

# Like a virus

## The coordinated spread of coronavirus disinformation

---

*Analysis of over 25.5 million tweets over 10 days identifies 5,752 accounts that coordinated 6,559 times to spread mis- and disinformation regarding the coronavirus for either commercial or political purposes. Almost all politically motivated activity promoted right wing governments or parties. Coordinated spreading of the China bioweapon conspiracy theory is estimated to have made over 5 million impressions on Twitter users, spread by mainly pro-Trump, partisan conservative and/or QAnon accounts.*

### Authors

Timothy Graham, Digital Media Research Centre (DMRC),

Queensland University of Technology

Axel Bruns, DMRC

Guangnan Zhu, DMRC

Rod Campbell, The Australia Institute

May 2020

## About The Australia Institute

The Australia Institute is an independent public policy think tank based in Canberra. It is funded by donations from philanthropic trusts and individuals and commissioned research. We barrack for ideas, not political parties or candidates. Since its launch in 1994, the Institute has carried out highly influential research on a broad range of economic, social and environmental issues.

## About the Centre for Responsible Technology

The Australia Institute established the Centre for Responsible Technology to give people greater influence over the way technology is rapidly changing our world. The Centre will collaborate with academics, activists, civil society and business to shape policy and practice around network technology by raising public awareness about the broader impacts and implications of data-driven change and advocating policies that promote the common good.

## Our philosophy

As we begin the 21st century, new dilemmas confront our society and our planet. Unprecedented levels of consumption co-exist with extreme poverty. Through new technology we are more connected than we have ever been, yet civic engagement is declining. Environmental neglect continues despite heightened ecological awareness. A better balance is urgently needed.

The Australia Institute's directors, staff and supporters represent a broad range of views and priorities. What unites us is a belief that through a combination of research and creativity we can promote new solutions and ways of thinking.

## Our purpose - 'Research that matters'

The Institute publishes research that contributes to a more just, sustainable and peaceful society. Our goal is to gather, interpret and communicate evidence in order to both diagnose the problems we face and propose new solutions to tackle them.

The Institute is wholly independent and not affiliated with any other organisation. Donations to its Research Fund are tax deductible for the donor. Anyone wishing to donate can do so via the website at <https://www.tai.org.au> or by calling the Institute on 02 6130 0530. Our secure and user-friendly website allows donors to make either one-off or regular monthly donations and we encourage everyone who can to donate in this way as it assists our research in the most significant manner.

Level 1, Endeavour House, 1 Franklin St  
Canberra, ACT 2601  
Tel: (02) 61300530  
Email: [mail@tai.org.au](mailto:mail@tai.org.au)  
Website: [www.tai.org.au](http://www.tai.org.au)  
ISSN: 1836-9014

# Summary

The coronavirus pandemic has been paralleled by an ‘infodemic’ of mis- and disinformation and conspiracy theories about the virus. This study analysed 2.6 million tweets relating to coronavirus and their 25.5 million retweets over 10 days from late March 2020. Two different approaches were taken.

First, accounts that retweet identical coronavirus-related content repeatedly within one second of each other were identified. Such accounts are very likely to be highly-automated or ‘bot’ accounts, controlled by computers rather than humans. 5,752 accounts retweeted material in a coordinated way (co-retweet) 6,559 times.

Ten different bot-like networks are described, some of which appear to push political agendas while others seem designed to promote commercial sites via coronavirus related Twitter trends. Political use of coronavirus from these most obviously bot-like accounts was prominent among Turkish accounts, focusing on disputes between President Erdoğan and other Turkish leaders leading up to the country’s curfew at the end of the analysis period. Paraguayan accounts also appear to co-retweet links and tweets to astroturf for the government, or to magnify fear about mortality and infection rates within Paraguay. Another bot account network promotes positive tweets about the Saudi Arabian Government and the Saudi Crown Prince, while another promotes hyper-partisan criticism of the Spanish Government, which is led by the Spanish Socialist Workers’ Party.

The second part of the analysis identified co-ordinated efforts to promote a particular coronavirus conspiracy theory – that it was engineered as a bioweapon by China. With this narrower scope, the method to identify co-ordination was expanded from one second to one minute and so likely captures fully automated bot accounts as well as more hybrid automated-human accounts and perhaps some fully human accounts operating in close co-ordination. A co-retweet network of 2,903 accounts and 4,125 links or co-retweets between them was identified. Within this network, the top 30 clusters were analysed manually to identify group identities within each cluster.

Of these 30 clusters, 28 identify as Pro-Trump, QAnon, and/or Republican partisan. The two exceptions were a cluster of broadly pro-Democrat accounts and a cluster of ‘Baloch National Movement’ accounts that heavily pushed anti-Pakistan content.

The co-ordinated efforts to promote the bioweapon conspiracy theory focused on 882 original tweets, which were retweeted 18,498 times and liked 31,783 times, creating an estimated 5 million impressions on Twitter users. Similar research in January suggests there is a sustained, co-ordinated effort to promote this theory by pro-Trump, Republican and aligned networks of accounts.

Whether the coordinated inauthentic behaviours we have observed for the bioweapon conspiracy are orchestrated by the hard core of participants in these groups themselves, or are designed by external operators to target and exploit the worldviews of such groups, the net effect is often the same: the themes and topics promoted by coordinated inauthentic activity are taken up by the wider fringe community, and thereby gains amplification and authenticity: the mis- and disinformation contained in the initial messages is no longer distributed solely by bots and other accounts that may be identified as acting in coordinated and inauthentic ways, but also and to a potentially greater extent by ordinary, authentic human users.

From this phase, mis- and disinformation can easily cross into the wider public when mainstream media and prominent social media actors engaged with the conspiracy theory, even critically. Engagement by celebrities, journalists, politicians, media outlets, state authorities, and others with large followings substantially amplifies the visibility of the conspiracy theory. Official denials and corrections can perversely be exploited by the conspiracy theorists to claim that authorities are covering up 'the real truth'. In Australia, for example, the effects of this vicious circle are now being observed in the sharp rise in concerns about 5G technology, at least in part as a result of the circulation of the conspiracy theories about links between COVID-19 and 5G that we have documented in this report.

Governments, non-government actors and technology platforms should address co-ordinated disinformation campaigns such as these through:

- **Increased detection and mitigation.** Our analysis demonstrates that such behaviour can be detected by technical means, and most social media platforms are using detection tools and suspensions to a greater or lesser extent. Independent critical investigation by scholarly social media researchers is crucial in this context, both to develop new and innovative detection approaches and to track and evaluate the activities of the platform operators themselves.
- **Digital literacy.** While there are a number of digital media literacy initiatives in train, there is a significant need for further funding and institutional support for such initiatives at all levels, and for all age groups.
- **Mainstream media.** Media outlets should be encouraged to reduce click-bait conspiracy theory coverage, which puts substantial new audiences in contact with mis- and disinformation. Coverage of official responses to this content also needs to be cautious not to contribute to the spreading of conspiracy theories.
- **Scholarly Research.** Studies such as this draw on advanced computational methods and forensic qualitative analysis of large-scale, real-time social media data. Such work requires secure funding and access to data. Data access is increasingly constrained by the leading social media platforms.

# Introduction

The trajectory of the coronavirus outbreak, epidemic, and eventual pandemic since late 2019 is paralleled by the emergence and increasing circulation of rumours, conjecture, inadvertent misinformation and deliberate disinformation about the causes, spread, and severity of, and potential remedies for the virus. The circulation of such information of dubious origins and limited veracity around the world has grown to such an extent that the World Health Organisation (WHO) has adopted the term 'infodemic' (United Nations, 2020) to describe it - and although it is easy to overstretch such metaphors, some of the mis- and disinformation associated with the pandemic has spread through social networks and mainstream media in similar ways as the pandemic itself has spread through the population.

In this report we examine a major conspiracy theory related to the pandemic, namely that China bioengineered the coronavirus as a weapon and it was either accidentally or strategically released from a virology lab in Wuhan. We highlight especially the role of what appear to be coordinated campaigns to kickstart the spread of such conspiracy theories, including the '5G internet causes coronavirus' hoax, pointing to the activities of bots on Twitter that seek to make specific themes and stories visible to a larger population of social media users by targeting the platform's 'trending topics' algorithms.

