Beyond news: Polarization of left, middle, and right-wing news channels on YouTube

Project Paper

presented

by

Peer Saleth and Raphaela Keßler

at the

Universität Konstanz

Faculty of Politics
Department of Social and Economic Data Science

1. Evaluated by Prof. Dr. David Garcia

https://github.com/peersal/Beyond-News

Konstanz, 2023

Peer Saleth and Raphaela Keßler

Beyond news: Polarization of left, middle, and right-wing news channels on YouTube

Abstract

In today's political landscape, social media platforms have emerged as important channels for political discourse, exerting significant influence on public opinion formation and ideological exchange. Among these platforms, YouTube has gained notable importance as an effective tool for political communication. This paper examines the role of YouTube as a platform for political discourse and its impact on polarization within German politics.

The first research question investigates whether YouTube channels of different political directions (left, centre, right) show notable differences in polarity and toxicity in their community. Additionally we analyze the sentiment of content published by the channels. In the second part, we look at the factors that contribute to polarization in German political YouTube channels. We hypothesised about the extent of discussions, the mood in the comments and the focus on individual politicians.

This research uses sentiment analysis, toxicity detection and network analysis on three selected YouTube channels: "Achtung, Reichelt!" (right-leaning), "Jung & Naiv" (left-leaning) and "Deutsche Welle (DW)" (centre). Logistic regression is used to predict channel affiliation based on sentiment and toxicity metrics. This study also uses linear regression to understand the relationship between the sentiment/toxicity of comments and the extent of discussions.

We find that left and right channels have differences in negativity and toxicity, with the right channel comments being more toxic and negatively charged. The structural balance analysis suggests that the social networks around the left channels are more polarised. The hypotheses regarding the second research question has to be rejected. The analysis shows no linear relationship between negative sentiment/personal attacks and the extent of discussion. However, it's essential to note that these findings must always be interpreted within the context of the given data and the methodological approach employed. You will find all code segments, data and additional plots at our GitHub repository.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Polarization in German Politics: The Influence of Social Media, Particularly YouTube

In recent years, social media platforms have become influential channels for political discourse, amplifying voices and shaping public opinion. Among the various social media platforms, YouTube stands out as a particularly powerful tool for political communication and engagement. This paper delves into the concept of polarization in German politics, focusing on its manifestation and amplification through YouTube, and the implications it carries for the country's political landscape. YouTube's ascent as a political platform stems from its wide user base and accessibility, making it an ideal space for politicians, activists, and commentators to share their ideas and engage with audiences. The following paragraphs will highlight why we have chosen YouTube as a medium and why the exploration of polarization is of central importance.

1.1 News Channels on YouTube

YouTube has emerged as a prominent platform for consuming and spreading news. Its unique characteristics have prompted both creators and viewers to engage with news content in new and diverse ways. YouTube's extended video length sets it apart, enabling comprehensive exploration of complex political topics, unlike the constraints of platforms such as TikTok. YouTube increasingly younger users and thus differ from traditional print newspapers. It is notable that 35% of YouTube users are under the age of 35 (Statista, 2023).

Spreading news on YouTube has advantages and disadvantages. YouTube records a high level of interactivity and offers viewers the opportunity to actively participate in discussions. With 2.5 billion monthly logged-in users, YouTube provides an extensive platform for news channels to reach a global audience (Statista, 2023). Unlike traditional media, which is often paid-for, YouTube's accessibility ensures that news reaches a wider range of users (Ksiazek et al., 2016).

But this accessibility also fosters the disadvantage of limited organisational control. The decentralised nature of YouTube means that there is no complete control over the interpretation and distribution of its content. Also, the resources of the channel owners are often too limited to engage in broad and educational journalism. Videos need to get as many views as possible in order to earn money through advertising. The platform's emphasis on shareability sometimes leads to "soft news" being prioritised over in-depth analysis and complex issues are presented as entertainment. Some creators prioritise sensationalism over objective reporting, potentially trivialising important political issues (Lichtenstein, 2018). Ksiazek et al. can show that digital journalism follows traditional standards in terms of production elements, but lacks objectivity of content. That highlights that the news is produced to the same production standards like natural sound, audio, music and narration as traditional news media, but do not set the same standards (fairness, sourcing, and agenda) for the objectivity of their content (Limor and Thomas B., 2011; Ksiazek et al., 2016). This makes it difficult for users to reflect on what they see and form an educated opinion.

1.2 Polarization on YouTube

In summary, YouTube has become a powerful medium for news sharing, offering advantages such as interactivity, a large user base and accessibility. However, the lack of organisational control and proneness to sensationalism pose challenges. When there is no balance between engaging content and responsible journalism on YouTube news channels, it can lead to polarization, which is the motivation for our research question.

Polarization is a phenomenon in which different viewpoints become increasingly clear and entrenched, often leading to increased social division. The consequences of polarization for democracy are profound. Democracy thrives on diversity of opinion and the ability to engage in civil discourse (Barber, 2009). Polarization leads to political stagnation through solidified political camps, decreasing trust in institutions and the rise of populist movements (Jones, 2015) (Casal Bértoa and Rama, 2021).

1.3 Research Questions and Hypotheses

From these insights, we derive two central research questions, which are followed by corresponding hypotheses.

Research Question 1: Do YouTube channels from the left, middle, and right exhibit significant differences in polarity? This question aims to find out whether there are significant differences in polarity between YouTube channels with different political leanings. The research hypothesizes the following:

H1a: Right-wing and left-wing YouTubers are more likely to present polarizing topics with negative connotations than public broadcasters. This hypothesis refers to the sensationalism mentioned above. It is assumed that political channels that lean to the right or to the left are more likely to discuss controversial issues and often present them in a negative light. This means that such channels are more likely to present extreme viewpoints that lead to divisions.

H1b: Right-wing and left-wing YouTube channels have more toxicity in the comments than public broadcast channels. With this, we want to check whether comment sections on right-wing and left-wing channels are more likely to show toxic behaviour and hostile arguments compared to public channels. This is an indicator of increased polarization and hostility within these ideological spaces.

H1c: The comment community of right- and left-wing YouTube channels have a higher degree of homophily and therefore are more polarized than the community of public broadcasting. This hypothesis proposes that the comment communities surrounding political YouTube channels on the right and left are characterized by greater polarization and a lack of balance in the exchange of ideas, as opposed to the relatively more neutral and balanced environment found around public broadcasting channels.

Research Question 2: What are the underlying factors that contribute to the polarity of German political YouTube channels from the left, middle, and right? This question delves into the driving forces behind the polarity observed in the content and discussions of political YouTube channels. The following hypotheses are formulated:

H2a: As negative the connotations of the comments are, as extensive are the discussions. The hypothesis suggests that the extent of discussions around polarizing topics is directly correlated with the negativity of the comments. In other words, more negative comments are indicative of more intense and extensive debates around these subjects.

H2b: If individual politicians get put into the center, the discussions get more extensive. This also follows the thesis that "hard news" moves into the background and charisma (or no charisma) and characteristics of the politicians plays a role. Even though Lichtenstein (2018) did not find a personalization in YouTube information channels, we would like to review this for our data.

