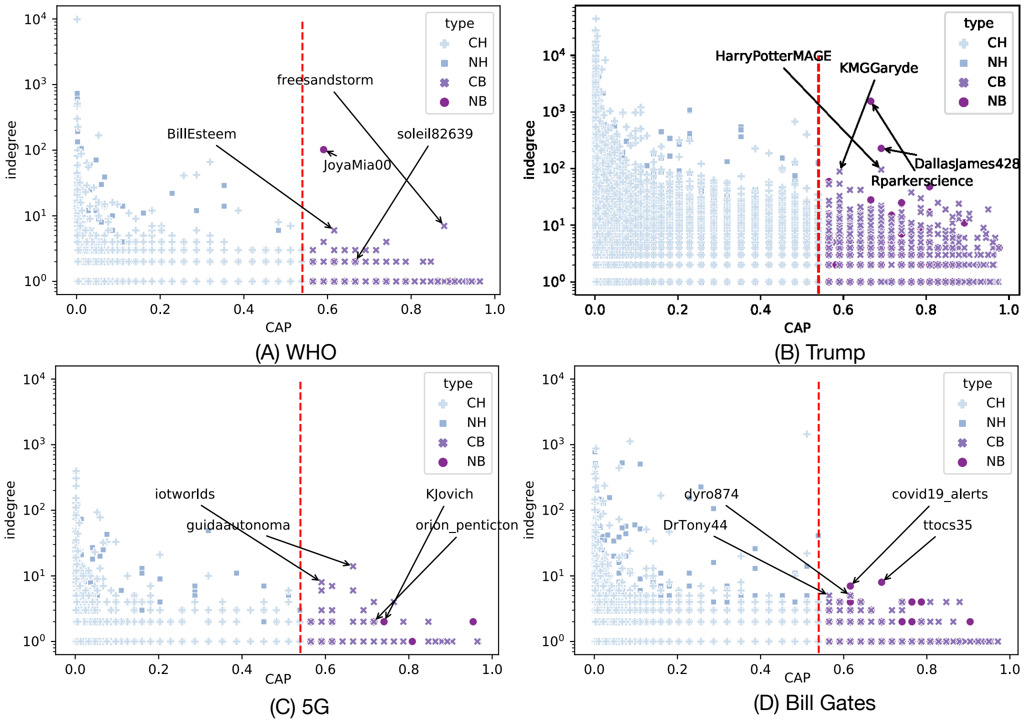
(d) Bill Gates

Figure 4.1: Retweet networks related to “WHO,”“Trump,” “5G,” and “Bill Gates.” Red nodes indicate non-credible bots; green nodes indicate credible bots. Edges represent retweets among bots, and between bots and popular accounts (labelled).

In terms of "5G," two distinct groups were observed once more in the retweet network ($n = 67,523$), with 1.62% of the labelled users being non-credible bots and 8.82% being credible bots (Fig. 4.1c). On top of the network, one can find "@davidicke" (David Icke) and "@davidkurten" (David Kurten). The former is a conspiracy theorist, while the latter has been a member of the UK Independence Party (right-wing populist party) since 2016 [224–226]. They were the two most-retweeted users of the 5G conspiracy topic. In Fig. 4.1c, the mainstream British media "@BBCWorld" and "@WHO" were located on the opposite side of the network. On the side of "@davidicke," there were more non-credible bots, while on the other side, there were more credible bots. Even though "5G" was considered to be a popular conspiracy theory in the early days of the COVID-19 pandemic, its incredible bots were less than the other topics.

"Bill Gates" is another topic that emerged as a conspiracy theory topic. Regarding this topic, 5.95% of the users labeled in the retweet network ($n = 94,584$) were not credible bots, while 18.0% were credible bots (Fig. 4.1d). Again, "@davidicke" and "@DailyCaller" were occupied by non-credible bots on the right, while "@business" and "@thehill" were among the credible bots in the cluster on the left. Furthermore, "@davidicke" was seen in both "5G" and "Bill Gates" conspiracy theory topics; this account was suspended by Twitter and is no longer available. There are no evident relationships between these bots in the "Bill Gates" network. This is because these bots and labeled accounts did not have mutual retweeting linkages; for example, "@DailyCaller" was retweeted by 336 humans, 294 of whom were non-credible humans.

Then, indegrees (the number of retweeted posts by different users, a metric of engagement) as a function of the CAP (Completed Automation Probability, a type of bot score) was measured (Fig. 4.2). The complementary cumulative distribution function (CCDF) figures of indegrees of each topic are visualized in Appendix (Fig. B.1). Most users were definitely credible humans. It turns out that indegrees tend to be inversely related to the bot score, and that indegrees for humans are, on average, greater than those for bots across all issues. There, bots were less engaging in retweets than humans in general. However, average indegrees of non-credible bots are higher than those of credible bots (t-test, $p = 0.00047$). In each topic shown in Fig. 4.2, there are bot outliers with higher indegrees. For examples, "@soleil82639" (NB) retweets messages in at least three languages, including English, Japanese, and French; these posts are relevant to history, politics, and COVID-19. "@freesandstorm" (CB) was constantly retweeting climate and environment-related tweets ("@BillEsteem" (CB) is closed) as Fig. 4.2a shows. The user, "@guidaautonoma" (CB) was a fan of



(A) WHO

(B) Trump

(C) 5G

(D) Bill Gates

Figure 4.2: Degrees vs. Bot Score (CAP) in “WHO,” “Trump,” “5G” and “Bill Gates” topics. The red dashed line is the threshold for bots/humans classification (CAP=0.54).

autonomous cars; “@orion_pentict” (NB) retweeted anti-Biden tweets, and “@KJovic” was an anti-lockdown supporter (Fig. 4.2c). In the “Bill Gates” topic, “@dyro874” (CB) has been suspended; “@DrTony44” was retweeting entertainment-related tweets; “@covid19 alert” (NB) was retweeting COVID-19-related tweets; and “@ttocs35” has been suspended (Fig. 4.2d). These non-credible bots were actually retweeted as many times as those of humans. Therefore, they were as equally influential as humans. It is confirmed that some outlier bots were actually suspended by Twitter and are no longer active. Although the number of retweets by non-credible humans may be more than the number of retweets by non-credible bots, the effects of the latter are still non-negligible because of the existence of the outliers as well as the parasitic nature of non-credible bots, which will be discussed later.

Above all, two segregated networks of information-spreading emerged in all the topics examined. It turns out that one side of the dense connected components was propagating credible sources, mostly by credible bots. By contrast, the other component was diffusing non-credible information, mainly amplified by non-credible bots.

4.3.2 Parasitic natures of bots

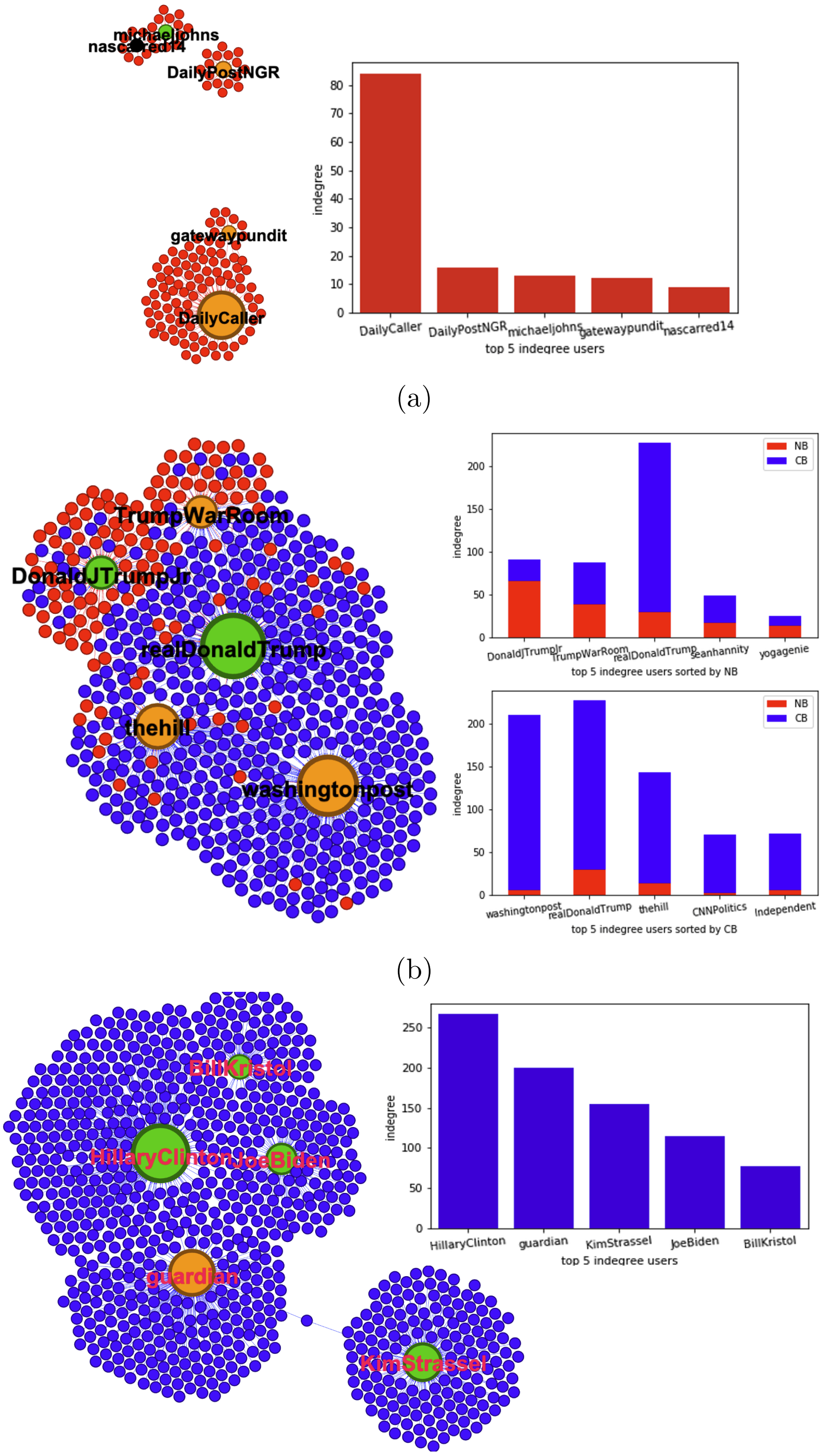
Next, how bots selectively amplified voices from media and celebrity accounts in the aggregated retweet networks mentioned in Sec. 4.2.4 was examined. Fig. 4.3 shows the results by bot category. The top five most-retweeted accounts were plotted alongside. In Fig. 4.3a, it was evident that non-credible bots were parasitic on far-right mass media accounts such as “@DailyCaller” ($k_{in} = 84$), but the indegrees of the other four were significantly lower. Note that “@DailyPostNGR” is a local newspaper in Nigeria; “@michaeljohns” is an American conservative pundit, speechwriter, and policy analyst; and “@nascarred14” was suspended.

Examining the accounts retweeted by both credible and non-credible bots in Fig. 4.3b revealed that the majority of non-credible bots were parasitic on right-leaning celebrity and media accounts, such as “@DonaldJTrumpJr” ($k_{in} = 61$). Trump’s most popular article was a repost from the website “DailyCaller” titled “Chinese Government Finally Admits Underreporting of Coronavirus Cases.” Other accounts included “@TrumpWarRoom” (“The office of Donald J. Trump”) [5], “@seanhannity” (an American talk show host and conservative political pundit), and “@yogagenie” (suspended). On the contrary, most of the credible bots amplified the voices of legitimate media accounts, such as “@washingtonpost” ($k_{in} = 205$), “@thehill,” “@CNNPolitics,” and “@Independent”. A notable exception is “@realDonaldTrump” whose posts were largely shared not only by non-credible bots ($k_{in} = 29$) but also by credible bots ($k_{in} = 199$).

Fig. 4.3c shows that major celebrities and media were selectively retweeted only by credible bots, regardless of their political stance. The celebrities included “@Kimstrassel” (conservative American author, $k_{in} = 155$) and “@HillaryClinton” (American politician, $k_{in} = 267$). The other accounts included “@guardian” (a British newspaper), “@JoeBiden,” and “@BillKristol” (an American neoconservative political analyst).

These results further support that non-credible bots are not neglectable in the context of the COVID-19 infodemic due to their parasitic natures toward popular right-leaning users.

[5] `https://donaldjtrump.com`



(a)

(b)

(c)

Figure 4.3: Celebrities and media accounts retweeted by (a) non-credible bots only; (b) both credible and non-credible bots; (c) credible bots only. Green denotes celebrities, and orange denotes media. Accounts with black labels were all suspended by Twitter (as of Aug. 7, 2021). The size of a node is proportional to its indegee.

4.3.3 Temporal patterns of retweets in humans and bots

It is assumed that the behaviors of non-credible bots are correlated with non-credible humans rather than credible humans, because the intention of non-credible bots would be to amplify the spread of misinformation including conspiracy theories. Thus, the temporal patterns of human and bot retweet behaviors were quantified to verify the hypothesis. Daily retweet counts were scaled within $[0, 1]$ for comparisons among credible/non-credible humans and bots. Fig. 4.4 shows the daily series of retweets by humans and bots for each topic, in which the patterns of retweets increase following similar trends.

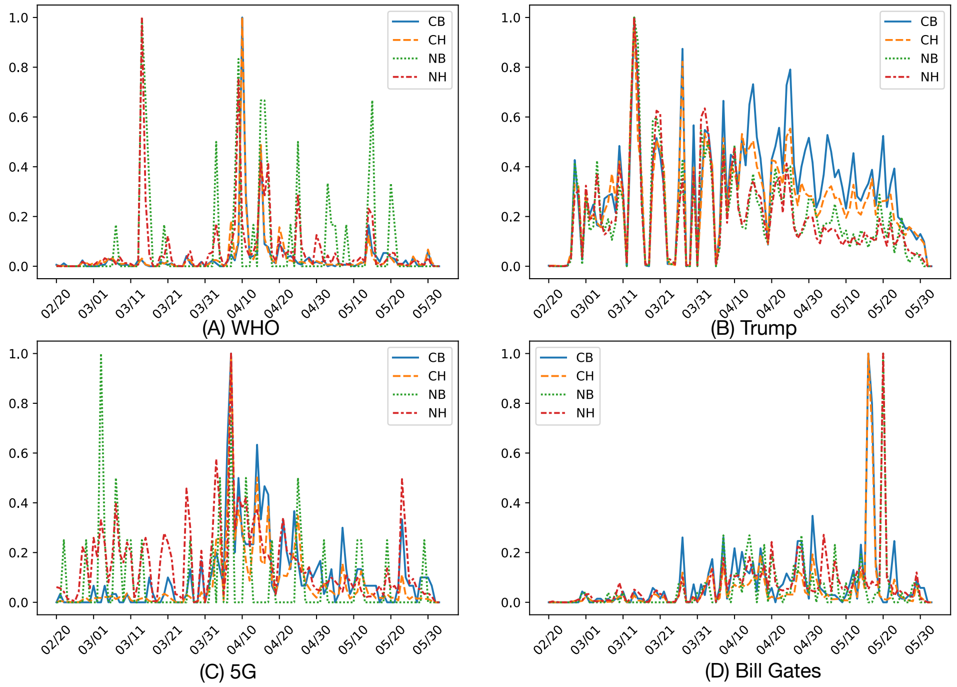


Figure 4.4: Retweet count series (scaled 0-1) generated for bots and humans in four topics.