While we focus here on Twitter, it should be noted that such coordinated inauthentic behaviour is by no means limited to this particular platform (indeed, the term 'coordinated inauthentic behaviour' itself was coined by Facebook to describe similar activities there; Gleicher, 2018), nor likely to occur only in the opening stages of major events; our analysis here therefore also serves as a demonstration both of a much more general concern, and of the research approaches that can uncover such coordinated activity.

Similarly, infodemics such as those surrounding the COVID-19 crisis are not solely a concern for social media platforms and their operators: although the early spread of conspiracy theories and other mis- and disinformation may occur here first, they reach larger audiences only once they are also amplified by reporting in mainstream media and endorsements by celebrities and other influencers. Indeed, even critical reporting, official corrections, and journalistic fact-checking may be ineffective in slowing the spread of such mis- and disinformation: research shows that conspiracy theorists are emboldened in their activities and hardened in their worldviews by pushback from the authorities that they already regard as part of 'deep state' conspiracies (e.g. Bail et al., 2018; Krämer, 2017).

This means that, although the detection of coordinated inauthentic behaviour designed to spread conspiracy theories is a crucial element of the overall response, infodemics cannot be prevented by such means alone. Platform providers, as well as research institutions in digital and social media fields, can and do play a critical role in developing and deploying

such detection mechanisms and stopping inauthentic activities before they can cause harm; however, in addition to such deliberate and inauthentic campaigns to spread mis- and disinformation through automated means, the authentic activities of ordinary social media users, and especially of individual and institutional influencers who have amassed large social media followings, also contribute substantially to the dissemination of problematic content. The latter cannot be addressed effectively by technological solutions, but instead requires a longer-term effort aimed especially at developing advanced digital media literacies throughout the online population.

In short, coordinated inauthentic behaviour on social media platforms is problematic not simply because it is designed to circulate mis- and disinformation in its own right - rather, its insidious nature lies in the fact that it aims to kickstart the circulation of conspiracy theories and other mis- and disinformation amongst real, authentic users. The coordinated inauthentic activities we describe here seek to exploit the concerns, fears, and prejudices of ordinary, human social media users, and thereby to enrol these users in doing the work of the conspiracy theorists themselves.

# Methods

## DATA COLLECTION

The dataset in this study was collected via the Twitter Streaming Application Programming Interface (API), spanning just over 10 days from 30 Mar. 2020 to 9 Apr. 2020. It includes 25,464,662 retweets of 2,626,429 tweets. The collector retrieved a streaming sample of tweets containing the general coronavirus-related hashtags listed below. These hashtags were selected through a snowballing approach that began with identifying the globally trending hashtags at the time (#coronavirus and #COVID—19) and iteratively analysing tweets containing these hashtags to find other relevant general hashtags that were being used to organise the coronavirus-related discussions on Twitter, including in Australia.

Hashtags and key terms in the data collection via the Twitter Streaming API were:

- #coronavirus
- #coronacrisis
- #covid19
- #COVID—19
- #COVID-19
- #covid19aus

## BOT OR NOT

In this analysis we search for, and uncover, evidence of bots coordinating together to retweet coronavirus-related content. Specifically, we analyse Twitter accounts who retweet the same tweets *within 1 second of each other* multiple times, a form of coordinated communication known as ‘co-retweeting’ (Keller et al., 2020). Whilst it is perfectly normal for multiple accounts to retweet the same tweet (e.g. when a public figure or celebrity such as @elonmusk sends a tweet), it is very unlikely that two accounts will frequently retweet the same tweet (Keller et al., 2020) or the same URL (cf. Giglietto et al., 2019; 2020) within a short period of time of each other. *Even less likely* are two accounts co-retweeting within 1 second of each other (i.e. at the exact same time or 1 second apart).

Accounts that co-retweet at inhuman speeds are very likely to be controlled by a computer, known as ‘bot’ accounts (Graham & Ackland, 2016), or at the least partly computer

controlled (i.e. hybrid accounts) whereby a human operator relies on automated tools such as tweet schedulers and scripting techniques to produce content on the account. The result of this analysis is what we term as a '*bot-like*' *co-retweet network* (shown in Figure 1), which consists of 5,752 accounts and a total of 6,559 co-retweets.

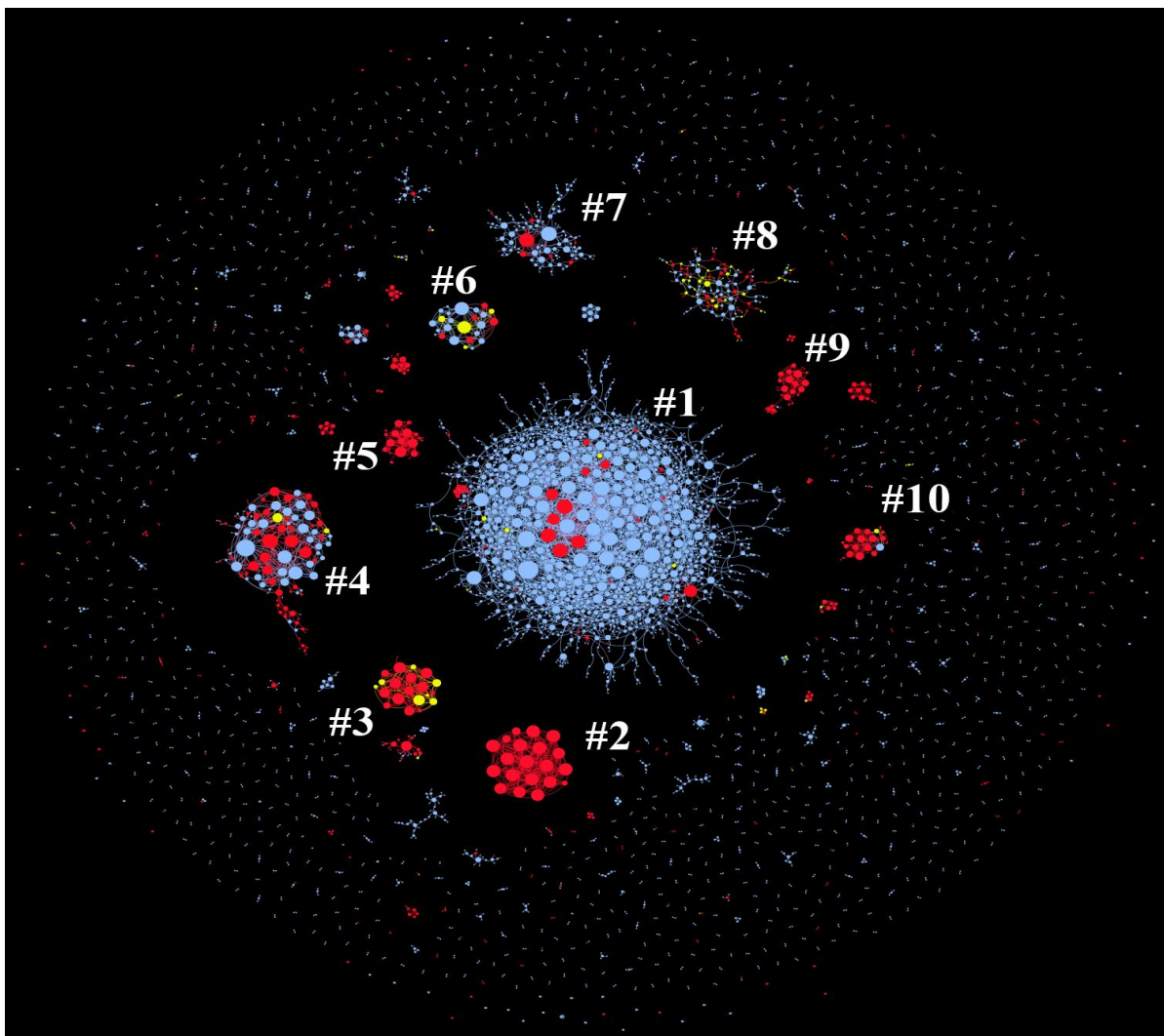
To detect bots we can also use software tools such as Botometer and TweetBotOrNot2, which use machine learning to analyse a given account and produce a metric that estimates how likely it is to be an automated account. Although these tools cannot uncover evidence of coordination, they are useful at the individual account level to help distinguish bots and humans. To augment our analysis, we use TweetBotOrNot2 to analyse each account in the co-retweet network and produce a bot score between 0 and 1, where 0 is extremely unlikely to be a bot and 1 is extremely likely to be a bot. TweetBotOrNot2 is a very conservative bot detection tool - it has state-of-the-art accuracy, but tends to find less bots. In short, it tends to only flag accounts as bots that are genuinely automated accounts, rather than hybrid accounts or human-controlled troll accounts that behave inauthentically but without the use of automation.



