

Unraveling the Dynamics of Television Debates and Social Media Engagement: Insights from an Indian News Show

Kiran Garimella¹, Abhilash Datta²

¹Rutgers University

²Indian Institute of Technology Kharagpur

kiran.garimella@rutgers.edu, abhilashdatta8224@gmail.com

Abstract

The relationship between television shows and social media has become increasingly intertwined in recent years. Social media platforms, particularly Twitter, have emerged as significant sources of public opinion and discourse on topics discussed in television shows. In India, news debates leverage the popularity of social media to promote hashtags and engage users in discussions and debates on a daily basis.

This paper focuses on the analysis of one of India's most prominent and widely-watched TV news debate shows: "Arnab Goswami – The Debate". The study examines the content of the show by analyzing the hashtags used to promote it and the social media data corresponding to these hashtags. The goal is to understand the composition of the audience engaged in social media discussions related to the show. The findings reveal that the show exhibits a strong bias towards the ruling Bharatiya Janata Party (BJP), with over 60% of the debates featuring either pro-BJP or anti-opposition content. Social media support for the show primarily comes from BJP supporters. Notably, BJP leaders and influencers play a significant role in promoting the show on social media, leveraging their existing networks and resources to artificially trend specific hashtags. Furthermore, the study uncovers a reciprocal flow of information between the TV show and social media. We find evidence that the show's choice of topics is linked to social media posts made by party workers, suggesting a dynamic interplay between traditional media and online platforms.

By exploring the complex interaction between television debates and social media support, this study contributes to a deeper understanding of the evolving relationship between these two domains in the digital age. The findings hold implications for media researchers and practitioners, offering insights into the ways in which social media can influence traditional media and vice versa.

1 Introduction

Television and social media platforms have emerged as influential players in shaping public opinion and political discourse. The interdependence between these two spheres has garnered considerable attention in recent years, as they collectively contribute to the agenda-setting process and influence public perceptions (Brojakowski et al. 2015; Valenzuela, Puente, and Flores 2017).

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

In this paper, we focus on the popular prime time debate show "Arnab Goswami – The Debate" on Republic TV, which is the most-watched English language news channel in India (Broadcast Audience Research Council 2022). Broadcasting from 9 to 11pm India time on weekdays, this show garners a significant viewership of over 5 million individuals across the country. An intriguing aspect of this show is its active engagement with social media platforms. Each day, the show's producers promote a specific hashtag related to the debate topic, urging social media users to actively participate in discussing and deliberating upon the subject matter. This integration of television programming with social media interaction forms the basis of our investigation.

By examining the content discussed on the show and analyzing the corresponding social media data, we aim to comprehensively understand the intertwined nature of television news debates, and social media support, and political influence, specifically in the context of the BJP in India. While previous research has explored the impact of media and social media on political communication (Bimber 2014; Gerodimos and Justinussen 2015), there remains a dearth of studies that investigate the connection between television shows and a political party's online support base. Our study fills this gap by examining the content, audience composition, and influence of social media discussions surrounding the television news debate show closely aligned with the BJP.

We show that there is a clear bias in the topics chosen by the show and they favor the BJP significantly. Next, we show that the audience who engages with the show's content on social media is mostly made up of BJP supporters, with official BJP spokespeople also actively promoting the show. These BJP supporters specifically amplify pro BJP and anti opposition narratives from the show that benefit their agenda. Given the influence and infrastructure the BJP has on social media, this provides a significant boost to the show. Finally, and most importantly, this relationship between the show and the party does not seem to be just one way. We find evidence that some of the hashtags being used on the show were previously used by BJP supporters, indicating that the show might be picking up topics to debate through content generated on social media.

This research is of paramount importance due to several reasons. Firstly, understanding the relationship between tele-

vision news debates and social media support is crucial in comprehending the contemporary media landscape and its impact on political communication. The influence of partisan media on public opinion and agenda setting has far-reaching implications for democratic processes. Secondly, analyzing the BJP's utilization of social media support and its symbiotic relationship with the television show sheds light on the tactics employed by political parties to shape public discourse and maintain their political influence. Such political influence might be an age old practice typically done through back channels and lobbying, but our study provides some evidence of its existence.

Previous work in this space either look at the topics of discussion on the show (Subhajt and Pegu 2021) or on the influence of the news on social media (De and Pal 2022). Such approaches that solely focus on either television news debates or social media discussions fail to capture the interdependence and feedback loop that exists between these two domains. By integrating content analysis of the television show, examination of social media support, and understanding the influence of political narratives, our study offers a comprehensive perspective on this intricate interplay. Also, even though literature on political communication and media effects have talked about the impact of traditional and new media (Guo 2019), quantitative evidence of the specific feedback loop observed between television news debates and political party supporters may not have existed prior to our work.

Understanding these dynamics is essential for fostering informed public debate, promoting media literacy, and safeguarding the democratic ideals that underpin a healthy media ecosystem.

2 Related Work

Use of hashtags on TV shows. The use of hashtags on television shows has proven to be an effective means of connecting with viewers on social media platforms. By incorporating hashtags, TV shows can encourage audience engagement, facilitate discussions, and promote the show to a wider audience. This approach not only generates free promotion and marketing for the show but also fosters a sense of community among viewers, ensuring the longevity of the program (Mukherjee and Jansen 2014).

Audience engagement has emerged as a vital aspect, especially with the rise of user-generated content (Newman 2009). In leading news organizations, audiences are increasingly becoming active contributors to the way stories are researched and told. This lends credence to our findings regarding the strong role of party workers and supporters in shaping the discussion topics of the debate show we study in this paper. User-generated posts often lead news bulletins, pointing to a historic shift in control towards individual consumers and grassroots political actors.

Scholars in the field of communication have long explored the relationship between television and social media, recognizing social media as a platform for re-articulating audience engagement. However, capturing and understanding audience engagement, especially with the proliferation of media

channels, has posed challenges for researchers (Moe, Poell, and Van Dijck 2016).

The emergence of social media as a “back channel” for television, as described in the book by (Proulx and Shepatin 2012), has opened up new opportunities for marketers to reach and engage audiences in innovative ways. The concept of “Social TV” gained attention in the early 2010s, highlighting the potential for viewers to connect with others who share similar interests and engage in synchronous discussions during television viewing (Wohn and Na 2011). However, despite initial enthusiasm, many television shows have since abandoned the use of social media channels as a primary means of promoting their shows (Hu et al. 2014; Guo 2019).

Numerous television shows have embraced the use of hashtags to engage their audiences. For example, shows like #TheEllenShow, #Survivor (a reality game show on CBS), and #TheTonightShow have encouraged fans to create and participate in discussions using specific hashtags. Popular television series such as #GameofThrones and #StrangerThings have also witnessed active hashtag-based conversations among dedicated viewers. Several news shows, including CNN's Anderson Cooper 360 (#AC360) and Fox News's Tucker Carlson Tonight (#TuckerCarlsonTonight), utilize hashtags as a promotional tool. However, these news programs do not employ distinct hashtags for individual episodes or segments.

One notable application of hashtags in the television realm has been during US presidential debates, starting in 2014. Robertson et al. (2019) argue that social media hashtags have played a democratic role in spreading information about political debates, while Lin et al. (2013) explore the social dynamics of emergent hashtags. Khosla et al. (2019) demonstrate that the use of debate hashtags on Twitter can even predict users' voting behavior.

