Indian Legacy and Digital Media on YouTube: A Comparative Analysis of News Narratives and User Engagement

Piyush Ghasiya¹ and Kazutoshi Sasahara School of Environment and Society, Tokyo Institute of Technology, Tokyo, Japan

Abstract

The current news media ecosystem in India can be divided into two groups: legacy media and digital media. The former dominate TV space, while the latter operate primarily on the web and social media. Recent literature has posited potential differences in journalistic standards and content objectivity between these two categories, but these assertions lack robust empirical validation. To address this research gap, we analyzed the characteristics of news content and user engagement on YouTube. We collected 1074 news videos uploaded by the two types of media on YouTube, along with more than 400,000 comments. Using network and text analysis, we found that users form two highly polarized groups with minimal interaction between them, suggesting the presence of echo chambers. Furthermore, toxicity analysis shows that digital media hosted more hateful and toxic comments than legacy media. Our findings suggest complex dynamics in the Indian media ecosystem, potentially influenced by factors such as ownership structures, content strategies, and audience segmentation. The results indicate potential correlations between corporate ownership models and alignment of content with governmental narratives in legacy media, while digital media demonstrates a trend toward content diversification.

Keywords: Echo Chamber; India; News Narratives; News Media; User Engagement; YouTube.

Introduction

Significance of News Media for Democracy

News media play a critical role in informing citizens about public affairs, government functioning, and politics. The information provided by news media helps to inform decision-making. By serving as a public representative and as a watchdog of government performance, the media operate as critical mechanisms for democratic accountability (Goidel, Gaddie, &

¹ CONTACT Piyush Ghasiya. Email: ghasiya.p.aa@m.titech.ac.jp

Ehrl, 2017). This is why news media is called the 'fourth pillar' of democracy (the other three pillars are: the executive, the legislative, and the judiciary). This pillar can positively impact democracy by sticking solely to objective journalism, the fundamental tenets of which include detachment, nonpartisanship, facticity, the inverted pyramid writing style, and balance (Benkler, Faris, & Roberts, 2018).

In contrast, if news media do not follow the principles of objective journalism—e.g., by taking sides, by presenting only one perspective, by being a 'lapdog' rather than a watchdog—democracy can be harmed. State-controlled media in countries such as China, North Korea, and (to a certain extent) Russia are a few examples of lapdog media; in these countries, media mainly spread government propaganda, and independent media in these countries struggle to exist due to various kind of censorship, imprisonment, and even threats of death. Hence, free press is one of the main criteria on which the health of a democracy is judged. Nevertheless, this does not mean that media in democracies always follow the tenets of objective journalism. In 1988, in their seminal work *Manufacturing Consent: The Political Economy of the Mass Media*, Edward S. Herman and Noam Chomsky proposed the propaganda model and showed how US media carried out system-supportive propaganda (Herman, & Chomsky, 2008).

Propaganda often uses misinformation. In the era of social media, creating misinformation has become extremely easy, while at the same time, the broadcasting of misinformation by news channels is becoming a significant problem. The virality of content does not depend on its truthfulness, and sometimes news channels—in their eagerness to capitalize on virality—pick up viral content from social media and present it as a 'prime time exclusive'; since such exclusives are time-sensitive, the due process of verification and fact-checking can be overlooked by news channels. There have been several cases in India where viral fake news or disinformation from social media found its way to prime-time slots on news channels. Additionally, an agenda-driven mentality of some media outlets further contributes to the spread of misinformation.² Previous studies have shown that such false content is partially responsible for the lack of trust in news media (Ognyanova et al., 2020).

Current News Media Ecosystem in India

Understanding the Indian media ecosystem is essential, due to its status as the world's largest democracy and to the pivotal role its media plays in shaping public opinion and reflecting

⁻

² https://thewire.in/media/2017s-top-fake-news-stories-circulated-by-the-indian-media

societal issues. Unlike other democracies (such as the US, the UK, Brazil, or Japan), India's media ecosystem is characterized by linguistic diversity, regional complexities, and rapidly-evolving digital adoption. In particular, the interplay of political and corporate influences on news narratives and the widespread adoption of digital media, even in rural areas, make India a unique case study for media, politics, and public engagement, potentially offering insights for other democracies.

According to the Government of India data, there are 388 news and current affairs channels in India.³ Regional language channels and private (corporate) players dominate the news channel market in India. However, high internet penetration and high-speed internet (4G and 5G) are gradually changing Indians' news consumption behavior: in past, cable TV was the main source of news consumption, but now, smartphones is becoming the main source of news access.⁴ However, when it comes to the content and the journalistic practices of these hundreds of legacy media news channels, the situation is very dire, leading some insiders (i.e., journalists and editors) to call this "the dark age of journalism" in India.⁵

Currently in India, news media can be divided into two main groups: legacy media and digital media. Two major differences set these groups apart. The first major difference is ownership. Legacy media channels are owned by large media houses; for example, Zee News is owned by Subhash Chandra's Zee group, Republic TV is owned by Arnab Goswami's SARG Media Holding, CNN-News18 is owned by Mukesh Ambani's Reliance group, NDTV⁶ is owned by Gautam Adani's Adani Group, and Aaj Tak is owned by Aroon Purie's India Today group. In contrast, digital media such as The Wire, The Print, Newslaundry, The Quint, and NewsClick are funded by a combination of non-profit organizations, individuals, and users (through subscription models).⁷ The second major difference is operational medium: legacy media operate largely on cable TV, while the internet (websites and YouTube) is the operational medium of digital media.⁸

³

 $[\]frac{https://sansad.in/getFile/loksabhaquestions/annex/175/AU857.pdf?source=pqals\#:\sim:text=There\%20are\%20388\%20News\%20and,\%2Fwebpage\%2DUser\%2Dtvchannels.\&text=(c)\%20Since\%201.1.$

⁴ https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021/india

⁵ https://niemanreports.org/articles/india-ndtv-modi/

⁶ In December 2022, the founders of NDTV, Prannoy and Radhika Roy, sold their share to Adani group, making the latter the majority shareholder (64.71%) of NDTV news network.

⁷ For detailed information about ownership in media in India, please refer to The Media Ownership Monitor website: http://india.mom-gmr.org/en/

⁸ Legacy news media also utilize the digital medium by uploading their content on their websites, YouTube, and other social media platforms, such as Facebook. This gives them the advantage of being available both on TV and on the internet.