CHAPTER 2
Data

In this section, we will outline the steps involved in sampling, generating the data using the YouTube Data API and the preprocessing.

2.1 Sampling

For this project on journalism on YouTube, a variety of analysis of the comments of German-language YouTube channels was carried out. When selecting our channels, we considered a number of factors such as the number of views, the number of videos and the attribution of a political orientation to an external source. For this purpose, we selected three channels that can be located in the political spectrum of left, centre and right (status: may 2023): "Achtung, Reichelt!" with 351,000 subscribers and 682 videos is criticised by several newspapers as a right-wing propaganda channel (Fernsehen, 2023; Nowak, 2023). The owner Tilo Jung of "Jung&Naiv", with 518,000 subscribers and 3746 videos, distances himself from right-wing and conservative ideology (Unfried and Welzer, 2018). The channel was also recognised as a left channel by the Rosa Luxemburg Foundation (LIEDTKE and MARWECKI, 2019) [1]. In contrast, "Deutsche Welle (DW)" with 878,000 subscribers and 30,203 videos serves as a public service broadcasting channel financed by federal tax money. Thus we placed "Deutsche Welle" in the middle of the political spectrum. The bar plot "Channel Statistics" shows that the channel "Achtung Reichelt!" has the most comments, likes and views. All channels have a mean view count per video greater than 50000.

[1] The Rosa Luxemburg Foundation is a party-affiliated foundation of the Left Party.

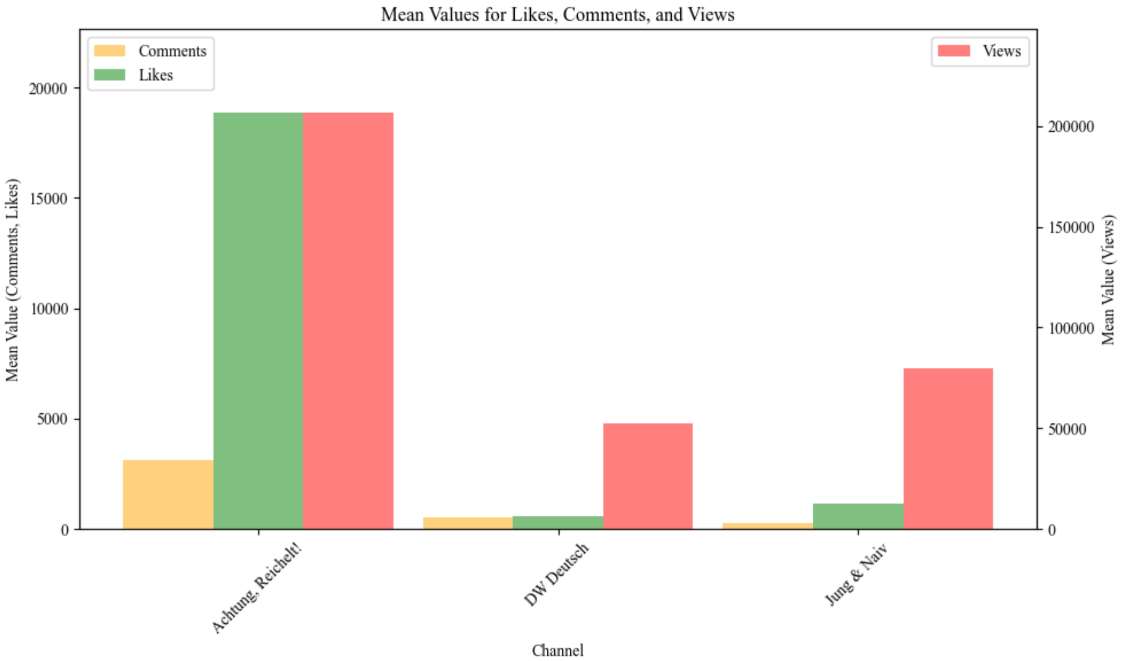


Figure 1. Channel Statistics

2.2 Data Generation Process

The YouTube Data API facilitates the extraction of key insights from videos, comment threads and comments on YouTube. This process involves several steps. To collect video statistics, the API's search endpoint is used with specific channel IDs to collect data such as video titles, publishing dates, likes, number of views and other statistics. Then, comment threads for each video are collected via the commentThreads endpoint by specifying video IDs as parameters. Each comment thread is broken down to obtain individual comments and associated metadata, which includes text, author details, publishing dates, likes and replies.

2.3 Preprocessing

In order to compare videos with a similar topic from different channels, we use the cosine similarity. Videos the same topic but of more than one channels get produced in different ways and get put in different frames. The approach of cosine similarity allows us to identify in a weaker way the similarity of videos and following this making our model more comparable in comparison to use just random videos of different channels. For that we take the video transcript and their vector representation. Word vectorization or word embedding is a method in NLP for mapping words onto a corresponding vector to find word predictions or word similarities. With iterating over every video j description d we can represent each description with a vector μ_d, calculate the dotproduct of each video description and maximize the similarity. The smaller the cosine of the angle between the comparing vectors (cosine distance), the higher the similarity.

$$argmax_{\mu_d} = \frac{exp(\mu_d * \vec{v})}{\sum_j exp(\mu_j * \vec{v})}$$

This task is perfomed using the pretrained "de_core_news_md" pipeline from spaCy [2] which includes multiple NLP methods most important tok2vec. It is a linguistically sophisticated model designed for German natural language processing tasks. The pipeline is trained on a diverse range of text sources, including news articles, websites, and other textual materials in the German language. Its training corpus encompasses a vast spectrum of domains and topics, enabling it to effectively capture the nuances of linguistic structures, syntax, and semantics prevalent in real-world German text.

Using this method, we're effectively examining 153 videos and their related comments on the same political subjects. This approach allows us to draw better conclusions about toxicity, sentiment, and polarization, particularly since different subjects can provoke diverse emotional reactions. In addition, the comments were tokenised, which breaks down the comments into individual tokens or words. Text cleaning includes the removal of special characters, punctuation and non-alphanumeric symbols. In a further step, converting text to lower case letters ensured consistency of the text and reduced redundancy in the analysis. By removing stop words, frequently used words with little contextual meaning are eliminated. To capture a quick overview of the main topics addressed, we generate word clouds for each channel. This preliminary analysis highlights a significant overlap in political themes. Key terms such as "Ukraine", "Putin", "war", "the Green Party", "Afd" and relevant political figures consistently emerge as some of the most frequently used words.

