ABSTRACT

Title of Thesis:	EXPLORATORY GRAPH BASED BOT DETECTION APPLICATION ON REDDIT SUBNETWORKS
	Gabriel Cruz, Master of Information Management, 2021
Thesis directed by:	Professor Jennifer Golbeck University of Maryland, College of Information Studies

Methods for detecting bots have traditionally focused on implementing machine learning systems to classify abnormal behavior. We focus on abnormal term usage as a marker of bot behavior around politically charged language regarding the Covid-19 pandemic on Reddit. We look at multiplex networks abstracted from different subreddits around six terms. We then use novel measures to quantify the differences between the layers in all of these multiplex networks to detect abnormalities in term usage over time and to quantify the differences between subreddit aggregated networks. We conclude that there is not enough evidence to declare that any one term investigated demonstrated an abnormal rate of usage over time. Additionally, none of the aggregated networks demonstrated differences between them indicating that the usage of the terms themselves is not different. We hope to demonstrate the efficacy of this graph-based technique to potentially detect botnet structures on social media.

EXPLORATORY GRAPH BASED BOT DETECTION APPLICATION ON REDDIT SUBNETWORKS

by

Gabriel Cruz

Master's thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Information Management
2021

Advisory Committee:
Professor Jennifer Golbeck, Chair
Professor Beth St. Jean
Professor Ping Wang

Dedication

To my parents who have taught me everything I know, and to my brothers who inspire me to work harder and harder each day.

Table of Contents

Chapter 1: Introduction

One of the aspects of social media that makes it so attractive to users is its ability to easily connect people. Social networks make it easy for people to engage in meaningful conversations with individuals that they otherwise would not be able to engage with for any reason. During the coronavirus pandemic, social media platforms have been a tool to help people replace face-to-face interaction with a socially distanced alternative. Apart from connecting people for conversation, social media platforms allow users to align themselves with political or social movements. For example, in the case of the Black Lives Matter movement, analysis showed that Twitter users mainly engaged with Black Lives Matter hashtags for 3 reasons. To demonstrate solidarity, to help organize small grievances into a larger mass movement, and to actively engage in the counter-movement. [38]

Recently, the role that social media influences have in the political sphere has come into focus. It is well known that bots played a role in influencing public opinion during the 2016 election. [8, 14] Other movements such as the Black Lives Matter movement have provided opportunities for bot and troll accounts to try to influence public opinion. In their goal to influence public opinion, these bots align themselves with counter-movements and in effect may aid in further polarizing political ideologies. In the case of Black Lives Matter, for example, conservative social bots adopted the phrase ”All Lives Matter” to demonstrate alignment with conservative sentiment.

Political alignment tends to be a theme wherein bots make it clear that they belong to the different wings of the political spectrum. Recently with the emergence of the coronavirus pandemic, bots have taken it on themselves to spread misinformation about the virus. We should note that bots do not usually create content on their own. Instead, they spread information and links to webpages that have been posted or created by other accounts.

Because bots typically do not create their own content, we would expect bots to act a certain way, specifically, act like the political figures whose rhetoric they are taking after. We would expect them to use controversial phrases such as ”All Lives Matter” or ”Blue Lives Matter” in response to the Black Lives Matter movement. In the case of the coronavirus pandemic, we would expect them to imitate right-wing thought leaders by referring to Covid-19 with derrogatory terms such as ”China virus”.

This behavior of course is problematic since we know that repeated exposure to false information makes it more likely that a person will believe that information [51] This kind of coordinated effort aimed at shifting public opinion around different topics has its other dangers as well. It is known that certain rhetoric could increase cases of violence and prejudice against members of the Asian community as it pertains to conversations around the Covid pandemic. Being able to detect coordinated efforts for misinformation campaigns as they happen is therefore important as being the first step in mitigating the effects of misinformation or inflammatory speech.

Detecting these coordinated efforts comes with its own set of challenges. Computational methods for detecting whether an account may be a bot or not could be computationally expensive. These same computational methods tend to instead take a nuclear approach by focusing on the content of the posts, follower accounts, etc. Using these detection methods may not tell us if an entire online community is actively involved in spreading misinformation. If they are, there are additional challenges in understanding what sentiment these bots carry. Additionally, such methods may not consider the interactions of an online community around a particular topic over time. The question would then be how are those botnets spreading misinformation?

Given recent political events in the US, the importance of understanding how misinformation campaigns lead to radicalization online in real-time is becoming clearer. Given the fact that some aggressors gain their beliefs or ideologies from online forums, it is becoming clear that there is a necessity for computational methods to stop the spread of misinformation online. In some cases, new conspiracy theories can emerge or communities can shift their way of thinking overnight. For example, this was noted during the presidential inauguration when online supporters of the Qanon conspiracy theory demonstrated disappointment after their theory was proven to be false on inauguration day.

This paper aims to investigate whether network differentiation measures can be applied in the context of social media networks to find coordinated malicious behavior on Reddit based on key term usage. This paper will seek to apply novel network science algorithms to this modern problem to detect anomalies in the behaviors of these botnets. We hope that the techniques used here demonstrate the viability of using measures that quantify the differences between the layers in a multiplex network to detect network abnormalities on social media. We also hope that this paves a path for using simple methods to detect coordinated online behavior. We are going to look at two characteristics of the network structure using novel algorithms that do the following:

- Quantify differences between layers of a multiplex network.
  - This can be used to determine the stability of some terms, in this case, the term stability of some key terms in the networks over time.
- Quantify differences between networks using the topology of the network.
  - Used to determine a ”difference” matrix of the aggregated networks of the different terms.

Both of these will be used to answer the following questions:

- Is it possible to detect abnormalities in term usage rates over time within social media communities using an algorithm that quantifies node structure similarities between layers of a multiplex network?
- How do communities with dishonest moderating practices differ from widely accessible communities on Reddit when differences between networks are quantified?

We intend for this form of term and graph based exploratory analysis to supplement social network analysis in later studies. Within Reddit, we can not assume that coordinated bot behavior might be present in the first place, given its moderated structure. If we do find abnormal term rates or term usage it would then raise a flag since this would mean that moderation efforts have not caught bot behavior on the platform. Abnormal behavior might also mean high rates of term adoption between bot members of a botnet. Therefore, given the terms we are investigating, we can argue that we could expect stable usage of terms over time, and when compared between subreddits.

Our hypotheses are as follows:

- Regarding Research Question 1:
  - In regards to the *D* coefficient for determining differences between network layers:
    - H null: *D* coefficient does not differ between different monthly aggregated layers for each subreddits and each term.
    - H1: *D* coefficient is statistically different between different monthly aggregated layers for any subreddits and any term.
- Regarding Research Question 2:
  - In regards to the Network Portrait Divergence coefficient for determining differences between networks:

* H null: NPD coefficient does not differ between different aggregated layers for all subreddits in each term.
* H1: NPD coefficient is statistically different between different aggregated subreddits networks of any term.

We begin this paper by reviewing current techniques to detect social bots and how the limitations of these current methods demonstrate the need for new graph based methods for bot detection. We then review the state of the field in detecting botnets and define a place of fit for our research. Next, we elaborate on our data collection process to explain how we abstracted Reddit data to test our hypothesis. This is followed by a discussion on the findings, limitations, and implications of our results.

Chapter 2: Literature Review

2.1 Social Bot Detection

Because of the proliferation of social bots in recent years, interest in the detection of social bots has spiked. Current methods for detecting bots are usually based on the nature of bots that are being detected [50]. These techniques to detect different types of bots are divided into 4 major subgroups as described by [50]:

- Graph Based
- Machine Learning Based
- Crowd Sourcing Based
- Anomaly Based

2.1.1 Crowdsourcing Based

The simplest method to understand is Crowd Sourcing based social bot detection since this is simply the act of detecting bots with the help of people. There are clear issues with this method in that it can be expensive and may not scale well. Consider [27] wherein postgraduate students had to tag a data set of Twitter users by considering their user-profiles and determine if they are bots or not. In this study, two trained students were used to tag posts while one student was tasked to consolidate disagreements between the two. Even in that setting wherein, trained individuals are classifying bots, there is disagreement between the raters which had to be settled by a third rater later on.

2.1.2 Machine Learning Based

While we will not be using machine learning based methods for our study, it is important to understand the landscape of machine learning based detection methods and how it fits in with social botnet detection as well. Machine learning methods have become a popular way to computationally detect bots online. Consider for example BotOrNot [28], an online service that helps users determine if a Twitter account is controlled by a human or a bot. In the case of BotOrNot, a user provides a Twitter username, which is then used to determine over one thousand features about that user. Some of these features are related to the content of the tweets that a user outputs such as time between tweets, tweet sentiment, and Natural Language Processing items such as part of speech tagging. This information is fed into a Random Forest classifier which returns the likelihood that the Twitter account served is controlled by a bot.

Other models also use the behavior of accounts and content to detect bots. CATS [7] uses a very low amount of features that can detect bots within focused domains, like for example bots that post links to shopping sites. This approach makes for a scale-able approach to detecting bots that are part of botnets. Such a model is limited however in this functionality because, to apply this algorithm towards the intended purpose, we would need to know how bots are set up to function. Meaning, we need to have some context on how bots behave online already. Additionally, bots are becoming smarter in their approach to evade detection and to try to circumvent the terms of service of their respective platforms. [7]

Other approaches use the content of social media posts to predict whether or not an account is a bot or a human. In the case of [49], the topics that a tweet contains are used to train a classification model. And in an approach that is similar to this study, [12] uses the similarities between the content of tweets between two users to determine if an account is a bot or not.

Other machine learning models use unsupervised learning to detect bots. In the case of DeBot, [23] an assumption is being made about the behavior of bots in that they will demonstrate correlated behaviors over extended periods. Similarly to CATS, there is an assumption being made that bots act the same if they are part of a larger botnet structure. However, this may not always be the case as bot behavior may change in response to avoiding new techniques for detection.

2.1.3 Graph and Anomaly Based

Machine learning approaches are favorable in that they take into account, a holistic view by considering the metadata of tweets and accounts along with related information. However, one other limitation is that they may not consider the topological structure of the botnet which may be trying to influence user behavior online. While in certain cases the similarities between bots are considered [7, 12, 49], simply classifying bots is not enough to tell us about potentially synchronized behavior online. A combination of methods would then be able to tell us something more meaningful about what botnets are trying to achieve through their actions online.

