



OPEN ACCESS

EDITED BY

Lara Lengel,
Bowling Green State University, United States

REVIEWED BY

Koldobika Meso Ayerdi,
University of the Basque Country, Spain
Valquiria Aparecida Passos Kneipp,
Federal University of Rio Grande do Norte,
Brazil

*CORRESPONDENCE

Jorge Cruz-Silva
✉ jacruz@puce.edu.ec
Marco López-Paredes
✉ mvlopez@puce.edu.ec

RECEIVED 15 October 2025

REVISED 07 January 2026

ACCEPTED 20 January 2026

PUBLISHED 30 January 2026

CITATION

Cruz-Silva J and López-Paredes M (2026) Tracing the discursive drift from news framing to discriminatory expressions in YouTube comments.

Front. Commun. 11:1726119.

doi: 10.3389/fcomm.2026.1726119

COPYRIGHT

© 2026 Cruz-Silva and López-Paredes. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Tracing the discursive drift from news framing to discriminatory expressions in YouTube comments

Jorge Cruz-Silva* and Marco López-Paredes*

Observatorio de Comunicación, Pontificia Universidad Católica del Ecuador, Quito, Ecuador

In the algorithm-mediated ecosystem, comment spaces have become arenas for discursive reinterpretation and confrontation. This study analyzes how journalistic discourse is reconfigured in these environments through the phenomenon of discursive drift, understood as the semantic and affective shift between the original framing of the news story and public responses. The case study is news videos about the Constituent Assembly in Ecuador, examined using a mixed-method approach combining text mining, lexical analysis, and semantic modeling. Over 1,600 comments from Ecuadorian media outlets were analyzed. Lexical dictionaries were constructed to detect incivility, the Hostile Environment Effect, and hate speech, and a TF-IDF model with cosine similarity was applied to measure the semantic distance between headlines and comments. The results show that 43% of the comments exhibit significant discursive drift, evidencing a systematic disconnect from the original journalistic frame. The coexistence of incivility and perceived media bias suggests that discursive drift is not a marginal phenomenon, but rather a structural condition of the digital public space, where affective polarization redefines the communicative function of journalism in contemporary democracies.

KEYWORDS

discursive drift, hate speech, digital communication, hostile media effect, news framing

1 Introduction

In the era of permanent connectivity, traditional media no longer monopolize the production of news discourse. Among them, YouTube occupies a privileged place as a hybrid space that combines the logic of entertainment, professional journalism, and spontaneous public participation. In this algorithmic environment, each piece of news becomes a starting point for collective reinterpretation, where meaning is unstable and negotiated.

Public debates on issues of national interest—such as the Constituent Assembly in Ecuador—move into the digital space, where user comments actively reinterpret journalistic frameworks. In this process, phenomena such as incivility, hate speech, and the Hostile Media Effect emerge, revealing an emotional dimension of online news consumption. Such manifestations not only express discontent or distrust toward the media but also contribute to the erosion of journalistic authority and the consolidation of polarized discursive communities, where interaction becomes confrontation and deliberation becomes a performance of identity.

This study aims to examine how journalistic coverage of the Constituent Assembly on YouTube gave rise to a network of comments intertwining criticism, distrust, and the redefinition of news discourse. The research hypothesizes that discursive drift constitutes a structural mechanism of collective reinterpretation in digital environments, through which

audiences not only comment on the news but also reproduce it in ideological and emotional terms.

2 Methods

The research was conducted using a mixed design, with the objective of examining the interaction between news framing, incivility, hate speech, and discursive drift in YouTube comments about the proposed Constituent Assembly in Ecuador.

The corpus for analysis consisted of a selection of journalistic videos published on YouTube that addressed the Constituent Assembly. These videos were produced by Ecuadorian media outlets with diverse political orientations, allowing for the observation of public reception in different ideological contexts. More than 1,600 comments were extracted from these videos, which formed the basis of the discursive analysis. Relevant metadata—such as the source channel, date, and video title—were identified for each comment to link the content of the responses to their news source and allow for comparisons between media outlets with different editorial lines.

