

Ukraine-Russia on Twitter: Network Analysis of Influencers and Bots

By:

James Anderson

Parthiv Borgohain

Rishabh Tiwari

Saurabh Arora

Shubhada Kapre



Agenda

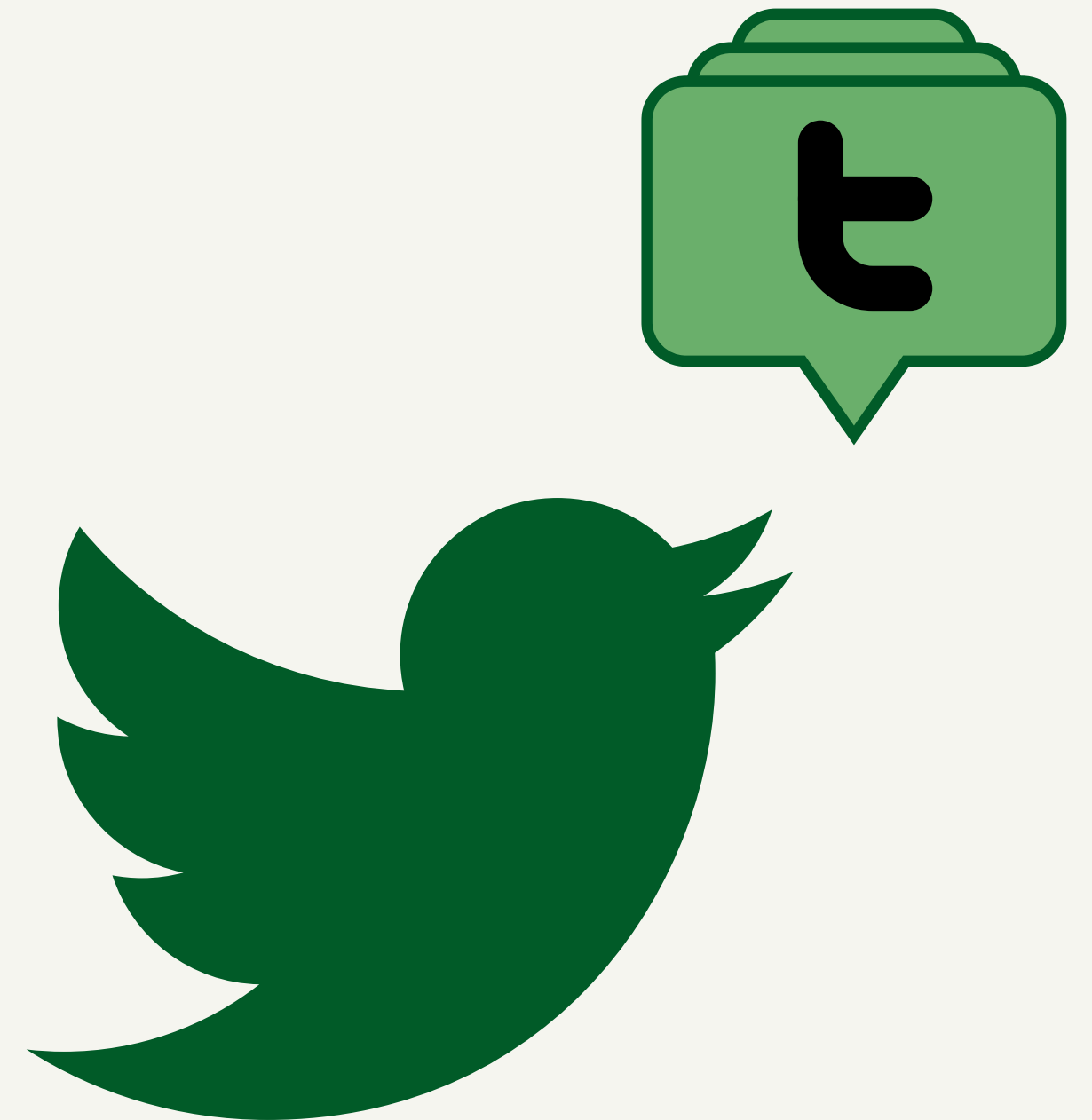
- Data Collection and Description
- Bot Detection Model
- Networks
- Topic Modelling
- Insights
- Conclusion and Recommendations

