EasyChair Preprint

№ 16017

# Navigating the Disagreement Space: A Case Study on Persistent YouTube Users' Interactions in Immigration-Related Discussions

Davide Bassi, Martín Pereira-Fariña and Renata Vieira



April 19, 2026

# Navigating the Disagreement Space: A Case Study on Persistent YouTube Users' Interactions in Immigration-Related Discussions

Bassi Davide
Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) - University of Santiago de Compostela
Santiago de Compostela, Spain
davide.bassi@usc.es

Renata Vieira
Centro Interdisciplinar de História, Culturas e Sociedades (CIDEHUS) - University of Évora
Évora, Portugal

Martín Pereira-Fariña
Instituto de Investigación en Humanidades (iHUS) - University of Santiago de Compostela
Santiago de Compostela, Spain

## Abstract

YouTube comment sections constitute volatile, deindividualized arenas, where low accountability and antagonistic dynamics hinder constructive political discussion. This paper investigate how persistent users, those who repeatedly return to these contentious spaces, navigate immigration-related debates and sustain their stance positioning over time. Drawing on a longitudinal corpus of U.S. immigration-related YouTube videos comments (2020–2024), we introduce a Natural Language Inference (NLI) based stance detection pipeline that scales to YouTube's conversational structure, enabling fine-grained classification of user issue positioning. Across user clusters, we identify diverse interactive strategies: some reinforce in-group alignment, others deliberately seek cross-cutting arenas, and still others combine antagonism with curiosity in ways that challenge conventional accounts of echo chambers. Our results highlight that, among YouTube persistent users, polarization is neither uniform nor inevitable: while some groups radicalize, others sustain margins of openness through cross-cutting interaction with diverse users. The findings illustrate how issue positioning, communicative style, and strategic engagement choices co-evolve in deindividualized online settings. Together, this paper advances theoretical understanding of online polarization and contributes new resources for computational analysis of controversial debates. Complete results, datasets and scripts available at https://github.com/BassiDavide/Navigating_Disagreement.git

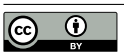
## CCS Concepts

• **Applied computing** → *Sociology*.

## Keywords

Stance Detection, Social Network Analysis, Controversy Analysis, Natural Language Inference, Polarization, YouTube

**ACM Reference Format:**
Bassi Davide, Renata Vieira, and Martín Pereira-Fariña. 2026. Navigating the Disagreement Space: A Case Study on Persistent YouTube Users' Interactions in Immigration-Related Discussions. In *18th ACM Web Science Conference (WebSci '26), May 26–29, 2026, Braunschweig, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3795766.3799757




*WebSci '26, Braunschweig, Germany*

ACM ISBN 979-8-4007-2504-3/2026/05
https://doi.org/10.1145/3795766.3799757

## 1 Introduction

The discursive climate on contemporary social networks (SNs) is deteriorating: antagonism and incivility now pervade political exchange. As such hostile norms consolidate, engaging in online public discourse becomes increasingly costly, leading many users to retreat from participation. This contraction of the deliberative space can heighten the risks of policy gridlock, social division, and real-world violence against vulnerable targets [3, 23, 24]. Understanding how these environments shape political interaction, and how users exercise agency in navigating and negotiating increasingly antagonistic dynamics, is therefore essential.

SNs, in fact, function as environments where users actively engage in peer-to-peer interaction to construct, negotiate, and consolidate their social identities [3]. Polylogical spaces like YouTube's comment section have increasingly become arenas of participatory identity performance [8], where users exploit disagreement to position themselves and perform group membership through stance-taking, relational alignment, and (dis)affiliative moves [5, 11]. YouTube provides a unique environment for observing these strategies. While often associated with in-group echoing and norm consolidation [18], recent work shows that its comment sections also sustain substantial cross-cutting interaction [14], allowing users to engage in both affiliative and confrontational identity work.

YouTube's volatile participation rules enable users to "take or leave the floor" with minimal social obligation [12]. Interactions remain largely anonymous, and identity is primarily enacted through salient social categories [1]. At the same time, the discursive environment of YouTube comment sections is shaped by the characteristics of the channel, both in terms of audience composition and engagement [22, 26] and with respect to the civility and the targets of comments [31, 33]. Users exercise agency not only through what they say, but also through where they choose to post, thereby generating an interactional space: the set of discursive arenas (YouTube channels, in our case) in which users express and contest opinions. Despite these distinctive features, little is known about sustained participation dynamics in such volatile environment.

In light of this, the present study examines how users' political stances relates to the ways they navigate YouTube's volatile discursive environment and sustain their positioning over time. Given the lack of prior longitudinal work on sustained participation in volatile comment spaces, we adopt a theory-driven exploratory design aimed at mapping systematic patterns rather than testing directional hypotheses. Specifically, we aim to answer the following research questions:

- RQ1: What is the relationship between persistent user's stance and their choice of interactional spaces on YouTube?
- RQ2: Does the interaction between stance and chosen interactional spaces relates to users' engagement modalities?
- RQ3: Does the interaction between stance and chosen interactional spaces map onto users' communicative style?
- RQ4: How does sustained participation in these volatile commenting environments correspond to issue-polarization trajectories?

Our main contributions are the following: (1) We present a longitudinal dataset of comments from nearly 3,500 US immigration-related YouTube videos spanning 2020–2024 and (2) a machine learning model to classify user positions on immigration, demonstrating the effectiveness of NLI formulation for stance detection in YouTube. Combining stance with interactional choices, we identify distinct clusters of users who display persistent participation patterns and analyze their strategies of discussion around this controversial topic. (3) Our findings reveal distinct interactive strategies: while pro-immigration users exhibit confirmation patterns, commenting mostly on left-leaning content, anti-immigration users display heterogeneous approaches, ranging from echo-chamber interaction to deliberate cross-cutting exposure. (4) These groups differ also in the way they engage with others: contra-immigration confirmation clusters tend to reinforce affinity within like-minded spaces, pro-immigration and contra-immigration cross-cutting clusters combine higher antagonism with greater nuance, revealing the complex picture of online political discourse as a site where identity affirmation and contestation coexist. (5) Finally, we compare polarization trajectories of persistent users with general audience. We show that, while this latter polarize predominantly toward contra-immigration positions, users engaged in cross-cutting exposure and discussions, even if in hostile ones, show lower polarization rates, suggesting a protective function of such behavior.

## 2 Related Works

*Selective Exposure Theory and Recommendation Algorithms* emphasize how message content, platform architecture, and algorithmic recommendations structure users' information environments [32]. According to selective exposure theory [30], these systems interact with users' preferences, reinforcing attitude-consistent and identity-driven behaviors, ultimately leading to the generation of filter bubbles, where engagement with like-minded peers foster ideological segregation [2, 19]. Within this view, YouTube has often been portrayed as a driver of radicalization [9, 27].

Yet, a growing body of work complicates this interpretation. Extremist material reaches only a minority of users [20], algorithmic effects are often weaker than previously assumed [4], and attempts to burst ideological bubbles can even intensify polarization [3]. Moreover, despite these dynamics, cross-cutting interactions persist in YouTube [15, 19], underscoring how polarization emerges not only from platform structures but from users' active navigation choices [16]. Together, this research shifts the emphasis toward models that integrate user agency and bottom-up identity work.

*Interpersonal Pragmatics.* shows that social media use is driven less by informational purposes than by relational and expressive forms of identity work [21]. On YouTube, peer-to-peer exchanges provide ad hoc opportunities for sociocultural comparison and value affirmation through patterns of (dis)affiliation with specific groups [11]. Prior work demonstrates that such identity work frequently relies on impoliteness and divisive rhetoric, which both reflect and reinforce group boundaries [1, 10, 11, 13]. In online controversies, users adapt their rhetorical choices to their degree of commitment and to the stance of interlocutor, strategically positioning themselves within the interactional landscape [5].

These studies, however, typically capture only snapshot of interaction. What remains underexplored is how identity work evolves for users who repeatedly return to these volatile, often hostile arenas. For persistent commenters, sustained engagement is not incidental: it becomes part of how they construct, defend, and perform social identity over time.