There is another difference between these two groups, and that is their sincerity in following the principles of objective journalism. News media can be led to set aside these principles and behave like a lapdog for several reasons, but ownership and advertising licenses play a major role (Herman, & Chomsky, 2008). Similarly, analysis by Champlin and Knoedler showed that news media have two responsibilities: as an institution to protect the public interest, and as a capitalist enterprise. Meeting these dual responsibilities has always been challenging; however, the rise of large media conglomerates and the resulting concentration of ownership in the media industry have greatly tipped the scales toward the latter and away from the former (Champlin, & Knoedler, 2002). In this context, legacy media in India appears to prioritize profit margins, which may affect their adherence to traditional journalistic principles and thus result in alignment with the current government's narratives. Onversely, digital media, with its different economic models, tends to focus on public interest journalism and critically engaging with those in power.

In addition to the above issues related to journalistic practices, another critical problem has arisen in the era of social media: the issue of echo chambers. An *echo chamber* can be defined as an environment in which participants encounter beliefs that strengthen their preexisting beliefs, with communication and reiteration occurring inside a closed system and shielded from refutation (Cinelli et al., 2021; Nguyen, 2020; Sasahara et al., 2021) As the patterns of news consumption change (especially for young people, who like to receive their news via social media), and given the polarizing socio-political environment in India, the probability of the existence of echo chambers among users of legacy media and digital media is very high. Politically biased news in itself is bad for a democracy, but getting that news via social media with auto-recommendation systems can further polarize society and consumers.

In summary, in the rapidly evolving media ecosystem of India, the rise of digital media has impacted how information is disseminated and consumed, and the shift from legacy media to digital media, together with legacy media's profit-seeking and politically biased (un-)journalistic practices, raise crucial questions about the nature of news coverage, audience engagement, and the characteristics of media users' biases. Understanding these dynamics is essential to comprehending the broader implications of media evolution for societal polarization, diversity of viewpoints, and the role of social media in shaping public discourse.

-

4

⁹ Apart from profit margin, other potential reasons behind legacy media's current positioning in India include owner's relation/affiliation with a specific ideology (Hindu Nationalist/Hindutva), fear of state clampdown, and losing state contracts in other areas, among others.

¹⁰ Ibid.

To address these issues, this study poses the following research questions (RQs):

RQ1: How do legacy media and digital media differ in their news coverage in India?

RQ2: How do users engage with news content from legacy media and digital media?

RQ3: What are the structural properties of social interactions between users on legacy media and on digital media?

RQ1 aims to uncover the thematic and stylistic differences between legacy and digital media in their news reporting, highlighting how each medium prioritizes and presents information. RQ2 examines user engagement patterns and preferences on social media, revealing the influence of the different media formats on public participation and discourse. RQ3 investigates the composition and characteristics of media audiences, exploring whether the shift to digital media platforms contributes to echo chambers or promotes a more diverse information ecosystem.

The organization of the rest of this paper is as follows. Section 2 reviews previous research on the themes of decreasing standards in Indian media, the impact of social media, and the rise of independent journalism in India. Section 3 explains the methods of data collection and the computational analyses used. Sections 4 and 5 offers the results and the discussions, respectively, of the key findings, and Section 6 summarizes the significance of this study.

Literature Review

In this section, we review relevant research to highlight the research gaps that our research seeks to fill.

Indian media has historically been subject to immense external pressures, from corporate owners and the political powers that be (Maheshwari & Sparks, 2021; Chadha, 2017); however, the authoritarian and right-wing nature of the current government has changed the media ecosystem in India. Support for Prime Minister Modi and the BJP comes from innumerable mainstream media houses, television channels, and newspapers establishments—in contrast to earlier times, when the RSS mouthpiece *The Organiser* was their sole supporter (Philipose, 2021). Moreover, in terms of content and characterization, Indian media has become

aggressive, abusive, and judiciary¹¹ (Aryal & Bharti, 2022); further commentary shows how legacy-media TV debate shows in India have become increasing uncivil and impolite (Kumar, & Dutta, 2023). One study analyzed legacy media's two biggest news channels (Republic TV and Times Now), specifically their primetime debates on a specific event (the India–China border conflict), and it found that the narratives of both channels were congruent with the extreme right-wing Hindutva ideology of the government and that both channels had adopted a hyper-nationalist style of presentation (Kumar, 2023).

All the above-cited studies are qualitative and analytical, and they focus only on legacy media. In contrast, research on digital media in India is scant, and it primarily examines digital media literacy and consumption. The 2019 India Digital News Report¹² revealed that English-language internet users in India are embracing mobile-first, platform-dominated media for their news consumption; the report also found that users have serious concerns about the possible implications of expressing their political views and are worried about disinformation (Aneez et al., 2019). In addition, a large study (N≈50,000) examining online news consumption in India found no evidence that digital-born media have succeeded in challenging the dominance of legacy media, as the median audience reach of digital-born outlets did not grow significantly over the four-year period of analysis (2014–18) (Mukerjee, 2024).

While numerous studies have analyzed YouTube data globally, relatively few have investigated it specifically in the Indian context. One study revealed that approximately 69% of the news videos it collected (N=14,506) contained toxic comments, and it found that religion and violence/crime-related news have the highest rates of toxic comments while economy-related news has the lowest rate of toxic comments (Alshamrani et al., 2020). Similarly, another study found that topics such as racism, war and conflict, and Israel–Palestine received more toxicity than science and technology, environment and weather, and art and culture (Salminen, 2020).

This review shows a lack of comparative studies of legacy and digital media in India. Our research aims to reduce this gap by analyzing news content and user engagement on YouTube.

Research Methods

_

¹¹ TV news channels starts a media trial and does not wait for legal procedure to declare innocence or culpability.

¹² This report is supported by The Hindu, The Indian Express, PTI (Press Trust of India), and The Quint. We would like to mention here that The Quint is one of the digital media platforms we studied in this research.

We collected data from two sources. First, we accessed Facebook to compile a list of popular YouTube channels for Indian legacy media and Indian digital media, identified as such based on the number of Facebook followers. We chose Facebook for this task because it is the most widely used social media platform in India. Then, we collected all the Facebook posts (total interactions $^{14} \ge 10$) containing YouTube links from those news media channels. Second, we collected all the available data from YouTube, including user comments and engagement statistics, in order to analyze news narratives and user engagement in Indian legacy and digital media. The following sections explain the method in more detail.

YouTube Channel Lists from Facebook

Using CrowdTangle¹⁵, we created two lists: one of news channels for legacy media, and one of news sites for digital media. Each list included five of the most popular news channels/sites. The legacy media list included the Facebook pages of India Today, NDTV, Republic, TIMES NOW, and Zee News English; the digital media list included the Facebook pages of NewsClick.in, Newslaundry, The Quint, The Print, and TheWire.in. Figures 1 and 2 show the legacy and digital media lists, along with the number of followers on Facebook on March 8, 2023 (date of data collection). From these pages, we collected all the posts with YouTube links posted between June 1, 2019 (starting date of second Modi government) and March 8, 2023. Consequently, we collected 3432 posts, of which 1840 were from legacy media and 1592 were from digital media. Subsequently, we removed all the posts that had less than 10 total interactions, such as likes, shares, and comments (n=1584). This left us with 1848 posts (868 posts for legacy media and 980 posts for digital media). In the next step, we removed all the posts with duplicated YouTube links (n=700), leaving us with 1148 posts (n=571 for legacy media and n=577 for digital media). Lastly, we removed all the entertainment-related posts (n=74), leaving us with 1074 posts in total, 516 posts for legacy media and 558 posts for digital media.

