UFC

Uncovering Coordinated Communities on Twitter During the 2020 U.S. Election

Renan S Linhares, José M Rosa, Carlos HG Ferreira, Fabricio Murai, Gabriel Nobre, Jussara Almeida

Vaux Gomes [1]

[1]Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Campus Jaguaribe

Presentation Feb/2025

Table of Contents

#	Value
Title	Uncovering coordinated communities on twitter during the 2020 us election
Year	2022
Event	The international conference on Advances in Social Network Analysis and Mining (ASONAM)
Authors	Renan S Linhares, José M Rosa, Carlos HG Ferreira, Fabricio Murai, Gabriel Nobre, Jussara Almeida

UFC

Social Media [Backlinko]

As of January 2025, there are over **5.24 billion people using social media worldwide**, and the **average user accesses 6.83 social media platforms monthly** [1].

[1] Source: Backlinko. Social Media Usage & Growth Statistics. Last update Feb 2025.

- Social networks drive a wide range of **coordinated activities**, both positive and negative.

Positive Coordinated Activities	**Negative Coordinated Activities**
Social Movements and Protests	Misinformation and Disinformation Campaigns
Disaster Relief and Humanitarian Aid	Hate Speech and Online Harassment
Crowdfunding and Fundraising	Cyberbullying and Online Mobs
Collaborative Projects	**Extremism and Radicalization**

Introduction

Paper context

Quote

Indeed, the literature displays strong evidence that **many events during the U.S. election were influenced by coordinated user actions on Twitter**

- During these events, many **account suspensions** happened in many different platforms

Introduction

Paper context

- After the 2020 U.S. election there was a **riot at the Capitol**
- The event occurred while **Trump was in change of the presidency**

UFC

- Two big questions arise:
  - How effective were the account suspensions on **preventing negative coordinated activities** and related content?
  - How these suspensions **helped protecting (avoiding harm over) the electoral process**?

Introduction

Previous works

Previous Papers
H. Tran, "Studying the community of trump supporters on twitter during the 2020 us presidential election via hashtags# maga and# trump2020," Journalism and Media, 2021.
Y. Dai, "Using 2020 u.s. presidential election to study patterns of user influence, community formation and behaviors on twitter," Ph.D. dissertation, The Pennsylvania State University, 2021.
A. Abilov, Y. Hua, H. Matatov, O. Amir, and M. Naaman, "Voterfraud2020: a multi-modal dataset of election fraud claims on twitter," in International Conference on Web and Social Media, 2021.
K. Sharma, E. Ferrara, and Y. Liu, "Characterizing online engagement with disinformation and conspiracies in the 2020 u.s. presidential election," International AAAI Conference on Web and Social Media, 2022.

- Previous' works approaches
  - Most approaches rely on **users links** to find tightly connected groups (**communities**)
  - Their evidence of an existing group is the **promotion of the same peace of content**

- Neglected aspects:
  1. Potentially large presence of **Noisy Edges**
     - **Marginal edges** that are not related to the coordination
  2. **Partial view** of the network:
     - Data sampling can lead to misinterpretation[2]

[2] The whole net is never seen!

Introduction

Paper's approach

Quote

(...) we propose a novel approach to **identify edges connecting groups** of users who retweeted the same content **with strong evidence of consistent and coordinated behavior**.

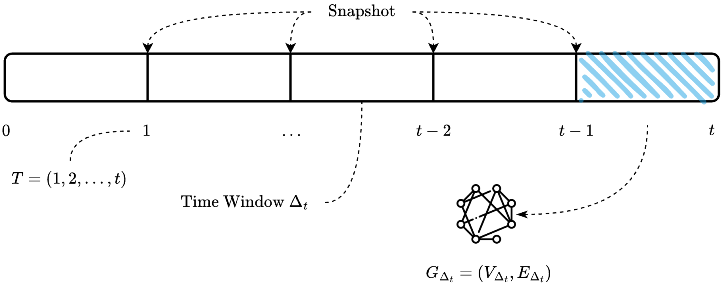
- Applied techniques:
  - Disparity Filter
  - Neighborhood Overlap

Methodology

1. Co-Retweet Network Model
2. Backbone Extraction
3. Community Detection and Characterization

Methodology

Co-Retweet Network Model



- V_{Δ_t} is the set of nodes, representing users who **retweeted** during Δ_t; and
- E_{Δ_t} is the set of **weighted edges** connecting pairs of nodes v_i and v_j
    - w is the **number of retweets** in common between v_i and v_j in time window Δ_t

Methodology

Co-Retweet Network Model

Quote

The **co-retweet** network has the ability to **explicitly represent connections among users who aimed at promoting certain pieces of content** (a.k.a. original tweets) by retweeting them