To statistically confirm this observation, the Pearson correlation coefficient of these users' retweet temporal oscillations were calculated. The findings are summarized in Table 4.2. This demonstrates that non-credible bot retweets correlated with non-credible human retweets to a much higher degree than credible human retweets across all topics. Based on the analysis result of Pearson correlation coefficient, the preceding assumption that

Table 4.2: Pearson's correlation coefficients of retweets between NB and NH and between NB and CH.

Topic	Type	Correlation Coefficient	P-value
WHO	NB & NH	0.84	6.05E-29
	NB & CH	0.15	0.1250527
Trump	NB & NH	0.96	4.65E-61
	NB & CH	0.82	1.81E-26
5G	NB & NH	0.45	1.31E-06
	NB & CH	0.32	0.001031208
Bill Gates	NB & NH	0.91	1.46E-41
	NB & CH	0.04	0.712220673

the behaviors of non-credible bots are correlated with non-credible humans rather than credible humans, is partially supported. To further verify the correlation between NB and NH, CB and CH, commonality in retweets generated by bots and humans were further considered in the following section.

4.3.4 Commonality in retweets by humans and bots

Finally, terms (nouns), domains (URLs), and users that appeared frequently in retweets generated by humans and bots were examined. Take the 5G topic for example. Fig. 4.5 compares the importance of terms (as measured by TF-IDF) of 5G-related articles retweeted by bots and humans. The non-credible bots selectively retweeted articles that included China-related terms such as "wuhan," "china," and "'chinese," as shown in the red bars of Fig. 4.5 (a) and (b), and the non-credible humans did the same. The following are some examples of articles:

"The China Coronavirus 5G Connection is a very important factor when trying to comprehend the coronavirus (formerly abbreviated 2019-nCoV, now COVID-19) outbreak." [6]

Such articles attempted to subjectively link China with coronavirus and highlighted Wuhan as a test city for China's 5G rollout. The blue bars in Fig. 4.5 (a) and (b) show that credible bots and humans retweeted articles with the word "conspiracy." This suggests that both credible bots and humans

[6] `https://worldtruth.tv`

retweeted articles intended to alert readers to the conspiracy theory. Here is an example of the article:

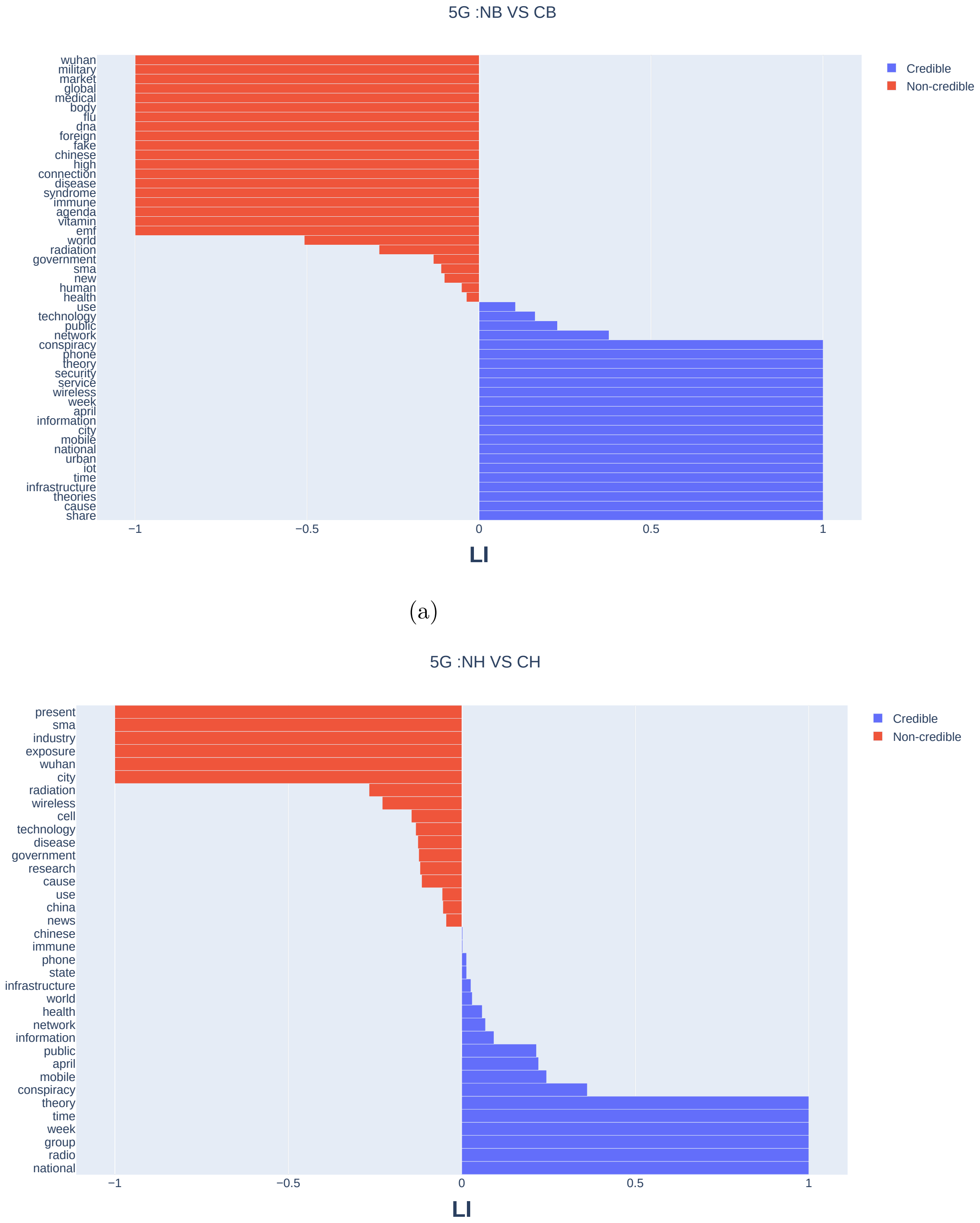
> *"One theory claims that the novel coronavirus originated in Wuhan because the Chinese city had recently been rolling out 5G. It's now supposedly spread to other cities that are also using 5G. These false conspiracy theories neglect to mention that a highly contagious virus would naturally spread more in densely populated cities with access to 5G, and that the coronavirus pandemic has hit countries like Iran and Japan where 5G isn't in use yet."* [7]

These results suggested that the 5G conspiracy theory spread within an echo chamber fabricated by non-credible users, while shutting down criticizing voices from a credible cluster.

The same linguistic analysis was performed inlcuding, "WHO," "Trump," and "Bill Gates," yielding similar linguistic features (Appendix B Fig. B.2) That is, in the Trump topic and the others, articles containing China-related terms were preferentially shared by both non-credible bots and humans. These consistent patterns indicate that non-credible humans were spreading China-related mis/disinformation and conspiracy theories, which were amplified by non-credible bots with political leanings.

In addition, the shared terms' ratio for each topic was calculated. Non-credible bots and humans shared 57%, 90%, 50%, and 30% of the terms in the retweeted articles for the "WHO," "Trump," "5G," and "Bill Gates," respectively (cf. credible bots and humans shared 73%, 93%, 70%, and 40% of terms, respectively). This implies that non-credible bots and humans are topically similar; hence it may be difficult to distinguish between humans and bots simply from the observation of several tweets. These findings imply that non-credible humans were disseminating China-related misinformation and false claims, and that non-credible bots were likely amplifying their effects. The ramifications will be revisited later.

[7] `https://www.theverge.com`



(a)

(b)

Figure 4.5: Term importance in retweeted articles in the "5G" topic: (a) non-credible bots vs. credible bots; (b) non-credible humans vs. credible humans. Red bars indicate word importance for non-credible users, whereas blue bars indicate credible users.

Furthermore, it is suggested that both non-credible bots and humans exhibit high commonality in retweeted domains (URLs) and retweeted users. In the 5G topic, for example, non-credible bots and humans shared many popular domains and users (Table 4.3). The same analyses were carried out for other topics, as well. Table 4.4 shows the shared domains and users that often occurred in retweets across all topics for each user type. The non-credible bots shared many in common with the top 10 retweeted domains and users, suggesting the commonality in retweet behaviors between non-credible bots and non-credible humans (as well as credible bots and credible humans). These results lend credence to the assumption that non-credible bots followed non-credible humans rather than credible humans.

4.4 Discussion

This study explored the involvement of bots by evaluating retweet networks, temporal patterns of retweets, and retweeted contents during the COVID-19 infodemic. Misinformation and conspiracy-theory-related topics were addressed for the study, such as “WHO,” “Trump,” “5G,” and “Bill Gates.” Two main conclusions are drawn from previous analyses: bots’ segmented retweet networks and their temporal and topical commonalities. In this section, it is necessary to re-examine the findings and analyze their implications.

First, retweet networks were segregated and parasitic, implying two types of voices or echo chambers in all issues (Fig. 4.1). The first symbolizes mainstream media and official institutions, while the second represents right-wing (self-)media and celebrities. The echo chamber effect may enhance mis/disinformation from non-credible sources while impeding the spread of information from legitimate sources. User indegree (Fig. 4.2) suggested that the basic influence of retweets by non-credible humans can be substantially more than that of credible bots. However, one can conclude that bots did not play a significant role in the COVID-19 infodemic as they did in prior political events, such as the 2016 U.S. presidential election. However, it is not easy to draw such an affirmative conclusion. The clustering of non-credible bots may be indicative of a partisan asymmetry and the fact that non-credible bots follow non-credible humans. In particular, non-credible bots were parasitic on far-right media and celebrity accounts, amplifying their voices (Fig. 4.3). It is speculated that non-credible bots targeted such powerful accounts in order to gain amplification from prestigious accounts with social capital and high follower counts, hence expanding their reach and engagement. Non-credible bots’ parasitic role helps to interact with accounts that may allow them to have their narratives or opinions amplified into the public scope.

Second, this study demonstrates that, for each topic, retweet actions and content were substantially associated between non-credible bots and non-credible humans (Figs. 4.4 and 4.5). It is discovered that non-credible bots distributed China-related terms (e.g., “wuhan,”“china,”“chinese”) from non-credible sites and users in the 5G and the other topics (Fig. 4.5 and Table 4.3 in Appendix). These regular trends may indicate that many (but not all) non-credible humans were disseminating China-related false claims, while non-credible bots amplified their effects. It is suggested here that Trump’s racism slander phrase “China virus” aiming at Chinese and Asian Americans encouraged Asian hate and Sinophobia (anti-Chinese emotions) across Western nations. Therefore, it is suggested that bots were used to distribute such lies during the early stages of the COVID-19 infodemic. Even though many malicious accounts were suspended on Twitter during the COVID-19 infodemic, sophisticated bots remained active and selectively parasitic on the partisan clusters. Overall, these results suggest that the contribution of non-credible bots in the COVID-19 infodemic should not be ignored.

The role of bots as weapons for online manipulation and political interference should be reevaluated in light of this evidence. This means that the information ecosystem of bots needs to be constantly examined. This is especially important when it is necessary to figure out the bots’ synergetic mechanism for mis/disinformation diffusing, which is not investigated in the current settings. But this could still be a real threat to society in the future. The WHO mentioned that an infodemic is a “second disease” that came out simultaneously as COVID-19. It is important to take action right away to stop this infodemic. As this study demonstrates, social media analysis is crucial to get an overview of an infodemic and gain insights into a mitigation strategy.

This study has some limitations that need to be solved in future research. The study did not obtain a complete picture of how users interact with each other because Twitter has been removing “malicious” users. In addition, more information about the sources of credible and non-credible domains (URLs) is necessary to label more domains in the study. A collective effort is needed to address the availability of a credible/non-credible domain list. Despite these limitations, this study contributes to our understanding of the role that bots play in the propagation of misinformation during an infodemic during a global healthcare crisis. It also highlights the need to develop an effective method to deal with the behaviors of malicious bots.

Table 4.3: Popular retweeted domains and users in the 5G topic. (a) Top 10 domains retweeted by credible humans (CH), credible bots (CB), non-credible humans (NH), and non-credible bots (NB). (b) Top 10 users retweeted by CH, CB, NH, and NB. Green denotes credible domains or users; red denotes non-credible domains or users; blue denotes unknown domains or users.

(a) Top 10 domains retweeted by

	NB	CB	NH	CH
1	beforeitsnews.com	bbc.co.uk	worldtruth.tv	theguardian.com
2	dailypost.ng	theguardian.com	express.co.uk	bbc.co.uk
3	worldtruth.tv	bbc.com	infowars.com	bbc.com
4	zerohedge.com	reuters.com	beforeitsnews.com	theverge.com
5	dailyrecord.co.uk	theverge.com	humansarefree.com	businessinsider.com
6	today.ng	cnn.com	neonnettle.com	cnn.com
7	infowars.com	businessinsider.com	thelastamericanvagabond.com	reuters.com
8	banned.video	nytimes.com	dailycaller.com	ft.com
9	thetruthaboutcancer.com	newsweek.com	paulcraigroberts.org	vox.com
10	rt.com	vox.com	thesun.co.uk	nytimes.com

(b) Top 10 users retweeted by

	NB	CB	NH	CH
1	@shinethelight17	@Reuters	@WorldTruthTV	@guardian
2	@DailyPostNGR	@guidaautonoma	@BILDERBERG_GP	@rooshv
3	@Laurel700	@Exchange5g	@davidicke	@guardiannews
4	@davidicke	@rooshv	@shinethelight17	@verge
5	@freezerohedge	@HaroldSinnott	@TornadoNewsLink	@Omojuwa
6	@NigeriaNewsdesk	@verge	@boblister_poole	@Reuters
7	@BANNEDdotVIDEO	@ipfconline1	@DailyPostNGR	@davidicke
8	@WorldTruthTV	@Shirastweet	@BANNEDdotVIDEO	@Exchange5g
9	@owhy3	@nuskitconsultan	@davidkurten	@davidkurten
10		@guardian	@buttscornershop	@ruskin147

Table 4.4: Common domains and users retweeted by non-credible users (NB and NH) and by credible users (CB and CH).