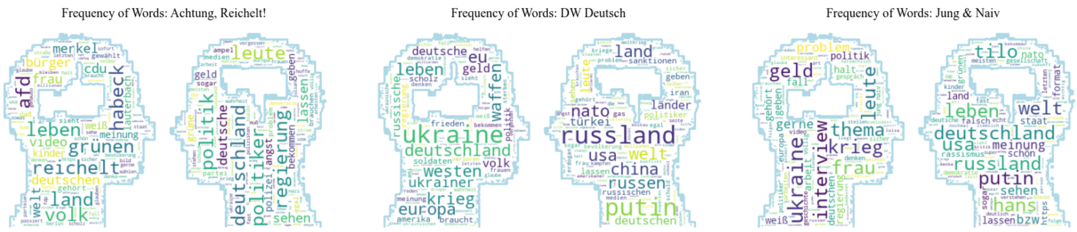


Figure 2. Wordcloud of Channels

Additionally some interaction Values had to be reprocessed. Given the specific configuration of the YouTube Data API, the spectrum of interaction is constrained to two tiers. A commentary can manifest either as a primary level comment or as a secondary level comment. In particular, secondary level comments always refer to the corresponding primary level comment, regardless of the fact that they might be addressed to another user. To redress this problem, we subject each comment to an analytical process employing regular expressions, namely, `r'@([\S]+)'` to detect usernames preceded by the '@' symbol devoid of interstitial whitespace, and `r'@(\S+\s+\S+)'` to identify usernames containing a space as a separator while sharing identical initial substrings. These identified matches are subsequently cross-referenced against our repository of users and integrated into the graphical framework as edges. This method thereby introduces an additional

[2] https://spacy.io/models/de

dimension to our prevailing data structure.

CHAPTER 3
Methods

In the first research question, we have the YouTube channels (with their political orientation) as dependent variable and polarization in the transcript of the video, polarization in the comments and structural balance in clusters of comments as independent variables. As statistical method we used a logistic regression to model the probability that the independent variables can predict which political leaning a channel has. The hypotheses for the second research question focus on the scope of the discussions in the comment section. Extended comment sections are the dependent variable and the sentiment of comments within extended discussions and personalised comments are the independent variables. In order to test the hypothesis we used a linear regression. In the following chapter, the operationalization of the variables mentioned and the methods used to obtain them are explained in more detail.

3.1 Measuring Polarization

In a simple way we could have measured polarization with the ratio of likes and dislikes, but dislikes got restricted from the Google company. That is the reason why we had to evolve own measurements to scale the concept of polarization. We operationalized Polarization on YouTube in three different ways.

First, we apply a BERT classifier to the transcript of the videos (H1a). This gives us information about the sentiments (positive, negative, neutral) of the content itself. A BERT sentiment classifier employs the Bidirectional Encoder Representations from Transformers (BERT)[1] model to analyze and classify sentiment in text. BERT comprehends the context of words by considering both preceding and subsequent words, facilitating a nuanced understanding of language. This enables the classifier to capture intricate sentiment nuances in text, distinguishing positive, negative, and neutral sentiments with high accuracy. We used a model that uses Google's Bert architecture but was trained with 1.834 million German-language samples (Twitter, Facebook and movie, app and hotel reviews) (Guhr et al., 2020). Finally, we extracted the probability that the transcript is negative (prob_neg), this variable has a range between 0 and 1.

Second, we operationalized polarization with toxicity because we assume that online discussions below the videos are not only divergent but also potentially harmful and offensive (H1b). This allows for the creation of a toxicity-based polarization index, where higher scores indicate greater polarization. The Perspective API [2] is a powerful tool for detecting toxicity in comments. It uses machine learning and assesses the potential toxicity of a text by analysing various linguistic cues such as obscenities, insults and personal attacks. The score does not indicate a pure increase or decrease in toxicity, but the likelihood that a comment will be perceived as toxic. The publishers defined toxicity as "a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion" (Jigsaw and team, 2023). The API is typically used by platforms to help moderate and filter out harmful comments to promote healthier online discussion. The variable "toxicity" also has a range between 0 and 1. Examining the normalized histograms of individual channels seen

[1]https://huggingface.co/oliverguhr/german-sentiment-bert
[2]https://perspectiveapi.com

in Figure 3, a discernible pattern emerges wherein Julian Reichelt garners the highest frequency of toxic comments. Following closely in frequency are DW Deutsch's comments with a relatively elevated toxicity level, whereas Jung & Naiv's comments exhibit the lowest frequency of toxicity.

It is noteworthy to underscore that the specific threshold value, plays a pivotal role in this analysis; however, the outlined observation generally remains consistent across varying thresholds. A crucial point of clarification pertains to the nature of the toxicity metric, which reflects not the extent of toxicity, but rather the predictive probability assigned by the model to a given textual string being toxic. While one might infer that comments rated as more toxic possess a heightened likelihood of being toxic, such a direct correlation cannot be definitively established.

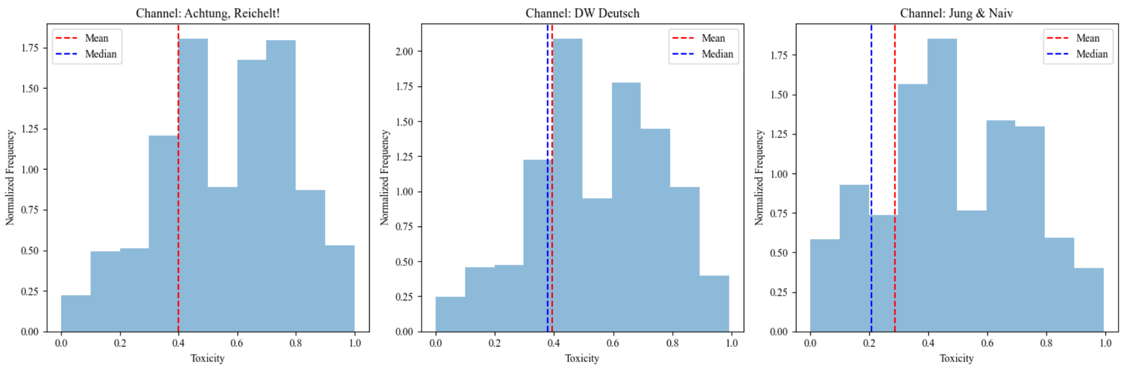


Figure 3. Toxicity Histograms

Thirdly, our investigation entails an examination of the social networks inherent to each channel community, wherein we scrutinize their structural balance. To accomplish this, we construct network models utilizing the "networkx" package in Python (Hagberg et al., 2008). One fundamental inquiry that confronted our investigation pertains to the proposition of establishing connections between all secondary-level comments and their respective first-level comments, even in cases where such connections might not be explicitly directed. It is likely that the users which wrote a second level comment do not interact with users of the initial first level comment. In our approach, however, we have chosen to consider these connections together with those originally identified by the application of regular expressions. This decision was primarily motivated by the observation that contributors at the first level of commenting frequently engage in the discourse present within the subsequent second-level comments. Furthermore, the thematic trajectory of these discussions tends to exhibit a discernible influence stemming from the subjects broached in the corresponding first-level comments. We utilize the spring layout algorithm from "networkx" to visually illustrate the network structures [3].

