

Characterizing Political Campaigning with Lexical Mutants on Indian Social Media

Shruti Phadke¹, Tanushree Mitra²

¹ University of Texas at Austin

² University of Washington

shruti.ph@gmail.com, tmitra@uw.edu

Abstract

Increasingly online platforms are becoming popular arenas of political amplification in India. With known instances of pre-organized coordinated operations, researchers are questioning the legitimacy of political expression and its consequences on the democratic processes in India. In this paper, we study an evolved form of political amplification by first identifying and then characterizing political campaigns with lexical mutations. By lexical mutation, we mean content that is reframed, paraphrased, or altered while preserving the same underlying message. Using multilingual embeddings and network analysis, we detect over 3.8K political campaigns with text mutations spanning multiple languages and social media platforms in India. By further assessing the political leanings of accounts repeatedly involved in such amplification campaigns, we contribute a broader understanding of how political amplification is used across various political parties in India. Moreover, our temporal analysis of the largest amplification campaigns suggests that political campaigning can evolve as temporally ordered arguments and counter-arguments between groups with competing political interests. Overall, our work contributes insights into how lexical mutations can be leveraged to bypass the platform manipulation policies and how such competing campaigning can provide an exaggerated sense of political divide on Indian social media.

Introduction

Online political activism is rapidly becoming a weapon of mass influence on Indian social media. Starting from the Indian general elections in 2014, social media has been used as a campaigning arena in significant political events over the last several years (Sohal and Kaur 2018; Akbar et al. 2021; Dash et al. 2021). Especially given the recent allegations (Singh 2021; Bureau 2021; Rajgarhia 2020) of platform manipulation by political parties in India, it is important to invest in computational research that studies online influence operations outside of the West. In fact, a recent study uncovered organized political influence where party supporters received tweet templates through WhatsApp and Google docs and were encouraged to create cospasta campaigns to influence public opinion (Jakesch et al. 2021). One

message quoted in their study also suggests that users may be instructed to tweak the messaging template without directly copy-pasting on social media [(Jakesch et al. 2021) Pg. 9].

Upholding humanitarian values, CAA empowers persecuted minorities by giving Indian citizenship. #WeSupportCAA

CAA respects humanitarian values by empowering persecuted minorities #StandwithCAA

CAA upholds humanitarian values empoweing minorieis

BJP government upholds humanitarian values through CAA. Indian citizenship for persecuted minorities #BJP4India

BJP respects humanitarian values through CAA by offering refuge to persecuted minorities #BJP4India

Figure 1: Examples of messages from an amplification campaign with lexical mutants. There are 932 more messages similar to this in different languages (English, Hindi, Punjabi, and Marathi), distributed by different Twitter accounts and Facebook groups.

For example, consider messages in Figure 1 about a controversial amendment to Citizenship Act in India that offered asylum to religious minorities to surrounding countries, excluding the Muslim population. All messages have the same underlying content, however, each message varies lexically. We argue that subtle lexical mutations such as this, especially spanning over multiple languages, could contribute to a larger political amplification campaign which is harder to detect than a simple cospasta¹. This paper focuses on detecting and analyzing political amplification campaigns with lexical mutations. By “amplification campaigns”, we mean coordinated efforts at amplifying certain opinions on online social media (Weber and Neumann 2021). While analyzing amplification, we specifically study coordination in language through lexical mutants. We use the term “*lexical mutant*” (re-purposed from Meme mutation (Simmons, Adamic, and Adar 2011; Leskovec, Backstrom, and Kleinberg 2009)) to mean content that is reframed, paraphrased, rephrased, or altered while preserving the same underlying message. Moreover, we extend this definition to incorporate mutations across Indian languages. To first find evidence of

¹a pattern of different users posting identical messages to possibly amplify a certain opinion (Weber and Neumann 2021)

and then assess the extent of amplification campaigns with lexical mutations, we ask:

RQ1: How can we identify political message amplification campaigns with lexical mutants on Indian social media?

We focus on two recent national events in India: the introduction of the Farm Bills and the Citizenship Amendment Act CAA (explained in detail in the Data section) and curate a cross-platform dataset of Tweets and Facebook group posts. Considering the use of multiple languages along with English on Indian social media, we use multilingual sentence embeddings with subsequent network analysis and identify over 3.8K political amplification campaigns with lexical mutations. By further establishing that nearly 34% of the messages in amplification campaigns are unique lexical mutants, our work provides an essential context into the actual expanse of political amplification beyond cospasta.

