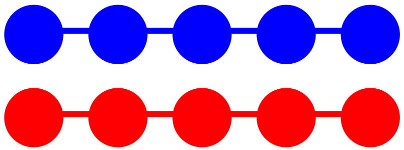
Echo Chambers, Filter Bubble and Polarization on Youtube: the USA’s post-electoral debate

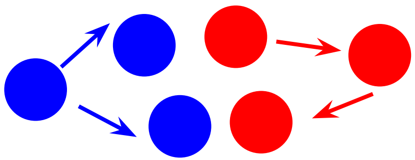
Salvatore Romano, Davide Beraldo, Giovanni Rossetti, Bruno Sotic,
Paul Grua, Armand Bazin, Maxime Bertaux, Youcef Taiati, Antonella Autuori, Andrea Elena Febres Medina, Wen Li, Inga Luchs, Annelien Smets, Lynge Asbjørn Møller, Alexandra Elliott, Matthieu Comoy, Ali El Amrani, Eirini Nikopoulou, Nicolas Pogeant, Yamina Boubekeur, Arthur Lezer, Mehdi Bessalah, Andrea Angulo Granda, Tcheutga Corine, Lisa Lan, Kaothar Zehar, Dong Pha Pham, Josue Charles, June Camille Ménard, Minhee KYOUNG,Hangchen Liu,Yiran Zhao

Theoretical framework

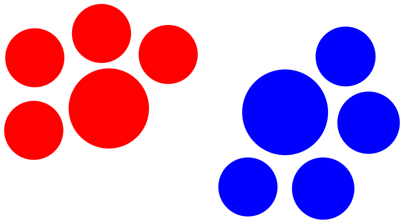
Echo Chamber → The political ideology of the subject and the consequent watching choices. Based on ideological affinity, are created by users (Dubois ’18).



Filter bubble → Direct effect of algorithmic personalization, based on the users’ behavior (Eli Pariser ’11).

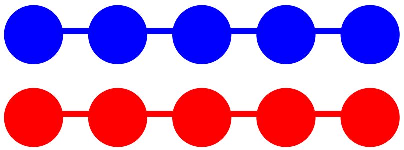


Polarization → Process of increased segregation into distinct social groups, separated along racial, economic, political, religious or other lines (Gallacher & Heerdink ‘19).

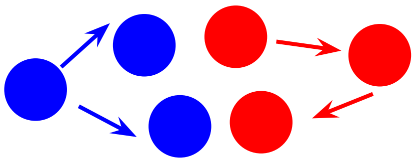


Theoretical framework

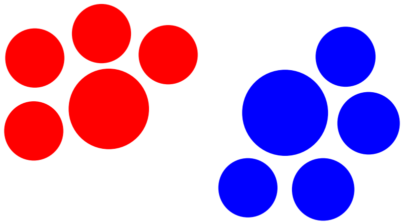
Echo Chamber → The political ideology of the subject and the consequent watching choices. Based on ideological affinity, are created by users (Dubois ’18).



Filter bubble → Direct effect of algorithmic personalization, based on the users’ behavior (Eli Pariser ’11).



Polarization → Process of increased segregation into distinct social groups, separated along racial, economic, political, religious or other lines (Gallacher & Heerdink ‘19).



Research Questions

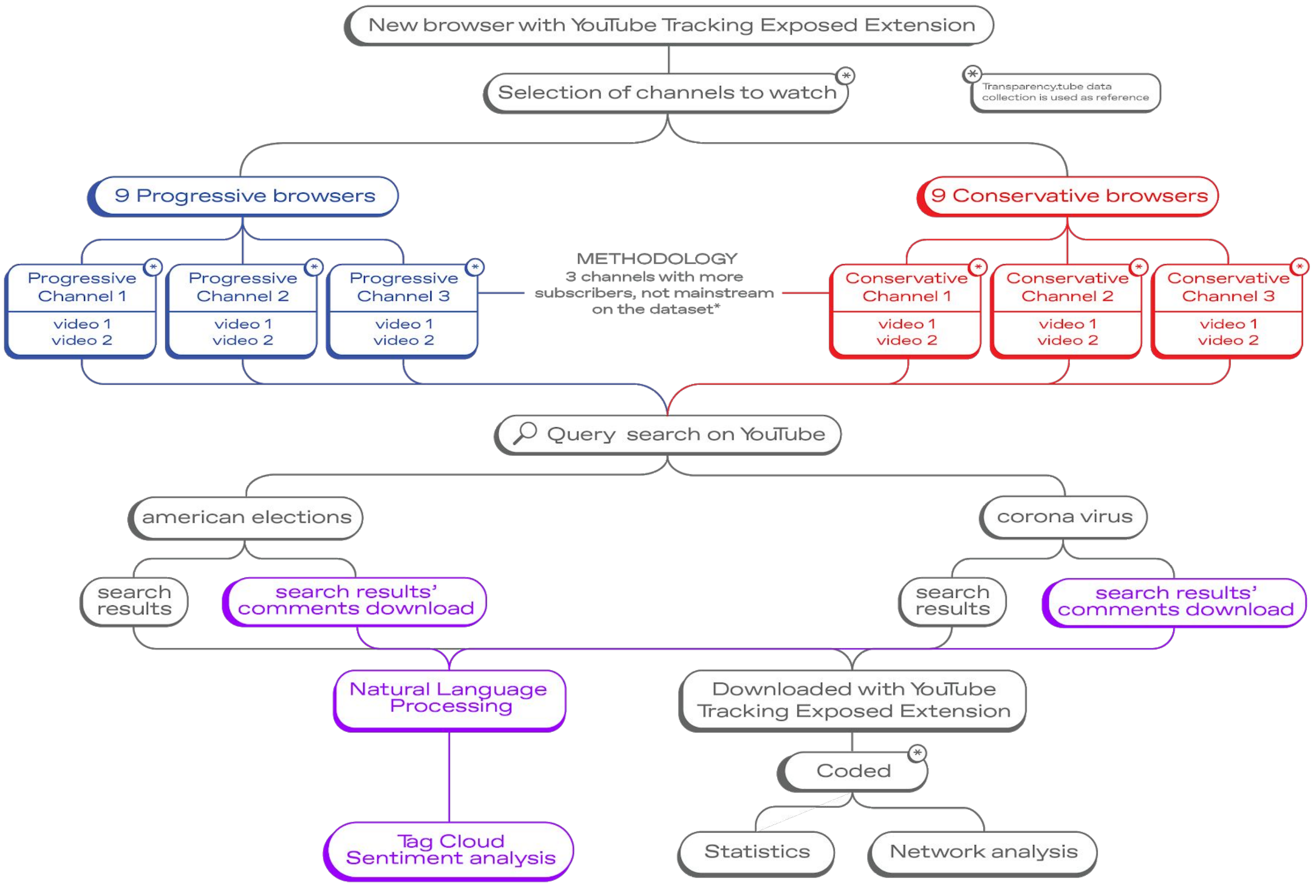
RQ: *Does YouTube's **algorithm** enforce a **filter bubble** and **polarization** patterns based on an (artificially generated) **echo chamber**?*