# Coordinated bot behaviour

Figure 1 shows the bot-like co-retweet network, where nodes are pairs of accounts who are connected together if they both retweeted the same tweet within 1 second of each other. The network is weighted, so the links (or 'edges') in the network are thicker if the accounts co-retweeted more often. The colours represent human (blue; score less than 0.2), bot (red; score greater than 0.8) and indeterminate (yellow; score between 0.2 and 0.8). The numbers in Figure 1 represent different clusters of interest within the network. Each of the ten clusters highlighted reveals various kinds of coordinated bot-like activity, some of which is inauthentic.

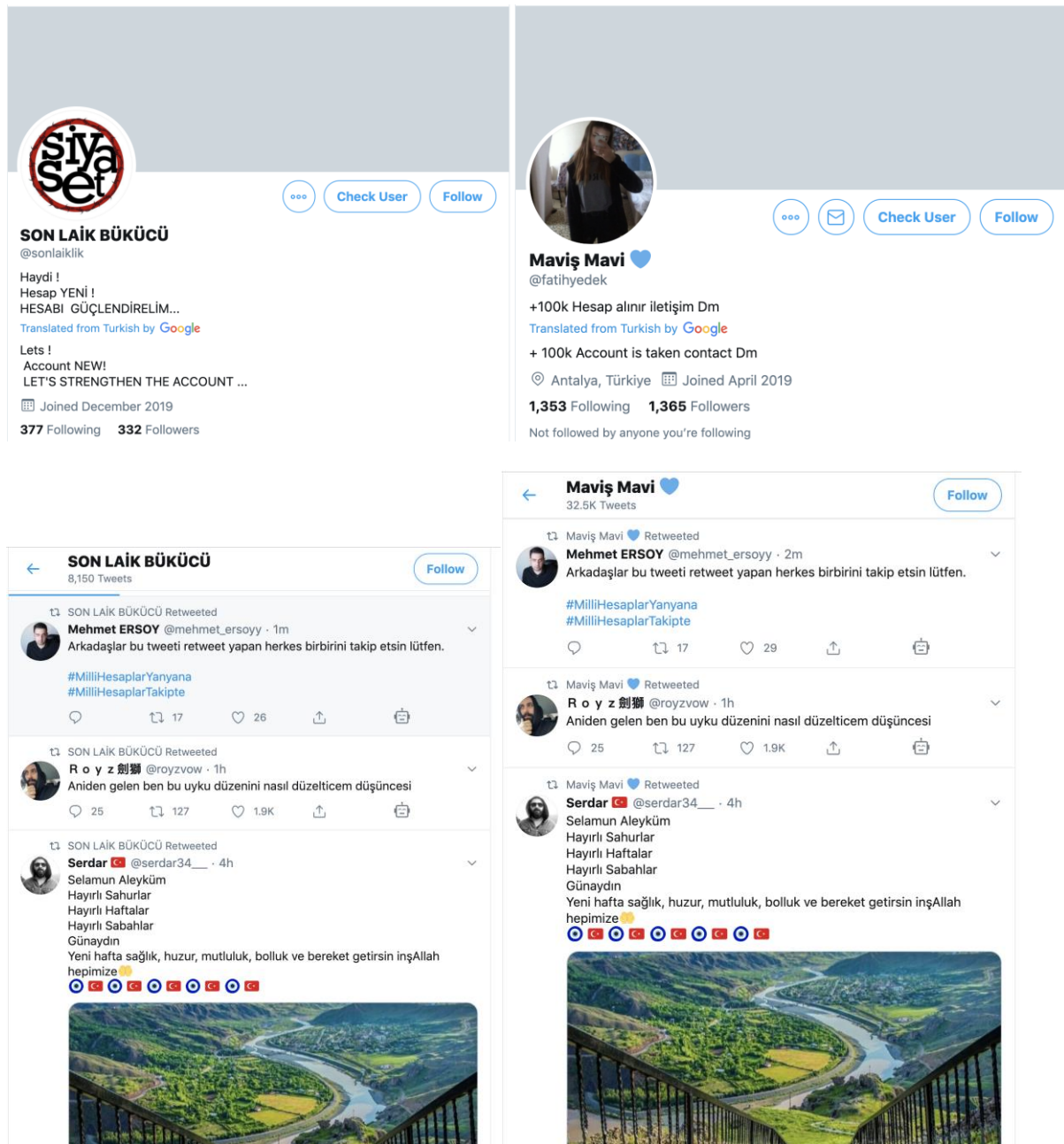
**Figure 1 - Bot-like co-retweet network**



The largest cluster (Cluster #1) in the middle of the network consists mainly of Paraguayan and Turkish accounts. Many of these accounts, including the most active ones, appear to be compromised accounts that may have been hacked and repurposed into bot networks -

they appear like real accounts but their recent retweeting behaviour and frequency is not characteristic of genuine organic activity. For example, at the centre of this cluster are networks of accounts such as @sonlaiklik and @fatihyedek (Figure 2 below) that post identical retweets at exactly the same time. These are ‘textbook’ examples of bots that automatically and coordinatedly retweet the same content, presumably to build up follower counts and be rented out or sold to paying customers for political and/or economic purposes.

Figure 2 - Turkish retweet bots and example of their recent coordinated activity



The central Paraguayan accounts in this cluster, including those also flagged as bots by tweetbotornot2, tend to focus on negative content such as death and infections rate of COVID-19, and they often retweet official accounts of Ministers and agencies with the right-

wing Paraguayan government. For example, they often retweet graphs and statistics tweeted by Julio Mazzoleni (<https://twitter.com/MazzoleniJulio>), Minister of Public Health and Social Welfare of the Republic of Paraguay, and <https://twitter.com/msaludpy>, the official account of the Ministry of Public Health and Social Welfare of Paraguay. We speculate that this behaviour serves to either ‘astroturf’ Twitter to show support for the current government and/or magnify fear and sow discord about the coronavirus and its mortality and infection rates within Paraguay. Many of these accounts retweet sites including *Russia Today (RT)*, which is a Russian state media outlet. Note: at the time of writing there are conflicting reports about whether Russia is engaged in a coronavirus disinformation campaign.

The central Turkish accounts in this cluster tend to focus on aggravating political tensions and divisions relating to how the Turkish government handled restrictions and curfews during the coronavirus pandemic. These accounts focus on the dispute between far-right President Recep Tayyip Erdoğan and Istanbul's more moderate opposition mayor, Ekrem Imamoglu, who both launched their own campaigns to fund pandemic relief efforts and had differing opinions about when and how to implement curfews to stop the spread of the coronavirus. The dataset in this study covers roughly the 10 day period leading up to a 48-hour curfew that was imposed in 31 Turkish cities, instigated by Minister Soylu on 10 April 2020. The curfew was announced only a few hours before its starting time and caused a panic buying in the country (Dwyer, 2020), leading to Minister Soylu announcing his resignation (although this was ultimately refused by President Erdoğan). We speculate that the coordinated bots and suspicious hybrid accounts discovered in this analysis were aiming to magnify political tensions in Turkey and sow confusion and anxiety about the spread of the coronavirus and its effective response from different (and competing) levels of government in the country.