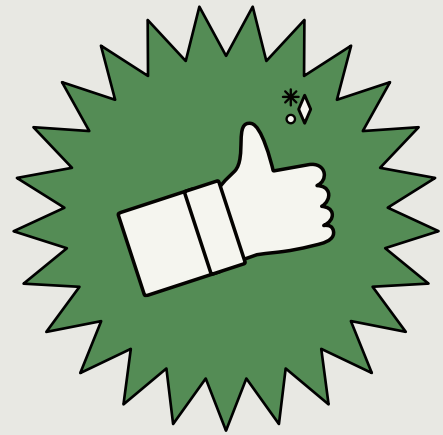


Data Collection

Scraped 13k tweets with the following 9 hashtags:

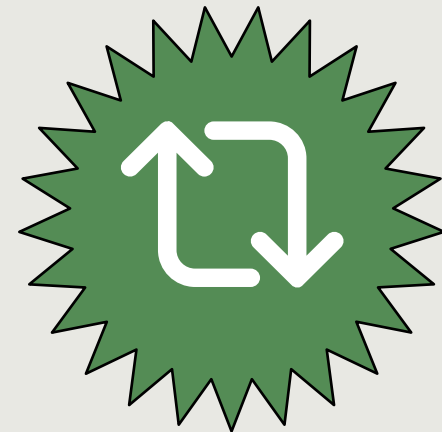
- # "#Russia"
- # "#Ukraine"
- # "#RussianUkrainianWar"
- # "#UkraineWar"
- # "#UkraineRussianWar"
- # "#RussiansANaziState"
- # "#SlavaUkraïni"
- # "#UkraineRussiaWar", "#RussiaUkraineWar"