Data Collection from YouTube

¹

¹³ There are 378 million users in India, highest in the World and almost double than the US, the second ranked in this list with 193 million users.

¹⁴ Total interaction is the combined number of likes, comments, shares, and reactions ('love', 'wow', 'funny',

^{&#}x27;sad', 'angry', and 'care') that a post received at the time of data collection.

¹⁵ https://help.crowdtangle.com/en/articles/4201940-about-us

We collected data from the YouTube links in the collected Facebook posts (n=1074) using YouTube Data API v3.¹⁶ This API can provide not only the engagement statistics, such as the number of views, likes, and comments, but also all the comments posted on a video.

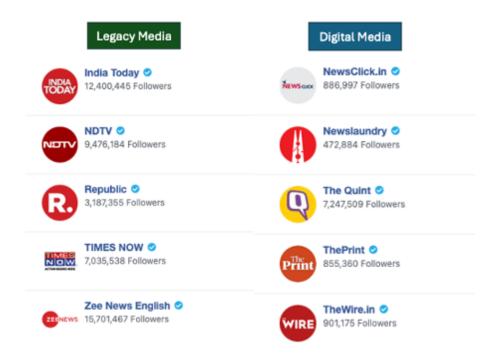


Figure 1: List of legacy and digital media, along with their number of Facebook followers on March 8, 2023

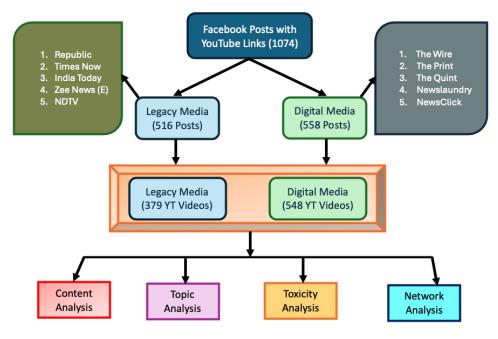


Figure 2: Research methods

8

¹⁶ <u>https://developers.google.com/youtube/v3/getting-started</u>

The collected YouTube data include the engagement statistics and the comments of 927 videos (378 videos from legacy media, 548 videos from digital media). This sample is 148 videos fewer than the number of links we collected from Facebook; this is due to the fact that 138 legacy-media videos and 10 digital-media videos were unavailable, removed either by YouTube for violation of community standards or by the channel/site itself for unknown reasons.

Table 1 shows the YouTube subscriber bases of legacy media channels and of digital media sites. Digital media has a relatively larger user base on YouTube than on Facebook, which is understandable given that YouTube is the main platform for digital media. Among the legacy media, Zee News English has a surprisingly low user count (n=16,500).

Table 1: YouTube subscriber counts¹⁷ of legacy and digital media

Legacy Media	YouTube	Subscribers	Digital Media	Subscribers
	Channel			
Danublia	Republic	5.52 million	The Wire	4.37 million
Republic	World			
	TIMES NOW,	3.02 million,	Newslaundry	1.6 million
Times Now ¹⁸	Zoom,	6.39 million,		
	ET Now	573 K		
India Today	India Today	7.03 million	The Print	2.16 million
Zee News	Zee News	16.5 k	The Quint	3.58 million
English	English			
NDTV	NDTV	12.4 million	NewsClick	3.82 million

Table 2: Individual share of each channel/site in collected YouTube data

Legacy Media	Posts ¹⁹	Digital Media	Posts ²⁰
Republic	341	The Wire	329
Times Now	14	Newslaundry	142
India Today	19	The Print	48
Zee News English	2	The Quint	16
NDTV	1	NewsClick	12

9

¹⁷ The subscriber count is as of March 8, 2023.

¹⁸ Zoom, ET Now, and Times Now are owned by Times Groups. Since some of the videos in our dataset are also from Zoom and ET Now, we have combined subscribers from these channels.

¹⁹ One video from UP government and 1 from R Bharat (Hindi news channel of Republic TV).

²⁰ One video from World News Day.

Table 3: Engagement with collected posts on YouTube videos

Legacy Media (378 videos) ²¹		Digital Media (548 videos)	
View Count	31,631,239 (Avg. 83,680)	View Count	95,006,279 (Avg. 173,369)
Like Count	79,489 (Avg. 210)	Like Count	3,459,703 (Avg. 6,313)
Comments Count	49,124 (Avg. 130)	Comments Count	545,126 (Avg. 995)

Table 2 shows the individual share of each channel/site in our YouTube dataset, and Table 3 presents the engagement statistics for all the videos. Although legacy media have larger subscriber bases than digital media on YouTube, digital media is way ahead of legacy media in terms of engagement: for example, digital media received an average of 995 comments on 548 videos, compared to legacy media's average of 130 comments on 379 videos.

For the final step, we collected 34,184 comments from 378 legacy media videos (55 videos have no comments) and 375,880 comments from 548 digital media videos (6 videos have no comments). For subsequent analyses, we manually labeled all the YouTube videos (of both legacy and digital media) according to two classifications: the format, such as whether the video is an interview or speech or debate; and the theme, such as whether the video is about foreign affairs or elections or the economy. We also looked at the most commented video and the video that received the highest overall engagement on both Facebook and YouTube.

Data Analysis

We employed a multi-faceted analytical approach (Figure 2). Content and topic analysis identifies what is being reported, while toxicity analysis delves into user engagement and discourse quality, and network analysis examines the user community structure. Together, these methods can offer a detailed and multi-dimensional understanding of the differences and dynamics between legacy media and digital media.

1. Content Analysis and Topic Analysis: The first step consists of comparing thematic differences between legacy media and digital media in news coverage on YouTube, which enables us to address the characteristics of their news narratives (related to RQ1). We used BERTopic, a Python library (Grootendorst, 2022), to perform topic modeling on the headlines (i.e., YouTube video titles). BERTopic leverages transformers and c-TF-IDF to

²¹ In legacy media, there was one advertisement by Mankind Pharmaceutical Company (COVID-related) that received over a million views. We removed that from our engagement statistics.

create dense clusters, allowing for easy interpretable topics while keeping important words in the topic descriptions.