Lastly, understanding the multifaceted relationship between media platforms and their publics adds another layer of complexity (Coleman and Ross 2010). Various publics are increasingly finding ways to subvert the gatekeeping functions of mainstream media to articulate their voice. This phenomenon deepens our understanding of the intricate dynamics we observed between the media, public discourse, and political engagement concerning the studied television debate show.

Impact of social media on newsroom content. The existing literature on the interrelationship between traditional news media and social media platforms is extensive and multidisciplinary, covering communications, media studies, and political science. In recent years, both news organizations and social media platforms have found that their missions increasingly overlap, each moving into domains traditionally ascribed to the other (Braun and Gillespie 2011). This confluence resonates with our study, where we observe a reciprocal flow of information between the TV show and its social media audience. In similar lines, news organizations are grappling with the need to host user communities that may not abide by traditional journalistic norms, while social media platforms are dealing with content that resembles news and challenges their established user guidelines.

Furthermore, the disruptive impact of social media on traditional news media’s business models is another area of concern (Newman 2011). This paper’s findings extend the context of our analysis by highlighting the need to understand the commercial implications of the interplay between social media and television debates. News organizations are becoming increasingly cautious about the influence of social media on their economic sustainability, which complements our observation on how the popularity of television debates is bolstered by targeted social media activities, particularly from political supporters.

The connection between social media and the choices editors make in newsrooms have also been studied previously in various other contexts. Sen and Yildirim (2015) study the causal relationship between social media popularity and how that influences editorial decisions in newsrooms. They show that editors prefer to schedule follow up content for stories which are popular on social media much more. Mukerjee, Yang, and Peng (2023) also show similar trends for editorial choices for newsrooms to post content on Facebook. Editors typically prefer posting more content on popular topics which are being engaged highly because of the advertising dollars associated with engagement. Petre (2020) explores the dual nature of social media metrics in journalism. While the metrics serve as a form of managerial surveillance and can lead to increased pressure on journalists to improve their metrics, some journalists also find ways to leverage metrics to assert their professional value and autonomy, challenging the perception of metrics as solely disempowering. The book offers insights into data-driven journalism and offers a glimpse into how the metrics revolution could impact other professions in the future.

Indian television debates and social media. Indian television debates have come under scrutiny recently due to their poor quality and biased nature (NewsLaundry 2020). In the study by Akbar et al. (2022), the authors demonstrate how television debates enabled the formation of social media communities to contribute to the politicization of a death by suicide. Chakravarty and Roy (2015); Subhagit and Pegu (2021) found that a majority of prime time debate shows in India exhibit a pro-government bias. Bhat and Chadha (2022) conducted an analysis of television shows in India and discovered that instead of facilitating meaningful deliberation on important civic issues, these shows often prioritize promoting religious majoritarianism, defending the policies of the Modi government, and advocating hyper-nationalism. The hosts of most TV debate shows employ a combative style and polarizing tone to stifle dissenting voices and impede free expression.

On social media, De and Pal (2022) found that the BJP engages more with anchors and commentators from these debate shows. It is therefore expected that the support for these shows predominantly comes from an audience sympathetic to the BJP’s agenda. Previous research (Jakesch et al. 2021) also reveals the existence of an organized machinery of users on social media who actively tweet and amplify BJP hashtags by making use of a complex network of WhatsApp groups and pre-written templates to trend hashtags.

These findings shed light on the complexities and bi-

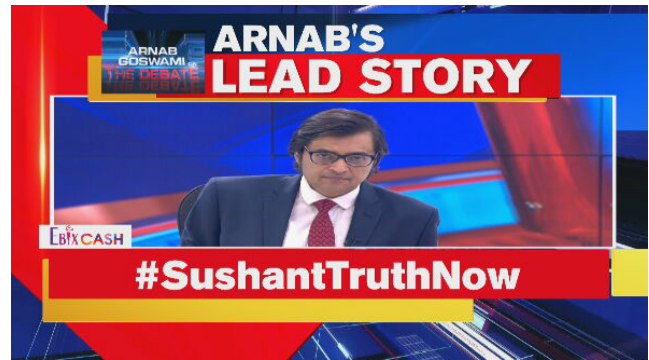


Figure 1: Example hashtag being promoted at the beginning of the show

ases within Indian television debate shows and social media, highlighting the need for critical examination and fostering of more inclusive and balanced discussions.

3 Datasets

3.1 Background

Republic TV is an Indian English-language news channel founded by the journalist Arnab Goswami in May 2017. The channel has been the most-watched English news channel in India since its inception, according to data from the Broadcast Audience Research Council (BARC) with an average viewership of 40% (Broadcast Audience Research Council 2022). Republic TV is known for its sensationalization of news, and its controversial anchor Arnab Goswami who has been accused of being biased and of consistently taking a pro-hindu, pro-nationalist and pro-government tone (Subhagit and Pegu 2021).

Republic TV’s flagship shows are the prime time debate shows hosted by Arnab Goswami at 9pm and 10pm every week day with over 5 million daily viewers. Every debate has a unique hashtag which is promoted by the host of the show. Since its inception, these debate shows have been using hashtags to promote discussion about the topic on social media. The anchors often remind viewers about using these hashtags to engage with them on social media. Figure 1 shows an example hashtag (#SushantTruthNow) being promoted at the beginning of the show.

3.2 Debate Hashtags Dataset

We first start by collecting all debate shows that aired on Republic TV in 2022 by scraping their website.¹ To isolate the effects of Republic TV debates on Twitter users and to remove the influence of other news channels, common hashtags that were used by both Republic TV and other news channels were removed from the dataset. We also manually went over the hashtags and removed any generic hashtags which could be used outside the show. This removed 28 hashtags. We were left with 636 hashtags which we can be sure that they were unique and used only by the show.

¹<https://www.republicworld.com/the-debate>

Next, using the Twitter API V2, we searched for these hashtags and obtained all the users who tweeted about these hashtags. We also apply a temporal filter and restrict our search to tweets that were created within a 3-day window before and after date the debate show aired using the hashtag. This additional step ensures that the tweets captured in the dataset are directly related to the debate show, and not due to other external factors.

Typically, hashtags used in debates have a relatively short lifespan, lasting anywhere from a few hours to a few days. Most tweets start immediately after the 9pm show airs and have a lifespan of a few hours. However, the longevity of a hashtag depends on several factors, including the popularity and relevance of the debate topic, the level of engagement from users, and the overall visibility of the hashtag across social media platforms. Some hashtags may continue to be used for several days or even weeks if they spark a broader public conversation or if they are picked up by influential social media users and amplified to a wider audience.

Next, we got the 3,200 most recent tweets for the users who have tweeted at least three of these hashtags. This cut off ensures that the users in our dataset were actively engaged in discussing the topics covered by the debates.² We also collected the profile information of these users.

