

ISIStwitter
Network

Analysis (Project 7)

6.5.2021

Niklas Saari

Project Members - Me?

I like coding.

Let's see how the writing goes...

Structure of the presentation

- Background of the project topic review based on prior research questions (Radical/Extremist Groups in the Social Media)
- Case: ISIS Network Analysis
- Dataset description
- Potential possibilities of the given data
- Project assignment limitations
- Results so far
- What next?

Project Background

Public social networks grow -> attention by extremist or radical organizations and individuals raises who try to influence other people.

They often disseminate their ideologies and destructive plans while trying to acquire new members.

As these organizations are often violent and radical hate groups, some approaches could be implemented to detect members, possibly to prevent spreading of propaganda and hateful

ideologies to even prevent war between peaceful societies.

Background - continues

Many parties could be interested to find ways how to observe and identify behaviour of these radical groups, to make actions based on observed information

Big social network companies are at some level applying content moderation, to make sure that people are applying into their terms of use (ToS).

We don't want hate speech, inciting to violence, etc.
don't we?

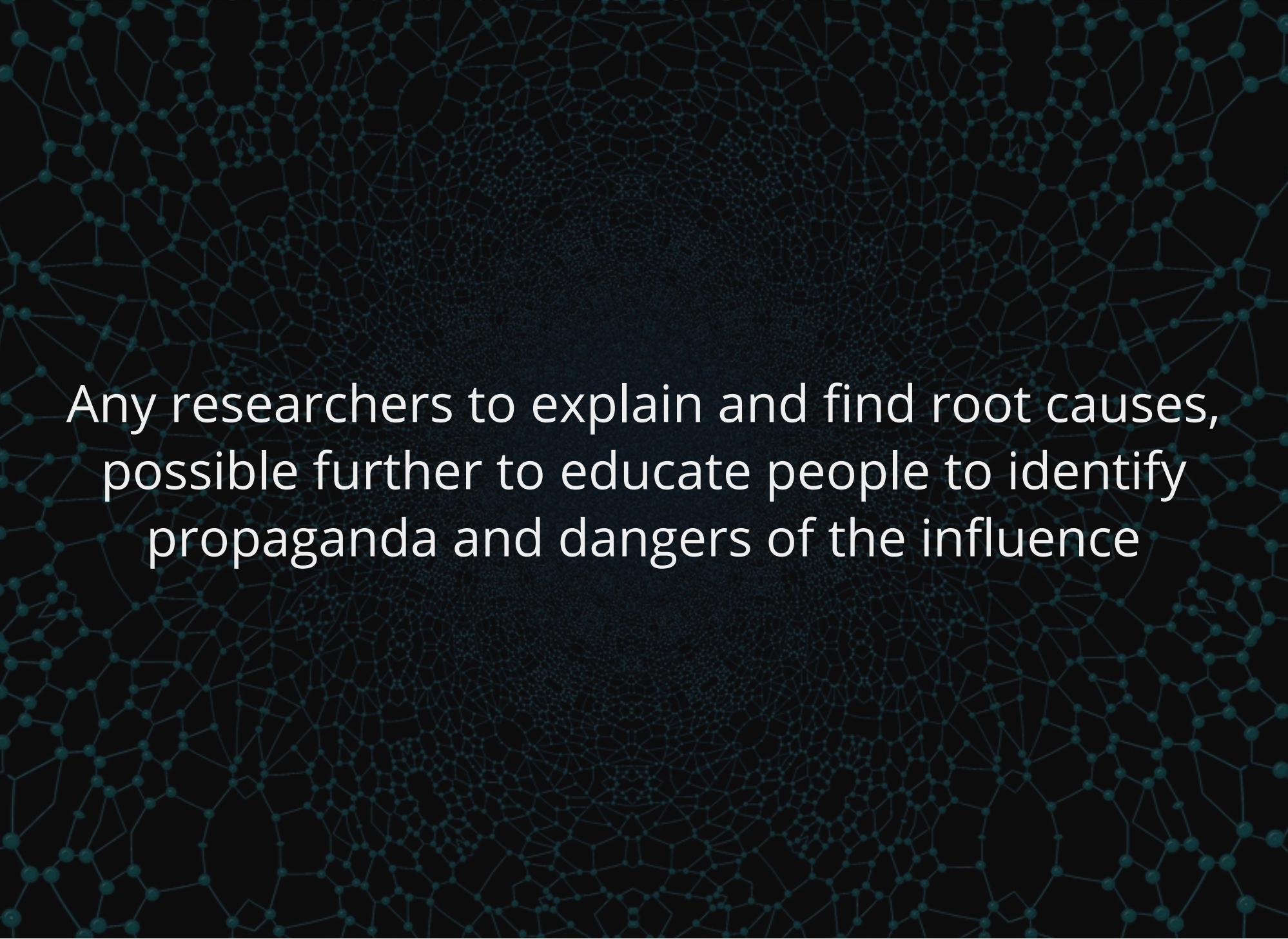
- Currently they are not hold responsible for applying content moderation in many countries, but still many are applying it (luckily).
- How to automate it at some level and find root causes? Of course, there are concerns for free-speech but we have should have find methods for identifying radical parts.