- 2. Toxicity Analysis: The second step involves quantifying the level of toxicity in user comments. This is used as a proxy for the discourse quality and the potential for polarization among audiences (related to RQ2). To this end, we used Google's Perspective API²² via Communalytic²³. Perspective API measures toxicity, severe toxicity, profanity, identity attacks, insults, and threats in a given text. Although the interface has several acknowledged limitations (Hosseini et al., 2017), Perspective API is one of the most reliable tools to measure toxicity on social media discourse. The Man–Whitney U test is applied to the toxicity results to assess toxicity distributions among legacy and digital media users.
- 3. Network Analysis: The third step consists of mapping social interactions between users and news sources on YouTube. The resulting networks are used to evaluate the presence of echo chambers, or to determine whether the networks are homogenous and polarized or diverse and decentralized (related to RQ3). We used NetworkX Python library to construct the networks, and we visualized these networks using Gephi (Bastian et al., 2009).

Results

Content Analysis

Content analysis revealed distinctive patterns in program category (i.e., 'live broadcast', 'debate', 'discussion', 'ground report', etc.) in legacy media that are absent from digital media. In legacy media, 83 of 378 videos were very long live broadcasts that these channels directly aired on YouTube; the average duration of these 83 videos is over 6 hours, with the longest video clocking in at 11 hours 55 minutes. We labeled these videos as 'miscellaneous live', because they include several shows on different themes hosted by one channel (similar to watching news on cable TV, but on YouTube or Facebook). These are not the only long videos: specials such as Prime Minister (PM) Modi's "Mann ki Baat" or "Economic Summit" were also broadcasted live on YouTube by legacy media news channels²⁴. In contrast, the categories 'interview' (210) and 'ground report' (207) dominated digital media videos. These patterns point to the entirely different outlooks of legacy and digital media toward Facebook and YouTube. Social media is secondary to legacy news channels (primary is cable TV); they are

²² https://www.perspectiveapi.com/

²³ https://communalytic.org/

²⁴ In legacy news media, 51 videos were unavailable, but the number of views and comments (if any) were available. This is why we did not remove these unavailable videos from our dataset.

only interested in being present on social media, rather than capturing this space. For digital media, by contrast, social media is the one and only medium to reach their audience. Figure 3 shows the share of videos per category in legacy media (left) and digital media (right).

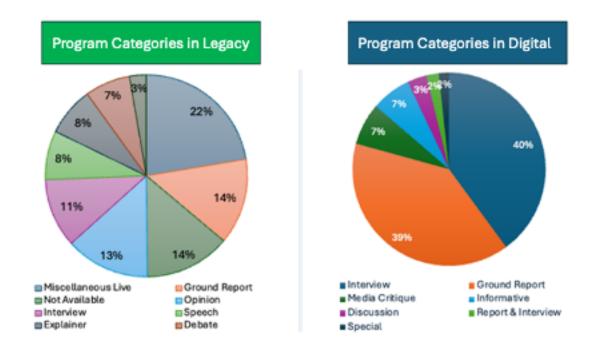


Figure 3: Share of videos per topic in legacy media (left) and digital media (right).

When it came to labeling for themes, 'miscellaneous live' videos have several shows on different themes (as described above), so we tagged these videos as 'miscellaneous' (83). Another 51 videos were unavailable, so we tagged these as 'NA' (51). Other than these, some of the top themes in legacy media include: foreign affairs (29), elections (22), COVID-19 (21), armed forces (18), and Taliban 2.0 (18). For digital media, some of the top themes include: COVID-19 (101), elections (49), media critique²⁵ (37), farmers' protest (23), and the economy (22). It is difficult to compare legacy and digital media based on themes, because a large share (46%) of legacy videos are either 'miscellaneous' or 'NA'. However, it is clear that digital media focused extensively on COVID-19.

Table 4 shows the posts with the most comments on Facebook for both legacy and digital media. The "Supreme Court judgment on the issue of Hijab ban in schools in Karnataka" program on *Republic World* (Republic's English news channel) received the highest number of comments (485) among legacy media posts; the story involved a split verdict, and a future

12

²⁵ 'Media critique' is a theme assigned to Newslaundry's weekly satire show titled "TV Newsance." In this show, the Newslaundry team review some of the most outrageous, silly, and propagandist news and reporting from legacy news media.

direction was not decided at that time.²⁶ For digital media, the program "Delhi Riots 2020", where *The Wire* showed Hindu right-wing extremist leader Yati Narsinghanand Saraswati's recurrent hate speeches and call for violence against Muslims (in the context of CAA-NRC protests), received highest number of comments (248); even though there was clear evidence against Mr. Narsinghanand (as this program showed), no legal action was taken against him.

Table 4: Posts with most comments on Facebook

Media	Headline	Channel	Comments
Legacy	Big Hijab Controversy LIVE: 'Not	RepublicWorld	485
	Wearing Hijab Leads to Misbehaviour,'		
	Says Samajwadi Party MP		
Digital	'Unki Ankhein Phod Di Jayengi'27: The	The Wire	248
	Anti-Muslim Cleric Who Called for		
	Final War in Delhi		

Table 5 shows the two posts that received the greatest engagement on Facebook for legacy media and digital media. An interview of Fareed Zakaria on *India Today* on "Russian invasion of Ukraine" received the highest engagement for legacy media on Facebook (5609); this program was aired just a few days after the Russian invasion on February 24, 2022, and the host tried to understand this conflict. For digital media, highest engagement (1842) was received by a four-minute video clip on *The Wire* in which Indian actress Swara Bhasker explained the CAA-NRC act and the protests against this act. In both terms of most comments and most engagement, legacy media has higher numbers than digital media on Facebook.

Table 5: Posts with highest engagement on Facebook

Media	Headline	Channel	Total Interactions
Legacy	What Next For Russia-Ukarine Crisis?	India Today	5609
	Political Scientist Fareed Zakaria		
	Speaks To Rahul Kanwal		
Digital	Swara Bhasker's Quick NRC-CAA	The Wire	1842
	Guide: The Protests and the Act,		
	Explained		

13

²⁶ https://www.scobserver.in/journal/sco-daily-supreme-court-issues-split-verdict-in-hijab-ban-case/

²⁷ English translation of "Unki ankhein phod di jayengi" is "Their eyes will be blown out."

Table 6: Videos with most comments on YouTube

Media	Headline	Channel	Comments
Legacy	ज्ञानवापी में मिल गए 'बाबा'? ²⁸ Poochta Hai Bharat Debate With Arnab Goswami Republic Bharat TV	Republic Bharat	2099
Digital	जो हिंदुस्तान का भविष्य है, वही मुसलमानों काभविष्य है: आरिफ़ मोहम्मद ख़ान ²⁹	The Wire	30,372

The most commented video on YouTube for legacy media was *Republic Bharat* debate on controversial Gyanwapi Mosque issue, while for digital media, the most commented video on YouTube was an interview given to *The Wire* by Arif Mohammed Khan (Governor of Kerala), discussing situation of Muslim in the present government (Table 6).

