

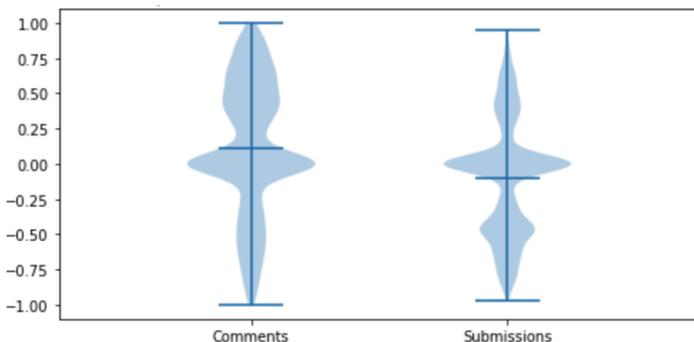
# An Exploratory Analysis of Suspicious Reddit User Accounts based on Sentiment and Interactions



Eunjin Lee (IBM UK), James Ashford (Cardiff), Malgorzata Turala (ARL), Liam Turner (Cardiff), Vera Liao (IBM US), Rachel Bellamy (IBM US), Geeth de Mel (IBM UK), Roger Whitaker (Cardiff).

## Objectives

- Coalition operations require a better understanding of how individuals and social groups interact in varying contexts.
- Online social networks shaped by user activity offer a unique opportunity for us to observe such group behaviors.
- We explore user behavior on **Reddit**, a large social network site designed as a discussion forum for various topics.



The compound sentiment scores of the comment bodies and submission titles

## Approach

- Usernames of 944 suspicious accounts were taken from a **pre-compiled list** published on the Reddit Wiki, which was a part of Reddit's 2017 transparency report.
- Using the subreddit activity and comment history collected through the Python Reddit API Wrapper (PRAW), the time at which a user commented or submitted content was aggregated into a **timeline graph**.
- The body of the content and the submission titles of each activity were passed through the Valence Aware Dictionary and Sentiment Reasoner (VADER) to calculate the **compound sentiment scores**.

## Results

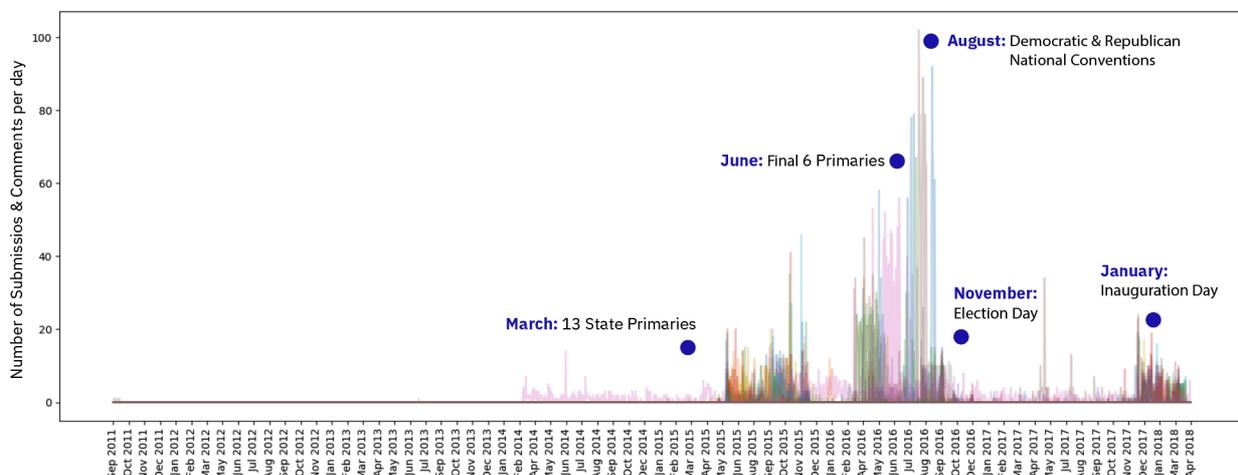
- Of the 944 accounts, **66% (659)** did not engage in discussion or post any content to Reddit. They may have up- and down-voted content posted by others.
- The timeline shows that **activity increases running up to major events of the US 2016 Presidential Election**, where the largest spike of activity coincides with the National Conventions
- Many accounts were **active many years prior** to the election.
- The compound sentiment of the comments and submissions shows that for both activities, the median score is neutral. The sentiment scores for the submission titles is more **negative** than that of the comment bodies.

## Summary & Future Work

- To understand temporal behavior patterns in different account types, we will **compare** the suspicious user account patterns to a class of 'normal' accounts who were active in the same time period.
- To understand the impact of the suspicious activity on the Reddit community, we will carry out further statistical analysis on how much **engagement** each comment or submission received and also run sentiment analysis on the reply **responses** from the community.
- To characterize the activity patterns of the both the 'normal' and suspicious accounts, we will use **network motif analysis**.

## Military & Coalition Relevance

- Malicious users spreading misinformation can disrupt groups and societies relevant to coalition operations.
- Generalizable, machine-driven methods can help in identifying potential influential users.



Timeline of suspicious accounts' activity with key events from the 2016 US Presidential Elections.