Twitter bots and COVID-19 Conspiracy Theories

Project Report for Information Security Course

Payam Tavakoli
payta322@student.liu.se
Linköping University, Sweden

August Carlson
augca570@student.liu.se
Linköping University, Sweden

Supervisor: Alireza Mohammadinodooshan
alireza.mohammadinodooshan@liu.se
Linköping University, Sweden

***Abstract*—The discussions on Twitter about COVID-19 have been ongoing since the start of the pandemic, in these discussions several conspiracy theories are being spread. In this report we have investigated how users create and interact with tweets concerning COVID-19 conspiracy theories on Twitter based on their likelihood of being bots. Over a two week period over 3500 tweets and 15 000 users have been gathered and analyzed, and our results suggests that accounts that are more likely to be bots interact more each other and accounts that are more likely to be human interact more with each other.**

I. Introduction

Social bots are becoming an increasingly serious security threat on social media. Between 9% and 15% of Twitter users are estimated to be bots [1]. On the other hand, the pandemic of COVID-19 has been accompanied by an outpouring of misinformation on social media (e.g., the 5G wireless network conspiracy theory). Sweden's Civil Contingencies Agency (MSB) has also paid considerable attention to this subject, owing to the security implications for the society [2].

Research on COVID-19 conspiracy theories have previously been conducted, see section V. This previous research does however not focus on user interactions on the tweets which will be done in this research paper and is the main contribution of this paper.

In this project, we will examine tweets about COVID-19 conspiracy theories made by bots. Research will be conducted to determine how humans as well as other bots interact with tweets made by bots about conspiracy theories within COVID-19. We will analyze the extent to which these bots interact with each other to contribute to increased spread and visibility of conspiracy theories about COVID-19.

A. Research Questions

1) To what extent does humans interact with the bots tweets spreading conspiracy theories?
2) To what extent does other bots interact with the bots tweets spreading conspiracy theories?

II. Methods and Data

A. Tools

1) Official Twitter API: The Official Twitter API [3] allows Twitter accounts with developer status to perform programmatic operations on Twitter users and tweets. For this project we will use it to retrieve tweets and their authors as well as users that have interacted with these tweets.

We chose to use the Official Twitter API over an unofficial Twitter API like Tweepy because of the excellent documentation and straightforwardness of the Official Twitter API.

2) Botometer: The project needed a way to analyze if a user is a bot or not. Writing a bot analysis tool was out of scope, so we needed a free tool that is able to do this part of the project for us. We chose Botometer for several reasons, not only was it recommended to us by the project supervisor, but it is also well known on the internet, has been used since 2014 by several researchers and is well maintained by the developers still to this day.

Botometer [4] is a tool that analyzes a Twitter user and returns a probability between 0-1 of the user being a bot. Botometer works by extracting over a thousand features from the given Twitter user account, these are then used by various machine learning models to compute a classification scores.

When using Botometer to query a user the API will return the following four major blocks of data, **CAP**, **display_scores**, **raw_scores** and **user**, see figure 1. **CAP** stands for Complete Automation Probability, which is the main score returned by Botometer which gives a probability that the user is a bot based on the data in raw_scores. Inside **CAP** there is a score for English and one for universal. **Raw_scores** are different scores in different categories, see figure 2. For example **fake_follower** inside **raw_scores** returns the probability that this account is made to increase other accounts follower count. **Display_scores** are simply **raw_scores** multiplied by a factor of 5. User contains metadata for the account.

The score we will be focusing on is **CAP**, and depending on if the **majority_lang** field contains the language code for English(”en”) we will choose the English score, otherwise we will chose the universal score. This score will be refered to as bot score.

B. Data collection

1) Keywords: We first selected the conspiracy theory to analyze, we chose the conspiracy that the COVID-19 pandemic was planned or fake. We targeted tweets regarding this with the keywords **plandemic, convid, #plandemic, #convid**. When deciding on a conspiracy theory we stared by looking up the most popular conspiracy theories related to COVID-19 and related hashtags. The two most prevalent

```
{
    "cap": {
        "english": 0.8018818614025648,
        "universal": 0.5557322218336633
    },
    "display_scores": {
        "english": {…
        },
        "universal": {…
        }
    },
    "raw_scores": {
        "english": {…
        },
        "universal": {…
        }
    },
    "user": {
        "majority_lang": "en",
        "user_data": {
            "id_str": "11330",
            "screen_name": "test_screen_name"
        }
    }
}
```


