

Reddit Users' Reaction to Major US Political Events in 2020-2021

Volik Oleksii and Vashchenko Mykola

Faculty of Informatics, National University of «Kyiv-Mohyla Academy»

Computational Social Science

Teacher: Andrew Kurochkin

November 30, 2025

Kyiv

Content

1. Executive Summary	3
2. Dataset Description and Methodology	4
3. Analysis and Key Findings	6
4. Conclusions	16

1. Executive Summary

Using Reddit comments/submissions [dataset](#) (2020.09-2021.01), we analyzed comments and submissions from these subreddits:

Category	Subreddit
Anti-Trump, Liberal	r/politics, r/democrats, r/politicaldiscussion
Pro-Trump, Conservative	r/conservative, r/republicans, r/conspiracy
Neutral	r/news, r/worldnews, r/neutralpolitics

Main research questions:

- What events created the most active discussions?
- Did liberal communities dominate during the election and did activity shift to conservatives during the events of January?
- What was the emotional response of different groups during events?
- What groups were the most united and when?

Top-level findings:

- The Conservative community displayed division and weakness in its response to the January events.
- The Neutral community functioned as a source of objective information during periods of high polarization.
- The Liberal community demonstrated unity and consistency, allowing it to effectively dictate the dominant narrative and act as the leader of the situational response.

Why they matter:

- Liberals benefit from pushing ideas through the largest channel r/politics, which translates most of the narratives and has the biggest liberal community. That means we can use this subreddit to analyze general liberal activity in the future.
- Political campaigns should focus on the critic of opposition, as it drives the discussion more than stating achievements or responding to critics.
- Conservatives are very decentralized, which means that creating a big subreddit to consolidate republicans and push narratives through it will

make reddit more republican biased, with a big volume and strong narrative.

2. Dataset Description and Methodology

Data source and specifications:

- **Source:** raw [CSV](#) data (processing pipeline available [[here](#)]).
- **Timeframe:** October 2020 – March 2021 (6 months).
- **Volume:** 6 million records (approx. 2 GB).
- **Key Dimensions:** subreddit source, timestamps and textual content (submission titles and selftext, comment bodies).

Transformations applied:

- Converted Unix timestamp to datetime object for performing dates operations.
- Replaced damaged data with empty strings.
- Removed the most frequent words that were considered as spam and not relative information.
- Removed short comments less than 7 characters.
- Selected keywords related to research.
- Combined all text columns in one, converted to lowercase and removed spaces.
- Divided our initial dataset to comments and submissions datasets
- All in all, we reduced the size of the dataset by 6 times.

Methods used:

- Lemmatization at preprocessing step for further methods except from VADER sentiment analysis.
- Bigram creation using sklearn library to select main discussion topics.
- Overall sentiment analysis using VADER.
- Bigram sentiment analysis using VADER to show differences in reaction between groups.
- Topic analysis for finding groups by similar discussion.
- Jaccard similarity was calculated to quantify the divergence in vocabulary between opposing political groups.
- Measuring the volume of words category.

Limitations of the dataset and approach:

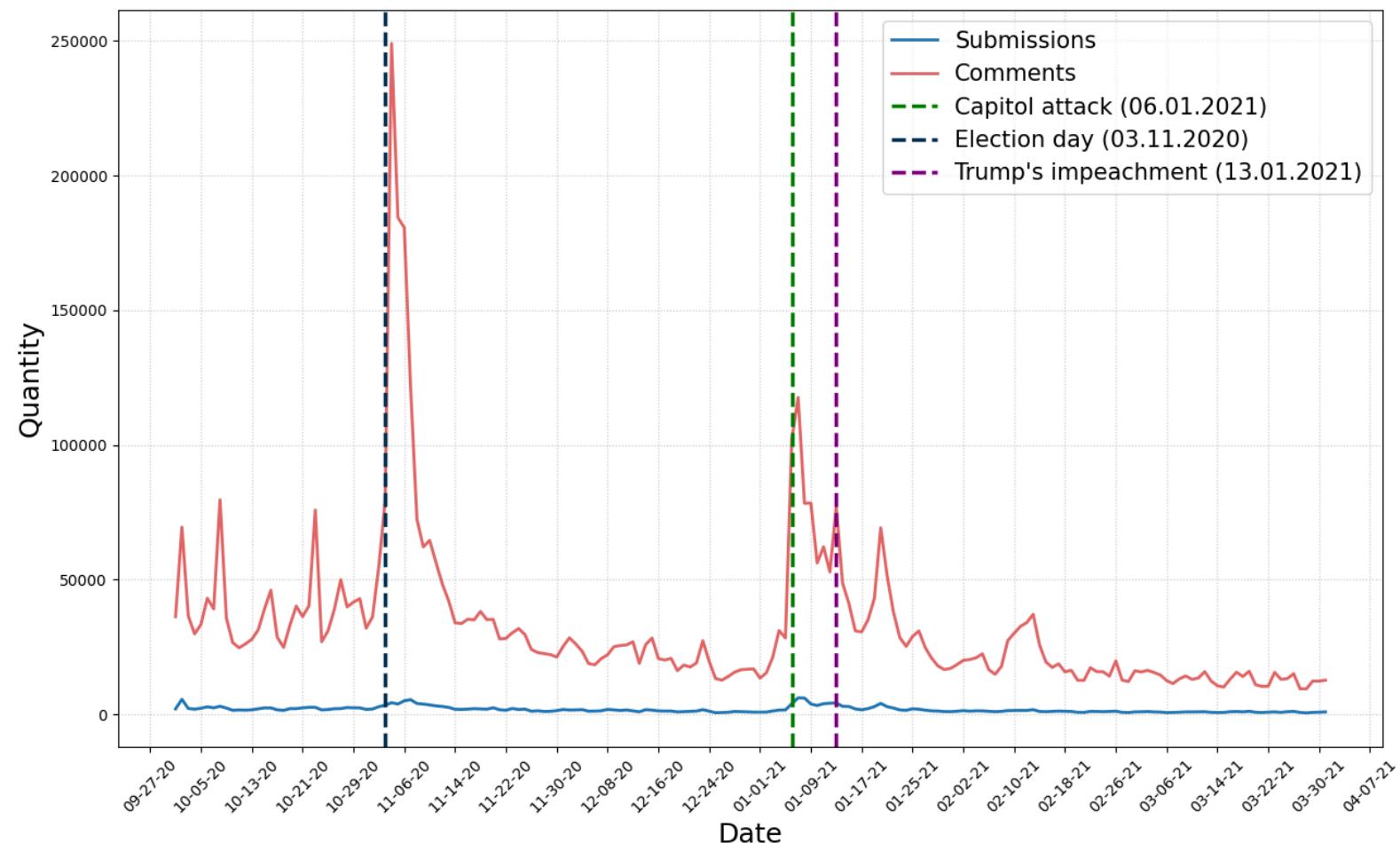
- **Demographic limitations:** data from dataset reflects only Reddit users' views. This platform is usually used by young people which is not representative of the US population.
- **Echo Chambers:** detected polarization between isolated and ideological clear subreddits (like r/Conservative and r/Democrats) may be exaggerated and may not reflect real US political discourse.
- **Author's unreliability:** reddit allows anonymous accounts, so it may be complicated to detect bots that post one and the same. (We tried as we can to remove all suspicious data from our dataset)

