Fake News in Online Social Networks: Community Takeover on Reddit

Abstract

This research analyzes several communities within Reddit that were infiltrated by the Russian Internet Research Agency. The IRA shared information with these subreddits and used their accounts to bypass the normal gatekeeping phase, ensuring a higher visibility to the overall community. By using an agent base model to simulate the interactions between users and posts (both organic and malicious) as well as subreddit-specific metrics, a method to determine the number of IRA accounts dedicated to infiltrating each given community is determined.

Introduction

Reddit Background:
- 6th most visited site in the United States
- 20th most visited site in the world
- 330 million monthly active users (Huffman & Ohanian, 2017).

How Reddit Operates:
- Users subscribe to any of the 138,000 active subcommunities called subreddits.
- Users post content (news articles, images, or text) that others within these communities may upvote or downvote.

Identified Russian Influence:
- Reddit's 2017 transparency report found 944 suspicious accounts linked to the Russian Internet Research Agency (Huffman, Reddit, 2018).
- These accounts were posting divisive content to 454 subreddits over the course of two years.
- The subreddits spanned from a wide range of interests including US politics, world politics, bitcoin, blockchain, and animal images.
- For this research, "Trolls" are the IRA accounts posting to Reddit.

Methods

Agent-Based Model
- Users operate as the agents
- Posts are resources
- Users can upvote or downvote posts
- Preferential attachment models visibility of the post based off its upvotes and downvotes.

Gatekeeping:
- Extra-active users sort by "new" to view all content posted to a subreddit.
- By upvoting early, they ensure the broader community sees the post.
- By downvoting, they can effectively bury the post.

The Trolls:
- Trolls immediately upvoted posts in an attempt to move them past the gatekeeping stage and into the view of the main community.
- The numbers of bots were varied to match the score distribution of posts from the IRA in various subreddit communities.

Validation

- Performed by matching the distribution of scores for troll posts to that of the IRA posts by subreddit.

Model Findings

- The agent-based model is able to find posts with the same average score as those from the IRA within subreddits; however, the variance in the scores from the model is much lower than that of the IRA posts.
- Additional attributes of the post need to be considered to determine virality.

Conclusions

- Infiltrating subreddits requires a nuanced understanding of the community's norms.
- Reddit behavior can be modeled using a preferential attachment agent-based model tailored to the metrics of the individual subreddits.
- Models have a harder time predicting the truly viral posts that amass thousands of upvotes.

Future Work

- Include attributes of post (image, news article, etc.) that affect upvote ratio
- Incorporate the ability to comment on posts