Likes

0.36 Likes per Tweet



Retweets

88 Retweets per User



Followers

47K followers per User



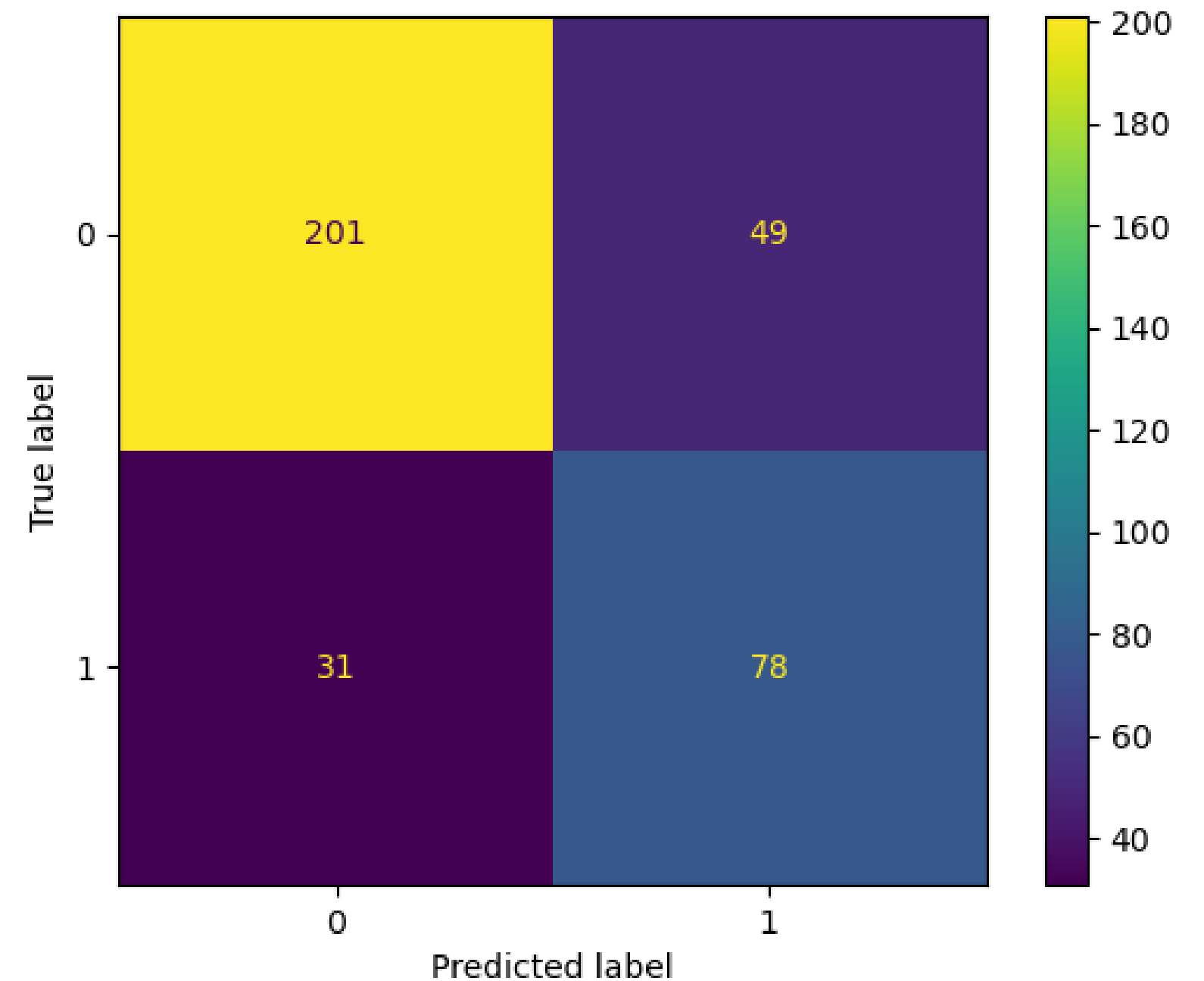
Following

1.7k following per User

Bot Detection

- Acquired a dataset from Kaggle containing 2 columns: "id" and a mechanical turk-based assessment of bot/human classification.
- Leveraged the Tweepy library to scrape information on approximately 2000 Twitter accounts.
- Decision variables include following, followers, and source of account.
- Developed and trained a random forest classification model with max_depth=2, random_state= 0 to distinguish between bots and humans.
- Model returned 77.7% accuracy on the test set.
- Applied the trained bot model to analyze the scraped Russia-Ukraine network users' information and accurately classify each user as either bot or human.

Confusion Matrix for RF Classifier is-



Influencers

- One subject on the tweet
- Pathos often referenced

"RT @igorsushko: #Russia: Bread line ðŸ“¸ of Ukrainians who weren't imprisoned or executed in fascist-occupied #Mariupol.
Welcome to Russkiy Mâ€™"

"RT @nexta_tv: The #US President to visit #Poland on February 20-22, White House says
Earlier, Polish authorities said they had sent Joe Biâ€™"

Bots

- Often includes links
- Prolific use of hastags & unique mention
- Multiple subjects in the tweet

@RaufHabibov @sandumaiamd @EmmanuelMacron @NATO Macron's favorite courtesan #Armenia
is #Iran regime's friend, whoâ€™ <https://t.co/2Zv98iloF3>