3. Analysis and Key Findings

What specific events cause the most active discussions and does the nature of the event change the volume of engagement?

Insight: The 2020 elections caused the most active discussions in the comments section, where users argued with each other. In contrast, the Capitol attack generated the biggest spike in the submissions, which almost consisted of titles about attack.

Figure 1: Daily activity volume on key political events

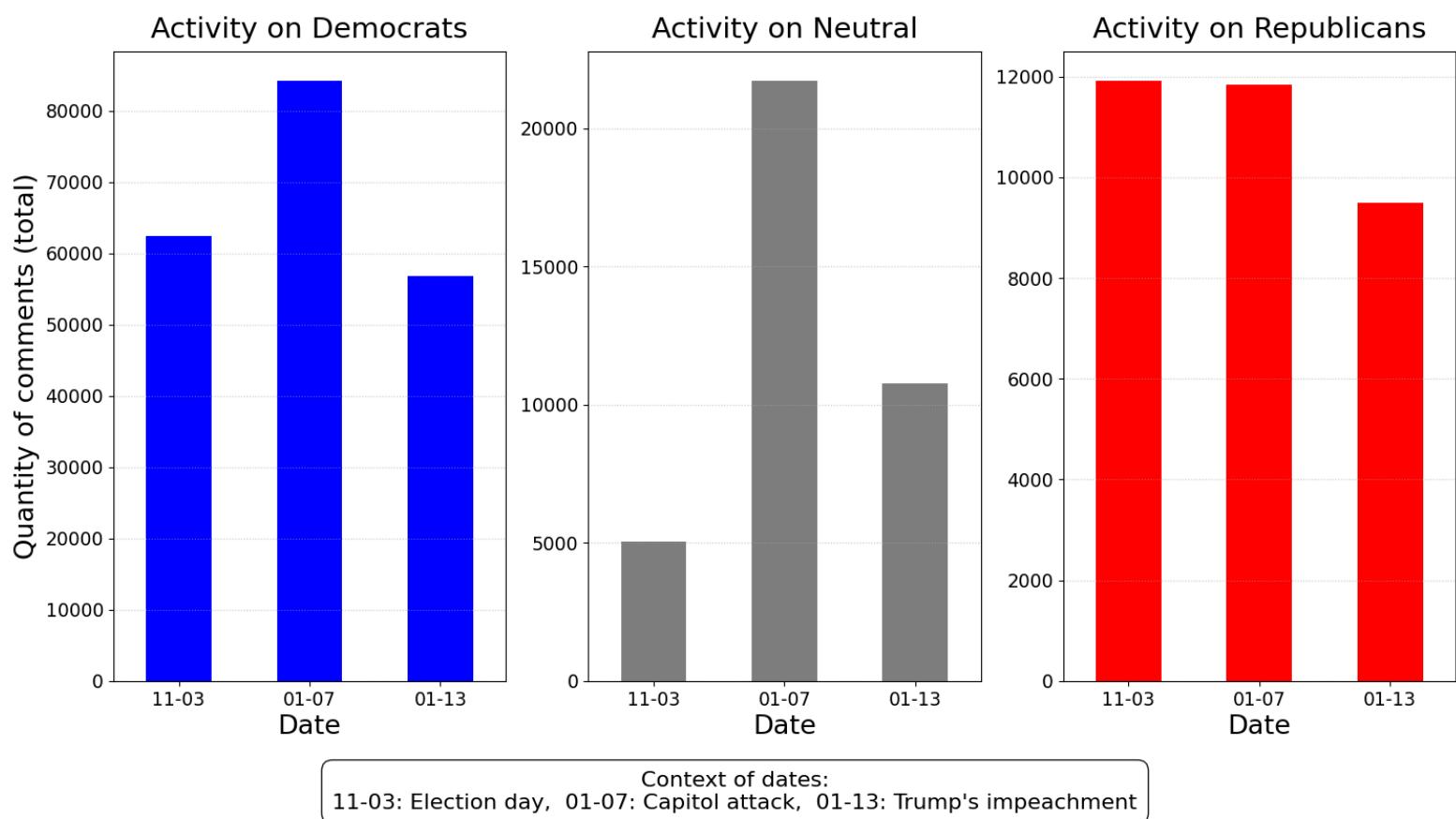


Meaning: There are huge peaks during major US politics events, the biggest are elections, Capitol attack, and Trump's impeachment. We can state that Reddit is a place where politics is discussed actively and the opinion of the target audience is represented in this social network.

What subreddits were the most active during what periods?

Insight: The 2020 elections topic was most popular among radical democrats and republicans. While neutral group, which was not influenced by right or left leaning sources, was most active during the Capitol raid and Trump's impeachment.