(a) Domains

Topic	**NB ∩ NH**	**CB ∩ CH**
WHO	6	8
Trump	10	8
5G	3	9
Bill Gates	6	8

(b) Users

Topic	**NB ∩ NH**	**CB ∩ CH**
WHO	8	6
Trump	7	7
5G	5	4
Bill Gates	7	7

Chapter 5

Conclusions and future work

5.1 General findings

The COVID-19 pandemic continues to evolve, and we do not know when the pandemic will be under control. Generic SNS users continue to suffer anxiety about coronavirus and the negative mental and emotional impacts of lockdowns, and to be affected by its effects on economics, social life, daily work, traveling, education, etc. Misinformation about the virus is believed to have emerged from the pandemic and its multiple impacts. Therefore, the COVID-19 infodemic could continue for an extended period. How to cope with the infodemic is still a problem for all stakeholders including SNSs, generic SNSs users, governments, platform developers, and researchers.

Here, we first briefly recall the findings of two studies. This study analyzed the features of the early stages of the COVID-19 infodemic using network techniques and machine learning tools. The study narrows the knowledge gaps about this COVID-19 infodemic in two aspects. Regarding RQ1: What kind of conspiracy theories evolved during the COVID-19 infodemic? It is suggested that the meta conspiracy theory QAnon is evolving. But its evolution is not limited by its own contextual meaning. QAnon has stretch itself to a much broader boundary, leading to a larger meta conspiracy theory. The dynamics of trending fake news related topics, such as "WHO," "Trump" and conspiracy theory related topics, "5G" and "Bill Gates" correlate with the evolution of QAnon. For example, both QAnon and "5G" show anti-Asian bias. Furthermore, the meta QAnon conspiracy theory seems to include every popular misinformation topic, including recent emergent misinformation topics, such as "#plademic" and anti-vaccine topics. To mitigate the negative effects of these topics, it may be necessary to gently inform less biased users about the truth about QAnon to protect them from being pushed to the pro-misinformation side. This strategy of studying the QAnon strategy could also be applied to other conspiracies during the COVID-19 pandemic. These results indicate that QAnon is a major evolving meta conspiracy theory during the COVID-19 infodemic.

Regarding RQ2: Did bots play a role in disseminating conspiracy theories as part of the COVID-19 infodemic? Bots were playing a role in QAnon ($\sim$ 25% QAnon users were bots). It is proposed here that bots play a role in the propagation of conspiracy theories during the COVID-19 infodemic. It is discovered that bots tended to be located in segregated clusters on topics of conspiracy theory and misinformation. However, unlike the 2016 presidential election, bots operating in the context of the COVID-19 infodemic tend to follow humans, indicating that they are not playing a leading role. Moreover, bots are likely to be parasitic towards influencers and institutional accounts in order to get more exposure on online SNSs. As far as QAnon is concerned, bots in the pro-QAnon cluster have been decreasing, but the QAnon topics have increased, indicating that simply removing malicious users, including bots, may not be enough to mitigate the evolution of QAnon conspiracy theory. Research on the role of bots helps people to broaden the scope of understanding the nature of bots and further helps to find additional solutions to mitigate the effects of malicious bots during the pandemic.

5.2 Implications

The overall research offers a hint for mitigating misinformation during COVID-19 infodemic. To mitigate QAnon spreading online, Twitter is simply removing malicious users, but one can easily register a new account again. Twitter did not release the technical details on how they remove them. In addition to this removal policy, we might be able to ban malicious hashtags. First, malicious hashtags promote misinformation topics by hijacking. Malicious hashtags tend to hijack a popular hashtag to promote a target topic that is significantly different from its contextual meaning [227, 228]. Recall that Fig. 3.5 (Chapter 3) shows that a large portion of hashtags of the QAnon conspiracy umbrella are not related to QAnon at all, which might be explained by hashtag hijacking. These hashtags were probably hijacked by QAnon. If malicious hashtags were removed, users who are spreading similar misinformation would not be able to connect malicious users by querying those hashtags, and the possibility of touching harmful retweeted articles and misinformation, which connect some potentially malicious hashtags, could be reduced. Second, a hashtag is linguistically unique. Creating a new meaningful hashtag and making this new hashtag known to the public is more difficult than creating a new account. Additionally, as a hashtag is usually uniquely symbolized, users require longer time to realize that a new hashtag has replaced the removed one. However, the hashtag is an important feature for almost all SNSs. Arbitrary suspension of hashtags could result in unpredictable harm to a component of an SNS ecosystem, such as freedom of speech. To some extent, the harm caused by hashtag suspension could be greater than user suspension. After all, recovering a suspended hashtag could be a problem. Therefore, we need more advanced algorithms and the necessary manual work to identify "malicious" hashtags. Furthermore, it could be necessary to establish a precaution system to warn users against potentially malicious hashtags. These hashtags can be tagged with some labels to remind people of their hazards. Before adopting any further measures, we should consider ethical risks very carefully.

It is noted that recent fake information can be expressed by any type of media, e.g., DeepFake is able to produce fake videos [229], and GPT-3 can produce human-like texts [17]. These misinformation are usually content-based. Recall that Chapter 4 found out that non-credible bots were identified as being parasitic on non-credible accounts, and they always retweeted misinformation. We could use this parasitism feature to identify misinformation. For example, an SNS is able to highlight a non-credible bot with a special tag to warn ordinary users against their diffused misinformation (both retweets and tweets). Moreover, the fact that non-credible bots are parasitic on right-wing celebrities suggests that a right-wing article (could be any fake images, fake videos, texts) could be retweeted many times by non-credible bots. That is to say, the more a retweeted article is retweeted by non-credible bots, the more likely the article belongs to misinformation. This strategy might help detect any type of AI-produced misinformation. Regarding the implication of role of bots, an SNS is able to highlight a non-credible bot with a special tag of the SNS interface to warn ordinary users against their diffused misinformation (both retweets and tweets). Moreover, the fact that non-credible bots are parasitic on right-wing celebrities (Fig. 4.3, Chapter 4) suggests that a right-wing article could be retweeted many times by non-credible bots. That is to say, the more frequently a retweeted article is retweeted by non-credible bots, the more likely the article belongs to misinformation category. This might contribute to reducing the effort to identify misinformation diffused by bots.

Above all, QAnon-meta conspiracy theory has become a major component during the COVID-19 infodemic. QAnon has raised its most passionate followers. For example, Trump supporters rushed into the U.S. Capitol, deepening the U.S. social cleavages and expanding QAnon's global impact. Moreover, QAnon's harmful impact will influence less-leaning users and the innocent public to convert their bias to the violent and redial conspiracy theory. Furthermore, young people could have been attracted by irrational conspiracy theory, which has already induced a negative effect on society, such as Asian-hate violence in western countries. QAnon cannot become the major threat during the COVID-19 infodemic without the help of bots. High-in-degree bots significantly help diffuse conspiracy theories and related misinformation retweeted from humans. To get exposure as much as possible, they were parasitic on controversial accounts like Trump and even famous writers. To undermine the influence of the current infodemic, intervening in misinformation hashtags could be an alternative. It is pointed out here that the data used in the research are limited to Twitter. However, the above-mentioned implications could also be applied to other SNS platforms, such as Facebook. It is known that bots are not forbidden on Twitter, but they are on Facebook. Although Facebook has developed an "official" method for bot detection [230], parasitism and the topology of the bot retweet network identified in this research could contribute to the development of an optimized method in the future.

5.3 Future work

Recall that pro-QAnon-leaning humans and right-wing influencers were active during this infodemic. Together with non-credible bots, they tended to diffuse misinformation and conspiracy theories. This kind of behavior makes it extremely difficult for the lay public, especially for less-leaning users, to extract helpful information during the pandemic. This problem is similar to the research question: How can malicious users on social platforms be removed during an infodemic? This could be a focus of future studies after completing this Ph.D. study. Ideally, an online system that can detect malicious users on popular platforms, such as Twitter, Facebook, Weibo, and others, is expected to be developed. The system could contribute to maintaining a healthier online ecosystem and help people distinguish misinformation and conspiracy theories from credible information.

Acknowledgement

Firstly, I would like to express my sincere gratitude to my academic advisor Prof. Kazutoshi Sasahara for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D. study. Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Reiji Suzuki, Prof. Eisuke Kita and Prof. Takaya Arita, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives. My sincere thanks also go to my wife, Bin. She married me during my graduate study. She has stood by me through all my travails, my absences, my fits of pique and impatience. I thank my fellow labmates in SASAHARA Lab for the stimulating discussions, for the rushing hours we were working together before deadlines, and for all the fun we have had in the last three years. Also I thank my colleagues in the CREST project. Last but not the least, I would like to thank Nagoya University for Ph.D. student research funding support and thank the Tokai National Higher Education and Research System (THERS) for THERS Interdisciplinary Frontier Next Generation Researcher funding support, implemented by the MEXT's Challenging Research Program for Next Generation Researchers.

Bibliography

[1] C. Cioffi-Revilla, “The scope of computational social science,” in *Handbook of computational social science. Theory, Case studies and Ethics*, U. Engel, A. Quan-Haase, S. X. Liu, and L. E. Lyberg, Eds., vol. 1. Routledge, 2021, pp. 17–32.

[2] C. Cioffi-Revilla, *Introduction to Computational Social Science: Principles and Applications.* Springer Publishing Company, Incorporated, 2014.

[3] P. R. Krishnaiah, *A Handbook of Statistics.* Motilal Banarsidass Publishe, 1980, vol. 1.

[4] H. Borko, *Computer Applications in the Behavioral Sciences.* Prentice-Hall, 1962.

[5] J. R. Macnamara, “Media content analysis: Its uses, benefits and best practice methodology,” *Asia Pacific Public Relations Journal*, vol. 6, no. 1, pp. 1–34, 2005.

[6] H. D. Lasswell, “Propaganda technique in the world war,” Ph.D. dissertation, The University of Chicago, 1926.

[7] B. Berelson, *Content Analysis in Communication Research.* Free press, 1952.

[8] H. Guetzkow, “A use of simulation in the study of inter-nation relations,” *Behavioral Science*, vol. 4, no. 3, pp. 183–191, 1959.

[9] M. D. Ward and H. Guetzkow, “Toward integrated global models: From economic engineering to social science modeling,” *Journal of Policy Modeling*, vol. 1, no. 3, pp. 445–464, 1979.

[10] H. Guetzkow and L. Jensen, “Research activities on simulated international processes,” *Background*, vol. 9, no. 4, pp. 261–274, 1966.

[11] H. R. Alker Jr and R. D. Brunner, “Simulating international conflict: A comparison of three approaches,” *International Studies Quarterly*, vol. 13, no. 1, pp. 70–110, 1969.

[12] T. F. Gieryn, “Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists,” *American Sociological Review*, vol. 48, no. 6, pp. 781–795, 1983.

[13] D. Fisher, “Boundary work and science: The relation between power and knowledge,” *Theories of Science in Society*, pp. 98–119, 1990.

[14] D. E. Chubin, “State of the field the conceptualization of scientific specialties,” *The Sociological Quarterly*, vol. 17, no. 4, pp. 448–476, 1976.

[15] C. L. Palmer, “Practices and conditions of boundary crossing research work : a study of scientists at an interdisciplinary institute,” Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1996.

[16] R. Conte, N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J.-P. Nadal, A. Sanchez *et al.*, “Manifesto of computational social science,” *The European Physical Journal Special Topics*, vol. 214, no. 1, pp. 325–346, 2012.

[17] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 1877–1901.

[18] J. Wu, L. Ouyang, D. M. Ziegler, N. Stiennon, R. Lowe, J. Leike, and P. Christiano, “Recursively Summarizing Books with Human Feedback,” *arXiv:2109.10862*, 2021.

[19] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI magazine*, vol. 17, no. 3, pp. 37–37, 1996.

[20] T. G. Lima, R. C. M. Correia, D. M. Eler, C. Olivete-Jr, and R. E. Garcia, “KDD processes in non-relational data: The case of the MinerаMongo tool,” in *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, 2017, pp. 1–6.

[21] E. Eifrem, “How graph technology can map patterns to mitigate money-laundering risk,” *Computer Fraud Security*, vol. 2019, no. 10, pp. 6–8, 2019.

[22] D. V. Kute, B. Pradhan, N. Shukla, and A. Alamri, “Deep learning and explainable artificial intelligence techniques applied for detecting money laundering–a critical review,” *IEEE Access*, vol. 9, pp. 82 300–82 317, 2021.

[23] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, p. 1–135, 2008.

[24] Y. Chen, J. Elenee Argentinis, and G. Weber, “IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research,” *Clinical Therapeutics*, vol. 38, no. 4, pp. 688–701, 2016.

[25] A. Bavelas, “Communication Patterns in Task-Oriented Groups,” *The Journal of the Acoustical Society of America*, vol. 22, no. 6, pp. 725–730, 1950.

[26] P. Erdős and A. Rényi, “On random graphs I,” *Publ. Math. (Debrecen)*, vol. 6, p. 290, 1959.

[27] E. N. Gilbert, “Random Graphs,” *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959.

[28] L. J. Brent, J. Lehmann, and G. Ramos-Fernández, “Social network analysis in the study of nonhuman primates: A historical perspective,” *American Journal of Primatology*, vol. 73, no. 8, pp. 720–730, 2011.

[29] M. Oliveira and J. Gama, “An overview of social network analysis,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 99–115, 2012.

[30] J. Alman and V. V. Williams, “A Refined Laser Method and Faster Matrix Multiplication,” *arXiv:2010.05846*, 2020.

[31] J. Johnson, A. Nowak, P. Ormerod, B. Rosewell, and Y.-C. Zhang, *Non-Equilibrium Social Science and Policy: Introduction and Essays on New and Changing Paradigms in Socio-Economic Thinking.* Springer Nature, 2017.