The most viewed video on YouTube for legacy media was a live broadcast of PM Modi's roadshow after BJP's win in the Gujarat state assembly elections of 2022; the video was viewed 2.7 million times and garnered 21,870 likes. For digital media, the most viewed video, with 3.6 million views, is the same video that received the most comments (Table 7).

Table 7: Most viewed video on YouTube

Media	Headline	Channel	Views
Legacy	LIVE: PM Modi in Gujarat day after Elections 2022 Success for Electric Victory Road-Show	Republic World	2,722,843
Digital	जो हिंदुस्तान का भविष्य है, वही मुसलमानों का भविष्य है: आरिफ़ मोहम्मद ख़ान ³⁰	The Wire	3,660,323

Of the eight videos in Tables 4–7, all except one (the most viewed video for legacy media) have a Hindu–Muslim binary element. This points to a current divisive and polarizing characteristic of the BJP's Hindu majoritarian politics, which keeps the pot boiling by whipping up the sentiments of Hindus for electoral gains. The BJP tries to convert every issue into a Hindu-vs.-Muslim contest. These issues were picked up by legacy media to spread the government's narrative and propaganda; to counter this, digital media also had to pick up these issues, to show an alternate narrative.

²⁸ English translation of "Gyanvapi me mil gaye baba?" is "Lord Shiva is found in Gyanvapi?"

²⁹ English translation of "*Jo Hindustan ka bhavishya hai, wahi Musalman ka bhavishya hai: Afir Mohammad Khan*" is "What the future of India is, is the future of Muslims: Arif Mohammad Khan." ³⁰ Ibid.

Topic Analysis

Next, we performed topic modeling on the YouTube video titles (i.e., headlines). For legacy media, 'live broadcast of PM Modi's speeches, rallies, and visits' is the most prominent topic (topic 0) (Figure 4). This suggests a significant focus on PM Modi by legacy/corporate media. Not a single topic about issues raised by opposition parties (e.g., unemployment, inflation) is featured among the top ten biggest topics. This indicates a potential imbalance in coverage, emphasizing activities related to PM Modi while giving less attention to opposition parties' activities or other important issues.

Additionally, top topics for legacy media appear to involve domestic issues where there is potential for polarization (e.g., Hindu-Muslim narratives), while topics such as unemployment and the economy receive less coverage. For example, the fourth most prominent topic (topic 3) in legacy media is 'Shraddha Walkar murder case,' which involved a Hindu woman murdered by her live-in partner, a Muslim man, in Delhi on May 18, 2022.³¹ This case received extensive coverage, potentially due to its religious context, which was framed by some media outlets as part of the 'love jihad' narrative—a conspiracy theory holding that Muslim men conspire to lure Hindu women into marriage in order to alter India's religious demography as part of a political takeover strategy (Frøystad, 2021). Similarly, issues such as 'hijab ban' and the 'gyanvapi mosque issue' (topic 8), both involving Hindu–Muslim contexts, are also among the top topics, along with 'Grand Strategy with Major General G.D. Bakshi' (topic 2) and 'Checkmate with Major Gaurav Arya (Retd)' (topic 6). The two latter topics are shows hosted by retired military officers on Republic TV that often present nationalistic themes and exhibit a perspective favorable to the current government; G.D. Bakshi discusses politics, defense, and security, while Gaurav Arya focuses primarily on Pakistan-related defense and security issues.

_

 $[\]frac{31}{https://indian express.com/article/cities/delhi/a-year-since-shraddha-walkar-murder-father-says-no-update-from-cops-court-9033927/$

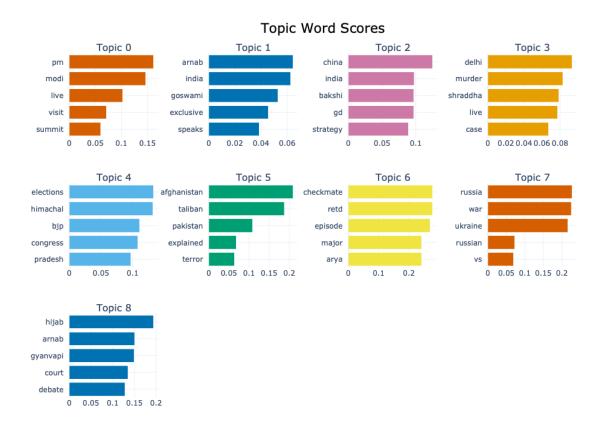


Figure 4: Top topics in legacy media YouTube videos

In contrast, the topics in digital media are more diverse, covering a wide range of issues including the budget, elections, protests, riots, COVID-19, and summit meetings (Figure 5). The most prominent topic is the show 'TV Newsance' (topic 0) (see footnote 14); other significant topics include the budget (topic 1), the Shaheen Bagh protest related to the Citizenship Amendment Act (CAA) (topic 2), state legislative elections (topic 3), the farmers' protests following the passing of three farm laws in 2020 (topic 4), Muslim-related issues (topic 5), and COVID-19-related issues (topics 6 and 7).

The topic analysis highlights the difference in news coverage between legacy media and digital media: legacy media focuses significantly on PM Modi, nationalistic themes, and polarizing issues, while digital media addresses a broader range of critical issues that impact the general population, such as COVID-19, farmer and CAA protests, and economic issues. It is important to note that legacy media often upload their most popular shows (typically prime time) on Facebook and YouTube, possibly because these shows, with their polarizing themes, attract substantial user engagement.

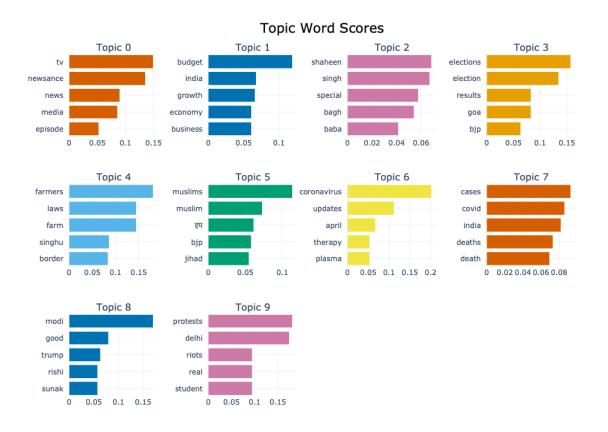


Figure 5: Top topics in digital media YouTube videos

Toxicity Analysis

To understand how users engage with news content, we performed a toxicity analysis. Our data consist primarily of English- and Hindi-language comments. For legacy media, we collected 23,861 English and 886 Hindi comments, while for digital media, we collected 211,127 English and 25,506 Hindi comments. In terms of toxicity results for English-language comments, the distribution of scores and the mean value were similar in legacy and digital media in all six categories (only 'insults' in digital media has a higher mean value than 'insults' in legacy media). The toxicity results for English comments are available in the Supplementary material.

