Analysis of Media Bias in Policy Discourse in India

Anirban Sen
t-anirbansen@microsoft.com
Microsoft Research
Bengaluru, India

Debanjan Ghatak
IIT Delhi
New Delhi, India

Sanket Dhakate
IIT Delhi
New Delhi, India

Kumari Rekha
IIT Delhi
New Delhi, India

Mehak Gupta
IIT Delhi
New Delhi, India

Gurjeet Singh Khanuja
IIT Delhi
New Delhi, India

Kartikeya Sharma
IIT Delhi
New Delhi, India

Aaditeshwar Seth
aseth@cse.iitd.ac.in
IIT Delhi
New Delhi, India

ABSTRACT

Many citizens consume information on government policies from the mass media. Consequently, biases existing in the policy discourse in media sources may influence citizens' understanding of the policies, about how they may affect diverse communities. These biases may also get amplified further through social media if it simply echoes the biases of mass media content. We build methods to quantify media bias in terms of preferred treatment given to certain issues corresponding to four economic policies, and alignment observed with the ideological stance of different political parties. We also examine how the social media community of followers of these media houses contribute to the policy discourse. Other than being one of the first large scale studies in the Indian context, our work contributes towards creating a standardized methodology to assess the ideological stance of a news-source, and its alignment with the social media discourse of its follower community. We find that the Indian mass media exhibits bias towards certain aspects or topics related to policy events. It also provides a significantly high coverage to aspects concerning the middle class and to political statements, neglecting the aspects directly relevant to the poor. Additionally, we find evidence of bias also in the representation provided to different political parties in the media. Social media seems to echo these biases rather than mitigate them. The tools and methods developed in this work can be useful for media watchdog institutions to call out biases in the media, and advocate for more complete coverage of issues across different news sources.

COMPASS '22, June 29-July 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9347-8/22/06…$15.00
https://doi.org/10.1145/3530190.3534798

CCS CONCEPTS

• **Information systems → Information systems applications**; • **Social and professional topics → Computing / technology policy**; • **Applied computing → Computers in other domains**.

KEYWORDS

Media bias analysis, Mass media bias, social media bias, social media analysis, content analysis

ACM Reference Format:
Anirban Sen, Debanjan Ghatak, Sanket Dhakate, Kumari Rekha, Mehak Gupta, Gurjeet Singh Khanuja, Kartikeya Sharma, and Aaditeshwar Seth. 2022. Analysis of Media Bias in Policy Discourse in India. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '22), June 29-July 1, 2022, Seattle, WA, USA*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3530190.3534798

1 INTRODUCTION

News media is known to inform public opinion on different government policies. It shapes how people think about these policies, and even what specific aspects (topics of discussion) of the policies people think about [31]. This influences the citizens' acceptability towards a policy. A biased media often prevents citizens from forming a balanced viewpoint related to government policies, which can skew voting decisions and impact public well being. Thus, analysis of media bias is an important area of study. Through this work, we intend to study the bias in the Indian news media.

Biases in media can take different forms such as coverage bias on how much attention is given to a policy or to its different aspects, selection bias on the amount of coverage given to different people or political parties, and sentiment bias on how positively or negatively different aspects and entities are represented in media [50]. Such biases inherently arise due to difference in ideologies, political affiliations, and commercial models of these outlets [18, 27, 43].

These biases influence the priorities that the readers place on various policies and their aspects. The biases in mass media, along with the impact they have on the prioritization of issues by the readers, is called *agenda setting*. To study this effect, we analyze different forms of biases in representation of the policy issues in Indian mass media, and compare it with the social media content on these policy issues to determine the concordance between these two spheres of expression of public opinion. We also study the effect of *framing* [12] in Indian mass media, by seeing how the newspapers talk about the policy from the perspective of five different constituencies of the *poor, middle class, informal sector, corporate, and government*. We use data from seven highly popular Indian news-sources in terms of circulation in this study, namely The Times of India (TOI), The Hindu (Hindu) Hindustan Times (HT), The Indian Express (IE), The Telegraph (TeleG), The New Indian Express (NIE), and Deccan Herald (DecH).

We examine the discourse on four economic policies in Indian mass media, and build a technological framework to quantify the bias existent in it. The policies considered in this study are *Demonetization* [32], *Aadhaar* [2], *GST* [14], and *Farmers' Protest* [29], each of which is an actively debated policy issue. We describe them briefly here: **(a) Demonetization**: A policy event where the government on 8 November, 2016 banned all 500 INR and 1000 INR banknotes with the motive of curtailing the use of illicit and counterfeit cash used to fund illegal activity and terrorism. The move was widely criticized owing to multiple problems caused to common people due to sudden depletion of liquidity, irregularities in norms of exchanging old currency notes, cash exhaustion in ATMs, etc. **(b) Aadhaar**: An initiative by the government to give every Indian resident a biometric-based unique identification number. The issue has been criticized owing to lack of security and privacy in citizens' data collection and storage mechanisms, and also because of an allegedly faulty implementation of the platform or use of the platform by different agencies. **(c) Farmers' protest**: A series of protests by farmers in India including the ones at Madhya Pradesh (Mandsaur protest) and Maharashtra (Kisan long march) demanding better prices for production of crops, loan waivers, and forest rights, among others. The issue is highly active politically with significant involvement of different politicians and political parties. **(d) Goods and Services Tax**: An indirect tax levied in India on the sale of goods and services at each step of the production value-chain with an effort towards formalization in the industry and simplification of multiple types of taxes which preceded the GST regime. Since its implementation there have been intense debates though on its complexity and problems in implementation which have impacted the overall growth of the economy.

We apply a mix of computational and qualitative analysis techniques to analyze the coverage, sentiment, and selection bias existent in mass media along three axes, namely ideological affiliation, political affiliation, and audience affiliation of news-sources. The first axis of ideological affiliation tells us the aspects or topics about policies that are dominantly covered by news-sources, and the different frames through which their articles are presented. With respect to this axis, we report the following research questions and findings: *(RQ-1a) Are news sources biased on the amount of coverage they give to different aspects about the policy?* Our analysis suggests that the mass media is significantly biased in terms of the coverage they provide to various aspects corresponding to each policy. *(RQ-1b) Do news-sources have a bias towards or against frames like pro/anti poor, pro/anti middle class, pro/anti government, pro/anti informal sector, and pro/anti corporate?* We do find evidence of mass media's bias towards the different frames that represent immediate concerns of different sections of people. Except *Farmers' Protests*, the news-sources show a conspicuous lack of coverage to the immediate aspects of the poor. On the other hand, all the other policies exhibit preferential coverage of political statements and aspects immediately impacting the middle class. A likely factor contributing to this may be the primary audience of the mass media, which does not include the poor, due to issues of literacy and digital access [26]. Our findings of the lack of coverage provided to aspects related to the poor are also corroborated by other reports [7].

The second axis of political affiliation tells us if the news-sources mostly prefer to write for or against a political party, and along this axis, we address the following research question: *(RQ-2) Is the mass media biased towards one of the two major political parties?* We find that the news-sources vary in the biases shown towards the two largest political parties in India, with respect to the coverage provided to the statements by and about the politicians belonging to them. These biases are also corroborated by the commonly believed political affiliations of these sources.

Finally, the last axis of audience affiliation tells us how the readers or followers of news-sources present their content on social media, if they propagate similar biases on social media, and exhibit diversity in news consumption. Along this axis, we report the following research questions and findings: *(RQ-3a) Are some news-sources more closely aligned with their readers (followers on social media) than others?* The social media readership is seen to be strongly aligned to their favorite media houses in terms of the aspects that they post about on the four policies. (RQ-3b) *Do people tend to follow multiple news sources or just a single source on social media?* A significant number of readers are found to however diversify their news consumption in terms of the number of news-sources followed. (RQ-3c) *Among those who follow multiple news sources, do they follow those which have similar ideological or political affiliations, or different ones?* Readers are also found to diversify their news consumption in terms of the political and ideological affiliations of these sources. Thus, while social media tends to lead to echo chambers [8, 46], we find that a significant fraction of users do follow news-sources with diverse political and ideological affiliations. The answers to these research questions provide strong hints towards the agenda setting and framing effects exercised by the Indian mass media that might lead towards influencing public opinion on key government policies. Our findings are indicative of biases in mass media and social media, and are consistent across the four economic policies considered.

Our key technical contribution is a computational framework that can analyze different forms of media bias for any news event using large scale web data. The framework is generalizable and can be used to perform a similar analysis for any event, by adapting it to use data obtained from any web based source. Our bias analysis framework can serve to provide indicators to algorithm developers of search engines and content recommendation platforms, about the biases embedded in web based content. It can also aid news readers to ensure diversity in news consumption by informing them about the biases inherent in multiple news-sources.

2 RELATED WORK

We divide studies related to media bias into three parts based on the three axes of bias analysis presented: (a) Ideological bias in the mass media, (b) Political bias in mass media, and (c) Diversity of news consumption in web and social media.

Ideological Bias in Mass Media: Journalists and news-sources shape public opinion by intentionally or inadvertently creating bias in their selection, writing, and distribution of news content, thus being called *gatekeepers* [47]. Scheufele et al. [54] discuss the concepts of agenda setting, framing, and priming in mass media that news sources can impose through their editorial gatekeeping processes. These three effects together play a significant role in influencing public opinion on socio-political aspects. Some examples of studying such ideological biases are as follows.

Some papers find evidence of selection and coverage biases towards events or news stories redleading to ideological biases. Saez-Trumper et al. [51] study selection and coverage biases for 80 news-sources and their social media communities, and establish that these biases reflect the regional priorities of the communities, rather than their political inclination. The problem of countering such biases through consumption of news from diverse sources is addressed by Bourgeois et al. [9], who compare selection of events across news-sources, and recommend a set of sources to ensure a diverse coverage of events.

Some other works find evidence of framing biases in mass media. Mudliar and Pal [41] study the coverage of the Indian indigenous, low cost tablet *Akash* in both Indian and international news-sources, and find that the reporting chiefly follows four technologically deterministic frames of presentation. In one of our previous studies [56], we find evidence of framing bias in the coverage of ICTD policies, most of which were found to be presented through a technologically deterministic frame. News-sources frame stories so as to appeal to their readers, which can also lead to ideological biases. Papacharissi et al. [44] study various frames through which terrorist attacks are reported in prominent US and UK newspapers, and how these frames vary between the two geographies. They find that other than differences in media ideology and journalism standards between the two geographies, audience attention is one of the major factors in deciding these frames. Semetko et al. [55] use content analysis of newspaper and television stories related to European politics to identify the dominant frames in them. They find that news-sources vary in terms of the dominant frames to attract audience, depending on whether they are serious or sensationalist outlets.

Our work builds on these studies by not only analyzing the coverage, selection, and statement biases in Indian mass media, but also combining them to infer the overall ideological alignment of prominent news outlets. The ideological axes help us understand the different constituencies of people the media writes about, and the way they are represented with respect to some nationally prominent economic policy issues. The proposed framework uses a mixed method approach on large scale news data for this purpose.

Political Bias in Mass Media: While several previous works on media bias consider political bias of media outlets as part of their ideological bias, we consider political bias to be a separate axes. There exist several studies on media bias that study political biases of media outlets. Chiang et al. [15] bring out evidence of endorsements provided to political candidates by mass media in the USA. Gentzkow et al. [25] similarly develop an index to define a measure of media slant by analyzing key phrases in news content specific to political ideologies. Munson et al. [48] assign a political bias score to each media outlet based on whether liberal or conservative candidates are over or under represented in these outlets. Budak et al. [11] use crowd-sourcing and machine learning techniques, and observe that the US media exhibits political biases by over-criticizing certain parties than others.

On similar lines, we develop a computational approach to study political affiliations of mass media. Given the objective nature of reporting of articles in the mainstream media houses considered, it is difficult to understand the political affiliation of news-sources at the article level. Thus, using a combination of dependency parsing (an NLP technique to identify grammatical dependencies in a sentence) and a recursive neural network (ReNN) based deep learning approach, we identify the political stance of the source at the statement level. We further aggregate them to obtain the political affiliation of the news-sources towards the two major political parties in India.