↓

Sub-RQ1: *Are there differences in the videos suggested as **search results** across different user types?*

Sub-RQ2: *Are there differences in **comments** to the videos suggested as search results across different user types?*

Methodology



Results overview

Users groups: Conservatives vs. Progressive
Queries: 'american elections'; 'corona virus'

Search results *(Visualizations + Networks)*

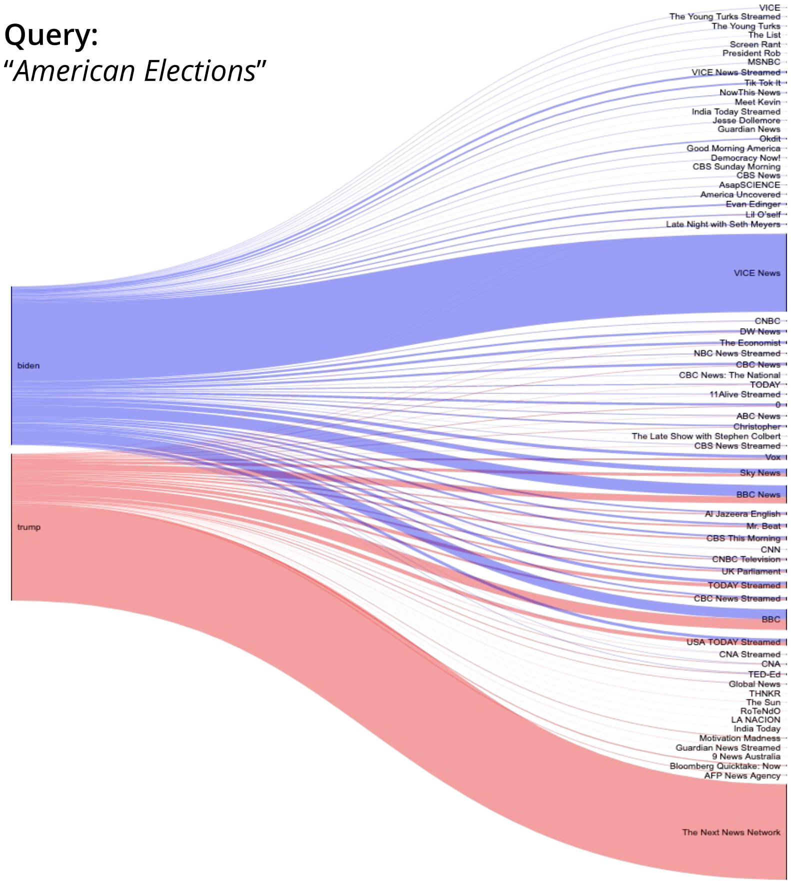
→ Media Type
→ Political Orientation

Comments *(Text analysis)*

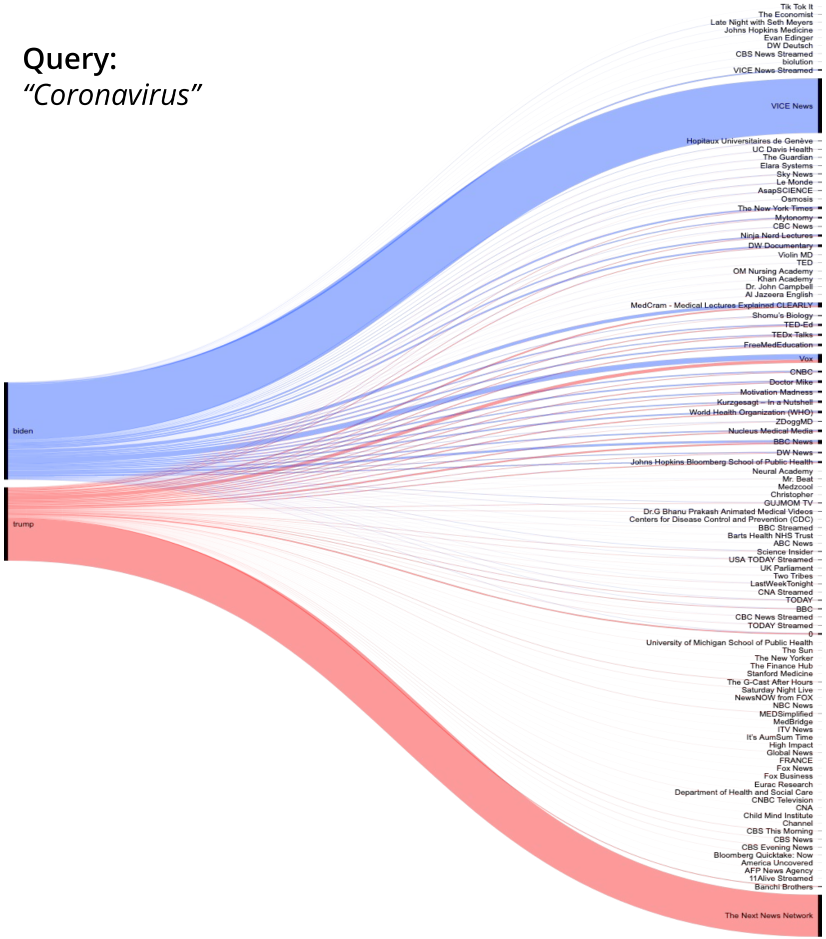
→ Content analysis
→ Sentiment analysis
→ Toxicity Index