Our final dataset contains 50,580 users with over 107 million tweets, of which 1.7 million contained hashtags used in the debates. 90% of the tweets were in either English or Hindi. Even though the show was only in English, we found roughly 40% of the tweets by the users were in Hindi.³

3.3 Classifying BJP Users

One of our primary hypothesis is that there is a connection between the BJP members and Republic TV shows. To study this connection, we need a classifier which can classify if a user is pro BJP or not. To identify users who are supporters of the BJP, we developed a machine learning model. Given a Twitter user, the model identifies whether the user is a BJP supporter or not.⁴ In the rest of the paper, we refer to these two classes as `pro-BJP` and `Other`.

To build a comprehensive dataset for analyzing political leanings on Twitter, we used two different types of sources. The dataset was annotated in a semi-supervised manner by leveraging the hashtags employed by users. We operated under the critical assumption that hashtags with established political leanings can serve as proxies for the user’s own political bias. This approach allowed us to acquire a large-scale

²The limit of three hashtags was decided based on the practical limitations of the API. Due to the uncertainty of the availability of the API at the time of the data collection (March 2023), we chose a cutoff that enabled us to get the most amount of data in a reasonable amount of time.

³ The language of the tweet was obtained directly from the Twitter API results (Twitter 2015).

⁴BJP supporters are broadly defined as supporters of causes that the party espouses, including Hindu religious accounts, and users who promote hyper-nationalism. The users who are not in the BJP supporter class could include a wide range of users who are either opposition party supporters or just news consumers.

	Num users
Coordinated posting (pro-BJP)	4,220
Trend4India (Other)	4,158
Nivaduck Pro-BJP	3,690
Nivaduck Other	3,780

Table 1: Number of users obtained from each dataset

dataset efficiently. We manually validate this assumption in later in the section.

1. Coordinated posting of party specific hashtags: For `pro-BJP`, we used the dataset from Jakesch et al. (2021), who show that these users coordinate and post BJP talking points on Twitter to get hashtags trending. Jakesch et al. provide concrete evidence (see Section 7.1 in their paper) that these accounts were indeed affiliated with the BJP. We obtained all tweets from these users. For `Other`, we used hashtags by the Twitter account `@trend4india`, which actively promotes anti BJP hashtags every day. Since there is no similar documented top-down operation run by the Congress party, we used the Twitter account `@trend4india`, which does not seem to be officially linked to the Congress but regularly promotes pro-Congress (and anti-BJP) hashtags and organizes trending raids similar to the BJP ones documented by Jakesch et al. (2021). A user is designated as `pro-BJP` or `Other` based on their utilization of a minimum of three unique hashtags from the corresponding sets.
2. We used the dataset of politician accounts from Panda et al. (2020) to obtain BJP and all other politicians for `pro-BJP` and `Other` respectively. We used hashtags used by opposition politicians to obtain additional accounts for the `Other` category.

Overall, this gave us 16k users (7.9k `pro-BJP` and 8k `Other`). We chose a stratified random sample of 12K users for training (6k of each class) and a stratified random sample of 4k users for validation (2k users of each class). The final number of users we used in our classification (after the sampling) is shown in Table 1. Next, we obtained the recent 3,200 tweets using the API for these users and pre-processed the data by removing stopwords, lower casing and lemmatization.

We created term frequency-inverse document frequency (tf-idf) vectors of the tweet text. We observed that mentions and hashtags particularly help in the classifier performance and hence created a separate vector of mentions, hashtags and urls for each user using td-idf. Following the technique presented in (Datta, Chakraborty, and Mukherjee 2022), we pass these mentions/hashtags/urls vectors through a five layer fully connected deep neural network and train it. Once the model has been properly trained, we take the output of the second last hidden layer of the neural network and call it the hashtag/mention/URL embedding for each user. According to Datta, Chakraborty, and Mukherjee (2022), this embedding captures the user’s personality features well, much more so than text.

	Precision	Recall	F1-score	Support
Other	0.93	0.89	0.91	1,600
Pro-BJP accuracy	0.89	0.93	0.91	3,170

Table 2: Classification report of the model on full validation set

For each user, we concatenated the embedding vector with the tf-idf vector from the tweets text and passed it to classical machine learning classifiers such as Random Forest Classifier, SVC, Logistic Regression, and XGBoost for classification. We found that Random Forest performs the best, scoring an accuracy of 0.91 and an F1 score of 0.91 on the validation set. Table 2 shows the results on the best model.

After the validation, we also manually tested our model on an unseen sample of 500 users. The testing data is selected randomly but is stratified to ensure that the class distribution is similar to the training and validation datasets. We obtain the predictions from the model on the testing data and then manually checked the user’s profile to see the correctness of the classification. The confusion matrix for the model from the manual testing is shown in Table 3. As we can see, the model performs reliably well on the task. We manually evaluated the small number of false positives and false negatives and found that the classifier confused journalists who mostly tweet for pro government stances (such as @OpIndia.com) as *Other*. In most false positive cases, users have used the word ‘Hindu’ in a non-political sense in some of their tweets.

		Prediction		Total
		pro-BJP	Other	
Actual	pro-BJP	207	31	238
	Other	9	253	262
Total		216	284	500

Table 3: Confusion Matrix on the test set

Finally, we applied the classifier over the entire Debate Hashtag Dataset and found that 24,149 (47.74%) of all users in the dataset are BJP supporters.

In our analysis, we further divide the BJP supporters to identify *active* BJP supporters as those who have tweeted more than 10 different hashtags of Republic debates. These users are not only supporters of BJP but are also actively engaged in discussions and debates related to Republic TV and BJP. In our dataset, there were 2,065 such users.

4 Analysis

4.1 Hashtag Analysis

We manually annotated all the 636 hashtags into different categories. One of the authors conducted the annotation through an iterative process. Initially, we established coding categories for hashtags, drawing inspiration from work

by Jafrelot, , and Jumle (2020). We then started qualitative coding to identify emergent themes, refining these through successive iterations until they could be aggregated into final codes. Most hashtags were unambiguously classifiable, but some necessitated group discussion among the authors for accurate categorization. We instituted a policy that each hashtag must be assigned a single ‘primary’ topic. In the few instances (less than 1% of cases) where a hashtag could belong to multiple categories, we allocated it to one primary topic for the sake of analytical rigor.

Upon annotation, we found that a significant portion of the debates covered anti opposition topics (34%), with other popular categories were being covered by the show pro BJP (28%), hyper-nationalism (9.5%)⁵, news (8.6%), Russia (6.7%) and Pakistan (3.5%). Only 2.1% of the shows featured hashtags that were anti BJP. This is in line with previous analysis which showed that the channel picks topics which are mostly pro government (Jafrelot, , and Jumle 2020).

Next, we analyze the activity of posting about these hashtags over time. Figure 2 shows the number of users who tweet the hashtag over time. We see that on average around 1,000-1,500 users post tweets about the debate every day. There are a few periods of high activity which are analyzed in Table 4. The red line (corresponding to the second y-axis) plots the ratio of the orange and blue bars, indicating the fraction of users on any day who are pro BJP. Interestingly, we see that this fraction is almost consistently around 60%.

By analyzing the news around the peaks in Figure 2, we want to understand whether there were any external events which drove an increase in engagement of the hashtags. As we can see (in Table 4), most of the peaks correspond to events which were popular politically and in many cases related to events where the BJP government was involved.