To enrich our analysis, we introduce edge weighting based on their corresponding toxicity levels. Consequently, we generate two distinct sets of plots: one encompassing all comments and the other exclusively featuring toxic comments. This dual-pronged approach enables us to track changes in the structural layout effectively. Moreover, we delve into cliques using the Louvain algorithm. This method aids in identifying clusters of highly interactive users (Blondel et al., 2008). By scrutinizing the dynamics of toxicity within these clusters, we can derive insightful conclusions regarding the

[3]This technique hinges on two fundamental principles:

1. **Spring Force:** Nodes connected by edges are drawn closer together. This effect is more pronounced for shorter edges, leading to the clustering of nodes with tighter connections.
2. **Electrical Repulsion:** Nodes lacking direct connections exert a repulsive force on each other, preventing undue proximity and facilitating the uniform dispersion of nodes.

prevailing discussion culture within each channel community.
When examining the entire network, one notable finding is that Julian Reichelt exhibits the highest level of toxicity, followed by DW and Jung und Naiv. This observation remains consistent even when accounting for variations in network sizes. Additionally, instances of higher-than-average toxicity are observed within clusters of users who interact with each other. Notably, this elevated toxicity is most pronounced in clusters connected to Julian Reichelt, followed by DW and Jung und Naiv. The deviation to mean toxicity of each big cluster can be seein in Figure 5 These findings suggest that intense discussions within all networks are more toxic especially for Julian Reichelt and DW.

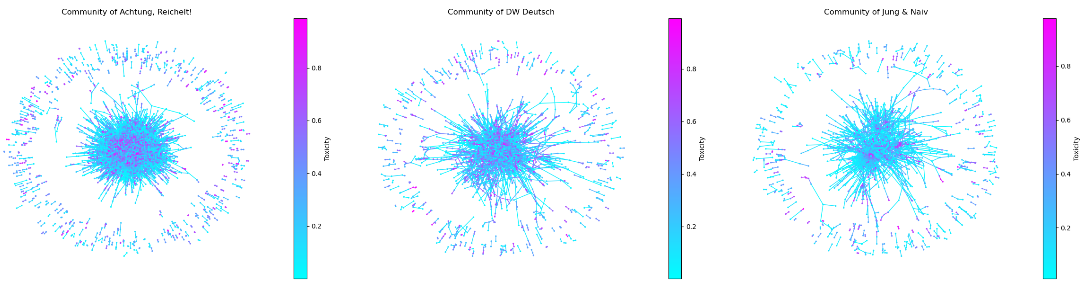


Figure 4. Community of Channels

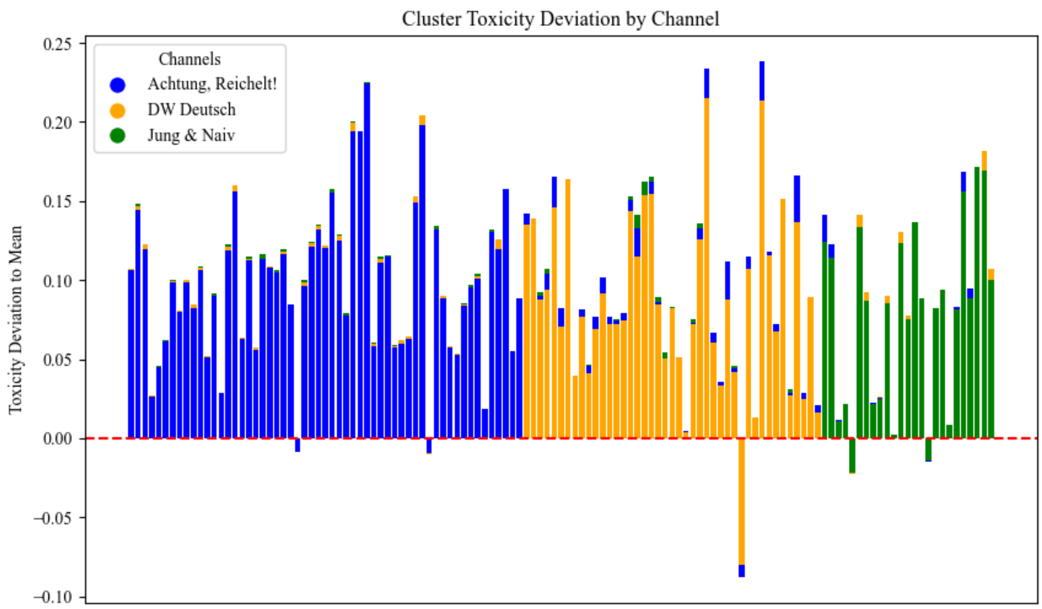


Figure 5. Characteristics of Cliques

Moreover, it is evident that there exist fewer clusters encompassing users from both the left and right wing channels in comparison to other channel combinations. A plausible explanation for this phenomenon is that segments of the communities aligned with the right and left wing channels exhibit lower levels of interaction between each other as opposed to their interaction with the central channel. When engaging in communication, the toxicity observed within this discourse consistently surpasses the average toxicity level. This stands as an indicator of a polarized conversational culture

between users affiliated with the left and right wing communities.
When we analyze the networks of comments labeled as toxic, a few things stand out. Despite Julian Reichelt having the most comments, there's a relatively higher number of toxic comments posted under his videos, which aligns with what we saw before. Upon closer inspection of the graphical representation, a conspicuously centralized cluster of individuals engaging in the composition of toxic remarks directed at other users becomes evident in the cases of Julian Reichelt and DW. This observation potentially alludes to the existence of a markedly polarized discourse milieu wherein a numerical minority vested with toxic behavior wields a disproportionate influence over the broader conversational landscape which is also supported by the findings of the clusters. While the variation in structure might stem from the higher volume of data[4] available for the Julian Reichelt network, we possess comparable data points for DW and Jung & Naiv. This enables us to infer that these noxious subgroups manifest in varying degrees across each channel.