Fig. 1. Botometer response.


```
"english": {
    "astroturf": 0.0,
    "fake_follower": 4.1,
    "financial": 1.5,
    "other": 4.7,
    "overall": 4.7,
    "self_declared": 3.2,
    "spammer": 2.8
},
"universal": {
    "astroturf": 0.3,
    "fake_follower": 3.2,
    "financial": 1.6,
    "other": 3.8,
    "overall": 3.8,
    "self_declared": 3.7,
    "spammer": 2.3
}
```


Fig. 2. Specific contents of display_scores and raw_scores.

conspiracy theories was about COVID-19 being planned and that the COVID-19 vaccine is harmful. We then tried to gather tweets for both of the conspiracy theories and chose the one about the pandemic being planned or fake due to it returning a much larger amount of tweets. We prioritized the amount of analyzable tweets to maximize the scope of the analysis of the conspiracy theory.

2) Fetching tweets: With the keywords specified, we used the Twitter API to retrieve original tweets (not replies or retweets) that included one or more of these keywords along with their authors and compiled them to a file. Then we looked at each of the tweets in the compiled file and for each one we fetched the users that had liked, retweeted or quoted(a retweet where the retweeter adds additional text) the original tweet and compiled that into another file.

We then had a file that contained the authors of a conspiracy theory tweet along with all of the users that had interacted with that tweet. With this data at hand we ran each of the entries in the file into the Botometer tool, which returned us a likelihood of the author and the interactors being a bot.

3) Timespan: The gathering of tweets was done over a two week period. The reason for not doing this during a larger timespan was the time limitations on the project.

III. Results

The following figures were drawn from the collected data. The results in each graph have been divided into five *probability buckets*, where each bucket represents 20%.

Figure 3 shows the total amount of interactions recorded. The x-axis represents the probability that the author of a tweet is a bot, and the y-axis represents the probability that the interactor to the tweet is a bot. However these raw numbers can be a bit misleading, so we generated figure 4 and figure 5 to data as clearly as possible. The reason for the numbers being misleading is the large difference of total tweets created by the users based on their bot scores as visualized in figure 7.

Figure 4 shows how many percent of interactions on the authors come from a certain bot score. For example we see that when the probability of the author being a bot is between 0.8 to 1, more than half of interactions come from interactors that have a probability between 0.6 to 0.8. For example, a misleading result from figure 3 is how interactors with a bot score between 0 to 0.2 interact with tweets made by a user with a bot score between 0 to 0.2 (3 entries) and 0.8 to 1 (5 entries). The total numbers show us that the interactors with as low bot score are more interactive with a user with a high bot score but when looking at figure 4 we see that the bucket with three entries cover a much larger percentage than the bucket with 5 entries (2.01% to 0.19%).

Figure 5 shows the same thing as figure 4, but flipped in such a way that it shows how many percent of the interactors interact with a tweet written by an author of a certain bot score. For example, a misleading result from figure 3 is how interactors with a bot score between 0 to 0.2 (3 entries) and 0.8 to 1 (20 entries) interact with tweets made by a user with a bot score between 0 to 0.2. The total numbers show us that the interactors with as high bot score are more interactive with a user with a low bot score but when looking at figure 5 we see that the bucket with three entries cover a much larger percentage than the bucket with 20 entries (2.31% to 0.7%).

For the tweets without interactions we represent them in figure 6. It shows how many interactionless tweets each probabilitybucket has.

We also extracted a histogram that shows the number of authors in a certain bot score with one or more interactions on their tweet. This is represented in Figure 7.

IV. Discussion

In our first heatmap seen in figure 3, as well as in our two histograms, seen in Figures 4, 5 we can see that users with a bot score between 0.6 and 0.8 are the most active, both in terms of creating new tweets and interacting with existing tweets. This data would suggest that users which are more likely to be bots are more active in spreading the chosen conspiracy

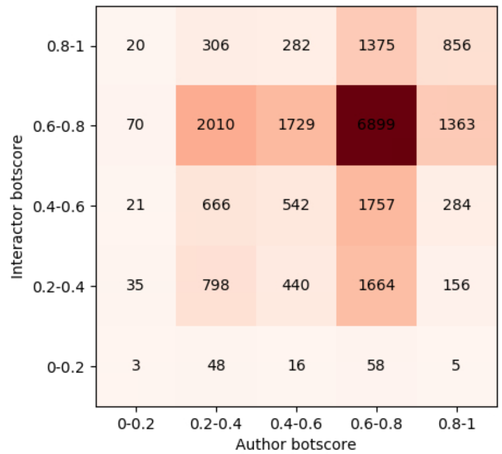

Fig. 3. Total number of interactions grouped by author and interactor bot score.