[32] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009.

[33] S. Milojević, "Power-law Distributions in Information Science - Making the Case for Logarithmic Binning," *arXiv:1011.1533*, 2010.

[34] T. C. Schelling, "Dynamic models of segregation," *The Journal of Mathematical Sociology*, vol. 1, no. 2, pp. 143–186, 1971.

[35] S. Franklin and A. Graesser, "Is It an agent, or just a program?: A taxonomy for autonomous agents," in *Intelligent Agents III Agent Theories, Architectures, and Languages*, J. P. Müller, M. J. Wooldridge, and N. R. Jennings, Eds. Berlin, Heidelberg: Springer, 1997, pp. 21–35.

[36] S. Bandini, S. Manzoni, and G. Vizzari, *Agent Based Modeling and Simulation.* New York, NY: Springer New York, 2009, pp. 184–197.

[37] M. Luck, P. McBurney, O. Shehory, and S. Willmott, *Agent Technology: Computing As Interaction; A Roadmap For Agent Based Computing.* University of Southampton on behalf of AgentLink III, 2005.

[38] F. Zambonelli and H. V. D. Parunak, "Signs of a revolution in computer science and software engineering," in *International Workshop on Engineering Societies in the Agents World.* Springer, 2002, pp. 13–28.

[39] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 3, pp. 7280–7287, 2002.

[40] C. D. Broad, *The Mind and its Place in Nature.* Routledge, 2008.

[41] E. Manley and T. Cheng, "Understanding road congestion as an emergent property of traffic networks," in *Proceedings of the 14th WMSCI*, Orlando, FL, USA, 2010.

[42] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann *et al.*, "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.

[43] B. Goodman and S. Flaxman, "European union regulations on algorithmic decision-making and a "right to explanation"," *AI Magazine*, vol. 38, no. 3, pp. 50–57, 2017.

[44] D. M. Lazer, A. Pentland, D. J. Watts, S. Aral, S. Athey, N. Contractor, D. Freelon, S. Gonzalez-Bailon, G. King, H. Margetts *et al.*, "Computational social science: Obstacles and opportunities," *Science*, vol. 369, no. 6507, pp. 1060–1062, 2020.

[45] M. Hannah, "QAnon and the information dark age," *First Monday*, vol. 26, no. 2, Jan 2021.

[46] C. Fuchs, "Everyday life and everyday communication in coronavirus capitalism," *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, vol. 18, no. 1, pp. 375–399, 2020.

[47] A. Önnerfors, *Conspiracy theories and COVID-19: The mechanisms behind a rapidly growing societal challenge.* The Swedish Civil Contingencies Agency, 2021.

[48] A. Bessi and E. Ferrara, "Social Bots Distort the 2016 US Presidential Election Online Discussion," *First Monday*, vol. 21, 2016.

[49] W. Xu and K. Sasahara, "A Network-Based Approach to QAnon User Dynamics and Topic Diversity During the COVID-19 Infodemic," *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, p. e17, 2022. [Online]. Available: http://dx.doi.org/10.1561/116.00000055

[50] W. Xu and K. Sasahara, "Characterizing the roles of bots on Twitter during the COVID-19 infodemic," *Journal of Computational Social Science*, vol. 5, no. 1, pp. 591–609, 2022.

[51] WHO, "Novel Coronavirus (2019-nCoV) Situation Report - 13," 2020.

[52] Asia Centre, ""INFODEMIC" AND SDGs Internet Freedoms in Southeast Asia," August 2021. [Online]. Available: https://asiacentre.org/infodemic-and-sdgs-internet-freedoms-in-southeast-asia/

[53] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on twitter during the 2016 us presidential election," *Science*, vol. 363, no. 6425, pp. 374–378, 2019.

[54] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–36, 2017.

[55] J. Roozenbeek, C. R. Schneider, S. Dryhurst, J. Kerr, A. L. Freeman, G. Recchia, A. M. Van Der Bles, and S. Van Der Linden, “Susceptibility to misinformation about covid-19 around the world,” *Royal Society Open Science*, vol. 7, no. 10, p. 201199, 2020.

[56] D. Romer and K. H. Jamieson, “Conspiracy theories as barriers to controlling the spread of covid-19 in the us,” *Social Science & Medicine*, vol. 263, p. 113356, 2020.

[57] R. Imhoff and P. Lamberty, “A bioweapon or a hoax? the link between distinct conspiracy beliefs about the coronavirus disease (covid-19) outbreak and pandemic behavior,” *Social Psychological and Personality Science*, vol. 11, no. 8, pp. 1110–1118, 2020.

[58] D. Freeman, F. Waite, L. Rosebrock, A. Petit, C. Causier, A. East, L. Jenner, A.-L. Teale, L. Carr, S. Mulhall, and et al., “Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England,” *Psychological Medicine*, vol. 52, no. 2, p. 251–263, 2022.

[59] J.-W. van Prooijen and K. M. Douglas, “Belief in conspiracy theories: Basic principles of an emerging research domain,” *European journal of social psychology*, vol. 48, no. 7, pp. 897–908, 2018.

[60] C. Benke, L. K. Autenrieth, E. Asselmann, and C. A. Pané-Farré, “Lockdown, quarantine measures, and social distancing: Associations with depression, anxiety and distress at the beginning of the COVID-19 pandemic among adults from Germany,” *Psychiatry Research*, vol. 293, p. 113462, 2020.

[61] M. Himelein-Wachowiak, S. Giorgi, A. Devoto, M. Rahman, L. Ungar, H. A. Schwartz, D. H. Epstein, L. Leggio, and B. Curtis, “Bots and Misinformation Spread on Social Media: Implications for COVID-19,” *J Med Internet Res*, vol. 23, no. 5, p. e26933, 2021.

[62] F. M. Magarini, M. Pinelli, A. Sinisi, S. Ferrari, G. L. De Fazio, and G. M. Galeazzi, “Irrational beliefs about covid-19: A scoping review,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 19, p. 9839, 2021.

[63] K. M. Douglas and R. M. Sutton, "Why conspiracy theories matter: A social psychological analysis," *European Review of Social Psychology*, vol. 29, no. 1, pp. 256–298, 2018.

[64] J.-W. van Prooijen and K. M. Douglas, "Conspiracy theories as part of history: The role of societal crisis situations," *Memory Studies*, vol. 10, no. 3, pp. 323–333, 2017, pMID: 29081831.

[65] J. E. Uscinski and J. M. Parent, *American Conspiracy Theories.* Oxford University Press, 2014.

[66] A. Cichocka, "To counter conspiracy theories, boost well-being," *Nature*, vol. 587, p. 177, 2020.

[67] K. Bierwiaczonek, J. R. Kunst, and O. Pich, "Belief in COVID-19 Conspiracy Theories Reduces Social Distancing over Time," *Applied Psychology: Health and Well-Being*, vol. 12, no. 4, pp. 1270–1285, 2020.

[68] F. Farinelli, *Conspiracy Theories and Right-wing Extremism – Insights and Recommendations for P/CVE.* EUROPEAN COMMISSION, 2021.

[69] M. Motta, D. Stecula, and C. Farhart, "How Right-Leaning Media Coverage of COVID-19 Facilitated the Spread of Misinformation in the Early Stages of the Pandemic in the U.S." *Canadian Journal of Political Science*, vol. 53, no. 2, p. 335–342, 2020.

[70] E. Pertwee, C. Simas, and H. J. Larson, "An epidemic of uncertainty: rumors, conspiracy theories and vaccine hesitancy," *Nature Medicine*, vol. 28, no. 3, pp. 456–459, 2022.

[71] W. Ahmed, J. Vidal-Alaball, J. Downing, and F. López Seguí, "COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data," *J Med Internet Res*, vol. 22, no. 5, p. e19458, 2020.

[72] N. Puri, E. A. Coomes, H. Haghbayan, and K. Gunaratne, "Social media and vaccine hesitancy: new updates for the era of covid-19 and globalized infectious diseases," *Human Vaccines & Immunotherapeutics*, vol. 16, no. 11, pp. 2586–2593, 2020.

[73] G. Eysenbach, "How to Fight an Infodemic: The Four Pillars of Infodemic Management," *J Med Internet Res*, vol. 22, no. 6, p. e21820, 2020.

[74] R. E. Park, “News as a Form of Knowledge: A Chapter in the Sociology of Knowledge,” *American Journal of Sociology*, vol. 45, no. 5, pp. 669–686, 1940.

[75] G. Overholser and K. H. Jamieson, *The press.* Oxford University Press, USA, 2005, vol. 2.

[76] L. Grossberg, E. Wartella, and D. Whitney, *MediaMaking: Mass Media in a Popular Culture.* Sage Publications, 1998.

[77] I. Schultz, “The journalistic gut feeling: Journalistic doxa, news habitus and orthodox news values,” *Journalism Practice*, vol. 1, no. 2, pp. 190–207, 2007.

[78] T. Harcup and D. O’Neill, “What is News?” *Journalism Studies*, vol. 18, no. 12, pp. 1470–1488, 2017.

[79] H. Davis and S. McLeod, “Why humans value sensational news: An evolutionary perspective,” *Evolution and Human Behavior*, vol. 24, no. 3, pp. 208–216, 2003.

[80] M. E. Grabe, S. Zhou, and B. Barnett, “Explicating Sensationalism in Television News: Content and the Bells and Whistles of Form,” *Journal of Broadcasting & Electronic Media*, vol. 45, no. 4, pp. 635–655, 2001.

[81] P. H. Vettehen, K. Nuijten, and J. Beentjes, “News in an Age of Competition: The Case of Sensationalism in Dutch Television News, 1995–2001,” *Journal of Broadcasting & Electronic Media*, vol. 49, no. 3, pp. 282–295, 2005.

[82] W. B. Frye, “A qualitative analysis of sensationalism in media,” Master’s thesis, School of Journalism, West Virginia University, 2005.

[83] L. Guo and C. Vargo, “‘Fake News’ and Emerging Online Media Ecosystem: An Integrated Intermedia Agenda-Setting Analysis of the 2016 U.S. Presidential Election,” *Communication Research*, vol. 47, no. 2, pp. 178–200, 2020.

[84] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[85] C. Martel, G. Pennycook, and D. G. Rand, “Reliance on emotion promotes belief in fake news,” *Cognitive research: principles and implications*, vol. 5, no. 1, pp. 1–20, 2020.

[86] K. D. Harber and D. J. Cohen, “The Emotional Broadcaster Theory of Social Sharing,” *Journal of Language and Social Psychology*, vol. 24, no. 4, pp. 382–400, 2005.

[87] S. Valenzuela, M. Piña, and J. Ramírez, “Behavioral Effects of Framing on Social Media Users: How Conflict, Economic, Human Interest, and Morality Frames Drive News Sharing,” *Journal of Communication*, vol. 67, no. 5, pp. 803–826, 08 2017.

[88] J. Berger and K. L. Milkman, “What Makes Online Content Viral?” *Journal of Marketing Research*, vol. 49, no. 2, pp. 192–205, 2012.

[89] The Lancet Infectious Diseases, “The COVID-19 infodemic,” *The Lancet. Infectious Diseases*, vol. 20, no. 8, p. 875, 2020.

[90] C. C. Gillispie and I. Grattan-Guinness, *Pierre-Simon Laplace, 1749-1827: A Life in Exact Science.* Princeton University Press, 2000.

[91] L. Bodenstein, “Regarding Anderegg et al. and climate change credibility,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 52, pp. E188–E189, 2010.

[92] W. R. Anderegg, J. W. Prall, J. Harold, and S. H. Schneider, “Expert credibility in climate change,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 27, pp. 12 107–12 109, 2010.

[93] X. Zhang and A. A. Ghorbani, “An overview of online fake news: Characterization, detection, and discussion,” *Information Processing & Management*, vol. 57, no. 2, p. 102025, 2020.

[94] E. C. Tandoc Jr, Z. W. Lim, and R. Ling, “Defining ‘Fake News’,” *Digital Journalism*, vol. 6, no. 2, pp. 137–153, 2018.

[95] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media: A Data Mining Perspective,” *SIGKDD Explor. Newsl.*, vol. 19, no. 1, p. 22–36, Sep 2017.

[96] G. Pennycook and D. G. Rand, “The Psychology of Fake News,” *Trends in Cognitive Sciences*, vol. 25, no. 5, pp. 388–402, 2021.

[97] F. Gu, Y. Wu, X. Hu, J. Guo, X. Yang, and X. Zhao, “The role of conspiracy theories in the spread of COVID-19 across the United States,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 7, p. 3843, 2021.

[98] A. Harris, “New Peer Reviews: Yan Report’s Claims that SARS-CoV-2 Was Created in a Chinese Lab Are Misleading and Unethical,” *Rapid Reviews COVID-19*, 2020.

[99] S. A. Baker, M. Wade, and M. J. Walsh, “The challenges of responding to misinformation during a pandemic: content moderation and the limitations of the concept of harm,” *Media International Australia*, vol. 177, no. 1, pp. 103–107, 2020.

[100] D. R. Boulware, M. F. Pullen, A. S. Bangdiwala, K. A. Pastick, S. M. Lofgren, E. C. Okafor, C. P. Skipper, A. A. Nascene, M. R. Nicol, M. Abassi *et al.*, “A randomized trial of hydroxychloroquine as post-exposure prophylaxis for covid-19,” *New England Journal of Medicine*, vol. 383, no. 6, pp. 517–525, 2020.

[101] M. S. Cohen, “Hydroxychloroquine for the prevention of COVID-19—searching for evidence,” *New England Journal of Medicine*, vol. 383, no. 6, pp. 585–586, 2020.

[102] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, “The COVID-19 social media infodemic,” *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020.

[103] B. Huang and K. M. Carley, “Disinformation and misinformation on twitter during the novel coronavirus outbreak,” *arXiv:2006.04278*, 2020.

[104] J. S. Brennen, F. M. Simon, P. N. Howard, and R. K. Nielsen, “Types, sources, and claims of covid-19 misinformation,” Ph.D. dissertation, University of Oxford, 2020.

[105] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, “The spread of low-credibility content by social bots,” *Nature Communications*, vol. 9, no. 1, p. 4787, 2018.

[106] X. Zhou, A. Jain, V. V. Phoha, and R. Zafarani, “Fake News Early Detection: A Theory-Driven Model,” *Digital Threats*, vol. 1, no. 2, 2020.