Comments were cleaned and normalized by removing links, emojis, and typographic noise. Video titles—representing the initial information frame—served as a reference to measure the degree of semantic deviation in the comments.

The study considered four analytical dimensions: incivility, the Hostile Media Effect, discriminatory discourse, and discursive drift. Each dimension was operationalized using specific lexical and semantic criteria. For incivility, a dictionary of recurrent insults and derogatory terms in Ecuadorian and Latin American Spanish was constructed. The Hostile Media Effect was defined through the presence of terms that directly accused the media of bias, manipulation, or corruption. Discriminatory discourse was identified using a lexicon that covered three axes: political-ideological, ethnic-cultural, and moral-sexual.

Discursive drift was conceptualized as the semantic distance between the original news frame and the public's response in the comments. Since the video titles condense the news frame, they were taken as a reference point for the original discourse. From them, the semantic similarity between the title and the comment was calculated using a vector text representation model based on weighted term frequencies (TF-IDF) and cosine similarity. The result was a continuous measure of "discursive distance" between 0 and 1, where values close to zero indicated thematic alignment with the title and values close to one indicated complete deviation.

The subsequent statistical analysis integrated these dimensions to explore the relationships between media drift, aggression, and ideology. Percentages of incivility, Hostile Media Effect, and hate speech were calculated by channel's political orientation and level of discursive drift. Co-occurrence tables were also developed to show the extent to which high drift coincided with hostile or discriminatory expressions, and comparisons between orientations were made to identify discursive patterns specific to each political spectrum.

All comments analyzed corresponded to public content and were anonymized, avoiding the verbatim reproduction of offensive expressions in the results. The lexicon-based detection method was chosen for its transparency and reproducibility, while being aware of its limitations compared to deep learning models (Table 1).

TABLE 1 Lexical and analytical dimensions.

Dimension	Analytical indicator	Example keywords/expressions	Type of measure
Incivility	Direct insults or offensive terms	Idiota, corrupto, imbécil	Binary (0/1)
Hostile Media Effect	Accusations of bias or manipulation	Vendido, manipulado	Binary (0/1)
Hate Speech	Discriminatory political, ethnic, or moral insults	Correísta, indio, maricón	Binary (0/1)
Discursive Drift	Semantic distance from original frame	TF-IDF cosine distance	Continuous (0–1)

3 Literature review

3.1 News framing, cognitive and affective effects

Framing involves promoting particular problem definitions, causal interpretations, moral evaluations, and treatment recommendations. The distinction between episodic and thematic framing is critical. Episodic frames focus on individual events, often decontextualizing systemic issues, whereas thematic frames present broader social patterns and institutional causes. [Dursun and Tunç \(2024\)](#) applied this framework in analyzing Turkish media's coverage of mental health, finding that episodic frames led audiences to personalize responsibility, while thematic framing encouraged systemic thinking. These findings are relevant for sensitive topics, as framing shapes empathy and policy support. Moreover, [Park et al. \(2021\)](#) demonstrated that emotional news framing, especially through infotainment formats, increases audience attention and emotional arousal. In emotionally charged frames, audiences are more likely to react strongly in comment sections, especially on platforms like YouTube.

3.2 Digital platforms and discursive drift

The concept of discursive drift refers to how original media content, once published on digital platforms like YouTube, is subject to reinterpretation, distortion, or recontextualization by its audience. YouTube, in particular, is a hybrid platform that combines professional news content with user-generated responses and algorithmic curation, allowing for an accelerated and unpredictable shift in narrative tone. [Canevez et al. \(2022\)](#) showed that after the publication of news on racially sensitive topics such as the George Floyd case, YouTube users re-framed these narratives in the comment sections, sometimes introducing misinformation or overtly racist sentiments. This drift is not random; it is influenced by the original framing, audience ideology, and platform features. [Ottoni et al. \(2018\)](#) highlighted how far-right YouTube channels systematically exploit this environment, using emotionally charged videos and unmoderated comments to reinforce and spread hate. The architecture of the platform contributes to the viral nature of discursive drift, amplifying hate speech and discriminatory expressions.