There have been several studies in the NLP domain on automatic ideology detection. Sim et al. [61] proposed a Hidden Markov Model (HMM) based model to understand the evolution of ideological rhetoric used by politicians during election campaigning, by inferring mixtures of ideological positions in documents. However, the model ignores intra-sentential contextual influences. Some other approaches [1, 33, 42] use topic models at the document level to analyze bias in news, blogs, and political speeches. Recent works use attention based models for ideology detection. Gangula et al. [22] detect political bias in news articles by using news headline attention using attention based mechanism alongside bidirectional LSTMs. Sanchez et al. [52] use attention based models like BERT, XLM-RoBERTa, and M-BERT to detect hyperpartisan news, using two different text masking techniques.

The advantage of using ReNNs for ideology detection is that they are capable of detecting bias polarity switches at higher levels in parse trees through phrase-level annotations. These phrase annotations allow ReNNs to to detect bias in complex sentences by capturing intra-sentential contextual influences. Additionally, attention based models require a significant amount of ground truth data and computational cost to perform well. On the contrary, ReNNs can leverage phrase level annotations to reduce the volume of annotated data required in training.

Diversity of News Consumption in Web and Social Media: Social media sites like Twitter have enabled people to easily obtain news from multiple sources. An et al. [4] find that the follower network of users in social media enable them to receive news from multiple media outlets on the same topic, and on multiple topics through their connections to journalists. Scharkow et al. [53] study the web browsing histories of two large sets of users, and establish that users are subjected to a significant variety in news consumption on social media. However, while these studies focus on the number of news outlets followed by users, they do not consider diversity in terms of the dominant political or ideological alignment of news content consumed by users. In this direction, some studies focus on studying the diversity in political alignment of the content presented to readers through web based or social media data. Fletcher and Nielsen [21] use data from *2017 Reuters Institute Digital News Report* to show that contrary to conventional wisdom, social media seems to add more diversity in terms of political affiliation, to the news consumed by users. They conduct a survey and divide the respondents into three groups of news users (those using social media for news), incidental users (those using social media for other purposes, but getting exposed to news on social media incidentally), and non-users (those who don't use social media). The authors find that compared to the non-users, not only do the incidental users get exposed to more news-sources on social media, but also tend to follow more sources from both left and right leaning ideological spectrum. Garrett [24] studies the users of two partisan online news sites, and claims that there is no evidence that readers abandon news with opposing political affiliations. Messing and Westwood [36] show that social media increases the probability that readers select news content with diverse political affiliations, reducing their chance of getting confined to partisan echo chambers.

However, there also exists a large body of work that contradict the claim that there exists significant diversity in the political alignment of web based news content. Bakshy et al. [6] study how Facebook users interact with socially shared news, and find that individual political affiliations play a strong role in limiting exposure to cross-cutting content. Dahlgren et al. [17] conduct a longitudinal study over a span of two years and find that in both print and online news media, people increasingly seek out news that is consistent with their political ideology. As argued by Feldman [19], political ideology is a highly stable characteristic in an individual, when considering the adult population. Our findings reveal that for the four policies considered, a significant number of followers of news-sources do tend to diversify both in terms of political and ideological affiliations of the sources they follow.

The studies mentioned in this section primarily focus on diversity in terms of political alignment of news content, we study diversity in terms of both political and ideological affiliations of news-sources. While political affiliation represents the preference of the source towards the two major ruling and opposition parties, ideological affiliation refers to the dominant frames through which these sources present their content.

3 METHODOLOGY

We developed a technological framework to analyze media bias along the three axes. The architecture is shown in Figure 1. We study biases present in the news presented by seven highly popular news-sources. These sources have been considered since all of them are national dailies in English, and the policy events studied in this work are nationally popular events with widespread coverage. The selected sources also provide us with a healthy mix of commonly believed political affiliations, which aids us in analyzing the biases across the political spectrum.

First, to understand the ideological affiliation of these sources, we see which aspects of a policy they cover dominantly, and the frames through which they are presented. We extract aspects from news articles for a policy using an automated method, calculate the coverage provided to them, and measure the sentiment slant of the majority articles belonging to them. Next, we map these aspects to the dominant frames using qualitative analysis (coding schema). The results are finally aggregated for each news-source to calculate its frame alignment.

Second, to find the political affiliation of news-sources, we extract statements about political parties from news articles, and calculate their pro/anti political stance at the statement level, unlike the article level sentiment classification used to understand the ideological affiliation of news-sources. We aggregate these political stance scores for each news-source, to calculate its political affiliation.

Finally, we find the audience affiliation of news-sources by extracting tweets posted by their followers that contain article URLs, and map these tweets to the aspects identified. We compare the aspect coverage of tweets with the news-sources' aspect coverage, helping us calculate the alignment of a news-source with its social media followers. We now elaborate each of these components in detail.

3.1 Ideological Affiliation: Aspect Extraction using LDA

To understand the ideological affiliation of news-sources, we study the various topics of discussion or *aspects* related to the policies covered by the news-sources, and the coverage they provide to these aspects. This helps us understand on which topics a source prefers to discuss more compared to others, given a policy event. A news article can belong to multiple aspects simultaneously, depending on the issues it discusses. Since we do not have these aspects pre-identified, we use an unsupervised method named Latent Dirichlet Allocation (LDA) to identify them. LDA is a commonly used unsupervised statistical modeling method that maps a set of documents to unobserved topics, helping cluster similar documents into topic clusters that can be manually examined and labeled. Here, the documents refer to media articles, which are mapped to different topic clusters, or *aspects*. Our approach in this direction is similar to the work by Yigit et al. [67] where the authors use LDA to cluster news events to various aspects using both news articles on the events and the user comments on them.

A problem with unsupervised methods like LDA is to identify the optimal number of clusters (or aspects) to be supplied as input to the method. Since the number of aspects is not known beforehand in our case, we had to evaluate the clustering performance by experimenting with a different number of possible aspects. For this purpose, we used the best performing topic coherence measure as suggested in the paper by Roder et al. [49], in conjunction with the PyLDAVis package [60], to visualize and infer the optimal number of clusters to be specified for each policy. Since LDA maps documents to topics probabilistically, in our case, an article is mapped to clusters if the probability of its belongingness as indicated by LDA is greater than 0.3 (experimentally determined). We further merged some of the resultant clusters together based on manual analysis of their articles, and obtained 16 aspects for *Demonetization*, 14 aspects for *Farmers' Protest*, 11 aspects for *GST*, and 17 aspects for *Aadhar*. We also named these clusters after going through this manual analysis.

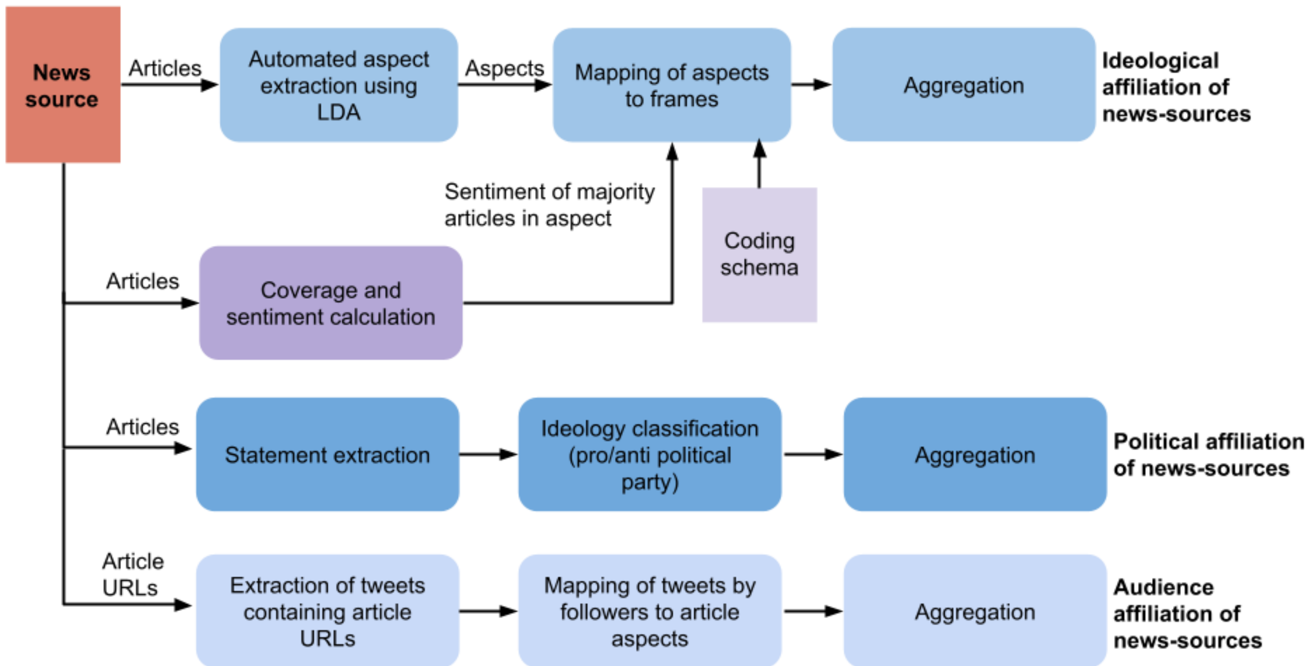


Figure 1: Framework to analyze media bias: Mass media bias is analyzed based on the three axes of ideological, political, and audience affiliation

To quantitatively see if the resultant clusters are accurate, we measure the aspect mapping accuracy, in which three annotators manually studied around 800 articles in total across the four policies. For each aspect in an event, the annotators went through majority articles in it to name the aspect unanimously. Next, they went through each article in the aspect to see if the article actually belonged to it. This exercise had an inter-coder agreement (measured using Cohen's Kappa) of $k >= 0.8$ for all events. We considered majority decision to determine the belongingness of an article to an aspect. Finally, the proportion of articles that actually belong to the aspects was calculated for each policy to evaluate the mapping accuracy, which were 85% for Demonetization, 96% for Aadhaar, 81% for GST and 76% for Farmers' Protest. Minimal manual analysis is required to evaluate the aspect mapping accuracy, since we start with LDA identified aspects, and evaluate a small subset of articles from them. We also mapped the tweets to the mass media aspects by studying the article URLs that they contain. First, tweets on a policy event are identified with the help of the media URLs. Next, these tweets are mapped to the aspects to which the articles in their URLs belong. Since tweets are concise and sometimes even grammatically incorrect, tweets without URLs are difficult to map to media aspects.

3.2 Ideological Affiliation: Article Level Sentiment Analysis

Alongside aspect coverage, we also need to see how news-sources write about the constituencies or sections of people, to understand their ideological affiliation. For this purpose, the sentiment slant of articles must be obtained. We use the Sentistrength tool for sentiment analysis of articles presented in the news-sources considered. This tool uses a lexicon based approach for sentiment analysis, and has often been used for sentiment analysis in political text [13]. While state of the art sentiment analysis methods use neural models like BERT for sentiment analysis, we used Sentistrength for this study since we do not have sufficient sentiment annotated data at the article level, and also since the tool provided satisfactory performance on the data collected. Sentistrength [65] reports TPOS (positivity) and TNEG (negativity) scores for each article. TPOS score is in the range of 1 (not positive) to 5 (extremely positive), and TNEG score is in the range of -1 (not negative) to -5 (extremely negative). The aggregate sentiment for an article is calculated as the sum of TPOS and TNEG. To evaluate the accuracy of Sentistrength in reflecting the article sentiments, we assigned three annotators the task of annotating the sentiment polarity of 200 articles from each policy to form the ground truth, using the labels *positive, negative*, and *neutral*. We discretized the Sentistrength scores to three polarity levels: positive ((1.5,5]), neutral ([-1.5,1.5]), and negative ([-5,-1.5)) based on manual analysis of 50 articles. The annotated polarities were then cross checked with the Sentistrength polarity levels for these 800 articles. The agreement was found to be 84% for Demonetization, 76% for Aadhaar, 60% for GST and 84% for Farmers' Protest, indicative of a decent performance by Sentistrength.The evaluation of accuracy was done on content that was nearly balanced across the categories of positive and negative.