- The authors believe that this approach has a **greater chance to model coordinated promotion** of content

Methodology

Co-Retweet Network Model

- Naturally, the co-retweet network has its problems
  - For instance, **very popular tweets may be retweeted by many users** acting independently
  - Likewise, **very active users** may retweet the same content as many other users by chance
- These problems generate a large number of weak edges

- Another problem, which is data related, is **the amount of data available by the APIs**
  - It is possible that the data available presents only a partial view of the net
- A solution for all these problems is achieved by **finding a suitable backbone** for the net

Methodology

Backbone Extraction: A two-step approach

Backbone

The set of edges **capturing co-retweeting patterns** that are more likely to reflect coordination (i.e. a subset of nodes and edges that characterizes communities most likely involved in coordinated actions).

- The backbone is extracted by two methods combined (previously mentioned)

Methodology

Backbone Extraction: A two-step approach

- First, extract initial backbone, retaining only **heavy edges** which can be used to identify strong evidence of coordination

$$\text{DisparityFilter}(G_{\Delta_t}, \alpha) \rightarrow B_{\Delta_t} \tag{1}$$

- Where $\alpha = 0.05$ is called the significance level

Methodology

Backbone Extraction: A two-step approach

Quote

The intuition is to **retain only edges connecting users with a number of retweets in common much larger than expected** given their typical behavior (...)

Disparity Filter Note

This is not only a counting threshold strategy. It is based on probability distribution calculations.

Methodology

Backbone Extraction: A two-step approach

- Secondly, by applying a graph metric called **Neighborhood Overlap** only nodes with a sufficient number of neighbors are kept (C_{Δ_t}):

$$C_{\Delta_t} \subseteq B_{\Delta_t} \subseteq G_{\Delta_t} \tag{2}$$

- Let (v_i, v_j) be an edge in B_{Δ_t} and $N(\cdot)$ a function that returns the set of neighbors of a node. Thus:

$$\text{JaccardSimilarity}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|} \tag{3}$$

- The Neighborhood Overlap method consists of **keeping the top $k\%$ edges** with larger JaccardSimilarity values

- Finally, it is time to **find communities** from C_{Δ_t}
- By definition these communities are going to be formed by **users that retweeted the same content in an unnatural frequency**

Louvain's community detection algorithm

Louvain's community detection is a greedy algorithm that aims at optimizing a metric called *modularity*

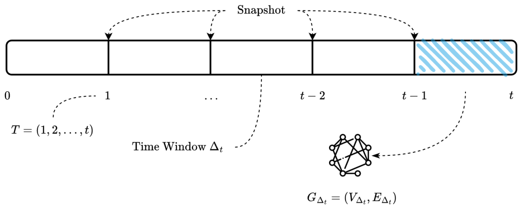
Modularity

Modularity evaluates the quality of the division of a graph in communities

- Modularity ranges from $[-1, 1]$. Values equal to 0.4 or higher are considered as **reliable indicatives of well-formed communities**

Methodology

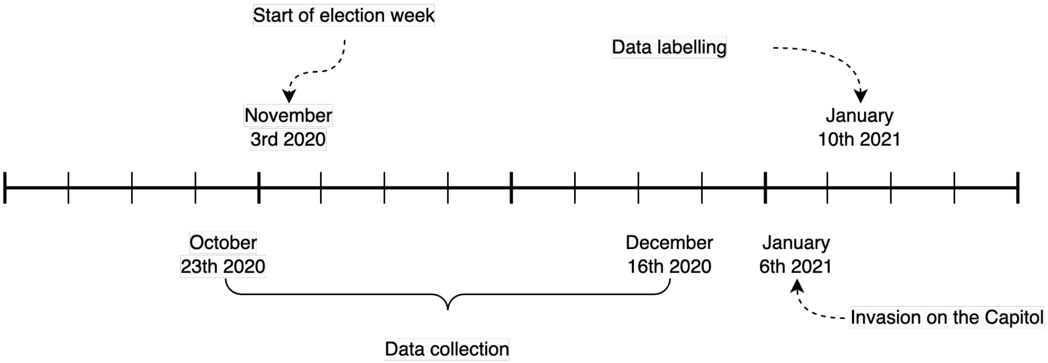
Community Detection and Characterization



- The **temporal dynamics** of the communities are measured by:
  - **Persistence:** Fraction of users in the backbone in window Δ_t who remained in the backbone in window Δ_{t+1}
  - **Normalized Mutual Information (NMI):** Assess whether users who appeared in the backbone across successive windows tend to remain in the same community or not

