

Submission to the Senate Select Committee on Foreign Interference through Social Media

On 5 December 2019, the Senate resolved to establish a Select Committee on Foreign Interference through Social Media to inquire into and report on the risk posed to Australia's democracy by foreign interference through social media.

Background

The News and Media Research Centre (N&MRC) at the University of Canberra (<http://www.canberra.edu.au/nmrc>) investigates the evolution of news, media, content and communication. The N&MRC is a national leader in the provision of expert commentary and analysis of social media manipulation in Australian politics. Formed in 2019 as part of the N&MRC, the Critical Conversations Lab investigates the way issues of social and political concern emerge through media and digital networks to enable public participation and influence political agendas.

The Virtual Observatory for the Study of Online Networks (VOSON) Lab at The Australian National University (<http://vosonlab.net/>) is a global leader in computational social science and big data analytics. Since 2005 the Lab has advanced the Social Science of the Internet through an innovative program of research, research tool development, teaching & research training.

The submission authors are happy to provide further information to the Inquiry if desired.

Submission authors:

Professor Robert Ackland, Leader VOSONLab, Research School of Social Sciences, The Australian National University

Associate Professor Michael Jensen, News & Media Research Centre & Institute of Governance and Public Administration, University of Canberra,

Associate Professor Mathieu O'Neil, Leader Critical Conversations Lab, News & Media Research Centre, University of Canberra,

Introduction: Foreign influence operations in the digital age

The threat of foreign influence is not new. Strategic thinking has long emphasised the importance of using informational means and other levers of coercion short of war in order to impose one's will on a target country. During a recent address before the Lowy Institute, the outgoing director of the Australian Security Intelligence Organisation, Duncan Lewis, declared espionage and foreign influence an "existential threat" to Australia and "far and away the most serious issue going forward" for Australian security (Lewis, 2019). Lewis' comments about the potential harms of foreign influence combined with espionage signal that foreign influence operations are often part of broader strategic objectives, utilising information obtained through espionage to inform influence activities.

The threat of foreign influence is uniquely pressing at the present for three reasons.

1. Digital networks play a central role in political communication

In contrast to conventional threats to national security, against which vast distances across the oceans have protected the country, online foreign influence negates the security provided by geography. Attackers can carry out foreign influence operations from outside the country and hide their origins and activity. Further, as society becomes more diverse and demands for government responsiveness to citizens and groups forming the public sphere increase, systematic distortions in public conversations can reverberate across other domains of political decision making (Luhmann 1982; Swanson and Mancini 1996). Public decisions must often be presented and defended within these spaces, which then inform other aspects of coverage across the media ecosystem. Hence, the centrality of digital networks to domestic political communication reduces entry barriers and likelihood of discovery for foreign adversaries and increases the risks for Australia's democracy.

2. The speed of social media renders information attacks hard to counter

Digital networks facilitate cost-effective access to communities, reducing the resources and time required to execute a sustained influence operation. They also enable timely interventions into political discussions which can be decisive in shaping outcomes (Kreiss 2014). Russia's covert Facebook advertising operation during the 2016 US election showed that operatives promoted ads coinciding with events on the same day, and that the median duration of these ads was just one day (Jensen 2019b). Influence operations capitalise on the fast temporalities of digital spaces which makes it hard to interrupt an operation in progress by suspending accounts: by the time they are reported, they likely have produced their intended effects. That is not to say suspending accounts is not worthwhile, particularly for accounts which have become highly influential, or as a means of slowing an operation's capacities. In summary digital networks enable foreign influence operations to scale-up much quicker than in the analogue age of communication. The creation of websites and social media posts, which is sometimes automated (Howard, Woolley, and Calo 2018), can participate in, and speed up, cascades of memes and URLs which then reach vast audiences (Starbird and Palen 2012; Zannettou et al. 2019).