3.3 Ideological Affiliation: Article Level Framing Analysis

Framing is defined as "the selection of a restricted number of thematically related attributes for inclusion on the media agenda when a particular object is discussed" [35]. In other words, framing refers to the way the media houses shape information to help the audience process it. One of the ways in which news-sources engage in framing is by orienting the news content towards or against specific constituencies. We analyze the ideological affiliation of news-sources by studying their representation of various frames through which their articles are presented. This is done by first identifying some dominant constituencies on which media articles on the four economic policies are written.

We identify five dominant constituencies that Indian media houses mostly cover for the policies considered. News around Constituency #1 [poor] consists of aspects about marginalization of the poor, and their necessary concerns. News about Constituency #2 [middle class] consists of aspects around lifestyle of the middle class citizens. News about Constituency #3 [informal sector and small trades] consists of aspects related to the impact of policies on the informal sector and small traders. Under Constituency #4 [corporate], news-sources discuss the role of industries in development, and how tax breaks and benefits for corporations affect the society on a whole. Finally, for Constituency #5 [government], news-sources discuss rational government policies for growth and development, and approval or criticism of government interventions. Three annotators went through a set of 100 articles from each policy event, and agreed on these five constituencies as they dominantly represent the audience or the people towards whom the mass media presents information about the events, although there can be many other constituencies that are covered by mass media pertaining to other policies. The presence of these constituencies in the Indian media narrative of these policies is also corroborated by earlier papers [16, 20, 58]. News-sources may present policy events through different frames, by dominantly supporting or opposing the cause of these constituencies in their articles. Thus, while *corporate* is a constituency, news-sources may present policy news through the *pro-corporate* or *anti-corporate* frames.

For each policy event, we map each aspect to frames based on whether the aspect supports or opposes or is not applicable to the cause of the constituencies. For example, for the *Demonetization* policy, the aspect on *Queues at banks and ATMs* is classified as pro-middle class because most articles on this aspect were negatively writing about the problems caused to the common people in getting cash at ATMs, and thus helping the middle class by drawing attention to their concerns. The same aspect is classified as anti-government because negative articles on these aspects generally criticize the government's apathy and lack of foresightedness in handling the issue. These aspect to frame mappings are then aggregated to obtain the dominant frames through which a news-source presents the constituencies, which tells us whether the way most articles are written in a news-source support or oppose the cause of a constituency.

We use qualitative analysis to map the aspects of a policy to one or more frames. This provides us information on whether majority articles in the aspect speak for or against the constituency. We developed a coding schema for each policy separately, for this purpose. The coding schema is a guide that helps the annotators to map policy aspects to one or more frames. The schema contains the normative definition of the constituencies, and example articles written on them. A snapshot of the coding schema for the policy Demonetization is presented in the appendix.

Using this schema, three annotators performed aspect to frame mapping for all policies, each of whom went through around 3000 articles in total (50 articles from each aspect for each policy event) for this exercise. The aspect to frame mapping is done in three steps: (a) We first find out the majority sentiment slant m of the articles in an aspect a (+1 indicates a majority positive or neutral sentiment, and -1 represents a negative majority), (b) A group of annotators manually determine if the majority articles of the aspect support or oppose the cause of a constituency c (*stance(a,c)*). We divide *stance(a,c)* by the majority sentiment (+1/-1), to get the alignment score (U) of the aspect w.r.t. the constituency. U can vary between -1, 0, and +1 for each (aspect, constituency) pair. These scores are presented in the appendix. (c) The alignment scores (U) are finally multiplied with the aggregate sentiment offset of the aspect (average of sentiment offset from mean sentiment score of articles belonging to the aspect) to obtain the aspect to frame mapping. These aspect to frame mappings are finally aggregated to obtain the ideological positioning of news-sources w.r.t. the constituencies, or the frames through which they present policies. This step is detailed in the *Results* section.

We evaluated the inter-coder percentage agreement for step (b), i.e., determining the stance of majority articles in an aspect towards a constituency, using the percentage agreement calculation method as described in [63]. The initial mapping exercise had an inter-coder agreement of 61.33% for Demonetization,76% for Aadhaar, 71% for GST, 74.3% for Farmers' Protests. We ran another round of moderation and due deliberation before coming up with the final mappings after this exercise. Our method of building the coding schema after rigorously going through the news articles by multiple annotators ensures that there is minimum bias and subjectivity in the mapping of aspects to frames. This process of mapping aspects to frames can be extended to any dataset – although the size of the entire data analyzed in this work is of the order of millions, the aspect to frame mapping has been done manually by studying just 3000 articles, i.e., randomly selecting 50 articles from each aspect for each event (there are around 15 aspects for each event). To build the coding schema, the annotators studied only 100 articles from each policy event (400 articles in total), which is a manageable number for qualitative analysis.

3.4 Political Affiliation: Statement Level Ideology Classification

Obtaining ideological affiliation of news sources, as described above, was done at an article level. We choose to determine the political affiliation by examining content at the sentence level, since a sentence level approach provides us a fine-grained understanding of a news-source's stance on a political party, which is difficult to obtain at article level.

We study the representation provided by the news-sources to the two major political parties in India with respect to these policies, by calculating the stance of the statements made by the news-sources about these political parties. The four policies considered in this study were either initiated or accelerated by BJP (the currently ruling party), while INC was in opposition. For this purpose, we first extracted sentences from articles where at least one of these parties was mentioned. Next, we built a dependency parse tree of each of these sentences using Stanford CoreNLP [34] to obtain the parts-of-speech (POS) tags. Rules based on POS tags and the dependency tree tags can be used to identify statements made about a party. If the statement is about a political party, the party name appears as an object (tags *dobj*/*pobj*) in the parse tree of the sentence. As reported in prior work [59], the accuracy of this method is reported to be more than 85% for the four policies considered.

The statements about political parties are next classified into *pro-party* and *anti-party*, using a Recursive Neural Network (ReNN) based ideology classifier developed by Sharma et al. [30]. While we use sentiment slant of media articles to analyze framing, we use the ReNN based ideology classifier for statements, since statements are short in length, and their sentiment slants often do not reflect their ideological stance towards or against a party. This occurs especially when they contain sarcasm, or complex phrases. The ReNN based classifier helps us capture such linguistic cues significantly better than sentiment analysis. The classifier provided an accuracy of 82% for the task of ideology classification, on a test set of 250 statements manually annotated for the four policies by the three annotators.

3.5 Audience Affiliation: Aspect Classification of Tweets

To understand the audience affiliation of mass media, we collect tweets related to the policy events using keywords pertinent to the policies (Appendix), and identify those made by the followers of the news-sources. Next, we select a subset of tweets that contain at least one article URL belonging to the set of policy articles collected from mass media. These tweets are then mapped to the aspects to which the article carried by them belongs. We study the coverage provided by these tweets to the various aspects corresponding to policies, to understand the audience affiliation of mass media houses.

4 IDEOLOGICAL AFFILIATION OF MASS MEDIA

This section covers our analysis of ideological affiliation of news-sources, in terms of their coverage bias towards various aspects and dominant frames of news presentation.

4.1 Aspect Coverage Bias

Data and Method: We attempt to answer the research question: *Are news sources biased on the amount of coverage they give to different aspects about the policies?* The question relates to the *agenda setting* effect of mass media as reported in literature [54]. Agenda setting is the idea that there exists a correlation between the emphasis placed by mass media on certain aspects, and the importance attributed by the readers to them. We study the emphasis that mass media places on various aspects in terms of the relative coverage given to these aspects.

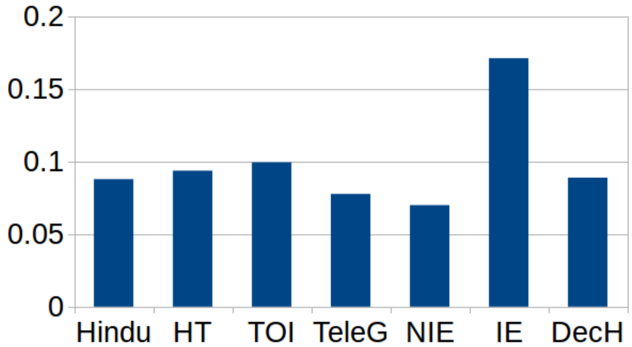
We extract news articles corresponding to the four policies using a keyword based approach. Non-OpEd news articles belonging to categories like National, International, Regional, Sports, and Business are collected, along with the URLs and their meta-data. We first supply a set of manually selected keywords corresponding to each policy. After extracting articles containing these keywords, the keyword set is augmented with newer, relevant keywords (top 20% scored on TF-IDF) from these articles. These two steps are repeated until the keyword set becomes static, and the final set of articles is used to perform our analysis. The final set of augmented keywords for each policy is shown in the Appendix. We finally use 22302, 13908, 22179, and 85486 articles respectively for Demonetization (Nov 2016 to Oct 2019), Aadhaar (2011 to 2019), GST (2011 to 2019), and Farmers' Protest (Nov 2016 to Oct 2019).

Relative aspect coverage is defined as the proportion of words in articles belonging to an aspect with respect to the total number of words across all aspects for a news-source[1].

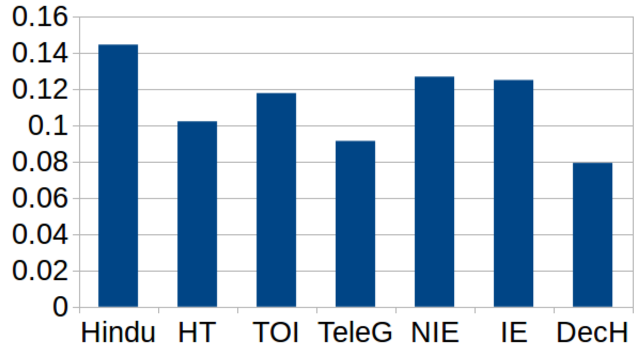
Analysis and Results: The aggregate coverage distribution of aspects (plots reported in the appendix) shows that there is bias in aspect coverage across news-sources. While the top aspect gets a coverage of more than 10% for most policies, the least covered aspect has an insignificant coverage. We observe this trend even when we take a look at the highest covered aspects for each event, on a per-news-source basis. For all policies considered, the top five aspects covered by every news-source remain more or less consistent.

To also observe if there exists a significant difference in the treatment of policy aspects by different news-sources, for each source, we create a vector of the aspect coverage percentage for each policy. We then obtain the mean aspect coverage percentage for the policy across all sources, and calculate the Euclidean distance between the two distributions. Lower the distance, closer is the news-source is to the mean aspect coverage distribution. The Euclidean distances for each source is shown in Figure 2. The highest deviation from mean aspect coverage is mostly shown by IE and Hindu, both of which are commonly believed to be pro-opposition news-sources [37] for the timelines considered. Apart from IE and Hindu, most sources lie significantly close to the mean aspect coverage trend, except in Aadhaar where a diversity of coverage seems to prevail across the different news-sources. These results are also corroborated by the Jensen-Shannon Divergence (JSD) between the aspect coverage and the mean aspect coverage distributions for each source. Considering all news-sources for each policy, we find the maximum values of JSD as 0.031 for Demonetization (IE), 0.04 for Aadhaar

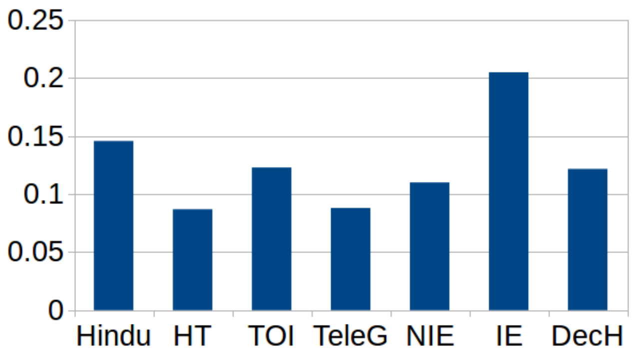
[1]Thus, we normalize the count of words per aspect, which handles the case of different news-sources containing different length of articles (and aspects).