There seems to be a give and take relationship between the party and the channel, for instance, when Republic was acquitted in a case, there was a huge support by BJP folks even though it was not related to BJP or politics. Given how popular the show is, this finding raises questions about the influence of media on public opinion formation and the potential for media outlets to shape the narrative and priorities of political discourse.

4.2 Users Posting Debate Hashtags

Next, we look at the users posting the debate hashtags by looking at properties of twitter usage in the three sets of users (active BJP supporters, BJP supporters and other).

Our analysis of tweet count and follower count reveals a significant difference between active supporters and the other two user sets ($p < 0.001$). Active supporters have a considerably higher number of tweets and followers compared to the other groups, as shown in Figure 4 and Figure 5. No-

⁵We specifically opted for the term ‘hyper-nationalism’ over the term ‘nationalism’ used by Jafrelot, , and Jumle (2020) because the traits expressed by the show –labelling anyone disagreeing with them as ‘anti-national’ or aggressively advocating for exclusionary practices that marginalize certain communities– are better captured by the former term.

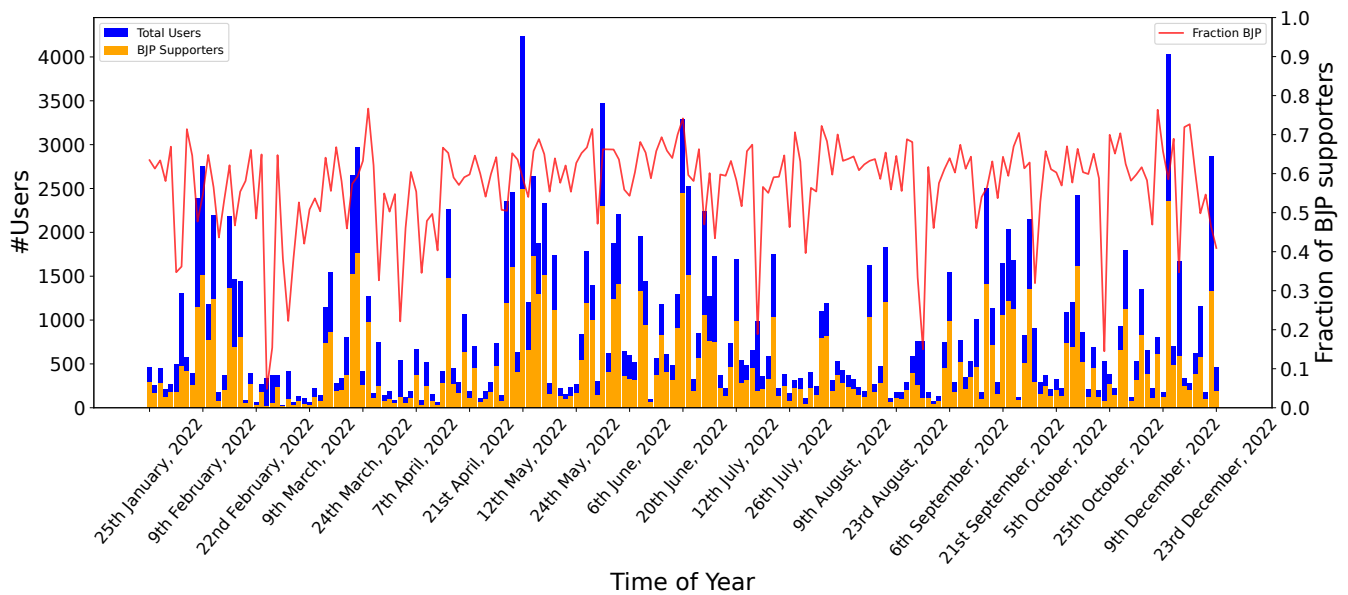


Figure 2: Number of users tweeting Republic TV debate hashtags over time. The blue bars show the total volume of users. The orange bars indicate BJP supporters. The red line (second y-axis) shows the fraction of the total users each day which was BJP supporters. We can observe that this fraction is consistently around 60%.

tably, the most followed and retweeted users within the active supporters group are predominantly BJP supporters or official spokespersons, such as @gauravbh, @Shehzad_Ind, and @shaziailmi. This finding establishes a clear connection between the individuals involved in the discussions and the amplification of messages related to these debates.

Additionally, the most followed and retweeted users in the general supporters category also consist mostly of BJP supporters or affiliates. However, it is important to note that their messages have limited reach, as indicated by the median number of retweets, which is 20 for supporters compared to 7 for non-supporters. This suggests that the influence and dissemination of messages are more concentrated among the active supporters and their affiliated accounts.

4.3 What Are They Amplifying?

The analysis reveals that these supporters selectively amplify content aligned with their own interests. Figure 3 provides insights into the preferred topics for each group, leading to several key findings:

- Supporters predominantly amplify pro-BJP and hyper-nationalism related content. Their focus is largely centered around promoting the party and its ideologies.
- Active supporters, in addition to pro-BJP and hyper-nationalism content, heavily engage in discussions related to the opposition, often taking an anti-opposition stance. Interestingly, they also show a significant interest in Bollywood content, such as criticizing Bollywood celebrities.
- Non-supporters, on the other hand, tend to focus more on general news and a diverse range of topics, including issues related to Russia.

The findings indicate a notable presence of BJP supporters on Twitter who actively engage in discussions surrounding Republic TV debates. This highlights the distinct preferences and priorities observed among different user groups. Supporters demonstrate a strong inclination to promote their preferred political party and its ideologies, whereas non-supporters engage in a wider range of discussions encompassing diverse news topics and international affairs.

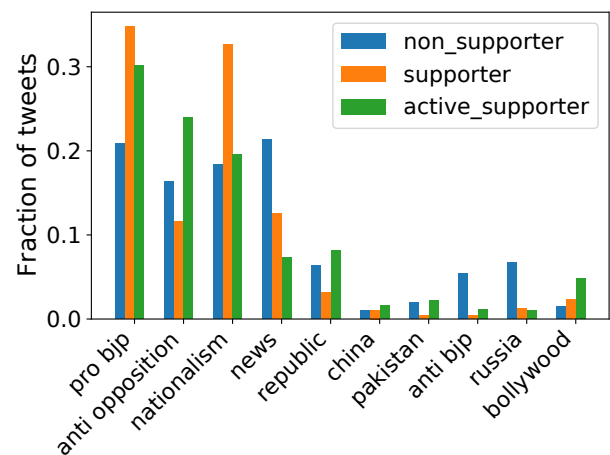


Figure 3: Topics being amplified by the three sets of users..

4.4 Coordinated Posting

Next, we looked at whether there was any coordinated posting promoting the debate hashtags. Jakesch et al.

Date Of Peak	Narrative
8th February 2022	Hijab ban ^a
18th March 2022	Kashmir Files ^b
14th April 2022	Religious tensions due to Ram Navami ^c
12th May 2022	Gyanvapi mosque ^d
1st June 2022	PFI Conspiracy ^e
20th June 2022	Opposition Government collapse in Maharashtra ^f
6th September 2022	Bengal violence ^g
20th September 2022	TRP report acquits Republic ^h
10th December 2022	India/China clashes ⁱ
22nd December 2022	Bharat jodo yatra ^j

^a <https://archive.is/DGg4V>

^b <https://archive.is/MTYUr>

^c <https://archive.is/31LXw>

^d <https://archive.is/geY1u>

^e <https://archive.is/ccOMl>

^f <https://archive.is/PoKkI>

^g <https://archive.is/2ECpn>

^h <https://archive.is/d7Tw0>

ⁱ <https://archive.is/VsJ4a>

^j <https://archive.is/70pgt>

Table 4: Narratives for the peaks in Figure 2.