Figure 2: Distribution of comments among different political groups



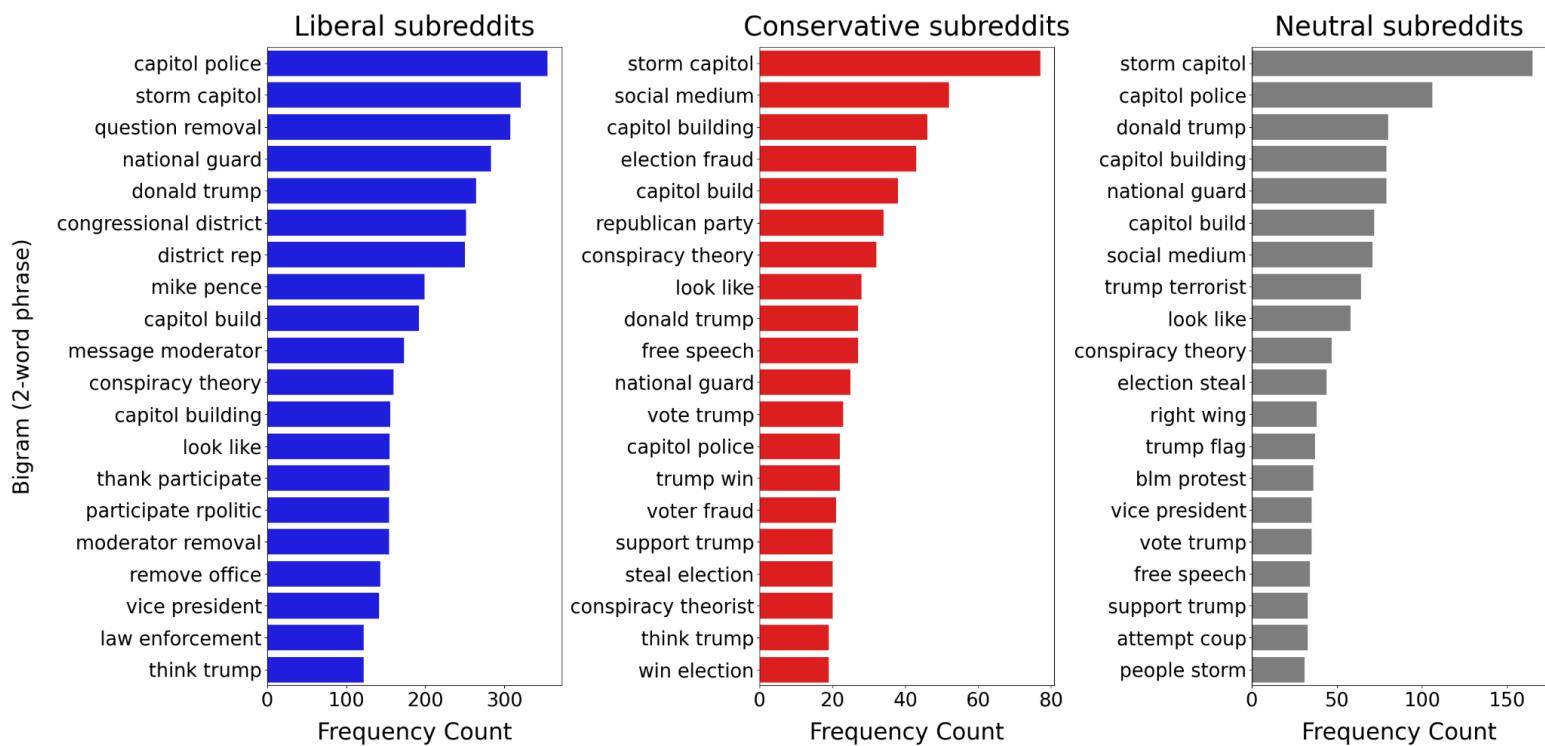
Meaning: By knowing this, we can state that the users without political leaning have more interest in extreme events and for leaning the neutral group, politicians must focus their messages on stability, democracy and crisis strategies rather than political ideology.

Contrary, for radical groups, elections were the most important event, so for mobilization of this group we need ideologically radical, clear messages.

What bigrams were the most frequently posted in the comments section?

Insight: During analysis we found out that republicans were commenting such bigrams as "conspiracy theory" and "election fraud". Liberals and neutral groups were focusing on facts and discussing the event and consequences.