*Computational Analysis of Online Debates* of online debates on YouTube are still emerging. Röchert et al. [29] combine sentiment analysis with network methods to assess whether controversial issues are discussed within homogeneous clusters. Yet, YouTube API limits restrict them to shallow reply structures (direct replies vs. top-level comments). To address these limitations, Bassi et al. [6] develops a pipeline that reconstructs YouTube conversation chains and build stance-based interaction networks to analyze YouTube's discussions dynamics. Bassi et al. [7] extended this line of inquiry introducing methods for automatically evaluating conversation quality, distinguishing constructive from destructive disagreement in YouTube debates.

## 3 Experimental Setting

### 3.1 Dataset

Using Social Blade[1] we identified the top 100 YouTube U.S. based channels for number of subscribers categorized under "News and Politics". We focus on the top 25 channels to capture highly visible, agenda-setting arenas where political polarization and sustained user conflict are most pronounced. Each channel was assigned a Left/Right political orientation based on All Sides Media Bias[2] and Media Bias Fact Check[3], and manually categorized as "Content Creator" or "Legacy Media". Our dataset covers from January 2020 to December 2024, a full presidential cycle characterized by intense debate about immigration discourse, including major policy shifts and border crises. Video collection focused on immigration-related videos, identified via immigration-specific YouTube API query searches (see Repository) combined with manual content verification to ensure topical relevance. Channels with disabled comments were excluded, which introduced a Left-skew in our sample. To mitigate this imbalance, we removed the least-commented Left-leaning news channels. Table 1 reports the final dataset composition. Accordingly, the dataset is not intended to be representative of immigration discourse on YouTube at large, but to capture highly politicized comment spaces where persistent engagement and stance contestation are most likely to emerge.

Rieder et al. [28] notes that YouTube Data API may prioritize recent uploads, introducing a recency bias. To test wether this affected our data, we audited channels by computing a monthly video

[1] https://socialblade.com/
[2] https://www.allsides.com/media-bias/ratings
[3] https://mediabiasfactcheck.com/

| Left-Leaning | | | Right-Leaning | | |
|---|---|---|---|---|---|
| **Channel** | **Video** | **Comments** | **Channel** | **Video** | **Comments** |
| *Legacy Media* | | | *Legacy Media* | | |
| ABC News | 614 | 233,374 | Fox Business | 689 | 301,095 |
| CBS News | 721 | 193,369 | LiveNOW Fox | 330 | 121,946 |
| VICE News | 149 | 138,573 | | | |
| Inside Edition | 111 | 112,285 | | | |
| *Content Creat.* | | | *Content Creat.* | | |
| Young Turks | 261 | 167,345 | Ben Shapiro | 122 | 93,562 |
| David Pakman | 220 | 122,922 | Megyn Kelly | 168 | 61,795 |
| Tyler Cohen | 40 | 58,651 | Matt Walsh | 38 | 38,042 |
| | | | Charlie Kirk | 17 | 17,897 |
| | | | DailyWire+ | 19 | 12,495 |
| **Sub-Total** | **2,116** | **1,026,519** | **Sub-Total** | **1,383** | **646,832** |

**Table 1: Left- and Right-Leaning Channels in the Dataset**

counts, and estimated linear and correlational trends. The results show no systematic loss of older videos: sloped cluster around zero with mixed directions, and correlations with time remain small (see Repository). These mixed, low-magnitude patterns indicate that retrieval did not preferentially favor recent uploads, ensuring stable temporal coverage the dataset.

## 3.2 Comments' Stance Detection

Stance detection in YouTube comments is challenging due to the limitations of YouTube API to return explicit parent-child relationships. Applying the pipeline proposed by Bassi et al. [6], we rebuilt conversation chains and identified parent comments. Because our objective was to detect stance toward immigration as a general topic, not toward the parent comments, contextual information was insufficient: replies may agree with their parent while expressing an opposing immigration stance, or vice versa. This required to explicitly fix the target. We therefore cast stance detection as a Natural Language Inference (NLI) task. As shown in Figure 1, for each instance, the child comment with its parent as context (and the video title when top-level) served as *premise*, while the *hypothesis* "The comment supports immigration"[4], was used to encode the stance target. Entailment, contradiction and neutral labels were mapped to pro-, contra- and other-immigration stances. This formulation ensures that stance is always inferred relative to immigration rather than the parent comments. To help the model prioritize the child comment while still benefiting from contextual cues, we implemented a dual-condition training scheme, in which each instance was presented both as "child only" and "child+parent". We fine-tuned pretrained encoders on the dataset from Bassi et al. [5], consisting of GPT-4o-labeled YouTube comments (macro-F1=78.8 on human gold data), which provides an upper bound for our task. For evaluation, we expanded the original test set: two annotators with backgrounds in political science and sociology manually labeled additional comments following the same guidelines (Cohen's $\kappa = 0.61$). Disagreements were resolved through discussion, yielding a gold-standard test set of 2,000 comments.

[4]Multiple hypothesis formulations produced highly consistent results agreement (93.9%), so this single hypothesis was applied to the full dataset.

We trained RoBERTa-Large and DeBERTa-V3-Large[5] under three conditions: (1) child comment only classification, (2) parent+child classification, and (3) parent+child classification with NLI formatting. Table 2 shows that contextual information improved performance across models. The NLI approach yields the best results with DeBERTa-V3 (Macro-F1 = 73.75 ± 0.08), likely due to its disentangled attention mechanism, which may better separate child from parent information. Given these results, we scaled stance detection to the full dataset.

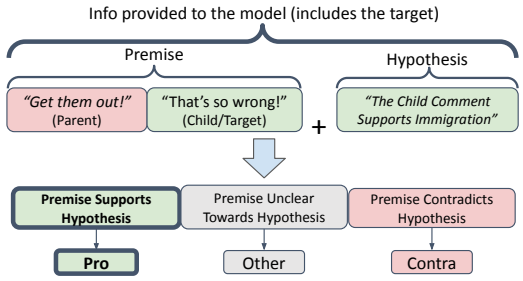



**Figure 1: NLI Stance Detection Approach**

## 3.3 Stance Based Users Classification

To identify the most persistent users we divided the dataset into quarters (4-month intervals), yielding 15 data points across the study period. To ensure a cohort of genuinely persistent users, we intentionally applied a strict inclusion criterion: users had to post at least one comment in at least 10/15 time periods. This returned us a set of 521 highly active users. This substantial reduction in users is explained both by our requirement for temporal consistency, and by the high churn characteristic of YouTube's deindividualized environment. To detect Pro/Contra-immigration users, we computed mean stance values ($\mu$) for each user across all time periods. As shown in Figure 2 the results revealed a bimodal distribution pattern in the aggregated data. To formally validate the bimodality and determine classification thresholds, we applied multiple statistical tests: a Gaussian Mixture Model identified two optimal components ($\mu_1 = 1.11$, $\mu_2 = 0.51$), Hartigan's bimodality coefficient (0.555) confirmed significant bimodality, and both AIC and BIC favored the two-component solution. The GMM components intersect at $\mu = 0.812$. However, to distinguish the group of users expressing an unclear stance, we employed a hybrid threshold approach: a lower boundary at the intersection point ($\mu = 0.8$) and an upper bound at the theoretical neutral position ($\mu = 1.0$). This methodology yielded the following user classification:

- Contra users ($\mu < 0.8$): 318 users (57.2%)
- Pro users ($\mu > 1.0$): 162 users (29.1%)
- Uncertain users ($0.8 \leq \mu \leq 1.0$): 76 users (13.7%)

## 3.4 Toxicity Detection

We analyzed comments' linguistic features using the Perspective API[6], which leverages models trained on millions of comments

[5]We used Optuna for hyperparameters calibration. Best setting can be found in the repository.