(2021) present how pro BJP hashtags are manipulated using a network of WhatsApp groups to coordinate posting and get hashtags trending on Twitter.⁶ The manipulation works by using a set of template tweets which are copy pasted by a network of BJP supporters at a pre agreed time. We explored whether any such coordinated posting exists in our dataset. We used fuzzy matching with Levenshtein distance to identify near similar tweets. We were able to identify seven Republic debate hashtags which were manipulated this way in our dataset by BJP supporters. The hashtags have specific themes like promoting development of Kashmir (#TirangaAtLalChowk), supporting BJP government decisions (#indiawithagniveer), supporting Hindu causes (#hinduholocaust, #kashmirfilestruth, #justiceforharsha), supporting BJP victories (#yogireturns), and amplifying anti opposition topics (#banpfi). Though this only represents 1% of all the hashtags we studied, this bolsters the evidence that there is a concerted connection between the hashtags the show uses and the hashtags official BJP accounts promote. We also obtained data of what hashtags were trending on the day these shows aired and found that six of the seven hashtags were trending on the day they were manipulated.

4.5 What Else Do They Tweet?

The tweets containing republic hashtags only make up roughly 1.5% of the tweets by the users in our dataset. What else do the users in our dataset tweet? This would help us learn more about these users and the topics they post. To

⁶An example of coordinated posting: <https://archive.is/v9wfi> promoting the an initiative by the BJP.

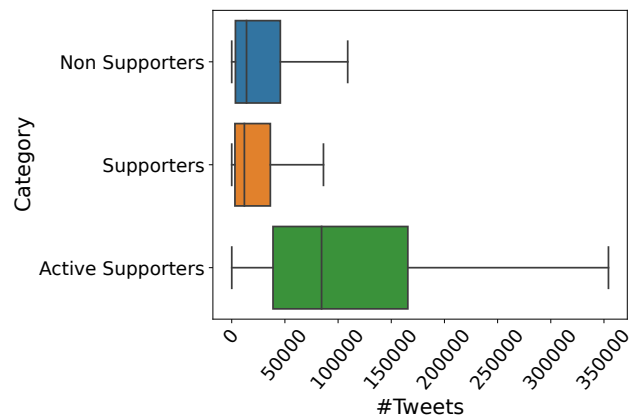


Figure 4: Comparison of the number of tweets for the three sets of users.

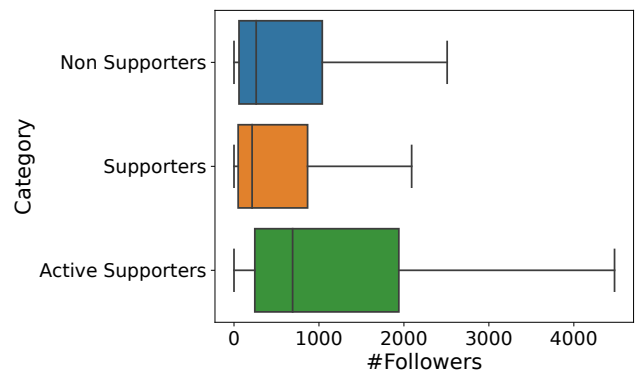


Figure 5: Comparison of the number of followers for the three sets of users.

identify these, we applied topic models on the users tweets. To identify topics, we first selected a tweets from each user category: active supporter, supporter, and non-supporter. We then employed Latent Dirichlet Allocation (LDA) for topic modeling on these sets. We use the LDA implementation provided by Gensim (Řehřek, Sojka et al. 2011) to obtain topics from our tweets. The LDA model generates a probability distribution over the topics for each tweet, indicating the likelihood of the tweet belonging to a particular topic. It also returns the most significant words belonging to a particular topic. We experimented with various topic numbers ranging from 10 to 150 and picked the one (N=50) which gave us the most coherent topics. We experimented with smaller numbers of topics, but found that they were insufficient to capture the richness and diversity of themes in the dataset. Topics were either too broad or overly fragmented, making it challenging to gain meaningful insights. On the other hand, when we tested with a higher number of topics, say 150, the topics became too granular and often exhibited significant overlap, making it harder to interpret and use the results effectively. After careful examination, we found that 50 topics struck a balance, providing a coherent and interpretable rep-

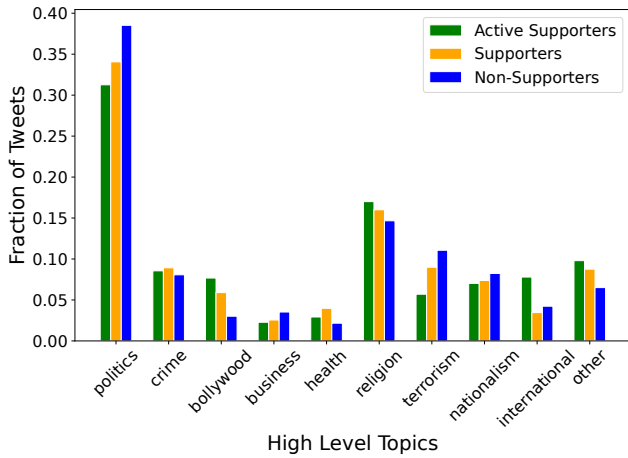


Figure 6: Comparison of tweet topics for the three user sets.

resentation of the underlying structures within the data without being overly detailed or redundant. We run separate topic models for the three different sets of users (active supporters, supporters and non supporters).

After generating 50 topics for each dataset, we manually label them with high-level categories. We chose 10 different high-level categories to label the topics. These are politics, religion, crime, business, health, hyper-nationalism, international, terrorism, bollywood and other. We ensure that the labeling is consistent across all the different user sets.⁷ To label the topics, we follow a two-step process. First, we analyze the most important words within each topic and identify the underlying theme. Then, we assign a high-level category to each topic based on the underlying theme. For example, if a topic has frequent words as ‘BJP’, ‘Modi’, and ‘Manifesto’, we assign it to the ‘politics’ category, and if a topic has frequent words as ‘Hindu’, ‘Shiva’, and ‘Muslim’, we assign it to the ‘religion’ category. If some topic, doesn’t match any of our high-level categories, we assign it ‘other’ category. We found that the topics we came up with in Section 4.1 were too narrow for this analysis. This is expected since the users might tweet many other topics outside of the topics covered by the debates.

Once we labeled the topics, we calculate the counts of tweets in each high-level category in our dataset. We combine all counts of topics belonging to a specific high-level category to find its total count. This provides insights into the most prevalent topic categories among the different user sets. A comparison of the tweet counts of different categories between the three user types is shown in Figure 6.

On comparing the tweet distributions of the three user types, we see that non supporters tweet much more about Politics and issues around terrorism. Supporters and active supporters tweet more on religion and Bollywood. This rhetoric of supporters of the BJP consistently targeting par-

⁷This was needed to be able to compare topics across the different user sets. However, we had to make a compromise and choose high level topics which had to be aggregated at a level which matches all three user sets.