Intelligence services to prevent terrorism



Any officials to prevent general violence and crime



Any researchers to explain and find root causes,
possible further to educate people to identify
propaganda and dangers of the influence

Case: ISIS Twitter Network

Islamic State of Iraq and the Levant (also known as Syria) (ISIS) is classified as radical group that follows a jihadist doctrine of Sunni Islam.

For years, known to spread its message and recruit advocates through social media.

ISIS Network Twitter data was published in the dark web.

Contains data from pro-ISIS fanboys from all over the world since the January 2015 (before November 2015 Paris Attacks) and after until May in 2016

Dataset description

- Contains 17410 different tweets
- The first tweet dates to 2015-01-06 and the last one 2016-05-13
- Has total of 112 different users, all claimed to be pro-ISIS fanboys

What it holds

Data content per tweet as following:

- Name
- Username (Twitter ID)
- Description
- Location (User supplied input)
- Followers (Amount)
- Number of statuses (Unsure what this means)
- Tweet time
- Tweet content

What we could learn from this specific data?

Who are the major players in the ISIS Twitter Network? E.g. who has the most influence power and how they behave?

Sentiment analysis for the content of the tweet; is it negative, neutral or positive and in which context.

Which keywords are mostly used? E.g. hashtags
and how they are connected

Are the spikes in the tweets on certain time?

Are location related to tweets or time? Or
something else?

Limitations for this specific assignment

Base on the given assignment, the scope is limited into

- Tweet data amount analysis - power law?
- Analysis based on mentions, retweets or hashtags
- Identify users with the most usage, analyse relationships
- Use VADER tool for identifying tweet sentiment (NEGATIVE, NEUTRAL, POSITIVE)

- Build social network from hashtags, summarize its properties, generate communities form the network by using Girvan-Newman algorithm
- For above, is it possible to connect follower amounts for these communities? Reduce the amount size of graph and plot relevant attribute values.