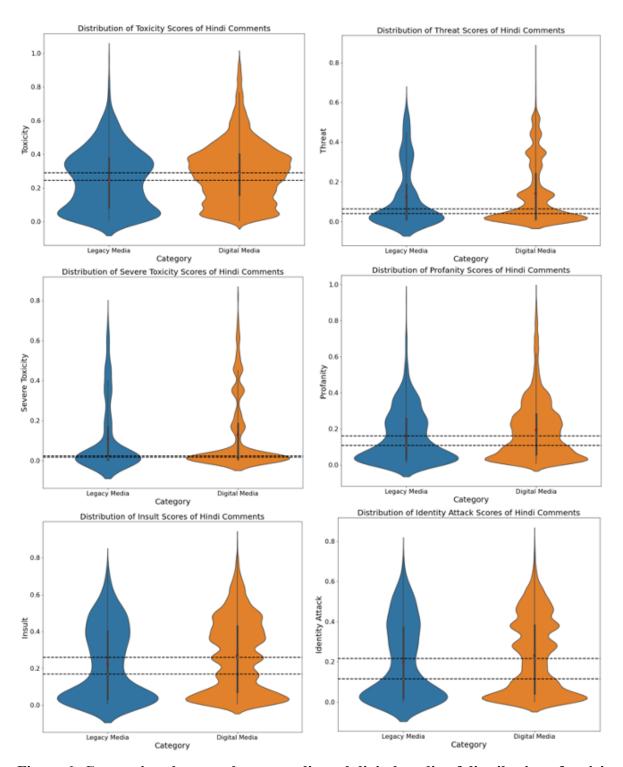


Figure 6: Comparison between legacy media and digital media of distribution of toxicity (all six categories) in Hindi comments.

The distributions of toxicity scores for Hindi comments (Figure 6) differ: the mean value in all categories of toxicity (except 'severe toxicity') is higher for digital media than legacy media. This suggests that YouTube users who are commenting in Hindi on digital media videos are using more toxic language than the users of legacy media videos. A possible explanation for this can be a diversity of opinions: previous research has shown that when

people encounter views that contradict their own, they may react with hostility and contempt (Zollo et al., 2017; Avalle et al., 2024). If the shared opinions are diverse, the higher the chances for disagreement and hence more heated debate; correspondingly, if a group is homogenous in their opinions, there will be no disagreement. The former situation seems to apply in digital media, while the latter seems to be the case for legacy media.

Network Analysis

To understand users in common to both media groups and their connections to the two, we created a bipartite network (Figure 7). There are two types of nodes: users and media channel (legacy and digital). An edge connects a user and a media if a user commented on a video. The degree (k) represents the number of comments given to separate videos; for example, if a user comments 10 times on one video, it is counted as k=1, and if a user comments on five different videos, it is counted as k=5. There are 210,968 users in total (nodes), of which legacy media has 24,441 users and digital media has 192,296 users. In this network, if a user commented on a video, they are connected to that media group, and the more a user commented, the stronger the link. Of the 210,968 users, 87.84% of users commented once (k=1), and 9.23% of users commented twice (k=2). Table 8 shows statistics related to Figure 7.

The green, blue, and black nodes that are present in the bridge that connects both legacy and digital media are the users that commented on both legacy and digital videos. We call these users 'common users', and there are 5769 such users (2.7%). Common users not only watch the content of both media groups but also interact and engage with people of different opinions; such interaction/discussion is imperative for a vibrant democracy. The small percentage of common users suggests that majority of users only follow the media conforming to their beliefs, which in turn suggests that there are few exchanges of different opinions between legacy users and digital users.

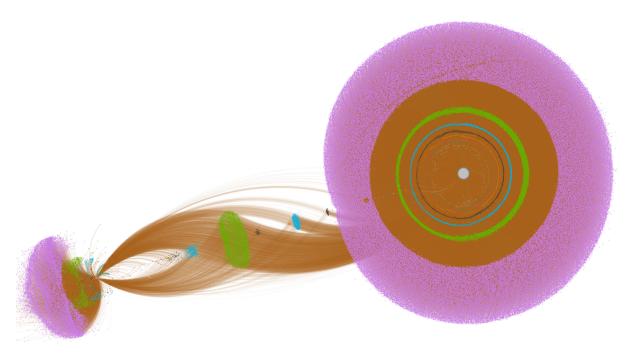


Figure 7: Bipartite network of users and their interactions with news media content. Purple (k=1), green (k=2), blue (k=3), black (k=4); k denotes the degree of a node.

Table 8: Network statistics for Figure 7

Nodes	210968
Edges	216735
Average Degree	2.055
Graph Density	0

Next, we selected the top 1000 users from each media group, based on the highest number of comments and constructed a network. In Figure 8, each node represents a user, and an edge represents a link between two users if both users commented on the same video; red edges represent a connection between users if they commented on digital media videos, while blue edges represent a connection between users if they commented on legacy media videos. Table 11 shows the network statistics for Figure 8. This network reveals that users of the two groups are poles apart and there is very little interaction between them, which suggests the presence of an echo chamber effect.

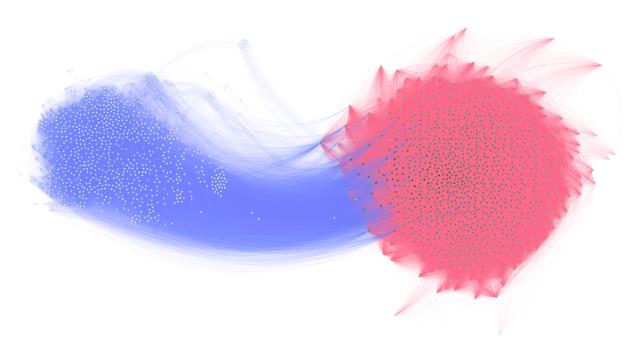


Figure 8: Network of top users, showing their interactions with each other

Table 9: Network statistics for Figure 8

Nodes	1892
Edges	460390
Average Degree	486.6
Graph Density	0.257

Discussion

In answer to RQ1, our content analysis reveals several characteristics of the two groups. Republic TV (341 videos out of 378) and The Wire (329 videos out of 548) are the most active in their respective groups in terms of uploading content on social media. As for category, 36% of the content from legacy media is either 'miscellaneous' or 'not available', while 79% of digital media videos are either 'interview' or 'ground report.' In seven of eight categories, the video with highest engagement featured Hindu–Muslim themes; this points to the significant role religion plays in India's current socio-political landscape and highlights how legacy media may use social media to encourage engagement and virality through polarizing content. Topic analysis further deepens our understanding of the news content of legacy media and digital media: the focus on PM Modi, nationalistic themes (shows on China and Pakistan), and communal issues (Shraddha Walkar murder case, Hijab ban issue) illustrate the thematic priorities of legacy media; in contrast, digital media cover a broader range of issues—including the economy, budget, riots, protests, elections, and media standards—despite their limited financial resources.

