VoterFraud2020
a Multi-modal Dataset of Election Fraud Claims on Twitter

Anton Abilov, Yiqing Hua, Hana Matatov, Ofra Amir, Mor Naaman

- Voterfraud2020 is a public Twitter dataset with **7.6M tweets** and **25.6M** retweets related to voter fraud claims collected between Oct 23rd and Dec 16th, 2020.

- The dataset can be used by researchers to **study the spread, reach, and dynamics of the voter fraud claims campaign** on Twitter.

- In addition to raw Twitter data, the dataset includes **community labels** of users computed from the retweet graph, perceptual **hash values of the tweeted images** and **aggregate data for links and Youtube videos** that appear in the tweets.

- The community detection revealed **five main distinct communities** engaged in the discussion of voter fraud claims, including promoters (55% of users) and detractors.

- **Twitter's suspensions** mostly affected a sub-community of **users who engaged more actively with QAnon content**; user suspension status is included in the data.

- In addition to the public data, due to the immense public interest we will **release the tweets of suspended users** to published academic researchers.

@AntonAbilov

@yiqqqing

@hanamatatov

@ofraam

@informor

CORNELL TECH
HOME OF THE JACOBS INSTITUTE

voterfraud2020.io