3.3 Counter-framing and collective hostility

Counter-framing is a discursive resistance tactic where audiences actively push back against the media's original framing. Users reframe narratives to align with their ideological positions, sometimes completely inverting the intended message. [Liu and McLeod \(2019\)](#) demonstrated how user comments can serve as alternative framing devices that influence how other readers interpret the news. These counter-frames often emerge in polarized environments, turning comment sections into battlegrounds of narrative contestation. [Murthy and Sharma \(2019\)](#) provided a network analysis of YouTube comment threads, revealing that hostility is not isolated but collective and interconnected.

3.4 Hostile media effect

The Hostile Media Effect (HME) reflects the cognitive bias in which partisans perceive neutral media content as biased against their beliefs. This psychological tendency is especially pronounced in polarized political climates and can result in aggressive backlash even when journalistic content strives for balance. [Liu \(2023\)](#) applied this concept to COVID-19 coverage and anti-Asian sentiment. The study found that emotionally negative frames triggered defensive responses among viewers, especially those predisposed to racial biases. Viewers experiencing HME were more likely to post accusatory or hateful comments.

3.5 Digital incivility and user latency

Digital incivility encompasses behaviors such as name-calling, sarcasm, insults, and dismissive language in online discourse. While traditionally seen as harmful, some scholars argue that incivility can increase audience engagement. However, the cost is a reduction in institutional trust and the quality of deliberative discourse. [Borah \(2013\)](#) found that uncivil comments in political blogs boosted reader interest but led to cynicism toward government and media. On YouTube, where moderation is limited, incivility is often rampant and becomes a norm rather than an exception. User latency—the continued engagement of individuals over time—further deepens this problem. Returning commenters often take on the role of influencers in threads, modeling aggressive behavior that others emulate. This persistence contributes to the entrenchment of hate speech within comment sections, making it a systemic issue rather than sporadic misconduct.

4 Results

4.1 Incivility and HME

The initial results reveal a pattern of moderate but significant incivility in the corpus, with 8.45% of comments featuring insults, derogatory comments, or aggressive language. This percentage is consistent with previous studies on interaction on open platforms such as YouTube, where the lack of moderation and the perception of anonymity reduce social barriers to hostile behavior. Incivility in this

context is not simply discursive noise: it acts as a marker of political emotionality and disenchantment with public figures, especially in times of polarization. The existence of this aggressive tone constitutes the fertile ground for the discursive drift described by Canevez et al. and Ottoni et al., in which users reconfigure news frames in a confrontational manner ([Table 2](#)).

The incidence of the Hostile Media Effect (HME), although lower (3.05%), is conceptually more relevant: it shows that part of the audience perceives the media as ideological adversaries, even when journalistic content seeks neutrality. The terms detected—“sellouts,” “paid for,” “liars”—confirm the existence of this biased interpretation. This phenomenon fits with [Liu's \(2023\)](#) proposal: in highly polarized environments, the interpretation of the message is filtered through the recipient's political predispositions, which triggers defensive or accusatory responses. In the Ecuadorian case, criticism of the media outlet can function as a discursive ritual of belonging, where accusing the channel is equivalent to reaffirming a political identity in front of other commentators.

The analysis by political orientation of the channels shows a significant difference: spaces associated with the left and center concentrate almost twice as much incivility and HME as those on the right. This suggests that the channels not only attract ideologically aligned audiences but also opposing ones who use the comments section as an arena for confrontation. In terms of framing, centrist and left-wing media outlets seem to activate a greater degree of emotionality or counter-framing, where users rewrite the narrative in a tone of distrust, delegitimization, or mockery. It is possible that the diversity of audiences in these spaces, along with the more institutional or reflexive nature of their messages, generates greater cognitive friction and, therefore, greater hostility in reception.