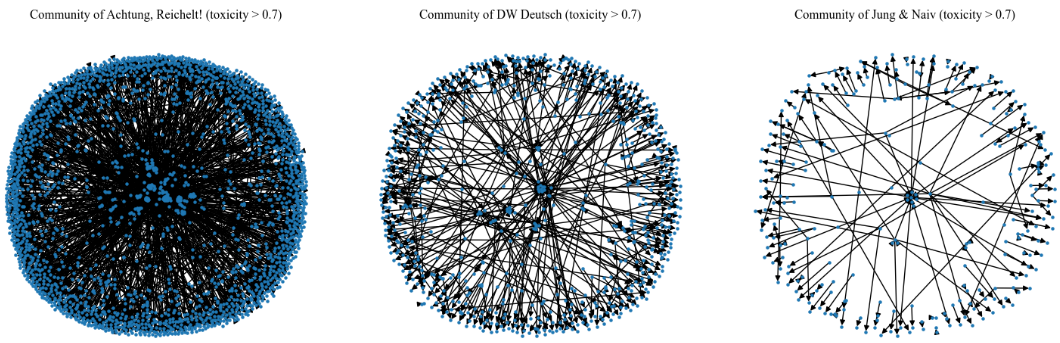


Figure 6. Community of Channels Toxicity

To further inspect polarization in the social networks we apply the triadic closure approach to operationalise polarization in social networks (H1c), which involves identifying and analysing connections between nodes within the network (Röchert et al., 2020). This technique focuses on the tendency of people with similar opinions to connect and form closed triangles. The theory behind the triadic closure approach can be traced back among others to Harary's "Balance Theorem" from the 1950s (Harary, 1959; Young Pedersen, 2019). The "Balance Theorem", a fundamental concept in graph theory, deals with the idea of edge-separated paths in a graph. It states that in a graph where each vertex has an even degree (an even number of connections to other vertices), a collection of edge-separated paths can be found that collectively cover all edges of the graph. This theorem can be applied to the field of the triadic closure approach (Easley and Kleinberg, 2010). In social networks, triadic closure refers to the tendency for two people who have a common friend to eventually connect. When individuals form ties based on shared beliefs, this can lead to the formation of clusters that reinforce those beliefs and potentially increase division between different groups with opposing views. By identifying such clusters, we can measure the degree of homophily and polarization within the network(Yarchi et al., 2021). Therefore, we examine all triangle structures within the networks and study their nature by looking at the edges. Triangles with 0 or 2 negative edges are considered stable/strong, while triangles with 1 or 3 negative edges are considered unstable/weak. A stable structure in a network is an indicator of polarization (Xia et al., 2016). This method is applied using the toxicity score with a threshold $t \geq 0.7$ for a negative edge and $t < 0.7$ for a positive edge. Additionally, we utilize the sentiment score from BERT and assign edges according to their sentiment where a negative connection is indicated by a negative sentiment. With a notable correlation coefficient of 0.2774 observed between the negative sentiment and toxicity scores, the outcomes demonstrate a comparable pattern. The structural balance, as

[4] Julian reichelt: 39538 interactive comments
DW: 7790 interactive comments
Jung & Naiv: 5220 interactive comments

Channel	Structural Balance (Toxicity)	Structural Balance (Sentiment)
Achtung, Reichelt!	84.098544 %	53.527436 %
DW Deutsch	87.455197 %	53.405018 %
Jung & Naiv	92.050209 %	62.343096 %

Table 1. Structural balance

depicted in the Table 1, is characterized by the relative prevalence of triangles featuring either 0 or 2 negative edges. Notably, the Jung & Naiv network exhibits the best structural balance, implying a heightened polarization within its network. Similarly, the structural balance for both the Achtung Reichelt and DW Deutsche networks is comparable and notably elevated, signifying a significant degree of polarization across all networks to varying extents.

3.2 Measuring the underlying Factors of Polarization

Assuming there exists polarization in the comments of a video, extensive discussions are of interest to us because they provide insight into the depth of polarization. We assume that negative comments challenge users and therefore lead to extensive interactions. As a consequence, we apply the BERT sentiment classifier again, but to the comments that have started extensive discussions (H2a). Heated debates also allow for the identification of strategies used to influence opinions and the evolution of debates over time. We assume that one of these strategies is the personalization of comments. We assume that by naming political actors, the debate is polarized and heated up. Therefore, we apply dictionary based matching to match the comments with the names of German politicians (H2b). For this we used a data set of all elected members of parliament from Germany from the Federal Statistical Office (Bundesamt, 2023). In addition to the self-coded variable "personal_attack" which refers to the dictionary approach of the German politicians, we supplemented the analysis with the variable "identity_attack" of the perspective API. The variable "identity_attack" is defined by the authors as "negative or hateful comments targeting someone because of their identity" (Jigsaw and team, 2023).

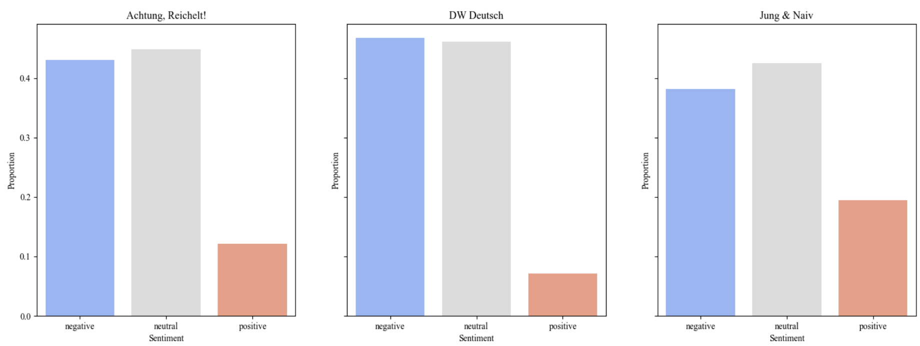


Figure 7. Sentiment Comments Histograms

Our research findings reveal that both right-wing and left-wing communities tend to refer more often to political figures. The data presented in Table 2 clearly demonstrates that 10.54% of comments posted on Julian Reichelt's an 9.76 % of comments posted on Jung & Naiv videos make reference to politicians. To investigate the impact of a comment mentioning a politician on negative sentiment and identity_attack values, we employ a calculation based on mean differences.

We compute the difference between the mean values of comments that reference politicians and the mean values of comments within the corresponding community of the specific channel. It can be seen that, comments containing mentions of politicians exhibit identity_attack scores that surpass the average. This observation aligns with our assumption that when politicians are referenced, there's a heightened likelihood of encountering hostile remarks. Analyzing channel-specific differences isn't straightforward due to our focus on deviations from individual channel averages rather than the overall mean. Attempting the latter could yield biased outcomes, given that the total average for negative sentiment (illustrated in Figure 7) and identity attack is notably below average.

channel	personal attack = 1	prob. neg. sentiment	identity attack
Achtung, Reichelt!	10.54 %	0.01	0.02
DW Deutsch	4.69 %	0.02	0.05
Jung & Naiv	9.76 %	-0.03	0.07

Table 2. Identity Attack Mean and Sentiment Mean of Personal Attacks

CHAPTER 4

Results

To draw a final conclusion we perform two chosen regression models using the variables we examined before.

4.1 Outcomes Logistic Regression

To compute the various impacts of variables that indicate polarization on a comment beeign from a specific channel we constructed following logistig regression:

$$y = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

where y is the dependent variable which is that video and its comment community belongs to specific channel, β_0 is the intercept, β_1 is the mean toxicity of written comments, β_2 is the mean probability of a negtaiv sentiment of the transcript of the videos and β_3 is the mean portion of comments which are part of a stable clique.

The pseudo R^2 of the MNLogit model shows the McFadden's R^2, which represents how much better the model predicts the dependent variable than the null model. McFadden (1977) states that values of 0.2 to 0.4 indicate an excellent fit. It follows that we have in the logit models for the channels of Jung & Naiv (0.5091) and Achtung, Reichelt! (0.2327) an excellent fit, but DW Deutsch does not fit that much better than the null model (see table 3).

