Detecting Malicious Information Campaigns

May 9, 2018

Traditional marketing has been supplanted by sophisticated, ultra-coordinated multi-channel information campaigns, but watchdog organizations are just beginning to catch up. How can we apply the latest machine learning developments in order to better flag malicious messages?

Challenges

Fact-checking sites have responded to the proliferation of fake news with gusto. However, false information is only one facet of malicious information campaigns. Messages can be true but based on stolen information. Messages can also be true but originate from a fake identity designed to mislead. On the other hand, messages can be false but shared by real citizens or journalists who believe their contents and are sharing without malicious intent.

Other watchdogs are focused on detecting bots, but automation in and of itself is not necessarily a sign of maliciousness. Of course, a bot network pretending to be a group of real people and disseminating fabricated facts is likely executing on a malicious information campaign. However, an organization can be using automated means in order to amplify real facts using informational bots that are clearly identified as bots. In contrast, a group of real people can be centrally directed to spread a malicious campaign through coordinated propagation timelines and channels.

Furthermore, the phrase “fake news” has been weaponized as a political tool, and as such any politically neutral groups working in the space must be careful and exact in their definition of malicious information campaigns.

Our Solution

Code for Democracy evaluates trending narratives to search for messages amplified through automated means, messages that fail fact checks, messages originating from false profiles, messages advocating for violence against specific people or groups, messages that have unnatural viral curves, and messages that correlate with known malicious narratives. Messages that raise two or more of these six flags are considered likely to be part of a malicious information campaign.

Further Reading

Atlantic Council Digital Forensic Research Lab. #TrollTracker: How To Spot Russian Trolls. (2018, March 29). Retrieved from https://medium.com/dfrlab/trolltracker-how-to-spot-russian-trolls-2f6d3d287eaa

Bessi, A., Zollo, F., Vicario, M. D., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Trend of Narratives in the Age of Misinformation. *Plos One,10*(8). doi:10.1371/journal.pone.0134641

Figueira, Á, & Oliveira, L. (2017). The current state of fake news: Challenges and opportunities. *Procedia Computer Science,121*, 817-825. doi:10.1016/j.procs.2017.11.106

Meyer, R. (2018, March 08). The Grim Conclusions of the Largest-Ever Study of Fake News. Retrieved from https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/

Rosenberger, L., & Berger, J. (2017, August 2). Hamilton 68: A New Tool to Track Russian Disinformation on Twitter. Retrieved from http://securingdemocracy.gmfus.org/blog/2017/08/02/hamilton-68-new-tool-track-russian-disinformation-twitter