# Those Trumpin' Tweet Trolls: Analyzing Behaviors of Russian Twitter Trolls and Genuine Twitter Users Through Use of Hashtags

Bharti Goel, Arup Kanti Dey and Anthony Windmon
*Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA*
*Emails: bharti@mail.usf.edu, arupkantidey@mail.usf.edu, awindmon@mail.usf.edu*

***Abstract*—Following the 2016 United States Presidential Election, evidence has surfaced proving that election results were tainted by online actors, called Trolls. Trolls, commonly housed on social media networks (i.e., Twitter, Facebook, Reddit, etc.), spread propaganda through these networks, intending to shift user's perspective to match theirs. In 2017, Twitter released data alleging trolls of Russian descent played a major role in the outcome of U.S. Presidential Election. In this paper we analyze the Russian troll data released by twitter and data collected by us on the basis of hashtags used by the Russian trolls. We used user description text, tweet text, hashtags, botometer score [7], language used for tweets and user profile language.**

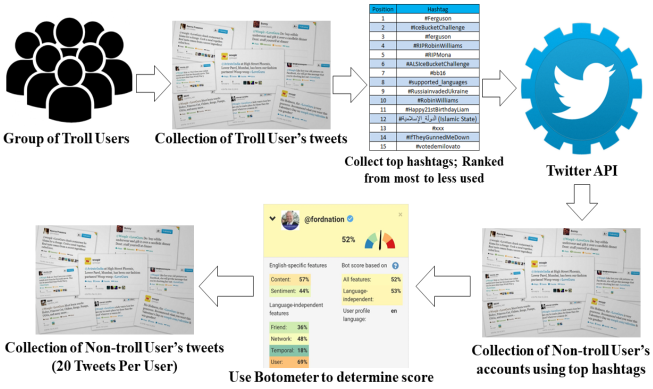
***Keywords*-Twitter, Hashtags, Russian trolls, Genuine users, Social networks, Data mining.**

## I. Introduction

Social media has granted its users opportunities to engage in a plethora of conversations, surrounding a variety of popular topics. Rather users are engaging in dialogue surrounding public school mass shootings, pop culture events (i.e., The MTV Video Music Awards) or The White House's most recent outlandish political scandals, social media networks (i.e., Twitter, Facebook, Reddit, etc.) regularly house these conversations. An ongoing discussion, on social media, is one regarding the final verdict of the 2016 U.S. Presidential Election, where Donald Trump came out victorious. There is substantial evidence proving the election results were achieved using manipulation and influence tactics via social media [1]. Furthermore, is evidence that suggests such tactics were executed by Sponsored Trolls [1].

In 2017, Twitter and Reddit both released data to corroborate these findings. Inside their released datasets, included nearly 10 million public, non-deleted tweets and media (i.e., images and videos), from $3,841$ Russian troll accounts, allegedly involved in persuading public opinion concerning the election [3]. Also, there are 770 Iran accounts included in the data, however these accounts were not in support of Donald Trump. Russian troll accounts, on the contrary, were in support of him. In this paper, we utilize Twitter's public data for Russian accounts to create an analysis enabling us to differentiate between genuine and troll account behaviors. Specifically, we did the following,

- Extracted the top $k$ hashtags from the Russian Troll data, generated by selecting hashtags covering about 90% of total tweets.
- Extracted tweets from $1,000$ genuine users, who incorporated the top hashtags in their tweets.
- Applied Botometer [7], which estimates the likelihood a user is genuine or not.