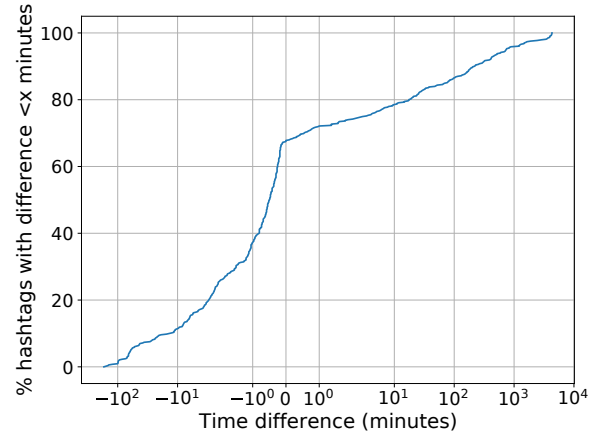


Figure 7: CDF of the time difference (in minutes) between when a hashtag was first shared on social media and when @republic or @republic_bharat tweeted the hashtag.

ticular Bollywood celebrities and movies has been documented previously (Kommiya Mothilal et al. 2022; Akbar et al. 2022). Overall, our analysis closely aligns with previous literature. It is expected that active BJP supporters would be highly engaged in political discourse and religious matters, while non-BJP supporters would be more concerned with social and political issues.

4.6 Impact of Social Media on Debate Hashtag Choice

Finally, we study whether the hashtags chosen by the show are in any way connected to user’s activity on social media. To study this, we check whether the hashtags used by the show were used prior by anyone on social media. Since most of the hashtags used by the show are peculiar and unique, the chances of them being used earlier, even by BJP supporters should be rare. Concretely, we computed the following: for each hashtag used by Republic, we get the time when it was first posted by any of Republic’s Twitter accounts (denoted by t_{TV}^h) and the first usage of the hashtag by a BJP supporter (denoted by t_{SM}^h). We compute difference $t_{TV}^h - t_{SM}^h$. If the difference is negative, this means that Republic posted the hashtag first. If the difference is positive, a BJP supporter posted the hashtag first before it appeared on Republic TV. Figure 7 shows a CDF of this difference over all the hashtags. We can see that around 35% of the hashtags were used on social media prior to being used by Republic TV. Around 15% of the hashtags were used at least 2 hours before Republic accounts used them.

Qualitative analysis of these hashtags reveals interesting observations. Take the case of the hashtag #AbuseNo92, a hashtag keeping count of the number of abuses by the Congress party on Prime Minister Narendra Modi.⁸ The hashtag is very specific and there is a little chance that someone else had tweeted it earlier to it being chosen. The first

⁸<https://archive.is/y1XBm>

tweet for this hashtag was done at 4pm India time by a user @SMedia4⁹ and then there was a show using the hashtag at 10pm. A few more examples of this pattern include: (i) #HinduphobiasReal tweeted 3 days prior to the show using the hashtag (by a user who mostly tweets anti Congress topics), (ii) #KashmirFilesPolitics was tweeted 2 days earlier by a user who claims to be ‘Proud to be an Indian, BJP supporter’, (iii) #BanPFINow being used by a user 2 days earlier by a user who says in his bio ‘Luv my India and Narendra Modi Sir’.¹⁰

We aggregated the topics which were posted earlier by social media users and Figure 8 shows that most of them were anti opposition or pro BJP hashtags. Note that it is important to note that we do not claim that there is a causal connection here since identifying a causal connection is not possible given our data. However, given that close to a 100 hashtags in the past year seem to have been used by BJP supporters prior to being used on the show does provide a signal that the show might be taking inspiration from social media users.

To find further evidence of this, we used topics labeled in Section 4.5 and created a time series of topic fractions for each of the three user sets per week. For each week, we compute the fraction of tweets belonging to a certain topic. We also labeled all the hashtags used by Republic into the same topics and created time series of the shows by Republic per week. Next, we applied Granger Causality tests (Granger 1969) on these time series to check if any of the topic time series can help predict the Republic TV show time series. Granger Causality test is a statistical test that is used to determine if a given time series and its lags are helpful in explaining the value of another series. If a certain time series t_1 ‘Granger causes’ another time series t_2 , it statistically means that the values in t_2 lag the values in t_1 by a certain amount and that t_1 can be used to predict values in t_2 . We set a lag of 4 weeks and run Granger Causality tests for all the 10 topics for the three user sets. Using a $p < 0.01$, we find that only three of the time series have statistically significant values. These are topic time series for Bollywood and hyper-nationalism for active supporters and hyper-nationalism for the supporters group. This analysis provides more evidence that certain types of topics used by BJP supporters are correlated with hashtag choices made by Republic TV.

5 Discussion

The study reveals that social media platforms, particularly Twitter, have emerged as influential channels for public opinion and discourse on issues discussed in television shows. The popularity of these platforms has been effectively harnessed by news debates in India, as they actively promote hashtags and encourage users to engage in discussions and debates on social media. Though this nexus between social media and television is over a decade old (Proulx and Shepatin 2012), there is not much study in

⁹<https://archive.is/OcUqG> The user claims on their bio ‘My TL reaches Modi ji ...’

¹⁰All three users have less than a thousand followers. To protect the privacy of these users, we do not include their screen name.

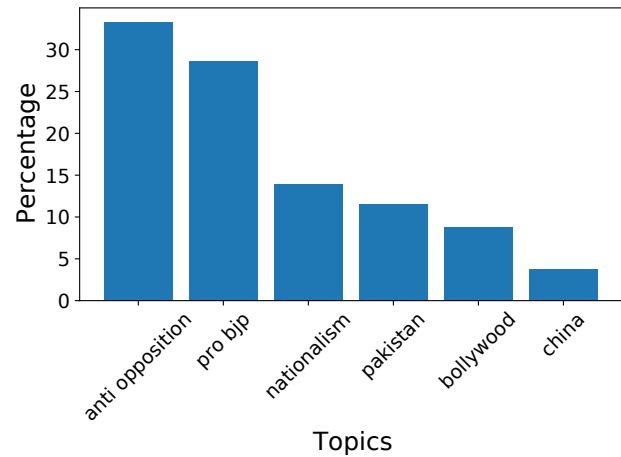


Figure 8: Topics posted earlier

the context described in this paper — using social media to study the bias of a news show.

The analysis of hashtags used by the news debate show indicates a significant bias towards the ruling Bharatiya Janata Party (BJP). Over 60% of the debates featured content that either favored the BJP or criticized the opposition. By dominating the content and narrative of the show, the BJP effectively controls the information flow and influences the issues that are highlighted and discussed. This enables the party to set the agenda and shape public opinion in alignment with its own political interests. While the initial impression of the finding may seem ‘trivial’, its significance becomes apparent when considering our subsequent two findings.

Our next significant finding is the role of social media support in amplifying the show’s reach and promoting specific hashtags. It is evident that the show garners substantial support from BJP supporters on social media platforms. BJP leaders and influencers actively participate in promoting the show, leveraging their existing networks and resources to artificially amplify and trend specific hashtags (Jakesch et al. 2021).