Toxicity analysis was used to determine users' participation in/interactions with the uploaded news content (RQ2). The mean toxicity for English-language comments is highly similar for both legacy and digital media, while Hindi comments on digital media are more toxic than those on legacy media in five of the six toxicity categories. Previous studies have shown that Perspective API is biased toward Hindi (Ghosh et al., 2021) and that Hindi-speaking users on Twitter tend to post more toxic, abusive, and offensive content (Agarwal et al., 2017; Rudra et al., 2016). One of our previous research papers also indicated that Hindutva supporters were more toxic in Hindi than in English (Ghasiya et al., 2023). However, this study compares the toxicity of users commenting in Hindi (only ~11% comments were in Hindi) for digital media versus legacy media. We identified a probable explanation for this phenomenon: the relative diversity among digital media users, compared to the homogeneity of legacy media users, could be a factor. Digital media attracts both pro-government users and those critical of government policies, whereas legacy media is primarily viewed by pro-government users. Progovernment users watching digital media often express strong dislike for the content or the journalists, especially when the video feature Muslims or Muslim-related issues; this also indicates the influence and dominance of BJP's communal politics, as well as legacy media's pro-government communal propaganda in the 'Hindi belt.' Table 10 shows a few examples of the most toxic comments posted under digital media videos.

Table 10: Most toxic Hindi comments posted under digital media videos (per toxicity category)

Author	Text ³²	Toxicity
no_user_name	मां चुदाने गया अलाह ईसलाम	0.968
	(Translation: Allah went to fuck my mother)	
no_user_name	बहन चोद के टांग पर क्यों मारा दिमाग पर मारना चाहिए था (Translation:	0.964
	This sister fucker should be hit on the head rather than the leg)	
no_user_name	ये साली कुतिया है। (Translation: She is a bloody bitch)	0.964
no_user_name	गद्दारो, तुम्हारी तो गांड में गोली मारनी चाहिए। (Translation: Traitors, you	0.960
	should be shot in the ass)	

In contrast, legacy media users tend to share similar beliefs, which are often conservative, nationalistic, and supportive of Hindutva ideology and the BJP. Since legacy media predominantly uploads content that aligns with these beliefs, users with different perspectives (such as those critical of the government) are less likely to engage with or comment on these

-

³² The translations from Hindi to English were performed by the author.

videos. The nature of toxic comments posted under legacy media videos differs, often targeting Muslims, Islam, opposition party figures, and Pakistan; examples are shown in Table 11.

Table 11: Most toxic Hindi comments posted under legacy media videos (per toxicity category)

Author	Text ³³	Toxicity
no_user_name	ओवेसी अपनी बकवास बंद करे।	0.840
	वरना मुँह बोलने के काबिल नहीं रहेगा (Translation: Owaisi shut your	
	mouth, otherwise you will not able to speak)	
no_user_name	भाई अगर सुब्रमण्यम स्वामी उसमे घुसे है तो सबकी मा चुदनी वाली है	0.787
	(Translation: If Subramanian Swamy is taking up this issue then	
	everybody is fucked)	
no_user_name	पाकिस्तान बलात्कार देश हैं 1971 पढ़ लो (Translation: Pakistan is a rape	0.687
	nation, read 1971)	
no_user_name	अल्लाह , इस्लाम, मुसलमान, सेकुलर,अब कभी भरोसा मत करना ।ये मक्कार	0.637
	दोगले झूठे (Translation: Do not trust Allah, Islam, Muslim, seculars	
	ever. They all are deceitful liars)	

Finally, network analysis provided the answer to RQ3, regarding user interaction between the two media groups. One of the networks (Figure 7) shows the existence of 'common users' who comment on videos from both media groups, though their number is a tiny (2.7%) proportion of the total number of commenters. These results, combined with our understanding of the diversity among digital media users, suggest that BJP supporters, Hindutva supporters, and Muslim haters are likely contributing to the higher mean toxicity in the comments sections of digital media. Another network (Figure 8) shows that the most active users form segregated network structures, resulting in the circulation of homogeneous information within the group. This indicates the presence of an 'echo chamber' among commenters of the two groups. There is low interaction between the top commenters from each group, creating polarizing environments. There may be several reasons for this low level of interactions between the two user groups, including lack of trust in 'other' media, confirmation biases, YouTube's algorithm (social media platforms' recommendation systems) (Alatawi et al., 2021; Rastegarpanah et al., 2019), and the belief that the 'others' cannot be persuaded; however, it is difficult to pinpoint the exact reason. While one might argue that low interaction is better than no interaction, in this case, cross-interaction tends to be highly toxic, rather than fostering healthy discussion.

-

³³ The translations from Hindi to English were performed by the author.

Limitations

Our study primarily focused on English-language news channels in India, with some Hindi content included due to shared digital platforms. This approach, while practical, has its limitations: by excluding channels in regional languages, we overlook the diverse perspectives and consumption patterns across India's many linguistic groups. This narrowed focus potentially impacts the comprehensiveness and representativeness of our findings, given India's rich linguistic diversity in its media landscape.

The language limitations affect objectivity in the Indian context and have implications for global generalizability. Media ecosystems, audience behaviors, and user engagement patterns vary significantly across countries and cultures, and our findings may not be directly applicable to countries with different media ecosystems and linguistic landscapes. For example, countries with a single dominant language might exhibit different dynamics in media consumption compared to multilingual societies like India. The impact of digital media can also vary based on such factors as internet penetration and cultural norms.

Future research could include a broader range of languages and could consider diverse media consumption patterns so as to enhance the objectivity and universality of the conclusions, both in India and globally. Moreover, it is crucial to expand the current analysis to encompass multiple social media platforms, for more robust conclusions and insights.

