Ethan Handelman

Professor Pierce

CSC 369

7 March 2025

"Final Data Analysis: Ukraine War Disinformation Campaigns on Twitter"

This data analysis is performed on a set of posts scraped from Twitter from February 2022-June 2023 that were filtered to contain only topics regarding the war in Ukraine. The purpose of this analysis is to investigate the hypothesis:

There is credible evidence on Twitter of coordinated disinformation campaigns linked to the Ukraine war, and the scale and organizational complexity of these campaigns strongly suggest involvement by state actors such as the Kremlin.

Average user context:

For some context about users, let's look at some stats about the average account within the dataset (See Table 1). The average tweet count seems extremely high given that this would mean the average user is tweeting >130 times per day. This indicates that there must be some outlier accounts tweeting far more often than others.

Statistic (mean)	Value
Follower Count	15648
Following Count	1702
Acct. Create Date	11/25/2016
Tweet Count	67223
Retweet Count	584

Table 1: Average account statistics
(calculated in mean_follow.py)

Investigating deeper, Table 2 shows the percentiles for tweet count (users > 100 tweets). The numbers up to the 95th percentile appear to be relatively linear until they absolutely spike in the 99.9th percentile, confirming that there are massive outliers within the dataset. Upon further analysis, there are 119 accounts with >10k tweets, and 6 with >50k.

%ile	Tweets
25	134
50	195
75	359
90	741
95	1227
99	3354
99.9	12221

Table 2: Tweet count percentiles for users with >100 tweets (calculated in tweet_count_percentiles.py)

Tweet frequency context:

One of the best ways to detect bot accounts is to examine the frequency of posts for any irregular patterns such as posting way too much or at predetermined intervals. Figure 1 shows the average time between tweets for users with >100 tweets. The strong right skew shows the majority tweeted more than once every two days.
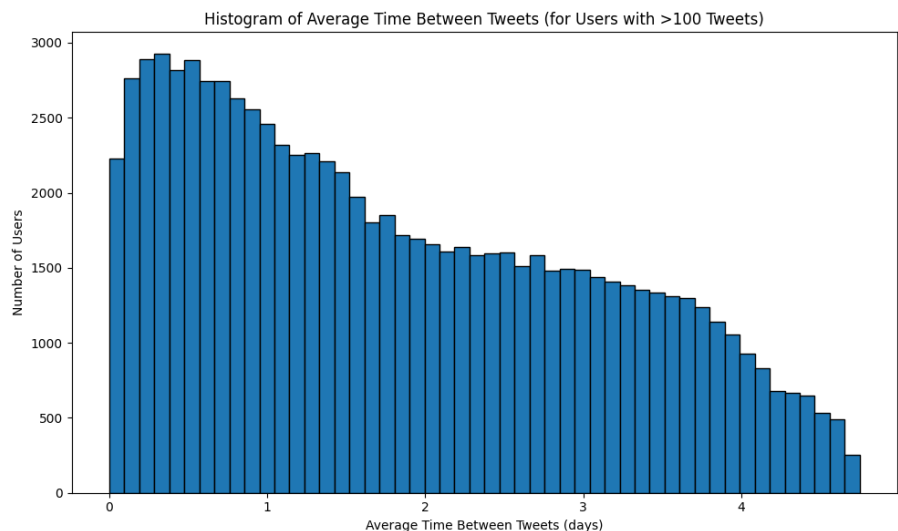
Figure 2: Histogram of user's average time between tweeting about the war (generated in time_between_tweets_histo.py)