For example, in the case of BotCamp [3], a combination of Machine Learning based classification and Graph based classification was used to detect bot campaigns. This is done by using DeBot [23] to determine if a Twitter account is that of a bot online. After accounts have been determined to be bots, graphs are abstracted from the interactions that they have with other bots. These interactions include retweets, mentions, likes, and comments. Once it does this, it can label each account to distinct community clusters. Later a supervised model is used to determine if a user is in agreement or disagreement with a sentiment. This information is what is then used to determine what the nodes in the different community clusters are trying to accomplish. Through this system, we can gain some meaning about what the bots are trying to accomplish. One drawback however is that this graph based system still depends on predetermined information on who is a bot. Such multifaceted approaches may not always be necessary.

For example, other graph based methods have taken simpler network analysis approaches. According to [37] it is easy to find bots when interactions on Reddit are conceptualized as a network. In these cases, bots are easily recognizable in that they contain high edge weights when compared to the non-bot counterparts. This abnormal behavior is because the interactions of these bots across the network happen at a high frequency. These bots however are known bots, usually with the word ”bot” in the username. While some of these might be harmful, the reason why there need to be advanced methods for the detection of botnets is that botnets are becoming smarter with time to overcome these detection methods.

2.2 Social Botnet Detection

Different methods are required computationally detecting botnets. Social botnets are defined as groups of social bots that are under the control of a single botmaster. [67] Bots within a botnet collaborate to engage in malicious behavior, oftentimes engaging with each other and with similar content. Additionally, according to [67], these bots try to mimic the actions and patterns of human users to avoid detection.

Because of the additional complexity involved, multifaceted approaches are usually required to detect these botnets. Consider the SBCD algorithm proposed by [45] to detect botnets. In this approach, a graph based and machine learning based approach is taken to classify social media accounts as part of a botnet network. This multilayered approach also takes into account social media post content to classify different botnet communities into respective categories that represent the bot's intent. These include groups such as spammers or fake followers which are also considered members of the botnets.

Another one of these methods is proposed in [62]. The first step in this method is to detect a "pivotal node". A pivotal node is a node that may represent the botmaster or a node that controls other bots in some way. The purpose of identifying pivotal nodes is that they will demonstrate unusually high interaction with the worker nodes. To determine the community of nodes that pertain to the botnets the modularity of the nodes is determined using a slightly modified algorithm for determining the modularity of a node. This multilayered method is considered to be a dual approach which is both anomaly based and graph based because of its need to detect anomalies first to inform the graph approach.

Lastly, there is SpamCom, a system proposed by [16] which is a system that aims to find spam botnets on Twitter. SpamCom boasts an efficient approach to detecting communities of botnets by conceptualizing Twitter data as multilayered networks with the end of using a graph based approach to find overlapping nodes and structural anomalies. These structural abnormalities are the abnormalities regarding the content of the tweets themselves along with the content similarity of a proposed spammer with other nodes that are near to it. By grouping these spammers into communities, SpamCom can identify communities of spammers on Twitter.

2.3 Identifying Malicious Behavior in Social Media

Because of the evolving nature of botnets and their ability to become smarter to avoid detection, efficient and novel methods are needed to detect abnormal behavior that informs further analysis. The practice of discovering groups of actors using only graph approaches has been explored in the recent past. One such application investigates persistent nodes in time series multiplex networks to find consistent nodes that transfer information between each other [13]. These nodes, known as hidden groups, are groups of nodes that stay connected over time and may try to disguise themselves by changing the mode of interaction between them over time. This method shows the ability to detect malicious behavior without the need for more information other than the structure of the multiplex networks.

Other frameworks base themselves on community features while also they implement machine learning models to tag nodes in the network [15, 17]. One of the proposed solutions is the Anomaly Detection On Multilayer Social networks (ADOMS) system [15]. ADOMS works by detecting anomalies in multilayered networks based on the structure of the multilayered networks. These structural abnormalities in the case of ADOMS are star-shaped node relationships that would mean the presence of spammers. ADOMS demonstrates a favorable unsupervised system, however, these systems are not without drawbacks. As mentioned, supervised systems such as the one in [17] have a drawback in that updating the models for new bot behaviors has overhead. ADOMS also assumes that cliques on their own are enough to represent abnormal network behavior. Additionally, as stated previously, these cliques or near cliques can change over time [13]. Therefore, changing network topology over time should not be ignored but instead embraced since it could also be potentially used to detect abnormal behavior.

2.4 Problem Identification

All of these methods aim to do the same thing, detect abnormal behavior by bots to detect how these bots interact with each other to spread like ideological messages. However, as noted, some of these methods have high overhead and might be blind to the societal and cultural context within which a botnet might be acting. For example, investigations into detecting botnets demonstrate that within social media, political botnets might support specific campaigns by taking after their messaging [3]. While it would be ideal for a detection method to understand and determine context, it is important as well that we know in which ways a botnet might be acting, for example, what messages they are spreading beforehand. We can then use this information to perform exploratory analysis in a network where we would expect this behavior to occur. For example, if we expect abnormalities to occur within a social network around some term or topic X, then we can narrow the scope of a social network to conversations about X to detect those actors that are particularly interested in X. Such a detection method remains largely unexplored so far. That is, using graph methods to detect abnormalities in a social network's topology and combining that with exploratory methods by filtering conversations around a single term. Abnormal behavior, in this case, might be, for example, bots that switch over to the topic of interest all at once (RQ1), or where the average community structure of conversation around that theme demonstrate abnormal topological changes (RQ2). This is what this paper intends to explore in further chapters.

Chapter 3: Background

3.1 Network Science

When speaking of a network, we need to be able to conceptualize the actors of some structure as the nodes, and the edges as some quantifiable relationship between the individual nodes. This set of nodes and the relationships between them, the edges, is what constitutes a network structure. Sometimes networks may have edges with special properties that represent special values. For example, edges may be weighted. Meaning they are mapped to values that represent some measure between the nodes. Edges may have a direction (directed) or they may undirected edges. Lastly, they may have labels that describe the relationship between two nodes.

Graph representations of real-world structures have proven to be helpful to analyze complex structures. For example, there has been a recent trend in the field of neuroscience towards learning how to apply network science models to solve challenges in the field of neuroscience. [59]. Outside of social media, conceptualizing other structures as graphs helps in identifying actionable insights [35]. For example, application of network science measures such as measures of centrality can be used to assess infrastructure systems [60]. In what is a meta-analysis, co-authorship networks can bring insights into what directions the field of network science is moving towards. [48]

Historically, the field of network science has been focused on single-layer network structures such as the applications mentioned. The reason is that many domains of science benefit from understanding related network structures as simple flat networks [36, 61]. Flat networks have a drawback however in that many real-world structures do not function as simple flat networks.

In reality, many real-world systems can be described as being multilayer or multiplex networks. Multiplex networks are networks in which nodes are members of different networks simultaneously. [19] In the case of multiplex networks, layers are associated with each other by some marker. In some networks, this marker is derived from the data itself. An example of this would be a multiplex network of sampled Facebook users where the nodes are the users, the edges are conversations between users, and the layers are differentiated by separating users that pertain to different Facebook groups. Layers in a multiplex networks could also be represented by time ranges. Using the Facebook example, the nodes of the multiplex network would represent users, and edges could represent the number of interactions between the nodes. The layers in the multiplex network would be differentiated by limiting the node-edge pairs to time ranges. [19] An example of such an application uses a 4-year multiplex network of Facebook friends that is aimed at understanding how the social interests of Facebook users change over time. [43] Multiplex networks do not only apply to social networks. Because of the complex nature of real-world structures, multiplex networks can capture multiple channels of connectivity between nodes and edges in networks. [19] Relevant application of multiplex network structure to analyze the underlying structure of a network is far-reaching. Examples include contagion studies, flight optimization studies, transportation networks, among others [18, 21, 26, 29, 33, 68].

3.1.1 Quantifying Multiplex Network Structure

According to [19], multiplex network structure can be identified with the following metrics:

- Interlayer degree correlations
  - Generally speaking these correlations can indicate if the hubs in one layer are also the hubs in another layer.
- Overlap and multi-degree
  - The node connectivity patterns can be correlated in two or more layers and these correlations can be captured by the overlap of the links. For example, we usually have a large fraction of friends with whom we communicate through multiple means of communication, such as phone, email, and instant messaging. This implies that the mobile phone social network has a significant overlap with the one of email communication or the one of instant messaging. The overlap of the links can be quantified by the global or local overlap between two layers, or by the multi degrees of the nodes that determine the specific overlapping pattern.
- Multi strengths and inverse multi participation ratio of weighted multiplex.
  - The weights of the links in the different layers can be correlated with other structural properties of the multiplex. For example, we tend to cite collaborators differently from other scientists. These types of correlations between structural properties of the multiplex network and the distribution of the weights are captured by the multi strengths and inverse multi participation ratio.
- Node pairwise multiplexity
  - When the nodes are not all active in all layers two nodes can have correlated activity patterns. For example, they can be active on the same, or different layers. These correlations are captured by the Node Pairwise Multiplexity.

- Layer pairwise multiplexity
  - When the nodes are not all active in all layers, two layers can have correlated activity patterns. For example, they can contain the same active nodes or different active nodes. These correlations are captured by the Layer Pairwise Multiplexity.

All of these different metrics can be used for different applications when performing investigations of the network. Because this study aims to explore changes in network structure over time we are only interested in measures that quantify inter-layer differences. Additionally, the metrics that we are interested in are meant to be applied in networks where the same nodes may be used in the different layers of the multiplex network. In our case, the networks that we will be working in may or may not share nodes between layers. An example of these kinds of social interactions online would be conceptualizing Facebook social networks as multiplex networks. If we were to consider a network of Facebook users, we are to expect different users to either delete their profiles or for more users to join over time. For such a network, this would mean that there are nodes that may not exist across layers that represent different time frames.

3.2 Social Media

3.2.1 Political Discussions

Over the past few months, the eyes of academics in the political space have been on social media. Specifically, the role of Former President Trump's use of Twitter has come into question. [53, 57] After the events on the Capitol building on January 6th, questions have arisen within the field of social media studies. For example, what role if any did social media play in radicalizing individuals on engaging in potentially treasonous acts such as the ones witnessed on January 6th?

While this question can be approached from many different disciplines, the motivations for this answering this question remain largely the same. Large social media platforms, through their terms of service, attempt to qualm the misuse of their platform by prohibiting users from engaging in illegal activities with their sites. Additionally, it should be noted that these platforms act as vehicles for individuals to exercise their constitutional right of free speech. Because of this gray area, it has become an area of interest on methods to detect harmful speech as it comes into their platforms. Some of these solutions involve informing users that the information that they may be viewing is unverified or may be false or misleading in some way. Instagram for example, uses third-party fact-checkers that allow users consuming information on the platform to know whether the information is verified. In some cases, Instagram may use this information to remove posts from their platform that does not meet their community guidelines. [2]

Identifying misinformation can be difficult at scale, however. As mentioned earlier, to reduce the amount of false information that exists online, efforts into methods for detecting false information have increased over the past few years.