3. Digital influence operations have low implementation costs

Finally, in contrast to other sophisticated weapons systems, the technological thresholds for influence campaigns are quite low. Unlike the technical hurdles involved in missile defence or nuclear weapons, influence operations can be carried out using a computer screen and an internet connection. Although information warfare tactics are often classified, general theories of how to carry off such operations can be found in marketing textbooks which abound in research on how to manipulate target audiences. Fabricating images, particularly in an era of “deep fakes” (Edwards and Livingston 2018) is now easier and cheaper than in an era where documents had to be forged, then physically transported to a target site.

Social media and other online communications are normally only one part of an influence campaign. Influence campaigns tend to be sustained, with an eye to impacting the course of a country’s politics beyond the next election cycle. Information operations supports other activities (Armistead 2004) which often include financing (which may be covert and illicit) and direct contacts with candidates and other party officials. It is therefore important that political parties, even at the local levels, receive training on how to handle approaches by persons acting on behalf of a foreign principal. Beyond political parties themselves, interest groups and other activist groups may be targeted through both online and offline outreach.

Internet Research Agency troll activity in the Australian political Twittersphere, 2015-2016

As part of Twitter’s investigation into use of social media by state-backed influence operations during the 2016 US presidential election, Twitter publicly released datasets containing tweets (and media linked to in tweets) identified with organisations such as the Russian Internet Research Agency (IRA).

While researchers have analysed IRA-authored tweets that relate to Australian politics (Jensen 2019a; Jensen and Sear 2018) we present here a new approach involving computational methods (network and text analysis) and data visualisation, that allows us to identify three specific styles of IRA troll account activity in the Australian political Twittersphere. The results of these analyses demonstrate that IRA trolling operation did not focus on persuasion and efforts to directly shift political views, nor did they generally seek to change the shape of online discussion. Rather, they tend to focus on a strategy of ‘resonance’ where they seek to embed themselves in a community and from there can work to activate at least certain sections of it for strategic aims (Clark 2017).

Our approach involved the following steps (the data and analysis tool come from the Virtual Observatory for the Study of Online Networks, VOSON):

1. The starting point for this analysis is defining or demarcating the Australian political Twittersphere: we used a large-scale Twitter dataset collected over a year (September 2015 to October 2016) that includes all the tweets authored by Australian federal politicians, and those tweets where the politicians were retweeted, replied to or mentioned.
2. We then constructed a subset of those users who tweeted at least once per month over this period (“active political tweeters”), and then identified a set of hashtags pertaining to

clearly-identifiable topics (events, issues, places) that were included in the tweets authored by these users. We refer to these as “issue hashtags”.

3. For each month, we produced a minimum spanning tree (MST) semantic network visualisation of the issue hashtags that allows us to see how the hashtags connect to each other semantically and cluster into key areas of public and policy interest, such as refugees and asylum seekers, the economy, health etc. Hashtags located close to one another on a branch of the tree map tend to be semantically related to one another, in that they were frequently co-located in tweets authored by the active political tweeters.
4. We then identified IRA-authored tweets that were: (1) created between September 2015 and October 2016; (2) contained the word “australia” or at least one of a set of hashtags that are clearly related to Australian politics (e.g. #auspol, #ausvotes, #qt, #qanda, #insiders) and (3) contained at least one of the issue hashtags identified above.
5. The final step was to map the troll data (what hashtags were used by troll accounts, and how these hashtags were co-located in their tweets) onto the MST semantic networks.

Figure 1 shows an MST semantic network (with troll account activity overlay) for one of the months, while Figure 2 shows a zoomed-in view, showing the issue hashtags that troll accounts included in their tweets, and how these hashtags were co-located in tweets.