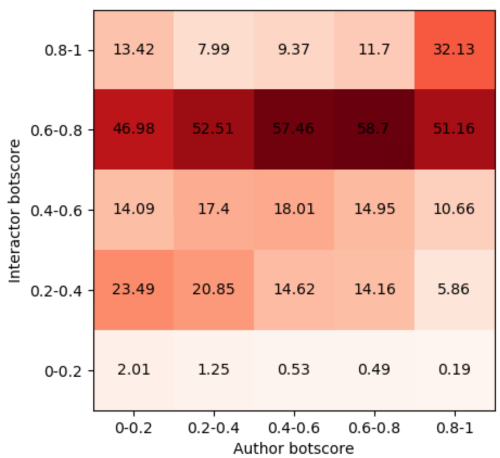

Fig. 4. Percentage of interactors based on their bot score to an author based on the authors bot score.

theory about COVID-19 being fake or planned on twitter than users which are less likely to be bots are. This assumes that users who get a bot score between 0.6 and 0.8 from Botometer are in fact bots.

Examining our two heatmaps based on percentages, figures 4 and 5, we can see patterns of the interactors by calculating their corresponding percentages, both in terms of the bot score of the original tweeter and the bot score of the interactors. From this data we can see that users generally tend to interact with tweets about our chosen conspiracy theory created by a user with a similar bot score to themselves. This means that tweeters that has a low bot score will get interactions from users that also has low bot scores, and tweeters with a high bot score will get interactions from users that also has high bot scores.

When examining our two histograms, figure 6 and 7, we can see some interesting patterns for all of the different bot score buckets. We can see that for all buckets about 50% of their tweets get at least one interaction. This shows that there is no

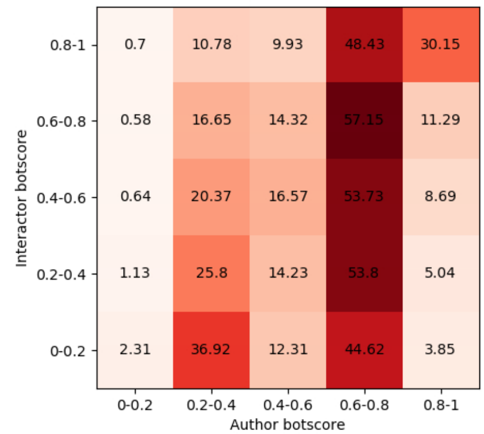

Fig. 5. Percentage of interactors based on their bot score to an author based on the authors bot score.

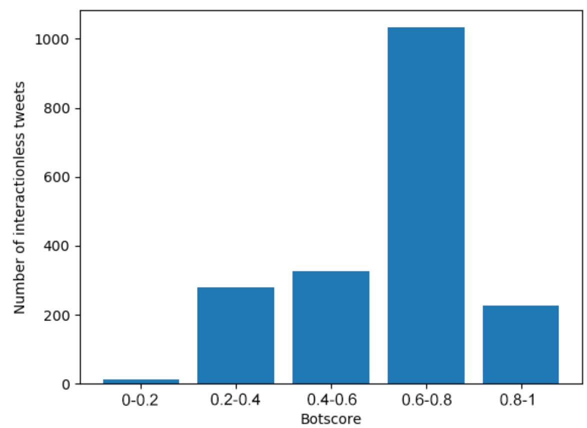

Fig. 6. Total number of tweets without an interaction grouped by authors bot score.

significant difference between interactions on tweets based on how likely their author is to be a bot. For example this means that is likely that bots don't make a conscious decision to look for unpopular tweets to interact with and promote for twitters algorithm.

V. Related Work

The main related work that have been used to inspire this project is "COVID-19 on Twitter: Bots, Conspiracies and Social media activism" by Emilio Ferrara [5], an article about how bots are used on Twitter to spread conspiracy theories about COVID-19. In this article tweets with keywords related to COVID-19 are collected using Twitter's search API. The authors of these tweets are then analyzed using Botometer API to determine the likelihood of them being bots. The data is then used to examine distribution of Botometer scores. The conclusions of Ferraras work is that they found early evidence of accounts with a high bot score promoting conspiracy theories about COVID-19.