Initial work

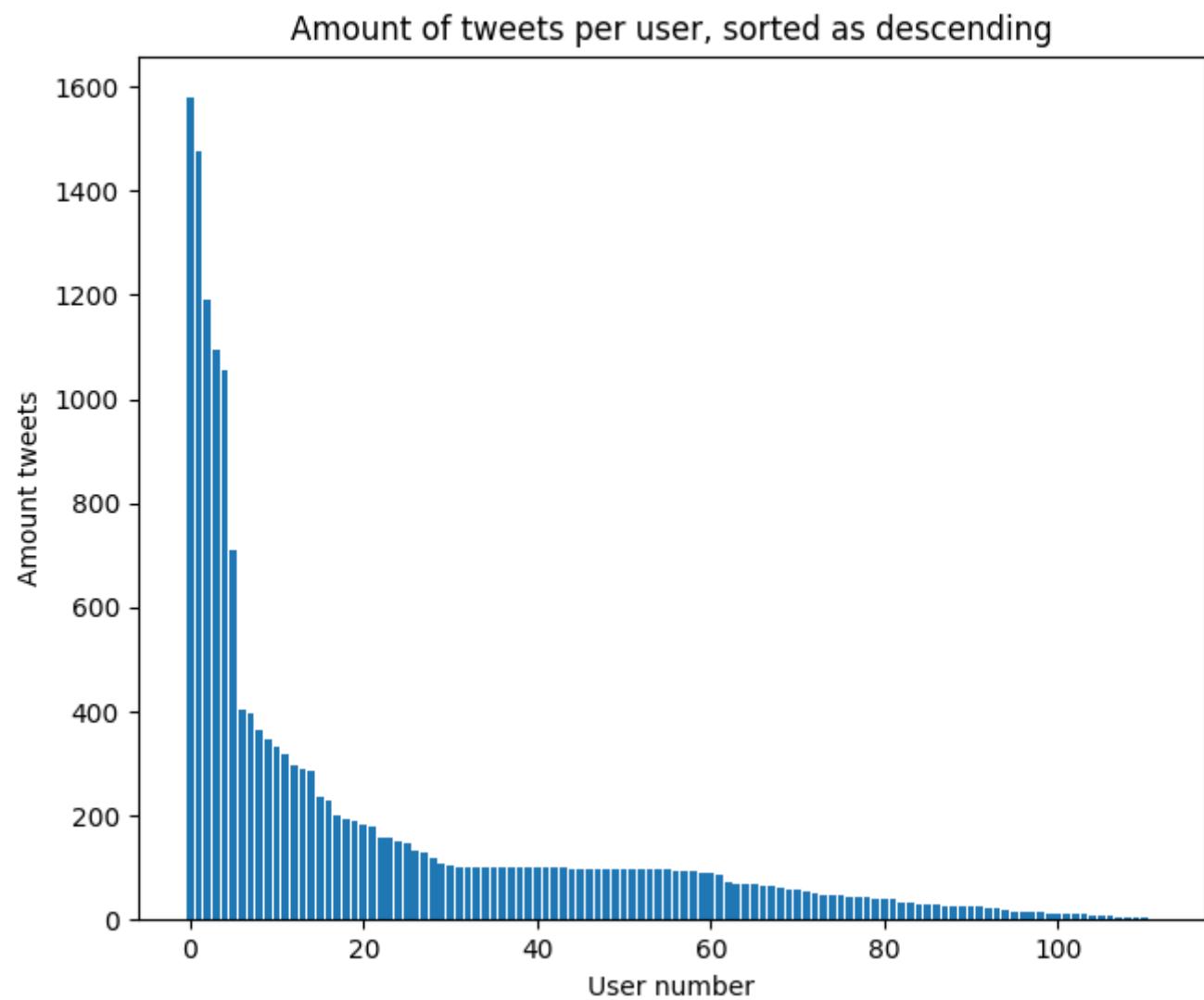
Data has been successfully parsed to get required information parts.

Test cases constructed for validating that implemented code methods work.

