

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

Eunjin Lee¹, James Ashford², Malgorzata Turalska³, Liam Turner², Vera Liao⁴,
Rachel Bellamy⁴, Geeth de Mel¹, and Roger Whitaker²

¹IBM Research (UK), Hursley, UK

²School of Computer Science and Informatics, Cardiff University, UK

³CCDC Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD, U.S.A.

⁴IBM TJ Watson Research Center, Yorktown Heights, NY, USA

Abstract—Online social networks shaped by user generated content and activity offer a unique opportunity for understanding how individuals and social groups interact. Insight into the dynamics of on-line communities could prove valuable in the military context helping with: evaluation of group stability and growth, identification of key influential users, and detection of attempts to spread disinformation. To this end, we explore user behaviour on Reddit, a large social network site designed as a discussion forum for various topics. We present an initial, exploratory investigation using sentiment and activity analysis, of the behaviour and content of 944 suspicious Reddit accounts.

Index Terms—Social Networks, Reddit, Misinformation, Sentiment Analysis

I. INTRODUCTION

The study of how groups of users interact online presents a unique insight into how social networks form and evolve. In particular, analysing the behaviour of users who spread disinformation on social media platforms is useful for inferring the patterns and strategies used by misinformation agencies. This is critical in the military context as understanding the structures of rapidly formed groups and the identification of adversarial agents is vital for the success of missions at hand, especially when looking into the future coalition multi domain operations [1]–[3].

Reddit is an online forum where users submit content to user-created boards referred to as *subreddits*—specifically they are topic-based communities covering issues such as news, politics and humour. Unlike traditional social networks, Reddit emphasises homophilous properties by means of a user-topic structure where users engage with communities of similar interests. As of July 2019, it is the fifth most visited website in the United States (US) and thirteenth in the world. There are over one million subreddits and around 330 million Reddit users [4]. In this paper, we use activity and sentiment analysis,

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorised to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

to explore the commenting and content submission behaviour of a set of known suspicious Reddit accounts.

II. METHODS

A. Activity Analysis

The usernames of the 944 suspicious accounts were taken from the pre-compiled list published on the Reddit Wiki [5], which was part of Reddits 2017 transparency report. These accounts were identified using a number of methods including suspicious creation patterns, usage patterns (including account sharing), voting collaboration and corroborated data from other social network platforms such as Twitter [6]. However, according to the best of our knowledge, sentiment expressed by the users were not considered in identifying these accounts.

There are two main activity types on Reddit—commenting and submitting. A user can comment on a submission created by themselves or by another user. A user can also upload content to share onto a subreddit, such as a link to a news site, a text-based post or an image. Using the subreddit activity and comment history collected from the The Python Reddit API Wrapper (PRAW) [7], the time at which a user commented or submitted content was aggregated into a timeline graph. The key events (e.g., the 2016 US Presidential Elections) were then overlaid onto this graph as shown in Figure 1.

TABLE I
THE TOP FIVE SUBREDDITS OF ACTIVITY BY THE SUSPICIOUS ACCOUNTS

| Activity Type | Subreddit | Activity Count |
|---------------|------------------|----------------|
| Comments | AskReddit | 791 |
| | CryptoCurrency | 581 |
| | worldnews | 269 |
| | Bad_Cop_No_Donut | 221 |
| | news | 199 |
| Submissions | Funny | 1159 |
| | uncen | 854 |
| | Bad_Cop_No_Donut | 532 |
| | PoliticalHumor | 458 |
| | gifs | 293 |

B. Sentiment Analysis

The body of the content and the submission titles of each activity was passed through the Valence Aware Dictionary and