Cluster #2 is a bot network that appears to be primarily economically motivated, as the accounts all retweet prize giveaway tweets (as shown below). However, in this dataset the accounts are all co-retweeting a tweet from a conspiracy theory troll account, @Greasy97597537 (<https://twitter.com/Greasy97597537>). The tweet has since been deleted or removed by Twitter, but the text is:

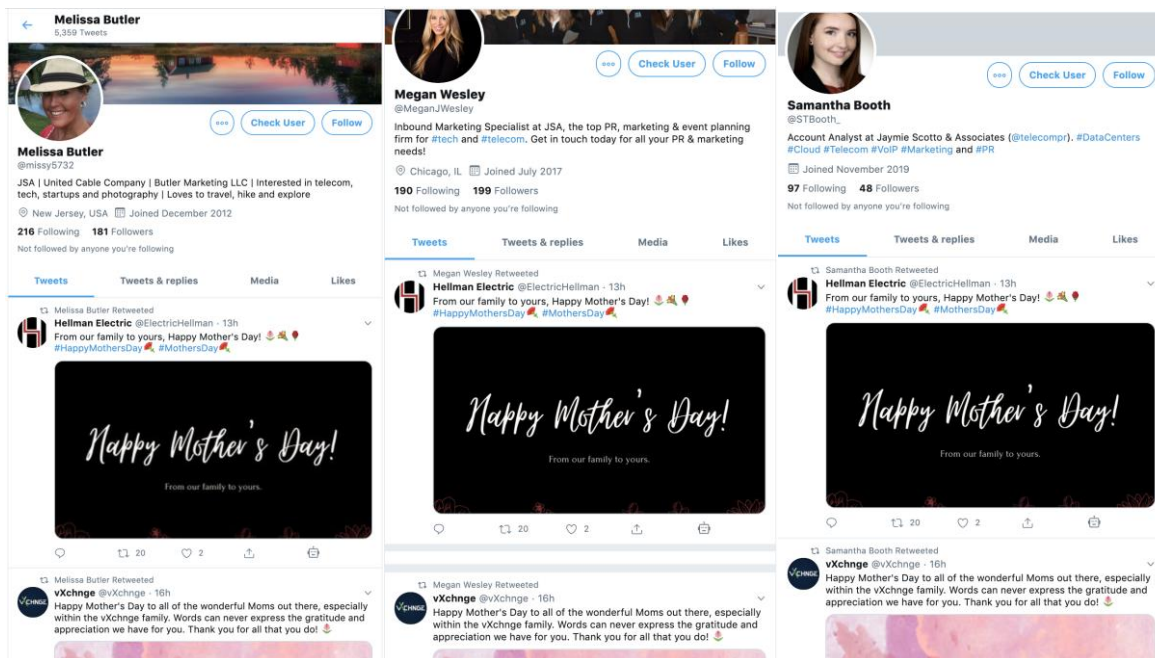
It's national #Vietnam #Veterans Day in #USA Time to look at the success of the #Vietnamese vs #CoronaVirus VIETNAMVN - Close to Wuhan CN VERY FEW INFECTED 🤒. Population 97 million 🧑 EVERY SINGLE PERSON: Wearing Masks 🧐 Carrying Sanitiser 🧼 #retweet #victory #example #win <https://t.co/uN8eb67Y06>

It is likely that the bots in Cluster #2 simply retweeted the tweet because it contained the hashtag “#win”. We therefore cannot prescribe any politically motivated malicious activity to this cluster, although this cluster demonstrates how bots can amplify unwanted or

problematic content on Twitter, and either Twitter doesn't care about the existence of this particular activity on its platform or (more likely) it isn't aware of the problem.

Cluster #3 is a bot network that spams content on behalf of the marketing company, *Jaymie Scott & Associates (JSA)*. Whilst all the accounts retweet the same tweets at the same time, strangely they all appear to be real employees of the company, despite having ceded complete control of their Twitter accounts to bots (see Figure 3). Perhaps the company requests its employees to allow them to take over control, i.e. to use the account as a network of self-promotional 'sockpuppets'. Few of the network's tweets relate directly to the coronavirus crisis itself, but some promote the company's online events during the crisis.

**Figure 3 - Commercial sockpuppet bots spamming identical content for JSA (a company)**



Cluster #4 is a Turkish bot network and many of the leading accounts are now suspended by Twitter, for example @baharrrrjjkk and @onurarslan181. These accounts primarily retweeted Zehra Neşe Kavak, a high profile doctor and obstetrician based in Istanbul (<https://twitter.com/zehranesekavak>) and Academic Hospital, a medical facility based in Istanbul (<https://twitter.com/AcademichSptl>). The motivations of this bot network are not clear, but it appears to be pushing a pro-Istanbul narrative to support the Mayor Ekrem Imamoglu's efforts to unilaterally respond to the coronavirus pandemic crisis within the city, in opposition to President Erdogan's national response that has attracted criticism. We speculate that the bot network seeks to astroturf false support for Imamoglu and/or to exacerbate political divisions through mis- and disinformation (spreading true content and news stories about Istanbul's success in fighting COVID-19 that drowns out other points of view or shows a biased perspective). In this way, Cluster #4 push similar narratives to many

of the Turkish bot accounts in the large central Cluster #1, although they focus on specific content - what appears to be positive medical information coming out of Istanbul.

Cluster #5 is a group of self-identified bots who retweet content with particular keywords or hashtags. There appears no coherent coordination of these accounts aside from their connection to several coronavirus related bots - they simply create spam and noise on Twitter, in some cases for humorous purposes, but otherwise have no clear group identity or motivations. Examples include:

- <https://twitter.com/coxinhabot>
- [https://twitter.com/Bot\\_Corona\\_V](https://twitter.com/Bot_Corona_V)

Cluster #6 is bots retweeting tweets about coronavirus treatments and how the virus works, however they also retweet a fake medical faculty account (<https://twitter.com/MedicinaUAQ>) that tweets extreme content including photos of children and adults with bodily disfigurements and diseases. There is also a focus on mis- and disinformation: retweeting statistics of coronavirus deaths in Spain and Mexico. However, it is not entirely clear if there is malicious intent driving this retweeting behaviour.

Cluster #7 is an unremarkable group of bots that retweet prize and giveaway competitions, but are otherwise a mix of different kinds of accounts with no links other than robotically responding to tweets with key-terms relating to giveaways. Some of the accounts might be hijacked sockpuppets, based on the profiles.

Cluster #8 is a pro-Saudi Arabic-language bot network that mainly retweets positive tweets about the Saudi government and the Crown Prince of Saudi Arabia, Mohammed bin Salman bin Abdulaziz al-Saud (see example tweet below, with English translation). Various accounts in this network also post a mix of Islamic religious messages and platitudes, as well as clickbait videos designed to gain attention and likes, such as funny videos of animals, but also accidents, traffic incidents and sometimes videos of assaults or crime. Many of these accounts pretend to be real people, but they post exactly the same content at the same time, as shown in the screenshots below in Figure 4.

Figure 4 - Pro-Saudi bot network

خيال Retweeted  
 علي العويبي @Ali\_alowifi

نثق بحكمة قيادتنا بإننا سنتجاوز هذه الأزمة العالمية فقد جعلت  
 #الإنسان\_أولاً وأكرمت #السعودي بالداخل والخارج واهتمت  
 بسلامته وصحته ووظيفته ليعيش بسلام .. فحفظ الله الوطن  
 وقيادته وأدام هذا الشعب الوفي الذي دائماً يثبت صدق الولاء  
 والانتماء لذلك نقول :  
 #سمعا\_وطاعه\_في\_المنشط\_والمكره