Some of these methods may include: [25, 42, 50]

- AI/ML models that are trained on the content of social media posts, along with features that include follower counts, follower types, engagements, etc.
- Detecting misinformation by looking at repeated language from other posts.
- Detecting misinformation by checking whether the engagement networks of users are contained within a densely interconnected network.

Conceptually one of the questions being explored in this study is whether it is possible to detect misinformation efforts computationally. Specifically, would malicious behavior by bots present itself within a network over time? As mentioned previously, actors playing a role in misinformation are not lone wolves. There is evidence to show that players in the business of spreading misinformation online will work in coordinated efforts to spread their message. For some, their goals are clear, to influence public opinion on some topic so that it benefits them. This behavior was noted during the 2016 election where foreign entities were meddling in the 2016 Presidential election to influence public opinion. [8, 58]

Because these are coordinated and calculated efforts, we would expect at least some of the accounts involved in the spread of misinformation to emerge from a common site. This could mean either one person or groups of people working towards the same goal to the effect where not every person individually controlling individual accounts. Because of this structure, we can confidently expect similar behavior from accounts that work to spread misinformation. [58] This type of group coordination through a botmaster conduit is what we aim to explore in the context of Reddit.

3.2.2 Reddit

Reddit is a forum where users can engage in conversations isolated within posts created by other users on the site. These created forms exist within structured subreddits that contain moderators. Posts within these subreddits revolve around a common theme. Some of these subreddits have very well-established communities with active members and hundreds of posts that users engage with.

Some subreddits exist for mountain bikers to talk about gear and trails that are popular. There are subreddits for people that are fans of different TV shows to share memes and to discuss the TV shows. Essentially, there are subreddits for any interests that a user might have. Within these subreddits, users can engage with others by either commenting on posts, voting on posts, or voting on comments. They can also upvote on posts, or give awards to other posts.

The Problem with Subreddits

In recent years, some of these subreddits have received pushback from Reddit for inciting violence or spreading misinformation. One example of these subreddits is r/TheDonald which was a forum for supporters of Donald Trump to engage in conversation around former President Donald Trump. TheDonald was eventually banned for failing to moderate violent rhetoric and misinformation. [32] Another such example, r/Incels was a subreddit that connected individuals that deemed themselves as being involuntarily celibate. They adopted tropes and ideas the lead members of the community feel like they were being treated unfairly. [46] Incels was also later banned for helping spread violent rhetoric.

Naturally one of the prevailing conversations currently is the conversation around the Covid-19 pandemic. Specifically, on Reddit, subreddits have been created that aim to be a one-stop-shop for information regarding the pandemic. This does not mean however that conversation around the pandemic is isolated to those few subreddits. Instead, because of the structure that Reddit is built around, there is nothing barring conversation regarding the coronavirus in most other subreddits. Additionally, because of this same structure, it can be noted that conversations around certain topics are done through the paradigm that the subreddit exists with. This behavior is not isolated to certain political subreddits. We may see this in subreddits such as liberal subreddits that may use terminology or phrasing that is consistent with the users of that community.

In this sense, Reddit acts as an avenue for information spread since users are more exposed to ideas and information within a forum that they trust. For example, within the topic of the coronavirus pandemic, several subreddits have emerged with the goal of aggregating news stories and encouraging conversations around a more narrow scope. Some of these subreddits have loosely moderated communities when compared to well-established subreddits that already exist on the site. It has been noted that within these communities there is significant information overlap between subreddits like r/ChinaFlu and other communities like: [66]

- r/Conspiracy
  - Subreddit devoted to discussing conspiracy theories around major global events.
- r/Collapse
  - Community devoted around discussions of an impending collapse of civilization.
- r/Wuhan_Flu
  - Community that claims to be a place for uncensored discussion pertaining to the coronavirus. This community was later quarantined by Reddit due to the hoax content that was being spread.

3.2.3 Echochambers

One of the dangers of social media is that, separately from connecting strangers together, it is also really good at focusing like-minded people into communities and spaces that ”echo” what the individual wants to hear. Echo-chambers are online groups or spaces that are divided ideologically and where the conversations that occur are between people that already agree with each other. It needs to be noted that while echo-chambers sound innocent, dangers can emerge in communities that are discussing large political or social events. For example, politicians may leverage the polarization of these communities by accepting populist ideas that cater to different ideological parties [41] Studies also noted that as ideological differences widened, the rate of communication lessened and in effect strengthened the echo-chamber being formed by the members of the respective communities. [20] Because the majority of conversation on these topics occur within echo-chambers [63] we must recognize the existence of them within the structure of popular social media sites.

3.2.4 Misinformation on Social Media

Another pressing issue with social media is its tendency to help spread misinformation online. One recent study noted that interactions with fake content on Facebook and Twitter hit a peak in 2016 and continues to grow on Twitter. [5] This can be concerning since social media users may not always be representative of the general public's feelings on politics and social issues. [47] This is to say, an information scientist's concern might be that the mixture of the social structure of social media along with the information being presented to the users that most depend on it would have adverse effects on society. Plenty of research has been done on this same topic with growing interest in specifically detection, dynamics, validation, and management of misinformation. [6] While work has been done in being able to calm the amount of misinformation that is present on these social media sites computationally [30, 64] current challenges present brand new obstacles. For example, the covid pandemic brings new questions and challenges which are just starting to emerge that provide insight into how misinformation related to the pandemic spreads [24]

3.2.5 Bots

The term bot is used to describe ”a diversity of software systems, such as systems that are designed to hold conversations with human, generally automated software agents and compromised accounts that are used in Command Control networks to launch attacks”.

[50]

One study examined the role that Twitter bots play within the highly polarized vaccination debate. In this case, the bots demonstrated high amounts of interaction with tweets that were within their opinion group. [65] The results of this unnaturally high level of interaction between bot and human can lead to bots having high popularity within their social network or bots being mistaken for human actors. [4] They are of concern for this paper is within the realm of highly politicized issues such as the COVID-19 pandemic. Additionally, bots can play a role in quickly disseminating false information by being super-spreaders of false information quickly increasing the virality or exposure of false information to populations online that would have not had access to it otherwise. [56]

Understanding that bots can have adverse effects on discourse in social media about highly politicized issues makes identifying dishonest behavior important. Especially since exposure to information from bots may cause some users to double down on their ideologies. [11] It has been demonstrated that news coverage that mentioned bots but did not explain how to recognize them did not ultimately help alleviate and instead caused harm compared to having no exposure to bots at all. [55] Identifying the structure then of these bots along with understanding their technique to influence social thought or belief should be a priority to helping alleviate some of the damage that these bots do in spreading misinformation online.

3.3 COVID-19

COVID-19 has completely changed daily life since early 2020. Since its detection as the cause of the Hubei province, the virus has placed governmental organizations on high alert. After the virus that causes COVID-19 was isolated and sequenced, it was named ”SARS-CoV-2”. The Coronavirus disease was officially declared a pandemic by the World

Health Organization on March 11th, 2020.

Because of the recency of the emergence of COVID-19, there is still so much that is unknown about it. The emergence of the virus has demonstrated the disparities in healthcare, financial inequalities and has helped bring racial injustice to the national conversation. [40] In regards to people and their interaction with the virus, it is important to note that COVID-19 has highlighted the fact that misinformation can spread very quickly online through social media sites. [24] Images of ”scientists” claiming that the pandemic is a hoax have appeared online and become viral. [31] Early on in the virus, there were images regarding home remedies that would promise to cure the virus. [34]

It is important to note that knowledge on how the virus works and how to treat the virus is still continuously emerging. [22] As more information comes out on the dynamics of the spread of the virus, the CDC has made updates to the messaging around proper safety measures to help minimize the rate of transmission of Covid-19. Along with this new treatment information (which is vital for healthcare workers to know), knowledge on how to talk about the virus in conversation has changed over time.

3.3.1 COVID-19 Terminology

The COVID-19 epidemic has demonstrated the challenges in managing public sentiment towards the disease and virus that cause the disease. As it relates to the current COVID-19 outbreak, previous experiences have demonstrated that public perception is greatly influenced by location and cultural fears. [52] To curb the spread of prejudice online, the World Health Organization (WHO) has published guidelines on proper virus nomenclature. [1] However, while the WHO works to inform government organizations on correct naming, the message may not reach all individuals. The terms used by the general public when speaking of the COVID-19, for example, may mistake the virus which causes the disease and the disease itself as being the same thing. Another issue is that as new information is being learned about how COVID-19 impacts individuals, the need for new terminology arises. One such example is the issue of properly stating what symptoms a Covid a long hauler has. In their case, while the Sars-Cov-2 virus is no longer present in their system, and they technically do not have Covid, they may demonstrate symptoms. [10]

In order to speak about the virus a person speaking to their members of their community might use the term ”coronavirus”, ”covid”, or ”Rona” to name a few popular examples. While the name use of COVID-19 does not necessarily indicate misinformation spread, the highly politicized climate surrounding the administration of health resources has lead to the use of inflammatory terms such as “killer virus”, “deadly virus”, and “Wuhan virus”. [39] According to Karalis Noel, the use of these terms ”promote fear and panic which propel prejudice, xenophobia, and discrimination”. [39] Because of this, we can treat these terms as misinformation since they frequently use misleading terminology that in this particular case, the WHO has deemed inappropriate for official use. The question arises, as this terminology is used in online spaces, what can the term used about a particularly polarizing topic in conversations tell us about the nature of misinformation spread on social media.

Chapter 4: Methodology

4.1 Data Collection

For this study, all of the data was collected from Reddit using the Pushshift API. Pushshift is a service that provides an API interface by which users can access information on Reddit posts. In our particular case, the Pushshift API was queried to gather all of the posts where the posts matched a particular input Subreddit. Additionally, only posts that have over 5 comments within them were gathered to get only posts that contain some conversation within them. We also noted that the amount of data that was being gathered when not filtering by the number of comments present was too high. By performing this initial sampling we were able to contain data to a manageable amount and filter to ensure that we gathered meaningful posts.

The API was queried using Python on a Raspberry Pi. In most cases, because there was a limit on the rate at which the API could be queried, this process took days. This is because in most cases the API had to be queried thousands of times to gather all of the necessary data. This is, at least once per post and in some cases, for upwards of 20 thousand posts per subreddit.