[6]https://developers.perspectiveapi.com/

| Model | Class | Support | Child Only | Parent+Child | Δ | Parent+Child (NLI) | Δ |
|---|---|---|---|---|---|---|---|
| DeBERTa-V3-Large | *Against* | 764 | 72.21 ± 0.76 | 72.46 ± 0.99 | +0.25 | **77.11 ± 0.53** | **+4.65** |
| | *Neutral* | 706 | 65.04 ± 1.07 | 69.07 ± 0.43 | +4.03 | **70.49 ± 0.06** | **+1.42** |
| | *Pro* | 534 | 68.73 ± 0.54 | 64.39 ± 0.85 | -4.34 | **73.66 ± 0.58** | **+9.27** |
| | Macro-Avg | 2004 | 68.66 ± 0.79 | 68.52 ± 0.07 | -0.14 | **73.75 ± 0.08** | **+5.23** |
| RoBERTa-Large | *Against* | 764 | 70.51 ± 0.71 | **72.63 ± 0.20** | **+2.12** | 72.63 ± 0.46 | 0.00 |
| | *Neutral* | 706 | 66.20 ± 1.00 | **70.97 ± 0.13** | **+4.77** | 69.75 ± 0.46 | -1.22 |
| | *Pro* | 534 | 66.45 ± 0.81 | 66.79 ± 1.10 | +0.34 | **68.62 ± 0.65** | **+1.83** |
| | Macro-Avg | 2004 | 67.72 ± 0.81 | 70.13 ± 0.41 | +2.41 | **70.33 ± 0.29** | **+0.20** |

**Table 2: Stance Detection Performance Across Different Training Conditions. Results show F1-scores with standard deviations over 3 runs. Best results per model and class are highlighted in bold. Improvements (Δ) show gains relative to the previous condition.**

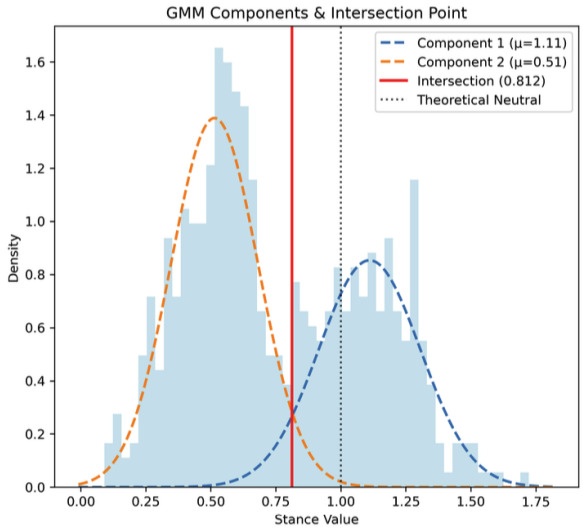



**Figure 2: Stance Distribution of Most Active Users**

annotated by multiple human raters to predict the perceived impact of comments on conversation quality. Among the many available attributes, we selected those most relevant to our research:

- *Toxicity*: Comments likely to make people leave a discussion due to rude, disrespectful, or unreasonable content
- *Insult*: Inflammatory or negative comments directed at individuals or groups.
- *Attack on commenter*: Direct attacks on fellow commenter.
- *Identity attack*: Negative comments targeting individuals based on identity characteristics.
- *Affinity* (experimental): References to shared interests, motivations, or outlooks between commenter.
- *Curiosity* (experimental): Attempts to seek clarification or ask follow-up questions to better understand others' perspectives.
- *Nuance* (experimental): Incorporating multiple points of view to provide detail or context.

# 4 Results

## 4.1 Choice of Commenting Channel Analysis

***RQ1: What is the relationship between persistent user's stance and their choice of interactional spaces on YouTube?*** With RQ1 we analyzed how stance-based groups distributed their activity across different interactional spaces. For each user, we tracked their commented channels and computed the proportion of comments posted on Left-leaning-News, Right-leaning-News, Left-leaning-Creators and Right-leaning-Creators, yielding four-dimensional interactive-community vectors. Given our aim of identifying ideal-typical engagement strategies - rather than density-driven micro-clusters - and the low dimensional, compositional structure of these vectors, we used K-means, which provides stable and interpretable centroid-based partitions aligned with our theoretical frame. Optimal cluster numbers were determined via silhouette scores[7]. Pro group yielded two clusters (k=2, silhouette = 0.844; n1=17, n2=145), Contra group yielded three clusters (k=3, silhouette = 0.685; n=98, n=135, n=85), and Control group yielded two clusters (k=2, silhouette = 0.875; n1=16; n2=60). To ensure statistical reliability, we removed the clusters with $n < 20$ (complete results including these clusters are available in the Repository and show no relevant differences). Average stance scores across groups remained consistent with initial classifications, with residual variations attributable to topic-specific opinion nuance and potential stance-detection uncertainty. Final user groups were labeled based to their characteristic stance-based interactive profile, see Table 3 (see Repository for PCA projection of users based on interactional-space features).

To compare commenting patterns across stance-based clusters, we computed all measures at the user level and then averaged them within clusters (see Appendix). The results in Figure 3 reveal distinct approaches YouTube users employ to navigate their social positioning on immigration. Pro group, the most active one, demonstrates a clear "confirmation strategy", commenting predominantly on left-leaning content. Interestingly, also the Control group shows the same pattern, despite not articulating an explicit stance on immigration in their commentary. In contrast, the Contra groups exhibit three differentiated strategies. Contra-Echo mirrors the Pro

[7]We assessed clustering robustness by re-running K-means under stricter user persistence thresholds (≥ 11 and ≥ 12 periods) and alternative feature normalization (z-scored vs raw interactional proportions). Cluster assignments remained highly consistent with the main analysis (ARI ≥ 0.90 across groups), and cluster centroids preserved the same qualitative interactional profiles. See Repository.

| Group | Statistics & Description |
|---|---|
| Pro | n=145; Stance=1.23±0.62; comm.=8509. Pro-immigration users consuming primarily left-leaning content. |
| Contra-Echo | n=135; Stance=0.48±0.59; comm.=7010. Anti-immigration users consuming primarily right-leaning content. |
| Contra-Balance | n=85; Stance=0.56±0.64; comm.=4798. Anti-immigration users with balanced cross-partisan consumption. |
| Contra-Cross | n=98; Stance=0.46±0.62; comm.=6947. Anti-immigration users consuming primarily left-leaning content. |
| Control | n=60; Stance=0.9±0.62; comm.=2994. Users holding an uncertain stance towards immigration, posting mostly on left channels. |

**Table 3: Interactive Stance-Based Groups**

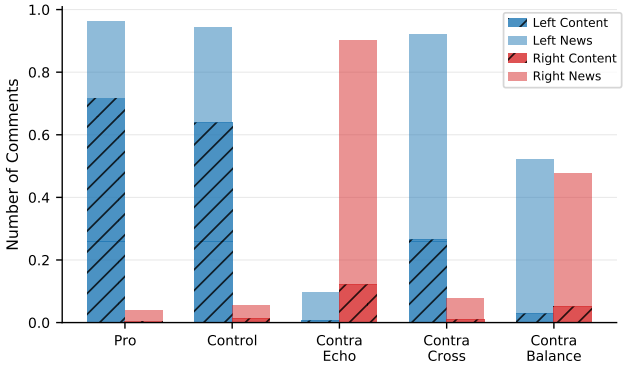



**Figure 3: Channels Consumption by Stance-Base Group**

group's confirmation approach, though from the opposite political side. Contra-Balance adopts a deliberate "cross-cutting exposure strategy" posting both on left and right content, mostly equally. Contra-Cross employs a counterintuitive strategy, commenting primarily on opposing channels (left-leaning) despite holding anti-immigration views, manifesting a deliberate strategy that challenges traditional echo chamber assumptions and indicates that some users may actively seek ideological opposed spaces. Finally, Pro and Control groups show strong preferences for Content Creators over News Channels across their respective political leanings, suggesting that the interactive spaces generated by such channels may better suits identity expression for users articulating such position within the broader "discursive environment". In stark contrast, Contra-Echo users demonstrate the opposite pattern: they comment significantly more in News Channels than Content Creators. This pattern is mirrored also by Contra-Balance and Contra-Cross users (although with some variance for the latter) indicating a preference for interacting in more institutionalized, formal channels for users holding opposing views towards immigration.

