CHARACTERIZING NETWORK PROPERTIES OF GAB USERS

Mentor: Binny Mathews

gab

A social network that champions free speech, individual liberty and the free flow of information online. All are welcome.

Join Gab

Email

Create Account

TOPICS COVERED

Problem Description

Dataset description

Analyzing the results

Findings

Future Works

PROBLEM STATEMENT

Objective: Finding the impact of posts of hateful users on other users over time.

As 45% of the data over gab.ai is images and videos, so only relying on text is not enough.

Lexicon will not capture all hateful users. So we rely on homophile to capture other hateful users using network properties.

Hence we are considering text and network properties to capture the behavior of users

Hypothesis : Do hateful users have high impact on the network?

EXTRACTING GAB.AI DATA

As mentioned by Brendan et. al. [1], we extracted data from gab.ai in a interval of 1 month and approximated the new followers of each of the user in that duration.

Extracted data for Oct 2016 to June 2018 (21 months)

Final Data (Cumulative Data)

Total Nodes in graph : 354947

Total edges in graph : 22761772

[1] We Know Who You Followed Last Summer: Inferring Social Link Creation Times In Twitter. Brendan Meeder, Brian Karrer, Amin Sayedi

FINDING HATEFUL POSTS

If a post contains:

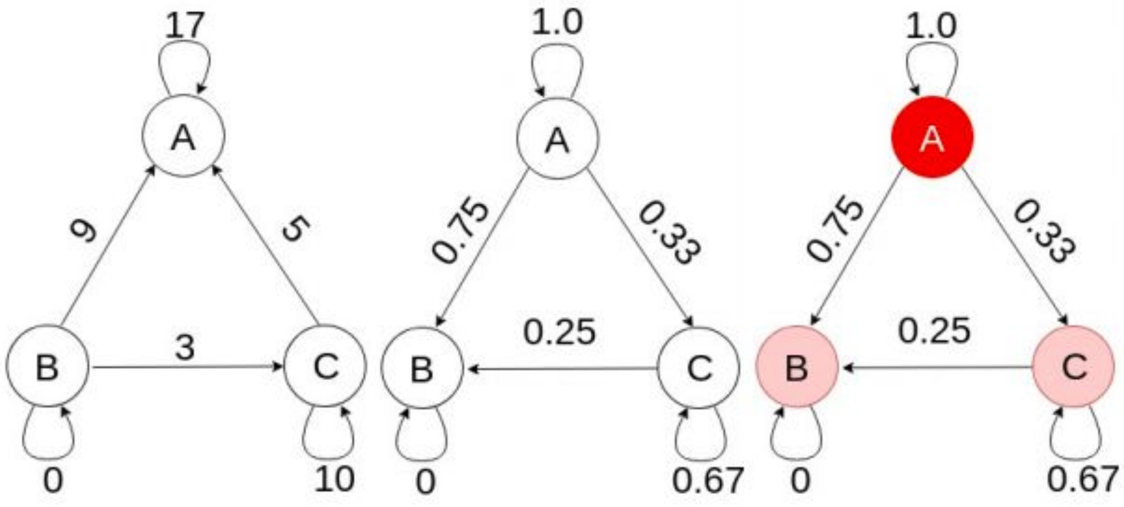
- atleast one hateful word from predefined lexicon[2] like “camel jockey”,”niglet” or
- Regex "\(\(\((.*?)\)\)\)”

then we considered the post as hateful.

[2] Spread of hate speech in online social media Binny Mathew, Ritam Dutt, Pawan Goyal, Animesh Mukherjee