[107] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, “Automatic detection of fake news,” in *Proceedings of the 27th International Conference on Computational Linguistics.* Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018, pp. 3391–3401.

[108] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.

[109] S. M. Mohammad, P. Sobhani, and S. Kiritchenko, "Stance and Sentiment in Tweets," *ACM Trans. Internet Technol.*, vol. 17, no. 3, 2017.

[110] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, 2020.

[111] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv:1810.04805*, 2018.

[112] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[113] Y. Ren and J. Zhang, "Fake news detection on news-oriented heterogeneous information networks through hierarchical graph attention," in *International Joint Conference on Neural Networks, IJCNN 2021, Shenzhen, China.* IEEE, 2021, pp. 1–8.

[114] Y. Ren, B. Wang, J. Zhang, and Y. Chang, "Adversarial active learning based heterogeneous graph neural network for fake news detection," in *2020 IEEE International Conference on Data Mining (ICDM).* IEEE, 2020, pp. 452–461.

[115] B. Ross, L. Pilz, B. Cabrera, F. Brachten, G. Neubaum, and S. Stieglitz, "Are social bots a real threat? an agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks," *European Journal of Information Systems*, vol. 28, no. 4, pp. 394–412, 2019.

[116] I. Pozzana and E. Ferrara, "Measuring Bot and Human Behavioral Dynamics," *Frontiers in Physics*, vol. 8, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fphy.2020.00125

[117] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.

[118] S. B. Stieglitz, F. Ross, Björn;, and A. Jung, “Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts,” *ACIS 2017 Proceedings*, vol. 89., 2017.

[119] C. Freitas, F. Benevenuto, S. Ghosh, and A. Veloso, “Reverse Engineering Socialbot Infiltration Strategies in Twitter,” in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ser. ASONAM ’15. New York, NY, USA: Association for Computing Machinery, 2015, p. 25–32.

[120] O. Varol, E. Ferrara, C. Davis, F. Menczer, and A. Flammini, “Online human-bot interactions: Detection, estimation, and characterization,” in *Proceedings of the International AAAI Conference on Web and Social Media.*

[121] F. Brachten, S. Stieglitz, L. Hofeditz, K. Kloppenborg, and A. Reimann, “Strategies and influence of social bots in a 2017 german state election-a case study on twitter,” *arXiv:1710.07562*, 2017.

[122] E. Ferrara, “Disinformation and social bot operations in the run up to the 2017 french presidential election,” *arXiv:1707.00086*, 2017.

[123] S. Bradshaw and P. Howard, “Troops, trolls and troublemakers: A global inventory of organized social media manipulation,” 2017.

[124] F. Brachten, M. Mirbabaie, S. Stieglitz, O. Berger, S. Bludau, and K. Schrickel, “Threat or opportunity?—Examining social bots in social media crisis communication,” *arXiv:1810.09159*, 2018.

[125] J. Messias, L. Schmidt, R. Rabelo, and F. Benevenuto, “You followed my bot! Transforming robots into influential users in Twitter,” *First Monday*, vol. 18, 07 2013.

[126] J. Zhang, R. Zhang, Y. Zhang, and G. Yan, “On the impact of social botnets for spam distribution and digital-influence manipulation,” in *2013 IEEE Conference on Communications and Network Security (CNS).* IEEE, 2013, pp. 46–54.

[127] A. B. Onuchowska, D. J. ;, and S. Samtani, “Rocket SHIP OR BLIMP? – IMPLICATIONS OF MALICIOUS ACCOUNTS REMOVAL ON TWITTER,” *In Proceedings of the 27th European Conference on Information Systems (ECIS)*, vol. 8-14, 2019.

[128] R. R. Rout, G. Lingam, and D. V. Somayajulu, “Detection of malicious social bots using learning automata with url features in twitter network,” *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 1004–1018, 2020.

[129] C. Zhang and B. Wu, “Social bot detection using” features fusion”,” in *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*. IEEE, 2020, pp. 626–629.

[130] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, “Botornot: A system to evaluate social bots,” in *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 273–274.

[131] X. Wang, Q. Zheng, K. Zheng, Y. Sui, S. Cao, and Y. Shi, “Detecting Social Media Bots with Variational AutoEncoder and k-Nearest Neighbor,” *Applied Sciences*, vol. 11, no. 12, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/12/5482

[132] S. Kudugunta and E. Ferrara, “Deep neural networks for bot detection,” *Information Sciences*, vol. 467, pp. 312–322, 2018.

[133] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, “Aiding the detection of fake accounts in large scale social online services,” in *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, 2012, pp. 197–210.

[134] G. Wang, M. Mohanlal, C. Wilson, X. Wang, M. Metzger, H. Zheng, and B. Y. Zhao, “Social turing tests: Crowdsourcing sybil detection,” *arXiv:1205.3856*, 2012.

[135] T.-D. Nguyen, T.-D. Cao, and L.-G. Nguyen, “Dga botnet detection using collaborative filtering and density-based clustering,” in *Proceedings of the Sixth International Symposium on Information and Communication Technology*, 2015, pp. 203–209.

[136] E. Ferrara, “What types of COVID-19 conspiracies are populated by Twitter bots?” *First Monday*, 2020. [Online]. Available: http://dx.doi.org/10.5210/fm.v25i6.10633

[137] J. L. Moreno, *Who shall survive?: A new approach to the problem of human interrelations.* Washington, Nervous and Mental Disease Pub. Co., 1934.

[138] K. Musiał and P. Kazienko, “Social networks on the internet,” *World Wide Web*, vol. 16, no. 1, pp. 31–72, 2013.

[139] B. Howard, “Analyzing online social networks,” *Communications of the ACM*, vol. 51, no. 11, pp. 14–16, 2008.

[140] B. Krishnamurthy and C. E. Wills, “On the leakage of personally identifiable information via online social networks,” in *Proceedings of the 2nd ACM Workshop on Online Social Networks*, 2009, pp. 7–12.

[141] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, “Microscopic evolution of social networks,” in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 462–470.

[142] J. Golbeck, J. A. Hendler *et al.*, “Filmtrust: movie recommendations using trust in web-based social networks.” in *CCNC*, vol. 2006. Citeseer, 2006, pp. 282–286.

[143] B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythornthwaite, “Computer networks as social networks: Collaborative work, telework, and virtual community,” *Annual Review of Sociology*, vol. 22, no. 1, pp. 213–238, 1996.

[144] M. Castells, *The Internet galaxy: Reflections on the Internet, business, and society.* Oxford University Press on Demand, 2002.

[145] K. Marx, *Selected writings in sociology & social philosophy.* Maidenhead, England: McGraw Hill Higher Education, Jan. 1963.

[146] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[147] P. Held, B. Krause, and R. Kruse, “Dynamic clustering in social networks using louvain and infomap method,” in *2016 Third European Network Intelligence Conference (ENIC)*, 2016, pp. 61–68.

[148] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *J. Mach. Learn. Res.*, vol. 3, p. 993–1022, 2003.

[149] L. Waltman and N. J. Van Eck, “A smart local moving algorithm for large-scale modularity-based community detection,” *The European Physical Journal B*, vol. 86, no. 11, pp. 1–14, 2013.

[150] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.

[151] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical Review E*, vol. 76, no. 3, p. 036106, 2007.

[152] A. Lancichinetti and S. Fortunato, "Community detection algorithms: a comparative analysis," *Physical Review E*, vol. 80, no. 5, p. 056117, 2009.

[153] A. N. Wickramasinghe and S. Muthukumarana, "Social network analysis and community detection on spread of covid-19," *Model Assisted Statistics and Applications*, vol. 16, no. 1, pp. 37–52, 2021.

[154] P. Pascual-Ferrá, N. Alperstein, and D. J. Barnett, "Social network analysis of covid-19 public discourse on twitter: Implications for risk communication," *Disaster Medicine and Public Health Preparedness*, p. 1–9, 2020.

[155] F. Durazzi, M. Müller, M. Salathé, and D. Remondini, "Clusters of science and health related twitter users become more isolated during the covid-19 pandemic," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[156] F. Rosenblatt, "The perceptron - a perceiving and recognizing automaton," Cornell Aeronautical Laboratory, Ithaca, New York, Tech. Rep. 85-460-1, January 1957.

[157] F. Rosenblatt, *Two theorems of statistical separability in the perceptron.* United States Department of Commerce, 1958.

[158] F. Rosenblatt, "Perceptron simulation experiments," in *Proceedings of the IRE*, vol. 48, no. 3. IEEE, 1960, pp. 301–309.

[159] A. B. Novikoff, "On convergence proofs for perceptrons," STANFORD RESEARCH INST MENLO PARK CA, Tech. Rep., 1963.

[160] P. Werbos, "Beyond regression:" new tools for prediction and analysis in the behavioral sciences," *Ph. D. dissertation, Harvard University*, 1974.

[161] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[162] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[163] C. D. Manning and H. Schutze, *Foundations of statistical natural language processing*, ser. The MIT Press. London, England: MIT Press, May 1999.

[164] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[165] K. Papineni, "Why inverse document frequency?" ser. NAACL '01. USA: Association for Computational Linguistics, 2001, p. 1–8.

[166] J. N. Singh and S. K. Dwivedi, "Comparative analysis of idf methods to determine word relevance in web document," *International Journal of Computer Science Issues (IJCSI)*, vol. 11, no. 1, p. 59, 2014.

[167] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*, vol. 242, no. 1. Citeseer, 2003, pp. 29–48.

[168] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, vol. 26, 2013.

[169] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, pp. 2493–2537, 2011.

[170] B. Castanho Silva, F. Vegetti, and L. Littvay, "The Elite Is Up to Something: Exploring the Relation Between Populism and Belief in

Conspiracy Theories," *Swiss Political Science Review*, vol. 23, no. 4, pp. 423–443, 2017.

[171] M. Aliapoulios, A. Papasavva, C. Ballard, E. De Cristofaro, G. Stringhini, S. Zannettou, and J. Blackburn, "The gospel according to Q: Understanding the QAnon conspiracy from the perspective of canonical information," *arXiv:2101.08750*, 2021.

[172] E. Zuckerman, "QAnon and the Emergence of the Unreal," *Journal of Design and Science*, no. 6, 2019.

[173] L. Bracewell, "Gender, Populism, and the QAnon Conspiracy Movement," *Frontiers in Sociology*, vol. 5, p. 134, 2021.

[174] Pew Research Center, "5 facts about the QAnon conspiracy theories," *Research Topics, Misinformation*, 2020.

[175] S. Planck, "Where we go one, we go all: QAnon and violent rhetoric on Twitter," *Locus: The Seton Hall Journal of Undergraduate Research*, vol. 3, Article 11, 2020.

[176] S. C. Briand, M. Cinelli, T. Nguyen, R. Lewis, D. Prybylski, C. M. Valensise, V. Colizza, A. E. Tozzi, N. Perra, A. Baronchelli, M. Tizzoni, F. Zollo, A. Scala, T. Purnat, C. Czerniak, A. J. Kucharski, A. Tshangela, L. Zhou, and W. Quattrociocchi, "Infodemics: A new challenge for public health," *Cell*, vol. 184, no. 25, pp. 6010–6014, 2021.

[177] S. Phadke, M. Samory, and T. Mitra, "Characterizing Social Imaginaries and Self-Disclosures of Dissonance in Online Conspiracy Discussion Communities," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW2, Oct 2021.

[178] M. Hoseini, P. Melo, F. Benevenuto, A. Feldmann, and S. Zannettou, "On the Globalization of the QAnon Conspiracy Theory Through Telegram," *arXiv:2105.13020*, 2021.

[179] A. Anwar, H. Ilyas, U. Yaqub, and S. Zaman, "Analyzing QAnon on Twitter in Context of US Elections 2020: Analysis of User Messages and Profiles Using VADER and BERT Topic Modeling," in *DG.O2021: The 22nd Annual International Conference on Digital Government Research*, ser. DG.O'21. New York, NY, USA: Association for Computing Machinery, 2021, p. 82–88.

[180] D. T. Miller, “Characterizing QAnon: Analysis of YouTube comments presents new conclusions about a popular conservative conspiracy,” *First Monday*, vol. 26, no. 2, 2021.

[181] Antonis Papasavva, Jeremy Blackburn, Gianluca Stringhini, Savvas Zannettou, and Emiliano De Cristofaro, ““Is it a Qoincidence?”: An Exploratory Study of QAnon on Voat.” in *WWW*. ACM / IW3C2, 2021.

[182] H. W. A. Hanley, D. Kumar, and Z. Durumeric, “No calm in the storm: Investigating QAnon website relationships,” *arXiv: 2106.15715*, 2021.

[183] F. A. Chowdhury, D. Saha, M. R. Hasan, K. Saha, and A. Mueen, “Examining factors associated with Twitter account suspension following the 2020 U.S. presidential election,” *arXiv: 2101.09575*, 2021.

[184] F. A. Chowdhury, L. Allen, M. Yousuf, and A. Mueen, “On Twitter purge: A retrospective analysis of suspended users,” *Companion Proceedings of the Web Conference 2020*, 2020.

[185] C. Giatsidis, D. M. Thilikos, and M. Vazirgiannis, “D-cores: Measuring collaboration of directed graphs based on degeneracy,” in *2011 IEEE 11th International Conference on Data Mining*, 2011, pp. 201–210.

[186] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[187] M. Sayyadiharikandeh, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, “Detection of novel social bots by ensembles of specialized classifiers,” *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020.

[188] C. Sievert and K. Shirley, “LDAvis: A method for visualizing and interpreting topics,” in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014, pp. 63–70.

[189] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, “Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software,” *PLOS ONE*, vol. 9, no. 6, pp. 1–12, 06 2014.

[190] R. J. G. B. Campello, D. Moulavi, and J. Sander, “Density-Based Clustering Based on Hierarchical Density Estimates,” in *Advances in*

Knowledge Discovery and Data Mining, J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, Eds. Berlin Heidelberg: Springer, 2013, pp. 160–172.

[191] L. McInnes, J. Healy, N. Saul, and L. Grossberger, "UMAP: Uniform Manifold Approximation and Projection," *The Journal of Open Source Software*, vol. 3, no. 29, p. 861, 2018.

[192] M. Teague, "'He wears the armor of God': evangelicals hail Trump's church photo op," *The Guardian*, accessed Jul 11, 2022.

[193] S. Nazar and T. Pieters, "Plandemic Revisited: A Product of Planned Disinformation Amplifying the COVID-19 "infodemic"," *Frontiers in Public Health*, vol. 9, p. 954, 2021.