Translated from Arabic by Google

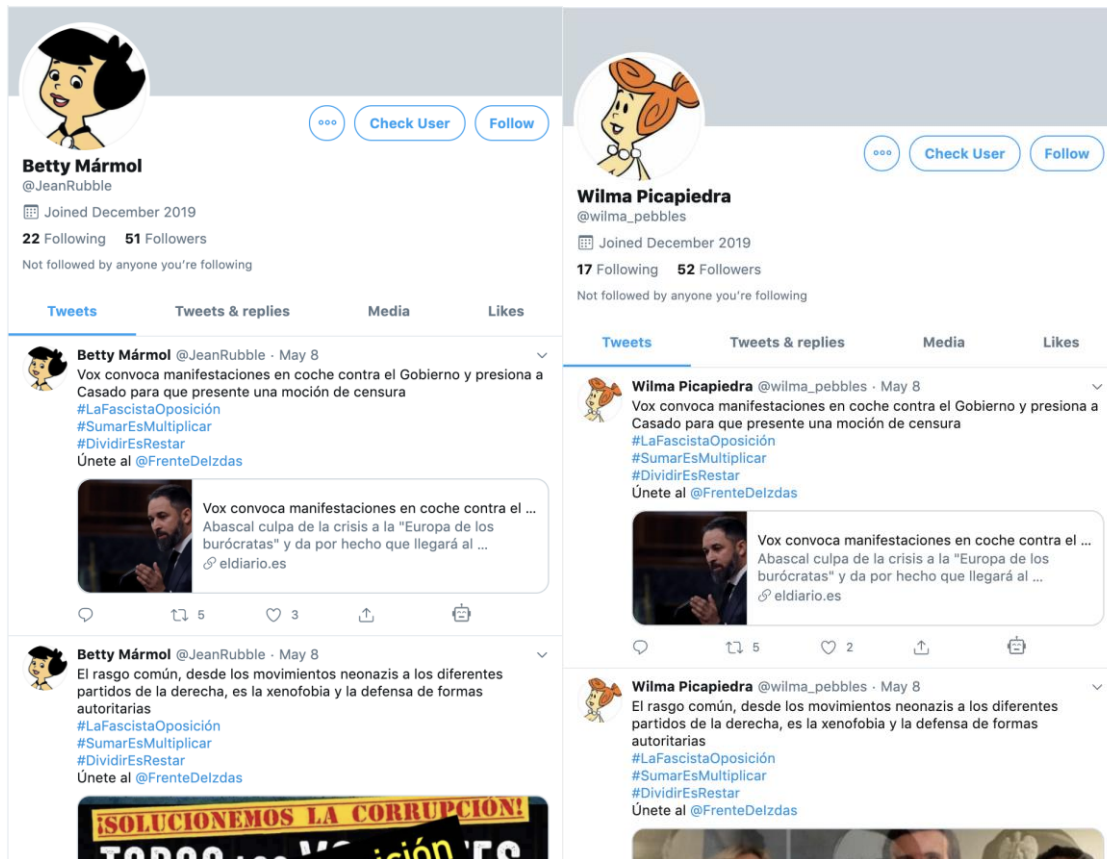
We believe in the wisdom of our leadership that we will overcome this global crisis it has made  
 #الإنسان\_أولاً And honored #السعودي At home and abroad, and cared for his safety, health and job to live in peace ..  
 God preserved the nation and its leadership and perpetuated this loyal people who always prove the sincerity of loyalty and affiliation so we say:  
 #سمعا\_وطاعه\_في\_المنشط\_والمكره



Cluster #9 is a network of self-identified bots that is part of a coding and data science community, including a ‘100 days of coding’ event that was occurring. These accounts retweet content focussed on coding and coronavirus, such as using AI to understand the spread of the virus and its mechanisms on the body. Some of the bot accounts also repost news articles about coding and IT security.

Cluster #10 is a network of bots that try to amplify political tensions in Spain by retweeting hyper-partisan content relating to the handling of the coronavirus by the government, and generally post hyper-partisan criticism of the government, including memes and visual imagery portraying the government as fascists. Several of the accounts retweet each other and also copy and paste text from each other whilst making small modifications to the content (e.g. swapping the images), as shown below in Figure 5.

Figure 5 - Spanish bots with Flintstones imagery as light-hearted cover for disinformation



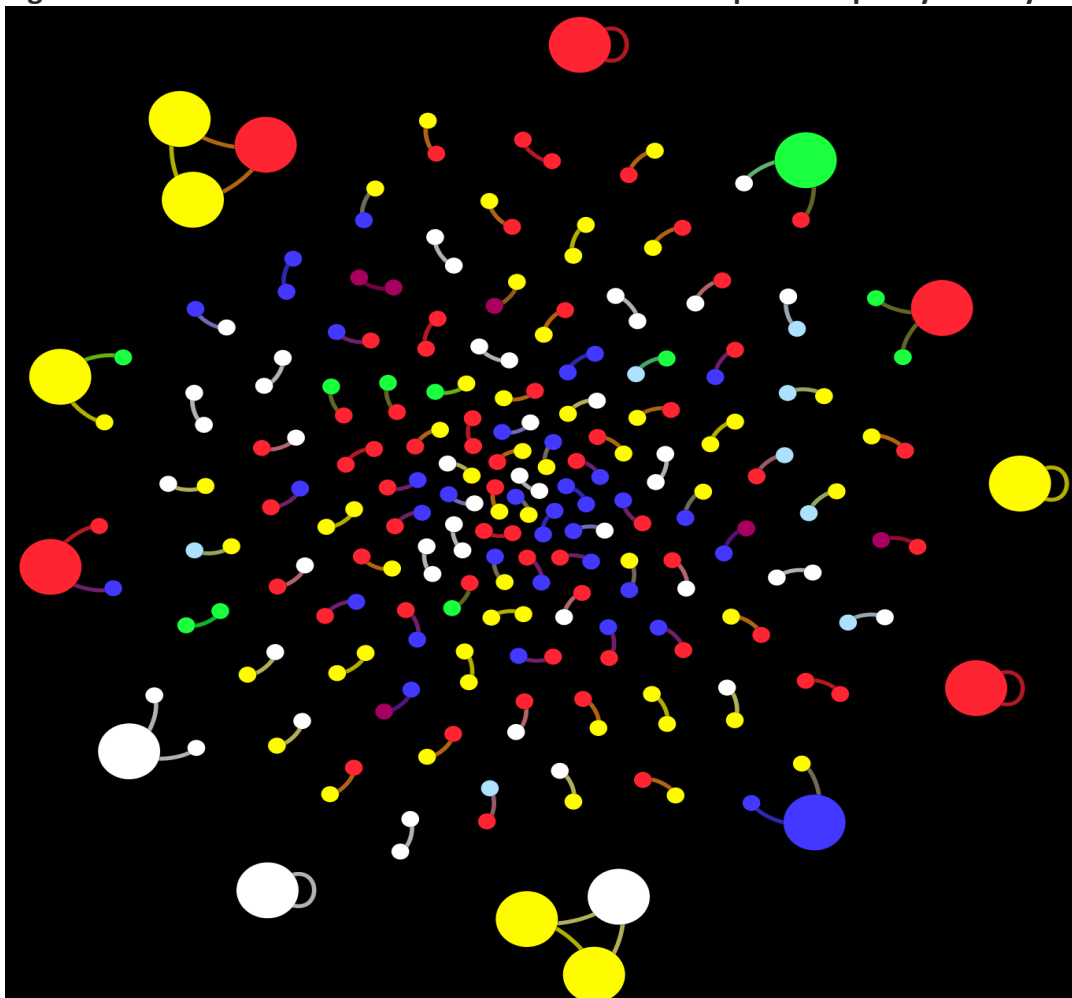
Note: picapiedra means 'Flintstones' in Spanish.

# Conspiracy theory analysis

## Coronavirus bioweapon conspiracy

This conspiracy theory emerged in January 2020, and claims that the coronavirus is an artificial biological weapon that was manufactured by the Chinese government in the Wuhan Institute of Virology (a.k.a. the Wuhan National Biosafety Laboratory). To date there is no scientific evidence to support this idea and it has come under criticism from WHO and other leading health organisations. We conducted two forms of co-retweet analysis to examine amplification of this conspiracy theory - using the 'bot' threshold (bot-like co-retweets in 1 second or less) and standard approach by Keller et al. (2020), which is co-retweets within one minute.