## 4.2 Engagement Patterns Analysis

***RQ2: Does the interaction between stance and chosen interactional spaces relates to users' engagement modalities?***

*Reply Pattern.* To examine whether stance based groups differ in the way they engage in discussions in YouTube, we measured the proportion of comments directed toward videos (top-level comments) versus comments directed toward other users (replies) across the five user groups. As before, proportions were computed at the

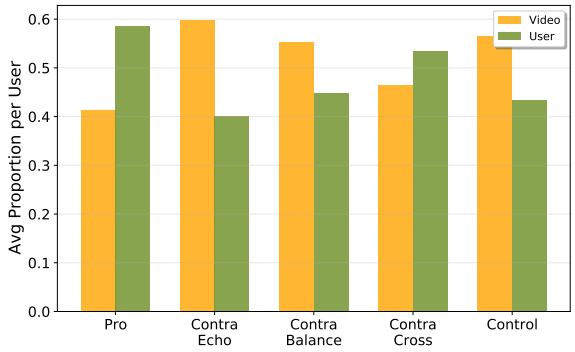



**Figure 4: Commenting Pattern Comparison: Video vs Users**

user level and then averaged within the groups. The analysis reveals systematic differences across groups (Figure 4). The Pro group demonstrates the highest propensity for user-directed engagement (58.7% vs. 41.3% video-directed), suggesting a preference for interpersonal discussion over direct video commentary. Conversely, Contra-Echo users demonstrate the highest propensity for video-directed engagement, reflecting a more declarative, broadcast-style mode of participation. Contra-Cross is the only Contra subgroup with a user-directed majority (53.5% vs. 46.5%; +7 pp), whereas Contra-Balance (55.2% vs. 44.8%; +10.4 pp) favor video-directed replies. These results align with these groups' more "complex social positioning profiles": Contra-Balance and Contra-Cross, in fact, represent departure from typical echo chamber dynamics, though in different ways. Finally, users with weaker issue positioning (Control) showed a more video-prone engagement approach (56.5% vs. 43.5%; +13.0 pp), suggesting a less confrontational style.

*Stance Targeting.* To examine whether users preferentially engage with ideologically aligned or opposing viewpoints, we analyzed stance-targeting behavior in user-to-user replies. To correct structural availability bias (i.e. when reply targets reflect stance prevalence in a video rather than user preference) we computed for each user the proportion of replies directed toward Contra (0), Neutral (1), and Pro (2) in a video, then normalized these by the stance distribution present in the video's comment section, representing the stances the user was exposed to when replying. We obtain a preference index representing the ratio of actual to expected targeting: values greater than 1.0 indicate over-targeting (active preference), values less than 1.0 indicate under-targeting (avoidance), and values approximately equal to 1.0 suggest engagement consistent with content availability. Figure 5 shows how Contra user groups exhibit a consistent preference for engaging with ideologically aligned comments. While this pattern is expected for Contra-Echo users operating within their ideological comfort zone, it proves more notable for Contra-Balance users who, despite participating across a broader set of commenting enviroments, still preferentially engage with like-minded interlocutors. Contra-Cross users display a distinct pattern: their over-targeting of contra-users suggest a strategy of reinforcing pre-existing views by "going behind enemy lines" and seeking out allies withing adversarial spaces. At the

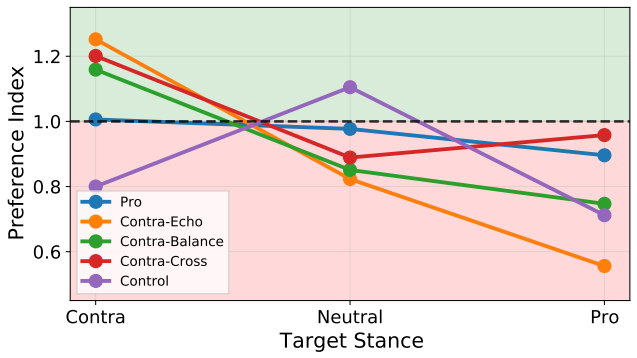



**Figure 5: Stance targeting preferences beyond random availability. Preference index $> 1.0$ indicates over-targeting (preference), $< 1.0$ under-targeting (avoidance), and $\approx 1.0$ availability bias based-targeting.**

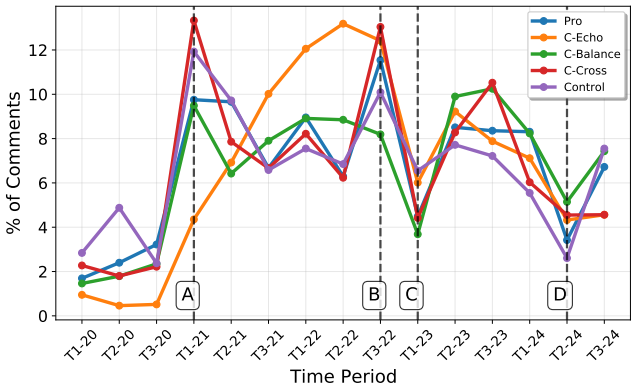



**Figure 6: Engagement Patterns in Function of Real-World Immigration Related Events**

same time, their weak avoidance of Pro commenters indicates openness to direct confrontation, reflecting a unique blend of selective exposure and engagement-seeking. The Pro group demonstrates engagement patterns consistent with availability bias. This behavior can be attributed to the minority status of Pro positions across most videos in the dataset (see Table 4), resulting in engagement patterns that reflect the relative scarcity of ideologically congruent content rather than deliberate selective exposure. Control group tendency to direct their replies toward neutral/other indicates users without strong ideological positioning tendency to gravitate toward moderate conversational spaces, avoiding partisan debate.

*Real-World Events' Reaction.* We examine how key immigration real-world events shaped users' interactive behaviors across groups from 2020–2024 (see Figure 6). Several distinct peaks correspond to major policy developments: the T1-21 surge (A) coincides with Biden's inauguration and immediate policy rollbacks; the T3-22 peak (B) corresponds to migrant bussing controversies initiated by Republican governors; the T1-23 rebound (C) aligns with Title 42's expiration in May 2023, though reduced media salience due to competing issues (debt ceiling, classified documents scandals) dampened responses; the subsequent T2-23 peak corresponds to record border crossings and intensified "border crisis" discourse; and the T2-24 decrease (D) might indicates media coverage broadening into general election campaigning (Trump vs. Biden rematch), where immigration was one of many issues (e.g. inflation, abortion, foreign policy). Notably, Contra-Echo exhibits distinctive temporal profile, suggesting that echo chamber dynamics produce distinct engagement rhythms. Between T1-21 and T3-22, also Contra-Balance shows a trajectory more aligned with Contra-Echo patterns than with other groups. We hypothesized that this convergence stems from their shared tendency to engage primarily within the same right-wing legacy media communities, making their engagement rhythms set more by these channel's agenda setting cycles rather than real-world events. We tested this hypothesis comparing C-Echo and C-Balance (target groups) against Pro, C-Cross, and Control groups. We first applied K-means clustering to users' temporal activity vectors and evaluated the association between cluster membership and group identity using a $\chi^2$ test, which revealed clear temporal structuring ($\chi^2 = 29.59$, $p < 0.001$) with 80.5% of target users falling in the same cluster. We then compared peak-activity timing distributions using a Kolmogorov–Smirnov test, showing that target groups tended to reach their activity peak one period later than other users (KS, $p < 0.001$). Finally, Mann–Whitney U tests run period-by-period identified significant differences in 10 of 20 intervals ($p < 0.05$), with medium–large effect sizes (Cohen's $d > 0.5$) during major immigration events. Overall, these results indicate that C-Echo and C-Balance share distinct temporal engagement patterns shaped by their common interactional environments.

## 4.3 Linguistic Analysis

***RQ3 Does the interaction between stance and chosen interactional spaces map onto users' communicative style?*** To examine how users' stance and interactional choices influence linguistic behavior, we conducted a stratified analysis of toxicity attributes derived from the Perspective API (Section 3.4). Our analytical approach addressed two key confounding factors. First, we stratified comments by length using tertile-based divisions (short, medium, long) to control for the correlation between comment length and toxicity scores, as we expected longer comments to provide more opportunities for toxic language expression. Second, recognizing the controversial communicative contexts in YouTube discussions, we separately analyzed video-directed comments (Level 0) and user-directed replies (Level> 0). The analysis employed user-level aggregation, where individual comment scores were first averaged by user, then by experimental group, before conducting ANOVA tests. This approach mitigates the influence of highly prolific users and ensures group-level comparisons reflect user behavior rather than comment volume. See Repository for complete results.

As shown in Figure 7, the length-stratified analysis reveals several key patterns. Most linguistic attributes demonstrate the anticipated length effect, with scores generally increasing from short to long comments. However, "attack on commenter" scores remain consistently elevated across all groups and lengths (ranging from 0.20-0.47), underscoring the persistently adversarial nature of

YouTube discussions on controversial topics. Across all measures, user-directed replies consistently exhibit higher toxicity levels than video comments, with particularly pronounced differences in "attack on commenter" scores (replies: 0.28-0.46 vs. video comments: 0.21-0.26) and "toxicity" (replies: 0.10-0.24 vs. video comments: 0.10-0.18), demonstrating that interpersonal exchanges escalate conflict beyond video-directed commentary.