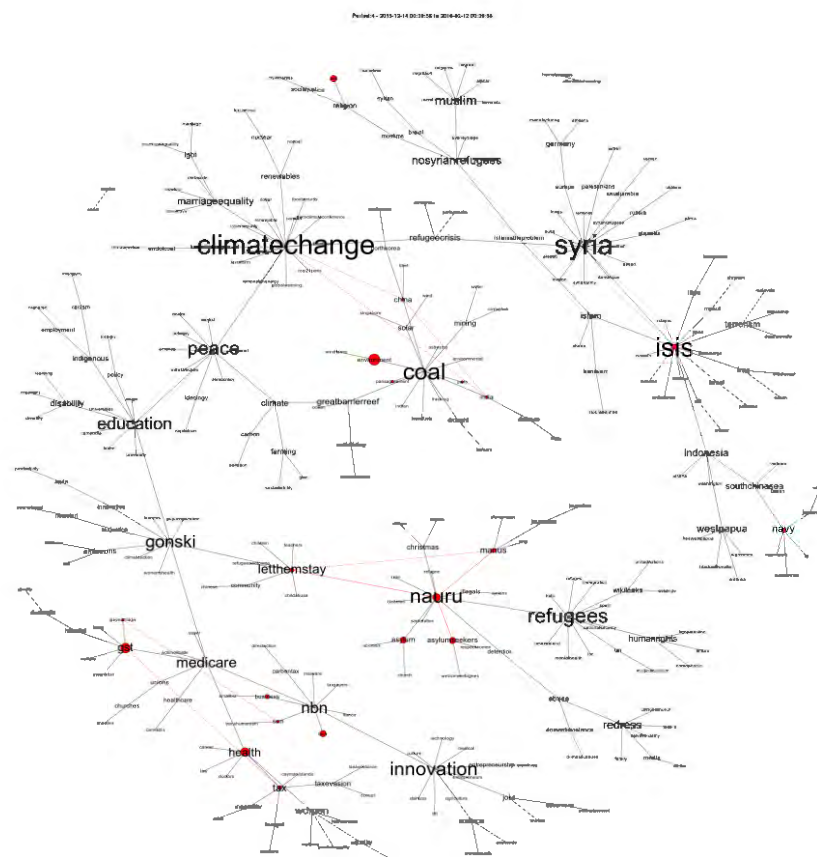


Figure 1: MST semantic network for active Australian political tweeters, with troll activity overlay.