Group of Troll Users
Collection of Troll User's tweets
Collect top hashtags; Ranked from most to less used
Twitter API
Collection of Non-troll User's accounts using top hashtags
Use Botometer to determine score
Collection of Non-troll User's tweets (20 Tweets Per User)

| Position | Hashtag |
|---|---|
| 1 | #Ferguson |
| 2 | #IceBucketChallenge |
| 3 | #ferguson |
| 4 | #RIPRobinWilliams |
| 5 | #RIPMona |
| 6 | #ALSIceBucketChallenge |
| 7 | #bb16 |
| 8 | #supported_languages |
| 9 | #RussiainvadedUkraine |
| 10 | #RobinWilliams |
| 11 | #Happy21stBirthdayLiam |
| 12 | #الدولة_الإسلامية (Islamic State) |
| 13 | #xxx |
| 14 | #IfTheyGunnedMeDown |
| 15 | #votedemilovato |

@fordnation
52%

English-specific features
Content: 57%
Sentiment: 44%

Language-independent features
Friend: 36%
Network: 48%
Temporal: 18%
User: 69%

Bot score based on
All features: 52%
Language-independent: 53%
User profile language: en

Figure 1. An illustrated step-by-step presentation of our analysis' workflow.

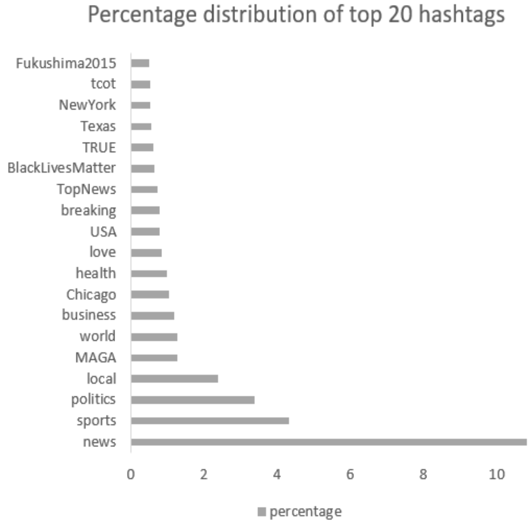


Figure 2. Top 20 hashtags, extracted from Twitter's public dataset on Russian Trolls.

Figure 3. Gephi graph showing the distribution of hashtags for Russian trolls and user data collected by us.

- Collected 20 tweets per genuine user for users with high Botometer scores and developed a comparative analysis based on common user behaviors.

The remainder of this paper is organized as follows. Section II will elaborate on our technical approach, Section III-C will discuss results and in Section IV we will highlight the relevance (related works) of our proposed analysis. Finally, we will finish with conclusions and future works in Section V.

## II. Data Collection & Analysis

The workflow of this project, which we previously described, is illustrated in Figure 1. In this section we will describe the same in detail.

### A. Hashtags Computation

In our analysis, we incorporated Twitter's public dataset, consisting of Russian Trolls supporting Donald Trump and genuine users, dataset collected by us using twitter api. To begin, we extracted the top $k$ hashtags from the Russian Troll data, where $k$ represents hashtags covering $90\%$ of the total tweets. These top hashtags included, ”news”, ”sports”, ”politics” and more, as shown in Figure 2, accompanied with percentage of tweets each hashtag covers. Such hashtags were extracted from Russian troll data using a Python script ($hashtags.py$), included on our Github [9] profile. In Figure 3 shows connection network for hashtags used by Russian trolls and users.

### B. Collecting User Details

Utilizing these hashtags, we implemented the (filter.json) filter [8], to extract real-time tweets from users who have applied the top hashtag(s) in their tweets. We collected real-time tweets for the hashtags generated in previous step to generate data for 1000 users. While doing this we found that some hashtags are very popular even now like news, politics, etc, but some hashtags are very specific and people were not tweeting them during our data collection eg. Fukushima2016, MAGA, etc.

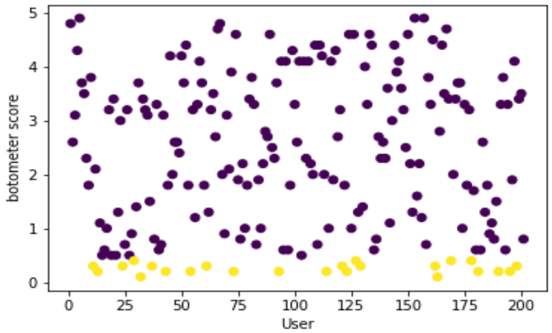


Figure 4. Botometer scores. Yellow users, were categorized as Trolls, and blue users were categorized a genuine users.