Channels overlap across groups

Query:
“American Elections”

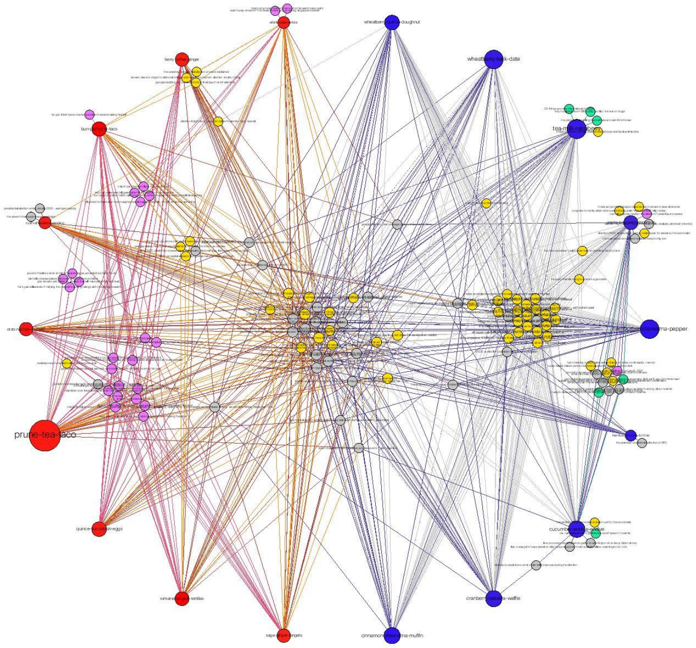


Query:
“Coronavirus”



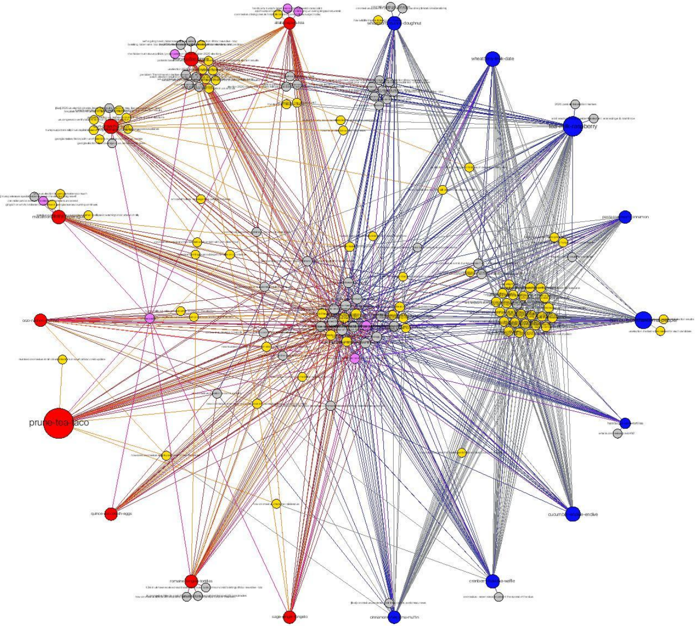
Search results according to *media type*

AMERICAN ELECTIONS_TITLE_MEDIA_FIRST 20 TOP VIDEOS



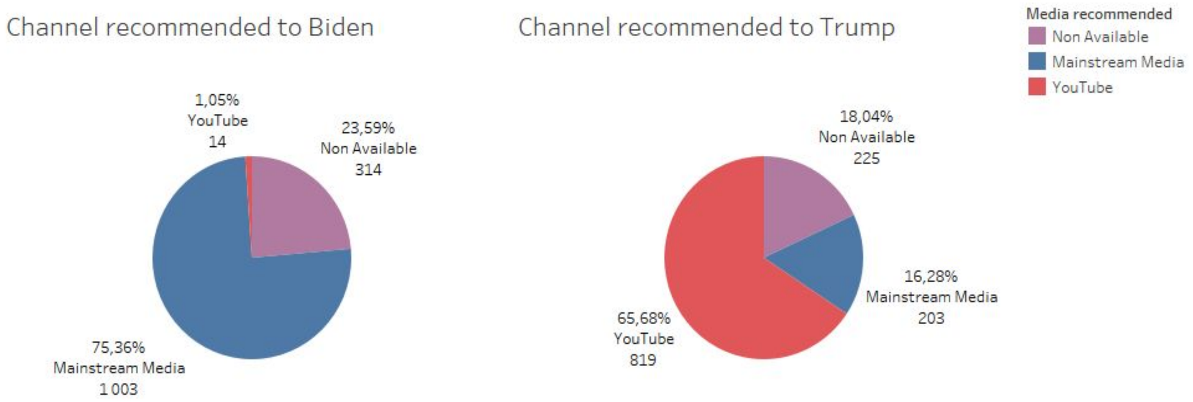
BIDEN PSEUDO
TRUMP PSEUDO
YOUTUBE
MAINSTREAM MEDIA
MISSING LINK MEDIA
UNCLASSIFIED