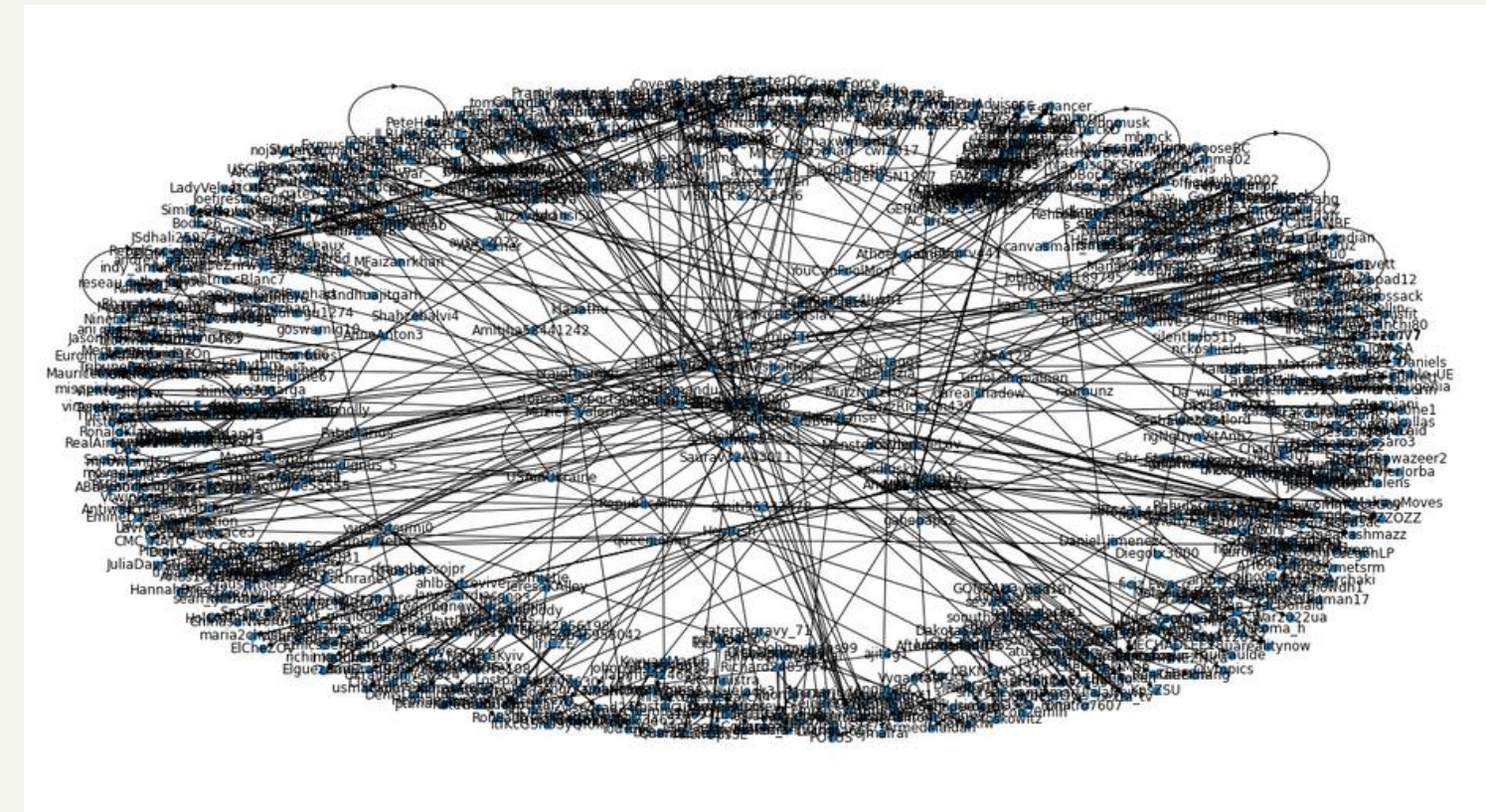
RT @GeorgeSirdar: Putin says 95% of Global Terrorism is done by the CIA

#Russia #Iran #India #Putin #USA #Biden #Germany #NATO #Ukraine
#â€™

18.1% of tweets were produced by accounts classified as bots.

17.2% of users in the network were classified as bots.

Our original data set had 2.3 hashtags per tweet. The dataset filtered to contain bots only had 2.9 hashtags per tweet. No considerable difference in the number of mentions, but the type of mentions used by bots was peculiar.



Topic Modelling

Topic 1

News

“

words near topic: https, co,
watch, pm , abpnews)

Topic 2

General War & Hashtags

“

(ukrainewar, ukrainewillwin,
ukrainerussianwar,
ukrainerussiawar, slavaukraini,
russianukrainianwar, war, forces)

Topic 3

Russian Embassy

“

Some topics that we got were as
follows: embassyofrussia,
mfa_russia, russian_consulate,
rusmission_eu, russianembassy

