

Over the past few years, journalists and politicians have often highlighted the presence of state-sponsored online trolls with the mission of swaying public opinion on particular issues. Researchers at Cyprus University of Technology, UCL, University of Alabama at Birmingham, and Boston University have taken a closer look at this phenomenon, hoping to achieve a better understanding of how these actors operate.

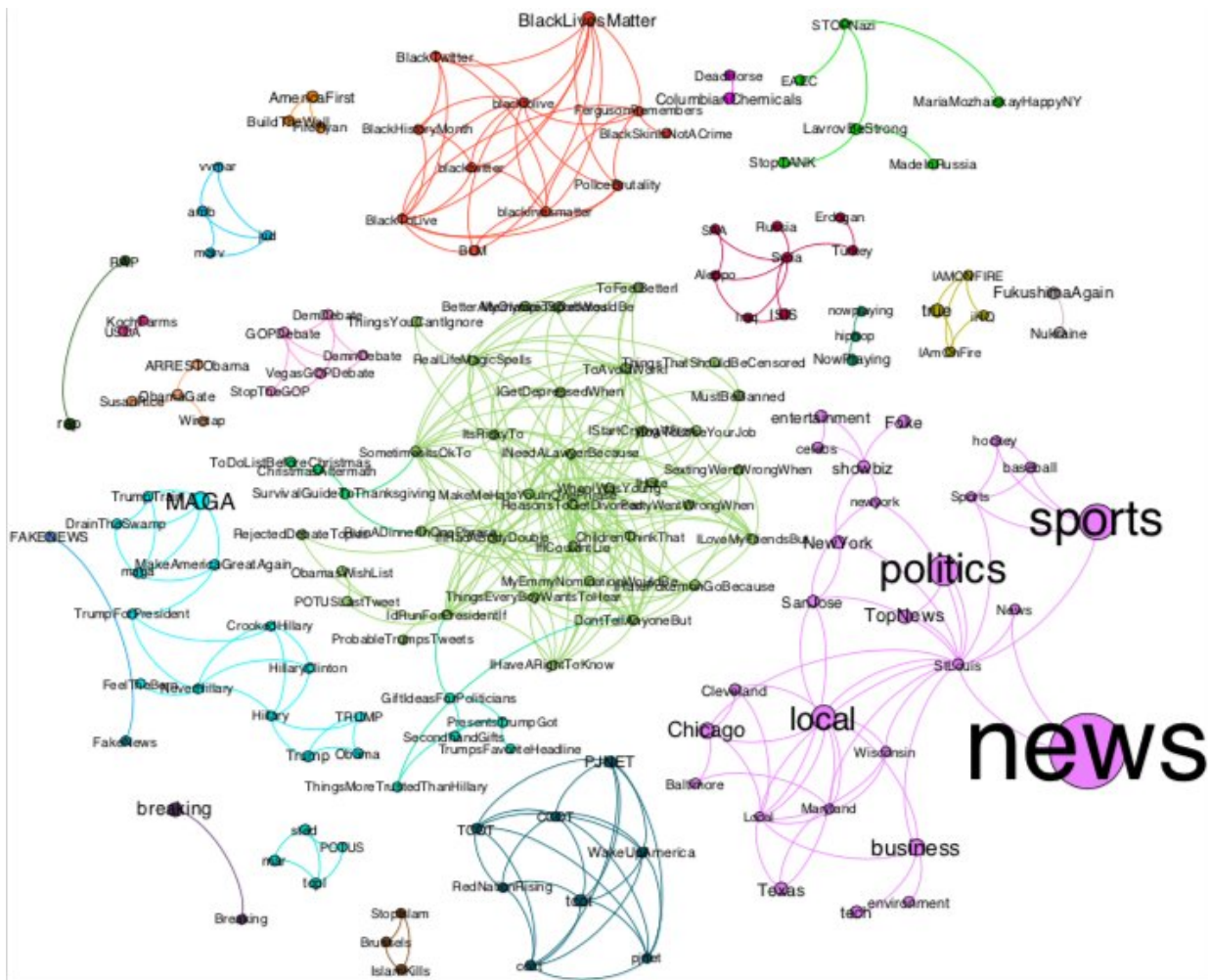
"Facebook has been tuning its systems to detect and block coordinated [malicious behavior](#) for several years," Gianluca Stringhini, assistant professor at Boston University, told TechXplore. "The threats faced by platforms, however, are rapidly changing. Until a couple of years ago, the main threat came from malware controlling fake accounts to spread spam and other frauds. This type of activity is controlled by programs and not by humans, and is therefore very precisely coordinated."