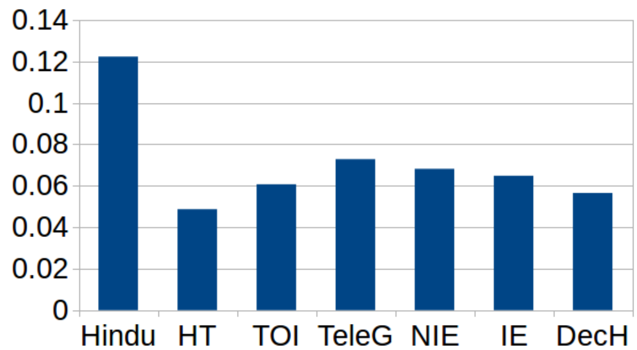
(a) Demonetization



(b) Aadhaar



(c) GST



(d) Farmers Protest

Figure 2: [RQ1a] Euclidean distance of relative coverage and mean relative coverage for the four policy events.

(Hindu), 0.054 for GST (IE), and 0.052 (Hindu). The significantly low values of JSD show the conformance of the sources with the mean trend, and also the maximum deviation exhibited by either IE or Hindu across policies. We also find the Kolmogorov-Smirnov 2-sample statistic for each news-source for each event, to see how significant the difference is between the news-source's coverage of aspects and the mean relative coverage. We find that the p-values lie in the range of [0.58,0.99] for Demonetization, [0.19,0.99] for Aadhaar, [0.74,0.99] for GST, and [0.26,0.86] for Farmers' Protest. The high p-values indicate that the difference between the relative coverage of aspects and the mean relative coverage is insignificant for all news-sources, indicating news-sources' tendency to follow the mean trend of aspect coverage. Thus, we conclude that while most news-sources seem to closely follow the mean aspect coverage distribution, there exist differences in the coverage they individually provide to each aspect.

4.2 Framing Bias

Data and Method: Here, we answer the research question: *Do news-sources have a bias towards or against frames like pro/anti poor, pro/anti middle class, pro/anti government, pro/anti informal sector, and pro/anti corporate?* We try to analyze the effect of *framing* in mass media [54]. We analyze this effect by automatically extracting aspects from the news articles using LDA, and then manually linking the aspects with one or more frames as described in section 3.3. To map a news-source to a frame, we first calculate the sentiment offset of each aspect it covers (equation 2), by aggregating the weighted sentiment offset of the articles present in the aspect. Next, these sentiment offsets are aggregated across aspects for the news-source, by weighing them with the aspect's alignment towards constituencies (equation 3 below). Thus, we aggregate the aspect to frame mappings to obtain the news-source to frame mapping. We calculate the (news-source,frame) alignment matrix M using the following equations:

$$C_{ian} = \frac{c_i}{\sum_{j\epsilon(n,a)} c_j} \tag{1}$$

$$S_{an} = \sum_{i\epsilon(n,a)} C_{ian} * (S_{ian} - S_{avg}(a)) \tag{2}$$

$$M(n,c) = \sum_{a\epsilon c} U[a,c] * S_{an} \tag{3}$$

where n represents a news-source, a an aspect, C_{ian} the relative coverage for the i^{th} article in news-source n belonging to aspect a, and S_{ian} the compound sentiment score of the i^{th} article for aspect a. $S_{avg}(a)$ is the average sentiment score of all articles in aspect a across all news-sources, and c is the constituency. Here, $(S_{ian} - S_{avg}(a))$ is the offset of the sentiment of the i^{th} article from the mean sentiment. $U[a,c]$ is the (aspect,constituency) alignment value $\epsilon[-1,0,+1]$ (calculation of U has been explained in section 3.3). Thus, the matrix M tells us the different frames through which the news-sources present their content by supporting or opposing the cause of a constituency.

Analysis and Results: To verify if there exists variations in terms of frame alignment of the sources, we perform a Principal Component Analysis (PCA) on the 5-dimensional mean frame vector (mean of the 5-dimensional frame vectors $M(n,)$ across all policies) for each news-source, for the four events. Performing PCA on the 7*5 matrix (M) provides us with two matrices into which it is decomposed – a 7*2 news-source matrix, and a 2*5 constituency matrix. We plot both news-sources and constituencies in the same 2D vector space as seen in Figure 3. A factor analysis is then done to interpret the two principal components. We map the news-source vectors The first component PC1 (x-axis) represents the *pro-informal sector*, *pro-middle class* and *pro-poor* frames on the positive side, and the *anti-government* and the *anti-corporate* frames on the negative side. The second component PC2 (y-axis) represents the *anti-government*, *anti-informal sector* and *anti-corporate*, frames to the negative side. We observe that TeleG, a commonly believed leftist news-source, is most aligned to the frames *pro-poor*, the *pro-informal sector*, and the *pro-middle class*, which is as expected. On the other hand, TOI is seen to be an outlier, and is aligned to *pro-government* and *pro-corporate* frames. DecH, IE, HT, and NIE, being close to the origin, are relatively balanced news-sources. Hindu is aligned more towards the *anti-corporate* and *anti-government* frames. We compare the mean relative coverage provided to constituencies by mass media across news-sources (plots in the appendix), and find that the coverage is consistently higher for the *middle class* (more than 50% coverage for Demonetization, Aadhaar, and GST) and *government* (more than 90% coverage) constituencies, when compared to the *poor* (less than 50% coverage for Demonetization, Aadhaar, and GST). Thus, in terms of coverage of aspects, we find that mass media in general provides less coverage to aspects related to the poor, and more coverage to middle class and political aspects. Our analysis thus indicates the presence of *framing* effects in Indian mass media. We find that (a) The news-sources are biased w.r.t. the frames through which they present their content, and (b) they provide consistently less coverage to aspects of the poor in general.

5 POLITICAL AFFILIATION OF MASS MEDIA

Data and Method: To study if the media houses are biased towards one of the two major political parties in India – the currently ruling Bharatiya Janta Party (BJP) and the opposition Indian National Congress (INC), we address the question: *Is the mass media biased towards one of the two major political parties?* To measure the alignment of the news-sources towards the two parties, we detect the ideological stance of statements made about them in the news articles. These statements are extracted, and classified into pro and anti BJP/INC classes by the ideology classifier described in the Methodology section. For each news-source, we aggregate the pro and anti ideology scores for the two parties, which provides us with its overall ideological stance.

Analysis and Results: We show the plot for aggregate political ideology scores for the news-sources in Figure 4. We report the aggregate political ideology scores for the (BJP,INC) slant for each news-source: Hindu (0.22,0.51), TOI (0.48,0.1), IE (-0.56,1), HT(-0.04,-0.25), TeleG (-0.56,-1.14), DecH (-0.56,-1.14), NIE (1,0.93). TOI and NIE seem to be strongly aligned towards BJP. Hindu and IE are pro-INC. HT comes out as a relatively neutral source, although it shows a slight anti-INC leaning. TeleG and DecH, while being more anti-INC, are against both parties. Thus, the Indian mass media indeed shows a bias in their political affiliation, based on the statements they cover about these parties.

6 AUDIENCE AFFILIATION OF MASS MEDIA

We try to understand the audience affiliation of news-sources from two perspectives, namely the alignment of sources with their follower community, and the political and ideological news preferences of these followers.

6.1 Alignment of Mass Media with the Audience

Data and Method: To answer the research question: *Are some news-sources more closely aligned with their readers (on social media) than others?* We analyze if the readers post more about those policy aspects that the mass media frequently covers. We consider the readership community of news-sources as the set of all followers of the news-source handles on Twitter (*TweetFol*). The number of tweets containing article URLs of the four events are 34521 for Aadhaar, 59489 for Demonetization, 38073 for GST, and 22820 for Farmers' Protest, which are used for analysis.

Analysis and Results: We take the followers of each source (even those who may be following other news sources in addition), and examine the URLs of the source that are tweeted by its followers. The euclidean distance between the coverage distributions of mass media and its social media followers (tweeted article URLs mapped to the aspects) is shown in Figure 5. The low values (less than 0.5 across all policies) indicate that the mass media and social media aspect coverages are significantly similar, although some media sources are closer than others to their followers in terms of aspect coverage. We also compute the Jensen-Shannon Divergence (JSD) between the distribution of a source's aspect coverage on a policy and that of its social media community.The JSD values for each news-source and policy are present in the appendix. We again find that the news-sources are significantly similar to their followers in terms of the aspect coverage distributions as is evident from the low values of JS divergence (ranging between 0.04 and 0.28).

To see if the differences in coverage are significant, we perform the Kolmogorov-Smirnov 2-sample test for each policy event, for mass media and social media coverage of aspects. We find that the p-values lie in the range of [0.30,0.89] for Demonetization, [0.19,0.99] for Aadhaar, [0.37,0.99] for GST, and [0.11,0.86] for Farmers' Protest. This shows that the coverage of aspects is significantly similar in mass media and social media, and that the most media houses align with their audience in terms of aspect coverage on policies.

6.2 Diversity in terms of Number of News-Sources Followed

Data and Method: The research question that we try to answer here is: *Do people tend to follow multiple news sources or just a single source on social media?* We measure the number of followers of

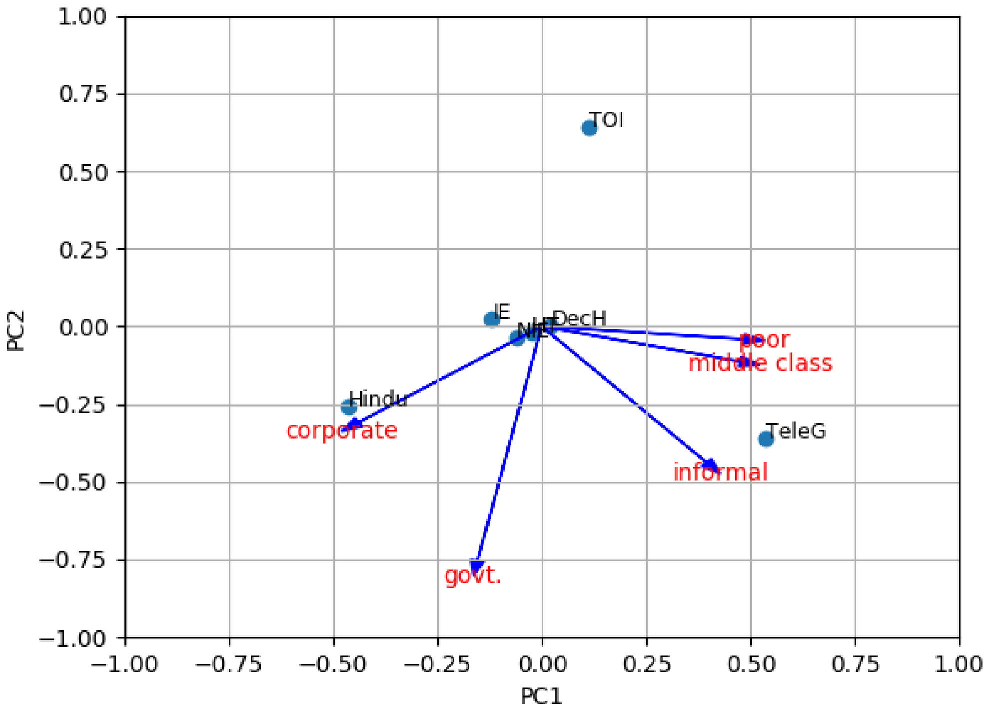


Figure 3: [RQ1b] PCA on frame vectors for the four events for all news-sources

each news-source, and the fraction of them who follow a single source (hereafter referred to as *dedicated followers*). We again use the *TweetFol* set for this analysis as described earlier.

Analysis and Results: We show in Table 1 the total number of followers and the fraction of dedicated followers for each news-source. We find that a majority of followers follow more than one news-source. TOI shows the highest dedicated follower count (48.5%), followed by NIE (30.5%) and DecH (29.5%). For all of the other sources, more than 80% of their followers follow multiple sources. This indicates that most social media followers are inclined to follow multiple sources on Twitter as also corroborated by previous work [53]. This may be because of the ease of following news-source handles on social media. However, it also indicates that news readers on social media have a tendency to diversify their news consumption in terms of the number of sources.