### C. Check for Bot

Since our collected tweets were from random users, without any prior knowledge of genuine versus troll users, we implemented Botometer for all the users before proceeding. The Botometer [7], returns a score estimating the likelihood of a user being genuine or not. We set our Botometer's threshold to $0.43$, as recommended in [10], and collected the top $1,000$ genuine users who demonstrated the best scores. In Figure 4, we depict a sample of these scores. As shown in the scatter plot, yellow users, towards the bottom, are ones whose Botometer scores were below the threshold. These users were categorized as trolls. On the contrary, the blue users, towards the top, produced significant scores. These users were categorized as genuine users.

### D. User Time-line Data Collection

Once we have user_id for probable genuine users we ran through their twitter time-line to collect 20 recent tweets generated by them. We used 20 as number to limit our data collection due to time and space constraints.

## III. Results

Through our analysis, we found several interesting discovers. As we were comparing between troll users data genuine

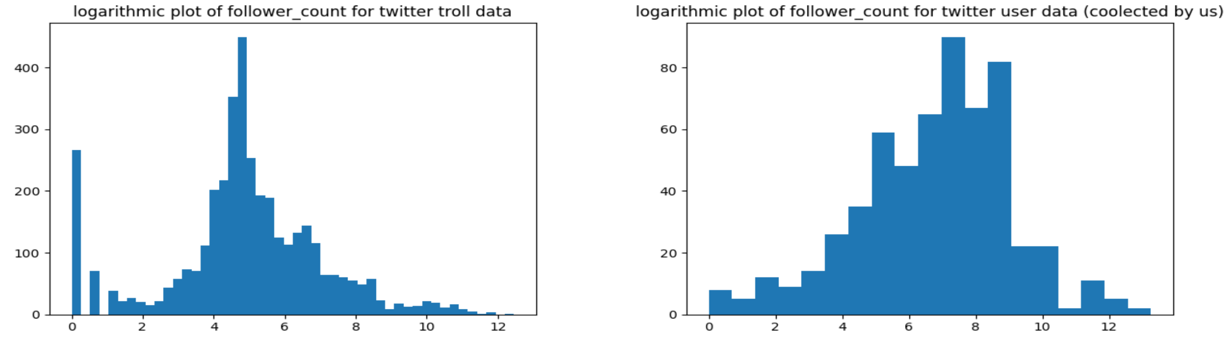


Figure 5. Logarithmic plot of follower count for Twitter Trolls (left) and Genuine Users (right).

users data, where we found that most of the column values are almost similar for both Russian trolls and genuine users. Again some data is available for genuine user but it is not available for troll user. Based on common data that we found in both users we first analyze number of follower count. As there is not significant difference between them so we tried with data distributions and statistical analysis.

### A. *Follower counts*

Below in the Figure 5 we can see histogram plot of follower count for genuine user and troll user. As we have some data imbalance and range of the follower count was very big so we tried those histogram plot with logarithmic value of follower count. We can see that follower count is dense for data range between 4 to 6 for troll user where as for genuine user its dense between 6 to 8. We can see that for genuine user histogram is right skewed but for troll data its a little bit left to the center.

### B. *User profile language and tweet language*

We also analyzed user profile language and user tweet language for both the users to find out statistical significant difference. From the Figure 6, we can see histogram plot of this analysis. From the figure can say that mostly used language for both profile language and tweet language is English, and number of user is almost similar for both troll and genuine category. But in case of Russian language number of tweets and user profile is much higher for troll user then genuine users. We can also see that troll users are limited to few languages, like Russian(ru), english(en), spanish(sp) etc., where as for genuine users user profile language and tweet language is available in different language means its not limited to certain languages. The most popular language for user accounts is English making more than 60% population, followed by Russian language for troll data. Also, for tweet language English seems to be most preferred language in general by both groups compounding to 65% tweets for normal users and around 50% for Russian trolls, followed by Russian language covering more than 40% of tweets by trolls.