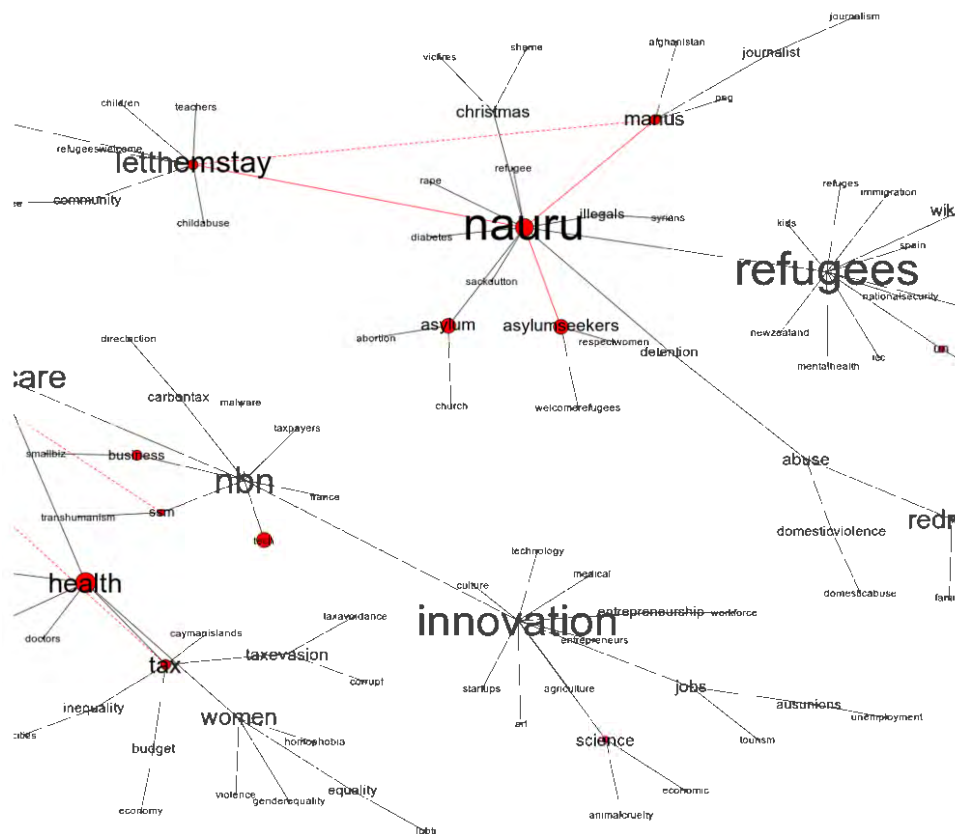


Figure 2: Detail from MST semantic network. Red nodes are hashtags tweeted by troll accounts (size proportional to number of tweets). Unbroken red line connects pairs of hashtags that both troll and non-troll accounts included in tweets. Broken red line connects pairs of hashtags that were included in tweets authored by troll accounts, but non-trolls did not include in tweets.

As with previous authors (Jensen 2019a; Jensen and Sear 2018), we found that IRA troll account activity in the Australian political twittersphere was not extensive. We found that there were 70 unique IRA troll accounts who authored a total of 535 tweets (or retweets) that were clearly focused on Australia and also featured one or more of our target issue hashtags. However, our MST semantic network visualisation approach allowed us to easily and quickly identify three examples of distinctive troll behaviour that match particular influence operation *modus operandi*.

Case 1: Audience building for future influence payoff

The first example of troll behaviour involved a series of tweets that were seemingly innocuous: they featured the hashtag #Periscope (a social networking application) and also hashtags for a number of countries (including Australia). An excerpt of one of the hashtags is: "... on #Periscope: 😊❤️ #sweden #australia #japan #africa #netherlands #dubai #belgium #china #korea ...". This behaviour clearly stood out in the MST semantic networks as it resulted in pairings of hashtags that were only being made by the troll accounts (the non-troll accounts were not including these pairs of hashtags in tweets) as seen in Figure 3.

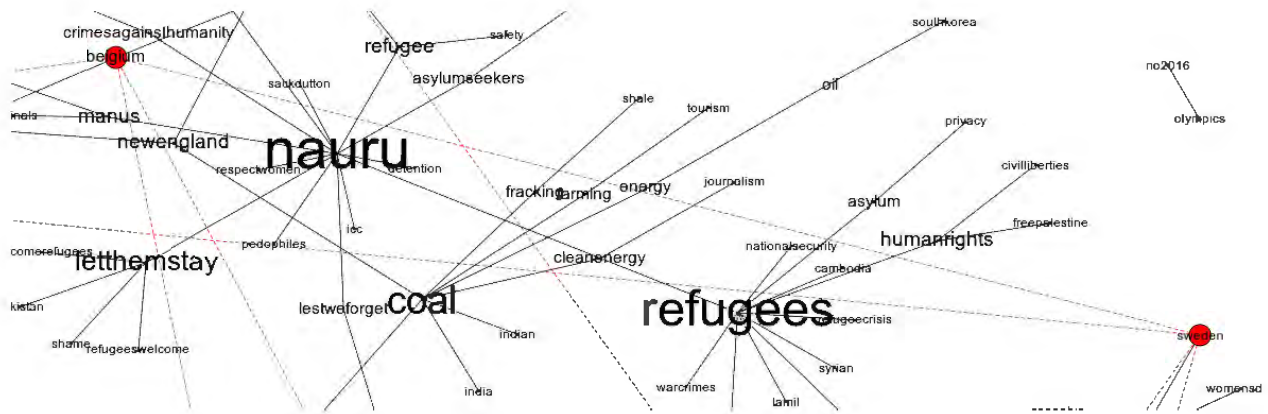


Figure 3: Troll activity is evidenced by the fact that hashtags for Belgium and Sweden are distant from one another in the MST semantic network (indicating that for non-troll accounts these hashtags are semantically distant) and there is an unbroken red line connecting the hashtags (indicating that only troll accounts were pairing these hashtags in tweets).