For our study, the following subreddits were queried to gather the comments for the posts within them:

- ”Conservative”

- ”Progressive”
- ”Democrats”
- ”Republicans”

The main goal was to find subreddits that exist on opposite political spectrums to exaggerate the found differences between the different subreddits. In the case of our study, the networks were separated to create the multiplex networks that we are looking for to answer RQ1. In total, we collected the comments for posts in all the aforementioned subreddits starting on January 01, 2020, until July 30, 2020. To create the layers of the multiplex network the overarching networks were split on the months so that multiplex networks were created from each network. This means that each layer of the multiplex network corresponded to each month mentioned previously.

4.2 Adjacency Matrices

The adjacency matrices generated from these data sets are undirected. These matrices are created by identifying conditions by which it is interpreted that conversations happen:

- For a term, if the term is used in the title of a post, then it is assumed that every comment that occurs within concerns the term that is present in the title.
  - In this case, a connection is present between the author of the post and the author of the comment for each comment present. This means that if a user commented twice within a post, they then have two connections with that author. In this same case, authors can have connections with themselves multiple times.

- For a term, if the term is used by a user when commentating, an edge is formed between the user commenting and the author of the post.
  - This stands whether or not the title of the post contains the term that we are searching for. This means that if a user has commented 4 times within a post, if the title of the post does not contain the term in question, and only 2 of the comments contained the term, then there are only 2 edges identified between the user who commented and the author of the post.

Once the structure for the adjacency matrices was identified, the adjacency matrices were created based on the time frame for research.

To answer Research Question 1, each month in this range was separated to create a network of each term of interest for each month for each subreddit. January was not included for any of the subreddits at the point of analysis simply because there was no use of the interest terms during that time.

For Research Question 2, the months of January to July were aggregated together for analysis. This data was aggregated to create the aggregated networks for each subreddit that the data was queried for. This means that each subreddit had one graph for each term of interest where the network contained all interactions regarding that term between January 1st and July 30th. The terms of interest are as follows:

- covid19
- covid
- chinavirus
- coronavirus
- corona

- rona

All of the data collection and manipulation was done with Python.

4.3 Quantifying Network Differences

The first algorithm used is gathered from [54] which is later applied in [44]. This measure is intended to provide a different measure between two different graphs. The resulting D measure is determined using the formula:

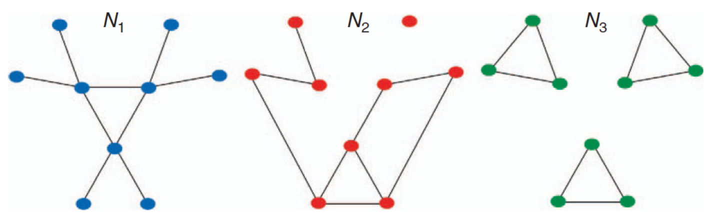
$$D = w_1\sqrt{\frac{J(\mu_G \parallel \mu_{G'})}{\log 2}} + w_2\left|\sqrt{NND(G)} + \sqrt{NND(G')}\right| + \frac{w_3}{2}\left(\sqrt{\frac{J(P_{\alpha G} \parallel P_{\alpha G'})}{\log 2}} + \sqrt{\frac{J(P_{\alpha G^c} \parallel P_{\alpha G^{c\prime}})}{\log 2}}\right)$$

This formula is comprised of three parts, each containing their coefficient weights indicated by the w_1 , w_2, and w_3 respectively. By default, they are set to 0.45, 0.45, and 0.1. The first part of this equation compares the networks' diameter, average distance, and other connectivity characteristics. The middle part determines the node's heterogeneity by determining the NND of graphs G. NND represents the network node's dispersion in terms of connectivity distances. This middle portion is an application the Jensen–Shannon divergence. The last part compares the nodes' centrality measures for each of the graphs being compared.

Put together, D can provide a dissimilarity measure ranging from 0 to 1 where a D value of 0 represents two graphs that are the same. A D measure of 1 would indicate that the two graphs are completely different in structure.

The application of this algorithm was in the form of an R script created by the authors of [54]. This R script was made publicly available on GitHub and was called from within Python to analyze results. One note to make about this measure is that this measure does not take into account edge weights.

This measure was chosen to quantify network differences over time because of its structure and advantages. Specifically, for our study we are interested not in the rate of usage of a term, but instead how that structure of conversation of a term changes over time and specifically if there are any months that demonstrate abnormal usages over others. Because of this, we are not interested in measures that use the quantity of nodes as part of computation, but instead how those nodes interact with each other. This means that if one month is different in structure from other months then the term that the graph is based on was potentially adopted by bots to engage in conversation. The main advantage of using this D measure is that it takes into account the topology of the graphs to compare. Specifically, topological in this case refers to not only nodes and edges, but the structure in which the edges exist. Consider the following example present in [54].



In this example, N_1, N_2, and N_3 all have the same amount of nodes and edges. The main difference between the three is that N_2 and N_3 contain nodes that are disconnected from all other nodes. In this case, the D measure does something that other dissimilarity measures may not do, and that quantifies measurable differences in the cases where there are disconnected nodes or cliques, for example. In the context of this study, high differences between two networks would be able to represent networks where the number of nodes and edges are the same but how the nodes and edges are connected are different in such a way that they interrupt the flow of information on the network. [54] Essentially stating that even though the quantity of conversation occurring in a network around a term may not entirely change over time, we should be able to capture how the structure of those conversations changed over time.

This is very different from the second algorithm used is gathered from [9] and is detailed below:

$$D_{\mathrm{JS}}(G, G') \equiv \frac{1}{2} KL(P \,||\, M) + \frac{1}{2} KL(Q \,||\, M)$$

This formula is in essence an application of the Jensen-Shannon divergence. In this formula, the KL divergence as it will be known from here on is implemented by applying the following formula.

$$KL(P(k,\ell) \,||\, Q(k,\ell)) = \sum_{\ell=0}^{\max(d,d')} \sum_{k=0}^{N} P(k,\ell) \log \frac{P(k,\ell)}{Q(k,\ell)}$$

The purpose of the single-valued KL divergence is to define a single value for differences between two portraits. To implement this formula, we need to determine the distribution of probabilities for two random nodes being connected.

The resulting network portrait divergence measure is a value between 0 and 1. This measure is also symmetric. This means that for Graph 1 and Graph 2, the value of $NPD(Graph1, Graph2)$ is the same value as $NPD(Graph2, Graph1)$. What makes this measure so desire-able for use is that: [9]

- NPD compares networks based on their topology.
- Because node matching is not needed, this measure is efficient and fast.
  - This also means that NPD does not assume that the graphs contain the same set of nodes.
- NPD can be used to compare weighted networks.

This makes NPD an appropriate measure to compare networks where we want to look at the topology and ”portraits” of the graphs. For example, while the D measure is able to capture topological changes that are within a community, we are not interested for RQ2 about topological changes over time but instead, we are interested in comparing two graphs wherein the only relation is the term that they were abstracted from. For example, in the case of RQ2, the nodes between the subreddits are not necessarily assumed to be shared between the layers. While they may be, it makes more sense that the nodes from within subreddits might match in some way over time, and because of that, we would be interested more so in how those nodes interact with each other within their respective communities. When we compare aggregated networks (ie. subreddit to subreddit) then we are inclined to care less about how the nodes in the networks differ in their interaction instead focusing on the overall landscapes of the networks for comparison.

Chapter 5: Analysis

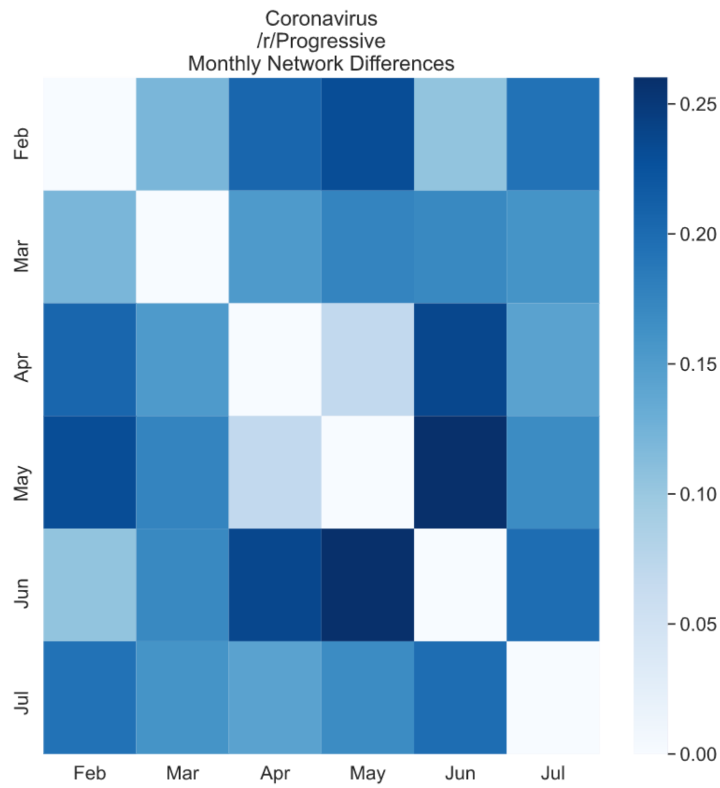
5.1 Research Question 1

To answer Research Question 1, the D measure was computed for each adjacency matrix which was created from aggregating monthly interactions so that each layer in the multiplex network corresponded to a month's worth of interactions. This meant that there was a matrix of D measures for each subreddit and each term. These matrices were visualized as heat maps to perform exploratory analysis and to find any anomalies. An example of such a heat map is demonstrated in figure 5.1.

In this case, the heat map represents the comparisons between monthly graphs created where the conversations occur in r/Progressive and interactions revolve around the term coronavirus. As it can be seen, this heat map demonstrates expected behavior in that none of the compared networks is largely different. Specifically, none of the networks have a D above 0.25. Remember that the D measure will be smaller as graphs become more similar and larger as they are different, with a range between 0 and 1. In this case, all of the values range between 0 and 0.25 which indicates that if there are large dissimilarities between networks, they are not large enough to indicate significant differences between the networks.

This type of heat map is completely expected (Figure 5.1). Consider that in the heat map, the network representing the month of February (row 1), as the compared graphs progress in time (move to the right along row 1) we see that the D measure increases up

Figure 5.1: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”coronavirus” within the r/Progressive subreddit from February to July



until we reach June. In June, the difference between the networks is less, meaning that the network for June looks more similar to February than May did to February. This is indicating that as time passes, and more conversation occurs on the site around the term ”coronavirus”, the more the graphs look differently from the initial graph of interactions in February. Additionally, if the difference were significantly larger, then this would point to some change in behavior at the point of June. We see this result because the graph in February is small and dense within r/Progressive regarding the term ”coronavirus”. Later months include more nodes and more complex network structures. For reference, the following images represent the February graph (Figure 5.2) and the corresponding July graph (Figure 5.3).