### C. *User description length and tweet length*

Having intuition that troll users will not give much emphasis on profile description we tried to dig down user profile description length and tweets length. We got some significant difference in the result that can be seen in Figure 7. We can see that troll user used to have much less profile description on the other hand genuine user have pretty long description. Meanwhile for the tweet text length the scenario is different. For tweet text troll users tend to post large tweet then genuine user.

## IV. Discussion

Social media networks are cultural stables in modern society, transforming ways in which information is delivered and received, effective methods of communication and ”normal” human interactions. Users of popular social networks (i.e., Instagram, Twitter, etc.) typically engage in sharing information and opinions on a large scale of different topics. Accompanied by these users, which we refer to as genuine users, are spam accounts associated with bot or troll users. These spam accounts, typically but not always, frame their account usability (i.e., profile image, types of tweets, engagement with other accounts) in a way where they can be easily identified and removed from these sites.

Data and computer scientists have been conducting a tremendous amount of research, collectively, developing new methods to capture and understand these spam accounts. In, [1] authors record and analyze behaviors associated with Russian and Iranian trolls on twitter, using the same dataset were using, through a span of several years. Through analysis, researchers concluded that trolls exhibit constantly evolving behaviors, making it difficult to develop an effective, long lasting detection algorithm to detect them. [4] applies a statistical analysis on the MacronLeaks disinformation campaign, which affected the outcome of the 2017 French Presidential Election. Researchers prove that a ”black market of reusable political disinformation bots” exist. And such bots, according to their analysis, were used in the 2016 U.S. Presidential Election and reused in the French election.

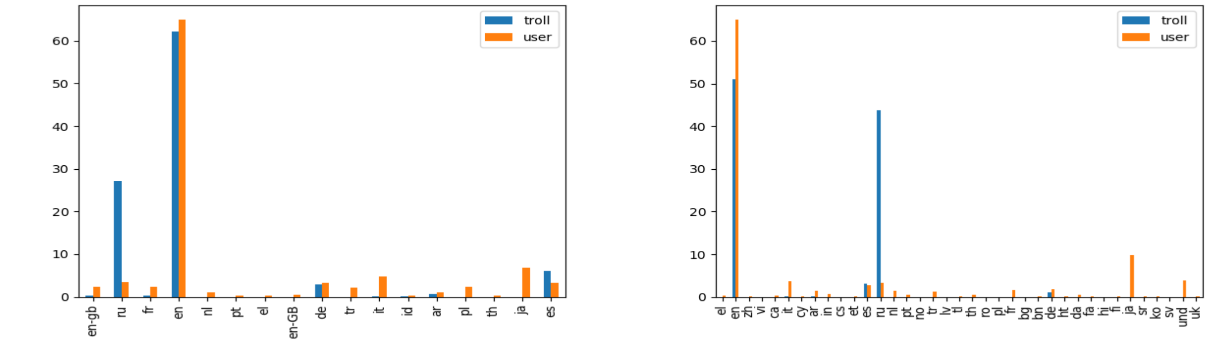


Figure 6. Distribution of hashtags separated by languages.

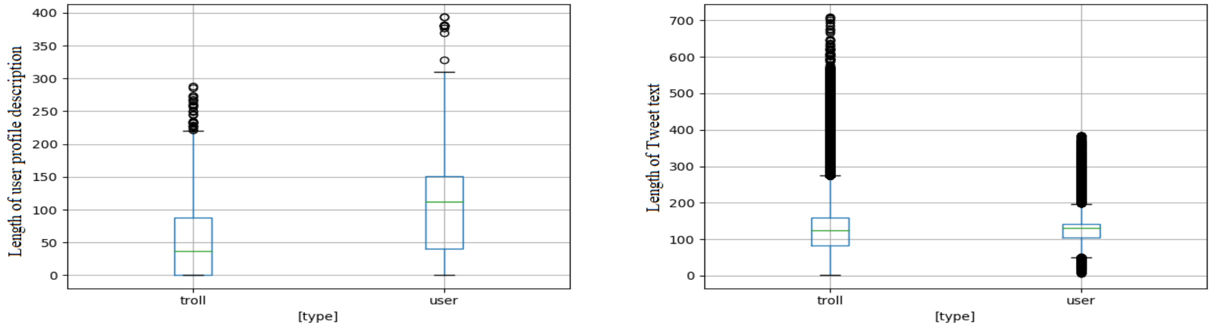


Figure 7. Length of user profile description (left) and length of tweet text (right).