6.3 Diversity in terms of Political and Ideological Preferences

Data and Method: To observe the diversity in terms of the ideological and political preferences of the readers following multiple sources, we address the question: *Among those who follow multiple news sources, do they follow those which have similar ideological or political affiliations, or different ones?* Unlike many earlier studies, in this work we differentiate between political and ideological affiliations of sources, and consider them as independent axes altogether – political affiliation refers to the alignment towards or against a set of political parties (section 5), while ideological affiliation refers to the alignment towards one or more frames through which content is presented in media (section 4.2).

From our analysis in sections 4.2 and 5, we first categorize the news-sources on their ideological and political affiliations as reported in Table 2. Followers of these sources can also be categorized based on the sources that they follow. We observe the follower community overlap for pairs of sources, and based on the political and ideological affiliations of the sources in these pairs, categorize their followers. We calculate the follower community overlap between pairs of news-sources as $|F_i \cap F_j|/|F_i \cup F_j|$ where F_i and F_j are follower sets of news-sources i and j, respectively. Based on the affiliations of the news-sources, we can categorize their followers into four categories based on whether a given pair of news-sources followed by them has the same or different political affiliations, and same or different ideological affiliations. These categories are: **[A]** Followers following sources with similar ideological and political affiliations, [B] Followers following sources with similar political but different ideological affiliations, [C] Followers following sources with different and political but similar ideological affiliations, and [D] Followers following sources with different political and ideological affiliations. We first enumerate 21 source pairs constructed

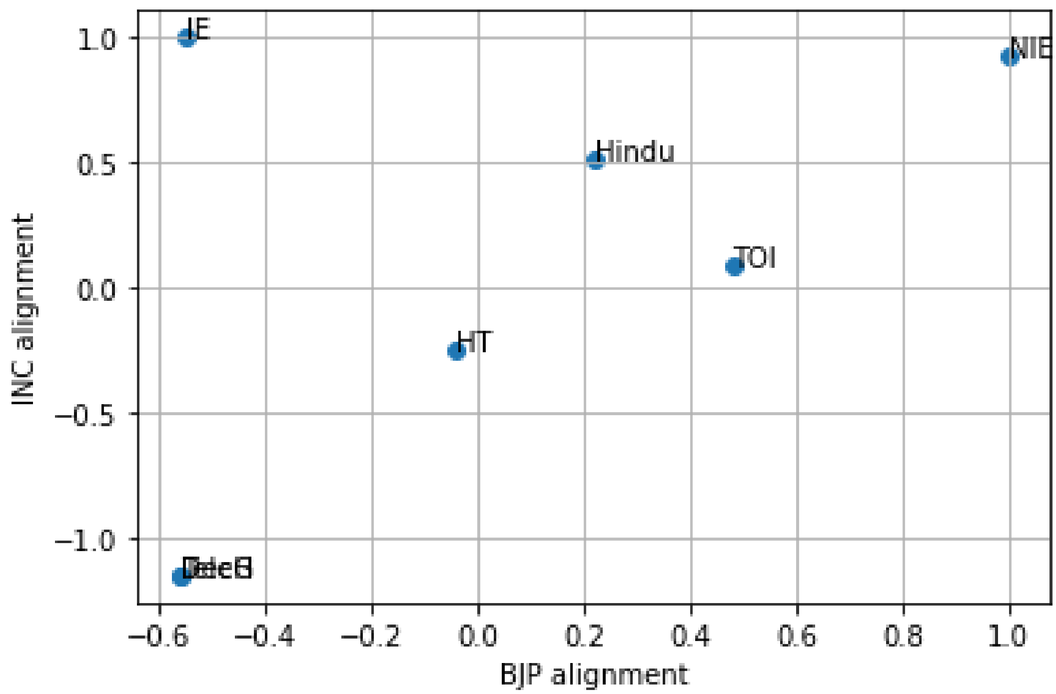


Figure 4: [RQ2] Aggregate political ideology scores of news-sources with respect to BJP and INC

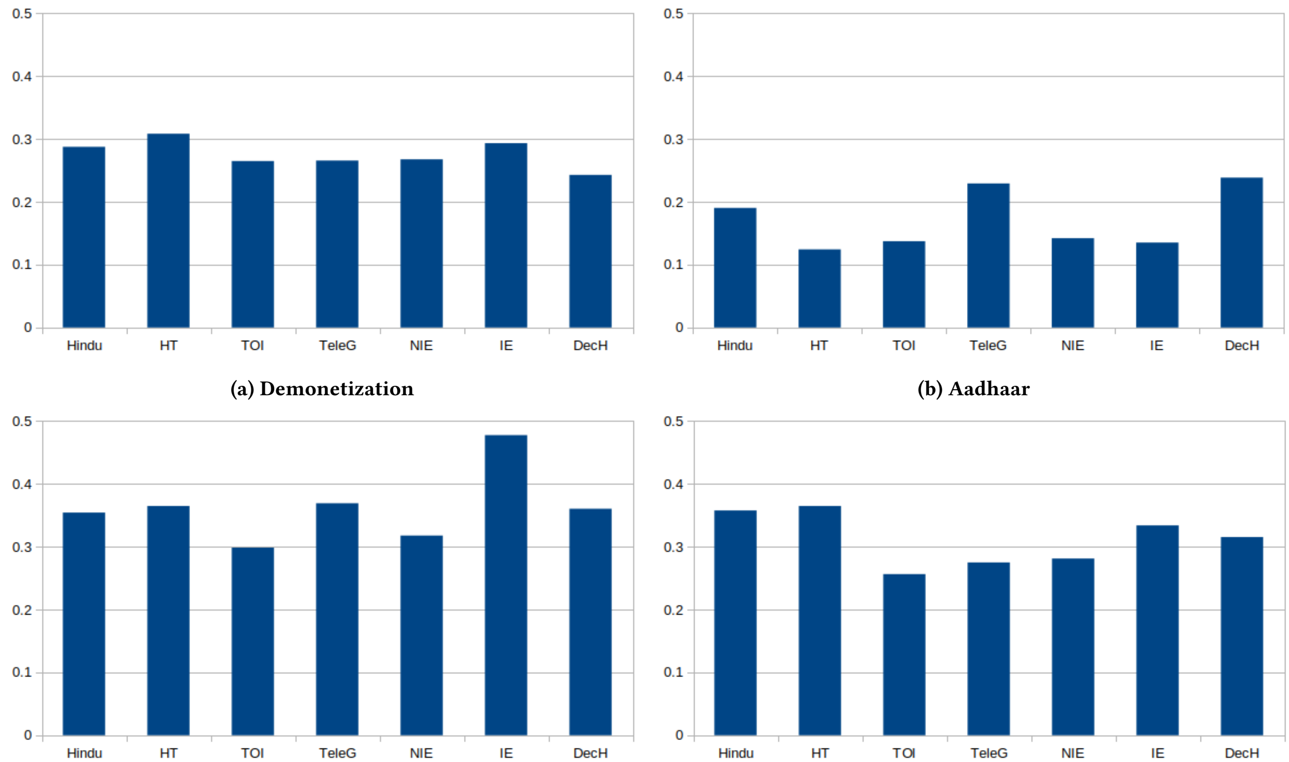
Source	# Total Followers	# Followers Following a Single Source
TOI	11026375	5346622 (48.5%)
HT	6299717	1134958 (18.0%)
HINDU	4842235	899340 (18.6%)
IE	2742133	258610 (9.4 %)
NIE	347149	105749 (30.5%)
TeleG	51885	10057 (19.4%)
DecH	24897	7338 (29.5%)

Table 1: [RQ3b] The total number of followers and the number of dedicated followers (and their percentage) for each source.

Source	Ideological Affiliation	Pol. Affiliation
TOI	pro-corporate, pro-government	pro-BJP
HT	Neutral	anti-INC
HINDU	anti-corporate, anti-government	pro-INC
IE	Neutral	pro-INC
NIE	Neutral	pro-BJP
TeleG	pro-poor, pro-middle class, pro-informal sector	anti-BJP, anti-INC
DecH	pro-poor, pro-middle class, pro-informal sector	anti-BJP, anti-INC

Table 2: Ideological and political affiliations of news sources as obtained from our analysis

from the seven news-sources, and then map each pair to the four categories. We show the percentage of followers following both sources in these pairs, given the unique followers for each pair in Table 3. We define a pair as a *dominant pair*, if the follower overlap percentage for the pair is more than 10% (decided based on the skew observed in each category and highlighted in blue in the table).



(a) Demonetization **(b) Aadhaar**

(c) GST **(d) Farmers Protest**

Figure 5: [RQ3a] Euclidean distance of relative coverage distributions between mass media and social media.

	Similar ideological	Different ideological
Similar political	**Category A** NIE,HT(163993,2.5%)	**Category B** TeleG,DecH(3123,4%) Hindu,IE(1874796,32.8%) TOI,HT(4538316,35.4%) TOI,NIE(178627,1.5%)
Different political	**Category C** DecH,NIE(20898,1.9%) DecH,IE(11300,0.4%) IE,NIE(161737,5.5%) DecH,HT(11460,0.1%) IE,HT(1910074,17.3%)	**Category D** TeleG,NIE(10449,2.6%) TeleG,IE(30712,1.1%) TeleG,Hindu(54233,0.6%) Hindu,DecH(12590,0.2%) Hindu,NIE(187217,3.7%) TeleG,HT(29626,0.4%) Hindu,HT(2887160,34.9%) DecH,TOI(13231,0.1%) TeleG,TOI(31976,0.2%) IE,TOI(2031058,17.3%) Hindu,TOI(3266012,25.9%)

Table 3: [RQ3c] Table showing the four categories of followers based on the affiliations of the news-sources they follow. The community size in terms of number of followers and percentage of total followers is indicated within braces.

Analysis and Results: We observe from Table 3 that the percentage overlap of followers, their absolute number, and the number of dominant pairs are all significantly higher for the category of followers following sources with different ideological affiliations (categories B and D), when compared to the category of followers following sources with similar ideological affiliations (categories A and C), indicating that followers prefer to follow ideologically diverse sources more than ideologically similar ones. A similar trend is found for political affiliation of sources as well (categories (C, D) and (A,B)). Thus, one of the reasons to follow multiple sources on social media may be to achieve ideological and political diversity in news consumption. From these two category groups, we also find that followers tend to diversify more in terms of ideological preferences than political preferences.

We also study if there exists any dependency between the political and ideological news preferences of followers. Table 3 shows that followers with similar ideological preferences do not vary much in terms of their political preferences (categories A and C), but the reverse is true (categories A and B). This observation may indicate that although some followers may have similar political preferences, they prefer to read the different frames through which a story is covered. These findings indicate that the political preferences of followers might depend on their ideological preferences.

We find the largest percentage and number of followers, for both overall and the dominant pairs for Category D, indicating that many followers want to break their echo chambers both ideologically and politically by diversifying news consumption. Therefore, there exist signs of diversity in the political and ideological preferences of social media followers of news-sources, who in general exhibit a weaker political than ideological diversity. Considering the ideological and political affiliation of sources as a proxy for followers' ideological and political preferences, we observe a dependency of political preference on ideological preference.

7 DISCUSSION AND CONCLUSION

The research question that we try to answer in this paper is: *Is the mass media biased in how it represents different policies?* We study three different axes of media bias, namely ideological affiliation, political affiliation, and audience affiliation. Our analysis shows that the Indian media is indeed biased corresponding to all three of these axes. Comparing the findings of prior work and our own findings, we see that our work validates the different theories of mass media bias and the way social media aids in furthering these biases.