Figure 3: The most frequent bigrams in comments on Capitol Attack day

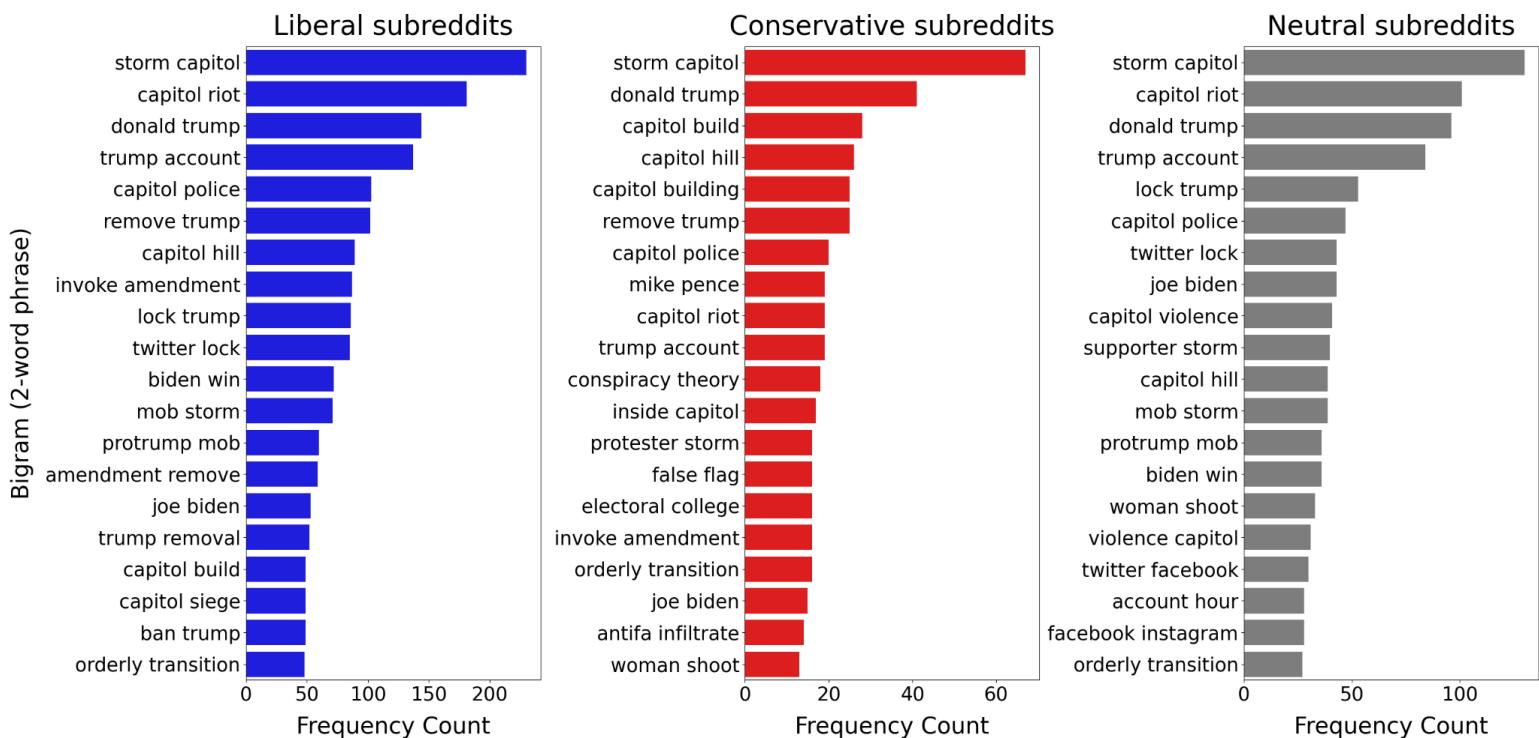


Meaning: Terms like “election fraud” and “conspiracy theory” are powerful points to mobilize conservatives. Liberals and neutral groups are concerned with law and consequences, so the best topics to mobilize this group are “capitol police” and “trump terrorist”.

Are comment bigrams influenced by submission bigrams?

Insight: We can see that reddit submissions bigrams are referring to general description of events rather than spreading the certain narrative. So users are forming their opinions in comments based on their beliefs, which are not influenced by submission content.

Figure 4: The most frequent bigrams in submissions on Capitol Attack day

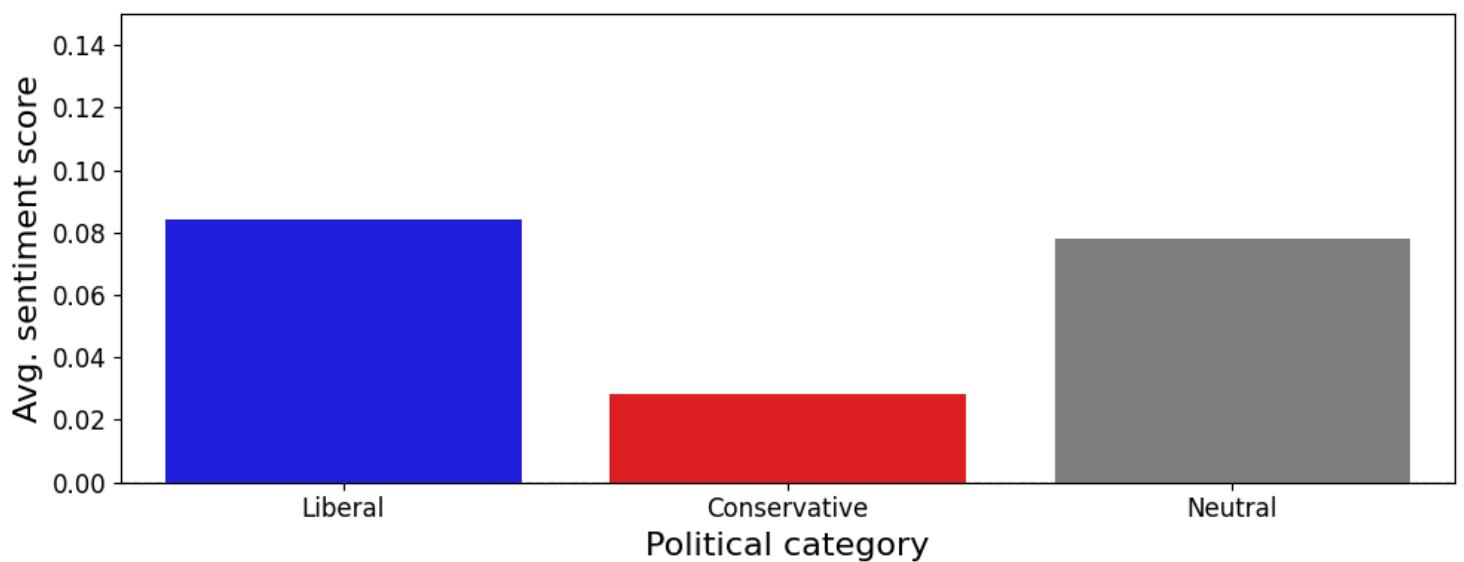


Meaning: On Reddit users' opinion is not influenced by subreddit submissions content, which means Reddit users choose the community approval of their thoughts rather than getting factual information.

How have users' sentiment changed during these events?