**Figure 6 - Bot-like co-retweet network of China bioweapon conspiracy activity**



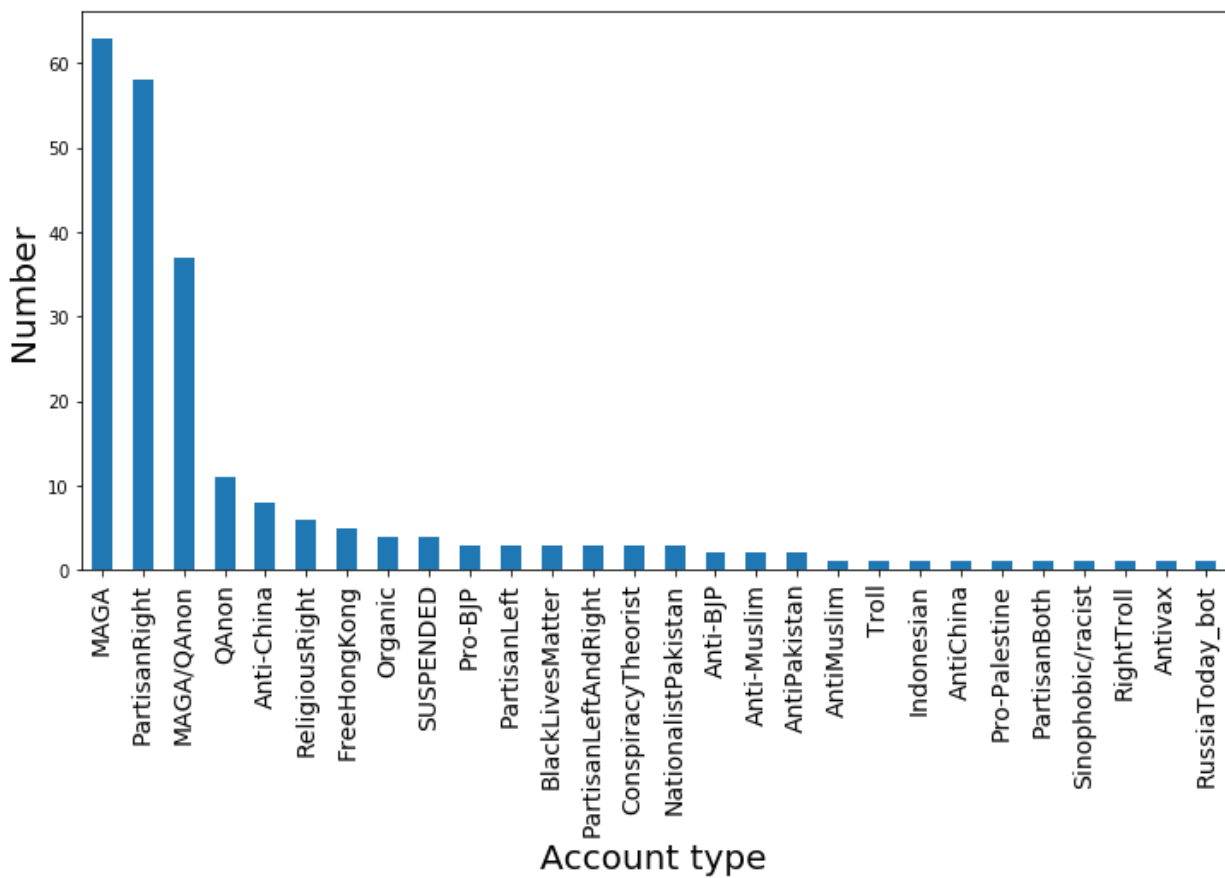
Note: The node colours represent what type of account it is (see Figure 8 for colour code index)

Figure 6 shows the bot-like co-retweet network for this conspiracy, where accounts are connected if they co-retweeted the same tweet within 1 second of each other. We use the



term 'bot' liberally to describe accounts that may range from completely automated (i.e. traditional scripted bots) through to hybrid accounts that utilise varying degrees of automation and/or scheduling software, thus producing extremely rapid (1 second or less) coordinated retweet activity. Each individual account (i.e. node) in this network was manually analysed to categorise them into types of accounts. This involved content analysis of the account profiles on Twitter using an inductive approach where the account categories or 'types' were derived from the data through an iterative process. This resulted in 28 categories of accounts, shown in Figure 7.

**Figure 7 - Categories and prevalence of accounts in the bot-like co-retweet network**



As Figure 8 shows, the majority of accounts self-identity through their profile descriptions and imagery as right wing and traditionalist conservative, with over two-thirds of the accounts being either typical MAGA troll accounts (27.39%), partisan right accounts that almost solely post biased conservative content and appear to be hyper-polarised (25.22%), or MAGA/QAnon accounts (16.09%) that explicitly support President Donald Trump in their profile descriptions and also identify with the QAnon far-right conspiracy group. The remaining one-third of accounts in the network represent much smaller proportions of group identities and account types. Notably, anti-Chinese accounts that mainly post criticism of the CCP and often racist and hate speech content constitute about 3.5% of the

accounts, with a further 2.6% of accounts aligning with religious (mostly Christian) right wing identities.

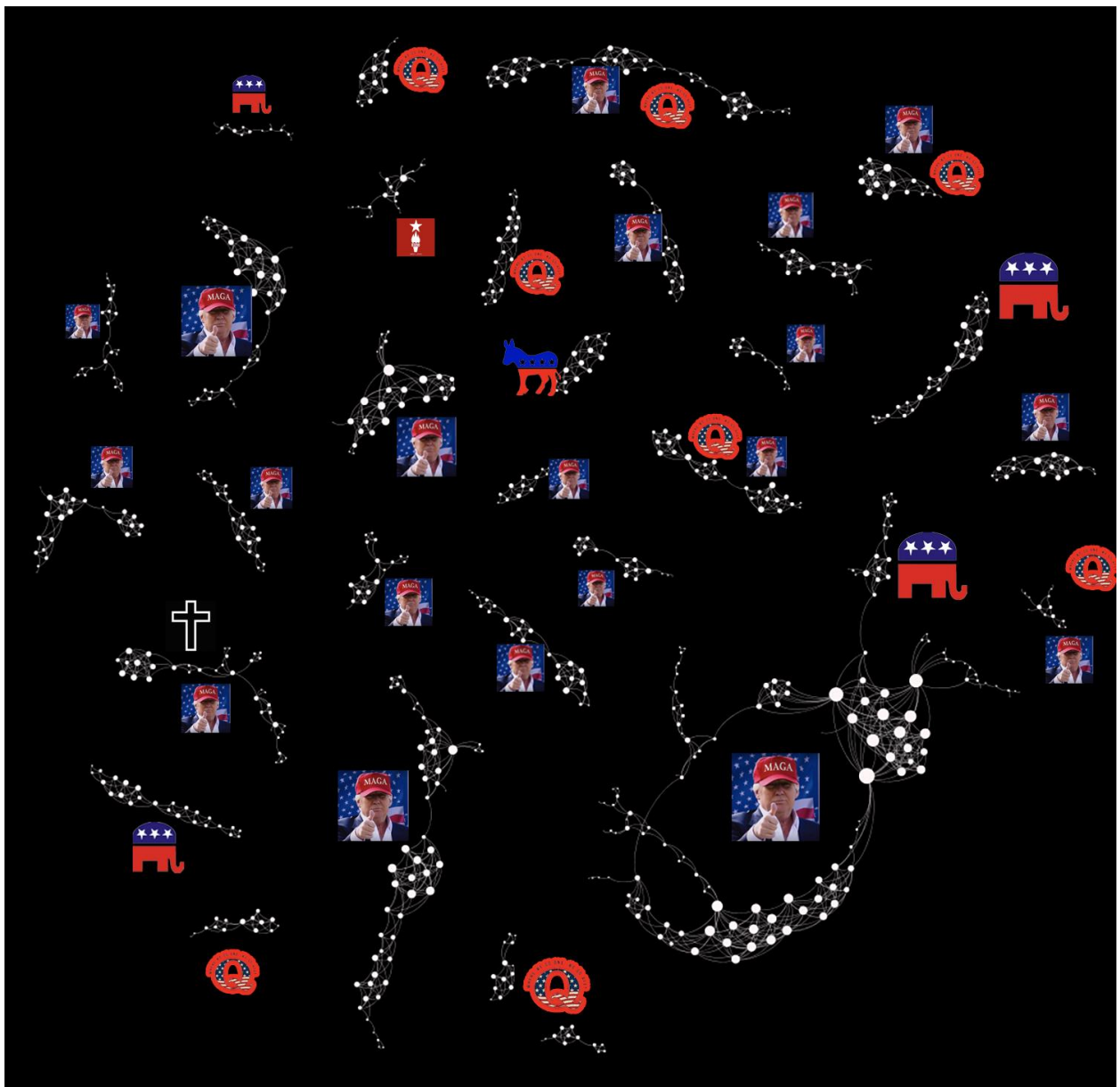
Figure 8 - Top six account types and colour codes in the bot-like co-retweet network



