POLARIZATION AND TOXICITY

DURING THE LATEST
PRESIDENTIAL ELECTION

THE IDEA:

- Now more than ever, YouTube is seen not only as a platform where to share and watch videos, but where exchanges between users can be had
- Political discussions and conversations can be seen flourishing under specialized news outlet's videos
- Threads under videos from these news outlets, belonging to the entire political ideology spectrum, can offer us a view on how users belonging to different political leanings interact with each other

THE QUESTION:

By analyzing such YouTube conversations, is it possible to identify a signicant correlation between their polarization and their level of toxicity?

WHICH COMMENTS?

- Comments under national, local and independent news channels' videos, together with organizations' channels videos, during the latest US Presidential Elections of 2020, which could be considered as fertile ground for political debates

GETTING THE DATA:

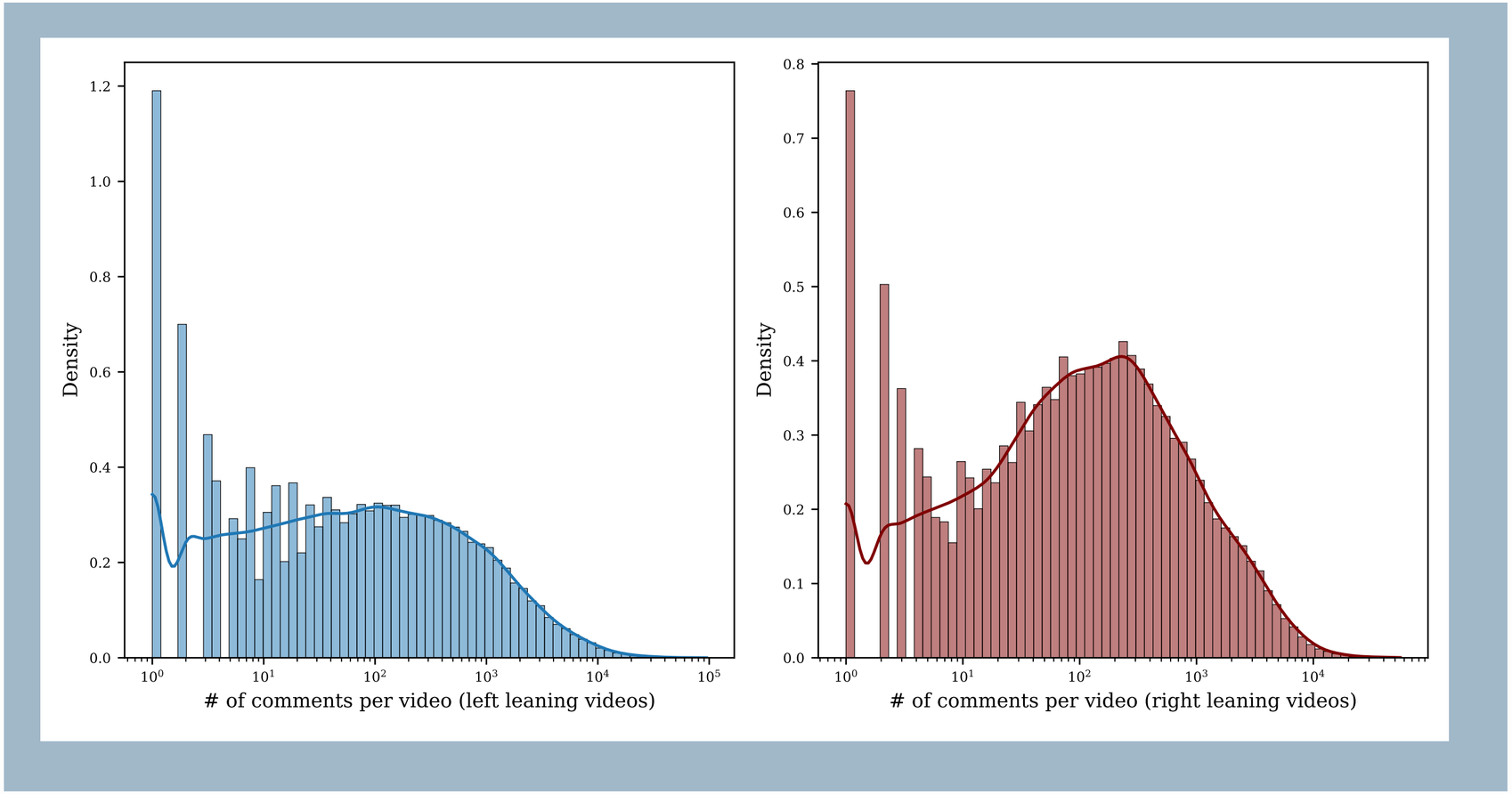
- Almost 1300 channels considered, classified either as left, right or center leaning
- A total 230 thousands videos uploaded between January and August 2020
- Full scrape of each video (root comments + replies)