CORONA VIRUS_TITLE_MEDIA_FIRST 20 TOP VIDEOS

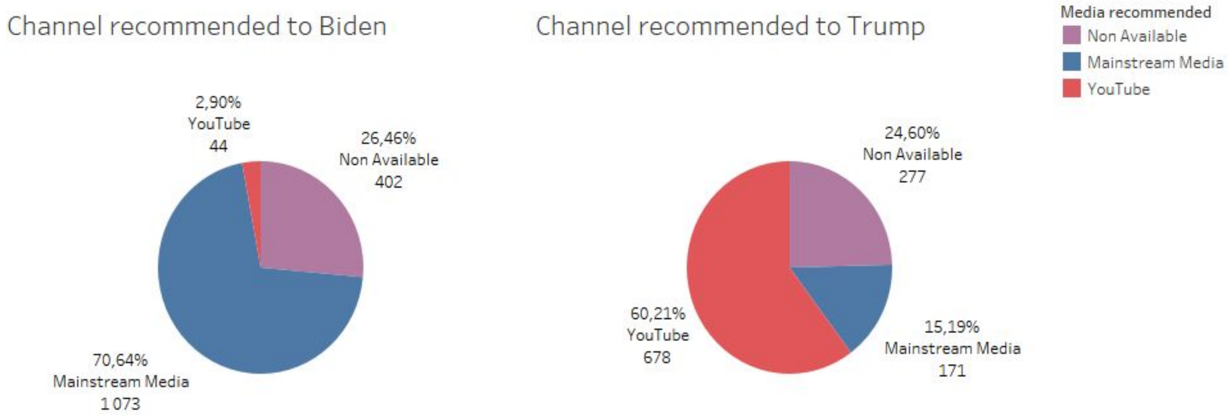


BIDEN PSEUDO
TRUMP PSEUDO
YOUTUBE
MAINSTREAM MEDIA
MISSING LINK MEDIA
UNCLASSIFIED

Query:
"American Elections"

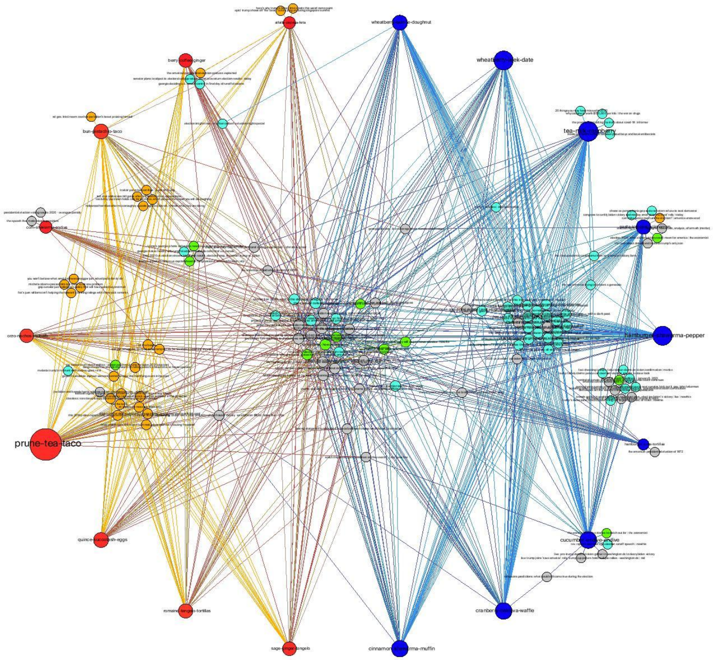


Query:
"Corona virus"



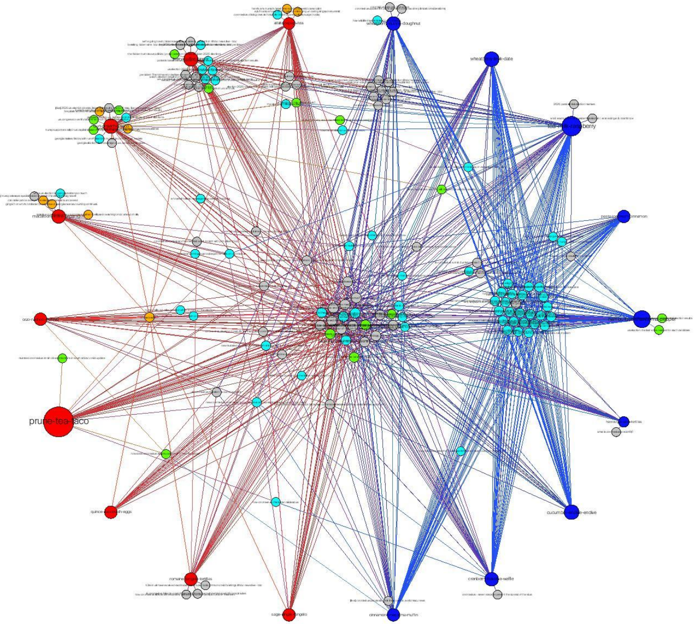
Search results according to *political orientation*

