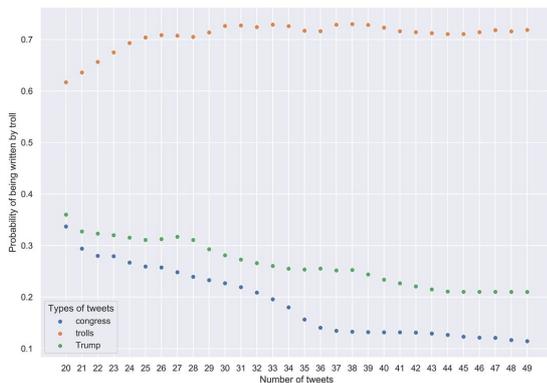# A novel strategy for quickly identifying twitter trolls

12 August 2020

Congress, troll, and Trump's tweets. Results of Bayesian inference for 50 random tweets. Credit: Monakhov, 2020 (PLOS ONE, CC BY)

Two algorithms that account for distinctive use of repeated words and word pairs require as few as 50 tweets to accurately distinguish deceptive "troll" messages from those posted by public figures. Sergei Monakhov of Friedrich Schiller University in Jena, Germany, presents these findings in the open-access journal *PLOS ONE* on August 12, 2020.

Troll internet messages aim to achieve a specific purpose, while also masking that purpose. For instance, in 2018, 13 Russian nationals were accused of using false personas to interfere with the 2016 U.S. presidential election via social media posts. While previous research has investigated distinguishing characteristics of troll tweets—such as timing, hashtags, and geographical location—few studies have examined linguistic features of the tweets themselves.

Monakhov took a sociolinguistic approach, focusing on the idea that trolls have a limited number of messages to convey, but must do so multiple times and with enough diversity of wording and topics to fool readers. Using a library of Russian troll tweets and genuine tweets from U.S. congresspeople, Monakhov showed that these troll-specific restrictions result in distinctive patterns of repeated words and word pairs that are different from patterns seen in genuine, non-troll tweets.

Then, Monakhov tested an algorithm that uses these distinctive patterns to distinguish between genuine tweets and troll tweets. He found that the algorithm required as few as 50 tweets for accurate identification of trolls versus congresspeople. He also found that the algorithm correctly distinguished troll tweets from tweets by Donald Trump—which although provocative and "potentially misleading," according to Twitter, are not crafted to hide his purpose.

This new strategy for quickly identifying troll tweets could help inform efforts to combat hybrid warfare while preserving freedom of speech. Further research will be needed to determine whether it can accurately distinguish troll tweets from other types of messages that are not posted by public figures.

Monakhov adds: "Though troll writing is usually thought of as being permeated with recurrent messages, its most characteristic trait is an anomalousdistribution of repeated words and word pairs. Using the ratio oftheir proportions as a quantitative measure, one needs as few as 50tweets for identifying internet troll accounts."

Provided by Public Library of Science

APA citation: A novel strategy for quickly identifying twitter trolls (2020, August 12) retrieved 3 December 2021 from https://techxplore.com/news/2020-08-strategy-quickly-twitter-trolls.html