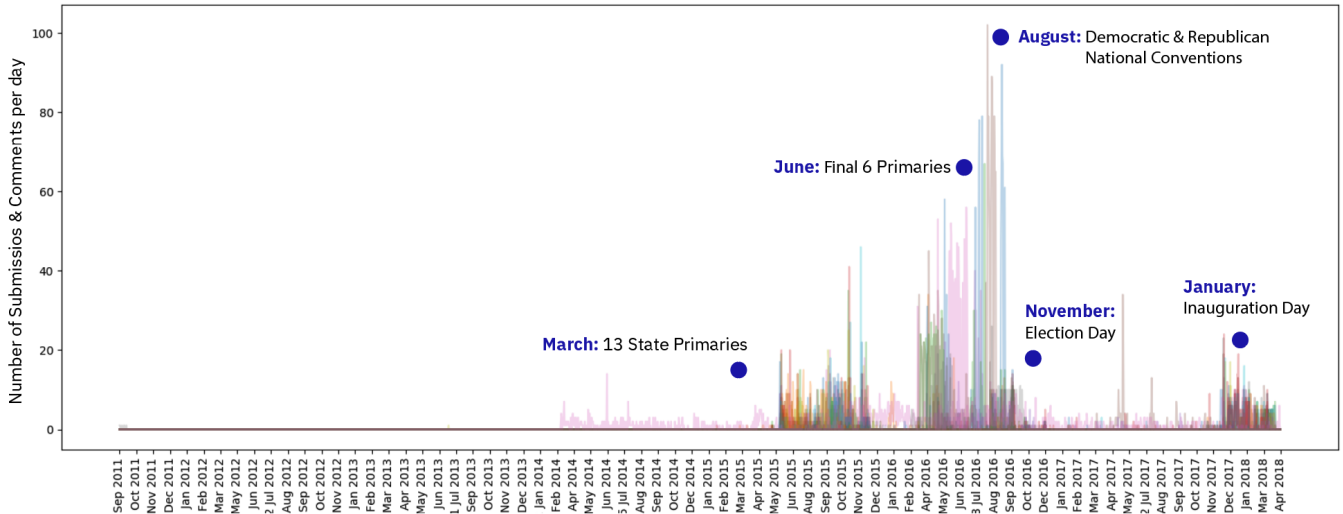


Fig. 1. Timeline of suspicious accounts' activity with key events from the 2016 US Presidential Elections.

Sentiment Reasoner (VADER), a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media [8]. The compounds score, a metric that calculates the sum of all the lexicon ratings which have been normalised between -1 and +1, was calculated for each activity and plotted as a violin-plot as shown Figure 2.

III. RESULTS

A. General

Out of the 944 accounts, 70% (662) accounts had zero karma. Karma is the Reddit-specific metric that is an aggregate of the up- and down- votes of the comment or content, received from other users. 66% (659) of these suspicious accounts did not engage in discussion or post any content to Reddit. A potential hypothesis may be that these dormant accounts were up- and down- voting others' content but there is no way to verify this information using the data that can be obtained using the PRAW API. The total number of submissions created by the accounts was 11,053, where 89.4% of these were links to external news and content websites. The total number of comments posted by the users was 6,711. Table I shows the top 5 subreddits where the accounts were commenting and submitting on.

B. Activity Analysis

Figure 1 shows the frequency of activity of the accounts in time alongside the 2016 US Presidential Elections. The data shows that some of the accounts were created and active up to five years before the elections. The timeline shows that the activities increase running up to major events of the election, where the largest spike of activity coincides with the National Conventions and through the months leading up to the election day in November.

C. Sentiment Analysis

The compound sentiment of the comments and submissions seen in Figure 2 shows that for both activities, the median score is neutral. The sentiment scores for submission titles is more negative than that of the comment bodies.

IV. DISCUSSION

The initial study of the suspicious accounts behaviour highlight some of the strategic patterns, especially where and how these accounts engage on the platform. It is at first surprising to see that there have been so many submissions to the /r/funny community (Table I). Although this and other subreddits are seemingly unrelated to the presidential elections, a known tactic of those trying to spread disinformation, is to establish the credibility of an account by gaining Karma points from posts to neutral, popular topics, and then use these now credible accounts for disinformation or for other manipulative agenda [9].

