

Decentralized Social Network Analysis

Yicheng “Catherine” Wang
Research Advisor: Dr. Oshani Seneviratne

Background

- Decentralized Social Networks are Online Social Network implemented on a distributed platform.
 - Reduce privacy concerns, censorship, rewarding mechanisms
- Steemit
 - More than 1,238,000 users, 200,000 pageviews per day worldwide

Research Statement

- How does Decentralized Social Networks users react to the significant world events?
- What kind of Decentralized Social Networks user interactions and activities are affected by users' geolocations?

Project Relevance

- With the rise of Decentralized Social Networks, particularly Steemit, it is important to analyze how users socialize and interact in such platforms.

Hypothesis

- On Decentralized Social Networks, users not only discuss topics related to cryptocurrencies and blockchain, but also discuss world events.

Work Plan

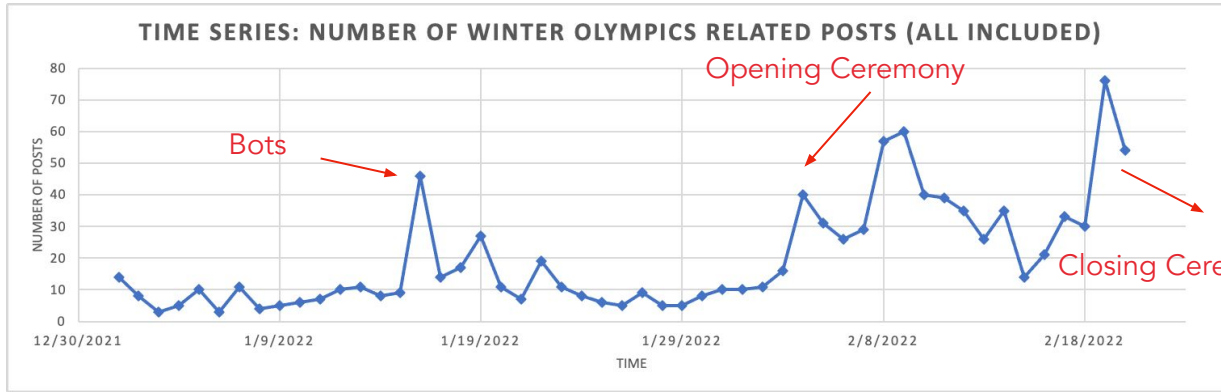
Spring 2020	Fall 2021	Spring 2022
2/15/2021 - 3/18/2021 Understand blockchain and how Ethereum smart contracts work	9/6/2021 — 9/20/2021 Literature reviews on Decentralized Social Networks	1/28/2022 — 2/11/2022 Geolocation information analysis for collected datasets
3/4/2021 — 3/18/2021 Literature Review on Ethereum smart contracts applications	9/20/2021 — 10/13/2021 Collect user data from Steemit and Mastodon for four predetermined events' period	2/11/2022 — 3/10/2022 Collect user data from Steemit from 12/6/2022 — 2/20/2022 and literature review on posts related textual analysis
3/18/2021 — 3/25/2021 Understand and determine data use agreements from covid-19 data sharing policies	10/13/2021 — 11/15/2021 Work with the collaborators from the architecture department for Geographical Information System (GIS) mapping and geolocation analysis	3/10/2022 — current Conduct data preprocessing process
3/25 — 5/6/2021 Implement written data sharing policies into a computational data use agreement on the Ethereum blockchain	11/15/2021 — 12/6/2021 Applying Natural Language Processing tools to conduct data analysis for specific events on continent basis	

Method and Materials

- **Extracted data from the Steemit public blockchain API.**
 - containing the Tokyo 2020 Summer Olympics, the release of Fukushima's wastewater in 2021, the Beijing 2022 Winter Olympics, and Russia-Ukraine conflict
 - Github Link:
- **Data Preprocessing and cleaning**
 - Translate comments and geolocations into English version
 - Removal all HTML tags
 - Remove all links
 - Remove all emoticons and special characters
 - Convert texts to lower case
 - Separate location information from json metadata
 - Convert geolocation into standardized format
 - Filtering events-related posts
- **Topic modeling, Name Entity Recognition, and Word Cloud applying Natural Language Processing Tools**

Results and Analysis

Winter Olympics related comments from 1/1/2022 to 2/20/2022



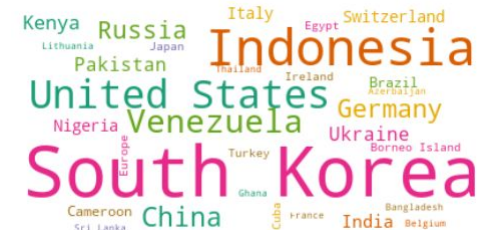
Comments include both users with geolocations and without geolocations (1005 comments)



Comments include both users with geolocations only (353 comments)



Continent-level word cloud with geolocation only



Country-level word cloud with geolocation only

Results and Analysis

Russia-Ukraine related comments from 1/1/2022 to 2/20/2022



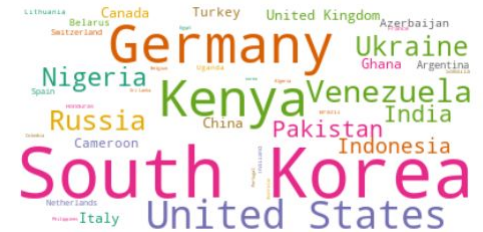
Comments include both users with geolocations and without geolocations (3340 comments)



Comments include both users with geolocations only (1170 comments)



Continent-level word cloud with geolocation only



Country-level word cloud with geolocation only

Conclusion

- A major finding in the study is while users are talking about topics related to blockchain and cryptocurrency, there appears to be correlations between world events and topics discussed during those time periods in the decentralized social media websites
- Future work
 - Change the analysis levels from continent levels to country or city levels to identify topics discussed by each specific regions
 - Challenges
 - Lack of good quality and quantity of geo-tagged posts and comments
 - Lack of reliability of user-provided geolocation information
 - Apply more different NLP techniques for the analysis.