Let's take a more in-depth look at the outlier accounts that seem to be tweeting at near inhuman rates (>50k tweets in the dataset):
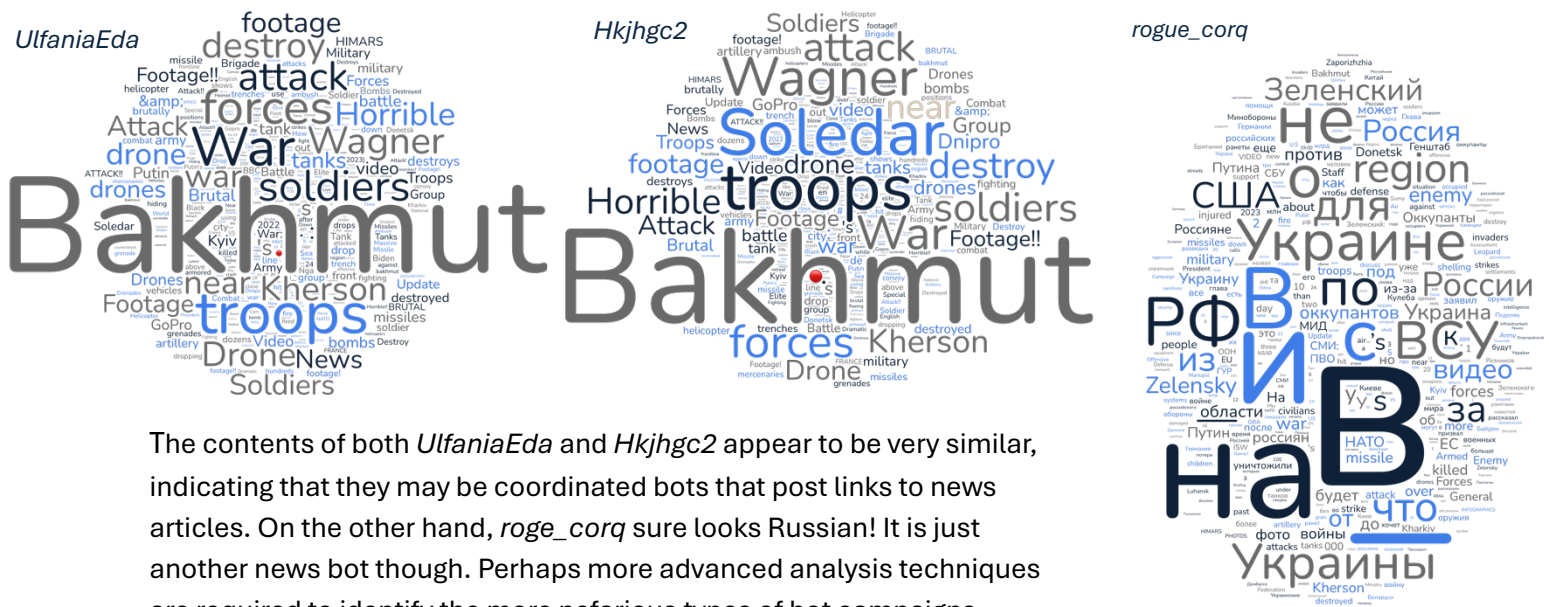
Besides the fact that the top posting account is named *F**kPutinBot* (probably not a Russian asset), there are several numbers that pop out. The median interval for *kanadianbest* is only 6 seconds (looking online this is a marketing bot for an ecommerce store).

User ID	Tweet Count	Mean Interval (s)	Median Interval (s)	Standard Deviation (s)
UlfaniaEda	53940	294.56	120.00	5300.06
Hkjhgc2	61482	321.46	94.00	20174.94
rogue_corq	62067	661.33	359.00	10062.63
kanadianbest	83594	426.81	6.00	24121.04
F*ckPutinBot	554244	72.23	60.00	2458.76

Table 3: Top five accounts in terms of tweet count and measures of the intervals between their tweets (calculated in user_time_stats.py)

Let's take a different approach and visualize the post date distribution of the remaining suspicious accounts (See Figure 2). Evidently, the account that was the most consistent was *roge_corq*, with both *UlfaniaEda* and *Hkjhgc2* posting almost exclusively between November 2022 and May 2023. This behavior could certainly align with an automated account, but what are they posting?

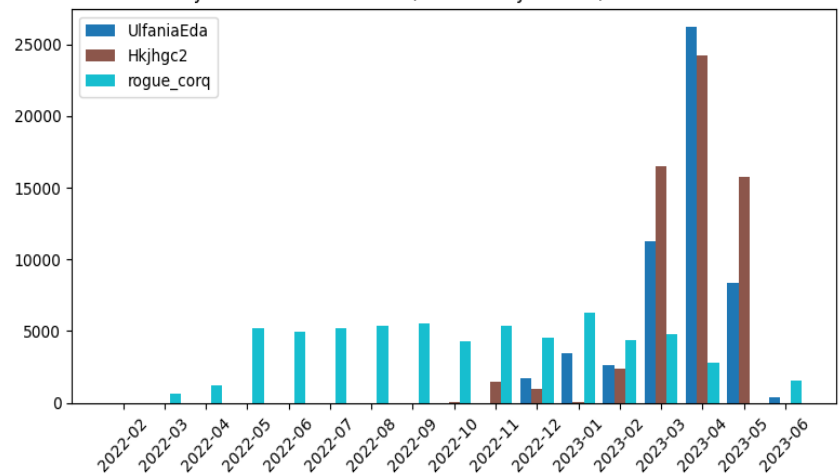*Figure 2: Histogram of the distribution of tweet dates for suspicious accounts with >50k posts (generated in* tweet_distrib_histo.py)

Tweet Content:

To examine the content of these suspicious accounts, we can extract all of their tweets (removing hashtags and common words like "Russian, "Ukrainian", etc.) and create word clouds:

UlfaniaEda

Hkjhgc2

rogue_corq

The contents of both *UlfaniaEda* and *Hkjhgc2* appear to be very similar, indicating that they may be coordinated bots that post links to news articles. On the other hand, *roge_corq* sure looks Russian! It is just another news bot though. Perhaps more advanced analysis techniques are required to identify the more nefarious types of bot campaigns.

Figures 3, 4, 5: Word cloud of tweets for accounts *(generated using* simplewordcloud.com, *extracted with* get_content.py)