# Network Based Analysis of Reddit troll Activity

EE 382V - M3 Poster Session
Dec 3, 2019

**Arka Sanka[1], and Sunny Sanyal[2]**
The University of Texas at Austin, USA[1,2]
arkasanka@utexas.edu[1] and sanyal.sunny@utexas.edu[2]

## 1. Introduction

Reddit is a social media platform where users form communities also known as subreddits to discuss various topics. A transparency report published by Reddit in 2017 states the presence of 944 Russian trolls. The report also found that these trolls manipulated public opinion during 2016 and 2018 US elections.

Goal: We identify trolls from a huge Reddit dataset and study their behavior through the lens of network science.

### Research Problems

- How to identify trolls in a dataset where the ground is not known to us?
- How do the trolls affect the overall network scenario?
- How does the troll behaviour changes over time?

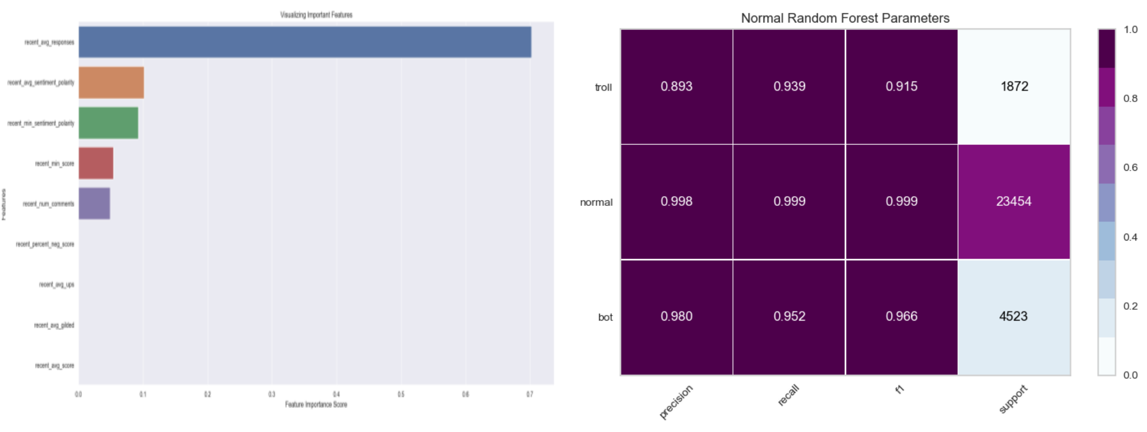
### Major contributions

- We built a random forest classifier to detect trolls.
- We make political inferences out of the network and temporal analysis.