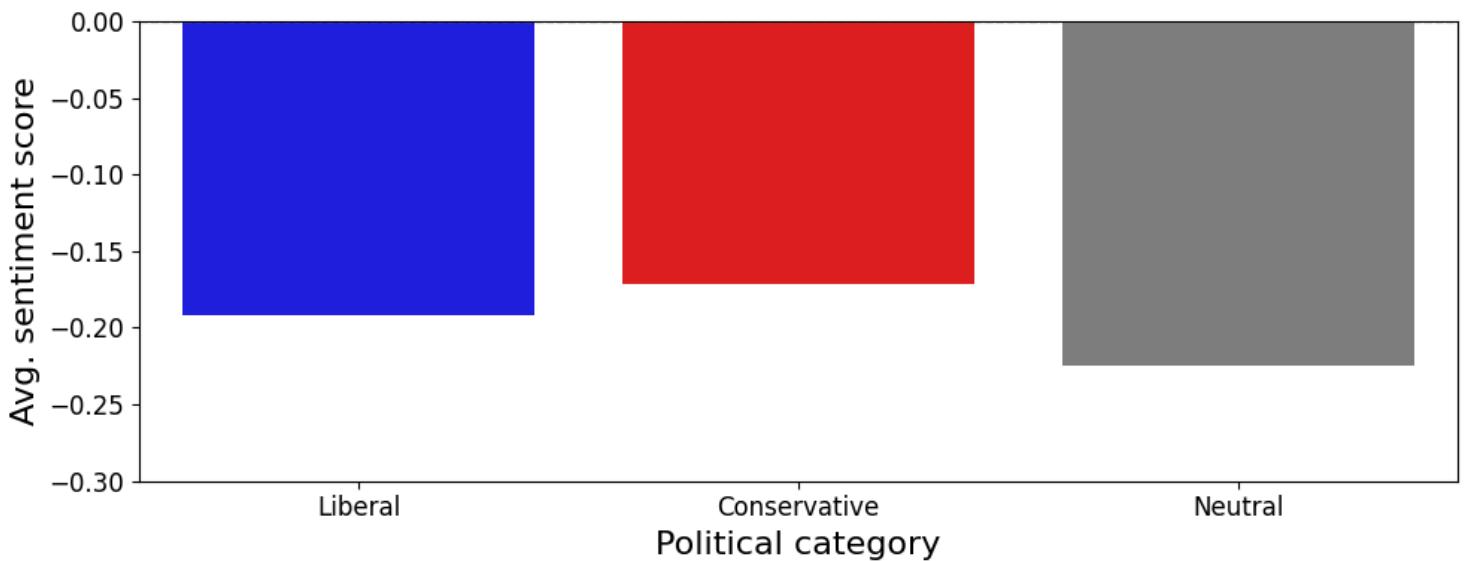
Insight: During Elections all groups had positive sentiment scores. Despite Donald T. lost elections, republican Reddit users from the “Conservative” group stayed about zero in sentiment score, which shows us that Republican discourse wasn't completely toxic.

Figure 5: The average sentiment of comments during Elections period



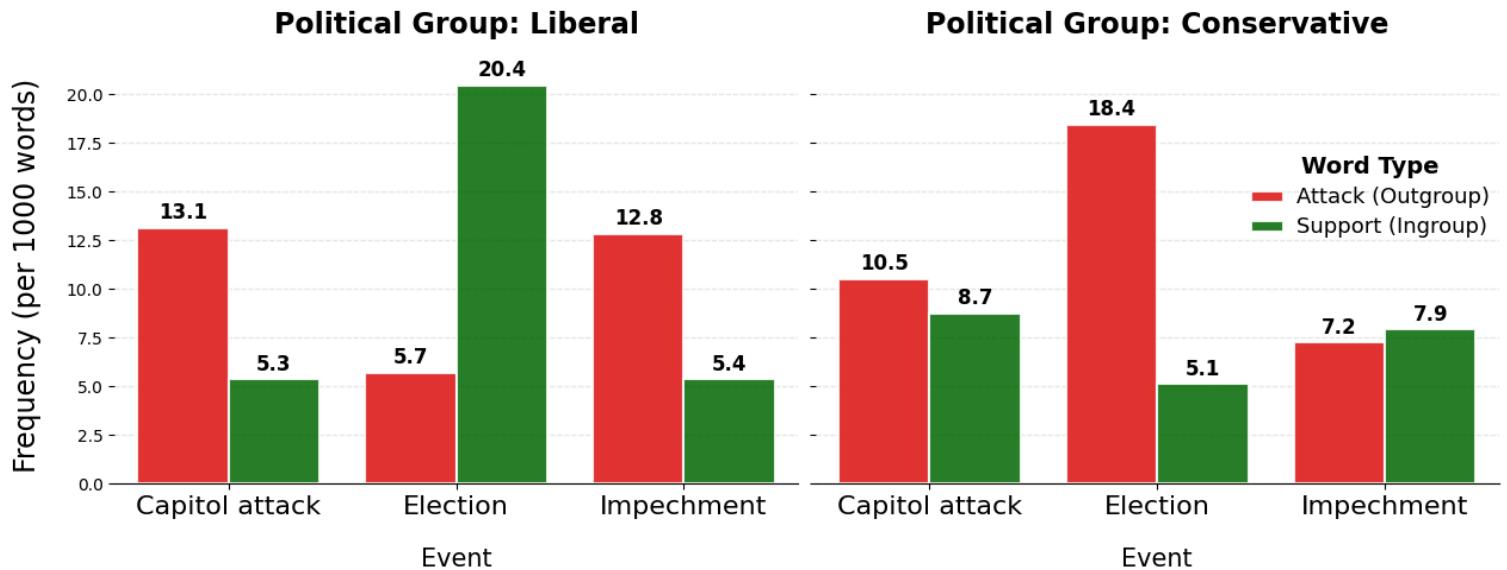
Insight: However, the capitol attack collapsed all good sentiment scores. Even Neutral sentiment score became the least from all groups. It proves that the shock of the event pushed back the majority of users more than republican-radicals.

Figure 6: The average sentiment of comments on Capitol Attack Day



Insight: So why was the sentiment score so negative? The data shows that “attack words” (negative words used against political opponents) were used much more frequently than “support words” (positive terms used for one's own group). This imbalance explains the overall negative sentiment in the discussions.

Figure 7: Both parties criticize opponents more than supporting their own side



General insight: We can see that during normal political processes every group has positive sentiment, they discuss topics from a positive side. After extreme unpredictable events users tend to speak negatively about topics.

Meaning: During crises, engagement is powered not only by shared values, but by shared enemies. The volume of “attack words” consistently overtakes “support words”. It proves that people on Reddit are mobilized more by hating the enemy than by supporting their own side.