In conclusion of our findings, we are able to reject the H1a hyptohesis as for both right and left wing channels the regression Results in Table 1 are negativ for the probability that a video transcript has a negativ sentiment (see table 4 to 6). It can be sayed that the effect is still higher for the right wing channel than the left wing channel. The odds ratio show, that if the average negativity of the transcript of a video increases by one, it is -1.1 times less likely that it originates

Table 3. Overview

Dep. Variable:	Jung & Naiv	**No. Observations:**	107
Model:	MNLogit	**Df Residuals:**	103
Method:	MLE	**Df Model:**	3
Log-Likelihood:	-33.869	**Pseudo R-squ.:**	0.5091
LLR p-value:	3.772e-15	**LL-Null:**	-68.994
Dep. Variable:	DW Deutsch	**No. Observations:**	107
Model:	MNLogit	**Df Residuals:**	103
Method:	MLE	**Df Model:**	3
Log-Likelihood:	-60.486	**Pseudo R-squ.:**	0.07340
LLR p-value:	0.02247	**LL-Null:**	-65.278
Dep. Variable:	Achtung, Reichelt!	**No. Observations:**	107
Model:	MNLogit	**Df Residuals:**	103
Method:	MLE	**Df Model:**	3
Log-Likelihood:	-53.410	**Pseudo R-squ.:**	0.2327
LLR p-value:	4.307e-07	**LL-Null:**	-69.611

Table 4. Log. Regression - Jung & Naiv

Jung & Naiv=1	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	8.2467	2.081	3.963	0.000	4.168	12.325
toxicity_mean	-30.3276	5.884	-5.154	0.000	-41.860	-18.795
prob_neg_transcript	-1.1139	1.686	-0.661	0.509	-4.418	2.190
stable_tox_mean	13.7374	6.061	2.266	0.023	1.858	25.617

Table 5. Log. Regression - DW Deutsch

DW Deutsch=1	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	-2.8331	1.473	-1.924	0.054	-5.720	0.053
toxicity_mean	6.4599	3.406	1.897	0.058	-0.215	13.135
prob_neg_transcript	1.4346	0.943	1.521	0.128	-0.414	3.283
stable_tox_mean	-4.6603	4.340	-1.074	0.283	-13.167	3.847

Table 6. Log. Regression - Achtung, Reichelt!

Achtung, Reichelt!=1	**coef**	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	-7.7325	2.116	-3.654	0.000	-11.880	-3.585
toxicity_mean	20.4280	4.943	4.133	0.000	10.740	30.116
prob_neg_transcript	-0.4908	1.036	-0.474	0.636	-2.521	1.539
stable_tox_mean	-4.3891	4.924	-0.891	0.373	-14.040	5.262

from Jung & Naiv than that it does not originate from the channel Jung & Naiv, while it is -0.491 times less likely that it can be assigned to the channel Achtung, Reichelt! than not to the channel of Achtung, Reichelt!. In contradiction the value is positive for the public broadcast channel DW Deutsch. If the average negativity of the transcript of a video increases by one, it is 1.435 times more likely that it originates from DW Deutsch. Additionally the p-values of the variable "prob_neg_transcript" indicate that the findings are not satistically significant.

We also have to reject hypothesis H1b, because not the right and the left channel can be better predicted by toxicity in the comments, but only for the right channel of Julian Reichelt the model gives a 20.428 times higher probability of correctly predicting the channel by toxicity. Our outcomes underscore an elevated magnitude of toxicity prevailing within the community surrounding Julian Reichelt's right-wing YouTube channel. The regression outputs substantiate that the mean toxicity within a video community wields a significantly positive influence on the likelihood that a given video originates from Julian Reichelt's right-wing channel. However, it is important to note that these results are not exclusive to the right-wing YouTube channel, but public service broadcasting has also shown a positive correlation (6.4599) with toxicity in the comments (see table 5). In other words, the left wing channel Jung & Naiv exhibits the most muted levels of toxicity. These observations are consistent with the findings of figure 7.

H1c must be rejected as well as we identify a well-established structural balance within the Jung & Naiv network, indicating increased polarization within that network (13.7374 in 4), but not for the right-wing channel (-4.3891 in table 6). Julian Reichelts network has about the same structural balance as DW Deutsch. Hence, this observation may indicate a potential lack of substantial interaction among diverse political factions within the context of Julian Reichelt's videos. However, further analysis is required to validate this hypothesis and draw definitive conclusions. The interpretation of these findings must be approached with caution, considering the unique configuration of interactions between first and second levels within the established networks.

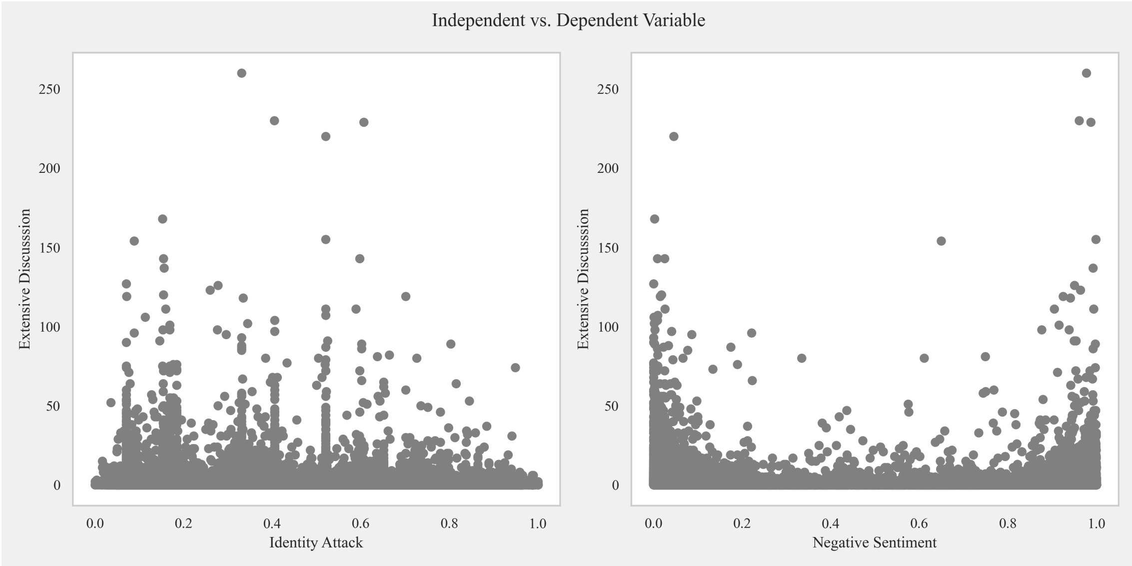
In summary, our assumptions are partly confirmed. The right-wing channel shows high toxicity and negativity, while the left-wing channel is highly polarized. However, the highest values for each factor do not occur simultaneously in both right-wing and left-wing channels.

4.2 Outcomes Linear Regression

Table 7. Linear Regression

Dep. Variable:	totalReplyCount	**R-squared (uncentered):**	0.012
Model:	OLS	**Adj. R-squared (uncentered):**	0.012
Method:	Least Squares	**F-statistic:**	530.8
Prob (F-statistic):	**Log-Likelihood:**	-3.6164e+05	
No. Observations:	134883	**AIC:**	7.233e+05
Df Residuals:	134880	**BIC:**	7.233e+05
Df Model:	3	**Covariance Type:**	nonrobust

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
prob_neg	0.2427	0.021	11.796	0.000	0.202	0.283
identity_attack	0.5301	0.027	19.709	0.000	0.477	0.583
personal_attack	0.1406	0.032	4.421	0.000	0.078	0.203



The Multivariate Linear Regression show statistically significant results. As negativ the comments, as higher the identity attacks and if a German politician gets mentioned discussions get more extensive. 1.2 percent variance of the dependent variable "extensive discussion" (reply count) can be explained by this linear regression.