AMERICAN ELECTIONS_TITLE_ORIENTATION_FIRST 20 TOP VIDEOS



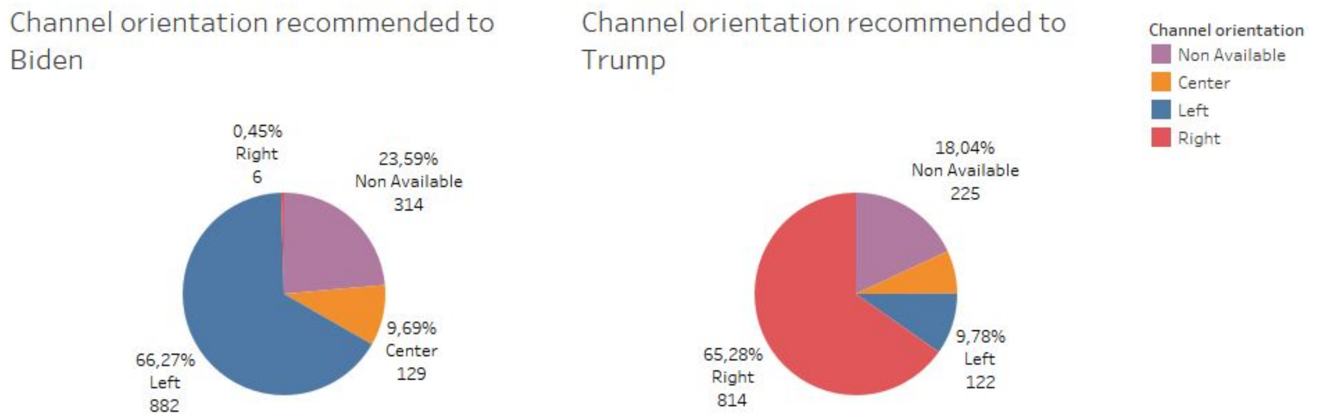
BIDEN PSEUDO
TRUMP PSEUDO
LEFT
RIGHT
CENTER
UNCLASSIFIED CHANNEL

CORONA VIRUS_TITLE_ORIENTATION_FIRST 20 TOP VIDEOS

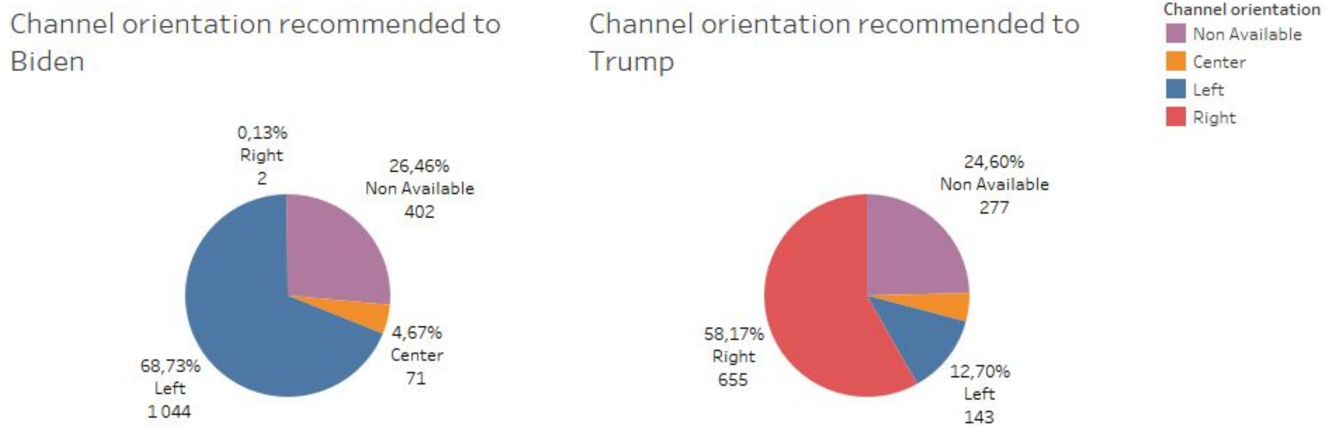


BIDEN PSEUDO
TRUMP PSEUDO
LEFT
RIGHT
CENTER
UNCLASSIFIED CHANNEL

Query:
“American Elections”

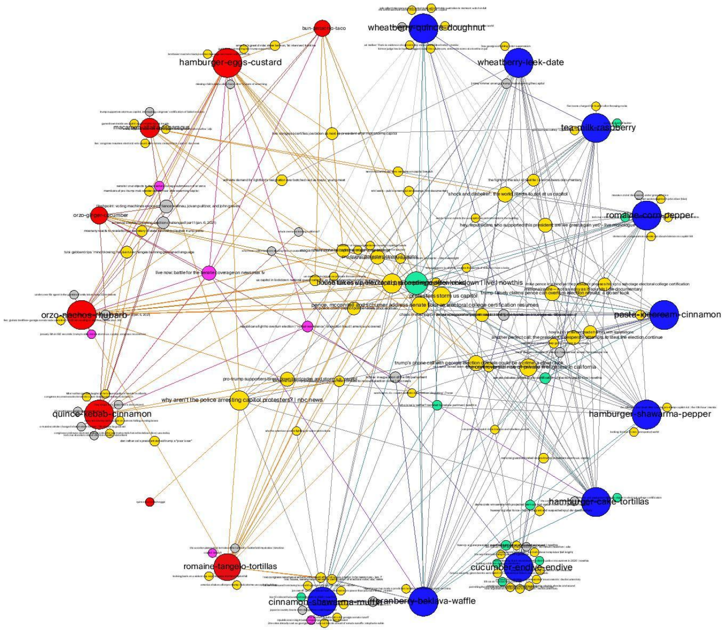


Query:
“Coronavirus”



Related videos: AP video about Capitol Hill

VIDEOAP_TITLE_MEDIA



BIDEN PSEUDO

TRUMP PSEUDO

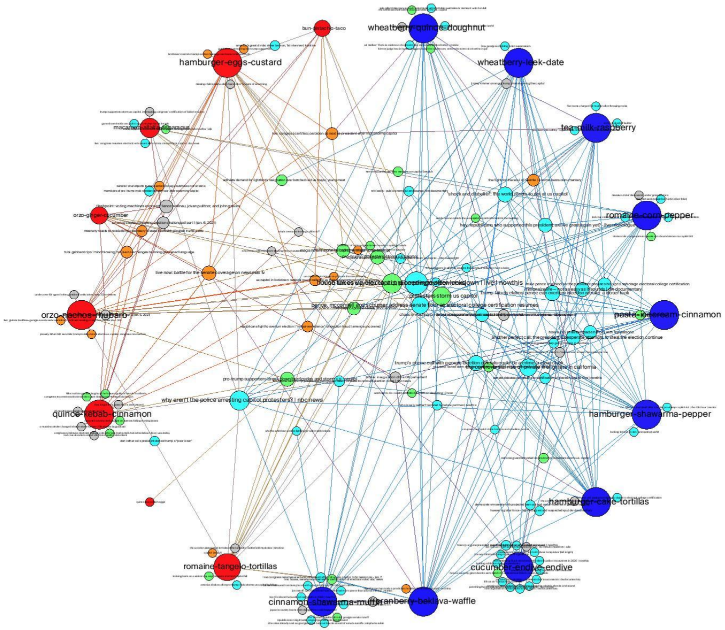
YOUTUBE

MAINSTREAM MEDIA

MISSING LINK MEDIA

UNCLASSIFIED

VIDEOAP_TITLE_ORIENTATION



BIDEN PSEUDO

TRUMP PSEUDO

LEFT

RIGHT

CENTER

UNCLASSIFIED CHANNEL

What’s in the comments ?

Let’s look at the comments of the recommended videos from each query with an NLP analysis.

→ Firstly, by looking at the most frequent words from each browser for each query.