Case Study

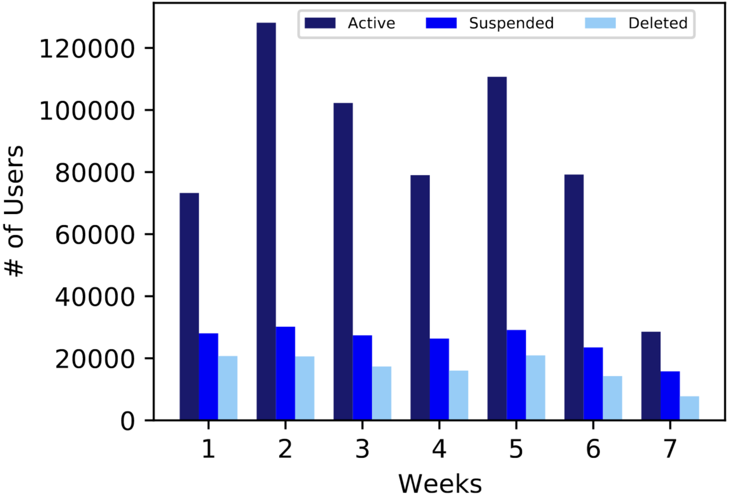
- The data used is VoterFraud Abilov et al. [2021]
- Tweets gathered by Twitter's API between October 23th and December 16th 2020.



Case Study

- The study used **7 consecutive weeks** of data focusing on the **most popular accounts** (users that receives at least 5000 re-tweets per week)
  - 186 users
  - 4,328 tweets
  - 4,545,021 re-tweets from 323,912 other users

Case Study


Active
Suspended
Deleted
120000
100000
80000
60000
40000
20000
0
# of Users
1
2
3
4
5
6
7
Weeks

Results

Backbone extraction

Network Model	#Nodes	#Edges	Avg.Deg.	Density	#C.C.	#Comm.	Mod.[3]
Original	121,992	651,835,172	19,686.5	0.0876	1	12	0.24
DF	28,310	1,709,735	120.8	0.0043	2	8	0.22
DF and NB	13,525	314,142	46.4	0.0030	26	89	0.51

Tabela 1: Network Model Statistics in Δ_1.

[3]Modularity values equal to 0.4 or higher are considered as reliable indicatives of well-formed communities

Results

Backbone extraction

- Disparity Filter (DF):
  - Large number of nodes and edges discarded
  - Fewer communities with weaker structures

Quote

- It suggests that the remaining nodes do not form well-structured communities
- It may be a side effect of the Twitter Sampling

Results

Backbone extraction

- Disparity Filter & Neighborhood Overlap (DF & NB, $k = 20\%$):
  - Less nodes and edges were kept
  - The number of communities is much higher with a stronger modularity value

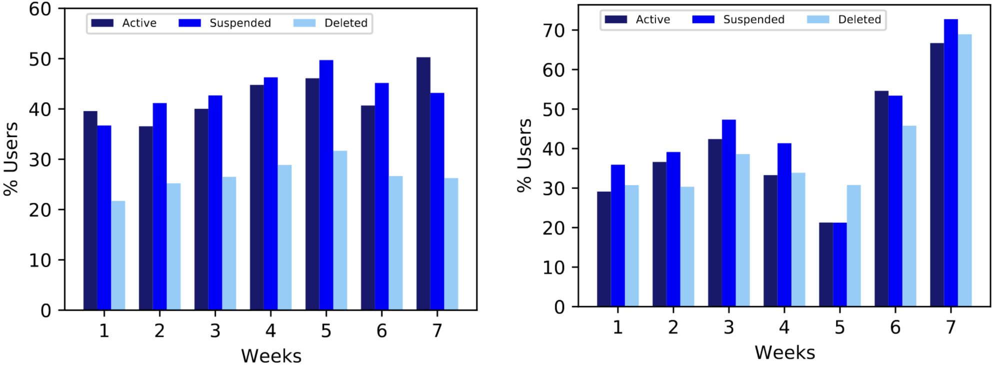
Results

Backbone extraction

Week	#Nodes	#Edges	Avg.Deg.	Density	Avg.Clust.	#C.C.	#Comm.	Mod.
1	13525	314142	46.4	0.003	0.35	26	89	0.51
2	23589	1540198	130.5	0.005	0.44	51	135	0.48
3	24662	1855157	150.4	0.006	0.40	45	74	0.44
4	18392	2241889	243.7	0.133	0.51	40	99	0.53
5	15978	2612256	326.9	0.021	0.55	35	106	0.44
6	25010	1887246	150.9	0.006	0.39	38	116	0.38
7	15944	317498	39.8	0.002	0.37	13	43	0.48

Results

Backbone extraction



(a) DF (b) DF and NB

Quote

Our backbones **capture the idea of users retweeting a large number of common tweets** while **still being connected to common neighbors** in the network.