Nevertheless, we cannot accept the hypotheses H2a and H2b because the variables do not fulfil the conditions for a linear regression. As the plots show, the dependent and independent variables do not have a linear relationship. With very low negativity in the comments and very high negativity, the discussion seems to become more intense and rather calm in the middle range. Interestingly, an opposite relationship emerges for the variable "identity attack". In the middle value range of the variable for attacks on individuals, the discussion in the comments seems to be the most extensive and therefore the most polarizing.

CHAPTER 5
Discussion

During our research, we encountered a multitude of problems and challenges. One of the key challenges was dealing with multicollinearity among independent variables, which posed a substantial issue for our study. We observed a VIF (Variance Inflation Factor) of 13.5 for the variable "toxicity," and a VIF of 13.1 for the "stable" variable, indicating an alarming level of collinearity between these two variables (see appendix 6). This result was attributed to the unfortunate weighting of the stable variable with toxicity, exacerbating the multicollinearity concern. We originally tried to include control variables, such as whether a comment is part of a louvain cluster or not (logistic regression), but the multicolinearity was too high. We also wanted to include control variables for the linear regression, namely which party a politician mentioned/attacked in the comments belongs to, or what gender he or she is. However, we had too many missing values for these variables ("personal_attack" = 0), which is why we dispensed with these control variables.

Furthermore, we grappled with the lack of independence in our observations, as the values of videos originating from the same YouTube channel were not entirely independent. This correlated nature of data points posed a challenge in maintaining the assumptions required for robust statistical analyses. Additionally, a central weakness in our sampling methodology became evident, as we restricted our evaluation to only three YouTube channels. This limitation could potentially affect the generalizability of our findings. To address this, future analyses should strive to replicate the results across a broader range of news channels or even extend the study to different platforms, ensuring greater diversity in the data sources.

A notable constraint surfaced in the context of our Logistic Regression analysis, where we were confronted with a dataset comprising a mere 100 observations. Such a limited sample size can severely impact the reliability and statistical power of our results. When we chose a dictionary-based approach to analyse mentions of politicians, the inclusion of a broader group of international politicians and political stakeholders could have provided deeper insights into the dynamics of personal attacks.

Moreover, our exploration revealed the absence of linearity between extensive discussions and personal attacks, challenging the assumption that an increase in discussion automatically corresponds to a rise in personal attacks.

The outcomes derived from our network analyses are open to discussed as well. The challenge arises from the disparity in sample sizes, which complicates direct visual comparisons. Moreover, the unique two-layer structure of YouTube could potentially impact the application of the triadic closure approach. In this context, the initial layer characterized by toxic connections ensures the existence of at least one shared negative edge within all cliques present in the corresponding second layer.

CHAPTER 6

Conclusion

The rejection of H1a contradicts the assumption that politically oriented social media channels polarise with negative content in order to generate likes, shares and comments through sensationalism. On the other hand, it may support the result of Lichtenstein (2018). Channels with a political orientation to the right or left have a less negative transcription value than DW Deutsch, because they may focus more on "soft news". Discussions about the hard facts, disasters and wars of the world are always connected with a certain negativity. But if (possibly due to a lack of resources) objective and in-depth reporting of "hard news" is avoided and simplified problems of world events are dealt with instead, there will be less negativity in the transcript.

The high toxicity in the comments below videos of right wing YouTube channels (H1b, H1c), shows that these users are more inclined to express their beliefs vehemently in comment sections. This group may feel frustrated and marginalised, which may lead to more toxic comments to express their dissatisfaction. When dissatisfaction creates an "us-vs-them" mentality, this can be an indicator of polarization.

The results regarding the second research question are interesting in the sense that an increased attack on identities and the mention of politicians trigger a more intense discussion (reply count). Attacks on identity often touch on deep issues such as racism, gender or religion. As these topics are very sensitive, outrage over such attacks can lead to more people replying to express their disagreement or lack of understanding. Identity attacks often trigger outrage or solidarity and are therefore polarizing. Examining the negativity in the comments in relation to the reply count shows that very low negative comments and very high negative comments lengthen the discussion. If a post is less negative, people who share similar views will be more inclined to express approval. This can lead to a variety of responses where readers share their support, stories, and experiences. On the other hand, posts with strong negative sentiment can trigger outrage or controversy. People might be inclined to reply to express disapproval or displeasure with the content. However, in-depth discussions may also arise here through a feeling of solidarity with particularly negative comments.

Social media platforms designed to connect people have become breeding grounds for polarization. Through algorithmic recommendations, users are often exposed to content that aligns with their existing beliefs, creating echo chambers that reinforce perspectives and limit exposure to dissenting viewpoints. This process can reinforce biases, prevent constructive dialogue and drive people further apart. Journalism, an important pillar of democracy, can be affected by polarization on social media. When social media display sensationalism or partisan bias, news becomes distorted and contributes to a fragmented information landscape. Audiences look for sources that confirm their beliefs and thus avoid engaging with critical reports, weakening the role of journalism as a source of objective information. Responsible journalism that focuses on balanced reporting can counteract sensationalism. Mastering the digital age requires a concerted effort to bridge the divides exacerbated by polarization and promote a better informed and united society. Over the last ten years, it has become evident that social media significantly influences polarization within the political landscape. The ascent of populist arguments and parties is observable across Europe. This surge in populism not only intensifies debates but also fosters the potential for criminal hate-driven activities as a repercussion of heightened polarization (Müller and Schwarz, 2021). A prominent illustration of this phenomenon is the assassination of Walter Lübcke by a right-wing extremist on June 1, 2019 . Our research indicates that the right-wing channel exhibits higher levels of toxicity and negative emotions in discussions. This serves as both a cautionary example and an urgent plea to political stakeholders, emphasizing the imperative nature of considering the potential consequences tied to their deliberate choice of language.

In this pursuit, we extend our sincere gratitude to Prof Dr Garcia for his invaluable contributions to our research. His guidance and recommendations regarding methodological approaches have significantly shaped the trajectory of our investigation.

Bibliography

Benjamin R. Barber. *Strong democracy: participatory politics for a new age.* University of California Press, Berkeley, 20. anniversary ed., with a new preface edition, 2009. ISBN 978-0-520-24233-3.

Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. 2008. doi: 10.48550/ARXIV.0803.0476. URL https://arxiv.org/abs/0803.0476. Publisher: arXiv Version Number: 2.

Statistisches Bundesamt. Gewählte Abgeordnete, May 2023. URL https://www.bundeswahlleiterin.de/bundestagswahlen/2021/ergebnisse/opendata.html#6420c916-0507-4d76-b752-33e8bdea15c9.