Finally, completing the cycle, we identify the reciprocal flow of information between social media and the topics picked for debate on the show. Our evidence suggests that the show’s choice of topics is influenced by social media posts made by party workers. While causality cannot be established in this particular setup, our results suggest a low probability of this occurrence happening by chance. This is notably influenced by the active participation of BJP supporters in promoting the channel’s content and the adoption of hashtags by influential users within their ranks.

These findings highlight the dynamic interplay between traditional media and online platforms, where political actors utilize social media as a tool to shape public discourse and extend their influence beyond the television screen and in turn social media discussions can influence the agenda-setting process of television debates. This nexus means that the BJP sets a narrative on social media which is picked up and ‘debated’ (amplified) by the TV show giving it legiti-

macy and prime time coverage, while the same BJP supporters amplify the debate narrative on social media promoting it to be an issue of national importance. It underscores the need to consider the complex inter-dependencies and feedback loops between these domains when analyzing the media landscape in the digital age.

Understanding the importance of our findings in the agenda setting by the BJP is crucial for several reasons. Firstly, our study documents the extent to which the relationship between the show and the party exist. It sheds light on the strategies employed by political parties to control the narrative and influence public opinion. It is important to note that such agenda setting on media might already be happening through back channels (such as personal connections or lobbying) and are not usually visible. Our study provides some quantitative evidence of this connection.

Secondly, it underscores the need for media literacy and critical engagement among the public. Awareness of the biases present in media outlets and the influence of political actors on agenda setting is essential for individuals to make informed judgments and engage in meaningful debates.

Lastly, our findings call for further research and scrutiny into the role of political parties in agenda setting and the implications for democratic processes. The ability of a political party to control the agenda and manipulate public discourse raises concerns about the diversity of voices and perspectives being represented in the media landscape. It prompts discussions on media ethics, journalistic independence, and the role of media regulators in maintaining a balanced and inclusive media environment.

Limitations and Broader Impact. While the study provides valuable insights into the relationship between a TV news debate show and social media support, there are several limitations that should be acknowledged:

Generalizability: The study focuses on one specific TV news debate show in India, which may limit the generalizability of the findings to other shows or contexts. The dynamics observed in this particular case may not be representative of the entire media ecosystem in India or other countries. Even for India, English news viewers make up a small fraction of the viewership of news. Given the prevalence of media bias in multiple countries, it would be interesting to see if similar patterns exist for other countries or languages.

Sample Selection Bias: The analysis of social media data relies on a specific selection of hashtags related to the TV show. This selection may introduce a bias by excluding other relevant discussions or perspectives that do not use those specific hashtags. On a similar note, the topic of the show and its leaning/bias is decided based on the hashtag and not the content of the show which may also introduce bias in our analysis.

Choices in user classification: Our analysis quantifying BJP support for the show hashtags depends on an automated classifier. Even though the classifier performs well in terms of accuracy, there are false positives, which might bias the results. Our results must be interpreted with a margin of error that could include these false positives. Our classifier identifies whether a Twitter user is a BJP supporter or not. Non supporters are not a homogenous group and con-

tain a wide range of users from diverse ideologies, including neutral users. A finer way of classifying users should be thought of in the future. Another issue which could impact our results is the misclassification. We thoroughly went through these examples of false positives and false negatives to understand where our classifier gets it wrong and tried to include features that capture such behavior. We found some cases where the classifier systematically gets it wrong. Our manual analysis indicates that the misclassifications are mostly towards classifying pro BJP as Other (the other way was quite rare). In most of these cases, the accounts belonged to journalists or activists who support the BJP, but also tweet about other news related stuff. With these cases in mind, we tried to understand the downstream impact of our findings. Since our paper mostly focuses on pro-BJP support to the TV show, if anything, our results are slightly under counting the support.

In grappling with the ethical, political, and interpretive challenges inherent to our methodology, we acknowledge the complex landscape of decisions that come into play. Data labeling is not a neutral act; the very act of categorizing users as ‘supporters’ or ‘non-supporters,’ for instance, could be interpreted as an oversimplification of a complex and nuanced spectrum of public opinions. Additionally, the allocation of topics to tweets and hashtags is intimately tied to the politics of representation, often reflecting latent power structures and sociopolitical orientations within the media and broader society. To mitigate these issues, we employed a rigorous process that involved consulting multiple sources of information, including academic literature and collective authorial discussions, especially in cases where the categorizations were ambiguous. We also established guidelines to ensure that each hashtag or topic would be allocated to a single ‘primary’ category, thereby maintaining analytical consistency and reducing interpretive bias.

On another note, our positionality in this research is anchored in a profound commitment to shed light on the anti-democratic tendencies manifest in the media landscape in India. As scholars, we are concerned with how specific media outlets may exert disproportionate influence in shaping public opinion, often in ways that align with the interests of ruling political parties, thus undermining the essential democratic dialogue. This informs not only the subject of our research but also the interpretive lens we adopt in analyzing our dataset. It is crucial to acknowledge this positionality to fully appreciate how our research questions were formulated, how the data were evaluated, and how our findings contribute to larger, ongoing debates about media ethics and the role of democracy in a digitized world.

Causality and Directionality: While the study suggests a reciprocal flow of information between the TV show and social media, it does not establish causality or the direction of influence. The observational setup we have in this paper does not allow us to make causal claims. Further research would be needed to determine whether the potential direction of influence we point out is actually a causal link.

Twitter focus: The study focuses on Twitter as a social network for convenience reasons. However, since Twitter is used only by elites in India, not focusing on other plat-

forms like WhatsApp or Facebook might bias our findings to a certain type of user. If Twitter is an elite social network with only a small fraction of the Indian electorate on it, why is a party like the BJP focused on creating infrastructure on Twitter? While the overall user base on Twitter may be smaller, it often includes influential individuals, journalists, opinion leaders, and celebrities who have a wide reach and can shape public opinion. Engaging with this influential subset of users can have a ripple effect on broader public discourse.