How did the core topics of discussion change from the Elections to the Capitol Attack?

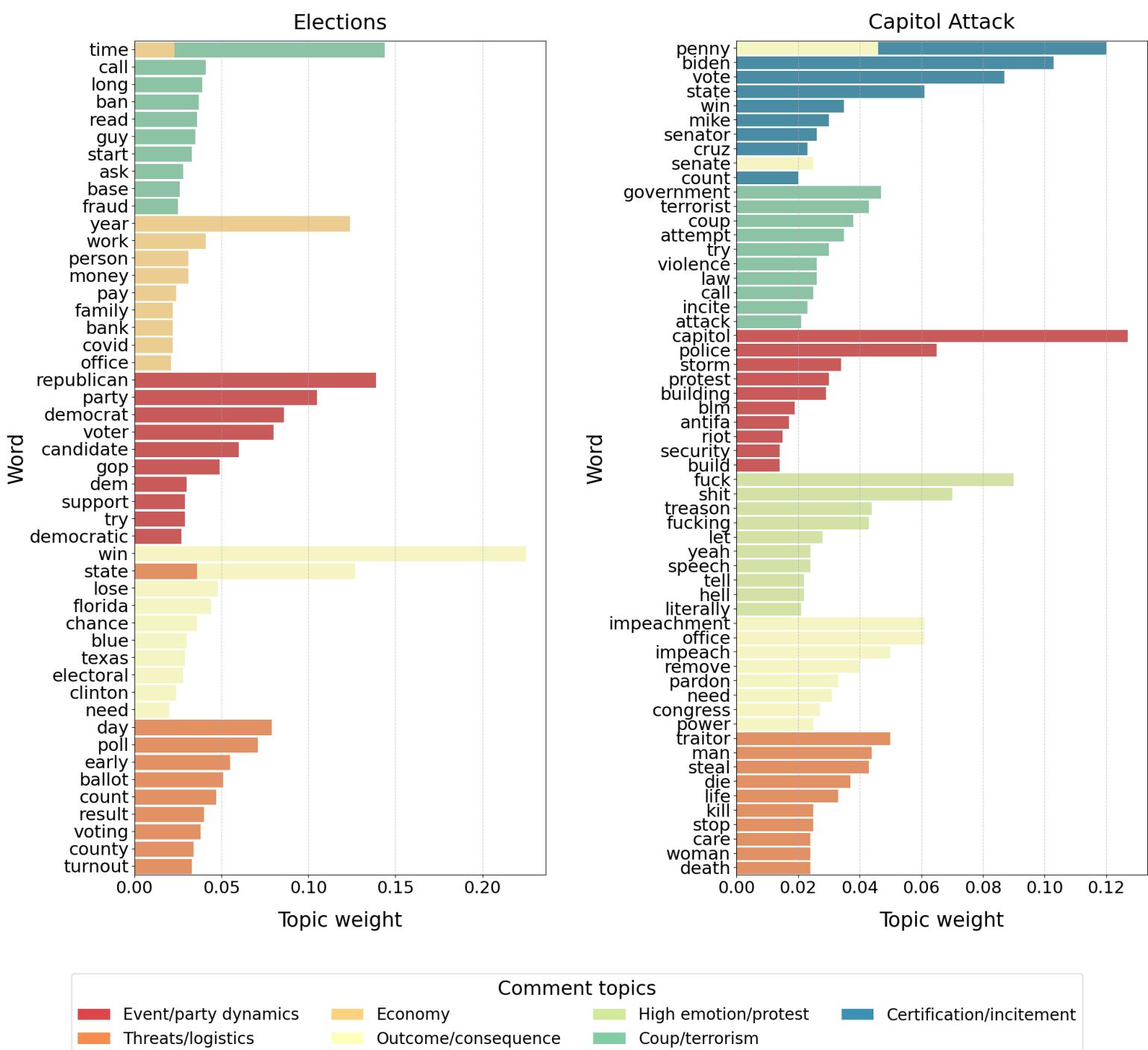
Insight: Topic modeling reveals a shift from default process to crisis.

The discourse around the elections was simple. Phrases like “win”, “poll” and “ballot” tell us that it was a common election with no surprises.

However, after capitol attack usual lexis changes cardinally to violent.

Comments were filled with words that describes crime (“terrorist”, “coup”, “treason”) and legal consequences (“remove”, “impeachment”)