While these troll tweets are seemingly benign, and in fact not even relevant to Australian politics, this troll behaviour is consistent with what we refer to as audience building and the process of embedding oneself within a target population: troll accounts attempt to build an audience and credibility on Twitter (by tweeting potentially appealing and useful content) with the aim of later directing politically-related or socially-disruptive content to this established audience.

The process involves a common tactic of spycraft, scaled to operate on a wider scale: connecting with people on grounds that flatter the target population's interests before shifting to terms that the influence agents will seek to produce effects on in the future. Recognizing that most people form social attachments outside of the context of political views, influence agents try to forge relationships with persons on nonpolitical grounds before shifting to political topics (Watts 2018).

Case 2: Issue payload injection - the example of refugees

The second example of troll behaviour related to engagement with the issue of refugees and asylum seekers. In an earlier period, both troll and non-troll accounts were engaging with this topic in a similar manner. In Figure 2 above (which pertains to the period December 2015 to January 2016), trolls and non-troll accounts used the same pairings of hashtags: #manus and #nauru, #nauru and #asylumseekers, #nauru and #letthemstay. Troll accounts also paired #nauru and #letthemstay, while non-trolls did not make this pairing (this is indicated by the fact that there is a red broken line joining the hashtags), but the hashtags are not semantically distant from one another and this again supports our contention that troll and non-troll engagement with the topic of refugees in this period was similar.

However in a later period (July to September 2016) the troll accounts made a connection (via tweets) between #manus and #isis; this connection was not being made by non-trolls and further, the non-troll tweeting activity was such that these two hashtags were semantically distant in this period (Figure 4). We point to this as evidence of “issue payload injection”: troll accounts attempted to influence the direction of discourse around the issue of refugees and asylum seekers by inferring that refugees being housed on Manus Island are potentially connected to Islamic State (and thus a security threat to Australia). This also serves to divide Australian society internally along “us vs them” lines.



Figure 4: Troll accounts are connecting (via tweets) #nauru and #isis, a connection that is not being made by non-troll accounts.

Case 3: Spreading bad vibes - a year in the life of troll account 2951506251

The final type of troll behaviour that we identified using the MST semantic network approach was that of highly active troll tweeters who regularly authored tweets about news events.

Account characteristics: number, frequency, and hashtags

Account 2951506251 was by far the most prolific troll in our dataset, with the number of tweets produced by this account (260) equal to that of all other troll accounts combined. This account produced more than five times as many tweets as the second most prolific account. All of these were original tweets, not retweets, with each tweet being a news item and including a single hashtag (#environment, #science or #tech).

The number of 2951506251's tweets per month steadily increased until it reached its peak in June 2016, just before the July 2 federal election. This number then halved, remained low for two months and slowly built back up, though never again reaching the levels immediately preceding the election.

The account did not adorn its tweets with the type of hashtags favoured by other IRA troll accounts active in the Australian Twittersphere, such as US-oriented hashtags (e.g. #blacklivesmatter, #blacktwitter, #guncontrol, #hillaryclinton, and #usa). These hashtags were part of an operation to

reach potential voters on the political left in the US, moving them to either vote for the Green Party candidate, Jill Stein, rather than Hillary Clinton, or to not vote at all. The latter involved efforts specifically directed against the African American community in an effort to convince them not to turn out to vote for Clinton. While the IRA trolls were far from the only factor - or even the most important factor - the 2016 election saw for the first time this century, a decline in the African American vote and, overall, there were four million voters from 2012 who failed to turn out in 2016 (Jamieson 2018). Other troll accounts in the Australian Twittersphere favoured divisive posts such as “Anti-#Islam rally is going on in Australia. People protest against islamisation of the country #ReclaimAustralia”.

Account content: Australia is a mess

In contrast, 2951506251 used hashtags which seemingly qualified post content without overt political or emotional value judgments (as stated before: #health, #tech, #environment). Under this “neutral” cover, 2951506251 provided links to news stories or headlines. We reproduce below a few representative examples.