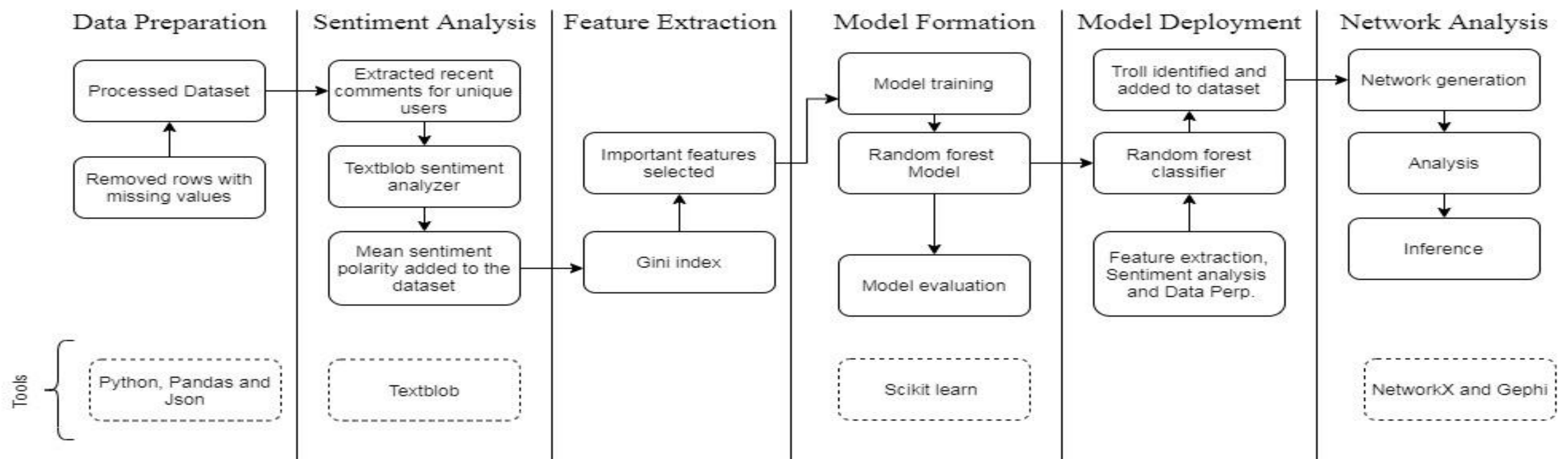
## 2. Approach

- We build a random forest classifier and apply it on a new dataset to detect trolls.
- We analyze the properties of the Russian trolls and the new trolls using network science.

Random Forest Classifier results



A high F1 score indicates that our model is highly accurate.