4.2 Discriminatory, exclusionary, or hateful language

The results of this stage show that nearly 8% of the comments contain discriminatory expressions, a high percentage for a corpus focused on news content. This finding confirms that the conversation about the Constituent Assembly is not limited to political disagreement but rather becomes a space for the reproduction of hierarchies and social exclusions. The expressions detected—ranging from ideological insults (“correísta,” “leftist”) to ethnic or moral offenses (“indio,” “faggot,” “scum”)—evidence a mix of political polarization and structural prejudice, where hate speech is naturalized as part of political criticism. This pattern coincides with what [Murthy and Sharma \(2019\)](#) called collective hostility: a type of aggression that is not directed solely at individuals but is collectively reinforced against symbolic groups.

TABLE 2 Summary statistics by political orientation.

Political orientation	Incivility (%)	HME (%)	Hate speech (%)	Avg. drift
Right	5.2	1.4	7.3	0.87
Center	9.8	3.1	8.4	0.93
Left	9.3	3.0	7.1	0.90

In theoretical terms, these results are part of the dynamics of discursive drift described by Canevez et al. (2022): journalistic frames of the political situation are recontextualized by the audience until they lose their informative character and transform into emotional narratives of opposition and contempt. Discriminatory comments do not emerge in a vacuum; they respond to the initial frame of interpretation of the news and the ideology of the digital community commenting on it.

The comparative analysis by political orientation reveals that centrist channels concentrate the highest proportion of hate speech (8.43%), followed by right-wing (7.28%) and left-wing (7.07%). Although the percentage differences are small, the trend is significant: spaces perceived as neutral or pluralistic attract the greatest cross-confrontation. This phenomenon can be explained by the Hostile Media Effect (Liu, 2023): when polarized audiences are confronted with a media outlet that does not confirm their biases, they interpret its neutrality as hostility, reacting with verbal or ideological aggression. In this sense, the presence of hate speech acts as a thermometer for the fragility of public discourse in environments where neutrality is no longer interpreted as objectivity, but as betrayal.

The concentration of discriminatory comments on international and digital channels—such as France 24 Español and DNews—suggests that hatred is not only directed toward local political actors, but also toward external or institutional sources perceived as outside the national debate. This xenophobic or anti-global component reinforces the idea that hate on social media is not a spontaneous outpouring, but rather a discursive mechanism for identity reaffirmation.

4.3 Discursive drift

The discursive drift analysis reveals that 43% of comments deviate significantly from the original framing proposed in the video title. This finding suggests that, in most cases, users do not engage in dialogue with journalistic content, but rather discursively reconfigure it to fit their own ideological logic. The observed semantic distance confirms the idea that YouTube not only redistributes media messages but also reconstructs their meanings through the interaction of heterogeneous audiences. This discursive drift constitutes a form of collective appropriation of public discourse, where the news ceases to be a point of arrival and becomes a narrative trigger for political, emotional, and moralizing expressions.

The coincidence between high drift and discursive aggression reinforces the interpretation that semantic drift is not neutral. Comments with greater distance from the headline also present higher levels of incivility, Hostile Media Effect, and discriminatory discourse, suggesting that drift operates as a space of opposition and resistance to institutional discourse. In other words, the further the comment is from the news frame, the more likely it is to adopt a confrontational or delegitimizing tone. This empirical pattern supports the concept of counter-framing (Liu and McLeod, 2019), according to which users reinterpret or invert the meaning of messages to reaffirm their ideological position vis-à-vis the media.

4.4 Visualization and correlations

The visual results confirm that discursive drift does not operate in isolation but is embedded in an ecology of conflict where incivility,

mistrust, and hatred interact at varying intensities. Comparing political orientations, it is observed that right-wing channels present the highest percentage of High Drift (51.5%), followed by the left (43.4%) and the center (35.1%). However, centrist and left-wing spaces concentrate more incivility and hate speech, suggesting that media perceived as neutral or pluralistic attract more cross-confrontation. This difference reveals that drift can have two modalities: a semantic one, associated with the ideological reinterpretation of the message, and an interactional one, linked to the aggressive tone of the response (Table 3).