Next, we constructed and analysed the standard *co-retweet network* (Keller et al. 2020), where nodes are connected if they both retweeted the same tweet within one minute. As discussed previously, this reveals patterns of coordination in the network because although it is normal for two accounts to both retweet the same tweet, it is far less likely that they will do so repeatedly and within a short space of time (i.e. one minute in this case). Unlike the bot-like co-retweet network discussed above, this co-retweet network consists of both human-controlled (e.g. trolls, sockpuppets) and automated accounts. The co-retweet network contained 2,903 accounts and 4,125 links or co-retweets between them. To focus on the most important and active accounts, we ran a community detection algorithm (Blondel et al. 2008) to identify the clusters within the network. We then filtered the network to show only the top 30 clusters by size, shown in Figure 9.

In addition to the co-retweet network analysis (Figure 9), we manually analysed each cluster to determine what kinds of accounts were participating and whether there were any clear group identities for each cluster. The results show that nearly all (28 out of 30) clusters were a mixture of MAGA/Pro-Trump (denoted by the Trump icon), QAnon (denoted by Q symbol), and/or Republican partisan (denoted by elephant logo). The two exceptions were a cluster of pro-Democrat and broadly left-wing accounts (denoted by the donkey logo) and a cluster of 'Baloch National Movement' accounts that heavily pushed anti-Pakistan narratives and content.

Figure 9 - Co-retweet network of the China bioweapon conspiracy amplification



Note: accounts are connected if they both retweeted the same tweet within a one-minute window; icons represent the group identity of each cluster.

The 4,125 co-retweets shown in Figure 9 were retweets of 882 original tweets about the China bioweapon conspiracy. This is about four times the number of original tweets about this conspiracy that the authors observed in earlier research on COVID conspiracies and disinformation that drew on data collected in late January (Graham and Bruns, 2020).

Of the original 882 tweets, only 814 could be accessed when we recollected their engagement metrics from the Twitter API, with the others likely deleted by the account or removed by Twitter. For the 814 available tweets we find that:

- They were retweeted 18,498 times (i.e. regular retweets, not co-retweets).
- They were liked 31,783 times.
- The average number of retweets was 22.72.
- The average number of likes was 39.04.
- The median number of retweets was 2.
- The median number of likes was 3.

From this engagement data, an estimate of the number of actual views by Twitter users, or 'impressions' can be made. Social media trade press routinely uses an engagement rate of around 1.0% (The Online Advertising Guide, 2020), meaning that for every 'engagement' via retweet or like around 100 impressions are made. (See also explanation in Graham et al., 2020).

By this methodology, the estimated total number of impressions is:  $(18,498 + 31,783) / 0.01 = 5,028,100$ .

In Graham et al. (2020), the estimated total number of impressions from co-retweets of this conspiracy theory was 3,583,200. This was measured over a seven-day period, suggesting a sustained level of coordinated amplification of the China bioweapon conspiracy through the early stages of the coronavirus pandemic. Interestingly, the January co-retweet activity was based on just 213 original tweets, compared to the 882 seen in March. Suggesting more activity in terms of co-retweeting of content and slightly less engagement from that activity.

# Discussion and Recommendations

## THE CHALLENGE OF COORDINATED INAUTHENTIC BEHAVIOUR

The analysis we have presented here, as well as further research conducted by our colleagues in the Digital Media Research Centre at Queensland University of Technology, clearly documents the presence of coordinated, inauthentic behaviour in discussions about COVID-19 on Twitter. Such activities are often designed to promote the spread of conspiracy theories and other mis- and disinformation about coronavirus and its origins, effects, and remedies, and tap into the pre-existing fears and concerns of a social media population forced to sift through contradictory statements even from state and medical authorities.

The emergence and spread of the conspiracy theories we have observed here parallels the outbreak of the virus itself, resulting in the World Health Organisation's description of that spread as an 'infodemic' (United Nations, 2020), but several of these conspiracy theories also connect with well-established fringe communities that variously oppose the roll-out of 5G mobile telephony networks, the vaccine research efforts funded by the Bill and Melinda Gates Foundation, or the rise of China as a global power. In these communities, some of which place themselves on the (far) right of U.S. politics in their account profile self-descriptions, new conspiracy theories that connect COVID-19 with such existing antipathies find ready support and amplification.

Whether the coordinated inauthentic behaviours we have observed here are orchestrated by the hard core of participants in these groups themselves, or are designed by external operators to target and exploit the worldviews of such groups, the net effect is often the same: the themes and topics promoted by coordinated inauthentic activity are taken up by the wider fringe community, and thereby gains amplification and authenticity: the mis- and disinformation contained in the initial messages is no longer distributed solely by bots and other accounts that may be identified as acting in coordinated and inauthentic ways, but also and to a potentially greater extent by ordinary, authentic human users.

This is a critical step in broadening the dissemination of conspiracy theories and related mis- and disinformation, but not necessarily the end goal for such activities: to attract a pre-existing group of conspiracy theorists on the fringes of the social network to a new conspiracy has little effect on overall public opinion. If the aim is to sow confusion and discord amongst the general public, then it is necessary to further increase the reach of such mis- and disinformation. This phase change takes place once more prominent mainstream media and social media actors engage with the conspiracy theory, even critically: such engagement (by celebrities, journalists, politicians, media outlets, state authorities, and

others with large followings) substantially amplifies the visibility of the conspiracy theory. It may also prompt official denials and corrections from political leaders and domain experts, and perversely these responses can in turn be exploited by the conspiracy theorists to claim that - from their perspective - there is now indeed evidence that 'the authorities' are covering up 'the real truth'. In Australia, for example, the effects of this vicious circle are now being observed in the sharp rise in concerns about future 5G technology rollouts (Taylor, 2020), at least in part as a result of the circulation of the conspiracy theories about links between COVID-19 and 5G that we have documented in this report.

This does not mean that authorities should never respond to conspiracy theories, of course, but it does point to the difficulties in addressing such mis- and disinformation without playing into the hands of those who spread it. In the following, we make a number of recommendations for a broad-based response to this challenge.

## RECOMMENDATIONS FOR ADDRESSING THE CHALLENGE

### Technological Measures

Coordinated inauthentic behaviour is in part a technical challenge; it can therefore also be addressed, at least in part, by technological means. Our analysis here has demonstrated that such behaviour can often be detected by technical means, and most mainstream social media platforms are now employing such detection tools to a greater or lesser extent.

However, our ability to observe the continued presence of bots and other inauthentic accounts on Twitter also shows that detection and mitigation measures remain limited (and the same is true for other leading social media platforms, too); while platforms do suspend and remove many inauthentic accounts and their content, they are engaged in an escalating arms race with the operators of such accounts as they continually adjust their activities to evade detection.

In addition to the in-house development efforts conducted by the platforms themselves, independent critical investigation by scholarly social media researchers is crucial in this context, both to develop new and innovative detection approaches and to track and evaluate the activities of the platform operators themselves; such research can also feed into new regulatory initiatives that ensure that platform provider interventions are effective as well as equitable in their operation.

We note again in this context that mistakes, overreach, and iniquities in the enactment of content and account takedowns by platform operators or state authorities are likely to backfire as they provide conspiracy theorists with a claim that there is indeed a conspiracy

to silence them. That said, however, the judicious and well-founded takedown or suppression of coordinated inauthentic behaviour, especially in the early stages of dissemination, can be critical in preventing it from gaining significant reach.