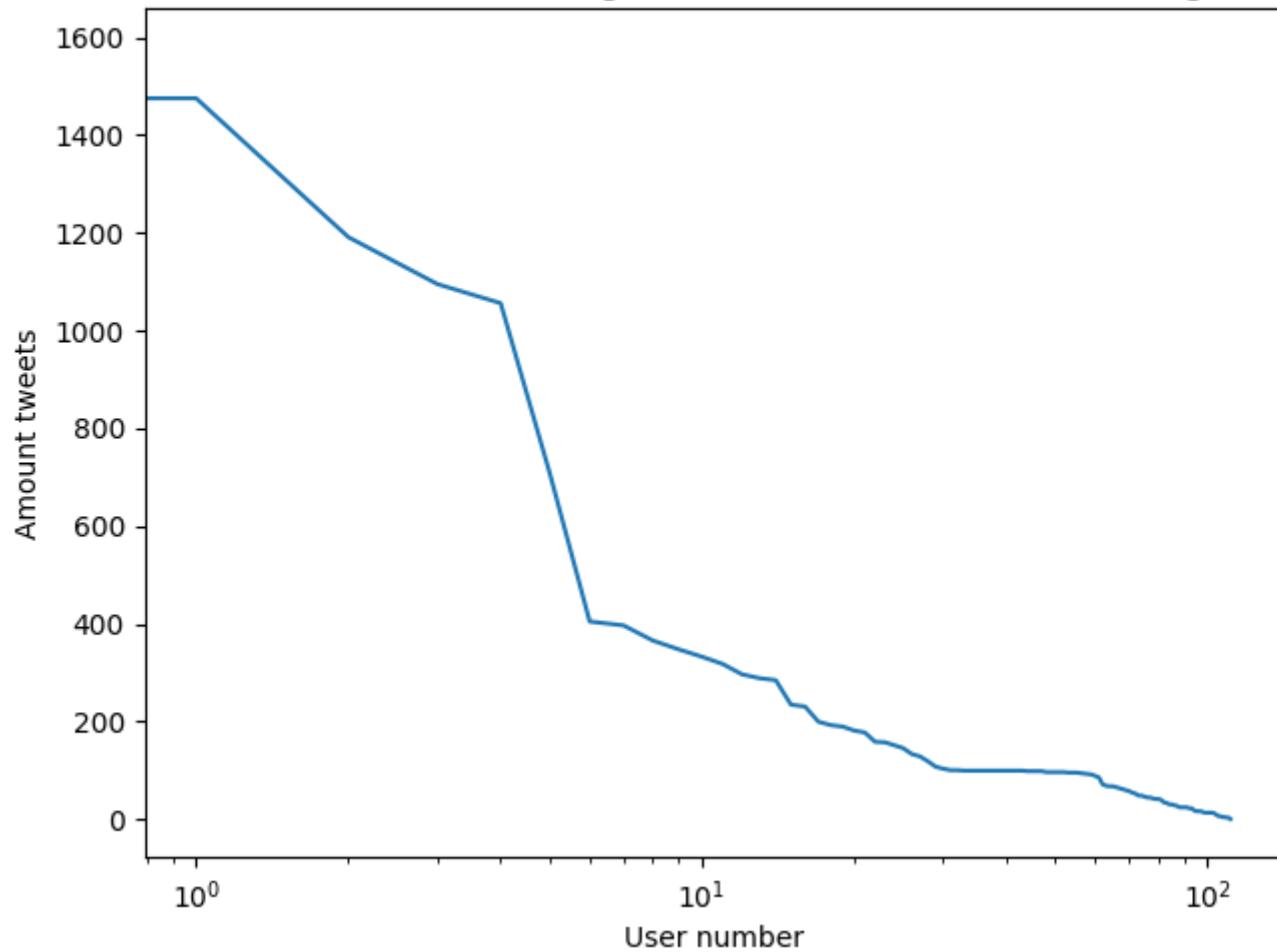
Using pytest framework to validate, that #hashtags, @mentions, RT retweets and VADER sentiments have been correctly identified from sample data; cases are manually validated.