Between-group differences emerged more strongly in replies than video comments, further suggesting that identity-driven dynamics are activated in interactive exchanges, whereas monologic, video-directed comments remain comparatively less differentiated. For short replies, significant group differences were observed across multiple toxicity dimensions, most notably "toxicity" (F=6.04, $p < 0.001$) and "attack on commenter" (F=11.20, $p < 0.001$). Medium-length replies showed significant differences in "attack on commenter" attribute (F=7.87, $p < 0.001$), while long replies displayed significant group effects for "inflammatory language" (F=5.43, $p < 0.001$) and "attack on commenter" (F=12.16, $p < 0.001$). Video comments showed fewer significant group differences, with the notable exceptions of "nuance" scores in long comments (F=5.08, $p < 0.001$).

Contra-Echo users show high "affinity" scores across all strata (0.42-0.60), consistent with echo-chamber interaction patterns, while simultaneously showing among the lowest "curiosity" scores (0.24-0.34), suggesting reduced openness to alternative perspectives. Contra-Cross users exhibit the most confrontational linguistic style, with elevated levels of "attack on commenter" (0.25-0.46) and "identity attack" (0.03-0.09), reflecting the adversarial dynamics that arise when engaging across ideological lines. Pro users register higher "toxicity" across contexts (video comments: 0.12-0.18; replies: 0.17 for short replies, highest among all groups), a pattern consistent with defensive engagement as a minority stance, as indicated by their stance targeting analysis in Section 4.2. Despite their elevated hostility, both Contra-Cross and Pro groups demonstrate the highest "nuance" scores, indicating that even contentious discussions can incorporate multiple perspectives and contextual elaboration.

We complemented these style-based measures with a lexical analysis of the expressions that are most salient for each group. Specifically, we extracted bigrams from all comments and identified those disproportionately associated with the group using a log-odds ratio with an informative Dirichlet prior, contrasting each group against all others. As shown in Figure 8, Pro users are characterized by bigrams emphasizing procedural and legal aspects of immigration (e.g., pathways to legal status, waiting times, citizenship), alongside references to asylum and migrants' living conditions. Contra-Echo users disproportionately employ expressions related to national identity, economic burden, frequently intertwined with partisan slogans and political figures. Contra-Balance users are marked by the prominence of electoral and partisan expressions, with immigration-related terms often embedded within broader political grievance discourse. Contra-Cross users exhibit a higher salience of security-related and moral prioritization language, including references to crime, cartels, and deservingness. Finally, Control users display comparatively descriptive and institutional language, referencing border procedures, legal entry, and historical or administrative contexts.

## 4.4 Issue Polarization Trends

**RQ4: How does sustained participation in these volatile commenting environments correspond to issue-polarization trajectories?** We conceptualize issue polarization on immigration as the concentration of expressed opinions at the pro and contra poles, accompanied by a decline in neutral or ambiguous positions. In our stance scheme, this corresponds to shifts in the relative frequency of Contra (0), Neutral (1), and Pro (2) comments away from the neutral category and toward the two extremes. Accordingly, for each channel (General Audience), we treated the yearly proportions of contra, neutral, and pro as an empirical approximation of the audience stance distribution, and estimated linear regression slopes on these proportions to capture the direction and rate of polarization over time. For persistent users, we applied the same logic at the user level, first computing yearly stance proportions for each user, and then averaging these within clusters, so that trajectories reflect typical user behavior rather than being skewed by differences in comment volume. From the results in Table 4, several patterns emerge: General audience comments exhibit a general trend of polarization, characterized by a decline in neutral stances. This shift is largely towards contra-immigration positions, except in Left-leaning content creators, where polarization is more evenly distributed across both directions. On the other hand, persistent users' groups display more heterogeneous dynamics: Pro-Immigration users gradually shift further left, consistent with interaction patterns that keep them embedded in like-minded discussion spaces. Contra-immigrationn users who remain within ideologically aligned environments likewise show steadily intensifying polarization. The users group with less pronounced issue positioning (Control), mirrors the trajectory of the general audience, despite engaging mostly in left-leaning channels. By contrast, contra-immigration users who actively engage across ideological boundaries (Contra Balanced and Contra Cross) exhibit the lowest levels of stance change over time. Taken together, these findings highlight two main insights. (1) Polarization trajectories depend on the combination of interactional choices and participation style: general audiences polarize broadly, while persistent commenters polarize in sharply differentiated ways — with echo-chamber engagement amplifying shifts, and cross-cutting exposure dampening them. (2) Initial stance clarity shapes susceptibility to change: users with weakly articulated positions (Control group) behave much like the general audience, showing pronounced losses in neutrality and movement toward polarized positions, whereas to other persistent users' trajectories are more shaped by how and with whom they choose to interact.

# 5 Discussions

This study examined how YouTube users navigate immigration debates, focusing on persistent commenters whose longitudinal activity reveals how stance and interactive styles co-evolve in YouTube. In the anonymous and volatile environment typical of YouTube, our results illuminate the different ways through which social identity is constructed in the absence of stable community norms.

**Contra-Echo.** This group exhibits the most prototypical echo chamber dynamics, showing a marked preference for interacting within ideologically aligned discussion spaces and for engaging