A selection of 2951506251’s #tech posts:

- News Corp's Australian Netflix challenger shuts up shop
- Australia government cyber attack came from foreign intelligence service: report
- IBM apologises for Australian e-census bungle, setting off blame game

A selection of 2951506251’s #health posts:

- Cancer overtakes heart disease as Australia's biggest killer
- Australian authorities spray Queensland hotel over Zika scare

A selection of 2951506251’s #environment posts:

- Australia's bushfires leave trail of death and destruction
- Australia's wheat crop threatened as La Nina climate indicator rises: analysts
- Australia scientists alarmed at new Great Barrier Reef coral bleaching
- One killed, thousands without power as storms hit Australia
- Statewide blackout in Australia raises questions over renewable energy
- Sinkhole swallows car in South Australia
- Australian explorer looking at grounds for lawsuit over fracking ban

Aggregating these posts creates the impression that Australia is a dreary place, where mostly bad things happen, or things don’t work, and where people are perpetually arguing about something or another. In the midst of all these bad news a minority of items were positive (“Australia sees agriculture output boost as El Nino fades #environment”, “Solar powered car racers set off in Australian challenge #science”). Such items served to legitimate the account as providing a balanced view.

We must provide two caveats to the above analysis. First, we have not conducted a systematic classification of account 2951506251’s 260 posts as “positive” or “negative”. Traditional sentiment-analysis tools might be unsuited to pick up the subtly bleak tone we have identified, and it was beyond the scope of this report (and would require further resources) to develop a specific automated content analytical tool for this task. Second, news organisations do tend to favour dramatic events and headlines over non-dramatic events. However it is undeniable that the overall picture created by the majority of 2951506251’s posts, under the cover of “neutral” hashtags, consistently leant towards the highly negative side of the news spectrum.

Parallel accounts

Another relatively highly active troll account tweeted 44 times during our period of study (tweets that included one or more of our target issue hashtags) and all of these tweets were similarly news items, but this time focused on health. Yet another active troll account specialised in news items about business events. The structure of the tweets authored by these three separate Twitter users was so similar that we have come to conclude that they were possibly being authored by the same person (or perhaps even a bot). The fact that the tweets only ever contained a single hashtag meant that the activity appeared in the MST semantic networks as hashtags that were not being connected or paired (by the troll) with one another (Figure 5).