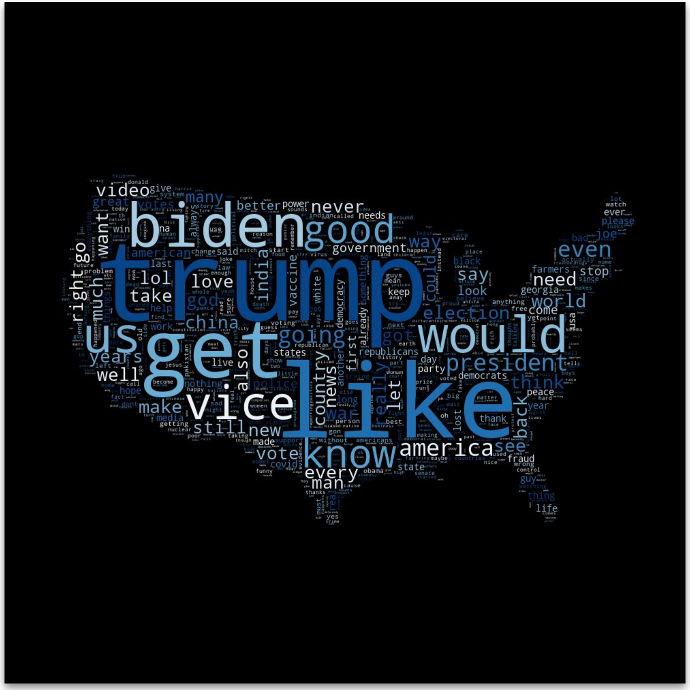
→ Then with a Topic Modeling approach (LDA) to see what comes out of these comments.

→ After that, a sentiment analysis with the sentiment polarity percentage (positive, negative, neutral) from these comments.

→ Finally, an aspect about toxicity in the comments between Conservatives and Progressives.

American Elections Query

The most frequent words of the youtube comments
(Progressives Browser)

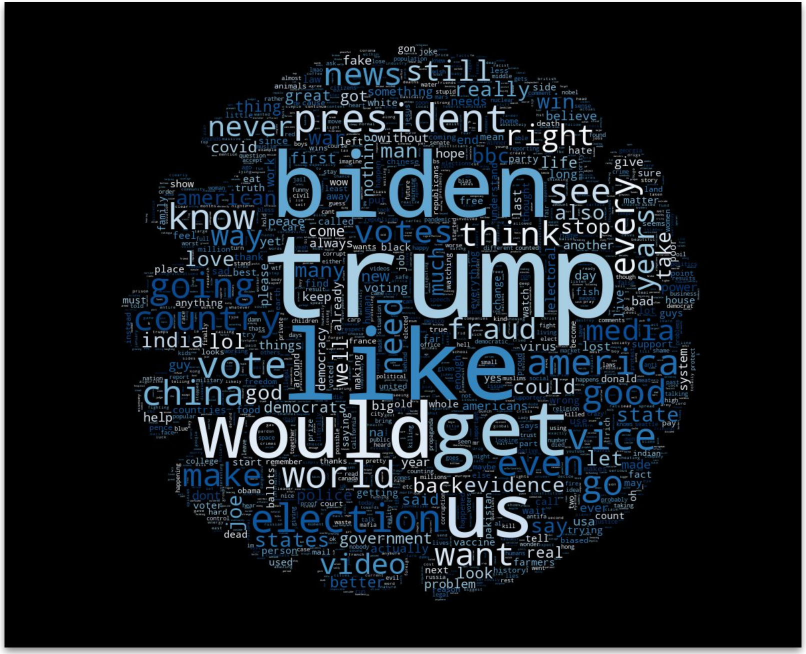


The most frequent words of the youtube comments
(Conservatives Browser)

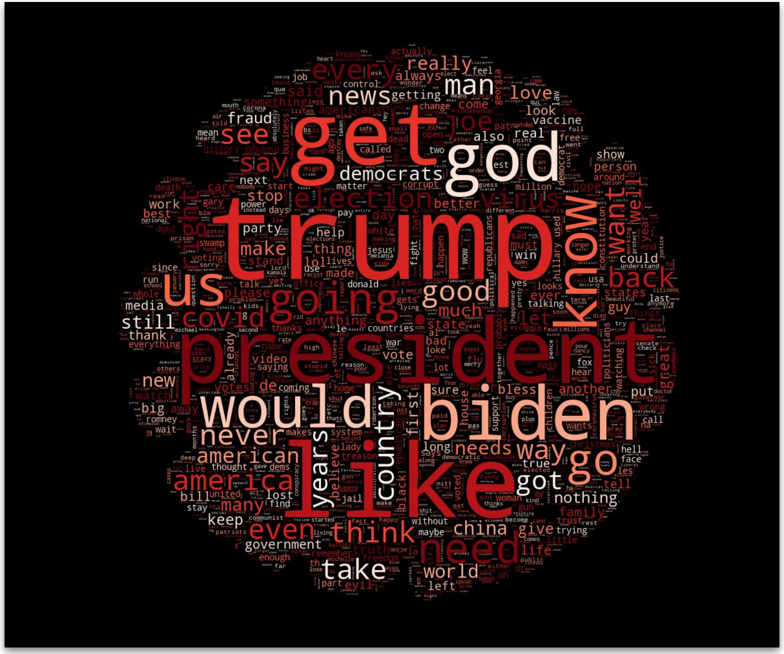
Data: The video comments come from recommended videos that are not in common between the two browsers

Corona Virus Query

The most frequent words of the youtube comments
(Progressives Browser)



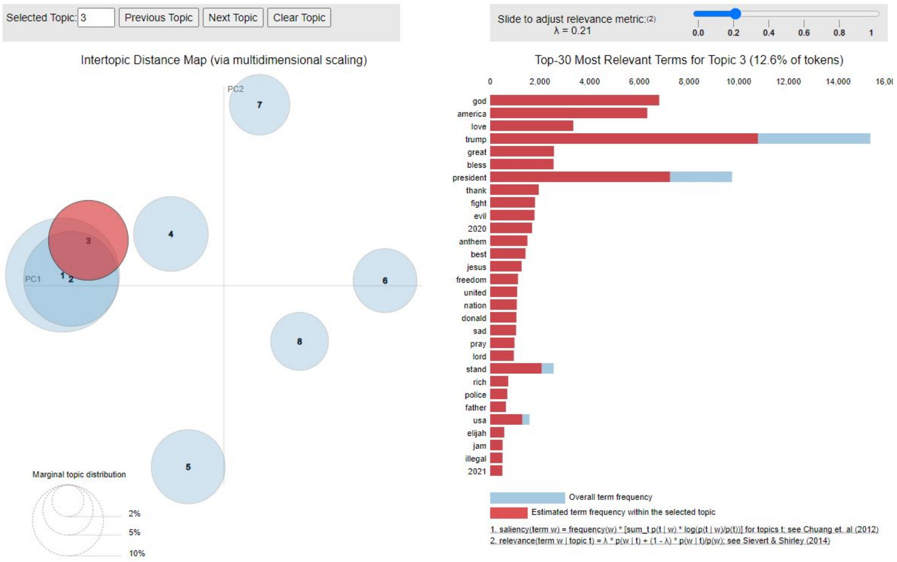
The most frequent words of the youtube comments
(Conservatives Browser)