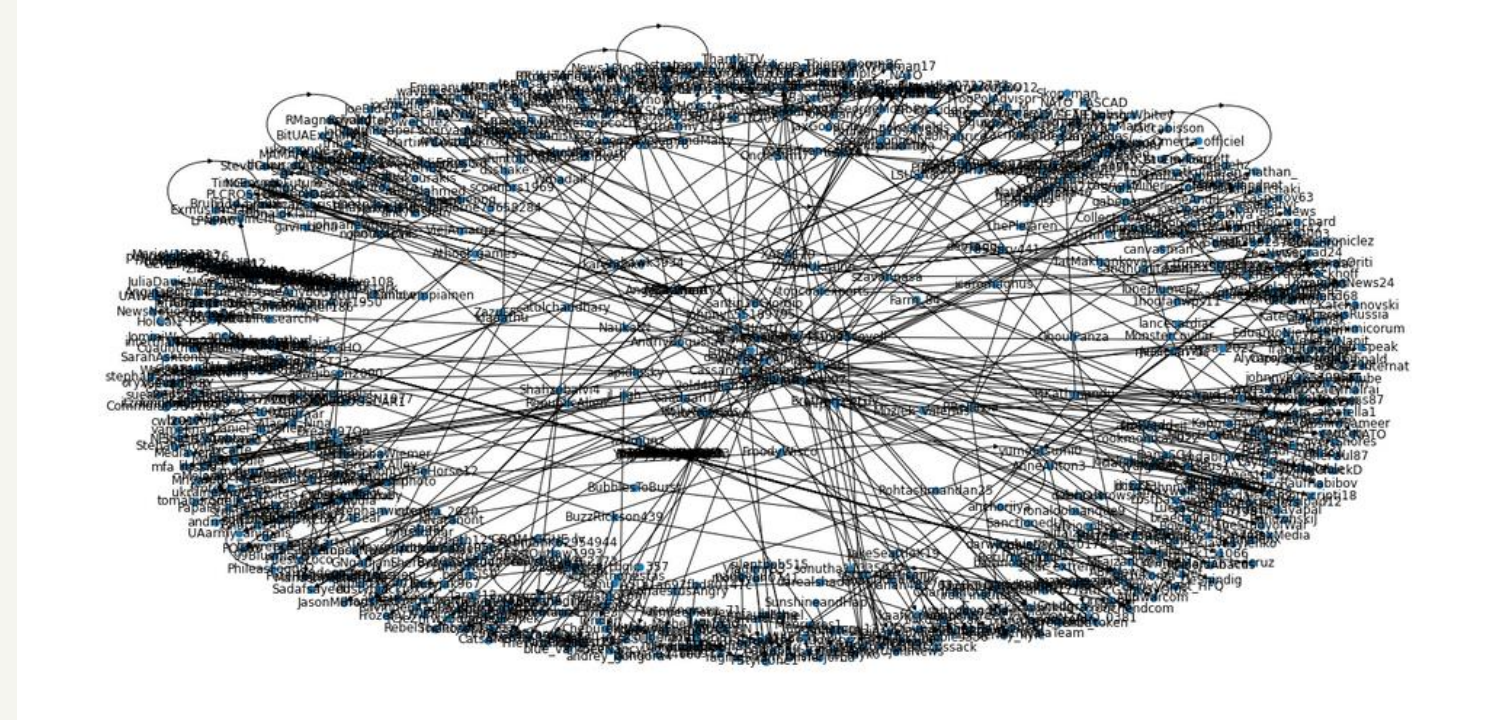
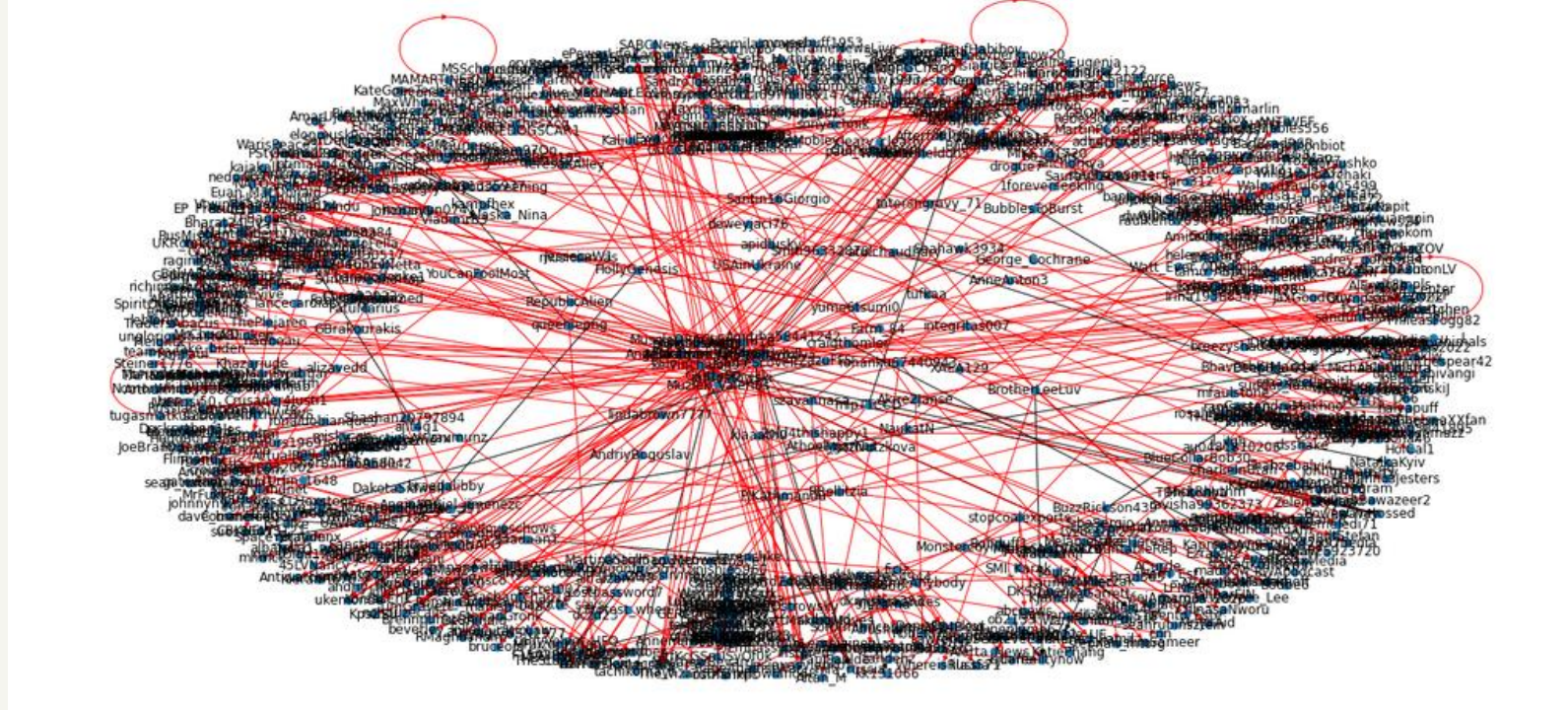
Topic Modeling for bots only did not return topic 3... Are bots not focused on "breaking news"?