FINDING HATEFUL AND NON HATEFUL USERS

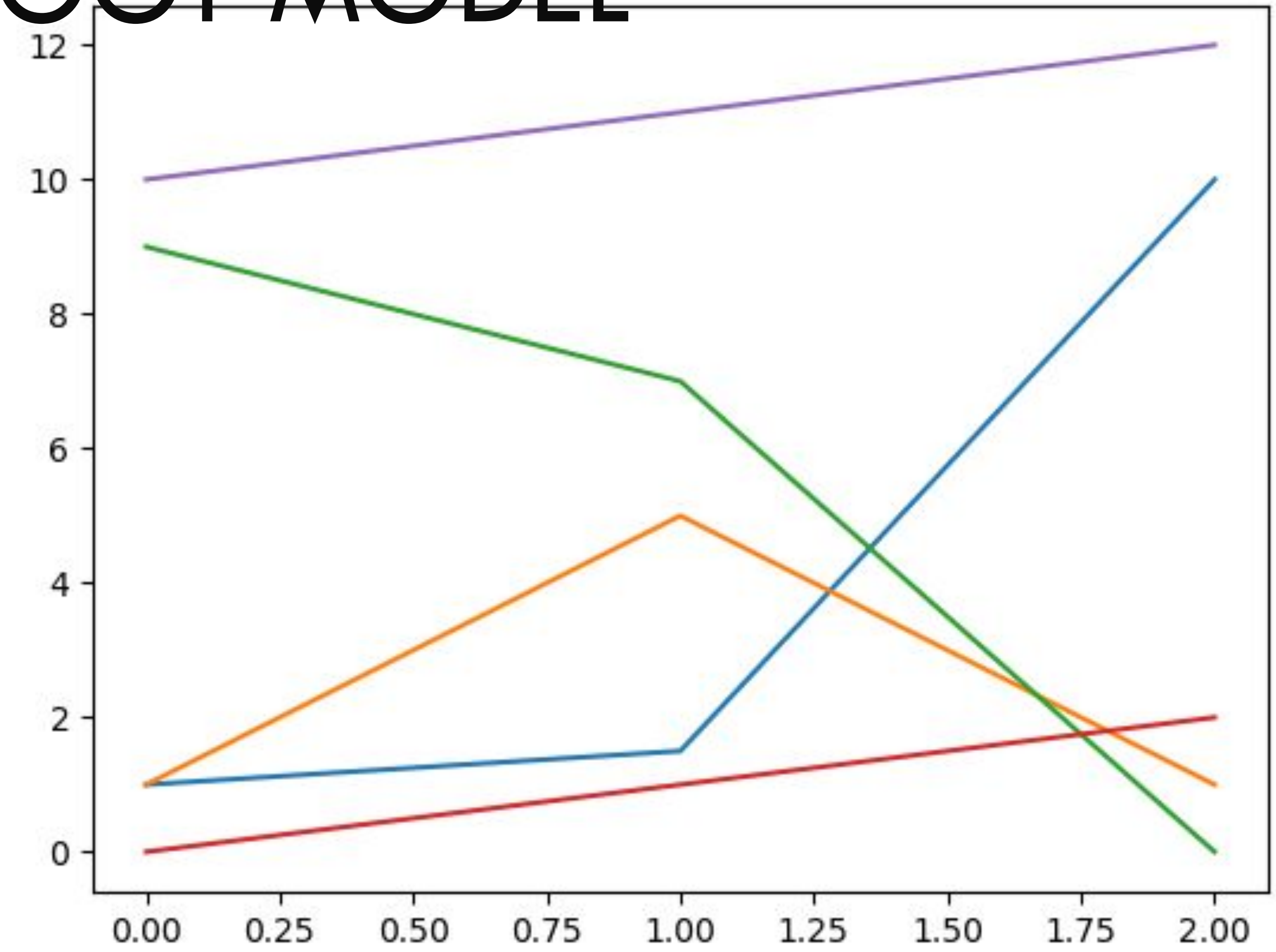
Using Degroot's model for information diffusion[2], we created a diffusion graph.
And ran degroot's for 5 epochs
All the nodes with final Degroot's value <0.05 were tagged as non hateful and value > 0.5 were tagged as hateful.