RESULTING DATASET:

- 166k videos were successfully scraped, retrieving almost 80M comments coming from more than 9M unique commenters

	Left leaning	Center leaning	Right leaning	Total:
Channels	427	165	419	1011
Videos	88k	22k	56k	166k
Comments	43M	7M	28M	78M

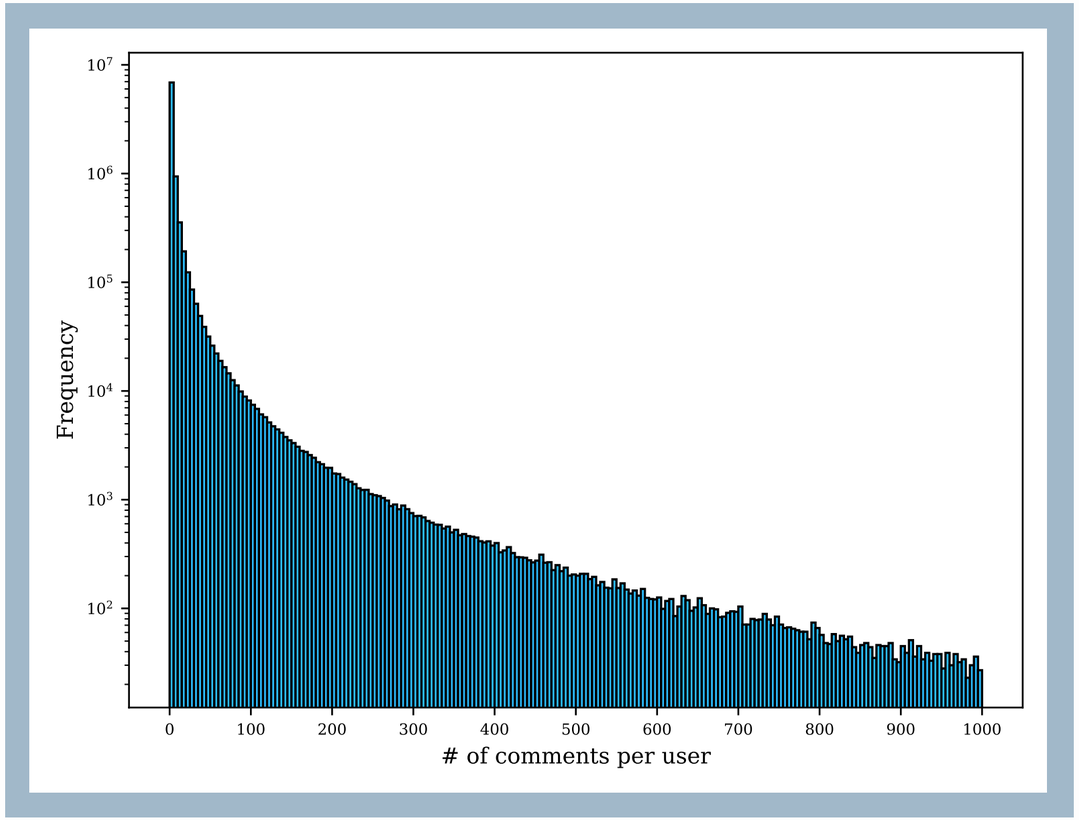
COMMENTS PER VIDEO:



- Even though there are more videos from left-leaning channels (and thus more total comments), the right-leaning videos have on average more comments per video (mean = 500, median = 91 vs mean = 493, median = 48)

COMMENTS PER USER:

- 48% of the 9+ million users only have a single comment throughout every video
- "Active" users (those with at least 5 comments) are 24% of the total



- While the average number of comments per user is 8.7 (median = 2), the distribution decreases in an exponential fashion

MODELING STEPS:

Labeling active users' leaning:

- Averaging the number of comments under left, right and center leaning channels

---

Scoring comments' toxicity:

- Google's Perspective API

Measuring toxicity of a thread of comments:

- (avg. of the ten percent most toxic comments in thread) - (avg. toxicity in thread)

MAPPING POLARIZATION:

- We can consider polarization of the users' leaning in a given thread of comments as the extent of users with opposed leanings taking part in it

- One possible way of quantifying it is checking whether the leanings' distribution deviates from a unimodal (unpolarized) distribution and resembles a bimodal (polarized) one

- Many measures and statistical tests are designed to check the bimodality of a distribution, each with its pros and shortcomings: we shall review three of them to find the best suited one for this analysis

MEASURES OF POLARIZATION:

Bimodality coefficient:

$$BC = \frac{m_3^2 + 1}{m_4 + 3\frac{(n-1)^2}{(n-2)(n-3)}}$$

- BC ≈ 0.55 for a uniform distribution. Higher values indicate bimodality

Distance from unimodality:

- Bins the data into K ordinal categories, and computes the deviation from the unimodality rule
- Any value higher than DFU = 0 indicates a deviation from unimodality

MEASURES OF POLARIZATION (CONT.):

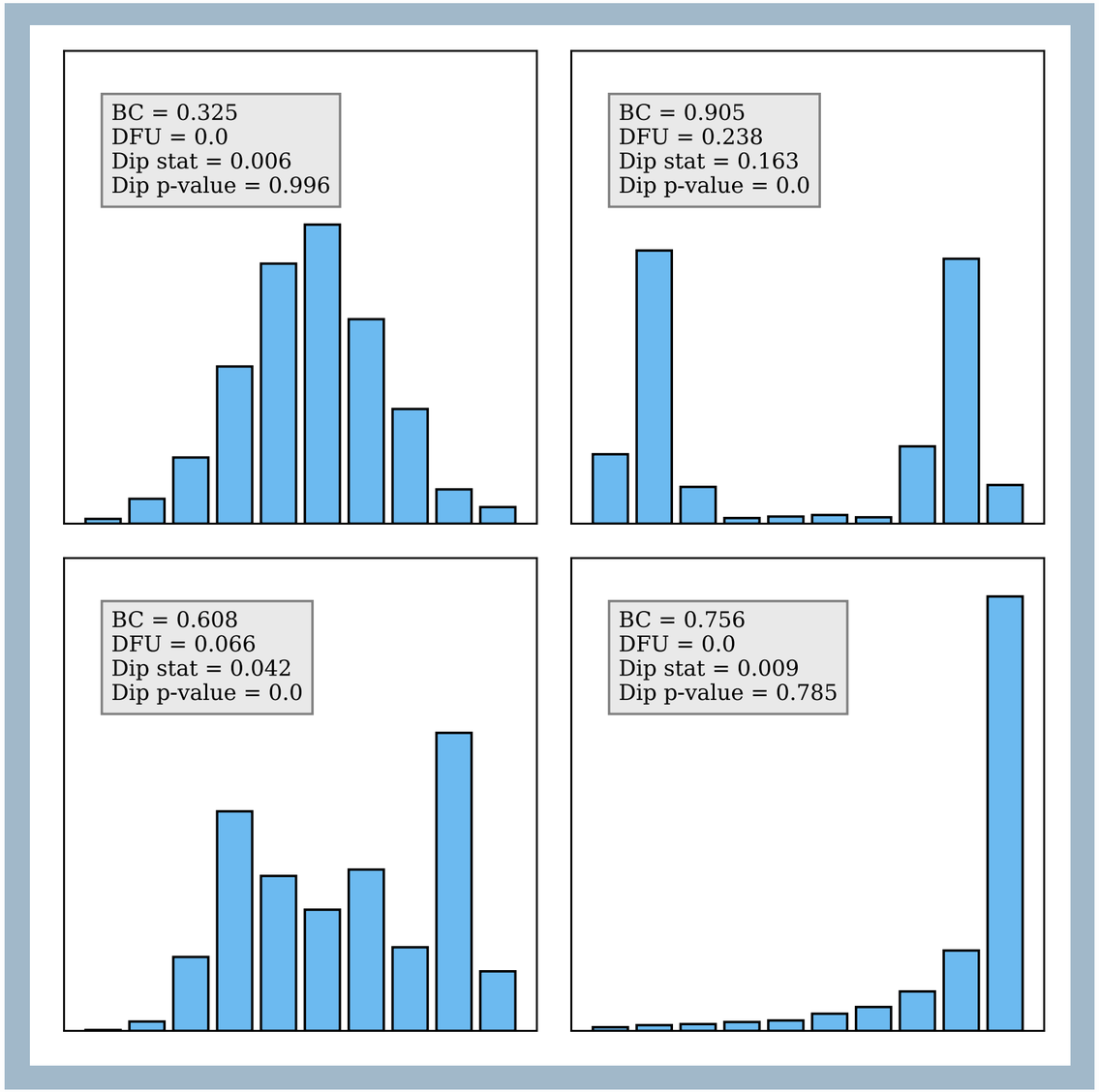
Hartigan's dip test:

- Measures multimodality in a sample by the maximum difference, over all sample points, between the empirical distribution function, and the unimodal distribution function that minimizes that maximum difference

- The dip test index is higher when the distribution approaches bimodality

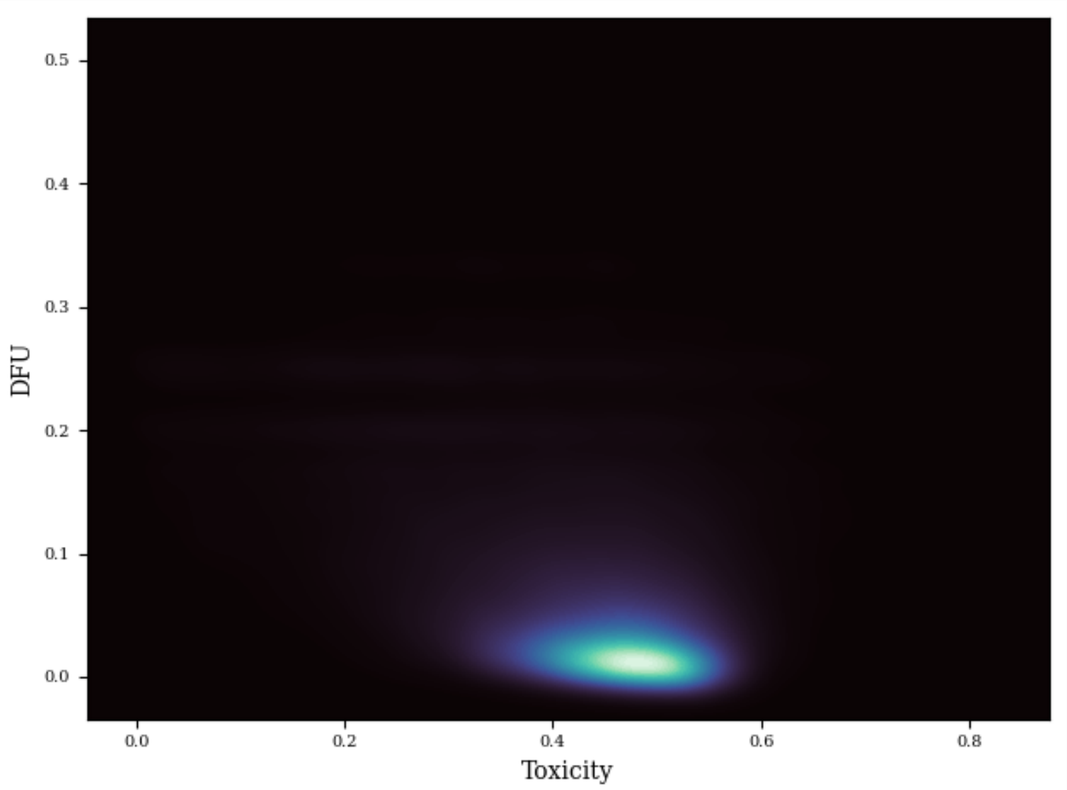
- For the dip test p-value, *"values less than 0.05 indicate significant bimodality and values greater than 0.05 but less than 0.10 suggest bimodality with marginal significance"*