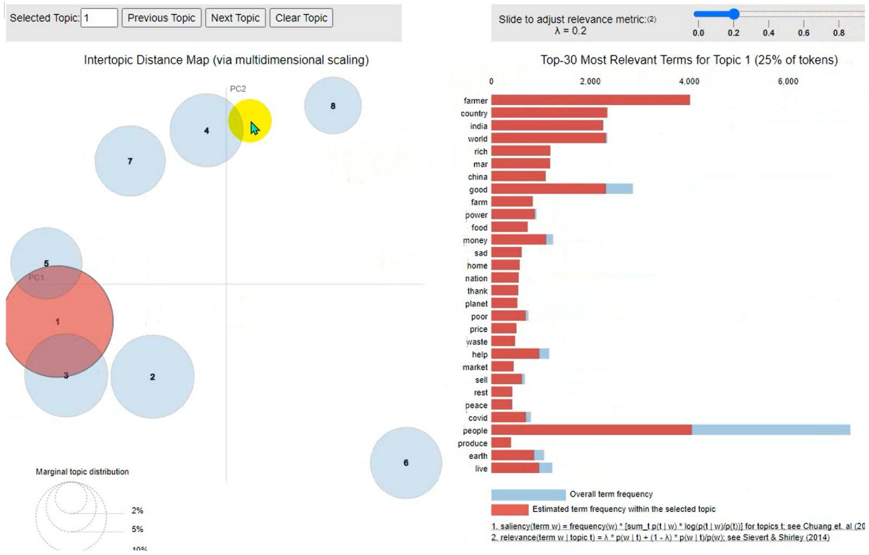
Data: The video comments come from recommended videos that are not in common between the two browsers

Topic modeling of the American Election Query

Conservative Side 14 topics



Progressive Side 8 topics



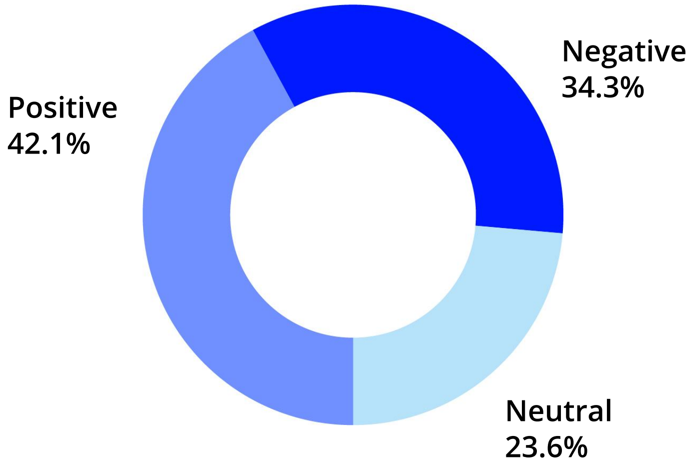
Common topics - vote, media, war, vaccine.

Exclusive topics

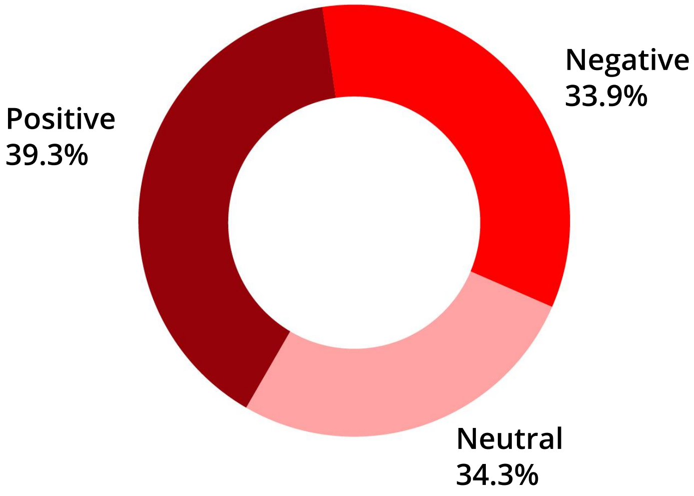
- *for Conservative side : religion, blm(trump supporters), gender issues*
- *for Progressive side : farmer, policy of candidates*

Sentiment of the American Election Query

Sentiment of youtube comments from the recommended videos (Progressives Browser)

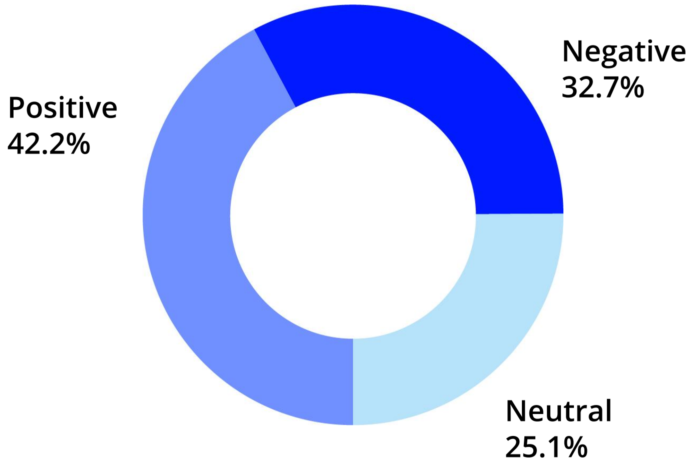


Sentiment of youtube comments from the recommended videos (Progressives Browser)