(a) Repost Network (b) Belief Network (c) Belief diffusion

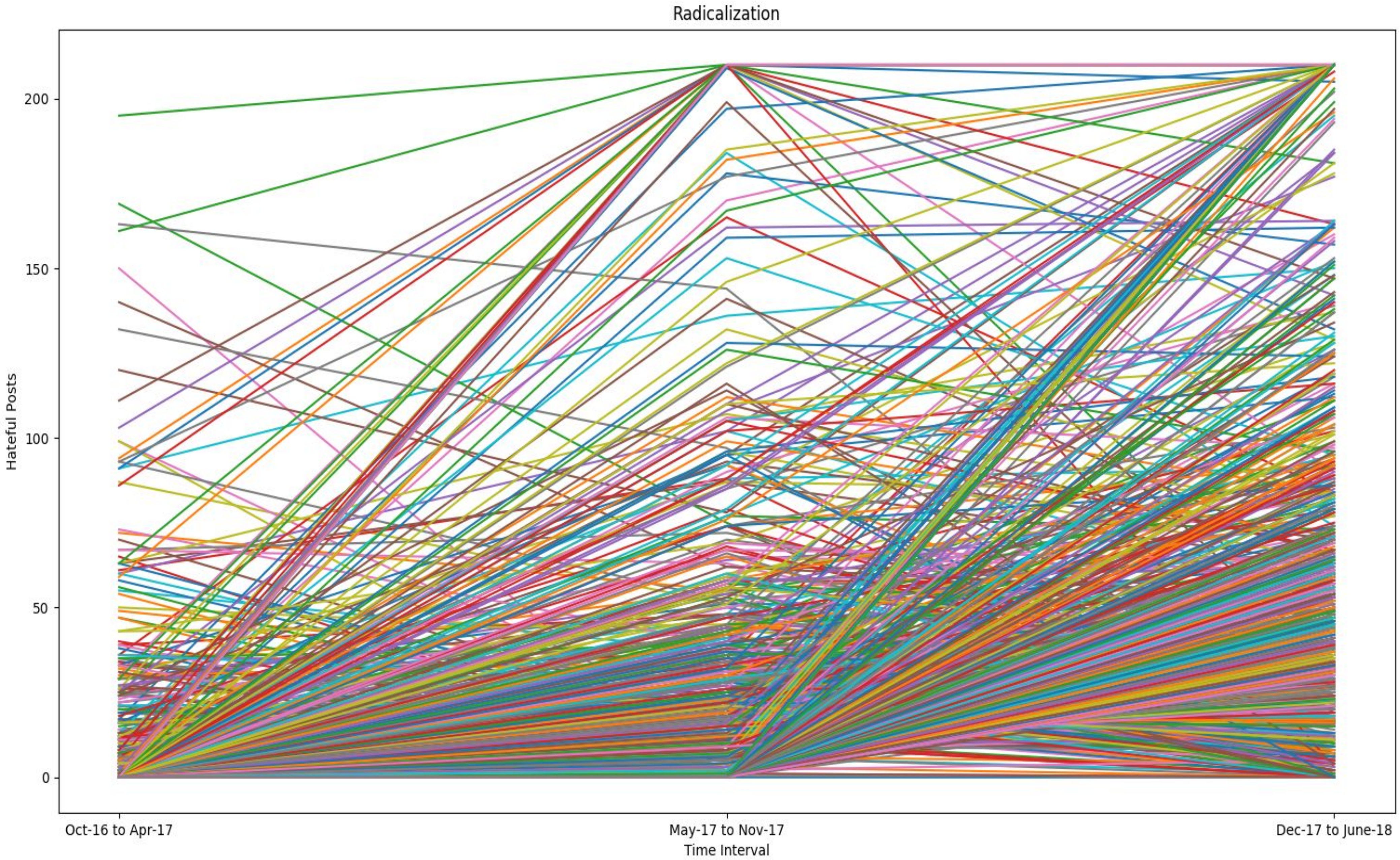
[2] Spread of hate speech in online social media Binny Mathew, Ritam Dutt, Pawan Goyal, Animesh Mukherjee

ANALYZING THE RESULTS OF DEGROOT MODEL



Graph for visualization purpose only

Trend over 21 months of followers of hateful users(898 users)



ANALYZING NETWORK PROPERTIES

Degree Distribution:

For hateful users (7365 users)