We have similarly to this article used Twitter API to gather tweets related to COVID-19 as well as used Botometer API

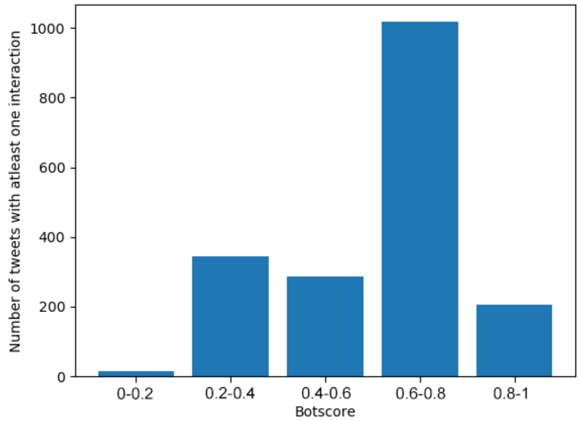

Fig. 7. Total number of tweets with one or more interactions grouped by authors bot score.

to analyze the likelihood of select users being bots. The main difference for our work is that we instead of only focusing on the authors of the tweets, we also focus on the interactors of our found tweets with the goal of finding the distribution of real users and bots who interact with tweets about conspiracy theories concerning COVID-19.

VI. Future Work

A. Implementing replies

Our initial plan was to also analyze replies to conspiracy theory tweets about COVID-19, after we began writing the functions for fetching tweets we realised the Twitter API does not directly support fetching replies to a given tweet. We didn't find a way to implement this given the projects time frame so we decided to exclude replies from our analysis.w

B. Additional Conspiracy Theories

One more conspiracy theory was planned to be analysed, we also wanted to analyse the anti-vaccine discussions on Twitter by using keywords like **vaccinedamage, vaxkills, antivax, vaccineinjury**. However once we realized how much the Twitter API and Botometer limit the number of tweets we are able to analyze we decided to stick with just one conspiracy theory.

C. Research Questions out of our scope

Two additional research questions were considered for this project but they were in the end not able to be researched and answered due to the time constraint of the project.

1) Is there a connection between a tweets age and the amount of bot versus human interactions on the tweet?
2) Is there a connection between a tweets popularity and the amount of bot versus human interactions on the tweet?

These research questions are interesting in relation to the two research questions handled in this report due to their ability to affect the conclusions drawn about said research questions. These new research questions would provide a wider perspective on how human and bot interaction on specific tweets varies based on different parameters. Therefore it would be an interesting way to improve this project to answer these research questions.

VII. Limitations

A. Dataset limitations

There were several factors that put a limit on the amount of tweets we were able to analyze.

1) Twitter API calls: The Twitter API only allowed us get tweets we could only retrieve tweets from the last 7 days from the time of our request. Conspiracy theories around COVID-19 were more popular a year ago when we were in the middle of the pandemic, there are less tweets talking about it now than before.

The Twitter API only allows for 300 calls per hour when it comes to retrieving users that have liked, retweeted or quoted a conspiracy theory tweet.

2) Botometer limit: Botometers free plan only allows for 500 calls per day, which put a hard cap on the amount of tweets we were able to analyze.

B. Replies

One of our goals was to include replies when analyzing the interactions on tweets, however since the Twitter API doesn't directly support this we decided to leave them out of our analysis since it would take time to implement a way to extract replies to a tweet.

C. Banned or removed users

In the time between that we fetched a user and that we analyze it through Botometer there is a chance that the user has removed the account or has gotten banned. This means that the data that user would have generated has been lost. We lost data on 245 users due to this which accounts for almost 1% of our dataset.

VIII. Conclusions

The conclusions we have drawn from the results is that the tweets about the chosen COVID-19 conspiracy theory that are made by accounts that are more likely to be bots are interacted with by accounts that are more likely to be bots. As well as the reverse, the tweets about the chosen COVID-19 conspiracy theory that are made by accounts that are less likely to be bots are interacted with by accounts that are less likely to be bots.

This means the answer to Research Question 1 is that users which are more likely to be humans do not interact extensively with tweets about the chosen conspiracy theory about COVID-19 made by users that are more likely to be bots. The answer to Research Question 2 is that users that are more likely to be bots do extensively interact with tweets about the chosen conspiracy theory about COVID-19 made by other users that are also more likely to be bots.

References

[1] O. Vanur et al., “Online Human-Bot Interactions: Detection, Estimation, and Characterization”, https://ojs.aaai.org/index.php/ICWSM/article/view/14871

[2] A. Önnerfors, ”Conspiracy theories and COVID-19: The mechanisms behind a rapidly growing societal challenge”, https://www.msb.se/siteassets/dokument/publikationer/english-publications/conspiracy-theories-and-covid-19.pdf, Accessed March 2022

[3] Official Twitter API documentation https://developer.twitter.com/en/docs

[4] Botometer by OSoMe https://botometer.osome.iu.edu/

[5] E. Ferrara, “What types of COVID-19 conspiracies are populated by Twitter bots? ”, https://firstmonday.org/article/view/10633/9548, Accessed March 2022