[194] A. Amarasingam and M.-A. Argentino, "The QAnon conspiracy theory: A security threat in the making," *CTC Sentinel*, vol. 13, no. 7, pp. 37–44, 2020.

[195] B. Nyhan and J. Reifler, "When corrections fail: The persistence of political misperceptions," *Political Behavior*, vol. 32, no. 2, pp. 303–330, 2010.

[196] F. Menczer, "Evolution of document networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. suppl 1, pp. 5261–5265, 2004.

[197] S. Redner, "How popular is your paper? An empirical study of the citation distribution," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 4, no. 2, pp. 131–134, 1998.

[198] T. R. Keller and U. Klinger, "Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications," *Political Communication*, vol. 36, no. 1, pp. 171–189, 2019.

[199] D. A. Broniatowski, A. M. Jamison, S. Qi, L. AlKulaib, T. Chen, A. Benton, S. C. Quinn, and M. Dredze, "Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate," pp. 1378–1384, 2018.

[200] R. Gallotti, F. Valle, N. Castaldo, P. Sacco, and M. De Domenico, "Assessing the risks of 'infodemics' in response to COVID-19 epidemics," *Nature Human Behaviour*, 2020.

[201] E. Ferrara, H. Chang, E. Chen, G. Muric, and J. Patel, “Characterizing Social Media Manipulation in the 2020 U.S. Presidential Election,” *First Monday*, vol. 25, no. 11, 2020.

[202] J. Uyheng and K. M. Carley, “Bots and Online Hate During the COVID-19 Pandemic: Case Studies in the United States and the Philippines,” *Journal of Computational Social Science*, vol. 3, no. 2, pp. 445–468, 2020.

[203] D. Boyd, S. Golder, and G. Lotan, “Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter,” *2010 43rd Hawaii International Conference on System Sciences*, pp. 1–10, 2010.

[204] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, “Measuring User Influence in Twitter: The Million Follower Fallacy,” *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4, no. 1, pp. 10–17, 2010.

[205] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, “Influence and Passivity in Social Media,” in *Machine Learning and Knowledge Discovery in Databases*, D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis, Eds. Berlin Heidelberg: Springer, 2011.

[206] L. Hong, O. Dan, and B. D. Davison, *Predicting Popular Messages in Twitter*. New York, USA: ACM Press, 2011.

[207] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, “Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network,” in *2010 IEEE Second International Conference on Social Computing*, 2010, pp. 177–184.

[208] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a Social Network or a News Media?” in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW ’10. New York, NY, USA: Association for Computing Machinery, 2010, p. 591–600.

[209] K. M. Douglas, J. E. Uscinski, R. M. Sutton, A. Cichocka, T. Nefes, C. S. Ang, and F. Deravi, “Understanding Conspiracy Theories,” *Political Psychology*, vol. 40, no. S1, pp. 3–35, 2019.

[210] J. Temperton, “How The 5G Coronavirus Conspiracy Theory Tore Through The Internet — WIRED UK,” https://www.wired.co.uk/article/5g-coronavirus-conspiracy-theoaccessedry, accessed Aug. 7, 2020.

[211] I. A. Hamilton, “77 phone masts attacked in UK due to coronavirus 5G conspiracy theory - Business Insider,” 2020, https://www.businessinsider.com/77-phone-masts-fire-coronavirus-5g-conspiracy-theory-2020-5, accessed Aug. 6, 2020.

[212] E. DisinfoLab. COVID-19 And 5G: A Case Study Of Platforms’ Content Moderation Of Conspiracy Theories. Https://t.co/s9Z8wnSSnP, accessed Apr. 21, 2020.

[213] M. Ketchell, “Coronavirus conspiracy theories are dangerous – here’s how to stop them spreading,” https://t.co/EUGn2vsHxb, accessed Apr. 21, 2020.

[214] F. C. Jack Goodman, “Coronavirus: Bill Gates ‘microchip’ conspiracy theory and other vaccine claims fact-checked - BBC News,” https://www.bbc.com/news/52847648, accessed Aug. 6, 2020.

[215] B. Jr., “Bill Gates denies conspiracy theories that say he wants to use coronavirus vaccines to implant tracking devices,” https://t.co/mclewrpRn9, accessed Aug. 6, 2020.

[216] A. Romano, “New Yahoo News/YouGov poll shows coronavirus conspiracy theories spreading on the right may hamper vaccine efforts,” 2020-09-10, https://t.co/RhbwZ16PRZ, accessed Sep. 10, 2020.

[217] S. Evanega, M. Lynas, J. Adams, and K. Smolenyak, “Quantifying Sources and Themes in The COVID-19 ’Infodemic’,” The Cornell Alliance for Science, Tech. Rep., 2020.

[218] M. Zimdars, “False, Misleading, Clickbait-y, and Satirical “News” Sources,” 2016. [Online]. Available: https://docs.google.com/document/d/10eA5-mCZLSS4MQY5QGb5ewC3VAL6pLkT53V_81ZyitM/preview

[219] J. Nørregaard, B. Horne, and S. Adalı, “NELA-GT-2018: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles,” pp. 630–638, 2019, 13th International Conference on Web and Social Media, ICWSM 2019 ; Conference date: 11-06-2019 Through 14-06-2019.

[220] M. Bastian, S. Heymann, and M. Jacomy, “Gephi: An Open Source Software for Exploring and Manipulating Networks,” *International AAAI Conference on Weblogs and Social Media*, 2009.

[221] K. Sasahara, “You are what you eat A social media study of food identity,” *Journal of Computational Social Science*, vol. 2, pp. 103–117, 2019.

[222] M. Stella, E. Ferrara, and M. D. Domenico, “Bots increase exposure to negative and inflammatory content in online social systems,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 49, pp. 12 435–12 440, 2018.

[223] M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, “Partisan asymmetries in online political activity,” *EPJ Data Science*, vol. 1, no. 1, pp. 1–19, 2012.

[224] BBC, “UKIP aiming to be ‘radical, populist’ party - Gerard Batten - BBC News,” https://www.bbc.com/news/uk-politics-45593648, accessed Aug. 28, 2020.

[225] Skynews, “Who is David Icke? The conspiracy theorist who claims he is the son of God,” https://news.sky.com/story/who-is-david-icke-the-conspiracy-theorist-who-claims-he-is-the-son-of-god-11982406, accessed Aug. 28, 2020.

[226] UropeanConservative, “David Kurten - European Conservative,” https://europeanconservative.com/authors/david-kurten/, accessed Aug. 28, 2020.

[227] C. VanDam and P.-N. Tan, “Detecting hashtag hijacking from twitter,” in *Proceedings of the 8th ACM Conference on Web Science*, 2016, pp. 370–371.

[228] P. Mousavi and J. Ouyang, “Detecting hashtag hijacking for hashtag activism,” in *Proceedings of the 1st Workshop on NLP for Positive Impact*, 2021, pp. 82–92.

[229] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, “Deep learning for deepfakes creation and detection,” *arXiv:1909.11573*, vol. 1, p. 2, 2019.

[230] T. Xu, G. Goossen, H. K. Cevahir, S. Khodeir, Y. Jin, F. Li, S. Shan, S. Patel, D. Freeman, and P. Pearce, “Deep Entity Classification: Abusive Account Detection for Online Social Networks,” in *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 2021, pp. 4097–4114.

Appendix A

Appendix for Chapter 3

A.1 Co-occurrence network clusters

The top three most popular topics are “US politics,” “News” and “Daily life,” described as follows.

A.1.1 class 0, $n = 84$, green : US politics

In addition to QAnon hashtags, such as “MAGA” (Make America Great Again), “WWG1WGA” (Where We Go One We Go All), and “WAKEUP,” political celebrities including “TRUMP,” “BILLGATES,” “JOEBIDEN” existed as well identified in the class. Misinformation hashtags such as “CONSPIRACY” ($k = 1,056$), “FAKENEWS’ ($k = 6,686$), “TRUTH” are identified as well. China-related conspiracy theory hashtags including ‘CHINAVIRUS’($k = 3,273$), “CHINESEVIRUS” ($k = 2,795$), ‘WUHANVIRUS” ($k = 2,757$) and human rights hashtags such as “BLACKLIVESMATTER” ($k = 6,390$) and “METOO” ($k = 1,073$) existed in the class as well.

A.1.2 class 2, $n = 93$, purple: News

The conspiracy-theory related hashtags, “WHO” ($k = 8,042$) and “5G” ($k = 3,364$) are spotted in the class. In addition, science-related hashtags such as “VACCINES” ($k = 10,843$), “SCIENCE” ($k = 4,698$), “RESEARCH’ ($k = 1,459$), “HEALTHCARE” ($k = 3,185$) and “CLIMATECHANGE” ($k = 3,123$) are spotted.

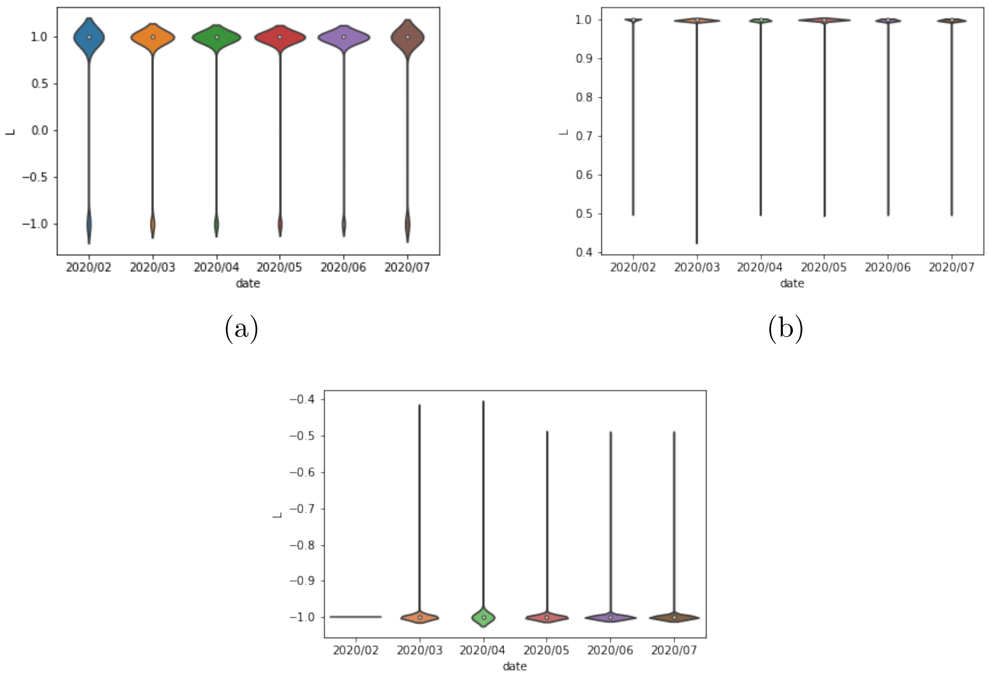
A.1.3 class 6, $n = 73$, cyan: Daily life

This class comprises people's daily life amid the pandemic including "STAYHOME" ($k = 17,960$), "SOCIALDISTANCING" ($k = 12,957$) and "QUARANTINE" ($k = 12,623$). Meanwhile, we identified religious hashtags, including "GOD" ($k = 1,112$) and "JESUS" ($k = 1,953$). Top 40 degree hashtags of modularity class 0, 2, 6 are shown in Table A.1.

Table A.1: Top 20 popular hashtags in class 0, 2, and 6 of QAnon hashtag co-occurrence network.

Rank	0	2	6
1	TRUMP	PANDEMIC	STAYHOME
2	USA	CHINA	QUARANTINE
3	COVIDIOTS	VACCINE	SOCIALDISTANCING
4	WEARAMASK	WHO	STAYATHOME
5	FAKENEWS	NEWS	MASKS
6	BLACKLIVESMATTER	HEALTH	MASK
7	AMERICA	US	NYC
8	DONALDTRUMP	UK	QUARANTINELIFE
9	WUHAN	VACCINES	TWITTER
10	MAGA	CANADA	TIKTOK
11	FLORIDA	ECONOMY	STAYHOMESAVELIVES
12	NEWYORK	HEALTHCARE	FACEMASK
13	CDC	SCIENCE	LOVE
14	BIDEN	CALIFORNIA	TRENDING
15	COVIDIOT	COVIDVACCINE	YOUTUBE
16	TEXAS	5G	MEMES
17	BLM	COVID19UK	CORONAPOCALYPSE
18	CORONAVIRUSUSA	MEDIA	THURSDAYTHOUGHTS
19	CNN	FACEMASKS	FRIDAYTHOUGHTS
20	BILLGATES	AUSTRALIA	LOCKDOWN2020

A.2 QAnon learning



(a)

(b)

(c)

Figure A.2: Distributions of QAnon-leaning (L) for (a) all users, (b) pro-leaning users, and (c) anti-leaning users. L for less-leaning users is shown in Fig. 3.3

A.3 The number of retweets

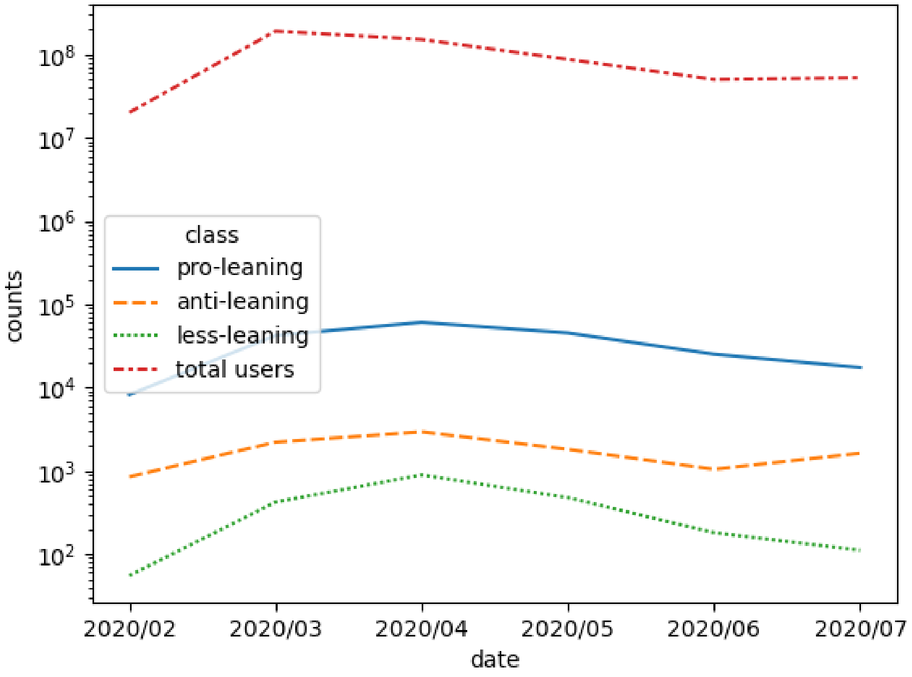


Figure A.3: The number of retweets for pro-leaning, anti-leaning, and less-leaning users and total.