Mean Degree: 2006

Min Degree: 4

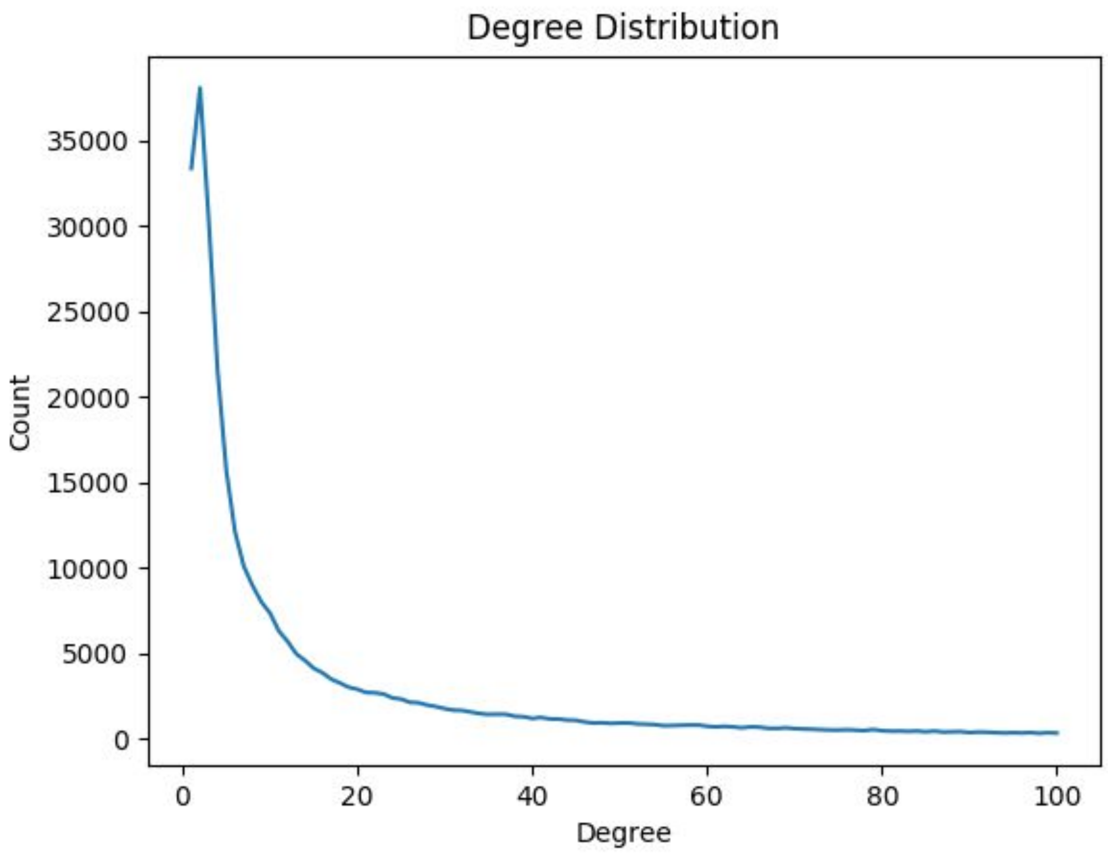
Max Degree: 31317

Users with degree > 1000 : 2465

For complete data(354947 users):