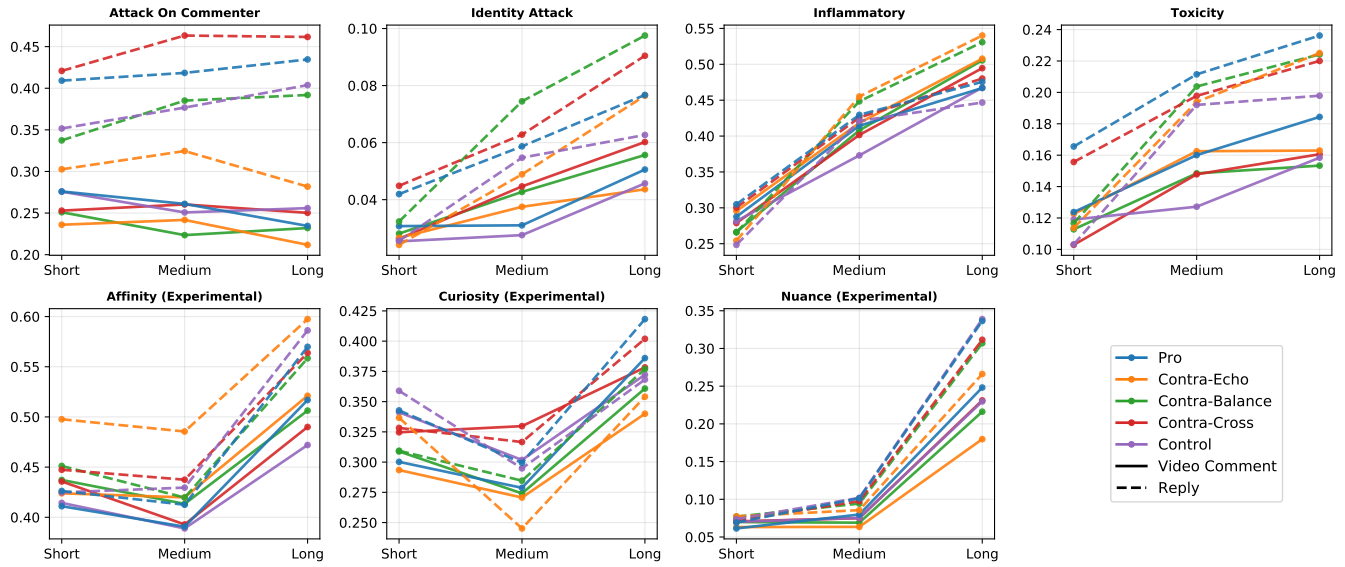



**Figure 7: Linguistic Attributes per Length Level**

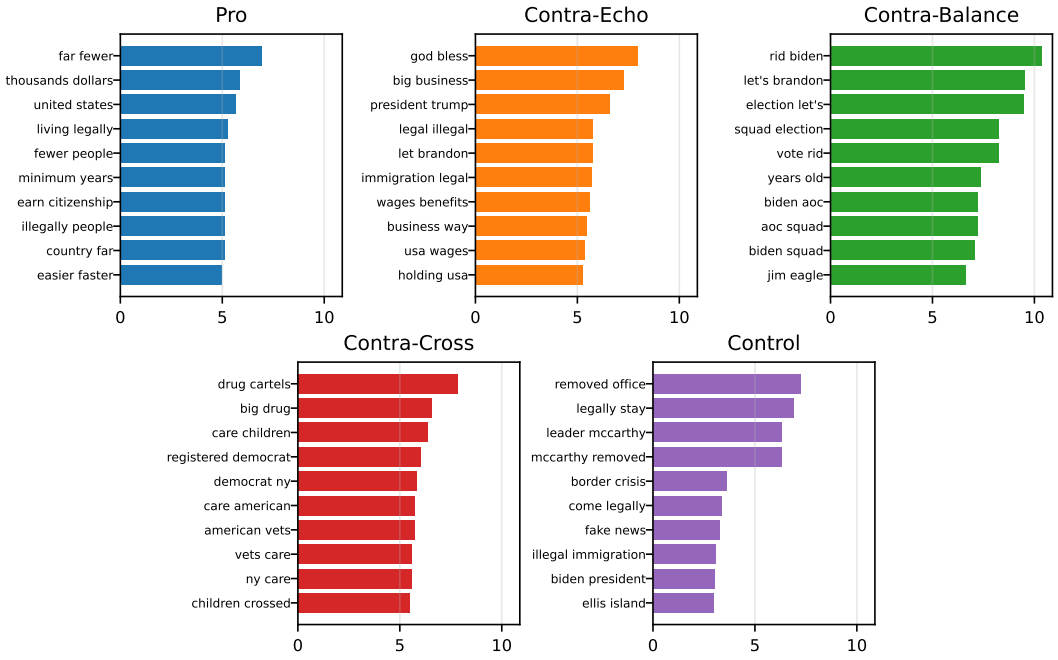



**Figure 8: Top 10 Group Bigrams Log-odds**

mostly with users reflecting their existing beliefs [19]. As shown by RQ1, they actively select right-leaning comment arenas as their primary sites of interaction, avoiding confrontational exchanges and preferring monologic, video-directed commenting over dialogic engagemtnt with other users. Coherently, when interacting with others, they restrict replies to like-minded users while avoiding opposing viewpoints. Linguistically, this is reflected in high affinity and low curiosity, coupled with minimal direct attacks. This "interactional closure" is also mirrored in their lexical emphasis on national identity and economic burden expressions, often intertwined with partisan slogans. Counterintuitively, their comments are also characterized by elevated toxicity, however research in interpersonal pragmatics demonstrates that such language can serve

| | Contra | | Neutral | | Pro | |
|---|---|---|---|---|---|---|
| **Group** | Mean | Slope | Mean | Slope | Mean | Slope |
| *General* | | | | | | |
| Right Creator | 26.9 | +4.16% | 64.7 | -4.98% | 8.3 | +0.82% |
| Left News | 30.0 | +4.61% | 53.2 | -5.59% | 16.8 | +0.99% |
| Right News | 44.7 | +3.17% | 44.8 | -3.92% | 10.5 | +0.76% |
| Left Content | 16.8 | +1.94% | 63.2 | -3.78% | 20.0 | +1.84% |
| *Active* | | | | | | |
| C-Echo | 54.1 | +4.11% | 42.2 | -4.24% | 3.7 | +0.13% |
| C-Balance | 56.4 | +1.86% | 36.4 | -0.91% | 7.2 | -0.95% |
| C-Cross | 53.4 | +1.41% | 38.3 | -1.53% | 8.3 | +0.11% |
| Control | 24.6 | +3.46% | 63.3 | -4.94% | 12.0 | +1.48% |
| Pro | 10.3 | +0.75% | 60.7 | -3.62% | 29.0 | +2.88% |

**Table 4: Channel Audience vs. User Groups: Stance Over Time**

as a form of in-group affiliative work than mere hostility [10, 11]. These echo chamber characteristics are further corroborated by RQ2 and RQ4: their participation patterns remain internally driven rather than event-responsive patterns (RQ2), and their stance trajectories show an increasing polarization toward anti-immigration stances over time (RQ4).

**Pro.** Although this group predominantly participate in like-minded channels (left leaning content creators), their behavior departs from a classic echo chamber profile. As RQ2 shows, they actively engage in user-to-user exchanges more frequently, often with opposing voices. This pattern can be connected to pro-immigration users being the minority across channels, RQ2 analysis of their targeting preferences, in fact, showed that, rather than displaying a systematic preference for aligned, their interactive partners largely depend on contextual availability. This structural asymmetry, combined with the consistent presence of contra-users across channels (see Table 7), helps explain their elevated levels of toxicity and direct attacks. Yet, despite this adversarial context, pro-immigration users also exhibit high levels of curiosity and nuance, indicating an openness when discussing, further differentiating the behavior of this group from the classic "echo chamber" model. Consistently, their discourse emphasizes procedural and humanitarian aspects of immigration, including legality, waiting times, and pathways to citizenship.

**Contra-Balance.** Despite holding a clear anti-immigration stance, this group displays a distinctive cross-exposure pattern, visiting both left and right leaning channels in near equal measure. Rather than signaling openness to debate, this pattern reflects a mode of participation centered on broad topical monitoring, as their behavior in RQ2 shows: they predominantly comment on videos rather than engage other users directly. Their lexical profile further reflects this orientation, with immigration-related terms frequently embedded within electoral and partisan grievance discourse rather than issue-specific framing. This breadth does not extend to interactional diversity: when they do enter conversations, they consistently avoid pro-immigration voices and seek alignment with contra peers. Linguistically, their elevated levels of toxicity and identity attacks mirror the affiliative impoliteness strategies for in-group bonding observed in Contra-Echo users [11]. At the macro level, their responsiveness to external events resembles more echo-chamber dynamics (e.g. Contra-Echo groups), but their relatively low polarization over time (RQ4) suggests that sustained exposure to ideologically mixed environments - even without direct cross-partisan engagement - may temper radicalization trajectories.

**Contra-Cross.** This group maintains a consistent anti-immigration stance yet interacts predominantly in left-leaning channels, making them atypical within the contra spectrum. Their linguistic profile, marked by high levels of identity and commenter attacks, might suggests classic trolling behavior [17], yet their interactive style reveals important nuances. In fact, while tending to interact with opposing views through inflammatory language (like a troll would do), when available, they also display one of the strongest preferences for interacting with like-minded users when such peers are present, indicating a structured alignment strategy rather than indiscriminate provocation. Crucially, as RQ3 shows, they exhibit the highest levels of curiosity in replies, a feature inconsistent with canonical troll behavior [25] and suggestive of situated openness within hostile exchanges. At the content level, this group disproportionately foregrounds security-related and moral prioritization language (e.g., crime, cartels, deservingness), consistent with confrontational engagement in opposing spaces. Finally, they display the lowest polarization among contra groups, and even a slight drift toward pro-immigration positions. This pattern suggests that engagement across ideological boundaries, while linguistically aggressive, does not preclude incremental stance flexibility over time.

**Control.** This group consists of users with weak stance orientations (Avg=0.9±0.62). Although they post predominantly in left-leaning content channels, their stance trajectories display one of the sharpest shifts toward contra-immigration positions over time. Interactionally, they tend to engage with other moderate/uncertain users, suggesting that their polarization may be driven less by YouTube dynamics than by influences external to the platform. Their lexical patterns are correspondingly descriptive and institutional, emphasizing legal entry procedures and administrative contexts.

Overall these results underscores that, among highly active commenters polarization on YouTube is not an inevitable outcome of algorithmic exposure alone, but emerges through users' strategic engagement across interactional spaces, communicative styles, and the thematic lenses through which immigration is articulated.