Initial analysis

The amount of tweets per user:



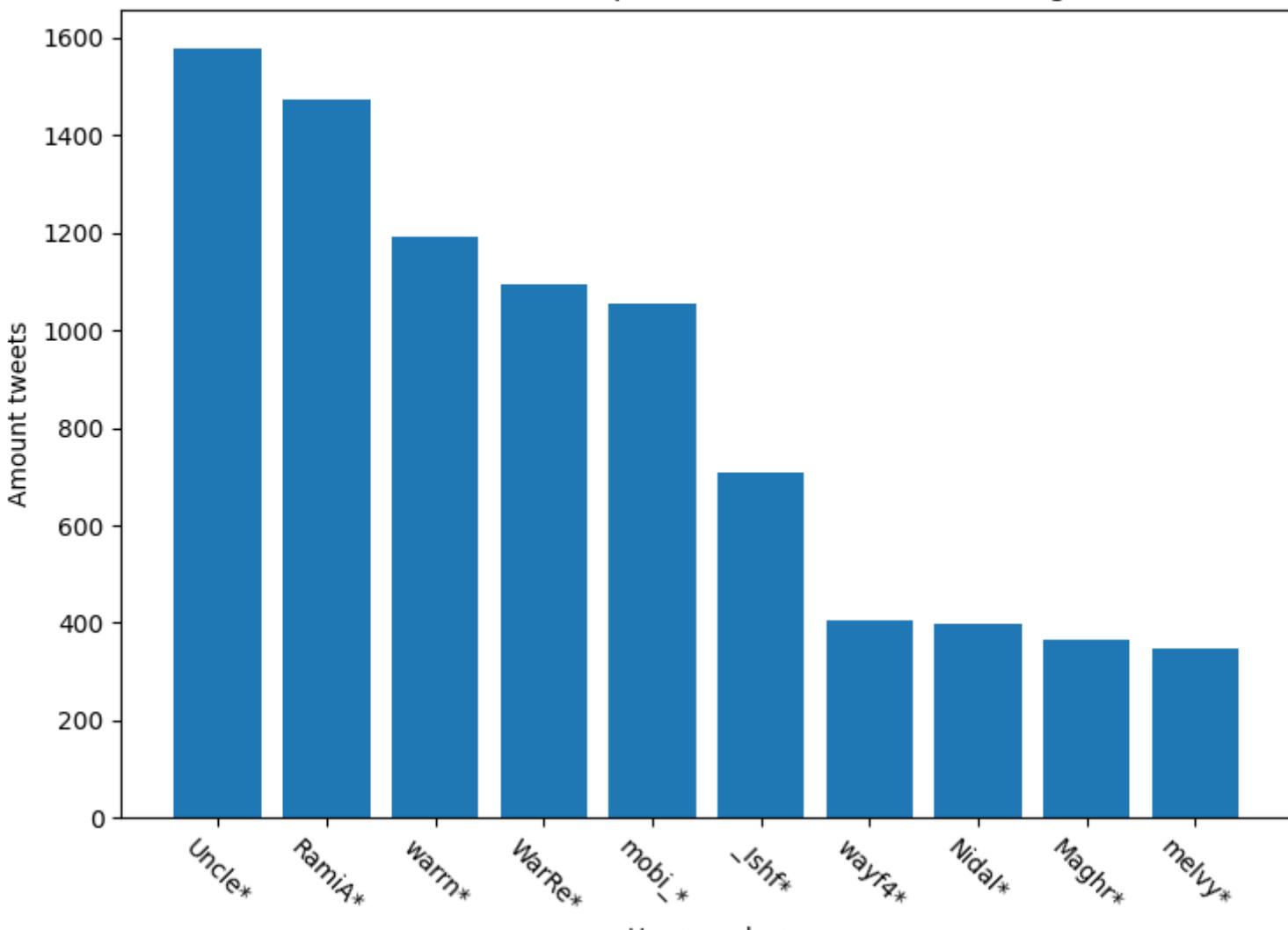
Amount of tweets in logarithm scale, sorted as descending