Average degree: 198.4

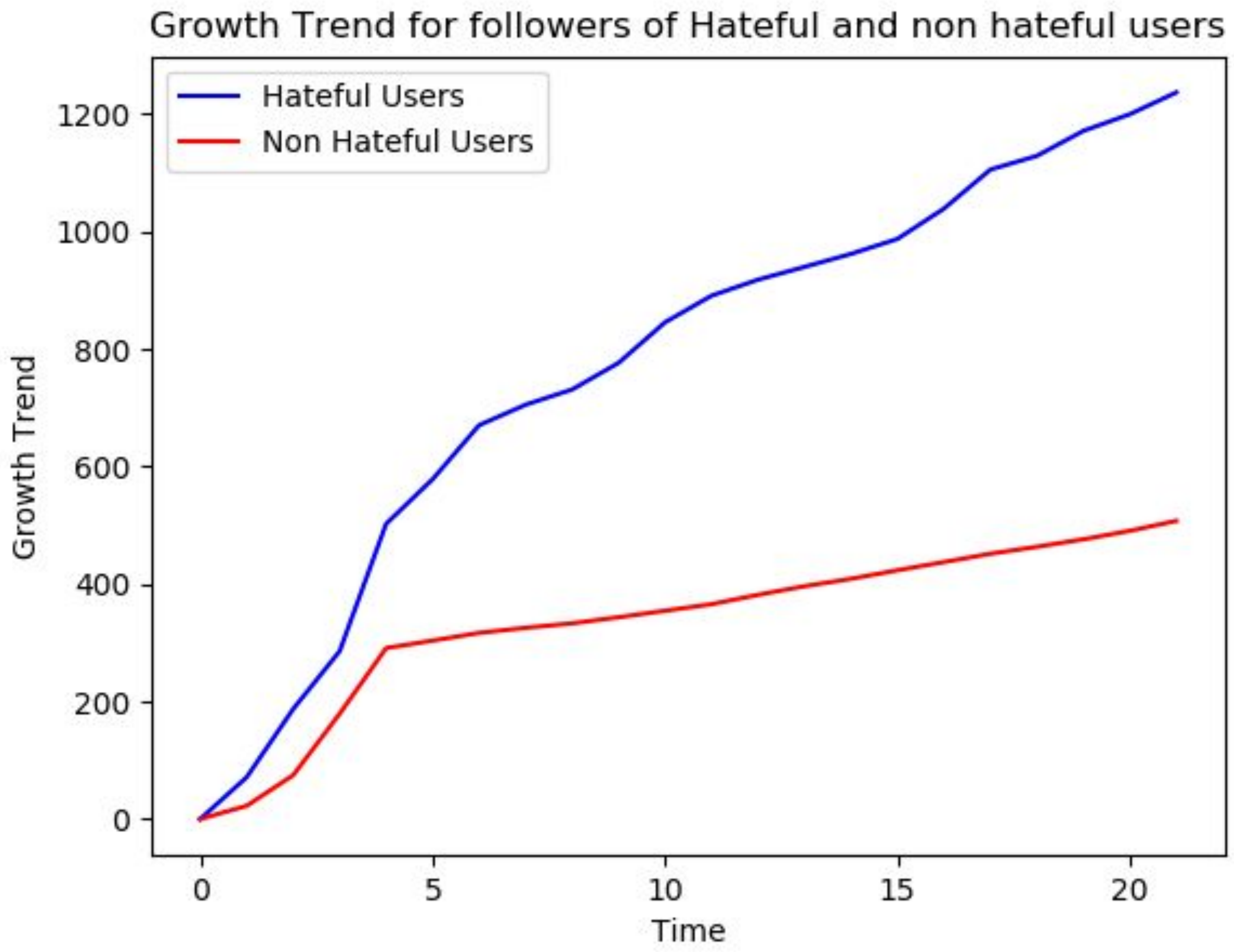
Max Degree: 31317