Comparison of dominant keywords during key political events in comments

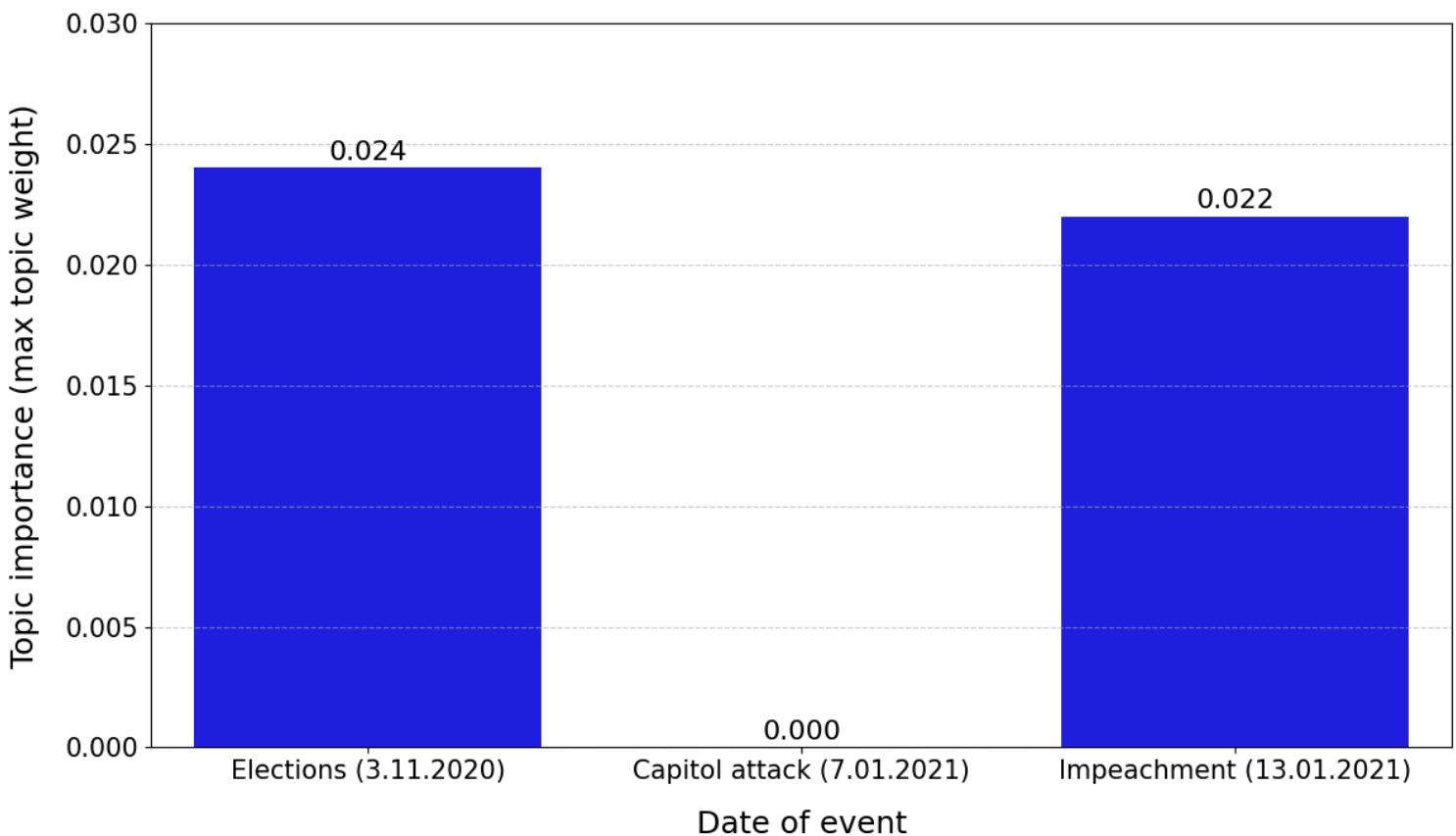


Meaning: This proves that during the capitol attack Reddit not just reacted with simple emotions, they made this event a global-level crisis followed with “appropriate” high emotional words. The Reddit community defined this event as a crime against all country rather than a common political protest.

What significant topics disappeared from discussion when the Capitol Attack occurred?

Insight: The capitol attack event was so important that it overshadowed other important things that time, for example, Covid-19. There were significantly fewer mentions of Covid-19 and words related to it than during elections or impeachment.

Weight of topics "Covid/pandemic/vaccine"



Meaning: This proves that the capitol attack was an event that almost consumed everything. Even a global crisis like Covid-19 was forgotten, as many other topics, by users.

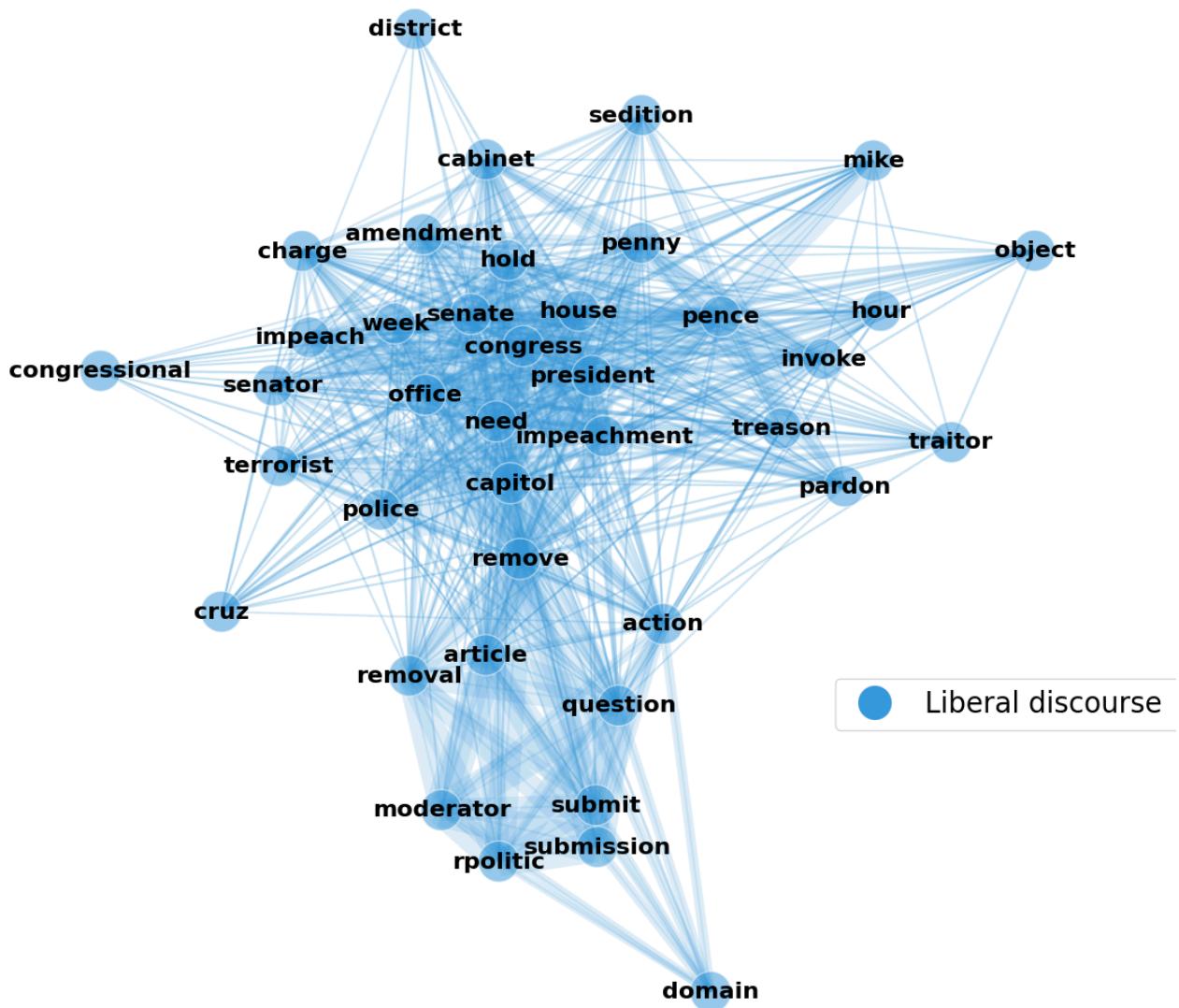
What words are used significantly more often by one group than by another?