We observe that the Indian mass media is ideologically biased in terms of the coverage it provides to different policy aspects, and the frames through which they are presented (RQ-1a,b). These findings are in line with several previous works that show that news-sources fine-tune their coverage of events based on the ideological inclination of their audience. Boykoff [10] performed content analysis on the coverage of a set of events related to the global justice movement in some prominent and influential newspapers and television networks in the US. They found five dominant frames through which such events were presented. The authors argued that the reason behind this framing bias was the need for the dissidents' need to gain peoples' attention through mass media, and mass media's need to cover such movements. Smith et al. [62] studied protest events in two prominent newspapers in the US, and showed that the mass media packaged protest events through ideological frames to appeal to the general public and influential third parties. Morstatter et al. [40] found evidence of framing in news-sources across multiple geographies for an event related to European defense, and found that news-sources from multiple geographies use these frames to attract audience.

Our findings are in line with these works. We find that the media mostly reports through frames that target the middle class, which forms their primary target audience. Several media houses also provide a significant coverage to political statements and rhetoric, since this too brings audience attention. However, nearly all media houses provide an insignificant coverage to the frames corresponding to the poor, since the poor do not form the primary audience for English mainstream media in India. Thus, we find that ideological biases in mass media lead to a deficiency in coverage of issues related to the poor.

We also find that the difference between the distributions of mass media and social media coverage of aspects is insignificant. Additionally, since the mass media provides an insignificant coverage to the immediate aspects of the poor [57], the social media community reflects a similar behavior as well. The results hint towards the fact that the Twitter follower community chiefly consists of the middle class, and is less keen on talking about the immediate aspects impacting the poor. Additionally, the poor often not being present on social media (especially Twitter), are unable to represent their own issues on it.

Mass media also exhibits political biases w.r.t. the two largest political parties in India (RQ-2). Budak et al. [11] presented a similar finding in the US context, i.e., they found a majority of the considered news outlets to be biased, in terms of the criticism they exhibit with respect to the two major political parties, while maintaining an objective reporting style. We found that Indian news-sources covered both pro- and anti-statements corresponding to a party [30]. This may also reflect the long-term or short-term political affiliations of mass media based on their ownership networks. Our results are corroborated also by commonly believed political affiliations of news-sources [38], e.g., TeleG is known to be a leftist news-source, and it does exhibit an *anti* alignment towards both parties, which do not follow the leftist ideology. NIE and TOI are both known to be pro-BJP as also shown by our analysis. IE, believed to be supported by the INC, shows a clear pro-INC alignment.

As observed by Saez-Trumper et al. [51], the social media follower community is seen to further these biases for the news-sources considered in our study. Contrary to the existing literature on echo chambers in social media [8, 46], however, we find some signs of followers trying to diversify their news consumption on social media w.r.t. their political and ideological preferences (RQ-3a,b,c). We also see that the followers tend to diversify more in terms of their ideological than political preferences, indicating that the political preferences might take a considerable amount of time to change as discussed in previous studies [45]. Thus, even in cases of majoritarian politics (like in India where currently the centrally ruling party BJP is the single largest party in the Parliament), there are indications of ideological diversity among readers, which may even translate into a change in political preferences in the future.

Our research also poses some interesting questions for further investigation. We showed in this work that the social media community of followers tend to follow news sources with different ideological and political biases. As part of future work, we also wish to find out the relationship between biases existing in news content and user engagement on social media, i.e., how and what type of users interact with biased mass media content. There have been several early studies on these lines [28, 64], which established users' tendencies to prefer polarized media outlets that suit their own political preferences. Garimella and Weber [23] established that the impact of this behavior is also reflected in social media, in the US context. The authors showed the steady increase in the following of polarized media houses by the social media audience, over a span of seven years.

The news-sources considered are all mainstream national dailies, and we do not currently consider opinions and editorials in this study. Thus, most of the articles that we analyzed exhibit an objective reporting. While we analyze the sentiment slant of articles to understand their ideological framing biases, this also motivates us to understand how the news-sources spin stories to frame them in a desired way. Recent work [5] on automated NLP based methods to detect issue specific and generic spin of political news can be leveraged in this respect. This analysis will be more relevant once we extend the current work to fringe outlets, opinions and editorials, and other regional news sources, which are known to exhibit a polar reporting style, and at times, to disseminate misinformation.

Our analysis of audience affiliation of news sources can also be extended to study the correlation between audience affiliation and the social media connections of the audience. For instance, to find out if followers of news sources with a particular political affiliation also tend to follow other social media users with the same affiliation, resulting in formation of polarized cohorts (tribalism in social media), and how these cohorts evolve over time, provided the occurrence of various political events. In this direction, Tokita et al. [66] established that social media gets increasingly sorted politically (leading to diminished cross-ideology connections among users), when media houses report differently on political issues. To understand the polarization among social media users, their social media connections, profile information, and the influential nodes that they are connected to can be leveraged. On similar lines, we can also study the role of Twitter bots in spreading biased news or misinformation in social media, and their impact on these cohorts. A similar study by Aldayel et al. [3] studied the roles of bots in polarizing the political discourse on social media.

Although we study economic policy aspects in this work, our methods are easily generalizable to any other domain of bias analysis like analysis of bias in policy documents, political speeches and debates, news belonging to other domains apart from politics, etc. To the best of our knowledge, this is the first framework for automated bias analysis of media along multiple axes of ideological, political, and audience affiliation that uses longitudinally analyzed web data of large scale. We also developed a method to identify the dominant frames in media discourse on policy events, using a mix of qualitative and computational techniques, which can be used to identify dominant frames for any set of events with a gentle amount of manual intervention. This method of frame identification can help understand the position of media houses with respect to different sections of citizens, aiding in self-regulation of media and provisioning appropriate representation to all sections of people. Additionally, it can aid news readers in further diversifying their news consumption, by providing them signals of media bias along different axes, thereby reducing the impact of bias originating from a single set of sources.

Limitations: Our work has a few limitations with respect to the study of audience affiliation of news sources. First, we have studied the news consumption preferences of only the social media follower community of the news-sources. Our findings thus cannot be generalized to the Indian population. Second, while we have studied the followers of the news-sources, and considered their followership as a proxy for their news consumption preferences, this is not necessarily true. It may be better to capture the engagement (reactions, retweets, and comments) of users with news handles on Twitter, or with tweets covered by news-sources to get a proper sense of their news consumption patterns.

There also exist some methodological limitations. The five constituencies studied in this work are determined by three annotators knowledgeable about the policies considered, after manually going through the news articles. This can still lead to subjectivity. A better approach might be to survey more readers to determine the constituencies in terms of topics and concerns. Another limitation is the limited amount of training data to identify statement/sentence level political biases of news sources. We can augment our dataset with other datasets that have been annotated for political affiliation at the sentence level. This will help us train our model with larger volume of data. Finally, since a large percentage of Indian readers consume news from regional sources, inclusion of regional and vernacular sources in our analysis is important.

REFERENCES

[1] Amr Ahmed and Eric Xing. 2010. Staying informed: supervised and semi-supervised multi-view topical analysis of ideological perspective. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. 1140–1150.
[2] Aayush Ailawadi. 2017. Your Aadhaar Data Is Now With Private Companies As Well. https://www.thequint.com/news/business/aadhar-data-with-private-companies
[3] Abeer Aldayel and Walid Magdy. 2022. Characterizing the role of bots' in polarized stance on social media. *Social Network Analysis and Mining* 12, 1 (2022), 1–24.
[4] Jisun An, Meeyoung Cha, P Krishna Gummadi, and Jon Crowcroft. 2011. Media Landscape in Twitter: A World of New Conventions and Political Diversity.. In *ICWSM*.
[5] Hayastan Avetisyan and David Broneske. 2021. Identifying and Understanding Game-Framing in Online News: BERT and Fine-Grained Linguistic Features. In *Proceedings of The Fourth International Conference on Natural Language and Speech Processing (ICNLSP 2021)*. 95–107.
[6] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.
[7] Samarth Bansal. 2018. "The Indian media is politically free but imprisoned by profit": P Sainath. https://bansalsamarth.medium.com/the-indian-media-is-politically-free-but-imprisoned-by-profit-p-sainath-5191b331c249
[8] W Lance Bennett and Shanto Iyengar. 2008. A new era of minimal effects? The changing foundations of political communication. *Journal of communication* 58, 4 (2008), 707–731.
[9] Dylan Bourgeois, Jérémie Rappaz, and Karl Aberer. 2018. Selection bias in news coverage: learning it, fighting it. In *Companion Proceedings of the The Web Conference 2018*. 535–543.
[10] Jules Boykoff. 2006. Framing dissent: Mass-media coverage of the global justice movement. *New Political Science* 28, 2 (2006), 201–228.
[11] Ceren Budak, Sharad Goel, and Justin M Rao. 2016. Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly* 80, S1 (2016), 250–271.
[12] Michael A Cacciatore, Ashley A Anderson, Doo-Hun Choi, Dominique Brossard, Dietram A Scheufele, Xuan Liang, Peter J Ladwig, Michael Xenos, and Anthony Dudo. 2012. Coverage of emerging technologies: A comparison between print and online media. *New media & society* 14, 6 (2012), 1039–1059.
[13] Josemar A Caetano, Hélder S Lima, Mateus F Santos, and Humberto T Marques-Neto. 2018. Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election. *Journal of internet services and applications* 9, 1 (2018), 1–15.
[14] Praveen Chakravarty and Vivek Dehejia. 2017. Will GST exacerbate regional divergence. *Economic and political weekly* 52, 25&26 (2017), 97–101.

[15] Chun-Fang Chiang and Brian Knight. 2011. Media bias and influence: Evidence from newspaper endorsements. *The Review of Economic Studies* 78, 3 (2011), 795–820.

[16] GST council. Last Modified in May 2020. Goods and Services Tax Council. http://gstcouncil.gov.in/compendium-articles-gst

[17] Peter M Dahlgren, Adam Shehata, and Jesper Strömbäck. 2019. Reinforcing spirals at work? Mutual influences between selective news exposure and ideological leaning. *European journal of communication* 34, 2 (2019), 159–174.

[18] Rakesh Dubbudu. 2018. The Central Government spent close to Rs 10000 crore on Publicity in the last 16 years. https://factly.in/the-central-government-spent-10000-crore-on-publicity-in-the-last-16-years/

[19] Stanley Feldman. 2013. Political ideology. *The Oxford handbook of political psychology* (2013).

[20] Giselle Gemma Fernandez and Roseline Florence Gomes. 2018. Framing and the construction of the recent demonetisation by the Indian mass media. *International Journal of Research and Analytical Reviews* 3, 1 (2018), 1–4.

[21] Richard Fletcher and Ramus Kleis Nielsen. 2017. Using social media appears to diversify your news diet, not narrow it. https://www.niemanlab.org/2017/06/using-social-media-appears-to-diversify-your-news-diet-not-narrow-it/

[22] Rama Rohit Reddy Gangula, Suma Reddy Duggenpudi, and Radhika Mamidi. 2019. Detecting political bias in news articles using headline attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. 77–84.

[23] Venkata Rama Kiran Garimella and Ingmar Weber. 2017. A long-term analysis of polarization on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11. 528–531.

[24] R Kelly Garrett. 2009. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of computer-mediated communication* 14, 2 (2009), 265–285.

[25] Matthew Gentzkow and Jesse M Shapiro. 2010. What drives media slant? Evidence from US daily newspapers. *Econometrica* 78, 1 (2010), 35–71.

[26] Ejaz Ghani and Saurabh Mishra. 2020. Closing the digital divide. https://www.financialexpress.com/opinion/closing-the-digital-divide/2126724/

[27] Edward S Herman. 1988. Manufacturing consent: The political economy of the mass media (2002, Edward S. Herman and Noam Chomsky; with a new introduction by the authors.; Updated ed. of: Manufacturing consent. c1988.; Includes bibliographical references and index. ed.).

[28] Shanto Iyengar and Kyu S Hahn. 2009. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of communication* 59, 1 (2009), 19–39.

[29] Surinder Jodhka. 2018. Rural change in times of 'distress'. *Economic and Political Weekly* 53, 26 (2018), 5–7.

[30] Navreet Kaur and Ankur Sharma. 2020. *Ideology Detection in the Indian Mass Media.* Technical Report. EasyChair.