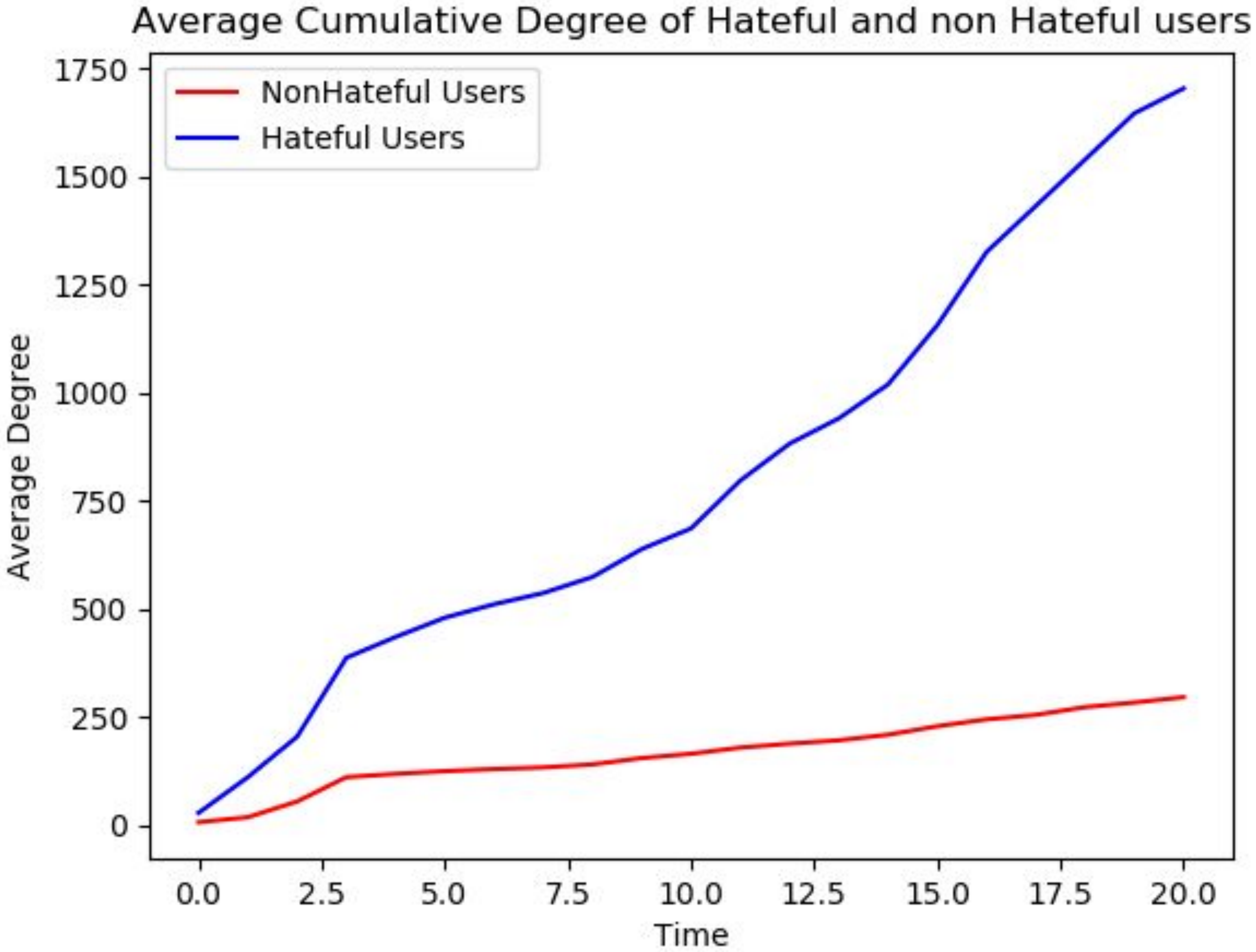
Note: graph is cropped till degree 100 only