## 6 Conclusions

This study contributes to the understanding of online polarization by focusing on a rarely examined group: highly active and persistent YouTube commenters. By tracing how stance and interactive styles intersect in YouTube's volatile environment, we show that disagreement around controversial issues, such as immigration, is navigated through a combination of *where* users choose to engage, with *who* they decide to interact with and *how* they linguistically articulate their positions. Rather than exhibiting uniform dynamics, in fact, highly active users display heterogeneous engagement strategies, ranging from echo-oriented participation to adversarial cross-partisan interaction and broad topical monitoring. These results can help to shed light on the divergent conclusions on echo chambers and cross-cutting exposure effects, suggesting that these effects depend also on how users actively perform social positioning in controversial setting adjusting their participation

across spaces, discourses and modes of engagement. Attending to these intersections provides a more nuanced picture of online debates: polarization is neither uniform nor inevitable. Sustained engagement in ideologically mixed environments, even in hostile exchanges, can coexist with lower polarization trajectories and persistent openness, underscoring the importance of examining disagreement as a multi-dimensional and user-driven process. Practically, we contribute a large-scale dataset of YouTube comments and a methodological pipeline that reformulates stance detection as an NLI task, enabling scalable and context-sensitive classification in YouTube discussions. These resources and methods extend the analytical toolkit for computational social science, allowing political identity work to be traced longitudinally within complex online environments like YouTube.

## Limitations and Future Works

**Data**: our dataset reflects structural biases as it includes only channels with enabled comments, focuses on U.S. immigration debates, and captures a self-selected subset of highly active commenters rather than the broader viewing audience. Future studies should expand across topics, countries, and platforms, delving more into persistent vs occasional commenters comparisons.

**Methods**: stance detection, though improved via NLI reformulation, can misclassify sarcasm or implicit cues. Moreover it remains an abstraction of a more complex phenomenon: positions on immigration are not reducible to a binary pro/contra distinction and are articulated through multiple frames and semantic dimensions. Our lexical analysis partially addresses this limitation by illustrating systematic differences in thematic emphasis across interactional groups; however, future work should develop stance models that are explicitly sensitive to framing, gradience, and issue-specific nuance. Similarly, toxicity detection through the Perspective API captures surface-level patterns but may overlook irony and coded language. Future work should refine stance and toxicity detection with more context-aware models. Furthermore, while we considered where users commented, we did not analyze videos' content; future research should integrate video framing.

**Interpretation**: our analysis is correlational, so we cannot determine whether interaction patterns drive stance change or the reverse. Peaks in activity aligned with immigration events, yet other concurrent factors likely shaped behavior. Future research should combine longitudinal trace data with experimental or survey designs to test causal mechanisms and account for unobserved factors such as personality, demographics, or offline contexts.

**Trolls**: following Phillips [25], trolling represents an internet subculture with specific identity performance norms. Hence, while trolls engage in identity work consistent with our framework, future research should distinguish subcultural trolling practices to better understand how they shape antagonism, curiosity, or stance trajectories.

## Ethic Statement

This study advances understanding of polarization and cross-cutting engagement on YouTube, highlighting both radicalization risks and mechanisms that sustain openness. The findings may inform interventions fostering healthier online debates. We also acknowledge key risks. The dataset contains offensive, handled by minimizing direct quotation and reporting aggregate results. To avoid profiling, all longitudinal data were anonymized and analyzed at group level. Data collection complied with YouTube's Terms of Service.

## References


[1] Marta Andersson. 2021. The climate of climate change: Impoliteness as a hallmark of homophily in YouTube comment threads on Greta Thunberg's environmental activism. *Journal of Pragmatics* 178 (2021), 93–107. doi:j.pragma.2021.03.003

[2] Chen Avin, Hadassa Daltrophe, and Zvi Lotker. 2024. On the impossibility of breaking the echo chamber effect in social media using regulation. *Scientific reports* 14, 1 (2024), 1107.

[3] Chris Bail. 2022. *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. Princeton University Press, Princeton. doi:10.1515/9780691246499

[4] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[5] Davide Bassi, Giovanni Da San Martino, Renata Vieira, and Martin Pereira-Farina. 2025. Drawing digital lines: pattern analysis of divisive rhetoric in social network discussions. *Humanities and Social Sciences Communications* 12, 1 (2025), 2009. https://doi.org/10.1057/s41599-025-06277-7

[6] Davide Bassi, Michele Joshua Maggini, Renata Vieira, and Martín Pereira-Fariña. 2024. A Pipeline for the Analysis of User Interactions in YouTube Comments: A Hybridization of LLMs and Rule-Based Methods. In *2024 11th International Conference on Social Networks Analysis, Management and Security SNAMS*. 146–153. doi:10.1109/SNAMS64316.2024.10883781

[7] Davide Bassi, Erik Bran Marino, Renata Vieira, and Martin Pereira. 2025. Old but Gold: LLM-Based Features and Shallow Learning Methods for Fine-Grained Controversy Analysis in YouTube Comments. In *Proceedings of the 12th Argument mining Workshop*, Elena Chistova, Philipp Cimiano, Shohreh Haddadan, Gabriella Lapesa, and Ramon Ruiz-Dolz (Eds.). Association for Computational Linguistics, Vienna, Austria, 46–57. doi:10.18653/v1/2025.argmining-1.5

[8] Phil Benson. 2016. *The discourse of YouTube: Multimodal text in a global context*. Routledge.

[9] Omran Berjawi, Danilo Cavaliere, Giuseppe Fenza, and Vincenzo Loia. 2024. Understanding radicalization pathways: a framework for assessing diversity in YouTube recommendation systems. *Social Network Analysis and Mining* 14, 1 (2024), 233. doi:10.1007/s13278-024-01394-8

[10] Pilar Garcés-Conejos Blitvich. 2010. The YouTubification of politics, impoliteness and polarization. In *Handbook of research on discourse behavior and digital communication: Language structures and social interaction*. IGI Global, 540–563. doi:10.4018/978-1-61520-773-2.ch035

[11] Pilar Garcés-Conejos Blitvich, Nuria Lorenzo-Dus, Patricia Bou-Franch, Istvan Kecskes, and Jesús Romero-Trillo. 2013. Relational work in anonymous, asynchronous communication: A study of (dis) affiliation in YouTube. *Research trends in intercultural pragmatics* (2013), 343–366. doi:10.1515/9781614513735.343

[12] Patricia Bou-Franch and Pilar Garcés-Conejos Blitvich. 2014. Conflict management in massive polylogues: A case study from YouTube. *Journal of Pragmatics* 73 (2014), 19–36.

[13] Michael S Boyd. 2014. (New) participatory framework on YouTube? Commenter interaction in US political speeches. *Journal of Pragmatics* 72 (2014), 46–58. doi:10.1016/j.pragma.2014.03.002

[14] Seung Woo Chae and Sung Hyun Lee. 2024. Where do cross-cutting discussions happen?: Identifying cross-cutting comments on YouTube videos of political vloggers and mainstream news outlets. *Plos one* 19, 5 (2024), e0302030. doi:10.1371/journal.pone.0302030

[15] Gianmarco De Francisci Morales, Corrado Monti, and Michele Starnini. 2021. No echo in the chambers of political interactions on Reddit. *Scientific reports* 11, 1 (2021), 2818. doi:10.1038/s41598-021-81531-x

[16] Kiran Garimella, Tim Smith, Rebecca Weiss, and Robert West. 2021. Political Polarization in Online News Consumption. *Proceedings of the International AAAI Conference on Web and Social Media* (May 2021), 152–162.

[17] Claire Hardaker. 2013. "Uh.... not to be nitpicky, but… the past tense of drag is dragged, not drug.": An overview of trolling strategies. *Journal of Language Aggression and Conflict* 1, 1 (2013), 58–86.