## Digital Literacy

As we have highlighted throughout this report, conspiracy theories and other mis- and disinformation break out from their circulation in fringe communities usually only once they are picked up and recirculated by ordinary users. This is especially likely to occur during times of heightened anxiety and confusion, as part of the current coronavirus crisis or in other, similar acute events. At such times, even ordinary users may look beyond their usual sources and networks for additional material to fill the information vacuum, and may therefore be susceptible to problematic content. If these users encounter conspiracy theory content and share it on to their own networks, that content is thereby circulated to new networks, and implicitly endorsed by the sharers; it is no longer immediately identifiable as fringe content, but now carries the imprimatur of otherwise trusted and reliable social contacts. From this point onward, it may circulate more widely and more quickly, across more diverse networks.

Such on-sharing by ordinary users cannot be addressed through technological measures alone. Platforms such as Twitter and Facebook have now begun to display content warnings that are attached to content from dubious sites as it appears in users' newsfeeds, and these warnings are likely to suppress some of the on-sharing of such content by ordinary users. However, such warnings are usually attached only to posts that directly link to dubious sites, and not to posts that share screenshots from such sites or merely paraphrase the information contained in their articles; this means that unsourced or more indirectly sourced conspiracy theories and other mis- and disinformation are still able to circulate without intervention or correction.

Therefore, enhanced digital literacy training for ordinary social media users must be an important additional (if more indirect) component of any mitigation strategy for mis- and disinformation: ordinary users need to be more fully able to assess the veracity of the information they encounter through their social media feeds for themselves, and more fully aware of the potential consequences of their on-sharing of dubious information to their downstream social networks. A number of digital media literacy initiatives are already in train, including from the newly formed Australian Media Literacy Alliance (AMLA), and there is a significant need for further funding and institutional support for such initiatives at all levels, and for all age groups.

## Mainstream Media

Conspiracy theories and other mis- and disinformation originate in fringe communities, and are promoted by fringe media outlets associated with such communities. In the case of COVID-19 this includes, for instance, alternative medicine sites speculating about the origins of the virus or promoting unproven medical remedies, or hyperpartisan politics sites linking the virus to their extreme left- or right-wing agendas. Such media outlets largely speak to the already converted and do little to increase the spread of the conspiracy theories beyond these fringe communities.

Other, more mainstream media outlets play a more problematic role: for instance, we have observed tabloid news outlets and other unscrupulous news sites using conspiracy content as clickbait designed to attract audiences to their platforms. Such sites may frame the conspiracy theories as outlandish or laughable, but often present them without significant correction or fact-checking; as a result, such coverage puts substantial new audiences in contact with problematic content that they would not otherwise have encountered. Tabloid media can therefore represent an important pathway for conspiracy theories to enter more mainstream public debate.

Eventually this increased circulation of conspiracy theory content may also create the necessity for political leaders or public health experts to respond explicitly to them. Because they come from the central authorities during the crisis, such responses are then also covered by quality mainstream media, of course, but in doing so they also draw further public attention to the very conspiracy theories they seek to debunk, and citizens who are already predisposed to scepticism towards official sources may choose to take more note of the original mis- and disinformation than of the new statements that aim to correct it. Explicitly or implicitly, this is the endgame for conspiracy theorists: by forcing the authorities to respond to their claims, and the mainstream media to cover such responses, they have managed to substantially increase the circulation of their ideas throughout mainstream society - and they can point to such denials as evidence for their claims that there really is a conspiracy to hide 'the real truth'.

This places mainstream journalists and media outlets in an unenviable position: they cannot simply ignore the authorities' public responses to mis- and disinformation, but by covering them they indirectly also amplify the initial spread itself. In this context it may therefore be necessary for mainstream media to diverge from their usual ideals of impartial and disinterested coverage, to clearly take sides, and to forcefully contribute to the debunking of conspiracy theories.

## Scholarly Research

The work we have presented here draws substantially on a mixture of advanced computational methods and forensic qualitative analysis of large-scale, real-time social



media data; especially in the context of a still-unfolding, acute pandemic and infodemic, it is exceptionally time-critical and requires the rapid deployment of data gathering, processing, analysis, and visualisation tools and methods. Such work is hampered, however, by the increasingly constrained levels of access provided by the leading social media platforms to critical, independent, public-interest scholarly research projects. While as one of the leading digital media research centres of the southern hemisphere our own research centre continues to enjoy comparatively good access to large-scale social media data, such access is increasingly rare, and threatens the independent analysis and oversight of social media communication and its role in wider public debates.

While this needs to be balanced with the data and privacy rights of individual social media users and the commercial interests of the platforms, there is therefore a substantial and urgent need to provide more immediate access to large-scale social media datasets for social media researchers. Similarly, in light of the notable computational load of data gathering, storage, and processing there is a significant need for the further development of shared national infrastructure for scholarly work in this field.

At present, Australia is a recognised global leader in digital and social media research, but this position is threatened by the uneven distribution of access and resources across the Australian tertiary education sector; national and state infrastructures such as the National eResearch Collaboration Tools and Resources project (NeCTAR) or Queensland's cloud computing and storage infrastructure QRISCloud are enormously valuable, but provide only generic support; projects such as the QUT-led Tracking Infrastructure for Social Media in Australia (TrISMA), which developed a comprehensive Australian Twitter Collection, received only short-term funding and therefore cannot provide access to the broader national community of researchers.

Yet as the COVID-19 crisis and associated infodemic demonstrate, social media are now a critical component of the national and international public debate, for better or for worse. In order to effectively monitor and analyse the public conversation within these spaces and thereby safeguard the Australian public from mis- and disinformation, greater national support and coordination in the provision of state-of-the-art social media research infrastructure is needed.

# Acknowledgments

This research is funded in part by the Centre for Responsible Technology at The Australia Institute, the Australian Research Council projects DP200101317 Evaluating the Challenge of 'Fake News' and Other Malinformation and FT130100703 Understanding Intermedia Information Flows in the Australian Online Public Sphere, and by the Australian Department of Defence (Defence Science and Technology). Computational resources and services used in this work were provided by the eResearch Office, Division of Research and Innovation, Queensland University of Technology, Brisbane, Australia.

## REFERENCES

- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
- Dwyer, C. (2020, 11 Apr.). Turkey Imposes 48-Hour Coronavirus Curfew In Major Cities With Little Warning. *NPR*. <https://www.npr.org/sections/coronavirus-live-updates/2020/04/11/832424919/turkey-imposes-48-hour-coronavirus-curfew-in-major-cities-with-little-warning>
- Giglietto, F., Righetti, N., & Marino, G. (2019). Understanding Coordinated and Inauthentic Link Sharing Behavior on Facebook in the Run-up to 2018 General Election and 2019 European Election in Italy. *SocArXiv Papers*. <https://doi.org/10.31235/osf.io/3jteh>
- Giglietto, F., Righetti, N., Rossi, L., & Marino, G. (2020). It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. *Information, Communication and Society*, 1–25. <https://doi.org/10.1080/1369118X.2020.1739732>
- Gleicher, N. (2018, December 6). Coordinated Inauthentic Behavior Explained. *Facebook Newsroom*. <https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/>
- Graham, T., Keller, T., Angus, D., & Bruns, A. (2020). The bioweapon conspiracy: A coordinated right-wing disinformation push during the COVID-19 outbreak. Paper under review.
- Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political Astroturfing on Twitter: How to coordinate a disinformation Campaign. *Political Communication*, 37(2), 256-280.

Krämer, B. (2017). Populist Online Practices: The Function of the Internet in Right-Wing Populism. *Information, Communication & Society*, 20(9), 1293–1309.  
<https://doi.org/10.1080/1369118X.2017.1328520>

Online Advertising Guide. (2020). Twitter Engagement Rate Calculator.  
<https://theonlineadvertisingguide.com/ad-calculators/twitter-engagement-rate-calculator/> (last accessed 15 April 2020)

Taylor, J. (2020, 13 May). Australian public's confidence in 5G 'shaken' by misinformation campaign. *The Guardian*.  
<https://www.theguardian.com/technology/2020/may/13/australian-publics-confidence-in-5g-shaken-by-misinformation-campaign>

United Nations Department of Global Communications. (2020). UN tackles 'infodemic' of misinformation and cybercrime in COVID-19 crisis. <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19>