After identifying the amplification campaigns, we further characterize the use of political amplification across various dimensions. Given that most of the previous studies focus only on one right-wing political party BJP or, primarily focus on only one social media platform, Twitter, we analyze amplification across multiple Indian political parties and extend our analysis to Facebook.

RQ2: What are the characteristics of amplification campaigns?

RQ2a: How are political messages amplified across the Indian political spectrum?

RQ2b: How are political messages amplified across different platforms?

RQ2c: What dominant claims are amplified by accounts with various political leanings?

By manually assigning political leaning to accounts, we find that BJP, anti-BJP, and the accounts supporting other political parties (INC and AAP) all use political amplification equitably (Figure 6). Furthermore, we find that 30% of the amplification campaigns spread across platforms (Figure 6 (c)). More interestingly, by analyzing prominently amplified political claims, we find that most large-scale campaigns are reactions to claims made by the political opposition. For example, political claims amplified by anti-BJP accounts are later countered through a similar amplification campaign by BJP (Figure 7).

Overall, our results indicate that studies on political influence in India can benefit from a larger political context than focusing on only one party. Moreover, our findings indicate that amplification beyond a simple cospasta may thrive in exerting inauthentic influence on the mass in the absence of updated platform manipulation policies. Finally, our results also contribute a novel understanding of the reactionary landscape of online Indian political influence which can compound radicalization and reduce trust in genuine political expression online.

Background and Related Work

Online Political Influence in India

After the 2014 general elections, social media has emerged as an important battleground in Indian politics (Ahmed, Jaidka, and Cho 2016; Jakesch et al. 2021; Jaffrelot 2015). In

an attempt to reach out to the younger population, political parties have started deploying organized political campaigns on social media (Tiwari 2019). Researchers have studied social media manipulation in India during elections (Ahmed, Jaidka, and Cho 2016; Das and Schroeder 2021; Jakesch et al. 2021; Jaffrelot 2015; Sohal and Kaur 2018), and various civil conflicts such as farmers' protests, COVID-19 crisis (Akbar et al. 2021; Dash et al. 2021) and protests against Citizenship Amendment Act (CAA) (Edingo 2021).

These studies independently observe that during elections or civil unrest events, social media in India was flooded with various influential narratives promoting political propaganda. For example, during the early months of COVID-19, the issues related to the pandemic were used to frame anti-Muslim disinformation and populist narratives (Akbar et al. 2021). These disinformation campaigns reflect how political influence narratives in India are closely related to religious fundamentalist attitudes. An in-depth interview study of supporters of different political parties in India, including the BJP, INC, and Communist party, found that Indian social media users are concerned with an increasing amount of religious fundamentalist appeals and narratives on social media (Das and Schroeder 2021).

Computational Research on Online Political Amplification

While the presence of influential political rhetoric is evident in Indian social media, computational research exploring coordinated influence operations in India is limited. Jakesch et al. study 75 cospasta hashtag manipulation campaigns across 600 WhatsApp groups and political Twitter accounts in India and find evidence of centrally controlled political influence operations that benefit from the voluntary participation by the party-followers (Jakesch et al. 2021). While their study brings to light an important phenomenon of influence operations in India, it also motivates further research into identifying such coordinated message templates at scale across more popular platforms such as Facebook. Moreover, while previous research has primarily focused on the ruling political party BJP, it is also important to understand the prevalence of such coordinated influence operations across multiple political parties in India. Furthermore, despite Facebook being the most popular social media platform in India², most prior works on influence operations in India have focused on Twitter (Ahmed, Jaidka, and Cho 2016; Dash et al. 2021; Edingo 2021; Jakesch et al. 2021). Given the increasing engagement seen by polarizing Indian political Facebook groups (Frenkel and Alba 2021), it is important to investigate the coordinated influence activity on this platform.

Outside of India, researchers have studied political amplification in the West, by focusing on re-tweet or re-share or co-tweet networks (Gallagher et al. 2021) or by considering other posting characteristics, such as posting time, user similarity, user coordination in an ensemble (Cinelli et al. 2022; Hristakieva et al. 2022; Weber and Neumann 2021).