[18] Muhammad Haroon, Magdalena Wojcieszak, Anshuman Chhabra, Xin Liu, Prasant Mohapatra, and Zubair Shafiq. 2023. Auditing YouTube's recommendation system for ideologically congenial, extreme, and problematic recommendations. *Proceedings of the national academy of sciences* 120, 50 (2023), e2213020120. doi:10.1073/pnas.2213020120

[19] David Hartmann, Sonja Mei Wang, Lena Pohlmann, and Bettina Berendt. 2025. A systematic review of echo chamber research: comparative analysis of conceptualizations, operationalizations, and varying outcomes. *J. Comput. Soc. Sci.* 8, 2 (2025), 52. doi:10.1007/S42001-025-00381-Z

[20] Homa Hosseinmardi, Amir Ghasemian, Aaron Clauset, Markus Mobius, David M Rothschild, and Duncan J Watts. 2021. Examining the consumption of radical

content on YouTube. *Proceedings of the national academy of sciences* 118, 32 (2021), e2101967118.
[21] Su Jung Kim. 2023. The role of social media news usage and platforms in civic and political engagement: Focusing on types of usage and platforms. *Computers in Human Behavior* 138 (2023), 107475.
[22] Julien Labarre. 2024. French Fox News? Audience-level metrics for the comparative study of news audience hyperpartisanship. *Journal of Information Technology & Politics* 21, 4 (2024), 510–527. doi:10.1080/19331681.2023.2300845
[23] Jennifer McCoy, Tahmina Rahman, and Murat Somer. 2018. Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American behavioral scientist* 62, 1 (2018), 16–42. doi:10.1177/0002764218759576
[24] Jennifer McCoy and Murat Somer. 2019. Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies. *The Annals of the American Academy of Political and Social Science* 681, 1 (2019), 234–271. doi:10.1177/0002716218818782
[25] Whitney Phillips. 2015. *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture.* Mit Press.
[26] Eugenia Ha Rim Rho, Gloria Mark, and Melissa Mazmanian. 2018. Fostering Civil Discourse Online: Linguistic Behavior in Comments of MeToo Articles across Political Perspectives. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 147 (nov 2018), 28 pages. doi:10.1145/3274416
[27] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. 2020. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 131–141. doi:10.1145/3351095.3372879
[28] Bernhard Rieder, Adrián Padilla, and Oscar Coromina. 2025. Forgetful by design? A critical audit of YouTube's search API for academic research. *Information, Communication & Society* (2025), 1–20.
[29] Daniel Röchert, German Neubaum, Björn Ross, Florian Brachten, and Stefan Stieglitz. 2020. Opinion-based homogeneity on YouTube: Combining sentiment and social network analysis. *Computational Communication Research* 2, 1 (2020), 81–108.
[30] Natalie Jomini Stroud. 2010. Polarization and partisan selective exposure. *Journal of communication* 60, 3 (2010), 556–576.
[31] Leona Yi-Fan Su, Michael A Xenos, Kathleen M Rose, Christopher Wirz, Dietram A Scheufele, and Dominique Brossard. 2018. Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society* 20, 10 (2018), 3678–3699.
[32] J.J. Van Bavel, S. Rathje, E. Harris, C. Robertson, and A. Sternisko. 2021. How social media shapes polarization. *Trends in Cognitive Sciences* 25, 11 (2021), 913–916. doi:10.1016/j.tics.2021.07.013
[33] Xudong Yu, Magdalena Wojcieszak, and Andreu Casas. 2024. Partisanship on social media: In-party love among American politicians, greater engagement with out-party hate among ordinary users. *Political Behavior* 46, 2 (2024), 799–824.

# A Appendix

## A.1 Cluster Variability Analysis

Table 5 shows reply comment statistics per user. Coefficient of Variation (CV = SD/mean) indicates variability: $< 0.5$ (low), $0.5-1.0$ (moderate), $> 1.0$ (high), $> 2.0$ (extreme). All groups show $CV > 1.0$, justifying user-level rather than group-level analysis.

| Group | Users | Posts | User Mean | Std | CV | Max |
|---|---|---|---|---|---|---|
| C-Echo | 131 | 7010 | 53.5 | 52.8 | 0.99 | 481 |
| C-Balance | 83 | 4798 | 57.8 | 61.9 | 1.07 | 505 |
| C-Cross | 98 | 6947 | 70.9 | 53.7 | 0.76 | 311 |
| Control | 60 | 2994 | 49.9 | 49.1 | 0.98 | 264 |
| Pro | 145 | 8509 | 58.7 | 47.7 | 0.81 | 338 |

**Table 5: Comment Variability by Stance Group**

## A.2 Temporal Analysis Statistics

Table 6 shows statistical tests of whether C-Echo and C-Balance exhibit distinct temporal activity patterns.

| Analysis | Statistic | p-value |
|---|---|---|
| *Clustering Analysis* | | |
| Chi-square test | $\chi^2 = 29.59$ | $< 0.001$ |
| Clustering purity | 80.5% | — |
| Silhouette score | 0.119 | — |
| *Peak Timing Analysis* | | |
| KS test | $D = 0.199$ | $< 0.001$ |
| Target peak (median) | Period 7 | — |
| Other peak (median) | Period 6 | — |
| *Variance Analysis* | | |
| Mann-Whitney U | $U = 33,143$ | 0.913 |
| Target variance (median) | 50.96 | — |
| Other variance (median) | 51.26 | — |
| *Period-by-Period Analysis* | | |
| Significant periods | 10/20 | $< 0.05$ |

**Table 6: Statistical Analysis of Temporal Activity Patterns**

## A.3 Complete Stance Proportions

Table 7 reports complete yearly stance proportion across channels and users' groups.

| Group / Channel Type | Metric | 2020 | 2021 | 2022 | 2023 | 2024 |
|---|---|---|---|---|---|---|
| *User Groups* | | | | | | |
| C-Balance | N Comments | 270 | 1149 | 1239 | 1134 | 1000 |
| | Contra (%) | 50.5 | 60.4 | 52.3 | 58.0 | 61.0 |
| | Neutral (%) | 39.9 | 32.2 | 39.9 | 36.4 | 33.3 |
| | Pro (%) | 9.6 | 7.4 | 7.8 | 5.6 | 5.7 |
| C-Cross | N Comments | 437 | 1935 | 1910 | 1613 | 1052 |
| | Contra (%) | 51.2 | 52.9 | 50.1 | 56.1 | 56.7 |
| | Neutral (%) | 42.1 | 37.6 | 40.1 | 36.7 | 34.9 |
| | Pro (%) | 6.7 | 9.5 | 9.8 | 7.2 | 8.4 |
| C-Echo | N Comments | 131 | 1500 | 2661 | 1605 | 1105 |
| | Contra (%) | 42.9 | 54.7 | 53.9 | 56.0 | 62.8 |
| | Neutral (%) | 54.7 | 40.6 | 41.7 | 40.4 | 33.6 |
| | Pro (%) | 2.4 | 4.7 | 4.4 | 3.6 | 3.6 |
| Control | N Comments | 302 | 845 | 733 | 643 | 470 |
| | Contra (%) | 17.4 | 25.2 | 21.7 | 22.9 | 35.9 |
| | Neutral (%) | 74.7 | 64.7 | 62.5 | 64.8 | 49.9 |
| | Pro (%) | 7.8 | 10.2 | 15.8 | 12.3 | 14.2 |
| Pro | N Comments | 622 | 2220 | 2280 | 1817 | 1570 |
| | Contra (%) | 8.3 | 9.5 | 11.5 | 10.6 | 11.5 |
| | Neutral (%) | 70.5 | 60.5 | 60.2 | 59.6 | 52.8 |
| | Pro (%) | 21.3 | 30.0 | 28.3 | 29.8 | 35.7 |
| *Broad Audience* | | | | | | |
| *Left-Leaning Channels* | | | | | | |
| Content Creator | N Comments | 26933 | 60561 | 64404 | 90552 | 106468 |
| | Contra (%) | 15.0 | 13.9 | 15.1 | 17.0 | 23.1 |
| | Neutral (%) | 67.9 | 67.4 | 65.1 | 65.5 | 49.9 |
| | Pro (%) | 17.1 | 18.7 | 19.8 | 17.5 | 27.0 |
| News Channel | N Comments | 70115 | 196335 | 216017 | 106824 | 88310 |
| | Contra (%) | 22.2 | 26.6 | 22.8 | 39.9 | 38.7 |
| | Neutral (%) | 62.0 | 57.6 | 61.8 | 43.4 | 41.1 |
| | Pro (%) | 15.8 | 15.8 | 15.4 | 16.7 | 20.3 |
| *Right-Leaning Channels* | | | | | | |
| Content Creator | N Comments | 3287 | 40781 | 29037 | 55579 | 95107 |
| | Contra (%) | 21.2 | 25.1 | 20.7 | 26.1 | 41.5 |
| | Neutral (%) | 70.9 | 67.4 | 73.0 | 65.3 | 47.1 |
| | Pro (%) | 7.8 | 7.6 | 6.3 | 8.6 | 11.4 |
| News Channel | N Comments | 27276 | 92546 | 122796 | 93497 | 86926 |
| | Contra (%) | 38.7 | 42.2 | 43.3 | 47.6 | 51.8 |
| | Neutral (%) | 50.4 | 49.3 | 48.2 | 41.5 | 34.7 |
| | Pro (%) | 10.9 | 8.5 | 8.5 | 10.9 | 13.5 |

**Table 7: Yearly Stance Distribution by Groups and Channels**