6th International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2020, Massachusetts Institute of Technology

The Vagaries of Robotic Trolls

Keywords: Trolling Behavior, Twitter, Political Propaganda, Social Media, Online Communities

Extended Abstract

Robotic trolling instigated by instruments of the Russian Federation is widely believed to have influenced the 2016 Presidential election in the United States. It is, however, extremely difficult to measure or investigate actual political, let alone psychological and sociological, effects caused by trolling and other similar online behavior. Consequently, we have concentrated on quantifiable patterns of robotic troll behavior, both as a means of identification, but also as a way of investigating the political logic behind that behavior.

Data: The dataset of tweets that we used in this paper was developed by (Boatwright et al). This dataset is a log of tweets from June 2015 to December 2017 that were associated with Russia's Internet Research Agency (IRA). The IRA has been previously linked with activities to influence the US political agenda.

In this dataset (GitHub), the authors have categorized the 1,875,029 tweets originating from IRA as belonging to one of five categories: Right Troll, Left Troll, News Feed, Hashtag Gamer and Fear Monger. The categorization of these tweets are based on the content of the tweet. For our work, we focus on the behavior of a subset of these categories related to trolling activity, i.e. Left Trolls and Right Trolls. Work in (Boatwright et al.) categorizes Right Trolls as a category of tweets related to nativist and right-leaning populist messages. Commonly employed features of this category include MAGA (Make America Great Again), support of the candidacy of Donald Trump, denigrating the Democratic party and sending divisive messages about mainstream and moderate Republicans. The tweets belonging to the category of Left Trolls emphasized socially liberal messages focusing on identity. Prominent themes include Black Lives Matter, Blacktivists, and LGBTQ messages. These tweets supported the candidacy of Bernie Sanders, while denouncing the moderate and mainstream Democratic party candidates, particularly Hillary Clinton.

Analysis: In order to test the patterns we had discovered in the IRA data, we collected and analyzed data from sets of users for whom we had a high confidence that they were human and not Twitterbots. At one extreme we looked at the Twitter behavior of celebrities and well-known users. Such users, as we expected, did have significantly more followers than those they follow. Spam accounts invert this pattern, as is made explicit in Twitter's Spam policy – which states that spam accounts are flagged if "…if you have a small number of followers compared to the amount of people you are following." In between these two extremes we find ordinary users. These users, while they do tend to have more followers than those they follow, nevertheless have a much smaller differential between these two numbers than do celebrities. We capture this ordinary behavior by collecting Twitter accounts participating in the #MondayMotivation and #NewYearsResolution hashtags. The logarithmic ratio of numbers of follower to following accounts clearly shows the significant differences between human users and robotic trolls.

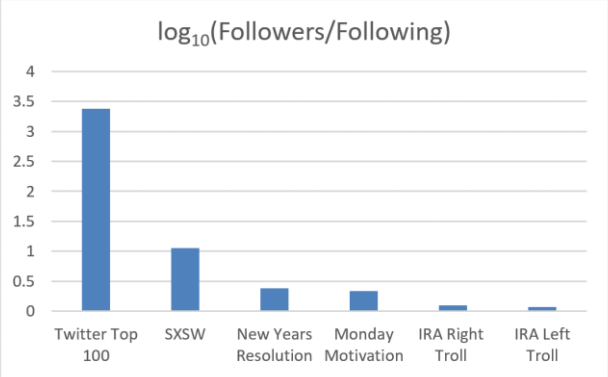

Figure 1. Logarithmic ratio of numbers of follower to following accounts.

Interpretative Analysis: The numbers of left and right troll accounts in the IRA data set were quite close and their follower/following ratios similar. The goal of IRA was clearly not to add significantly to the number of total tweets supporting either political persuasion. Given this fact, the information or content of the tweets served to confirm or provoke particular sets of Twitter users. The expression of leftist positions would be as likely to produce confirmation in right leaning Twitter users as would the expression of rightist positions. This is one explanation for the IRA investment in both kinds of trolls. This would be true of left leaning Twitter users, as well, but we are assuming given the current politics of the Russian Federation that this was an unlikely target audience. Alternatively, the goal might simply have been to encourage the polarizing discourse exemplified by both kinds of trolls.

Conclusion: Robotic trolls in this case demonstrate a follower/following profile remarkably different from a full range of human users. In the short-run these profile characteristics can be used to identify Twitterbots. It is harder, however, to understand the effects of such trolling behavior. The function of these IRA trolls is not to distribute fake news stories that would benefit one side or the other. If that were the goal, it would be unlikely that IRA would create a similar number of left and right troll accounts that are functionally equivalent. The more likely goal is provocation and confirmation, and more generally to increase the polarization of political discourse in the United States. That this action on the part of IRA supports the right leaning political positions is a conclusion that cannot be drawn from the data, but rather is a framework based on the political commitments of IRA and its role in the Russian Federation. In addition, it also suggests a need for more precise and careful categorization of data and more nuanced conceptualizations of political motivations and actions.

References

1. Boatwright, B. C., Linvill, D. L., & Warren, P. L. (2018). Troll factories: The Internet research agency and state-sponsored agenda building. *Resource Centre on Media Freedom in Europe*.

2. Twitter (2009b). The Twitter rules. http://status.twitter.com/post/136164828/restoringaccidentally-suspended-accounts.

3. GitHub. https://github.com/fivethirtyeight/russian-troll-tweets