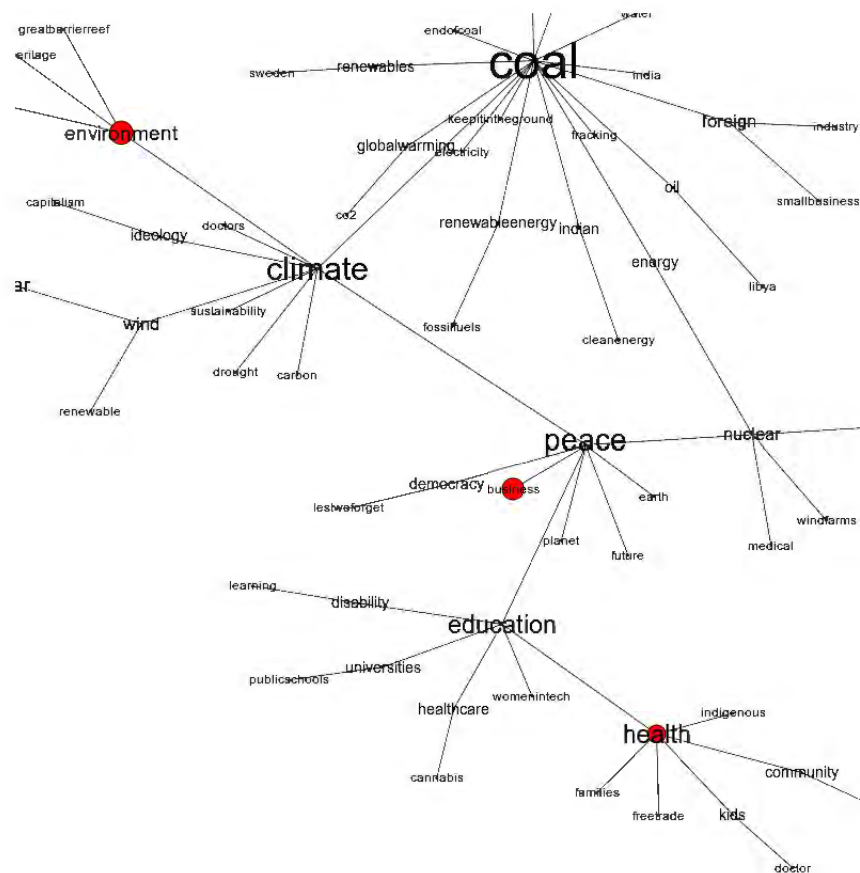


Figure 5: News reporting troll accounts authored tweets on social, economic and environmental news events. Their tweets typically only include a single hashtag and hence are not connected (via troll activity) in the MST semantic networks.

Conclusion

From our analysis of the dataset, it is apparent that Twitter “troll” accounts controlled by the Russian Internet Research Agency (IRA) pursued a multi-pronged engagement strategy in Australian social media. These tactics included developing relationships with users to enable future propaganda dissemination; injecting divisive content into existing debates; and attempting to colour online discussions of #Australia with negative content. The fact that it is difficult to evaluate whether these

interventions played a role in swaying Australians' opinions and beliefs, and to what extent, should not detract from the fact that these attempts occurred and are still occurring.

New research in this space includes developing stronger troll-identifying, troll-exposing, and troll-debunking tools. Central to this is a requirement for further development of software to enable the analysis of the behaviour of actors in online social spaces using computational approaches such as network and text analysis, and we point to the open source R packages developed by the VOSON Lab (e.g. Graham, Gertzel, Chan and Ackland 2019) as examples of such tools. While there has been a lot of research into the influence of social bots on Twitter during the 2016 US presidential election, most of this research conceptualised influence as contribution to information diffusion via Twitter retweet cascades (see, for example, Rizoiu, Graham, Zhang, Zhang, Ackland and Xie 2018). This submission has highlighted the usefulness of other computational approaches for conceptualising and measuring the potential societal impacts of foreign influence operations on social media.

Furthermore, the IRA is not the only active foreign influence agent active in the Australian social media space: our research also shows the distorting effects of WeChat 'official' or 'public' accounts targeting Chinese readers in Australia with news. Although some have suggested that WeChat news outlets are part of Australia's diverse and multicultural news environment, we have shown that these blogs replicate the Chinese Communist Party's censorship and opinion guidance practices (Jensen, Chen, and Sear 2018). This occurs even when operating in Australia via news outlets located here and targeting persons living in Australia. The consequence is that such outlets print not the news, but the news that serves the specific propaganda purposes of a foreign state.

The example of WeChat in Australia is only one instance where influence operates outside the boundaries of familiar Western social media platforms. It is also important to understand that social media operations do not happen in isolation of other activities carried out by foreign states. Social media operations are often guided by foreign intelligence agencies, explaining Duncan Lewis' framing of foreign influence and espionage as a common threat. More generally, social media operations are directed at amplifying other levers of state power - whether they be diplomatic, military, or economic. In a sense, this approach is nothing new. In 1948 George Kennan penned a 'political warfare' doctrine for the US which seized on the use of all measures short of war to advance America's strategic objectives. In an era of extended nuclear deterrence, when the risks of conventional military action are prohibitive as they might risk initiating an escalatory ladder, influence operations on social media may be one of the alternative preferred theatres of warfare.