The visual differences between the two graphs represent the finding in the differences D measure. As can be seen, the graph for February demonstrates that a graph that is sparsely connected with many fewer nodes than the corresponding July graph (Figure 5.3). On its own, this information may not explain the differences between the networks. Instead, this may help inform that the differences in the networks may have to do with simply increased term usage and not structural differences within the network.

5.1.1 Terms of Interest

It needs to be noted that for each of the terms of interest the visual analysis supports and helps explain the statistical results that were obtained. In this case, ANOVA tests were conducted on all of the matrices that were created from each subreddit and each term. The ANOVA tests were used to determine if any of the distributions of D measures between the monthly graphs demonstrated significant differences from other months. None of the matrices that were tested demonstrated statistical significance. This would then indicate that any analysis performed would have to try to manually and visually understand

Figure 5.2: February graph of conversations regarding to Coronavirus on r/Progressive

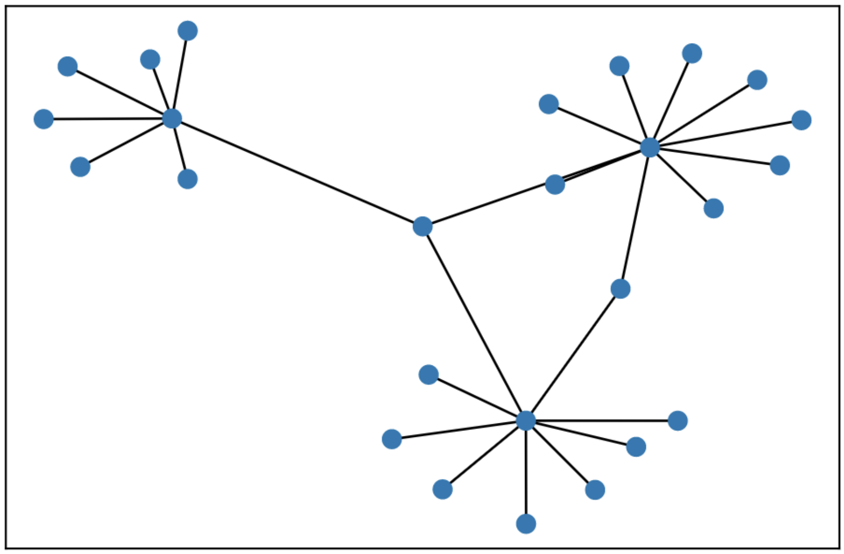
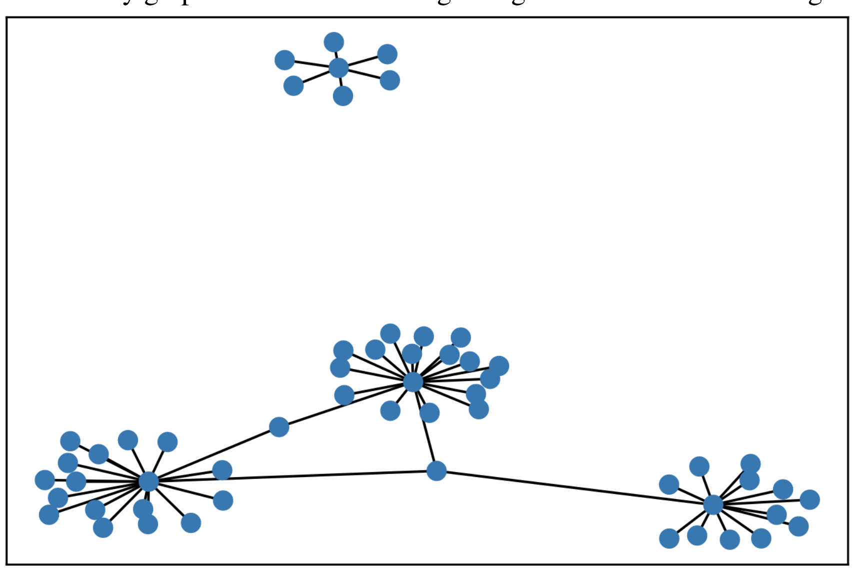


Figure 5.3: July graph of conversations regarding to Coronavirus on r/Progressive



the nuances within the heat map results. These nuances however are not mathematically supported.

China virus and Rona

One of the issues when interpreting the results of the China virus graphs is that there are not enough conversations regarding the term Chinavirus for the time frame investigated (figure 5.4). This means that in the topic of Chinavirus, over time and across the different subreddits, there is not enough evidence to state that the term usage was unstable. There is enough information to state however that the term was minimally used across the different subreddits and that conversations around this topic were limited. This may be because of the proactiveness of the different subreddits to moderate the information arriving at the subreddits. Either:

- The term was used and the comments or threads were later deleted.
- The term was never used in the subreddits in question in the first place.

There is a very similar case with the term Rona (Figure 5.5). The difference in the use of the word Rona is that with Rona there is no use of that term within the Progressive and very minimal use in Democrats overall. However, there is a difference when it comes to Conservative which has a consistent use of the term or presence after February.

While this amount of use is minimal to none, it needs to be noted that there is still some use and this is to be expected based on social media trends that indicate conservative communities making light of the conversations around the coronavirus. Republican demonstrates some use as well, albeit, much lower than Conservative.

Figure 5.4: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”chinavirus” within the r/Republican subreddit from February to July

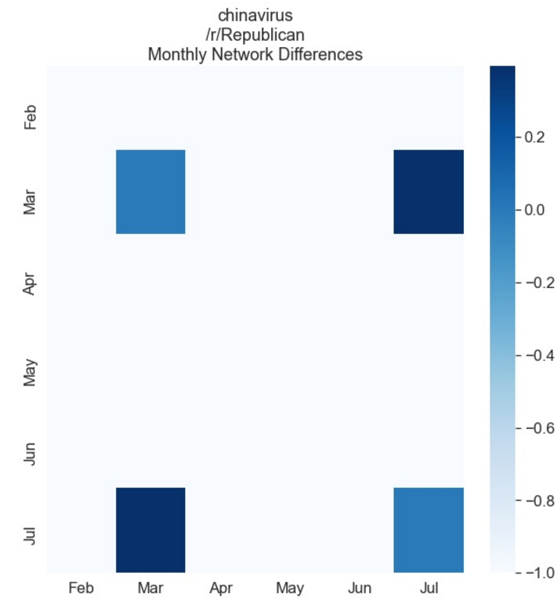
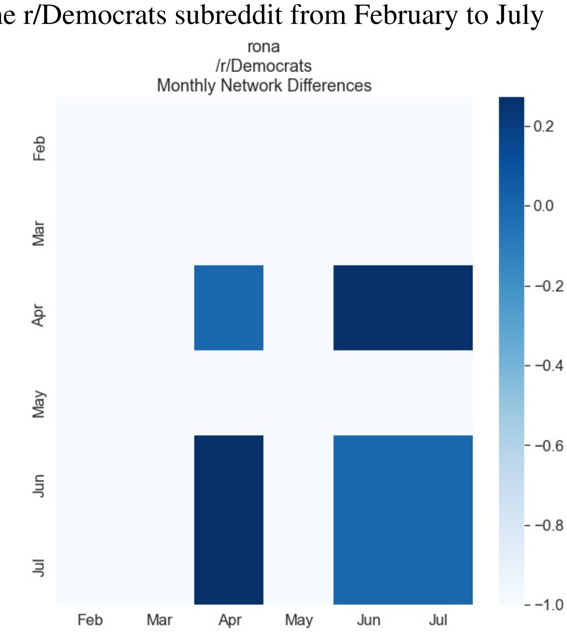


Figure 5.5: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”rona” within the r/Democrats subreddit from February to July



Corona

Corona can be considered a slightly more official term for referring to the novel coronavirus and in this case, behavior mostly matched what we expected to see based on social media trends. Specifically, within the progressive community, we see a lack of use of the term to refer to the novel coronavirus in (what appears to be) favor for more official terms.

The Republican subreddit demonstrates moderate differences (Figure 5.7) in the topological structure of the conversation indicating that as time progresses, the conversation around the topic becomes more robust. This would indicate a trend towards using ”corona” as the de facto term to refer to the pandemic. Interestingly, the Conservative subreddit demonstrates low changes (Figure 5.6) between the different networks and in this case, also demonstrates the expected results (as time passes, the differences between the networks become larger and larger).

Coronavirus

The most consistently used term was the term Coronavirus. Across all of the subreddits, Coronavirus remained stable over time. Even after performing an ANOVA test to check for variance, none of the months showed any particular variance that would indicate a large abnormality of use.

Covid and Covid19

The ”covid” term demonstrated some interesting trends that seem to divide the ”left-leaning” and the ”right-leaning” subreddits. For both r/Conservative and r/Republican, the term demonstrates a large amount of consistency with low D measures between networks in months after February. On the other side of the spectrum, ”left-leaning” subreddits

Figure 5.6: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”corona” within the r/Conservative subreddit from February to July

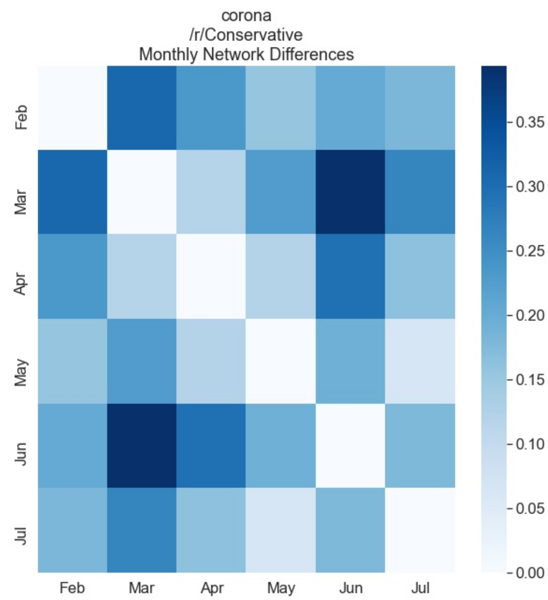
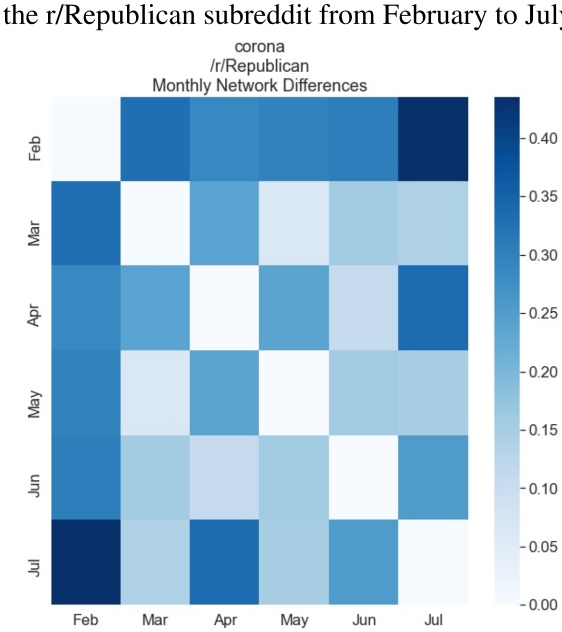


Figure 5.7: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”corona” within the r/Republican subreddit from February to July



demonstrate a low use of Covid within the sampled data until March where the D measure indicates moderate differences between the different networks.