Fernando Casal Bértoa and José Rama. Polarization: What Do We Know and What Can We Do About It? *Frontiers in Political Science*, 3:687695, June 2021. ISSN 2673-3145. doi: 10.3389/fpos.2021.687695. URL https://www.frontiersin.org/articles/10.3389/fpos.2021.687695/full.

David Easley and Jon Kleinberg. *Networks, crowds, and markets: reasoning about a highly connected world.* Cambridge University Press, New York, 2010. ISBN 978-0-521-19533-1. OCLC: ocn495616815.

Zweites Deutsches Fernsehen. BILD TV 2.0: Welcher Milliardär hinter "Achtung, Reichelt!" steckt - WALULIS DAILY, April 2023. URL https://www.zdf.de/funk/walulis-daily-12068/funk-bild-tv-20-welcher-milliardaer-hinter-achtung-reichelt-steckt---walulis-daily-102.html.

Oliver Guhr, Anne-Kathrin Schumann, Frank Bahrmann, and Hans Joachim Böhme. *Training a Broad-Coverage German Sentiment Classification Model for Dialog Systems.* European Language Resources Association, Marseille, France, March 2020. URL https://www.aclweb.org/anthology/2020.lrec-1.202.

Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

Frank Harary. On the measurement of structural balanceBehavioral Science. *Behavioral Science*, 4(216), 1959. URL https://www.proquest.com/docview/1301291792?pq-origsite=gscholar&fromopenview=true&imgSeq=1.

Jigsaw and Google's Counter Abuse Technology team. Perspecive API, 2023. URL https://perspectiveapi.com/research/.

David R. Jones. Declining Trust in Congress: Effects of Polarization and Consequences for Democracy. *The Forum*, 13(3), January 2015. ISSN 1540-8884, 2194-6183. doi: 10.1515/for-2015-0027. URL https://www.degruyter.com/document/doi/10.1515/for-2015-0027/html.

Thomas B Ksiazek, Limor Peer, and Kevin Lessard. User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments. *New Media & Society*, 18(3):502–520, March 2016. ISSN 1461-4448, 1461-7315. doi: 10.1177/1461444814545073. URL http://journals.sagepub.com/doi/10.1177/1461444814545073.

Dennis Lichtenstein. Political information on YouTube. An analysis of four political YouTube channels in Germany. *The Institute for higher social communication studies (IHECS)*, (Innovative educational and media practices for an inclusive and participatory Europe), 2018. ISSN 978-2-9602283-0-4.

MARIUS LIEDTKE and DANIEL MARWECKI. VON INFLUENCER*INNEN LERNEN, YOUTUBE & CO. ALS SPIELFELDER LINKER POLITIK UND BILDUNGSARBEIT. *Rosa-Luxemburg-Stiftung*, (07/2019), 2019.

Peer Limor and Ksiazek Thomas B. YOUTUBE AND THE CHALLENGE TO JOURNALISM: New standards for news videos online. *Journalism Studies*, 12(1):45–63, February 2011. ISSN 1461-670X, 1469-9699. doi: 10.1080/1461670X.2010.511951. URL http://www.tandfonline.com/doi/abs/10.1080/1461670X.2010.511951.

Daniel McFadden. Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments. *Cowles Foundation Discussion Papers*, (707), 1977. URL https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=1706&context=cowles-discussion-paper-series.

Karsten Müller and Carlo Schwarz. Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association*, 19(4):2131–2167, August 2021. ISSN 1542-4766, 1542-4774. doi: 10.1093/jeea/jvaa045. URL https://academic.oup.com/jeea/article/19/4/2131/5917396.

Peter Nowak. „Achtung, Reichelt!“ auf Youtube: Der schwarze Kanal. *TAZ*, 2023. URL https://taz.de/Achtung-Reichelt-auf-Youtube/!5914322/.

Daniel Röchert, German Neubaum, Björn Ross, Florian Brachten, and Stefan Stieglitz. Opinion-based Homogeneity on YouTube. 2(1), 2020.

Statista. YouTube: distribution of global audiences 2023, by age and gender, sources: We Are Social; DataReportal; Google; Meltwater, 2023. URL https://www.statista.com/statistics/1287137/youtube-global-users-age-gender-distribution/.

Peter Unfried and Harald Welzer. Tilo Jung im Interview: Angekotzt vom Stillstand. *TAZ*, October 2018. URL https://taz.de/Tilo-Jung-im-Interview/!167352/.

Weiguo Xia, Ming Cao, and Karl Henrik Johansson. Structural Balance and Opinion Separation in Trust–Mistrust Social Networks. *IEEE Transactions on Control of Network Systems*, 3(1): 46–56, March 2016. ISSN 2325-5870. doi: 10.1109/TCNS.2015.2437528.

Moran Yarchi, Christian Baden, and Neta Kligler-Vilenchik. Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication*, 38(1-2):98–139, March 2021. ISSN 1058-4609, 1091-7675. doi: 10.1080/10584609.2020.1785067. URL https://www.tandfonline.com/doi/full/10.1080/10584609.2020.1785067.

Mina Young Pedersen. Polarization and Echo Chambers: A Logical Analysis of Balance and Triadic Closure in Social Networks. Master's thesis, Universiteit van Amsterdam, 2019. URL https://eprints.illc.uva.nl/id/eprint/1700/1/MoL-2019-10.text.pdf.

CHAPTER A

Descriptive Tables

Table 8. Independent Variables - Log Regression

	toxicity mean	prob. neg. transcript	stable tox. mean
count	153	153	153
mean	0.359148	0.102240	0.125371
std	0.075860	0.206582	0.058916
min	0.132493	0.006049	0.000000
25%	0.297135	0.020132	0.088646
50%	0.378303	0.029650	0.113990
75%	0.416435	0.053154	0.162037
max	0.498905	0.980246	0.365854

Table 9. Independent Variables - Lin Regression

	prob. neg.	identity attack	personal attack
count	134883	134883	134883
mean	0.418578	0.394927	0.099471
std	0.452481	0.271002	0.299295
min	0.000004	0.000000	0.000000
25%	0.005772	0.158129	0.000000
50%	0.074081	0.337739	0.000000
75%	0.964480	0.604080	0.000000
max	0.999740	1.000000	1.000000

CHAPTER B

Regression Diagnostic Tables

Table 10. VIF - Log Regression

feature	VIF
toxicity mean	13.5
prob. neg. transcript	1.3
stable tox. mean	13.1

Table 11. Correlation Matrix - Log Regression

	toxicity mean	prob. neg. transcript	stable tox. mean
toxicity mean	1.000000	0.216647	-0.128439
prob. neg. transcript	0.216647	1.000000	-0.064327
stable tox. mean	-0.128439	-0.064327	1.000000

Table 12. VIF - Lin Regression

feature	VIF
prob. neg	1.7
identity attack	1.8
personal attack	1.1

Table 13. Correlation Matrix - Lin Regression

	prob. neg	identity attack	personal attack
prob. neg	1.000000	0.217969	0.006826
identity attack	0.217969	1.000000	0.032564
personal attack	0.006826	0.032564	1.000000