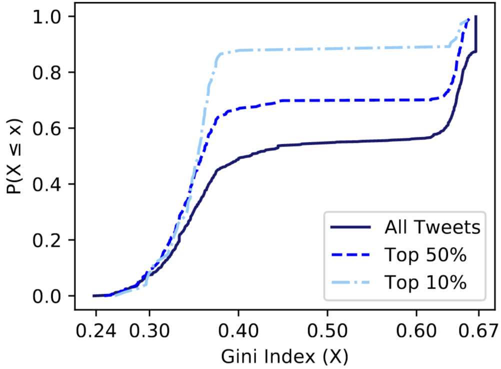
Results

Spread of information

- The spread of information was evidenced by the **Gini Index**:
  - The metric measures the **inequality in the distribution of retweets** among the three classes;
  - Closer to 0: **lower inequality**; retweets are **distributed more evenly** among different user classes;
  - Closer to 1: **higher inequality**; retweets are **concentrated** among a smaller number of user classes.

Results

Spread of information



- The Y-axis represents the **cumulative probability** of a tweet having a Gini index less than or equal to a specific value (x) on the X-axis.

- Overall interpretation
  - **General Inequality**: The graph shows that **retweet distribution is unequal**, as the Gini Index ranges from 0.24 to 0.67.
  - **Popular Tweets vs. All Tweets**: The "Top 10%" line is furthest to the left on the graph, indicating that **the most popular tweets (10%) have lower inequality in retweet distribution**
    - This suggests that popular tweets are **retweeted by a more diverse range of users**
  - **Top 50%**: The "Top 50%" line falls between the other two, showing that the 50% most popular tweets have an **intermediate level of inequality**.

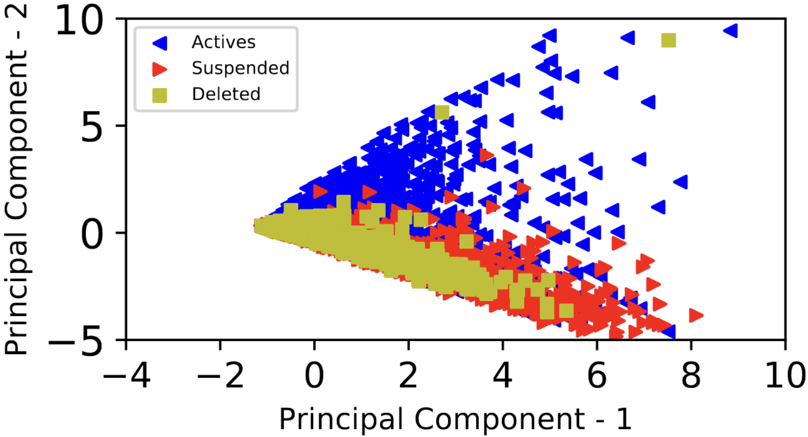
Results

Class behavior similarity

- Principal Component Analysis is used to summarize the similarity between in behavior for the users of each class
- Five metrics are calculated to characterize the users:
  (A) # tweets created
  (B) # **retweets**
  (C) # **words associated with extremist groups**
  (D) **degree centrality**
  (E) # weeks (out of 7) that the user stays on the backbone.

Results

Class behavior similarity



Results

Class behavior similarity

- The study shows that **(B) # retweets**, **(C) # words associated with extremist groups** and **(D) degree centrality** better separate the classes
- These features are <u>more common among users of class suspended and delete</u>

Results

Characterization of Potential Coordinated Communities

- It is shown by the persistence metric that users 25% (on average) of the users are kept in the backbone between window Δ_t and Δ_{t+1}
  - This suggests that there is some consistency on the spread of information by users
- At the same time, it is shown my the NMI metric that those users tend appear in different communities between windows
  - This suggests that their focus can shift

Conclusions

Quote

- We found that such communities consisted of (i) suspended, (ii) deleted, and (iii) still active users that exhibit behavior patterns similar to group (i)
- We also observed that the users who remained active during the observed period were found in backbones and even grouped in the same communities, thus exhibiting quite similar behavior
- Taken together, our **results suggest that the account ban imposed by Twitter after the 2020 election may not have effectively stopped the spread of information and conspiracy theories** and **should have captured a larger number of users**.

References

A. Abilov, Y. Hua, H. Matatov, O. Amir, and M. Naaman. Voterfraud2020: a multi-modal dataset of election fraud claims on twitter. In *Proceedings of the international AAAI conference on web and social media*, volume 15, pages 901–912, 2021.

R. S. Linhares, J. M. Rosa, C. H. Ferreira, F. Murai, G. Nobre, and J. Almeida. Uncovering coordinated communities on twitter during the 2020 us election. In *2022 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pages 80–87. IEEE, 2022.