References

- Armistead, Leigh. 2004. *Information Operations: Warfare and the Hard Reality of Soft Power*. Potomac Books, Inc.
- Clark, Howard Gambrell. 2017. *Information Warfare: The Lost Tradecraft*. Washington, DC: Narrative Strategies.
- Edwards, Scott, and Steven Livingston. 2018. "Fake News Is about to Get a Lot Worse. That Will Make It Easier to Violate Human Rights — and Get Away with It. - The Washington Post. *The Monkey Cage*. https://www.washingtonpost.com/news/monkey-cage/wp/2018/04/03/fake-news-is-about-to-get-a-lot-worse-that-will-make-it-easier-to-violate-human-rights-and-get-away-with-it/?utm_term=.9f6bc8abe30b (August 14, 2018).
- Graham, T., Gertzel, B., Chan, C-h & R. Ackland (2019). vosonSML: Collecting Social Media Data and Generating Networks for Analysis. <https://cran.r-project.org/package=vosonSML>.

- Howard, Philip N., Samuel Woolley, and Ryan Calo. 2018. "Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration." *Journal of Information Technology & Politics* 15(2): 81–93.
- Jamieson, Kathleen Hall. 2018. *Cyberwar: How Russian Hackers and Trolls Helped Elect a President What We Don't, Can't, and Do Know*. Oxford University Press.
- Jensen, Michael J. 2019a. "We've Been Hacked – so Will the Data Be Weaponised to Influence Election 2019? Here's What to Look for." *The Conversation*. <http://theconversation.com/weve-been-hacked-so-will-the-data-be-weaponised-to-influence-election-2019-heres-what-to-look-for-112130> (March 7, 2019).
- Jensen, Michael. 2019b. "'Fake News' Is Already Spreading Online in the Election Campaign – It's up to Us to Stop It." *The Conversation*. <http://theconversation.com/fake-news-is-already-spreading-online-in-the-election-campaign-its-up-to-us-to-stop-it-115455> (June 17, 2019).
- Jensen, Michael, and Tom Sear. 2018. "Russian Trolls Targeted Australian Voters on Twitter via #auspol and #MH17." *The Conversation*. <http://theconversation.com/russian-trolls-targeted-australian-voters-on-twitter-via-auspol-and-mh17-101386> (May 1, 2019).
- Jensen, Michael, Titus C. Chen, and Tom Sear. 2018. "How Digital Media Blur the Border between Australia and China." *The Conversation*. <http://theconversation.com/how-digital-media-blur-the-border-between-australia-and-china-101735> (May 1, 2019).
- Kreiss, Daniel. 2014. "Seizing the Moment: The Presidential Campaigns' Use of Twitter during the 2012 Electoral Cycle." *New Media & Society*: 1–18.
- Lewis, Duncan. 2019. 536 *The Lowy Institute: Live Events*. Lowy Institute Melbourne Australia. <https://soundcloud.com/lowyinstitute/an-address-by-asio-director-general-duncan-lewis> (September 12, 2019).
- Luhmann, Niklas. 1982. *The Differentiation of Society*. New York: Columbia University Press New York.
- Rizoiu, M.-A., Graham, T., Zhang, R., Zhang, Y., Ackland, R. and L. Xie (2018), #DebateNight: The Role and Influence of Socialbots on Twitter During the 1st 2016 U.S. Presidential Debate. In: International AAAI Conference on Web and Social Media (ICWSM '18).
- Starbird, Kate, and Leysia Palen. 2012. "(How) Will the Revolution Be Retweeted?: Information Diffusion and the 2011 Egyptian Uprising." In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12*, New York, NY, USA: ACM, 7–16. <http://doi.acm.org/10.1145/2145204.2145212> (August 12, 2015).
- Swanson, David L., and Paolo Mancini. 1996. *Politics, Media, and Modern Democracy: An International Study of Innovations in Electoral Campaigning and Their Consequences*. Greenwood Publishing Group.
- Watts, Clint. 2018. *Messing with the Enemy: Surviving in a Social Media World of Hackers, Terrorists, Russians, and Fake News*. New York: Harper.
- Zannettou, Savvas et al. 2019. "Who Let The Trolls Out?: Towards Understanding State-Sponsored Trolls." In *Proceedings of the 10th ACM Conference on Web Science, WebSci '19*, New York, NY, USA: ACM, 353–362. <http://doi.acm.org/10.1145/3292522.3326016> (August 28, 2019).