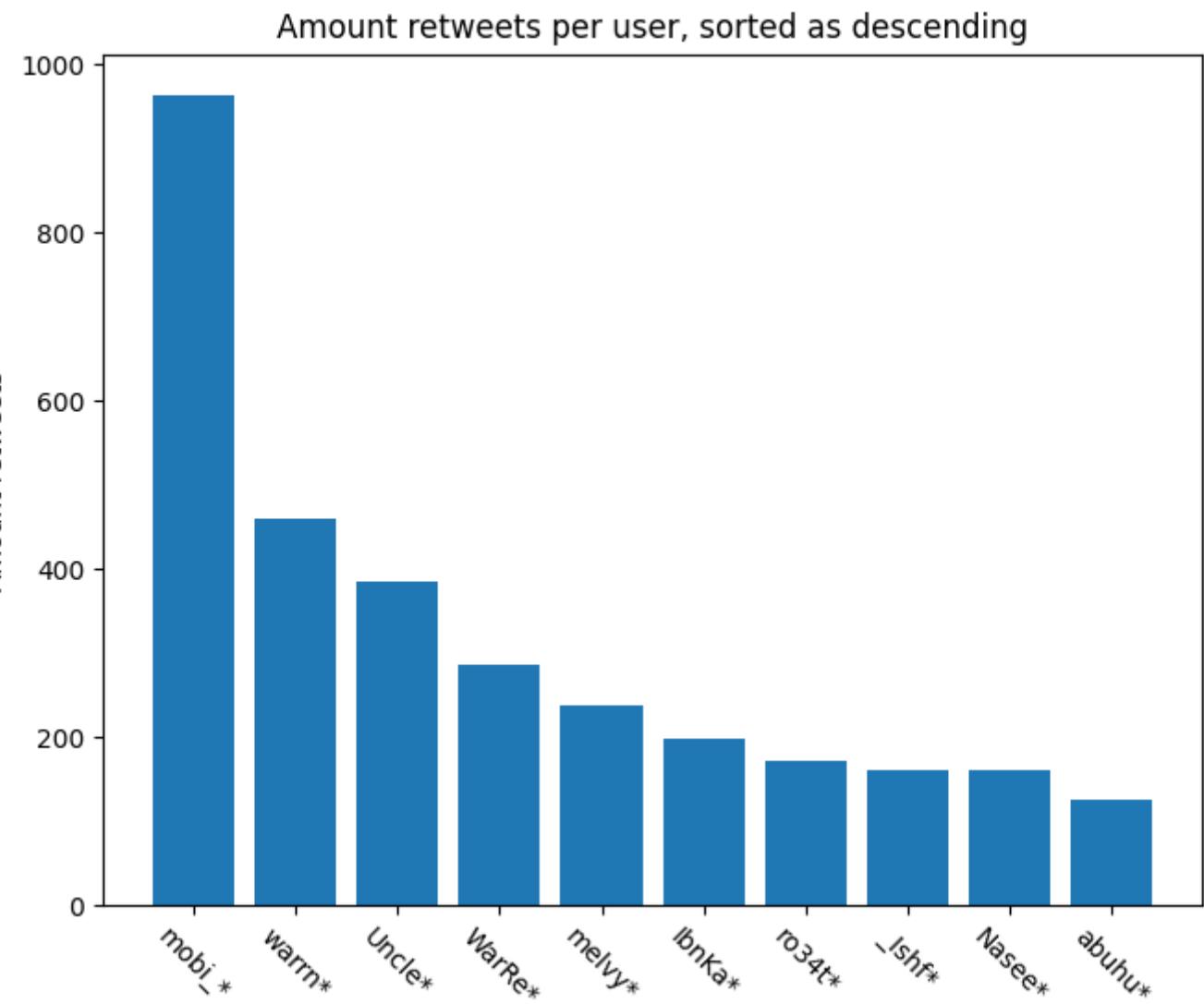
Seems to descent "almost" linearly.