Insight:

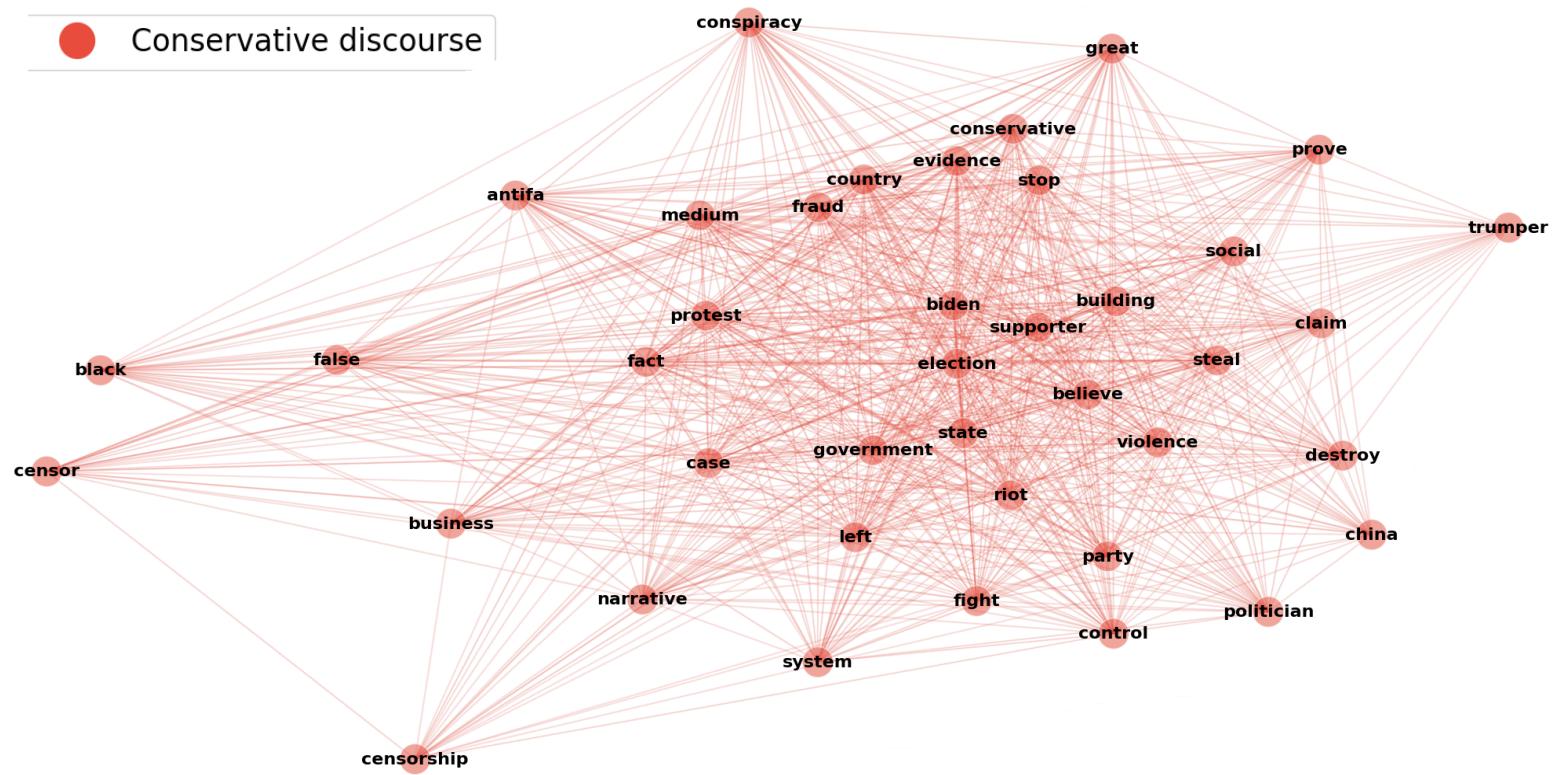
Liberal Discourse: High relations. The network is dense with strong connections, indicating a unified community consistently focused on specific events and a single shared narrative.

Conservative Discourse: Fewer relations between key terms. The network shows that topics are rather different than similar. This shows division inside the group and the lack of powerful unified narrative.

This reflects scientific researches about liberals and conservatives that they have different communication manners about political topics.



Conservative discourse



Meaning: This network graph visualizes the "semantic distance" between the two political groups. It forms two completely separate graphs with no connecting bridge between each other. Width of the edges indicates the frequency of the two vertices (two words) usage together.

4. Conclusions

To conclude, this research shows that Reddit political discourse is a place where events affect both the discussion topic and the sentiment. We found that critical events are getting the most attention and different groups inside of Reddit respond respectively based on their ideological beliefs. Reddit is a place where users share their opinion that is not influenced by submissions.

Although all groups used Attack words more, Liberal subreddits were united and consistent in their narrative, when conservatives showed fragmentation between ideas and did not manage to state a clear main narrative. The neutral group was the most active during extreme events, while being indifferent to standard events like elections. This tendency explains the negativity after the Capitol attack and impeachment.

However, some questions remain unanswered. How much of this behavior is generated by real users vs bots' generated content. We cannot also say whether Reddit's patterns are representative as a whole of US political attitudes or are mostly a feature of a young audience.

With more time given, further research would explore these problems by tracking bots activity and expanding on other online platforms to see whether these patterns and narratives are unique to Reddit or common to general political discourse.