This same is the case with Covid19 except for the fact that in the Progressive subreddit, the month of February is very different in structure than all of the other months.

All the other months, as mentioned previously, have a low-moderate amount of difference between them demonstrating no abnormalities between them.

5.2 Research Question 2

As mentioned previously, the second research question aims to investigate whether the networks of the use of terms in the different subreddits differ. Because of this, the research question hits some of the same issues as it pertains to different terms outlined in Research Question 1. As explained previously, the networks analyzed in this question are the aggregated networks over the range of time from January to July. Networks abstracted from certain terms such as Coronavirus demonstrated themselves to be very robust and expansive. The largest of these is the aggregated network of the use of Coronavirus within the r/Conservative subreddit which contained just over 100,000 edges. Other networks, such as those abstracted from Rona and Chinavirus are very small (with some subreddits containing no usage of those terms whatsoever), even though the time range aggregated is very quite large.

As mentioned previously, this little amount of use within certain subreddits could be due to a few different reasons. Further discussion on what this might mean for this study is expanded on in the conclusion.

Similar to Research Question 1, a comparison matrix was created by calculating the Network Portrait Divergence (known as NPD) value of each subreddit aggregated network of each term to each other. For example, as it pertained to Rona, the aggregated Rona

Figure 5.8: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”covid” within the r/Conservative subreddit from February to July

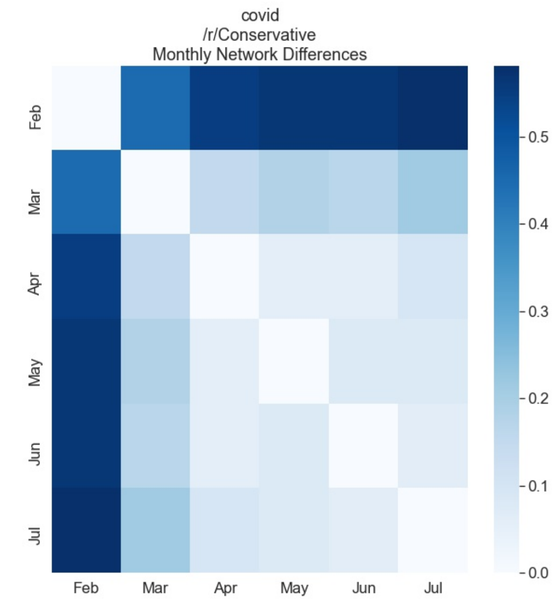


Figure 5.9: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”covid” within the r/Progressive subreddit from February to July

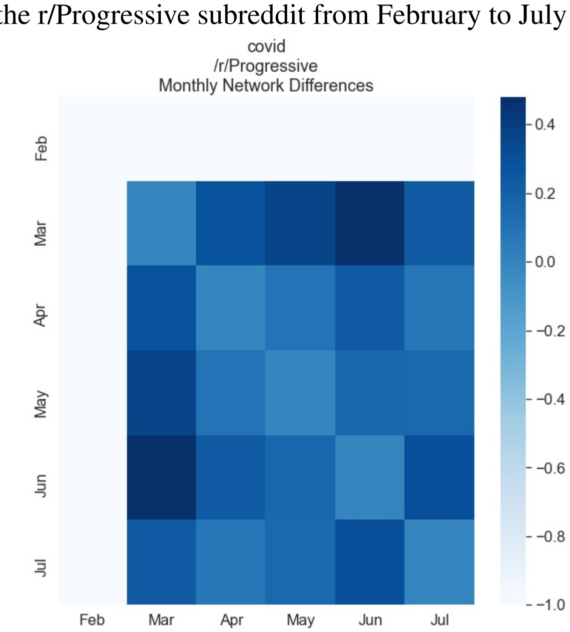


Figure 5.10: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”covid19” within the r/Conservative subreddit from February to July

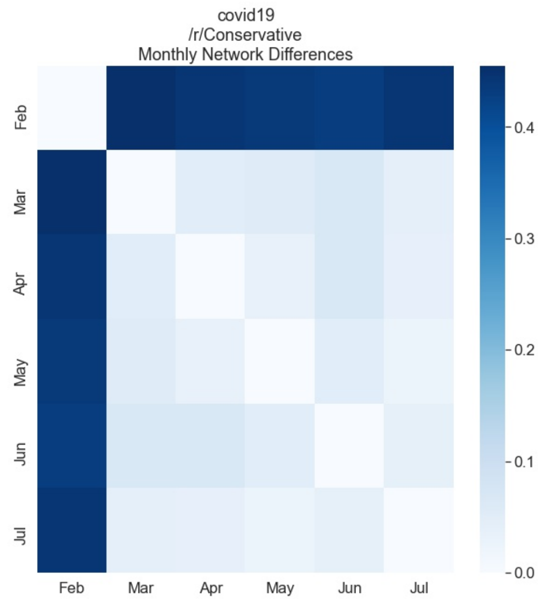
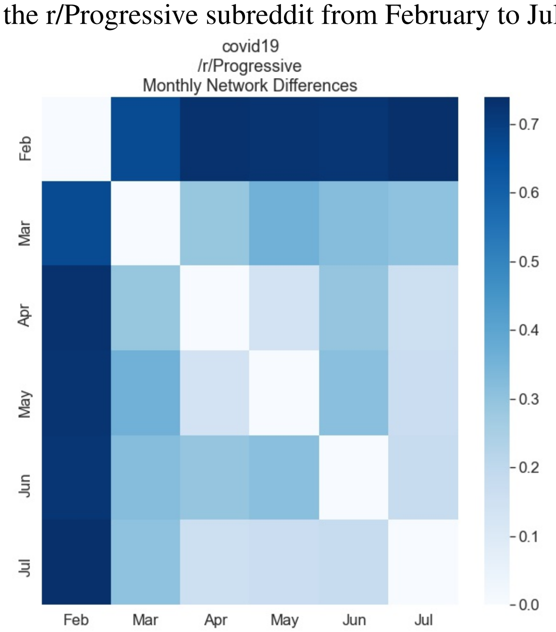


Figure 5.11: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”covid19” within the r/Progressive subreddit from February to July



networks for each subreddit were compared to each other and a matrix was created of the resulting NPD measure.

As mentioned previously, the NPD measure is a value between 0 and 1 where identical networks would produce an NPD of 1, and networks that are different would produce an NPD more towards 0. By plotting the graphs on a heat map the goal is to visually determine anomalies that might provide some insight into the behaviors of different terms in the different subreddits. Different from Research Question 1, RQ2 is not interested in looking at the layers in a multiplex network as being separated by time, instead, the goal is to look at each subreddit as different layers in what would otherwise be a flat network.

ANOVA tests were performed on each term individually to determine if any of the subreddits in question demonstrated different behavior. None of the subreddits demonstrated odd behavior which would indicate that among the subreddits sampled, statistically, there are no abnormalities present in the graph structure. This behavior can be seen in figures 5.8, 5.9, 5.10 5.11, 5.12 and 5.13.

5.2.1 Exploratory Analysis

In the case of the subreddits in question, the statistical analysis supports the visual exploration. None of the terms examined demonstrated any large visual anomalies that can be expanded upon.

All of the subreddits that were examined are known as being moderated communities. At the very least, there is some form of moderation that tries to push towards civilized conversation. This was by design, as the purpose is to find misuse of moderated open platforms by automated posters. We can note however that during initial data collection, a mistake was made on the name of one of the subreddits, namely Republicans. The mistake here is that the official moderated Republican is not the same as the unmoderated and

Figure 5.12: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”rona” between subreddits from January to July

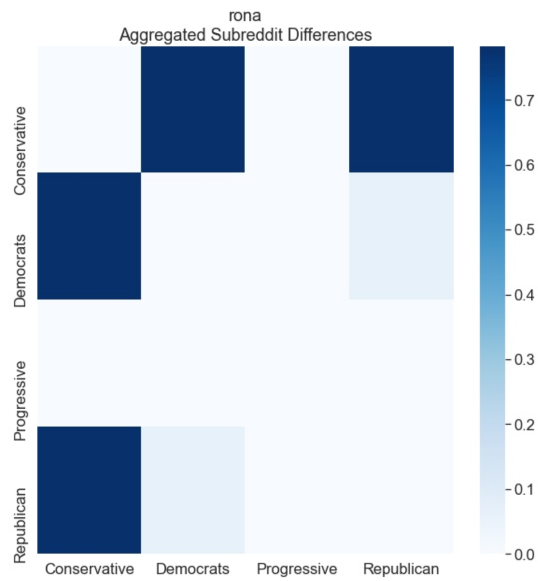
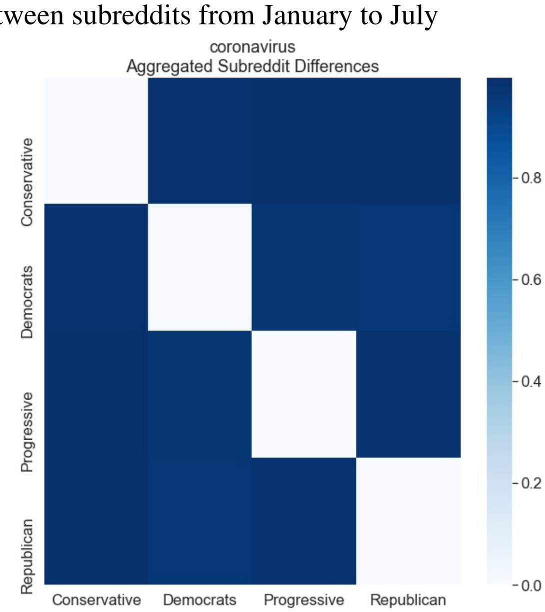


Figure 5.13: Heat map of multiplex layer differences in the multiplex graph of the usage around the term of ”coronavirus” between subreddits from January to July



seldom used Republicans. During this testing phase in which the data aggregated was much less, there were some visual differences between some of the networks and Republicans. This demonstrated that at least in this example, within moderated communities there seem to be differences between moderated and unmoderated communities.