ANALYZING THE TREND OVER 21 MONTHS

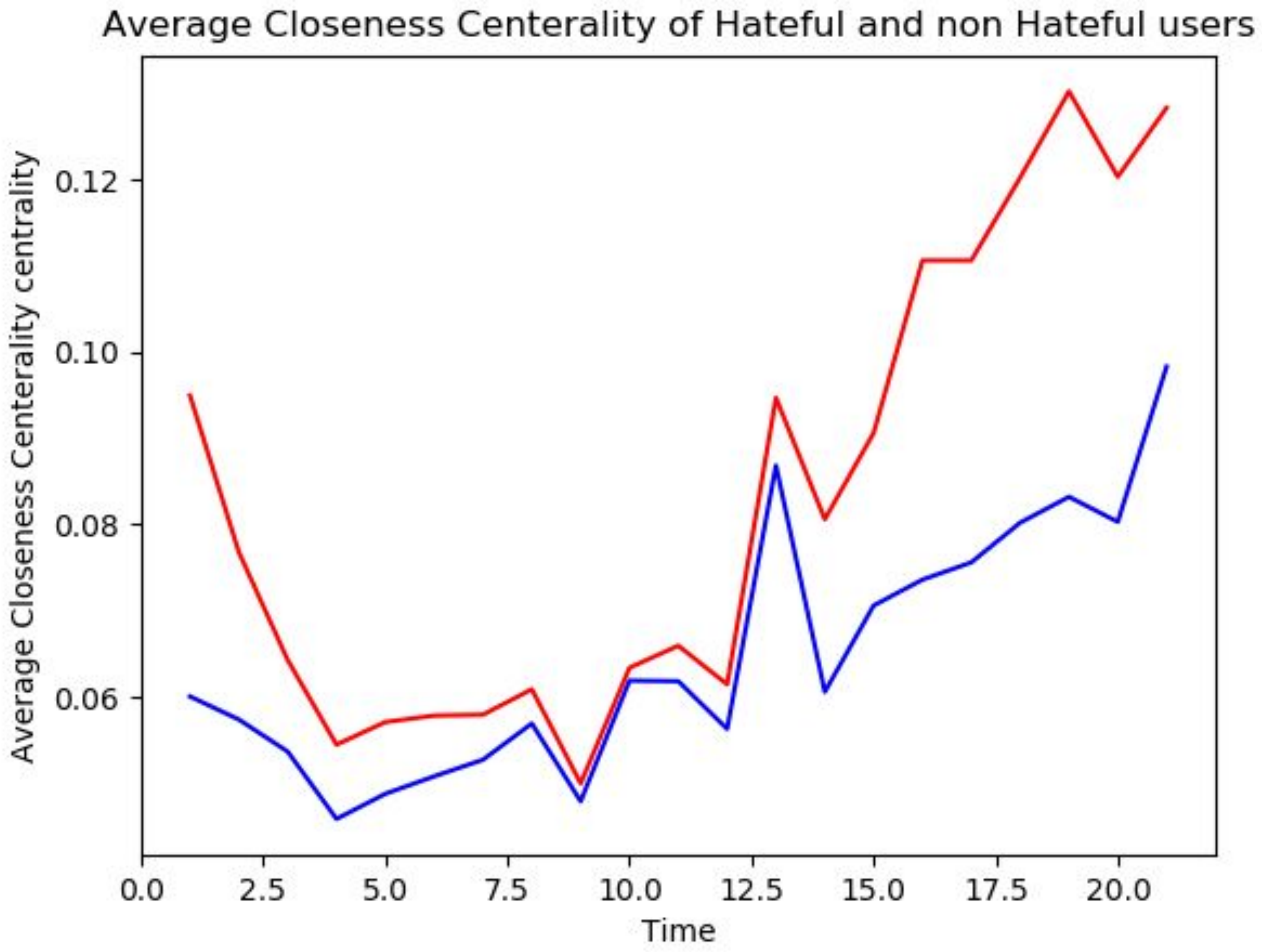
1) GROWTH TREND OF FOLLOWERS



2) DEGREE DISTRIBUTION

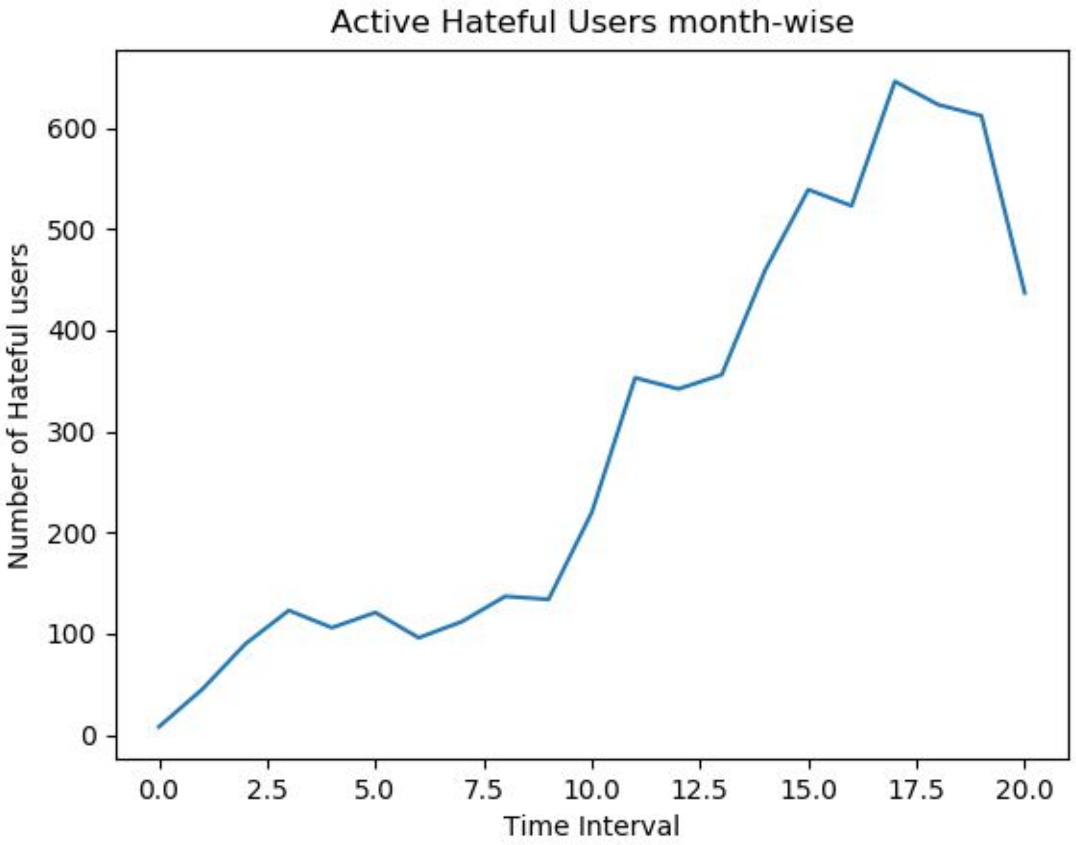


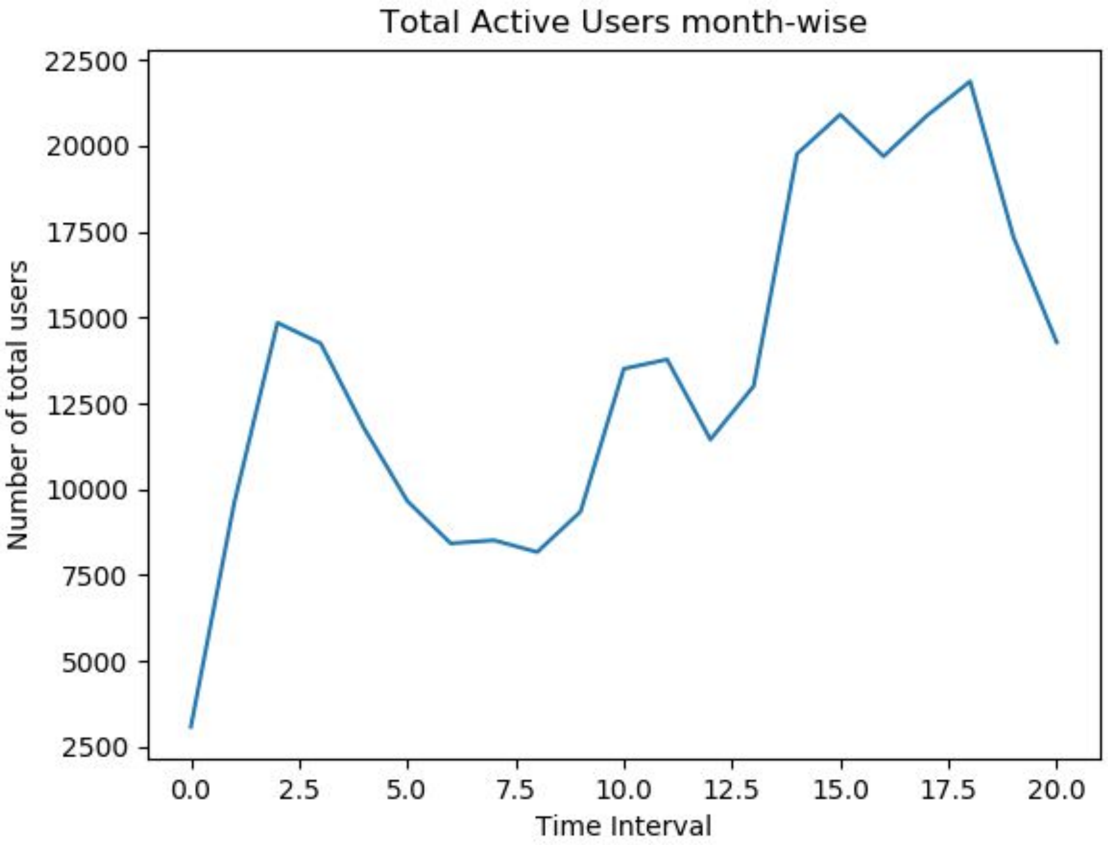
3) CLOSENESS CENTRALITY



4) TOTAL ACTIVE USERS (INCREMENTAL)

USERS WHO POSTED MORE THAN ONCE A MONTH WERE CONSIDERED ACTIVE





FINDINGS SUPPORTING OUR HYPOTHESIS

1) Hateful users = 7365 (2.4% of 300K approx.) had high impact on network.

2) Total followers of hateful users raised to 200K users.

3) Hateful users influence their followers to become hateful

- We tracked 898 non-hateful users who were following Hateful users
- Later on, almost all such users started posting (or reposting) hateful posts
- Through De-Groot model, we found that hate scores of almost all such users raised above hateful threshold score of 0.5 (i.e. they become hateful)

FUTURE WORKS

Most influential hateful user : Determine users who have influenced non-hateful user to radicalize.

Echo-chamber : By visiting an "echo chamber", people are able to seek out information which reinforces their existing views, potentially as an unconscious exercise of bias.

REFERENCES

[1] We Know Who You Followed Last Summer: Inferring Social Link Creation Times In Twitter. Brendan Meeder, Brian Karrer, Amin Sayedi

[2] Spread of hate speech in online social media Binny Mathew, Ritam Dutt, Pawan Goyal, Animesh Mukherjee