A.4 LDA topic modeling analysis

Table A.2: Topics extracted using LDA

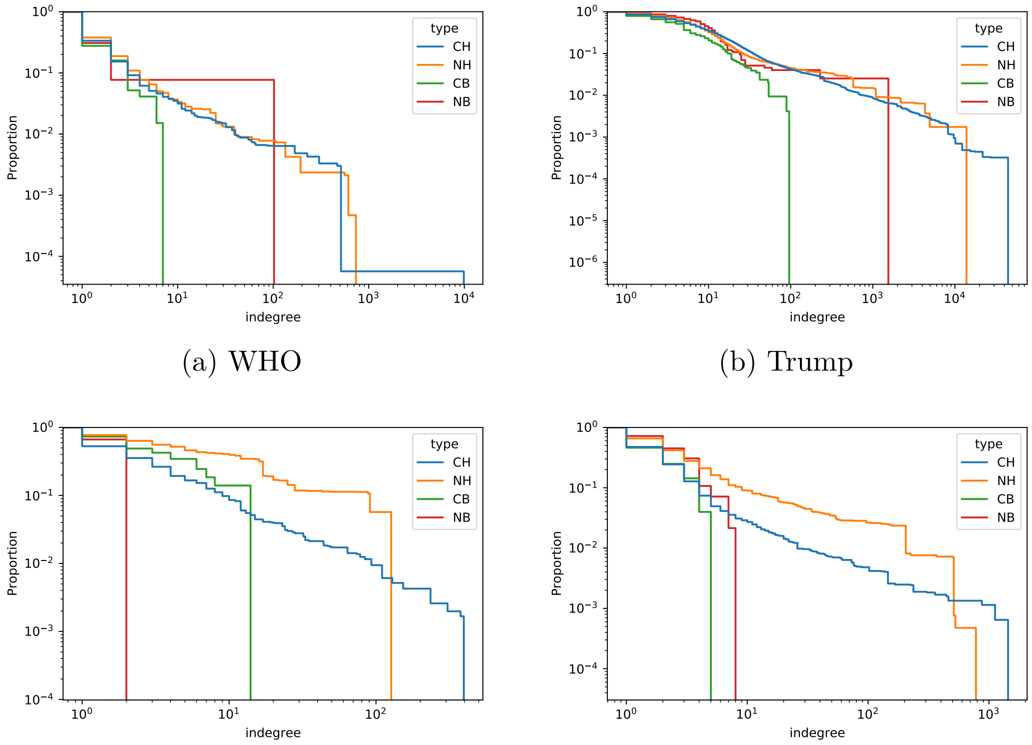
Topic ID	Topics
1	0.017*”trump” + 0.011*”health” + 0.009*”president” + 0.008*”national” + 0.007*”public” + 0.006*”guard” + 0.004*”house” + 0.004*”news” + 0.003*”world” + 0.003*”former”
2	0.010*”news” + 0.010*”april” + 0.008*”health” + 0.006*”vitamin” + 0.005*”national” + 0.005*”york” + 0.004*”trump” + 0.004*”last” + 0.004*”fox” + 0.004*”president”
3	0.010*”health” + 0.008*”trump” + 0.006*”chinese” + 0.006*”news” + 0.005*”president” + 0.005*”last” + 0.004*”government” + 0.004*”public” + 0.004*”china” + 0.004*”world”
4	0.013*”april” + 0.008*”march” + 0.007*”news” + 0.005*”member” + 0.005*”last” + 0.005*”health” + 0.005*”york” + 0.004*”december” + 0.004*”january” + 0.004*”first”
5	0.019*”health” + 0.011*”house” + 0.010*”national” + 0.007*”world” + 0.007*”committee” + 0.006*”public” + 0.006*”congress” + 0.006*”president” + 0.006*”law” + 0.006*”york”
6	0.018*”health” + 0.009*”news” + 0.007*”trump” + 0.006*”president” + 0.005*”house” + 0.005*”april” + 0.005*”county” + 0.005*”public” + 0.004*”government” + 0.004*”rate”
7	0.007*”york” + 0.006*”health” + 0.005*”april” + 0.005*”death” + 0.004*”rate” + 0.004*”trump” + 0.004*”president” + 0.004*”world” + 0.004*”news” + 0.004*”last”
8	0.005*”health” + 0.005*”april” + 0.004*”vitamin” + 0.004*”president” + 0.004*”york” + 0.003*”public” + 0.003*”news” + 0.003*”house” + 0.003*”trump” + 0.003*”first”
9	0.009*”house” + 0.007*”news” + 0.005*”care” + 0.004*”trump” + 0.004*”last” + 0.004*”york” + 0.004*”congress” + 0.004*”health” + 0.004*”april” + 0.004*”president”
10	0.007*”news” + 0.006*”april” + 0.004*”health” + 0.004*”wuhan” + 0.004*”first” + 0.003*”fox” + 0.003*”chinese” + 0.003*”president” + 0.003*”march” + 0.002*”mask”
11	0.011*”health” + 0.007*”york” + 0.006*”house” + 0.005*”president” + 0.005*”trump” + 0.005*”city” + 0.004*”government” + 0.003*”former” + 0.003*”national” + 0.003*”world”
12	0.012*”april” + 0.006*”member” + 0.006*”york” + 0.006*”health” + 0.006*”home” + 0.006*”national” + 0.005*”nursing” + 0.004*”president” + 0.004*”scholar” + 0.004*”news”

Coefficient values indicate the importance of each word in a topic.

Appendix B

Appendix for Chapter 4

B.1 CCDF



(a) WHO

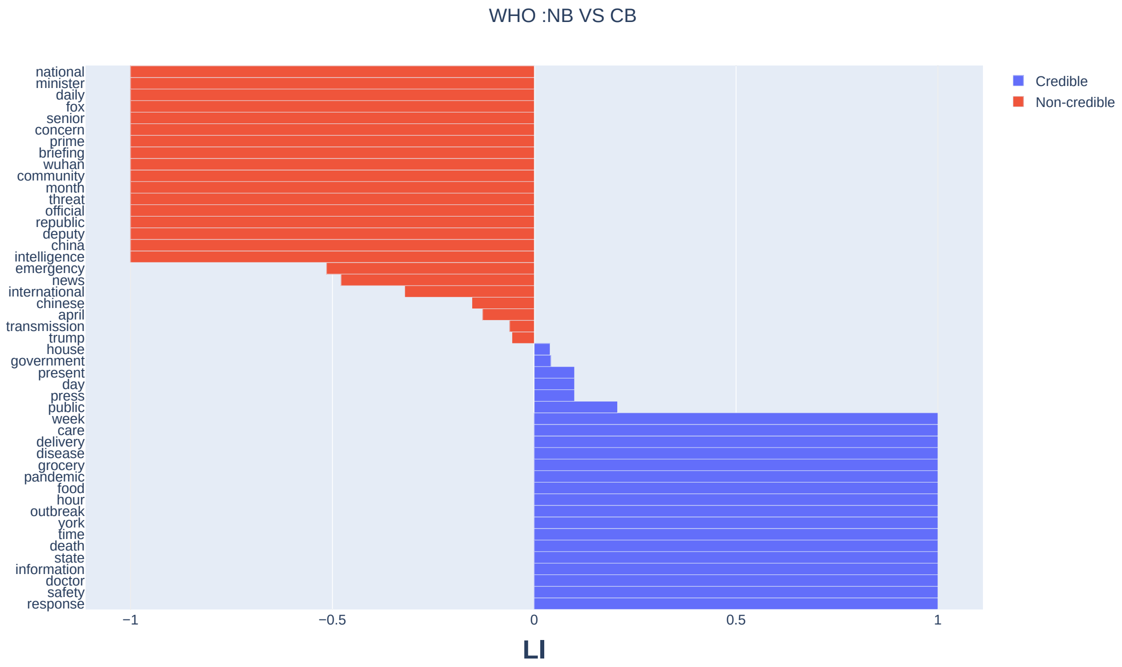
(b) Trump

(c) 5G

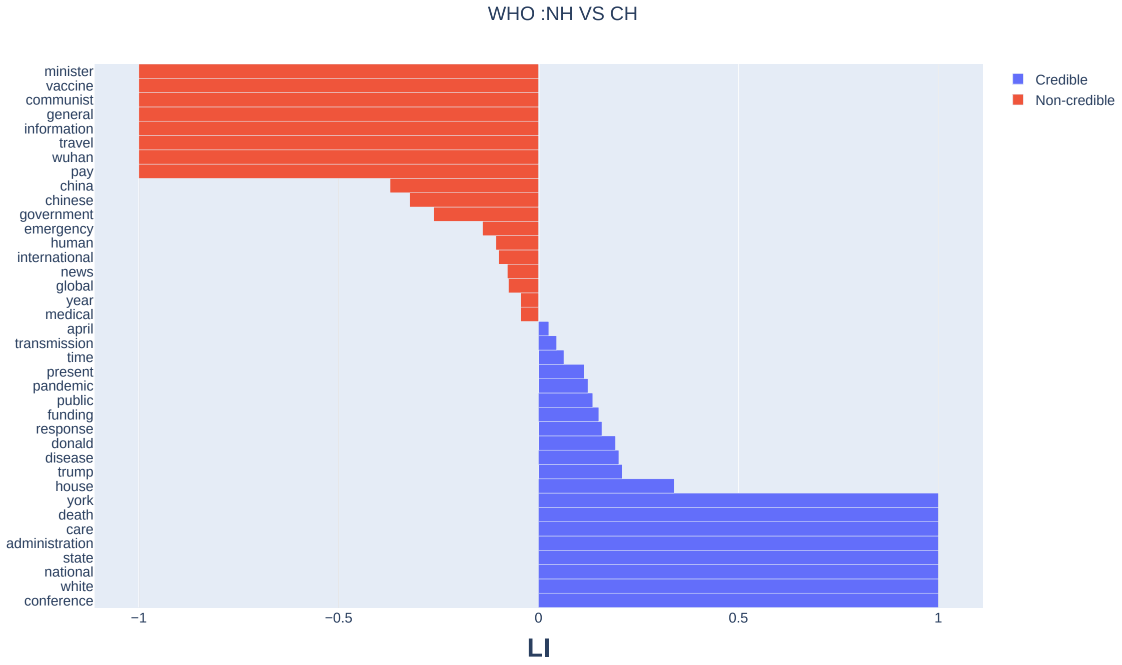
(d) Bill Gates

Figure B.1: CCDF curves for users of each topic.

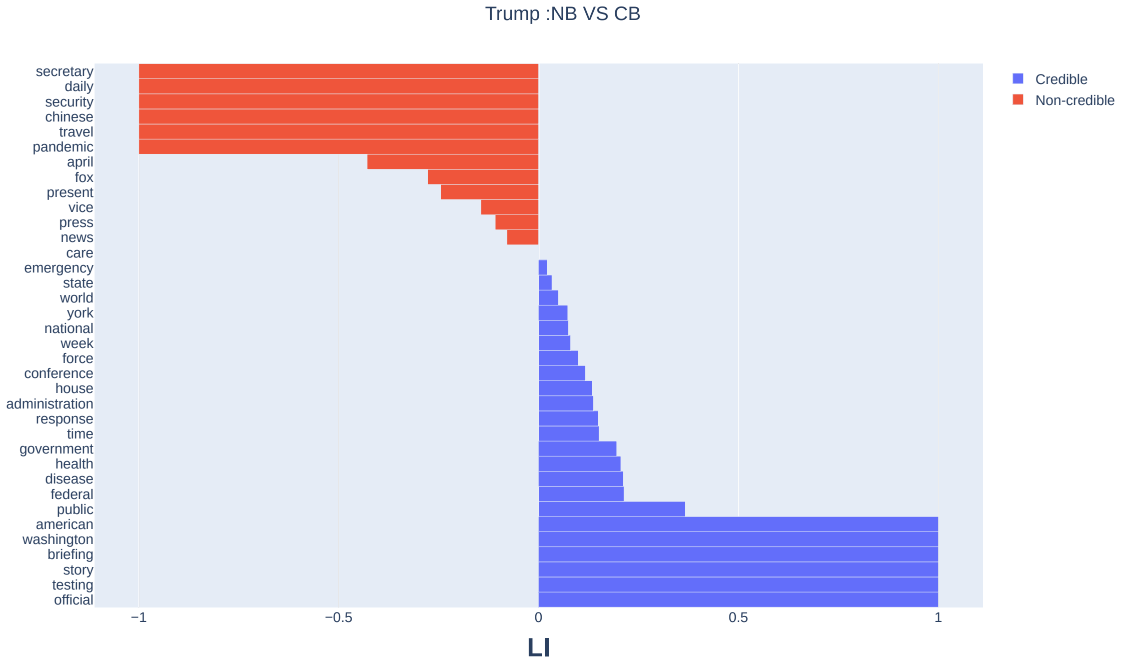
B.2 Term importance for each topic



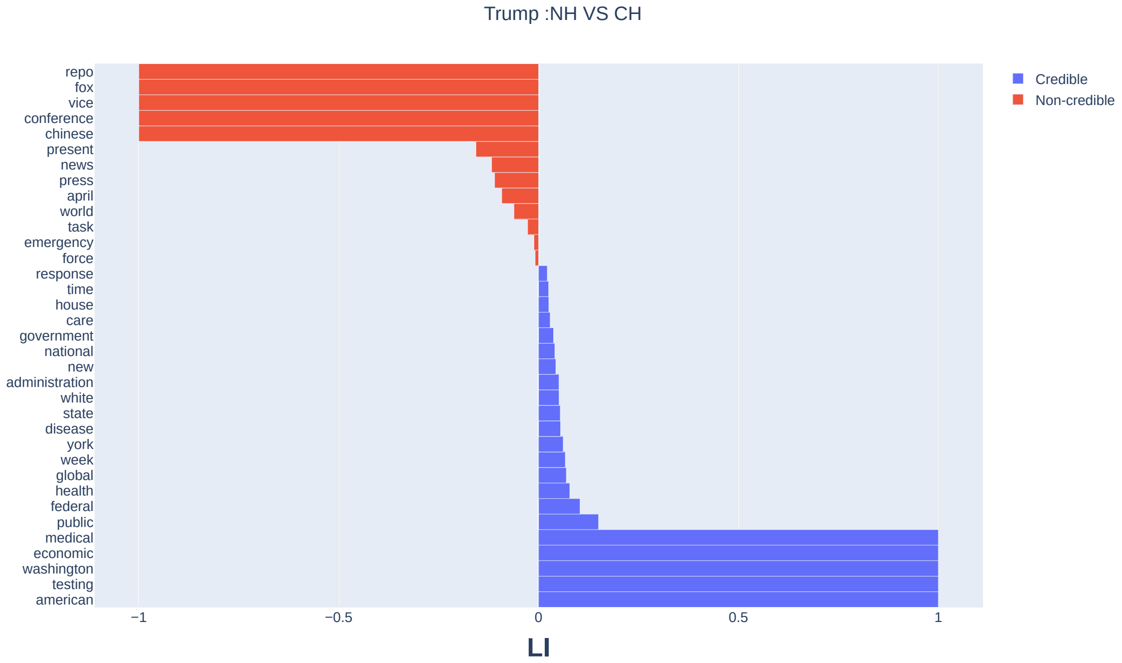
(a)