COMPARISON:



- In the first panel, the three measures agree on not considering the normal distribution as bimodal
- In the second and third one, the distributions are correctly classified as bimodal, even though the dip p-value doesn't give us any further insight on the extent of the bimodality
- In the fourth one, the BC classifies the skewed unimodal distribution as more polarized than the one in the third panel

POLARIZATION VS TOXITICY:



- Using DFU with 8 bins to compute the polarization score of a video, there seems to be a slight negative correlation between the toxicity and the polarization of the comments under a video

- It's important to notice how, while almost 30% of the videos have a DFU score = 0, the majority of videos with high polarization are those with relatively few comments

- The average value for the DFU scores is 0.05, the average value for the toxicity scores is 0.43

POLARIZATION VS TOXITICY:

- Furthermore, there's no appreciable difference between the average toxicity in the comments of left, right or center leaning videos, nor the correlation between polarization and toxicity

- Videos coming from center leaning channels however are, on average, more polarized than those coming from left or right leaning channels

SINGLE CONVERSATIONS:

- By extracting single conversations under the videos (composed by a root comment + its replies), we can notice how they are on average more polarized (0.12), but with a lower average toxicity score (0.32)

- The correlation between polarization and toxicity is less marked than when considering the whole corpus of comments

USERS' BEHAVIOUR:

Most toxic users:

- Only 2234 (0.1%) users of the 2M active users have an average toxicity score over their comments equal or higher than 0.7
- The leanings distribution of these users is more polarized than the one obtained considering the whole set of users. Nevertheless, left leaning users are more than double the right leaning ones
- In absolute values, the majority of toxic comments coming from these users are aimed toward left-leaning videos
- This can be explained by the fact that most of these users tend to post toxic comments under videos sharing their same leaning. A negligible percentage of users showed to be posting toxic comments almost exclusively under videos of the opposite faction
- The results are largely the same if the users having at least 70% of their comments identified as toxic are considered

USERS' BEHAVIOUR:

Most polarized users:

- 17.3% of the users have a leaning score (in absolute values) comprised between [0.8, 1)
- These users average toxicity and ratio of toxic comments is comparable to that of the less polarized group
- While again most of the toxic comments are towards one's own faction (in comparable measure between left and right leaning users), a small portion of users (around 1%) exclusively posts toxic comments under videos with opposing views. With this in mind, it's important to notice how these kind of toxic comments are an extremely small fraction of these users' activity, and as such it would be hard to classify such behaviour as a systematic one

USERS' BEHAVIOUR:

Systemic toxicity:

- More generally, looking at the whole dataset, 9% of the users only post toxic comments under left leaning videos, 5% do so exclusively under right leaning videos, and barely one percent devolve their toxicity exclusively to center leaning videos

- Once again, the vast majority of these toxic comments come from users having the same leaning of the videos they are commenting on

- This said, there exists a handful of generally very active users (78 average comments vs 8 for the whole sample), who seems to dedicate around 10% of their total activity to posting toxic comments exclusively under the opposition's videos.