The analysis of drift levels shows a pattern of progressive tension: comments with low semantic shift remain relatively respectful, while those in the middle range (0.8–0.99) register the greatest increase in incivility and hostility. It is at this intermediate point where the audience still engages with the original message but begins to resist or contradict its framing, transforming the dialogue into a dispute. In contrast, high-drift comments (≥ 0.99) tend to move so far away from the initial frame that they abandon direct confrontation and drift toward parallel or moralizing narratives. This shift marks the moment when the conversation ceases to be an extension of the media discourse and becomes a space for ideological self-affirmation.

The correlations between the variables reinforce this interpretation. The positive relationship between incivility and the Hostile Media Effect ($r = 0.25$) indicates that the perception of media bias drives the use of aggressive language, consolidating a cycle of distrust and confrontation. For its part, the low correlation between discursive drift and discursive aggression suggests that not all drift implies hostility: some shifts respond more to the recontextualization of the topic than to verbal violence. This nuance is key to drift theory: semantic distance can express resistance, irony, or discursive creativity, and not always violence.

5 Discussion

The results obtained in the five stages of this study clearly demonstrate how the contemporary digital ecosystem has transformed the relationship between media, audiences, and public truth. In the Ecuadorian case, coverage of the Constituent Assembly on YouTube not only circulated as news content, but also became territories of symbolic dispute, where journalistic meanings were redefined, overwhelmed, and, in many cases, inverted by audiences. The research demonstrates that platforms are not mere dissemination channels: they are spaces of multiple mediation, where information clashes with emotions, political identities, and algorithmic architectures that foster polarization.

From a theoretical perspective, the findings consolidate the concept of discursive drift as a first-rate analytical tool for studying the reconfiguration of discourse on digital platforms. The fact that more than 40% of the analyzed comments exhibit high drift reveals a

TABLE 3 Correlation matrix.

Variable	Incivility	HME	Hate speech	Discursive drift
Incivility	1.00	0.25	0.32	0.12
HME	0.25	1.00	0.29	0.08
Hate Speech	0.32	0.29	1.00	0.10
Discursive Drift	0.12	0.08	0.10	1.00

structural rupture between the original journalistic frame and its public reception. Discursive drift is a constitutive logic of digital communication, where meanings shift toward parallel or conflicting narratives. Unlike the traditional interpretation of the “agenda effect,” where the media define the topics of debate, drift shows that audiences reformulate the frames and hierarchies of meaning.

On the affective level, empirical data confirm that incivility and the Hostile Media Effect are recurring manifestations of this drift. The positive correlation between both variables ($r = 0.25$) suggests that the perception of media bias fuels verbal aggression, creating a cycle of antagonism in which the media are perceived not as sources of information, but as ideological adversaries. In this context, incivility cannot be understood solely as a lack of courtesy, but as a form of discursive resistance to a media system that users consider biased or manipulated. Hate speech, meanwhile, emerges as a more advanced phase of this communicative degradation, where criticism turns into dehumanization. However, the data show that hate speech does not dominate the space; rather, it appears as an extreme response within a broader dynamic of polarization.

Furthermore, the comparison between political orientations reveals a particularly relevant finding: centrist and left-wing channels concentrate more discursive aggression, while right-wing channels exhibit a more pronounced semantic drift. This contrast reveals two types of communicative ecosystems: some focused on direct confrontation (debate and hostility) and others on the symbolic reconfiguration of the message, where drift replaces confrontation. In sociopolitical terms, this difference can be interpreted as the digital manifestation of a structural ideological asymmetry: while the right tends to consolidate more homogeneous discursive bubbles, the center and left constitute spaces more permeable to dissonance and conflict.

The analysis of the qualitative examples—comments with high drift and aggressive language—illustrates how news discourse becomes raw material for delegitimization and irony. Institutional or neutral headlines, which attempt to maintain a professional frame, are read as “propaganda” or “lies,” reflecting a structural loss of trust in journalism. This finding coincides with what Liu (2023) calls “emotional media hostility,” where the public does not judge the content, but rather its perceived intentionality.