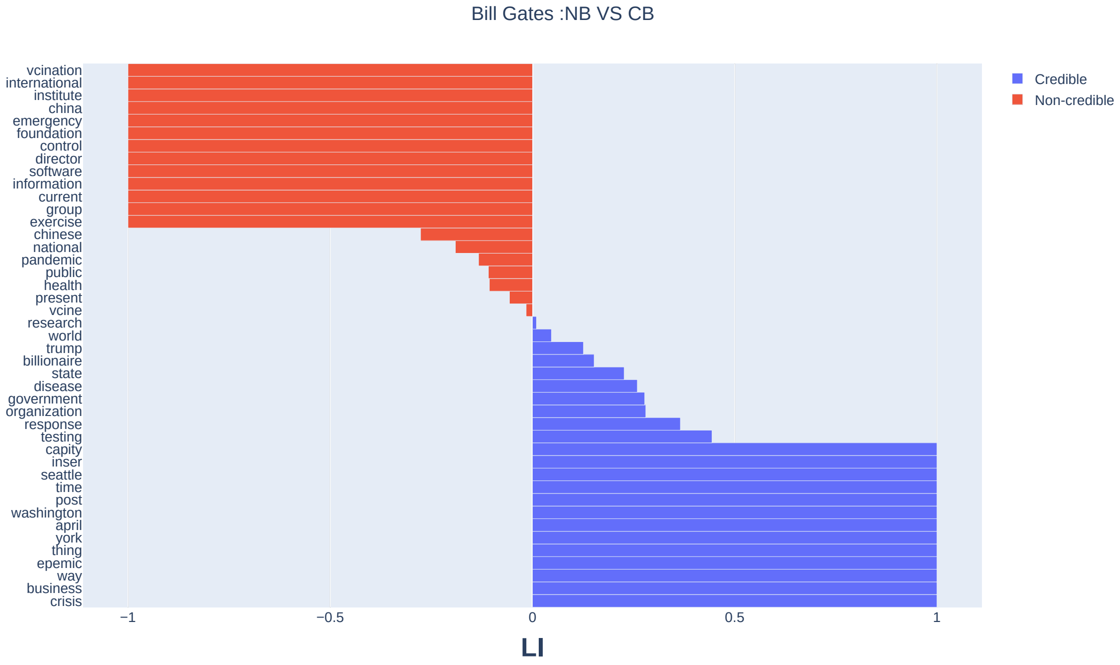
(b)


Trump :NB VS CB
secretary
daily
security
chinese
travel
pandemic
april
fox
present
vice
press
news
care
emergency
state
world
york
national
week
force
conference
house
administration
response
time
government
health
disease
federal
public
american
washington
briefing
story
testing
official
−1
−0.5
0
0.5
1
LI
Credible
Non-credible

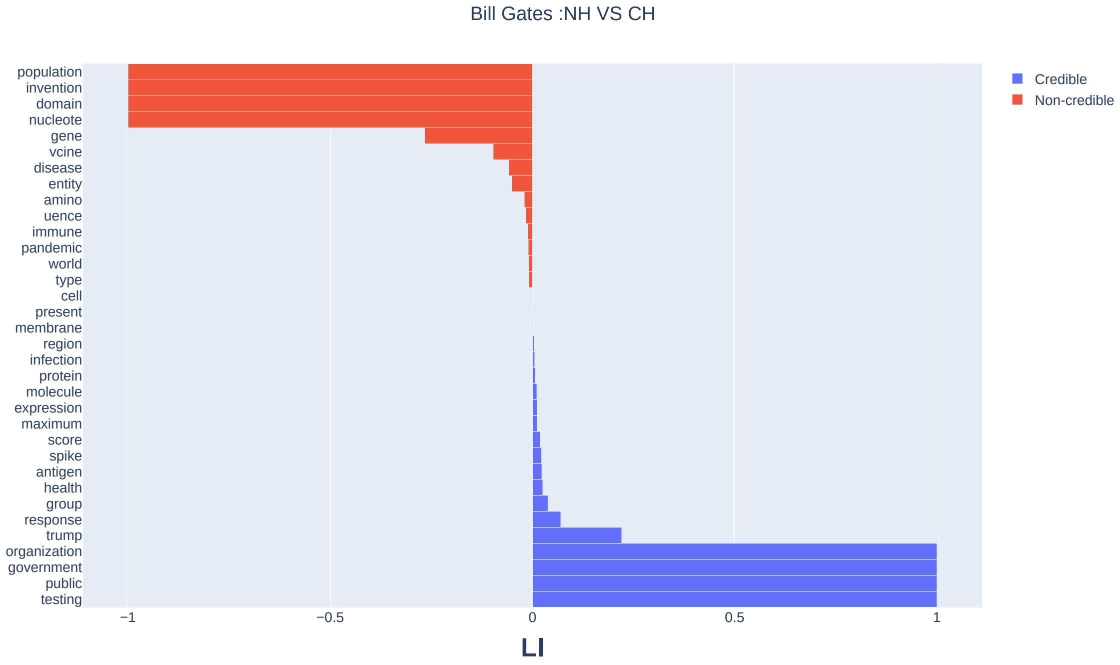
(c)


Trump :NH VS CH
repo
fox
vice
conference
chinese
present
news
press
april
world
task
emergency
force
response
time
house
care
government
national
new
administration
white
state
disease
york
week
global
health
federal
public
medical
economic
washington
testing
american
−1
−0.5
0
0.5
1
LI
Credible
Non-credible

(d)



(e)



(f)

Figure B.2: Top 30 featured terms by importance ranking of credible users and non-credible users' articles of "WHO", "Trump",and "Bill Gates" topics. Red bars indicate the term was from the articles retweeted by non-credible users; while blue bars indicate the term was from the articles retweeted by credible users. (We selected the top 30 terms from credible users category and top 30 terms from non-credible users category and then merged them without duplicates.)

B.3 Retweeted domains and users

Top 10 retweeted domains and users favored by credible/non-credible humans and bots. Green for credible users; red for non-credible users; blues for others. Column labels indicate the user type who retweeted.

Table B.1: WHO

(a) Top 10 domains retweeted by

	NB	CB	NH	CH
1	dailycaller.com	wsj.com	dailycaller.com	wsj.com
2	thegatewaypundit.com	reuters.com	thegatewaypundit.com	theguardian.com
3	wikileaks.org	theguardian.com	wikileaks.org	cnn.com
4	foxnews.com	bbc.com	presstv.com	washingtonpost.com
5	ahtribune.com	nytimes.com	ahtribune.com	nytimes.com
6	presstv.com	bbc.co.uk	foxnews.com	reuters.com
7	rt.com	cnn.com	breitbart.com	bbc.com
8	zerohedge.com	usatoday.com	justthenews.com	newsweek.com
9	saraacarter.com	washingtonpost.com	politicususa.com	bbc.co.uk
10	collective-evolution.com	foxnews.com	lifesitenews.com	justthenews.com

(b) Top 10 users retweeted by

	NB	CB	NH	CH
1	@gatewaypundit	@KimStrassel	@gatewaypundit	@KimStrassel
2	@ouchinagirl	@AmazonOmy	@wikileaks	@guardian
3	@wikileaks	@Vastuullisuus	@DailyCaller	@newtgingrich
4	@DailyCaller	@Reuters	@ouchinagirl	@jsolomonReports
5	@RoseGeorossi	@chidambara09	@Ian56789	@Newsweek
6	@Ian56789	@guardian	@PressTV	@chidambara09
7	@anuraag_saxena	@newtgingrich	@1776Stonewall	@TheElders
8	@PressTV	@jsolomonReports	@anuraag_saxena	@Reuters
9	@Doodisgirl	@ABC	@jsolomonReports	@mitchellvii
10	@RT_com	@verge	@Doodisgirl	@RWPUSA

Table B.2: Trump

(a) Top 10 domains retweeted by

	NB	CB	NH	CH
1	foxnews.com	nytimes.com	dailycaller.com	nytimes.com
2	dailycaller.com	cnn.com	foxnews.com	washingtonpost.com
3	breitbart.com	theguardian.com	breitbart.com	theguardian.com
4	dailywire.com	washingtonpost.com	justthenews.com	cnn.com
5	nypost.com	yahoo.com	dailywire.com	rawstory.com
6	justthenews.com	thehill.com	thegatewaypundit.com	thehill.com
7	thegatewaypundit.com	reuters.com	hannity.com	vox.com
8	CoronaVirus.gov	vox.com	politicususa.com	politico.com
9	politicususa.com	politico.com	nypost.com	independent.co.uk
10	hannity.com	cbsnews.com	CoronaVirus.gov	yahoo.com

(b) Top 10 users retweeted by

	NB	CB	NH	CH
1	@DailyCaller	@washingtonpost	@DailyCaller	@washingtonpost
2	@DonaldJTrumpJr	@thehill	@DonaldJTrumpJr	@thehill
3	@TrumpWarRoom	@CBSNews	@jsolomonReports	@HillaryClinton
4	@yogagenie	@guardian	@seanhannity	@TeaPainUSA
5	@realDonaldTrump	@realDonaldTrump	@TrumpWarRoom	@guardian
6	@Rparkerscience	@HillaryClinton	@marklevinshow	@JoeBiden
7	@seanhannity	@Independent	@realDonaldTrump	@BillKristol
8	@jsolomonReports	@voxdotcom	@RealJamesWoods	@Independent
9	@Bamafanaticfan1	@CNNPolitics	@thehill	@CNNPolitics
10	@thehill	@Thomas1774Paine	@Thomas1774Paine	@realDonaldTrump

Table B.3: 5G

(a) Top 10 domains retweeted by

	NB	CB	NH	CH
1	beforeitsnews.com	bbc.co.uk	worldtruth.tv	theguardian.com
2	dailypost.ng	theguardian.com	express.co.uk	bbc.co.uk
3	worldtruth.tv	bbc.com	infowars.com	bbc.com
4	zerohedge.com	reuters.com	beforeitsnews.com	theverge.com
5	dailyrecord.co.uk	theverge.com	humansarefree.com	businessinsider.com
6	today.ng	cnn.com	neonnettle.com	cnn.com
7	infowars.com	businessinsider.com	thelastamericanvagabond.com	reuters.com
8	banned.video	nytimes.com	dailycaller.com	ft.com
9	thetruthaboutcancer.com	newsweek.com	paulcraigroberts.org	vox.com
10	rt.com	vox.com	thesun.co.uk	nytimes.com

(b) Top 10 users retweeted by

	NB	CB	NH	CH
1	@shinethelight17	@Reuters	@WorldTruthTV	@guardian
2	@DailyPostNGR	@guidaautonoma	@BILDERBERG_GP	@rooshv
3	@Laurel700	@Exchange5g	@davidicke	@guardiannews
4	@davidicke	@rooshv	@shinethelight17	@verge
5	@freezerohedge	@HaroldSinnott	@TornadoNewsLink	@Omojuwa
6	@NigeriaNewsdesk	@verge	@boblister_poole	@Reuters
7	@BANNEDdotVIDEO	@ipfconline1	@DailyPostNGR	@davidicke
8	@WorldTruthTV	@Shirastweet	@BANNEDdotVIDEO	@Exchange5g
9	@owhy3	@nuskitconsultan	@davidkurten	@davidkurten
10		@guardian	@buttscornershop	@ruskin147

Table B.4: Bill Gates

(a) Top 10 domains retweeted by

	NB	CB	NH	CH
1	thegatewaypundit.com	thehill.com	thegatewaypundit.com	thehill.com
2	dailycaller.com	nytimes.com	activistpost.com	nytimes.com
3	newspunch.com	vox.com	express.co.uk	washingtonpost.com
4	nypost.com	washingtonpost.com	dailycaller.com	businessinsider.com
5	bloomberg.com	businessinsider.com	newspunch.com	vox.com
6	express.co.uk	wsj.com	nypost.com	theverge.com
7	americanthinker.com	cnn.com	worldtruth.tv	wsj.com
8	activistpost.com	theverge.com	zerohedge.com	newspunch.com
9	presstv.com	bbc.co.uk	montanadailygazette.com	cnn.com
10	naturalblaze.com	reuters.com	blacklistednews.com	thedailybeast.com

(b) Top 10 users retweeted by

	NB	CB	NH	CH
1	@EyesOnQ	@cjtruth	@o_rips	@cjtruth
2	@Ian56789	@EyesOnQ	@gatewaypundit	@EyesOnQ
3	@zsixkillerk	@joshrogin	@EyesOnQ	@carmindabrendel
4	@gatewaypundit	@WSJ	@bbusa617	@joshrogin
5	@gaye_gallops	@sapinker	@Ian56789	@WSJ
6	@taxfreeok	@HollyWilhelm4	@KarluskaP	@HollyWilhelm4
7	@SM0799	@businessinsider	@DailyCaller	@sapinker
8	@bbusa617	@carmindabrendel	@davidicke	@ezraklein
9	@DailyCaller	@voxdotcom	@WorldTruthTV	@stanveuger
10	@davidicke	@gaye_gallops	@zsixkillerk	@thehill