Similarly to the U.S. election, the bot groups were clear about their political stance.

Researchers in [2], employ a novel, supervised machine learning classifier capable of identifying accounts belonging to Twitter Bots and determining the magnitude in which theyre involved with online discussion. They implemented several features (i.e. followers-to-friends ratio and sentiment expression) and achieved a 2.25% miss-classification rate. Furthermore, [5] investigates the moods and discussions had on social networks, which can affect the troll behaviors. Researchers conclude that anyone, indeed, can be considered a troll, considering the state of their mood or the activities they engage with online. Finally, [6] considers several layers of behaviors attributed to online antisocial behaviors. First, they consider antisocial users who are ostracized from certain online communities. Next, they monitor how the behaviors of the blocked user changes over time. Lastly, the community's response to the user is taken into consideration.

## V. Conclusion and Future Works

In conclusion, we developed an analysis, which compares the behaviors of Russian Trolls and Genuine Users, through their commonly used hashtags. Due to time constraints, we were not able to produce the magnitude of results that we desired. However, we have previously described our results in Section III-C. During the course of our project, we encountered several challenges. Additionally, we faced challenges applying Twitter's Time-line Data Extraction API, since many of the Russian Trolls have been blocked from Twitter's website. Therefore, we were not able to calculate the Botometer scores for these users. Lastly, many Russian Troll users tweeted and used hashtags in both English and Russian languages, simultaneously. This issue created some difficulty when extracting the top hashtags.

In the future, we plan to develop a machine learning classifier component enough to automatically differentiate between genuine and troll users. Our immediate approach will be analyzing text of user profile data and tweets data to find out some potential significance. We will analyze parts of speech of tweets and urls to differentiate two classes of users. For the time constraint we could only collect few genuine users. We are planning to collect more users and more tweets to have a balance dataset. Through statistical analysis we couldnt find out any feature, which can give us most distinguishable value to identify classes, there is also multimedia data available for troll users data. In multimedia data it contains audio, video, images etc. We think this data will give us useful information to classify user properly. We'd also like to explore some different ways of collecting Twitter data, using Twitter's API features and use network centrality to measure influence of troll accounts.

## References

[1] S. Zannettou, T. Caultfield, W. Setzer, M. Sirivianos, G. Stringhini and J. Blackburn, ”Who Let The Trolls Out? Towards Understanding State-Sponsored Trolls”, *arXiv:1811.03130 [cs.SI]*, Nov. 2018.

[2] P.G. Efthimion, S. Payne and N. Proferes, ”Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots”, *SMU Data Science Review*, VOL.1, NO. 2, Art. 5, 2018.

[3] Twitter, ”Election Integrity - Twitters focus is on a healthy public conversation”, 2018.

[4] E. Ferrara, ”Disinformation and social bot operations in the run up the 2017 French Presidential Election”, *arXiv:1707.00086 [cs.SI]*, July 2017, DOI: 10.5210/fm.v22i8.8005.

[5] J. Cheng, M. Bernstein, C. Danescu-Niculescu-Mizil and J. Leskovec, ”Anyone can become a troll: causing of trolling behavior in online discussions”, *arXiv:1702.01119v1 [cs.SI]*, Feb. 2017.

[6] J. Cheng, C. Danescu-Niculescu-Mizil and J. Leskovec, ”Antisocial Behavior in Online Discussion Communities”, *arXiv:1504.00680v2 [cs.SI]*, May 2016.

[7] Botometer, Available Online: https://botometer.iuni.iu.edu/#!/

[8] Developer Twitter, Available Online: https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html

[9] https://github.com/bharti26/Identifying-twitter-trolls

[10] J. Gramlich, ”QA: How Pew Research Center identified bots on Twitter”, *Pew Research Center*, Apr. 2018.