Chapter 6: Conclusion

6.1 Research Questions

Earlier we mentioned methods that are focused on determining whether a particular actor online is a bot or not. This paper sits at the intersection between graph-based methods for bot detection and anomaly-based methods for bot detection. The difference between this paper and the methods mentioned previously is that while previous methods focus on predicting whether one account is a bot or not, we are using graph approaches to determine if the actions taken by these bots are detectable. Additionally, most of the studies mentioned are based on Twitter structure and data. This study focuses on anomalies in term usage on Reddit. Because of this, it lends itself to a graph-based method to understanding the behaviors of Social Botnets. This does not mean however that individual bot accounts cannot be identified through further analysis. Instead, this means that through this study we can at least try to detect coordinated efforts by manually narrowing the scope in which we expect the anomalies to exist with terms related to the coronavirus pandemic.

6.1.1 Research Question 1

Given the results of the analysis, we can conclude that we fail to reject the null hypothesis for RQ1 since we did not find any significant differences in the term use over time. Additionally, we noted a lack of use in terms that were investigated. This could be because the moderated nature of the communities allows for the deletion of problem terms before they are interacted with. We can see that snapshots taken of Reddit from the API used to query the data are taken at later times. Because the data in the API is not raw data including comments as they appear on the site, we may not be seeing the breadth of the conversation around the terms that we were interested in.

Given this data, we can state that within the subreddits that were investigated, none of the terms demonstrated abnormalities in network similarities over time. This means that within the subreddits that we were investigating, and given the terms that we were working with, there is not enough evidence to state that the use of any one term had large disruptions at any one point in time. As it pertains to term stability, this demonstrates that at least in the communities that were sampled, the terms themselves remain stable over time. This could be a testament to the moderation abilities of the communities in question since two of the subreddits in question have reputations as not being communities with clearly defined or stringent moderation policies when compared to their counterparts.

6.1.2 Research Question 2

As with RQ1, the analysis for RQ2 did not demonstrate any significant differences between the different graphs that were investigated. The NPD analysis performed between networks did not affirm any significant differences of any kind. In most cases, the lowest NPD in any of the matrices was 0.80. This meaning, the portrait of the graphs is very similar and structurally mimics each other.

This means that for RQ2 we fail to reject the null hypothesis which states that for each aggregated graph of the different terms there are no statistically significant differences between the NPD values.

Additionally as mentioned previously, there are no visual abnormalities between the aggregated networks that would warrant interpretation. The only interpretation that can be performed would derive from the fact that during the testing phase we noted that completely unmoderated communities would exhibit large differences in NPD and D measures from their moderated counterparts.

These results indicate that communities, whether loosely moderated or strictly moderated, do not demonstrate differences in the overall use of any given term. As mentioned previously the NPD measure is a measure intended to capture the differences between two networks based on the Jensen-Shannon divergence. To apply the Jensen-Shannon divergence, the networks are abstracted based on their topologies. Because this abstraction is not necessarily focused on term rates and instead, on the structure, we can say that while the rates in usage between the different networks may be different from subreddit to subreddit, the topological structures in how they are used are not different.

Lastly, because there are no differences in term usage behavior over time, and between the term usage behavior between the subreddits we are not able to state that any of the terms that were investigated had abnormal adoption rates. This is, since we could not find coordinated bot behavior, we cannot state that botnets decided to adopt or strictly stick to certain terms as part of their methods for influencing public opinion online. Additionally, because coordinated bot behavior was not detected, we cannot speak to adoption rates of human users on Reddit in response to being exposed to high rates of charged terms referring to the Covid-19 pandemic.

6.2 Discussion

Through this study, we were able to determine that there is another place for the application of algorithms that measure differences between networks. Albeit, while the results do not point to a hard difference from one online communities to another, the results are made clear through the limitations in that the lack of access to the entire unmoderated may have limited the findings.

One limitation in this study is that since we cannot scrape in real time and have to rely on the API as a proxy for data gathering we may not be able to capture real-time information as it arrives on the site. Additionally, we cannot capture the differences in moderation either, since we can only see the comments and threads until after the moderator either closes a conversation thread or deletes comments once anything has been posted on Reddit. This means that a lot of the data that we rely on to investigate illicit behavior is difficult to gather since it is deleted by the time that we get to it. Additionally, much of the differences in semantics that we are looking for (high rates of usage of terms that might spread misinformation) are not present in these moderated communities. Another limitation of this study is that the communities in which these discussions take place may get banned over time. Banning these communities means that there are challenges in accessing the data using the free API. This poses challenges, if we want to computationally detect abnormalities, there is an issue in that the abnormalities that we are looking for are in these mostly unmoderated communities. Alternatively, we had to make decisions in data collection about what data we wanted to collect in the different subreddits. We chose posts that had more than 5 comments simply because there are challenges in collecting all the posts. These challenges are more due to the time of data collection. The PushShift API allows for 20 API calls per second with hard limits on how many comments can be gathered at once per API call. For sampling and to make the data collection process sustainable in the case of any iterations, the limit had to be placed.

One of the motivations for this works is to try to understand how the emergence of the use of some term by an outside actor would impact term use in subsequent time frames. To that end, there has not been a solid conclusion to state that one term is being disseminated and that as it does it is adopted by other members of the community. This has to do with the fact that, visually, many comparison matrices demonstrated that once a term is coined for a concept (say Covid) it will persist. The structure of the graph pertaining to that term might change slightly over time. We can say that because there is some variation within the monthly matrices, however, none of these differences are significant enough to state that any one month was entirely different from others. Since there is no one month with large variations in the structure, we do not have enough grounding to comment on how term adoption changes after large disruptions in term usage.

6.3 Future Work

Future work in this area could involve investigating online communities that mimic the structure and usability of platforms such as Reddit but cater to right-wing individuals that have been pushed away from Reddit. Recent events have seen the emergence of Parler and TheDonald.win as platforms for Alt-Right members to convene and communicate. These communities would make a perfect place to find abnormalities because of their completely unmoderated nature which aims to ”fix” some of the issues in free speech that these individuals perceive to be present in popular platforms. Because of this, individuals using these platforms may feel more at liberty to spread ideas that are different from all other communities engaging in conversations around the same theme.

Another area for expansion would be towards the end of using snapshots provided by the PushShift API that includes snapshots of the entire Reddit site aggregated monthly. One issue with this kind of use would be the large overhead required towards computation of the resulting adjacency matrices that are created. With enough resources, any project can expand and use the algorithms' used here, among others, to apply different measures in a larger data set. This would also be a favorable way to compare algorithms' efficiency and efficacy in real-world networks.

References

[1] Naming the coronavirus disease (COVID-19) and the virus that causes it.

[2] Combatting Misinformation on Instagram, December 2019.

[3] Noor Abu-El-Rub and Abdullah Mueen. BotCamp: Bot-driven Interactions in Social Campaigns. In *The World Wide Web Conference on - WWW '19*, pages 2529–2535, San Francisco, CA, USA, 2019. ACM Press.

[4] Luca Maria Aiello, Martina Deplano, Rossano Schifanella, and Giancarlo Ruffo. People Are Strange When You're a Stranger. page 8.

[5] Hunt Allcott, Matthew Gentzkow, and Chuan Yu. Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2):2053168019848554, April 2019. Publisher: SAGE Publications Ltd.

[6] Malik Almaliki. Online Misinformation Spread: A Systematic Literature Map. In *Proceedings of the 2019 3rd International Conference on Information System and Data Mining - ICISDM 2019*, pages 171–178, Houston, TX, USA, 2019. ACM Press.

[7] A. A. Amleshwaram, N. Reddy, S. Yadav, G. Gu, and C. Yang. CATS: Characterizing automation of Twitter spammers. In *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS)*, pages 1–10, January 2013. ISSN: 2155-2509.

[8] A. Badawy, E. Ferrara, and K. Lerman. Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 258–265, August 2018. ISSN: 2473-991X.

[9] James P. Bagrow and Erik M. Bollt. An information-theoretic, all-scales approach to comparing networks. *Applied Network Science*, 4(1):45, December 2019. arXiv: 1804.03665.

[10] Abdul Mannan Baig. Chronic COVID syndrome: Need for an appropriate medical terminology for long-COVID and COVID long-haulers. *Journal of Medical Virology*, n/a(n/a), 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.26624.

[11] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221, September 2018. Publisher: National Academy of Sciences Section: Social Sciences.

[12] I. Bara, C. J. Fung, and T. Dinh. Enhancing Twitter spam accounts discovery using cross-account pattern mining. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pages 491–496, May 2015. ISSN: 1573-0077.

[13] Jeff Baumes, Mark Goldberg, Malik Magdon-Ismail, and William Al Wallace. Discovering Hidden Groups in Communication Networks. In Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Hsinchun Chen, Reagan Moore, Daniel D. Zeng, and John Leavitt, editors, *Intelligence and Security Informatics*, volume 3073, pages 378–389. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. Series Title: Lecture Notes in Computer Science.

[14] Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*, 21(11), November 2016.

[15] S. Y. Bhat and M. Abulaish. Community-based features for identifying spammers in Online Social Networks. In *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, pages 100–107, August 2013.

[16] P. V. Bindu, Rahul Mishra, and P. Santhi Thilagam. Discovering spammer communities in twitter. *Journal of Intelligent Information Systems*, 51(3):503–527, December 2018.

[17] P. V. Bindu, P. Santhi Thilagam, and Deepesh Ahuja. Discovering suspicious behavior in multilayer social networks. *Computers in Human Behavior*, 73:568–582, August 2017.

[18] Alessio Emanuele Biondo, Alessandro Pluchino, and Andrea Rapisarda. Informative Contagion Dynamics in a Multilayer Network Model of Financial Markets. *Italian Economic Journal*, 3(3):343–366, November 2017. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 3 Publisher: Springer International Publishing.

[19] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, November 2014.

[20] Jonathan Bright. Explaining the Emergence of Echo Chambers on Social Media: The Role of Ideology and Extremism. SSRN Scholarly Paper ID 2839728, Social Science Research Network, Rochester, NY, March 2017.

[21] Alessio Cardillo, Massimiliano Zanin, Jesús Gómez-Gardeñes, Miguel Romance, Alejandro J. García del Amo, and Stefano Boccaletti. Modeling the multi-layer nature of the European Air Transport Network: Resilience and passengers re-scheduling under random failures. *The European Physical Journal Special Topics*, 215(1):23–33, January 2013.