Top Influencers

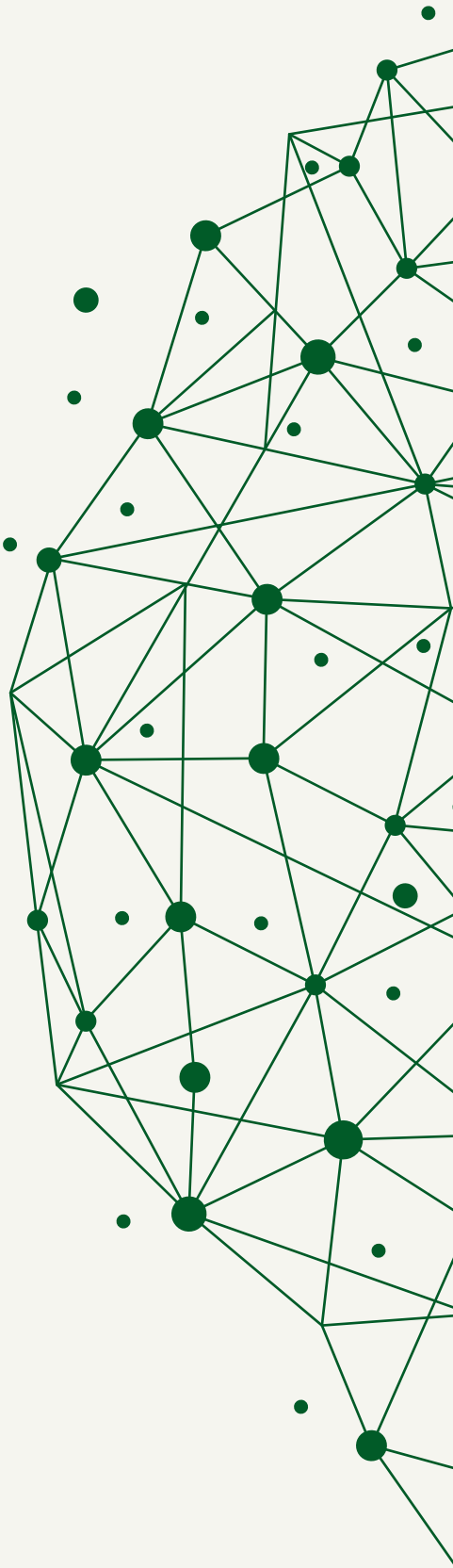
	Followers	Following	Retweet Count	is_retweet	is_mention
Id					
ABPNews	13254113.0	37.0	17.000000	0.000000	0.000000
ZeeNewsEnglish	5652959.0	26.0	1.000000	0.000000	0.000000
News18India	3094996.0	90.0	2.000000	0.000000	1.000000
AmarUjalaNews	2085040.0	86.0	1.000000	0.000000	0.000000
tv9gujarati	912951.0	64.0	2.000000	0.000000	0.000000
TheStudyofWar	581790.0	2761.0	1.500000	0.000000	0.000000
TV9Bharatvarsh	230429.0	21.0	1.750000	0.000000	0.750000
anjali_speak	105789.0	420.0	2.000000	1.000000	0.000000
RajLaveena	89029.0	266.0	2.000000	1.000000	0.000000
janrostowski	57341.0	457.0	73.000000	1.000000	0.000000
HillBeverlyhill	34277.0	33173.0	56.000000	1.000000	0.000000
kk131066	31488.0	26137.0	2.000000	1.000000	0.000000
RebHarms	35262.0	3195.0	19.000000	1.000000	0.000000
ob2133	23916.0	19889.0	81.000000	1.000000	0.000000