References

- Akbar, S. Z.; Sharma, A.; Mishra, D.; Mothilal, R. K.; Negi, H.; Nishal, S.; Panda, A.; and Pal, J. 2022. Devotees on an Astroturf: Media, Politics, and Outrage in the Suicide of a Popular FilmStar. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)*, 453–475.
- Bhat, P.; and Chadha, K. 2022. Expanding public debate? Examining the impact of India’s top English language political talk shows. *Media Asia*.
- Bimber, B. 2014. Digital media in the Obama campaigns of 2008 and 2012: Adaptation to the personalized political communication environment. *Journal of information technology & politics*, 11(2): 130–150.
- Braun, J.; and Gillespie, T. 2011. Hosting the public discourse, hosting the public: When online news and social media converge. *Journalism Practice*, 5(4): 383–398.
- Broadcast Audience Research Council. 2022. Broadcast Audience Research Council Data 2022.
- Brojakowski, B.; Burnett, J. R.; Cassella, R.; Castleberry, G.; Collins, M.; Dickinson, T. M.; Ducey, M.; Faltsek, D.; Fogle, K.; Geltzeiler, B.; et al. 2015. *Television, social media, and Fan Culture*. Lexington Books.
- Chakravarty, P.; and Roy, S. 2015. Mr. Modi goes to Delhi: Mediated populism and the 2014 Indian elections. *Television & New Media*, 16(4): 311–322.
- Coleman, S.; and Ross, K. 2010. *The media and the public: “them” and “us” in media discourse*, volume 9. John Wiley & Sons.
- Datta, A.; Chakraborty, S.; and Mukherjee, A. 2022. Personality Detection and Analysis using Twitter Data.
- De, S.; and Pal, J. 2022. Indian Journalists on Twitter and their Politician Footprint. <https://archive.is/wDkVA>. Accessed: 2024-04-01.
- Gerodimos, R.; and Justinussen, J. 2015. Obama’s 2012 Facebook campaign: Political communication in the age of the like button. *Journal of Information Technology & Politics*, 12(2): 113–132.
- Granger, C. W. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.
- Guo, M. 2019. Social television viewing with second screen platforms: Antecedents and consequences. *Media and Communication*, 7(1): 139–152.
- Hu, H.; Huang, J.; Zhao, H.; Wen, Y.; Chen, C. W.; and Chua, T.-S. 2014. Social tv analytics: a novel paradigm to transform tv watching experience. In *Proceedings of the 5th ACM Multimedia Systems Conference*, 172–175.
- Jafrelot, C.; ; and Jumle, V. 2020. A study of 1,779 Republic TV debates reveals how the channel champions Narendra Modi. *The Caravan*.
- Jakesch, M.; Garimella, K.; Eckles, D.; and Naaman, M. 2021. Trend alert: A cross-platform organization manipulated Twitter trends in the Indian general election. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2): 1–19.
- Khosla, S.; Chhaya, N.; Jindal, S.; Saha, O.; and Srivastava, M. 2019. Do events change opinions on social media? Studying the 2016 US Presidential debates. In *Social Informatics: 11th International Conference, SocInfo 2019, Doha, Qatar, November 18–21, 2019, Proceedings 11*, 287–297. Springer.
- Kommiya Mothilal, R.; Mishra, D.; Nishal, S.; Lalani, F. M.; and Pal, J. 2022. Voting with the Stars: Analyzing Partisan Engagement between Celebrities and Politicians in India. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1): 1–29.
- Lin, Y.-R.; Margolin, D.; Keegan, B.; Baronchelli, A.; and Lazer, D. 2013. # Bigbirds never die: Understanding social dynamics of emergent hashtags. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, 370–379.
- Moe, H.; Poell, T.; and Van Dijck, J. 2016. Rearticulating audience engagement: Social media and television. *Television & new media*, 17(2): 99–107.
- Mukerjee, S.; Yang, T.; and Peng, Y. 2023. Metrics in Action: How Social Media Metrics Shape News Production on Facebook. *Journal of Communication Special Issue on Social Media*.
- Mukherjee, P.; and Jansen, B. J. 2014. Social TV and the social soundtrack: significance of second screen interaction during television viewing. In *Social Computing, Behavioral-Cultural Modeling and Prediction: 7th International Conference, SBP 2014, Washington, DC, USA, April 1-4, 2014. Proceedings 7*, 317–324. Springer.
- Newman, N. 2009. The rise of social media and its impact on mainstream journalism.
- Newman, N. 2011. Mainstream media and the distribution of news in the age of social media.
- Newslaundry. 2020. Trial by Media: How Arnab and Navika Spent Over 65 Percent of Their Debates on SSR.
- Panda, A.; Gonawela, A.; Acharyya, S.; Mishra, D.; Mohapatra, M.; Chandrasekaran, R.; and Pal, J. 2020. Nivaduck-a scalable pipeline to build a database of political twitter handles for india and the united states. In *International Conference on Social Media and Society*, 200–209.
- Petre, C. 2020. *All the news that’s fit to click: How metrics are transforming the work of journalists*. Princeton University Press.

Proulx, M.; and Shepatin, S. 2012. *Social TV: how marketers can reach and engage audiences by connecting television to the web, social media, and mobile*. John Wiley & Sons.

Řehřek, R.; Sojka, P.; et al. 2011. Gensim—statistical semantics in python. Retrieved from genism.org.

Robertson, C. T.; Dutton, W. H.; Ackland, R.; and Peng, T.-Q. 2019. The democratic role of social media in political debates: The use of Twitter in the first televised US presidential debate of 2016. *Journal of Information Technology & Politics*, 16(2): 105–118.

Sen, A.; and Yildirim, P. 2015. Clicks bias in editorial decisions: How does popularity shape online news coverage? Available at SSRN 2619440.

Subhajit, P.; and Pegu, U. K. 2021. Media Polarization and Assertion of Majoritarianism in Indian News Media. *The Journal of Communication and Media Studies*, 6(2): 1.

Twitter. 2015. Evaluating language identification performance — blog.twitter.com. https://blog.twitter.com/engineering/en_us/a/2015/evaluating-language-identification-performance. [Accessed 11-09-2023].

Valenzuela, S.; Puente, S.; and Flores, P. M. 2017. Comparing disaster news on Twitter and television: An intermedia agenda setting perspective. *Journal of Broadcasting & Electronic Media*, 61(4): 615–637.

Wohn, D. Y.; and Na, E.-K. 2011. Tweeting about TV: Sharing television viewing experiences via social media message streams. *First Monday*.

Ethics checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **Yes**
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper’s contributions and scope? **Yes**
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes**
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes**. See Section 5 for a discussion on the biases and generalizability.
 - (e) Did you describe the limitations of your work? **Yes**. See Section 5, subsection “Limitations and Broader Impact” for a section on limitations.
 - (f) Did you discuss any potential negative societal impacts of your work? **Yes**. See Section 5, subsection “Limitations and Broader impact”.
 - (g) Did you discuss any potential misuse of your work? **Since the outcome involves public facing results (e.g. about a political party), we took extra care to make sure the results are not misinterpreted (e.g. see “Limitations and Broader Impact” subsection in Section 5.**
- (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **Yes**
- (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes**
2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? **NA**
 - (b) Have you provided justifications for all theoretical results? **NA**
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **NA**
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **NA**
 - (e) Did you address potential biases or limitations in your theoretical framework? **NA**
 - (f) Have you related your theoretical results to the existing literature in social science? **NA**
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? **NA**
3. Additionally, if you are including theoretical proofs...
 - (a) Did you state the full set of assumptions of all theoretical results? **NA**
 - (b) Did you include complete proofs of all theoretical results? **NA**
4. Additionally, if you ran machine learning experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **No**
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **Yes**
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **Yes**
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **NA**
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? **Yes**
 - (f) Do you discuss what is “the cost” of misclassification and fault (in)tolerance? **Yes**
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
 - (a) If your work uses existing assets, did you cite the creators? **Yes**
 - (b) Did you mention the license of the assets? **NA**
 - (c) Did you include any new assets in the supplemental material or as a URL? **NA**

- (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? **Yes**
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **Yes**
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see ?)? **Yes**
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see ?)? **NA**
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity...**
- (a) Did you include the full text of instructions given to participants and screenshots? **No**
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? **Yes**
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **No**
 - (d) Did you discuss how data is stored, shared, and de-identified? **Yes**