6 Conclusion

Taken together, results confirm that discursive drift constitutes a new analytical paradigm for digital communication. It is not merely a semantic phenomenon, but a social structure where meaning is redefined through interaction and conflict. On YouTube, each comment rewrites the news story from an affective or ideological perspective, transforming communication into a networked and emotionally charged process.

Methodologically, the use of indicators such as incivility, Hostile Media Effect, hate speech, and semantic distance allows us to quantify dimensions that were traditionally considered qualitative. This hybrid approach not only enriches communication analysis but also opens the possibility of building algorithmic observatories of public discourse, capable of monitoring in real time how news messages transform into arenas of symbolic dispute. In polarized contexts like Ecuador's, this perspective offers a crucial tool for studying the deterioration of public debate and possible strategies for rebuilding media trust.

Recognizing this transformation implies accepting that the role of journalism and communication institutions must shift from the production of certainties to the management of dissent and the critical literacy of audiences. In short, this study shows that digital communication not only reproduces existing social conflicts but also reorganizes and amplifies them through its own logic of participation. Understanding discursive drift, therefore, means understanding how media democracy is being reconfigured today, in an environment where each user has the capacity—and the responsibility—to produce meaning.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

Ethical approval was not required for the study involving human data in accordance with the local legislation and institutional requirements. Written informed consent was not required, for either participation in the study or for the publication of potentially/indirectly identifying information, in accordance with the local legislation and institutional requirements. The social media data was accessed and analyzed in accordance with the platform's terms of use and all relevant institutional/national regulations.

Author contributions

JC-S: Conceptualization, Data curation, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. ML-P: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work corresponds to the research project: Cultural, Institutional and Social Crises of Democracy in Latin America, QIPR0049-IBYA103261070, coordinated by Observatory of Communication at Pontifical Catholic University of Ecuador.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. ChatGPT 5 for simple translation of terms and for

the data set curation, and Jupiter Notebook 7.2.2 for the Web Scrapping phase.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Borah, P. (2013). Interactions of news frames and incivility in the political blogosphere: examining perceptual outcomes. *Polit. Commun.* 30, 456–473. doi: 10.1080/10584609.2012.737426

Canevez, R. N., Karabelnik, M., and Winter, J. S. (2022). Police brutality and racial justice narratives through multi-narrative framing: reporting and commenting on the George Floyd murder on YouTube. *Journalism Mass Commun. Q.* 99, 696–717. doi: 10.1177/10776990221108722

Dursun, A. D., and Tunç, A. (2024). "Mental health news coverage in Turkish national newspapers with respect to episodic and thematic framing: a retrospective content analysis" in Computational science and its applications – ICCSA 2024 workshops (lecture notes in computer science, 14815, 59–71). eds. O. Gervasi, B. Murgante, C. Garau, D. Taniar, A. M. A. C. Rocha and M. N. F. Lago (Switzerland: Springer).

Liu, J. (2023). Blaming Asians for coronavirus: the role of valenced framing and discrete emotions in hostile media effect. *Asian J. Commun.* 33, 68–86. doi: 10.1080/01292986.2023.2169722

Liu, J., and McLeod, D. M. 2019 Counter-framing effects of user comments *Int. J. Commun.* 13 2484–2503. Available online at: <https://ijoc.org/index.php/ijoc/article/view/11392>

Murthy, D., and Sharma, S. (2019). Visualizing YouTube's comment space: online hostility as a networked phenomena. *New Media Soc.* 21, 191–213. doi: 10.1177/1461444818792393

Ottoni, R., Bernardina, P., Cunha, E., Meira, W., Magno, G., and Almeida, V. (2018). "Analyzing right-wing YouTube channels: hate, violence and discrimination" in Proceedings of the 10th ACM conference on web science (323–332) (Association for Computing Machinery).

Park, S., Bier, L. M., and Park, H. W. (2021). The effects of infotainment on public reaction to North Korea using hybrid text mining: content analysis, machine learning-based sentiment analysis, and co-word analysis. *Prof. Inferm.* 30:e300306. doi: 10.3145/epi.2021.may.06