- Based on our previous influence score model, we assigned a score to every user based on their attributes like no. of followers, retweets, network metrics etc. Then we tried to analyze the Top 15 influencers as per this score
- Out of the 15 accounts we analyzed, 6 were Indian news channels, 1 was a political think-tank, and the remaining accounts were journalists and politicians.
- After conducting a preliminary analysis, we found that none of these accounts appeared to be bots.
- Upon initial inspection, we did not find any evidence to suggest that any of these accounts originated from Russia or Ukraine.

Network with Influencers



In the image on the left, we can see the entire network. The edges associated with the top influencers are highlighted in red. However, by removing the top 30 most influential users according to our model, we obtain the network on the right. Notably, we observe that the network remains highly dense, suggesting that it is self-sustaining. As a result, it is unlikely that solely going after bot accounts will not significantly impact the overall network and the prevailing narrative

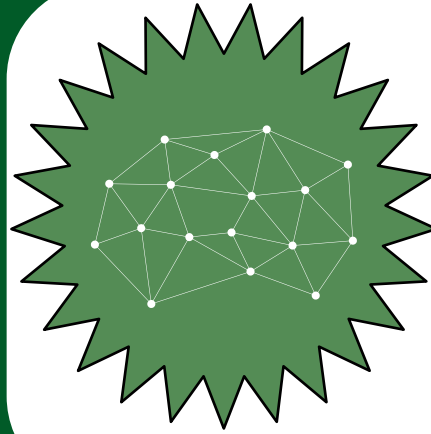




Conclusion

- 18.1% of tweets were produced by accounts classified as bots.
- Bots generally talk about the same subject matter as non-bots, however, they tend to not focus on “breaking news” (embassy example). Isolated incident
- May be able to add more parameters to the bot classifier. Add avg. number of mentions/hashtags used
- Topic Modeling showed a variety of related hashtags
- The Russia-Ukraine Network is self-sustaining. So, it is difficult to control the narrative by solely focusing on a few influencer accounts.
- Although a decent proportion of accounts seem to be bot accounts, they do not hold as much sway as the prevailing narrative regarding Russian bots seems to suggest
- Top influencers did not appear to originate from Russia or Ukraine

Recommendations



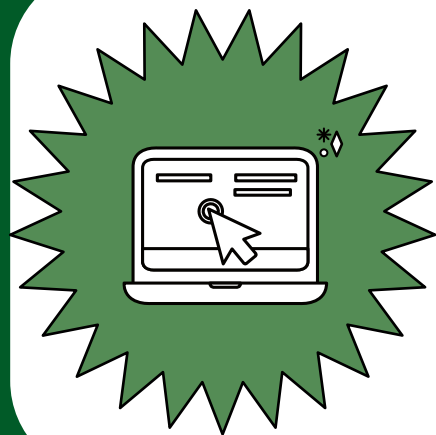
Self Sustaining Network

The network is self-sustaining. There is limited use to target influencers or bring more influencers into the network. Aside from there being the obvious ethical implications of the above.



Improve Bot Detection Software

Bot identification may be supplemented by topic modeling, named Entity Recognition, frequency of hashtags/mentions. Posing additional verification using a reCaptcha or other bot-prevention software test to minimizes false positives



Enhancing Search Engine

Use Topic Modeling to aggregate similar hashtags to give more comprehensive results.