In recent years, [social media platforms](#) have witnessed the rise of a new type of threat: coordinated state-sponsored [trolls](#) pushing a particular narrative on social media. This activity derives from humans rather than automated programs and is hence harder to tell apart from normal online chatter.

Currently, there is a growing divide in the U.S., much of which is caused by political tribalism said to be fueled by foreign actors. This theory was confirmed when Twitter and Reddit released information about Russian and Iranian trolls, who had been making bold statements on their platforms.

"Our goal was to take advantage of this precious ground truth dataset, analyzing it across several axes in order to better understand the behavior and the influence of these state-sponsored accounts," Savvas Zannettou, Ph.D. Candidate at Cyprus University of Technology, told TechXplore. "These insights are of paramount importance for developing robust solutions for detecting these actors, mitigating their impact on the web,

and also for raising public awareness regarding this emerging form of information warfare."



Visualization of top hashtags used by Russian trolls on Twitter. Credit: Zannettou et al.

Zannettou and his colleagues analyzed the data released by Twitter and Reddit across several different axes in an attempt to understand how these state-sponsored trolls operate, how their posts evolve over time,

who their targets are and how they influence the online community. Their [analysis](#) can be divided into three main themes.

"First, we analyzed various characteristics to assess the behavior of these accounts over time," Zannettou said. "Second, we used models called word embeddings to understand the linguistics and use of language of these state-sponsored accounts. Finally, we leveraged a statistical framework known as Hawkes Processes to model and quantify the influence that these accounts had on other communities, such as the rest of Twitter, Reddit, 4chan and Gab."

The analyses carried out by the researchers offer valuable insight on this widely discussed form of state-sponsored propaganda. Firstly, the researchers observed that the behavior of state-sponsored actors varied over time, making them increasingly harder to detect.

"By analyzing their behavior over time and their reported locations, we were also able to shed light on these agents' targets," Zannettou said. "Specifically, we found that Russian trolls were mainly posing as users from Russia, the U.S. and the U.K., while Iranian trolls were mainly posing as users from France, Brazil, the U.S., Saudi Arabia and Turkey."

Russian trolls were typically pro-Trump, while Iranian ones were against him. This suggests that regular users could be inadvertently caught in the middle of discussions between state-sponsored actors supporting conflicting ideologies. Moreover, the researchers observed that Russian-backed trolls were more sophisticated and subtle than Iranian-backed actors.



The map shows the reported locations of Russian trolls (red circles) and Iranian trolls (green triangles) on Twitter. The size of the shapes denotes the number of tweets that were posted from the specific reported location. Credit: Zannettou et al.

"This is evident from Russian trolls' dissemination of random and seemingly general purpose content, while Iranian trolls were more straightforward and focused on disseminating their propaganda," Zannettou explained. "Iranian campaigns are evidently motivated by real-world events. For example, we find that Iranians started communicating in French on November 2013, [around the time in which France opposed to the nuclear talks](#). The same applies to Arabic after [the attack on the Saudi Arabia embassy in Tehran](#), in January 2016."

Zannettou and his colleagues observed that Russian trolls were more proficient in spreading information on most online communities, with the exception of 4chan's Politically Incorrect board, where Iranian actors were more influential. So far, their work has focused on the text content

and metadata of tweets and online posts. Nonetheless, many state-sponsored accounts also share images and videos.

"We now aim to analyze images and videos to understand what kind of information trolls disseminate in these formats and whether these were part of their propaganda," Zannettou said. "Finally, we aim to assess the influence that these accounts had to the rest of the web over time."

Interactive graphs are available at trollspaper2018.github.io/trollspaper2018/#russians_graph_gexf and at trollspaper2018.github.io/trollspaper2018/#iranians_graph_gexf

More information: Who let the trolls out? Towards understanding state-sponsored trolls. arXiv:1811.03130 [cs.SI].
arxiv.org/abs/1811.03130

© 2018 Science X Network

Citation: Who let the trolls out? Researchers investigate state-sponsored trolls (2018, November 23) retrieved 26 April 2024 from <https://techxplore.com/news/2018-11-trolls-state-sponsored.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.