[31] Joseph T Klapper. 1960. The effects of mass communications. (1960).

[32] Amartya Lahiri. 2020. The great Indian demonetization. *Journal of Economic Perspectives* 34, 1 (2020), 55–74.

[33] Wei-Hao Lin, Eric Xing, and Alexander Hauptmann. 2008. A joint topic and perspective model for ideological discourse. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 17–32.

[34] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. http://www.aclweb.org/anthology/P/P14/P14-5010.

[35] Maxwell McCombs. 1997. New frontiers in agenda setting: Agendas of attributes and frames. *Mass communication review* 24, 1/2 (1997), 32–52.

[36] Solomon Messing and Sean J Westwood. 2014. Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication research* 41, 8 (2014), 1042–1063.

[37] Atul Kumar Mishra. 2015. Newspapers in India and their Political ideologies. https://rightlog.in/2015/07/newspapers-in-india-and-their-political-ideologies/

[38] Atul Kumar Mishra. July 2015. Newspapers in India and their Political ideologies. https://tfipost.com/2015/07/newspapers-in-india-and-their-political-ideologies/.

[39] Prasanna Mohanty. 2019. Labour reforms: No one knows the size of India's informal workforce, not even the govt. http://delivery.acm.org/10.1145/3210000/3209581/p54-baeza-yates.pdf?ip=115.97.131.182&id=3209581&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1549915037_a695eb7be552f3c0798bafed27b82ca7

[40] Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen R Corman, and Huan Liu. 2018. Identifying framing bias in online news. *ACM Transactions on Social Computing* 1, 2 (2018), 1–18.

[41] Preeti Mudliar and Joyojeet Pal. 2013. ICTD in the popular press: media discourse around Aakash, the'world's cheapest tablet'. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers-Volume 1*. 43–54.

[42] Viet-An Nguyen, Jordan L Ying, and Philip Resnik. 2013. Lexical and hierarchical topic regression. In *Advances in neural information processing systems*. 1106–1114.

[43] Sevanti Ninan. 2007. *Headlines from the heartland: Reinventing the Hindi public sphere.* Sage.

[44] Zizi Papacharissi and Maria de Fatima Oliveira. 2008. News frames terrorism: A comparative analysis of frames employed in terrorism coverage in US and UK newspapers. *The international journal of press/politics* 13, 1 (2008), 52–74.

[45] Johnathan C Peterson, Kevin B Smith, and John R Hibbing. 2020. Do people really become more conservative as they age? *The Journal of Politics* 82, 2 (2020), 600–611.

[46] Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. 2016. Echo chambers on Facebook. *Available at SSRN 2795110* (2016).

[47] Stephen D Reese, Tim P Vos, and Pamela J Shoemaker. 2009. Journalists as gatekeepers. In *The handbook of journalism studies*. Routledge, 93–107.

[48] Filipe N Ribeiro, Lucas Henriqueo, Fabricio Benevenutoo, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P Gummadi. 2018. Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale. (2018).

[49] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*. ACM, 399–408.

[50] Diego Sáez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: gatekeeping, coverage, and statement bias. In *CIKM*.

[51] Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 1679–1684.

[52] Javier Sánchez-Junquera, Paolo Rosso, Manuel Montes, and Simone Paolo Ponzetto. 2021. Masking and Transformer-based Models for Hyperpartisanship Detection in News. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*. 1244–1251.

[53] Michael Scharkow, Frank Mangold, Sebastian Stier, and Johannes Breuer. 2020. How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences* 117, 6 (2020), 2761–2763.

[54] Dietram A Scheufele and David Tewksbury. 2006. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication* 57, 1 (2006), 9–20.

[55] Holli A Semetko and Patti M Valkenburg. 2000. Framing European politics: A content analysis of press and television news. *Journal of communication* 50, 2 (2000), 93–109.

[56] Anirban Sen, Priya Chhillar, Pooja Aggarwal, Sravan Verma, Debanjan Ghatak, Priya Kumari, Manpreet Agandh, Aditya Guru, and Aaditeshwar Seth. 2019. An attempt at using mass media data to analyze the political economy around some key ICTD policies in India. 1–11. https://doi.org/10.1145/3287098.3287108

[57] Anirban Sen, Debanjan Ghatak, Kapil Kumar, Gurjeet Khanuja, Deepak Bansal, Mehak Gupta, Kumari Rekha, Saloni Bhogale, Priyamvada Trivedi, and Aaditeshwar Seth. 2019. Studying the discourse on economic policies in India using mass media, social media, and the parliamentary question hour data. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*. 234–247.

[58] Saif Shahin and Pei Zheng. 2020. Big data and the illusion of choice: Comparing the evolution of India's aadhaar and China's social credit system as technosocial discourses. *Social Science Computer Review* 38, 1 (2020), 25–41.

[59] Ankur Sharma, Navreet Kaur, Anirban Sen, and Aaditeshwar Seth. 2020. Ideology detection in the indian mass media. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 627–634.

[60] Carson Sievert and Kenneth Shirley. 2014. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*. 63–70.

[61] Yanchuan Sim, Brice DL Acree, Justin H Gross, and Noah A Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 91–101.

[62] Jackie Smith, John D McCarthy, Clark McPhail, and Boguslaw Augustyn. 2001. From protest to agenda building: Description bias in media coverage of protest events in Washington, DC. *Social Forces* 79, 4 (2001), 1397–1423.

[63] Stephanie. 2016. Inter-rater Reliability. https://www.statisticshowto.datasciencecentral.com/inter-rater-reliability/

[64] Natalie Jomini Stroud. 2010. Polarization and partisan selective exposure. *Journal of communication* 60, 3 (2010), 556–576.

[65] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment in Short Strength Detection Informal Text. *J. Am. Soc. Inf. Sci. Technol.* 61, 12 (Dec. 2010), 2544–2558. https://doi.org/10.1002/asi.v61:12

[66] Christopher K Tokita, Andrew M Guess, and Corina E Tarnita. 2021. Polarized information ecosystems can reorganize social networks via information cascades. *Proceedings of the National Academy of Sciences* 118, 50 (2021).

[67] Sevgi Yigit-Sert, Ismail Sengor Altingovde, and Özgür Ulusoy. 2016. Towards detecting media bias by utilizing user comments. In *Proceedings of the 8th ACM Conference on Web Science*. ACM, 374–375.

A APPENDIX

A.1 Keywords to Collect News Articles

The final set of keywords used to collect the articles pertaining to the four economic policies are shown in Table 4.

A.2 Relative Coverage of Aspects

In this section, we present the relative aspect coverage provided by the mass media and social media to different policies. We look at the aggregate coverage distributions. Aggregate aspect coverage across all news-sources is calculated by summing up the number of words belonging to the aspect across these sources, and then dividing this number by the total number of words across all aspects in all news-sources for that policy. Figure 6 shows the distributions. From the plots, it is evident that the social media follower community closely follows the coverage trend of aspects in the mass media. This is also established in our main paper. The Pearson correlation coefficients for the aspect coverage distributions of mass media and social media are: 0.92 for Demonetization, 0.90 for Aadhaar, 0.94 for GST, and 0.66 for Farmers'Protest. The low correlation for the last event arises from just two aspects: *Irrigation concerns and water pollution* and *Crimes and suicides in farmer community*. These two aspects are much highly posted about on Twitter (compared to the mass media), being sensitive issues related to the farmers. For the other three policies, the high values of correlation indicate towards a high alignment between the mass media and social media aspect coverage. We obtain similar trends even when we do this analysis for individual news-sources, which we do not report in this paper.

From the plots, it is also clear that both mass media and social media are biased in terms of aspect coverage, i.e., there exists a significant imbalance in the coverage of aspects. To state empirically, for Demonetization, Aadhaar, GST, and Farmers' Protests, the relative coverages in mass media for the highest and lowest covered aspects are (15.9%,0.8%), (9.9%,1.4%), (19%,2.3%), and (18%,0.1%), respectively. For the social media follower community, these are (26%,0.5%), (13.6%,0.6%), (14.1%,0.06%), and (31.6%,0.03%), respectively. The high inequities observed in these ordered pairs are indicative of the bias in aspect coverage exhibited by both mass media and social media.

A.3 Coverage of Constituencies by Mass Media

In this section, we analyze the relative coverage provided by mass media to the five constituencies of *poor, middle class, corporate, informal sector, and government*. For each constituency, we aggregate the relative coverages provided to the aspects belonging to the constituency using the following equation:

$$relative\_constituency\_coverage = \frac{\sum_{a \epsilon const} w_a}{\sum_{asp \epsilon A} w_{asp}} \tag{4}$$

where a and asp are aspects, A is the set of all aspects for a policy event, and w_a is the total number of words across all articles for aspect a. We show the constituency coverage for each policy even in figure 7. We observe from these plots that for Demonetization, Aadhaar, and GST, the coverage provided to the immediate problems of the poor are significantly lesser than that provided to the politics around a policy issue (represented by the *Government* constituency). Only in case of Farmers' Protests is the coverage provided to *poor* high. This is because most discussions on issues of farmers involve poor farmers and daily wagers, and both the ruling party and the opposition make a significant number of statements in the mass media on this sensitive issue. The least discussion happens for the constituency *informal sector*, which includes a significant portion of the total workforce in India [39], and includes workers, labourers, vendors, and small traders belonging to the unorganized sector with often low levels of income. In tables 5, 6, 7, and 8 we report the KS-statistics (2-sample test) of relative coverage for each pair of the five constituencies. The high values of KS-statistics along with the low p-values indicate that the difference in coverages are significant with above 99% significance level (that is, we can safely reject the null hypothesis that the coverages come from the same distribution for two different constituencies). This also indicates the existence of a constituency bias in mass media. The snapshot of our coding schema for Demonetization, and the aspect-to-constituency alignment matrices are shown in tables 11, 12, 13, and 14.

A.4 Similarity between Mass Media and Social Media in terms of Aspect Coverage

Table 15 shows the JS Divergence between the aspect coverage distribution of mass media sources and their corresponding social media communities, for each policy.

Keywords (manually selected)
Demonetization: demonitisation, demonitization, denomination note, cash withdrawal, swipe machine, unaccounted money, withdrawal limit, pos machine, fake currency, digital payment, digital transaction, cash transaction, cashless economy, black money, cash crunch, currency switch, long queue, demonetised note, cashless transaction, note ban, currency switch
Aadhaar: aadhar, aadhaar, adhar, adharcard, aadharcard, aadhaarcard, uidai, aadhar card, public distribution system, pds, ration card, ration, e-pos
GST: gst, goods and service tax, goods & services tax, gabbar singh tax, goods service tax, goods and services tax
Farmers' Protest: farm loan, crop loan, farmer suicide, debt waiver, waiver scheme, farming community, farmer agitation, plight farmer, distressed farmer, farmer issue, farmers protest, farmers' protest, agrarian crisis, agrarian unrest, farmers protests, farmers' protests, loan waivers, loan waiver, agriculture protest, farmers' march

Table 4: List of manually collected keywords used to extract articles (and tweets) corresponding to the economic policy events. Here, we only show the manually selected keywords after converting them to lowercase, and after pre-processing of the articles was done.