[22] Tak Kwong Chan. Universal masking for COVID-19: evidence, ethics and recommendations. *BMJ Global Health*, 5(5):e002819, May 2020.

[23] N. Chavoshi, H. Hamooni, and A. Mueen. DeBot: Twitter Bot Detection via Warped Correlation. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 817–822, December 2016. ISSN: 2374-8486.

[24] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The COVID-19 Social Media Infodemic. *Scientific Reports*, 10(1):16598, December 2020. arXiv: 2003.05004.

[25] Nadia K. Conroy, Victoria L. Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4, 2015. _eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/pra2.2015.145052010082.

[26] E. Cozzo, R. A. Baños, S. Meloni, and Y. Moreno. Contact-based Social Contagion in Multiplex Networks. *Physical Review E*, 88(5):050801, November 2013. arXiv: 1307.1656.

[27] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi. Social Fingerprinting: Detection of Spambot Groups Through DNA-Inspired Behavioral Modeling. *IEEE Transactions on Dependable and Secure Computing*, 15(4):561–576, July 2018. Conference Name: IEEE Transactions on Dependable and Secure Computing.

[28] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. BotOrNot: A System to Evaluate Social Bots. In *Proceedings of the 25th International Conference Companion on World Wide Web - WWW '16 Companion*, pages 273–274, Montréal, Québec, Canada, 2016. ACM Press.

[29] Wen-Bo Du, Xing-Lian Zhou, Marko Jusup, and Zhen Wang. Physics of transportation: Towards optimal capacity using the multilayer network framework. *Scientific Reports*, 6(1):19059, January 2016. Number: 1 Publisher: Nature Publishing Group.

[30] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The Rise of Social Bots. *Communications of the ACM*, 59(7):96–104, June 2016. arXiv: 1407.5225.

[31] Sheera Frenkel and Davey Alba. Misleading Virus Video, Pushed by the Trumps, Spreads Online. *The New York Times*, July 2020.

[32] Tiana Gaudette, Ryan Scrivens, Garth Davies, and Richard Frank. Upvoting extremism: Collective identity formation and the extreme right on Reddit. *New Media & Society*, page 1461444820958123, September 2020. Publisher: SAGE Publications.

[33] Sergio Gomez, Albert Diaz-Guilera, Jesus Gomez-Gardeñes, Conrad J. Perez-Vicente, Yamir Moreno, and Alex Arenas. Diffusion dynamics on multiplex networks. *Physical Review Letters*, 110(2):028701, January 2013. arXiv: 1207.2788.

[34] Christine Hauser and Johnny Diaz. F.D.A. Warns 7 Companies to Stop Claiming Silver and Other Products Treat Coronavirus. *The New York Times*, March 2020.

[35] S. Havlin, D. Y. Kenett, E. Ben-Jacob, A. Bunde, R. Cohen, H. Hermann, J. W. Kantelhardt, J. Kertész, S. Kirkpatrick, J. Kurths, J. Portugali, and S. Solomon. Challenges in network science: Applications to infrastructures, climate, social systems and economics. *The European Physical Journal Special Topics*, 214(1):273–293, November 2012.

[36] Alison L. Hill, David G. Rand, Martin A. Nowak, and Nicholas A. Christakis. Infectious Disease Modeling of Social Contagion in Networks. *PLoS Computational Biology*, 6(11), November 2010.

[37] Sofia Hurtado, Poushali Ray, and Radu Marculescu. Bot Detection in Reddit Political Discussion. In *Proceedings of the Fourth International Workshop on Social Sensing - SocialSense'19*, pages 30–35, Montreal, QC, Canada, 2019. ACM Press.

[38] Jelani Ince, Fabio Rojas, and Clayton A. Davis. The social media response to Black Lives Matter: how Twitter users interact with Black Lives Matter through hashtag use. *Ethnic and Racial Studies*, 40(11):1814–1830, September 2017. Publisher: Routledge _eprint: https://doi.org/10.1080/01419870.2017.1334931.

[39] Tiffany Karalis Noel. Conflating culture with COVID-19: Xenophobic repercussions of a global pandemic. *Social Sciences & Humanities Open*, 2(1):100044, January 2020.

[40] Sameed Ahmed M. Khatana and Peter W. Groeneveld. Health Disparities and the Coronavirus Disease 2019 (COVID-19) Pandemic in the USA. *Journal of General Internal Medicine*, 35(8):2431–2432, August 2020.

[41] Majid Khosravinik. Right Wing Populism in the West: Social Media Discourse and Echo Chambers. *Insight Turkey*, 19, July 2017.

[42] Sneha Kudugunta and Emilio Ferrara. Deep neural networks for bot detection. *Information Sciences*, 467:312–322, October 2018.

[43] K. Lewis, M. Gonzalez, and J. Kaufman. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1):68–72, January 2012.

[44] Liqiang Li and Jing Liu. The aggregation of multiplex networks based on the similarity of networks. *Physica A: Statistical Mechanics and its Applications*, 540:122976, February 2020.

[45] Greeshma Lingam, Rashmi Ranjan Rout, Dvln Somayajulu, and Sajal K. Das. Social Botnet Community Detection: A Novel Approach based on Behavioral Similarity in Twitter Network using Deep Learning. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*, pages 708–718, Taipei Taiwan, October 2020. ACM.

[46] December Maxwell, Sarah R. Robinson, Jessica R. Williams, and Craig Keaton. "A Short Story of a Lonely Guy": A Qualitative Thematic Analysis of Involuntary Celibacy Using Reddit. *Sexuality & Culture*, 24(6):1852–1874, December 2020.

[47] Jonathan Mellon and Christopher Prosser. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 4(3):2053168017720008, July 2017. Publisher: SAGE Publications Ltd.

[48] Roland Molontay and Marcell Nagy. Twenty Years of Network Science: A Bibliographic and Co-Authorship Network Analysis. *arXiv:2001.09006 [physics]*, June 2020. arXiv: 2001.09006.

[49] F. Morstatter, L. Wu, T. H. Nazer, K. M. Carley, and H. Liu. A new approach to bot detection: Striking the balance between precision and recall. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 533–540, August 2016.

[50] Mariam Orabi, Djedjiga Mouheb, Zaher Al Aghbari, and Ibrahim Kamel. Detection of Bots in Social Media: A Systematic Review. *Information Processing & Management*, 57(4):102250, July 2020.

[51] Gordon Pennycook, Tyrone D. Cannon, and David G. Rand. Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology. General*, 147(12):1865–1880, December 2018.

[52] Bobbie Person, Francisco Sy, Kelly Holton, Barbara Govert, Arthur Liang, Brenda Garza, Deborah Gould, Meredith Hickson, Marian McDonald, Cecilia Meijer, Julia Smith, Liza Veto, Walter Williams, and Laura Zauderer. Fear and Stigma: The Epidemic within the SARS Outbreak - Volume 10, Number 2—February 2004 - Emerging Infectious Diseases journal - CDC. 2004.

[53] Andrew S. Ross and Damian J. Rivers. Discursive Deflection: Accusation of "Fake News" and the Spread of Mis- and Disinformation in the Tweets of President Trump. *Social Media + Society*, 4(2):2056305118776010, April 2018. Publisher: SAGE Publications Ltd.

[54] Tiago A. Schieber, Laura Carpi, Albert Díaz-Guilera, Panos M. Pardalos, Cristina Masoller, and Martín G. Ravetti. Quantification of network structural dissimilarities. *Nature Communications*, 8(1):13928, January 2017. Number: 1 Publisher: Nature Publishing Group.

[55] Desirée Schmuck and Christian von Sikorski. Perceived threats from social bots: The media's role in supporting literacy. *Computers in Human Behavior*, 113:106507, December 2020.

[56] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kaicheng Yang, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature Communications*, 9(1):4787, December 2018. arXiv: 1707.07592.

[57] Galen Stolee and Steve Caton. Twitter, Trump, and the Base: A Shift to a New Form of Presidential Talk? *Signs and Society*, 6(1):147–165, January 2018.

[58] Denis Stukal, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. Detecting Bots on Russian Political Twitter. *Big Data*, 5(4):310–324, December 2017. Publisher: Mary Ann Liebert, Inc., publishers.

[59] Qawi K. Telesford, Sean L. Simpson, Jonathan H. Burdette, Satoru Hayasaka, and Paul J. Laurienti. The Brain as a Complex System: Using Network Science as a Tool for Understanding the Brain. *Brain Connectivity*, 1(4):295–308, October 2011. Publisher: Mary Ann Liebert, Inc., publishers.

[60] Aybike Ulusan and Ozlem Ergun. Restoration of services in disrupted infrastructure systems: A network science approach. *PLOS ONE*, 13(2):e0192272, February 2018. Publisher: Public Library of Science.

[61] Christophe Van den Bulte and Yogesh V. Joshi. New Product Diffusion with Influentials and Imitators. *Marketing Science*, 26(3):400–421, May 2007. Publisher: INFORMS.

[62] J. Wang and I. C. Paschalidis. Botnet Detection Based on Anomaly and Community Detection. *IEEE Transactions on Control of Network Systems*, 4(2):392–404, June 2017. Conference Name: IEEE Transactions on Control of Network Systems.

[63] Hywel T. P. Williams, James R. McMurray, Tim Kurz, and F. Hugo Lambert. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, 32:126–138, May 2015.

[64] Liang Wu, Fred Morstatter, Kathleen M. Carley, and Huan Liu. Misinformation in Social Media: Definition, Manipulation, and Detection. *ACM SIGKDD Explorations Newsletter*, 21(2):80–90, November 2019.

[65] Xiaoyi Yuan, Ross J. Schuchard, and Andrew T. Crooks. Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter. *Social Media + Society*, 5(3):2056305119865465, April 2019. Publisher: SAGE Publications Ltd.

[66] Jason Shuo Zhang, Brian C. Keegan, Qin Lv, and Chenhao Tan. A Tale of Two Communities: Characterizing Reddit Response to COVID-19 through /r/China_flu and /r/Coronavirus. *arXiv:2006.04816 [cs]*, June 2020. arXiv: 2006.04816.

[67] Jinxue Zhang, Rui Zhang, Yanchao Zhang, and Guanhua Yan. The Rise of Social Botnets: Attacks and Countermeasures. *arXiv:1603.02714 [cs]*, April 2016. arXiv: 1603.02714.

[68] Dawei Zhao, Lianhai Wang, Shudong Li, Zhen Wang, Lin Wang, and Bo Gao. Immunization of Epidemics in Multiplex Networks. *PLOS ONE*, 9(11):e112018, November 2014. Publisher: Public Library of Science.