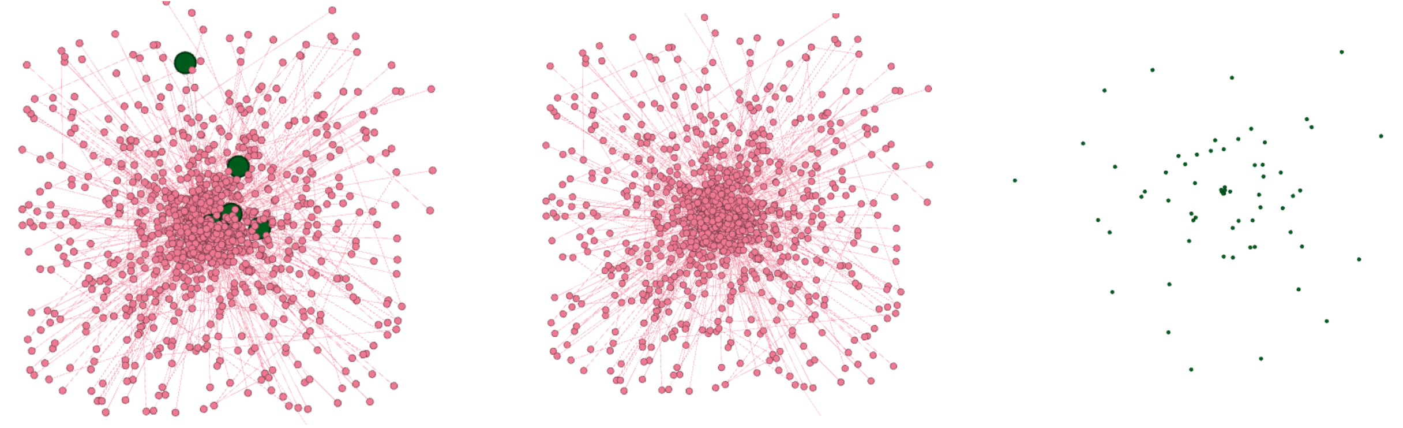


A Detailed flow of our approach

## 3. Results

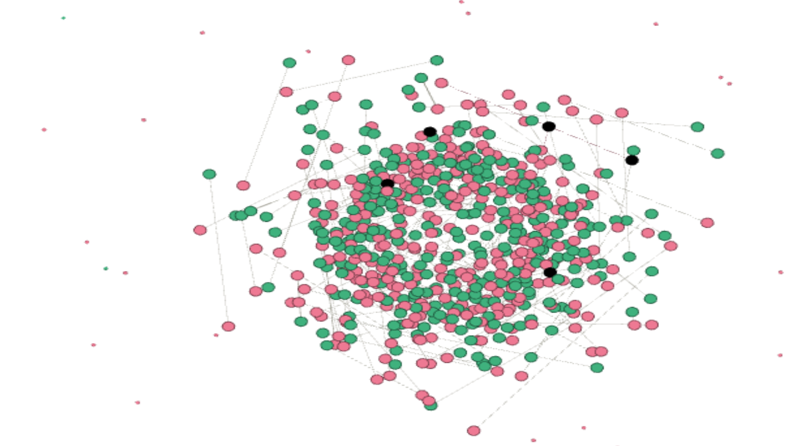
### Network Analysis

**The network properties are not affected by the presence of trolls.**



User to User networks based on Sep 2018 data.

| Network type | Avg. Degree | Avg. Clustering Coefficient | Avg Path lengh | # Nodes | # Edges |
|---|---|---|---|---|---|
| User and trolls | 177.544 | 1 | 1 | 19320 | 1715078 |
| User only | 177.434 | 1 | 1 | 19243 | 1707183 |
| Troll only | 0.234 | 0 | 1 | 77 | 9 |



User to Post network based on 2015- 2018 troll data.

| Network type | Avg. Degree | Avg. Clustering Coefficient | Avg. Path length | # Nodes | # Edges |
|---|---|---|---|---|---|
| Troll only | 0.952 | 0 | 1 | 643 | 306 |

## 4. Summary

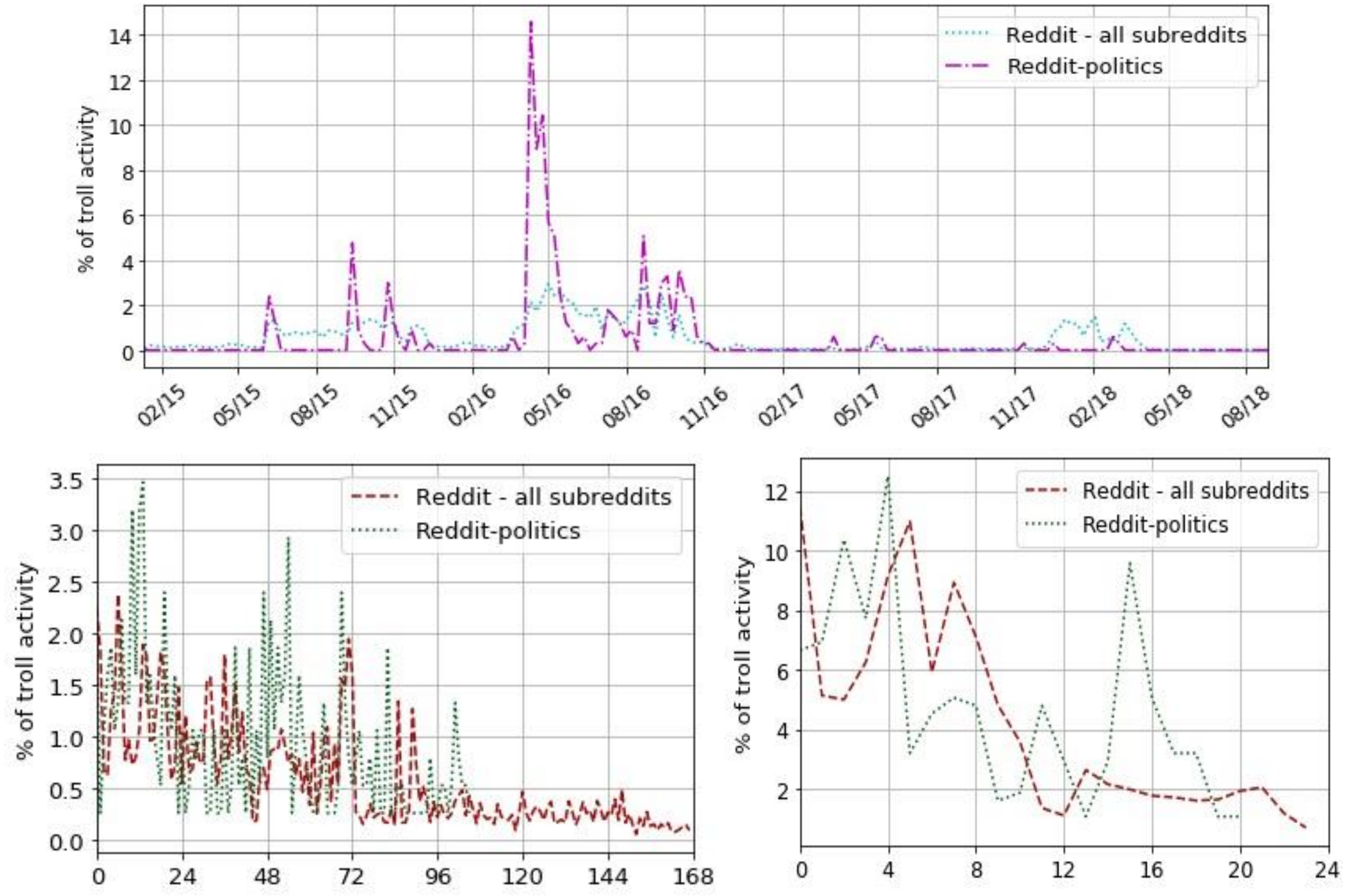
- In this work, we successfully identified and studied the troll behavior.
- We observed that the trolls do not affect the network properties and the behavior is time variant.