Let's see the partial names of top 10 users by
Tweets

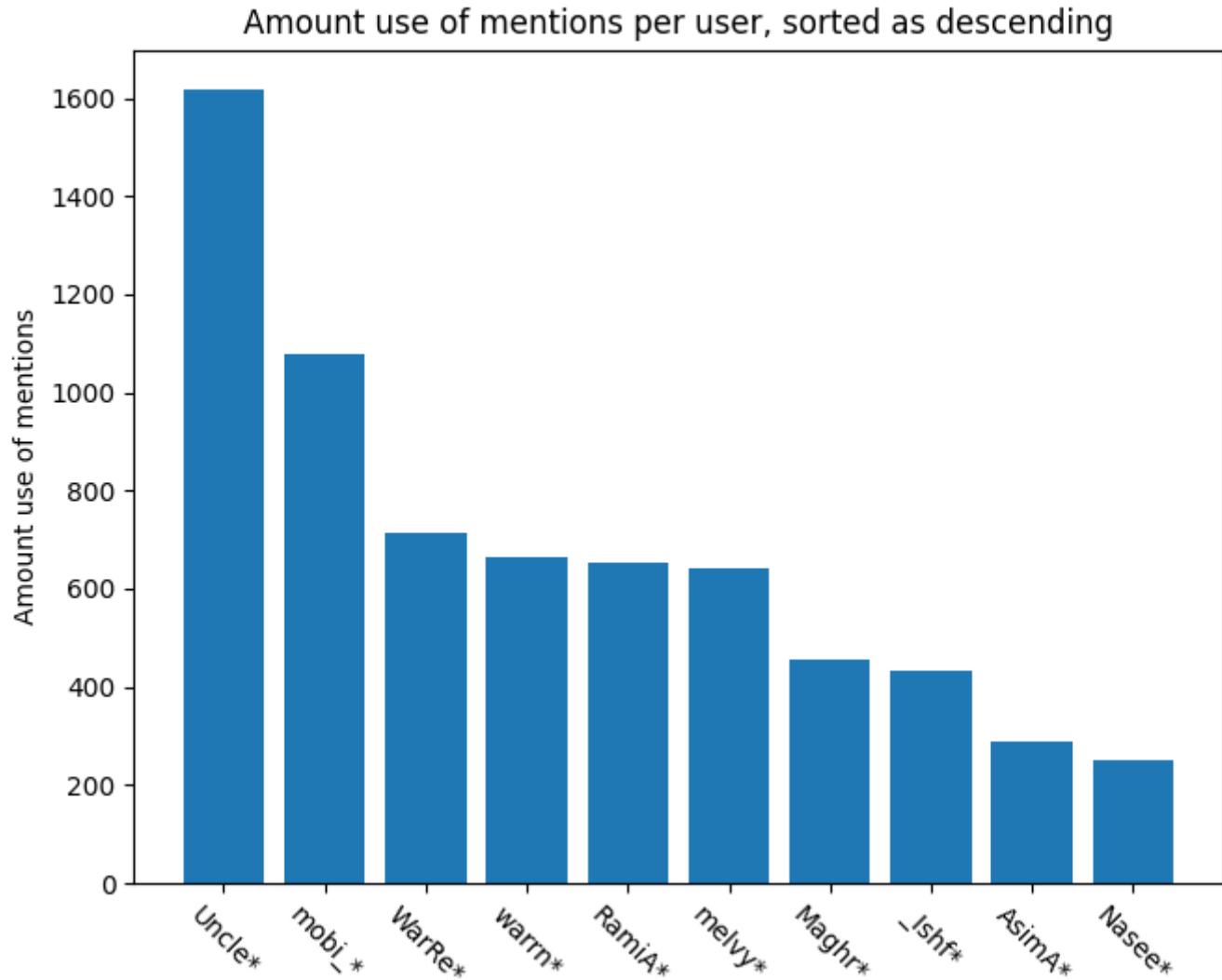
Amount of tweets per user, sorted as descending