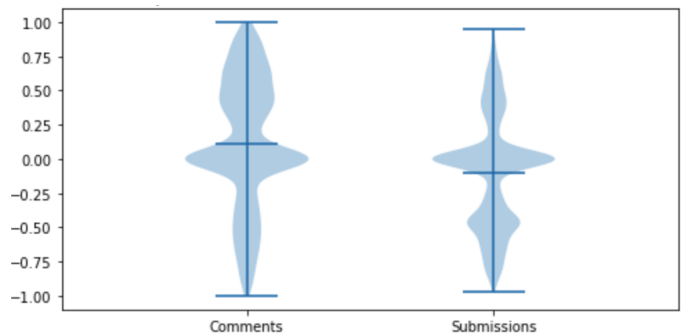


Fig. 2. The compound sentiment scores of the comment bodies and submission titles

Another tactic is to use accounts which have been on the platform for a while, as one known indicator of a suspicious

account is that the account was created moments before the comment or submission. The Figure 1 timeline indicates that some of these accounts may have been posting years in advance to increase their credibility. Another possibility is that these were originally credible Reddit accounts, that were then purchased for intentional misuse; another common tactic for obtaining credible account to use for spreading disinformation.

Provocative posts expressing a negative sentiment have been shown to receive more replies and engagement as they elicit a response from the community [10] [11]. This may explain why the sentiment for the submission titles are lower than that of the comments. Comments are people reacting to the provocative post. Further analysis is required to understand commenting and submitting strategies, when and for what purpose different strategies are used. For example, are their identifiable patterns that indicate an account switching from establishing credibility to carrying out malicious activity.

V. FUTURE WORK

A key future study to extend the work presented in this paper is to compare the suspicious user account patterns to a class of 'normal' accounts who were active in the same time period. This will show whether the patterns of activity of the suspicious accounts are indicative of the strategies used for disinformation spreading, or whether these trends are similar for users engaged around the election period. Norlander [9] did an analysis of comparing the suspicious accounts' creation time, common use and other factors with a class of 'normal' users and found distinctive differences between the accounts. He has not however, compared the sentiment or how the content was received between the two classes of accounts.

Analysing how the content was received by the communities is another important aspect we will investigate, as simply looking at the aggregate Karma score of the users and their content is not enough information to fully analyse their strategies. We will carry out statistical analysis on how much engagement each comment or submission received and also run sentiment analysis on the reply network response from the community, to analyse the impact of the suspicious activity on the Reddit community.

How the suspicious accounts interacted with each other and whether they coordinated efforts is another area we will look at in the future. We plan to use network motif theory to analyse the patterns of the both the normal and suspicious accounts activity. By analysing temporal and structural properties of conversations we will pursue two complementary paths of inquiry: one focused on users and a second one focused on group interaction. We are interested in identifying roles users play in the forum and in quantifying the sentiment of the discussion, so as to identify indicators of intra- and inter-group conflict, depending on the topic.

Possible military applications of our work include identification of malicious users, as well as early detection of topics that could be used to form or increase divisions within society.

REFERENCES

- [1] N. Marchal, "Automation, algorithms, and politics— when bots tweet: Toward a normative framework for bots on social networking sites (feature)," *International Journal of Communication*, vol. 10, no. 0, 2016.
- [2] A. J. Szep, "Algorithmic analysis & visualization of bots influence on us elections & your vote," Master's thesis, University of Arizona, 2019.
- [3] D. K. Spencer, S. Duncan, and A. Taliaferro, "Operationalizing artificial intelligence for multi-domain operations: a first look," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, vol. 11006. International Society for Optics and Photonics, 2019, p. 1100602.
- [4] Alexa Internet. (2019) Reddit competitive analysis, marketing mix and traffic. [Online]. Available: <https://www.alexa.com/siteinfo/reddit.com> (Visited on 29-Jul-2019)
- [5] Reddit.com. (2019) Suspicious accounts investigated by reddit. [Online]. Available: <https://www.reddit.com/wiki/suspiciousaccounts>
- [6] "Reddit.com". (2019) Reddits 2017 transparency report and suspect account findings. [Online]. Available: <https://www.reddit.com/r/announcements/comments/8bb85p>
- [7] Bryce Boe. (2016) Python reddit api wrapper. [Online]. Available: <https://praw.readthedocs.io/en/latest/>
- [8] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth international AAAI conference on weblogs and social media*, 2014.
- [9] B. Norlander, "Reddit bot classifier," 2018.
- [10] A. Guimaraes, O. Balalau, E. Terolli, and G. Weikum, "Analyzing the traits and anomalies of political discussions on reddit," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, no. 01, pp. 205–213, Jul. 2019.
- [11] D. García, A. Garas, and F. Schweitzer, "Positive words carry less information than negative words," *CoRR*, vol. abs/1110.4123, 2011.