Futurework: As a future work we plan to build a real-time suspected troll detection system to be used by human moderators. We also plan to analyze the trolls across other platforms such as Twitter and Facebook.

## 3. Results - continued

### Temporal Analysis

**The behavior of trolls vary over time.**

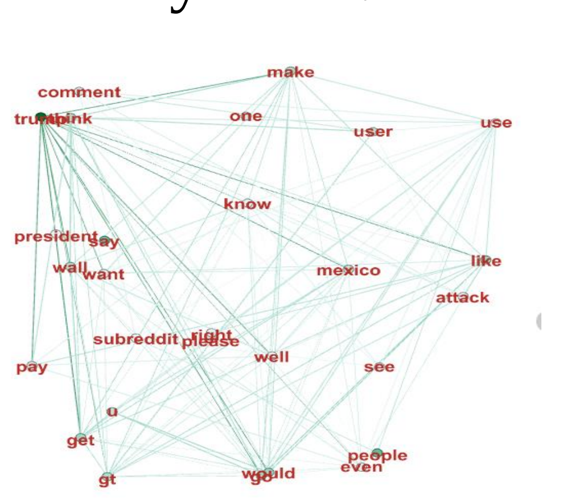


Temporal behavior of trolls as a function of weeks, hours of a week and hours of a day respectively.

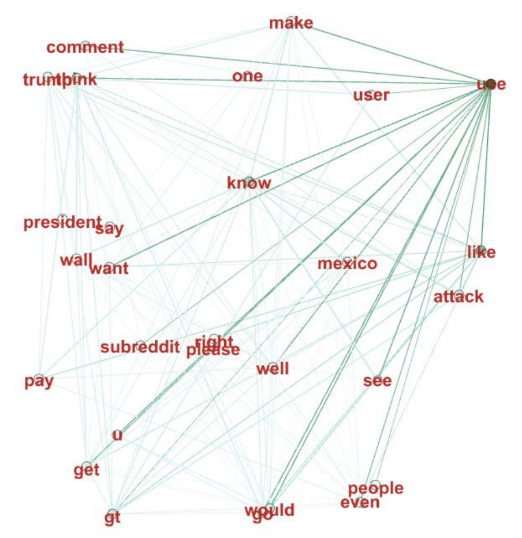
### Word Frequency Analysis

**Mostly the trolls are Trump supporters.**

The plots below represent the words used by trolls. Words are connected if their cosine similarity based on extracted Word2Vec embeddings is greater than 0.6. For both these plots we consider words used more than 1300 times by trolls.



The words were ranked by word-count and we see that the word trump is used very often by trolls.



The words were ranked by their betweenness centrality in the graph.