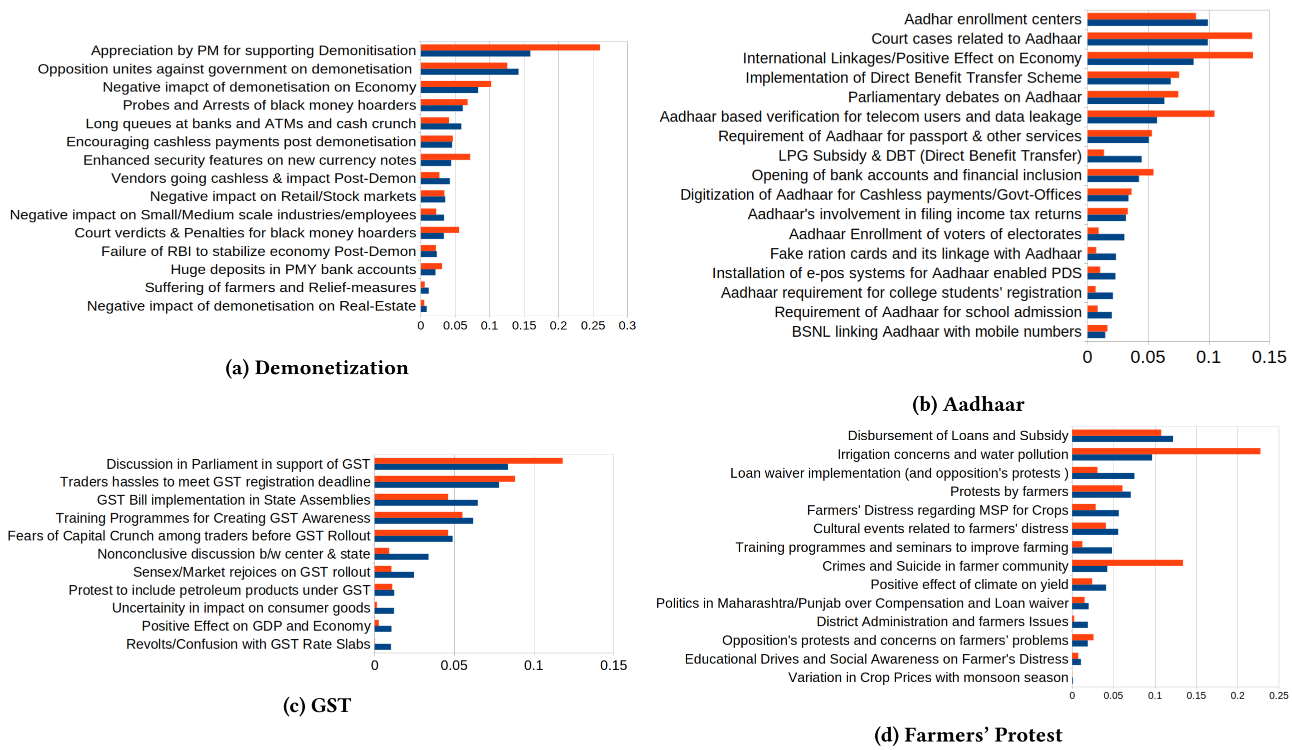


Figure 6: Aggregate relative coverage provided by the mass media and its social media follower community corresponding to each policy. The blue bars show the aggregate aspect coverage for mass media, while the red ones show the coverage for social media.

	Poor	Middle Class	Corporate	Informal Sector	Govt.
Poor	0	0.43	0.86	0.86	1
Middle Class	0.43	0	1	0.86	1
Corporate	0.86	1	0	0.43	1
Informal Sector	0.86	0.86	0.43	0	1
Government	1	1	1	1	0

Table 5: KS statistics (2-sample test) for relative coverage provided by the mass media to the five constituencies for Demonetization. All p-values lie below 0.05.

	Poor	**Middle Class**	**Corporate**	**Informal Sector**	**Govt.**
Poor	0	1	0.71	1	1
Middle Class	1	0	1	1	1
Corporate	0.71	1	0	1	1
Informal Sector	1	1	1	0	1
Government	1	1	1	1	0

Table 6: KS statistics (2-sample test) for relative coverage provided by the mass media to the five constituencies for Aadhaar. All p-values lie below 0.05.

	Poor	**Middle Class**	**Corporate**	**Informal Sector**	**Govt.**
Poor	0	1	1	0.57	1
Middle Class	1	0	0.43	1	1
Corporate	1	0.43	0	1	1
Informal Sector	0.57	1	1	0	1
Government	1	1	1	1	0

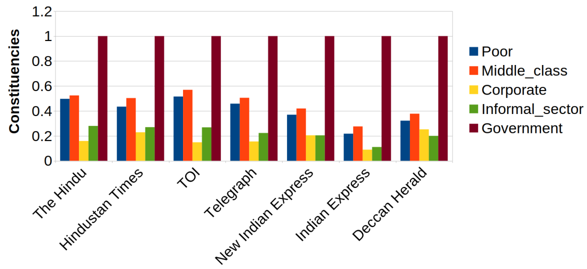
Table 7: KS statistics (2-sample test) for relative coverage provided by the mass media to the five constituencies for GST. All p-values lie below 0.05.

	Poor	**Middle Class**	**Corporate**	**Informal Sector**	**Govt.**
Poor	0	1	1	1	1
Middle Class	1	0	1	1	1
Corporate	1	1	0	0.71	1
Informal Sector	1	1	0.71	0	1
Government	1	1	1	1	0

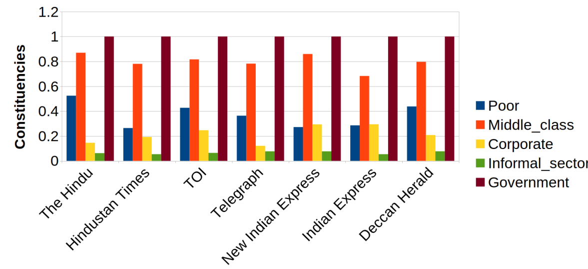
Table 8: KS statistics (2-sample test) for relative coverage provided by the mass media to the five constituencies for Farmers' Protests. All p-values lie below 0.05.

Constituency	**Does the article primarily target:**	**Normative Definition**
Poor	- Labourers, workers in factories and small mills (e.g., textile and diamond-cutting mills), migrant workers and labourers, poor people belonging to the lowest level of income, and workers without bank accounts	Poor people at the lowest levels of income. This includes labourers and factory workers without bank accounts. The welfare schemes which target the poor directly, like PMGKDS also come in the ambit of this class. Casual workers (working on contractual basis) with daily wage below 200 INR.
Middle class	- Workers employed in sectors with higher income range (e.g., daily wagers working in garment based activities like stitching), workers for whom absence of bank accounts is not specifically mentioned, ATMs, cash withdrawal limit, Note exchange at post offices, banks, customers, ...	Middle class people who suffered the immediate aftermath of the policy move like standing in long queues at ATMs, lack of money exchange at banks and post offices, and so on. Workers and daily wagers ...

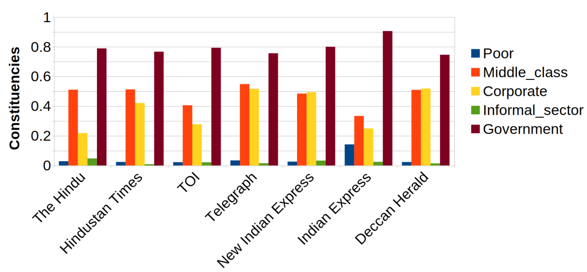
Table 9: Snapshot of coding schema for Demonetization (part 1): the schema is used to map aspects to the relevant constituencies accurately, with minimum subjectivity. The final schema has been updated after multiple rounds of discussion and due deliberation.



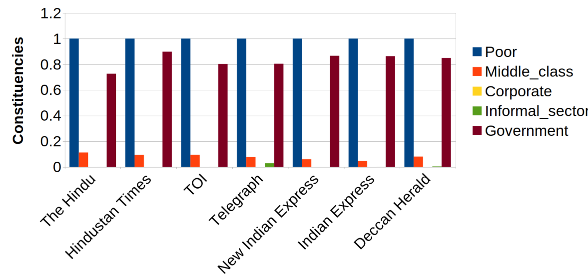
(a) Demonetization



(b) Aadhaar



(c) GST



(d) Farmers' Protest

Figure 7: Relative coverage provided by the mass media to each of the five constituencies for the four policies. The news-sources are shown in the x-axis from left to right in the order: The Hindu, Hindustan Times, The Times of India, Telegraph, New Indian Express, Indian Express, and Deccan Herald

Constituency	**Does the article primarily target:**	**Normative Definition**
Corporate	Manufacturing companies, industries, MSMEs, factories, multinationals, businesses, big real estate companies, Entrepreneurs, businessmen, bizmen, Import, export, raw material, brands, marketing, Sensex, investors, NSE, NIFTY, BSE, foreign capital ...	Big business houses, industrialists, SMEs and MNCs, and corporate business houses in general.
Informal sector and small traders	Unorganized sector, informal sector, companies not registered, unregistered enterprises, Small vendors/businesses, ...	Unorganized sector, unregistered companies, small traders, and vendors.
Government	State/Central government, state, centre, Name of prominent politician, minister, ministry, MP/MLA, their relatives, Names/positions of important government officials and designations ...	State and central government, policy makers, ministers, ministries, MPs, MLAs, and their relatives. Discussions in Parliament or assemblies about the narrative on Demonetization also come in this class.

Table 10: Snapshot of coding schema for Demonetization (part 2): the schema is used to map aspects to the relevant constituencies accurately, with minimum subjectivity. The final schema has been updated after multiple rounds of discussion and due deliberation.

Aspect	Poor	Middle Class	Corporate	Inf. Sector	Govt.
Failure of RBI to answer questions raised post demonetisation	0	-1	-1	0	1
Negative impact of Demonetisation on small and medium scale industries and its employees	-1	-1	-1	-1	1
Long queues at banks and ATMs and cash crunch	0	1	1	1	0
Court verdicts related to demonetisation and penalties issued for black money hoarders	1	1	-1	0	1
Negative impact on retail and stock markets and suspicious deposits in bank accounts	-1	-1	-1	0	1

Table 11: Snapshot of alignment matrix for Demonetization. We show only five aspects in this table out of the 16 aspects for Demonetization.

Aspect	Poor	Middle Class	Corporate	Inf. Sector	Govt.
Positive effect of climatic conditions on agriculture yield	-1	0	0	0	0
Opposition's protests and concerns on problems related to farmers (including Demonetization)	1	0	0	0	-1
Educational Drives and Social Awareness on Farmer's Distress	-1	0	0	0	0
Politics in Maharashtra/Punjab over Compensation and Loan waiver for farmers	-1	0	0	0	1
Variation in Crop Prices with monsoon season	-1	-1	0	-1	0

Table 12: Snapshot of alignment matrix for Farmers' Protests. We show only five aspects in this table out of the 14 aspects for Farmers' Protests.

Aspect	Poor	Middle Class	Corporate	Inf. Sector	Govt.
Requirement of Aadhaar for passport and other services (concessions)	0	1	0	0	1
Fake ration cards caught due to Aadhaar linkage, aiding in the good of poor and middle class	1	1	0	0	1
Installation of e-pos systems for Aadhaar enabled PDS causing resentment among poor and middle class	-1	-1	0	0	1
Digitization of Aadhaar enabled employees' provident fund, attendance systems at public offices, and cashless payments helping the middle class	0	1	0	0	1
Requirement of Aadhaar for school admission and the middle class	0	1	0	0	1

Table 13: Snapshot of alignment matrix for Aadhaar. We show only five aspects in this table out of the 17 aspects for Aadhaar.

Aspect	Poor	Middle Class	Corporate	Informal Sector	Government
Sensex/Market rejoices on GST rollout	0	1	1	1	1
GST Bill implementation in State Legislative Assemblies	0	1	1	1	0
Traders hassles to meet GST registration deadline and changes in sensex/nifty	0	0	-1	-1	-1
Fears of Capital Crunch among traders before GST Rollout	0	0	1	0	1
Training Programmes for Creating GST Awareness	0	0	1	1	1

Table 14: Snapshot of alignment matrix for GST. We show only five aspects in this table out of the 11 aspects for GST.

News Source	Demonetization	Aadhaar	GST	Farmers Protest
	TweetFol	TweetFol	TweetFol	TweetFol
Hindu	0.12	0.08	0.17	0.15
HT	0.13	0.03	0.18	0.07
IE	0.14	0.03	0.28	0.08
NIE	0.11	0.04	0.13	0.07
TeleG	–	0.11	–	0.07
TOI	0.11	0.04	0.12	0.04
DecH	0.12	0.10	0.15	0.11

Table 15: [RQ3] JS divergence showing difference in aspect coverage between mass media and social media: for TeleG, we could not find any tweet for Demonetization and GST. The Kolmogorov-Smirnov 2-sample test also suggest that the aspect coverage are significantly similar between the mass media and social media.