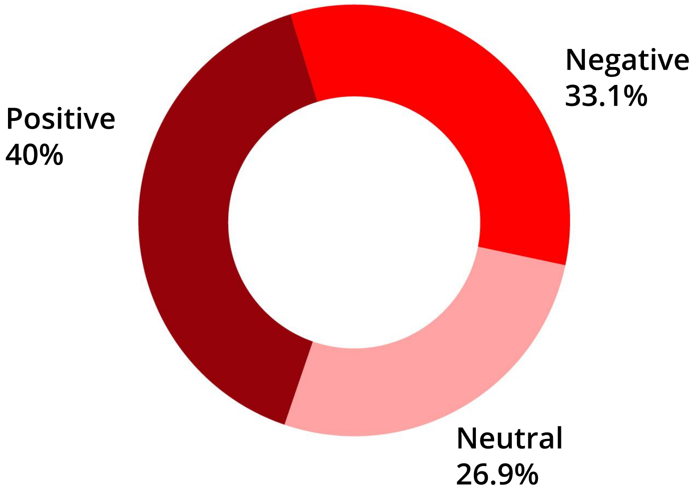


Sentiment of the Coronavirus Query

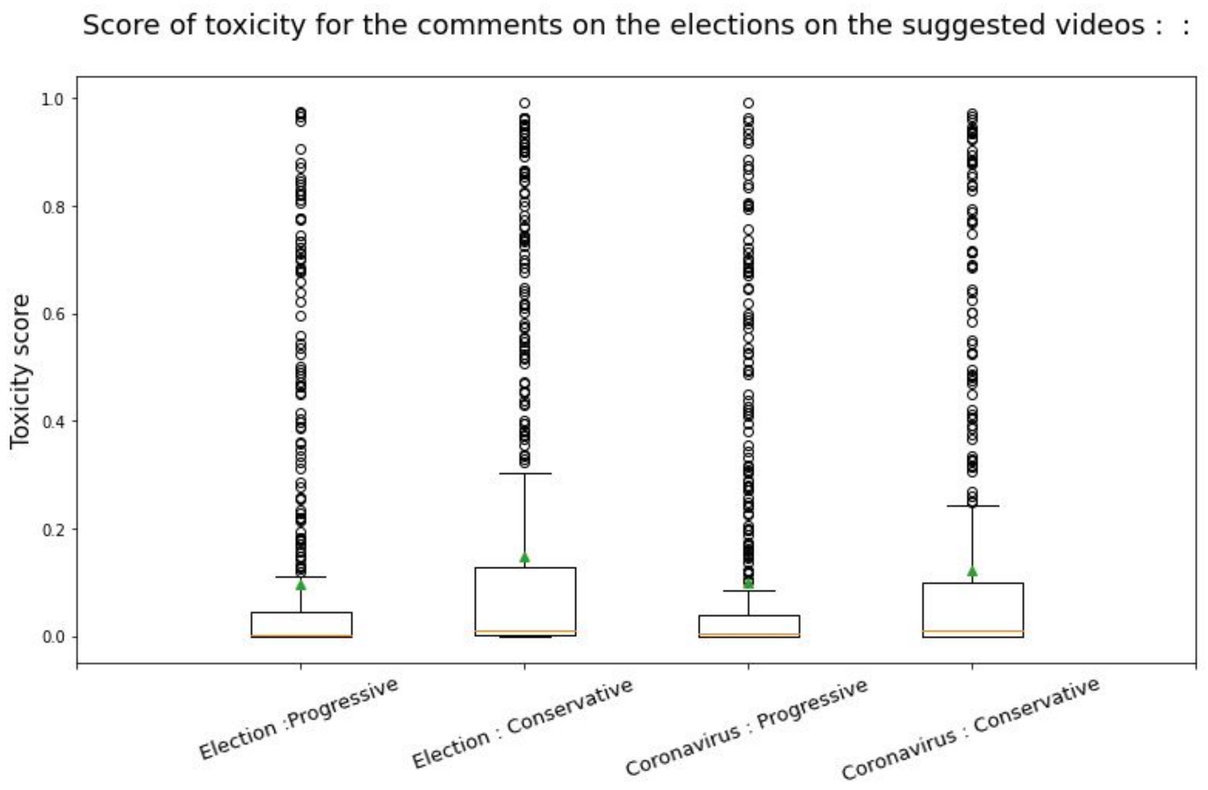
Sentiment of youtube comments from the recommended videos (Progressives Browser)



Sentiment of youtube comments from the recommended videos (Progressives Browser)



Toxicity Score from the two queries



↓↓ **Two examples:**

- ‘Trump make-up looks like someone spread some baby crap on his face' : score 0.9262657
- ' Waouh the project Echo-Chambers was so cool" score : 0.00063012395

The score is a prediction of toxicity based on the words used.
If words that are associated with swearing, insults or profanity are present in a comment, it is likely that it will be classified as toxic.
We can seen that half of the comments are considered as not toxic at all for progressive side and the conservative side. On the other hand for the other half, comments in the video suggested for the conservative browser are considered more toxic.

Final considerations - Conclusions

→ There is **some empirical evidence** of how the algorithm is **generating a filter bubble** possibly exacerbating an **echo chamber** behavior, both based on media type and political orientation of the channels suggested.

→ **However**, in our experiment a good part of this is due to **channels that have been watched** showing up in the results.

→ We could identify few interesting **differences** in the **topics** and in the degree of **toxicity** emerging from the **comments** (not in sentiment), that can base a study on users' **polarization**.

→ **However**, more **thorough sampling** and **analysis** is needed.

Final considerations - Further research

→ ***Scrolling through***: the more you scroll to include search results (e.g. after the top 20) the more results are influenced by the videos watched → collect more data by scrolling after the query.

→ ***Queries interaction***: when you perform queries in sequence, the previous queries interfere with the latter ones (e.g. election-related results in “corona virus” query because of previous query “american election”) → take this into consideration when comparing queries.

→ ***Number of videos watched***: the more videos a user watches, the clearer the results → simulating a stronger echo chamber by using automated users to watch several videos.

Baby Sharks say: “ciaooooo”