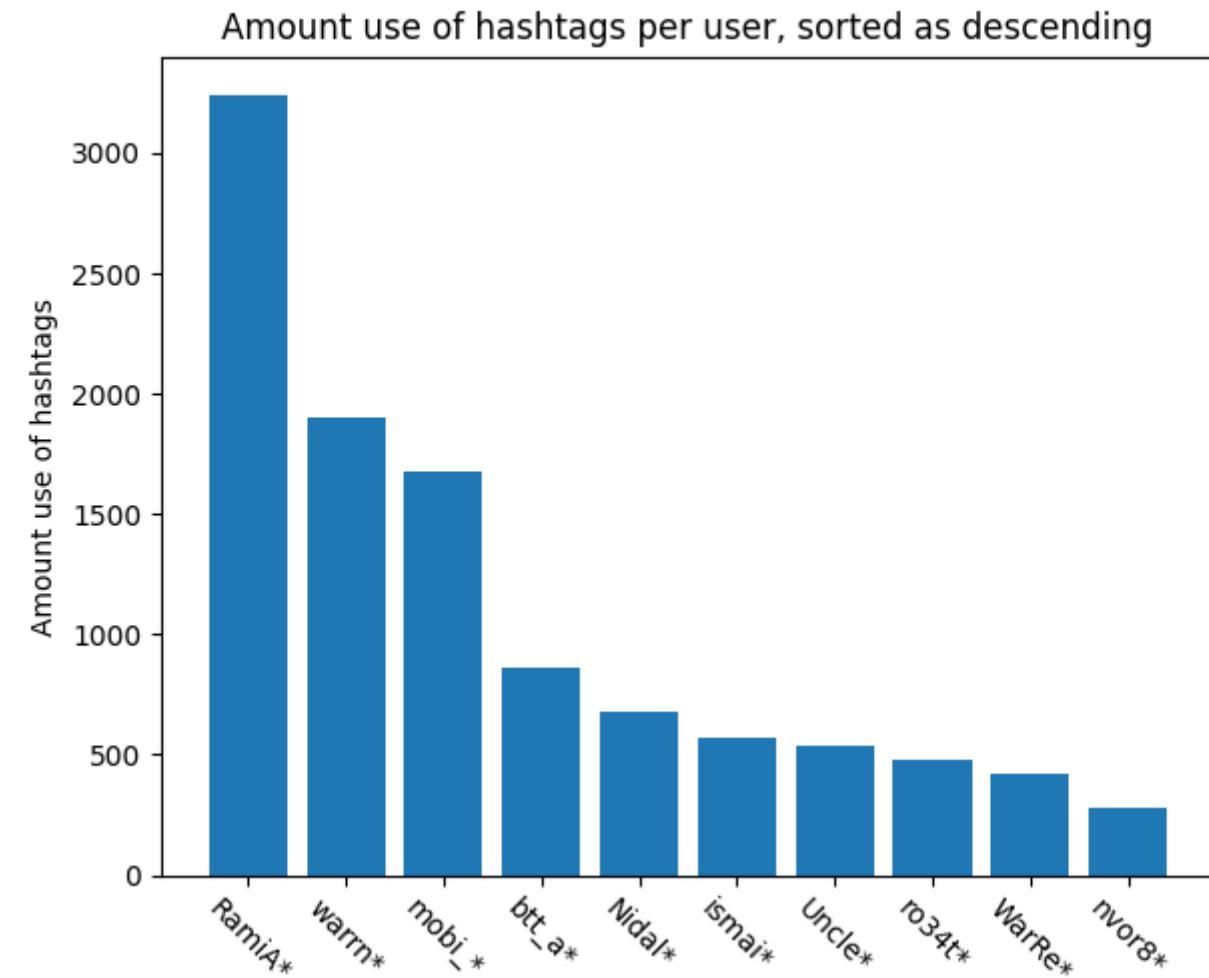
Let's see the partial names of top 10 users by
retweets



Top 10 users by the use of mentions



Top 10 users by the use of hashtags



Distribution of VADER sentiments

Amount of negative tweets: 6948

Amount of neutral tweets: 6657

Amount of positive tweets: 3805

Ternary plots work in progress...

What next?

Probably the biggest work ahead, still have to construct network from hashtags.

And should continue on writing..

Other requirements are probably close to met.