Conclusion

Our study represents the first comparative analysis of news coverage, user engagement, and user interaction for legacy media and digital media in India. We found that legacy media's news coverage on social media tends to be pro-government, highly nationalistic, and focused on Hindu–Muslim dynamics in Indian society; in contrast, digital media address issues of significant importance to the general public, such as the economy, public discontent with government policies, and administrative challenges. Additionally, our research revealed higher toxicity levels in Hindi comments on digital media, which can be attributed to users who support the government, believe in Hindutva ideology, and harbor negative sentiments toward Muslims, as these users often target digital media journalists due to ideological or religious differences. We also identified the presence of echo chambers in both media groups, characterized by low interaction between users of the two media groups. While some legacy media users comment on digital media videos to express dissent or hostility, this trend is not observed in the opposite direction.

Our findings have broader implications for the state of media in democratic societies around the world. The erosion of independent journalism and the rise of partisan media outlets are not unique to India but rather are part of a global trend wherein media organizations face similar pressures from political, economic, and social forces. In many democracies, media consolidation and the influence of powerful political entities threaten the fundamental role of the press as a check on government power. The situation in India underscores the importance of maintaining a diverse and independent media ecosystem, in order to ensure an informed citizenry and robust public discourse. The emergence of digital media as a potential counterbalance to traditional media highlights the need for continued support for independent journalism as well as the development of new media platforms that can uphold journalistic principles. These challenges are universal, and addressing them is crucial to the health and sustainability of democratic institutions globally.

Ethical statement

Facebook posts were collected using Meta's CrowdTangle platform and YouTube data was collected from publicly available YouTube API, hence complying with the privacy policy of both Facebook and YouTube.

Content Disclaimer

This manuscript includes explicit and abusive language encountered in the user comments. These examples are included to provide an accurate representation of the data and its implications. We have made every effort to handle this content ethically and sensitively, ensuring that the language is contextualized within the scope of the study. Readers are advised that the content may be distressing or offensive.

Disclosure statement

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was supported by JST CREST (JPMJCR20D3).

References

Goidel, K., Gaddie, K., & Ehrl, M. (2017). Watching the news and support for democracy: Why media systems matter. *Social Science Quarterly*, *98*(3), 836-855.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Herman, E. S., & Chomsky, N. (2008). *Manufacturing consent: The political economy of the mass media*. Random House.

Ognyanova, K., Lazer, D., Robertson, R. E., & Wilson, C. (2020). Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*.

Champlin, D., & Knoedler, J. (2002). Operating in the public interest or in pursuit of private profits? News in the age of media consolidation. *Journal of Economic Issues*, *36*(2), 459-468. Maheshwari, S., & Sparks, C. (2021). Political elites and journalistic practices in India: A case of institutionalized heteronomy. *Journalism*, *22*(1), 231-247.

Chadha, K. (2017). The Indian news media industry: Structural trends and journalistic implications. *Global Media and Communication*, *13*(2), 139-156.

Philipose, P. (2021). The Indian Media and Authoritarian Politics. アジア研究, 67(2), 58-68.

Aryal, S. K., & Bharti, S. S. (2022). Changing the media landscape in India under the Modi government: a case study based on the Narrative Policy Framework. *Studia z Polityki Publicznej*, 9(3 (35)), 47-64.

Kumar, A., & Dutta, P. (2023). The civility crisis: incivility, impoliteness and the descent of TV news debates in India. *Media Asia*, 1-7.

Kumar, A. (2023). Primetime Nationalism: Analysing Monologues on India's Republic TV and Times Now During Indo-China Border Conflict. *Journalism Studies*, *24*(12), 1518-1538.

Aneez, Z., T Neyazi, A., Kalogeropoulos, A., & Nielsen, R. (2019). India digital news report. *Reuters*.

Alatawi, F., Cheng, L., Tahir, A., Karami, M., Jiang, B., Black, T., & Liu, H. (2021). A survey on echo chambers on social media: Description, detection and mitigation. *arXiv* preprint *arXiv*:2112.05084.

Rastegarpanah, B., Gummadi, K. P., & Crovella, M. (2019, January). Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. In *Proceedings of the twelfth ACM international conference on web search and data mining* (pp. 231-239).

Mukerjee, S. (2024). Online news in India: a quantitative appraisal of the digital news consumption landscape in the world's largest democracy (2014–2018). *Information, Communication & Society*, 1-21.

Alshamrani, S., Abuhamad, M., Abusnaina, A., & Mohaisen, D. (2020, October). Investigating Online Toxicity in Users Interactions with the Mainstream Media Channels on YouTube. In *CIKM (Workshops)*.

Salminen, J., Sengün, S., Corporan, J., Jung, S. G., & Jansen, B. J. (2020). Topic-driven toxicity: Exploring the relationship between online toxicity and news topics. *PloS one*, *15*(2), e0228723.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* preprint arXiv:2203.05794.

Frøystad, K. (2021). Sound biting conspiracy: From india with "love jihad". *Religions*, *12*(12), 1064.

Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R. (2017). Deceiving google's perspective api built for detecting toxic comments. *arXiv preprint arXiv:1702.08138*.

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., Havlin, S., & Quattrociocchi, W. (2017). Debunking in a world of tribes. *PloS one*, *12*(7), e0181821.

Avalle, M., Di Marco, N., Etta, G., Sangiorgio, E., Alipour, S., Bonetti, A., Alvisi, L., Scala, A., Baronchelli, A., Cinelli, M., & Quattrociocchi, W. (2024). Persistent interaction patterns across social media platforms and over time. *Nature*, *628*(8008), 582-589.

Bastian, M., Heymann, S., & Jacomy, M. (2009, March). Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the international AAAI conference on web and social media* (Vol. 3, No. 1, pp. 361-362).

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141-161.

Mukherji, R. (2020). India's Illiberal Remedy. J. Democracy, 31, 91.

Sasahara, K., Chen, W., Peng, H., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2021). Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science*, 4(1), 381-402.

Ghosh, S., Baker, D., Jurgens, D., & Prabhakaran, V. (2021, November). Detecting Cross-Geographic Biases in Toxicity Modeling on Social Media. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)* (pp. 313-328).

Agarwal, P., Sharma, A., Grover, J., Sikka, M., Rudra, K., & Choudhury, M. (2017, January). I may talk in English but gaali toh Hindi mein hi denge: A study of English-Hindi codeswitching and swearing pattern on social networks. In 2017 9th international conference on communication systems and networks (comsnets) (pp. 554-557). IEEE.

Rudra, K., Rijhwani, S., Begum, R., Bali, K., Choudhury, M., & Ganguly, N. (2016, November). Understanding language preference for expression of opinion and sentiment: What do hindienglish speakers do on twitter?. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1131-1141).

Ghasiya, P., Ahnert, G., & Sasahara, K. (2023). Identifying Themes of Right-Wing Extremism in Hindutva Discourse on Twitter. *Social Media+ Society*, *9*(3), 20563051231199457.