Social influence campaign in the cyber information environment

By Kathleen M. Carley

Beliefs, opinions, and attitudes are shaped as people engage with others. Today, that is often through social media. A key feature of social media is that all information is digital, and all information is shared through devices by agents who need not be human. This has created a Wild West for information where it is as easy to create and share false information as true, where nonhuman agents are often better equipped than humans to share information, where technology arms races continually alter the landscape as to what is doable, and where human understanding of the phenomena and policy are lagging.

Lone wolves and large propaganda machines both engage with the public to disrupt civil discourse, sow discord, and spread disinformation. Bots, cyborgs, trolls, sock puppets, deepfakes, and memes are just a few of the technologies used in social engineering aimed at undermining civil society and supporting adversarial or business agendas. How can social discourse without undue influence persist in such an environment? What types of tools, theories, and policies are needed to support such open discourse?

In response to these cyber-mediated threats to democracy, a new scientific field has emerged: social cybersecurity. As noted by the National Academies of Science in 2019: Social cybersecurity is an applied computational social science with two objectives:

- “Characterize, understand, and forecast cyber-mediated changes in human behavior and in social, cultural, and political outcomes; and
- Build a social cyber infrastructure that will allow the essential character of a society to persist in a cyber-mediated information environment that is characterized by changing conditions, actual or imminent social cyberthreats, and cyber-mediated threats.”

As a computational social science, social cybersecurity is both science and engineering, with a focus on applied research that is shaped by, and is needed in order to shape, policy. Though tremendous advances have been made in this area, there are still key gaps. These gaps can be illustrated by looking at some findings from our research during the pandemic and the 2020 elections.

Most messages shared on Twitter were not disinformation, or even about the “disinformation issues.” However, stance detectors were able to identify sets of messages that were consistent with disinformation story lines. We found that in March the pandemic was being framed as a political issue by the White House and as a health and safety issue by the medical community. This framing led to different stances vis-à-vis the pandemic, which became aligned with political choices. Messages that are consistent with a disinformation storyline may not contain inaccurate facts. For example, many tweets about voter fraud and about vaccination implied false facts but did not state them. While this is not a discussion based on facts, it is not clear whether there should be policies to stop it. It is clear that such discussion can and has riled up groups, and led to protests and acts of violence.

Disinformation played a critical role in political activism but it required orchestrated influence campaigns to be effective. The following pattern was followed to foment political protests. Find a controversial issue (reopen America). Embed bots and trolls in the groups on each side of the message. Increase the apparent size of the group on each side (spike of new accounts created in late April). Use bots and trolls to increase cohesiveness among protestors, and to attack leaders of the opposition. (Bots sent messages linking reopen supporters to each other. Bots sent messages attacking Democratic leaders in Michigan, North Carolina, and Pennsylvania). Send messages fostering fear and upset among potential protesters (disinformation stories about empty hospitals, government actions, etc.). Use consistent messaging on the protest side. Promote lack of coordination among opposition (e.g., sending distraction messages). The result was that the pro-reopen side became more coordinated, more connected, used more common hashtags, and was suffering from dismay messages. Whereas the anti-reopen side grew larger and more disorganized, had its leaders attacked, and was inundated with distracting arguments. Being able to track these influence campaigns is now more possibly due to the BEND maneuver technology; however, this needs to be extended to other social media platforms.

Polarization campaigns were used repeatedly around health, safety, and job-related issues. In each case, bots and trolls were used strategically on both sides. Disinformation was embedded in master narratives to make it more believable and create a common theme. Issues were consistently reframed to become political. For example, consider the online face mask discussion. Early messaging began by building a pro- and anti- face mask group. Excite campaigns were used to encourage wearing masks. By April, enhance and explain campaigns began to dominate, arguing both why face masks do or do not work. Distraction messages were used to switch the discussion to being a right-to-choose issue, rather than a communal health and safety issue. In October, a surge of new accounts entered the scene, framing this as a right-to-choose and aligning this with political parties. Why distraction works and how to counter it needs further research.

Messages containing hate speech are generally sent by humans, not bots. But bots were used in March and April to link those spreading hate speech to each other, thereby creating communities of hate. Some of these groups were then redirected toward other issues such as face masks and fraudulent voting. The level of hate speech kept going up. The links to and support for QAnon messaging kept going up. It looked like things were being orchestrated for very violent activity. Then Twitter stepped in. Posts supporting QAnon messaging and containing hate speech decreased. Was this intervention positive? Is this just a correlation? Many of those actors appear to have gone “underground.” What might be the implications? Despite this, there was a spike in QAnon discussion on Oct. 17 and 18, and associated coordination among the right.

Messaging regarding postal voting and voter fraud increased throughout October. Fraud and anti-postal-voting discussion was approximately four times higher. Bots engaged in the discussion on both sides and represented around one-third of the users. They played different roles on each side. Coordination on the anti-postal-voting/voter-fraud side, and discoordination on the other. Very little of the discussion on election fraud was abusive. However the abusiveness spiked on Oct. 30. As with other abusive speech, these posts appeared to be trying to prime people for a fight. The vote fraud story is still ongoing. What we don’t know is the connection among these diverse events. We can tell that the online debate is being orchestrated — but by whom, and how this fits across platforms is not known.

Compared to 2016, in 2020 we are faster at tracking conversations, faster and more accurate at identifying bots, faster and more accurate at identifying hate speech, and we are beginning to be able to identify intent with the BEND maneuvers. However, the environment is much more volatile, the data more fragile, and there are new types of adversarial actors, such as cyborgs. We can tell that some of the messaging is targeting minority and at-risk communities — but identifying this automatically is not possible. Compared to 2016, we now know some things groups should do to promote a more secure cyberinformation environment — but it requires changing the way they operate, such as maintaining a social cybersecurity team.

Filling these gaps and addressing the related issues is not just a matter of more research. It is a matter of more coordination among researchers, of making research policy relevant, of changing the way research is conducted, of building a true science advisory for the cyberinformation environment.

Kathleen M. Carley is a professor in the School of Computer Science at Carnegie Mellon University.