²<https://www.statista.com/statistics/1115648/india-leading-social-media-sites-by-page-traffic/>





	Keywords	Timeline	#Messages	#accounts
Farmers' protests	FarmersProtest,KisanAndolan, किसान, आंदोलन, किसान, AIKSCC, farmer, farmers, Kisaan, किसान_बिल, बिल, Farmers, Kisaan, kisan, KisanBill, KisanProtest, TractorMarch, FarmerPolitics, "भारत बंद", bharatbandh, "bharat bandh", कृषि_कानून, कृषि, FarmersBill, BharatBand, भारत_बंद, खेती, farmlaws, Tractor2Twitter, कृषि_कानूनो, FarmLaws2020, FarmersDelhiProtest, फ़ार्मला	Oct 2020 -	 98,297	4,593
		Dec 2020	 701,345	171,345
CAA	CAAProtest, IndianCitizenshipActProtest, "citizenship amendment act", CAA_NRC, ShaheenBagh, शाहीनबाग, CAA, NRC, नागरिकता, नागरिकता_संशोधन_बिल, शाहीन, नागरिकतासंशोधनकानून, Shaheen, CAB, NRCBill, नागरिकताकानून, CitizenshipAmmendmentAct, CABBill, CitizenshipAmendmentBill, नागरिकता_संशोधन_विधेयक, "नागरिकता संशोधन विधेयक", एनआरसी, सीएए	Nov 2019 -	 94,489	4,218
		Jan 2020	 602, 967	125,827

Figure 2: Figure describing dataset keywords, timeline, and the number of posts. Note that we collect only original tweets and posts, excluding retweets and reshares.

Most similar to our work is research by Pacheco that investigates White Helmet coordinated influence networks using text similarity based on pattern recognition (Pacheco, Flammini, and Menczer 2020)

In this work, we extend the prior works by first, analyzing political amplification across multiple political parties and languages in India and second, by including previously understudied Indian Facebook. We further include lexical mutants in amplification campaigns that go beyond simple copy-paste and consider organic content tweaking that could be used to evade detection from moderation against platform manipulation. In the next section, we start by describing our cross-platform dataset centered around political events in India.

Data

Political Events and Keywords

In this paper, we study the political message amplification campaigns on Indian social media. To contextualize our analysis, we focus on two recent politically divisive events that affected the whole nation and attracted partisan debate on social media. We used keywords and hashtags related to these events to collect data from Facebook and Twitter.

Farmers' Protests: Farm acts, passed in the parliament of India in September 2020, sparked nationwide protests by farmers' organizations. Farm unions were primarily protesting the entry of corporations in crop trading facilitated by the farm acts along with legacy issues such as high farmer suicide rates and low agricultural income in India. The protests were highly politicized across the political spectrum in India with the ruling political right (BJP) standing in support of the farmers' bills while the oppositional left (INC, AAP, etc.) aligned with farmers' unions in the protests.

Citizenship Amendment Act (CAA): This amendment to the Citizenship Act was proposed by the Government of India under the leadership of the right-wing BJP party. It offered a pathway to Indian citizenship for persecuted religious minorities from Afghanistan, Bangladesh, and Pakistan who are not Muslims. The amendment was opposed by non-BJP politicians and student organizations, causing polarizing tensions across the nation which led to violent protests and rallies from both sides.

Selecting Keywords for Data Collection: We start building our list of keywords from previous works (Dash et al. 2021, 2022) around the Farmers' protests and CAAs. We aim to make the keyword list more generic to increase the coverage of the dataset, while still ensuring that we capture the relevant data. For example, we strategically include general words like "Kisan" or "सिएए" which are transliterations of the words "farmers" and "CAA" and are more likely to be used only in the Indian context. Table 2 records the keywords used in collecting the data.

Facebook Dataset

Facebook groups have been known to play key roles in recruiting supporters for politicians and political parties during elections (Woolley, Limperos, and Oliver 2010). In fact, Facebook cyber security expert commented that during the 2019 Parliamentary elections in India, Facebook groups and pages were designed to look independent but were actually linked to political parties trying to conceal their identities (Sambhav and Ranganathan 2022). Such Facebook groups were also found to be linked with fake or bot accounts that spread misinformation and influential content relevant to elections (Frenkel and Alba 2021).

We use CrowdTangle's post search API³ to extract posts made on Facebook pages and groups relevant to the political events described above. We use the list of keywords and hashtags mentioned in Figure 2 to collect public posts from Facebook groups and pages. Note that CrowdTangle or any official Facebook API does not offer any authorship information for the posts.

Twitter Dataset

We also collect tweets relevant to the two political events using the Twitter Academic Research API. Similar to the Facebook dataset, to analyze hidden amplification campaigns, we only preserve the original tweets, excluding retweets and quote tweets. Overall details of the dataset are available in Figure 2.

³<https://github.com/CrowdTangle/API/wiki/Search>

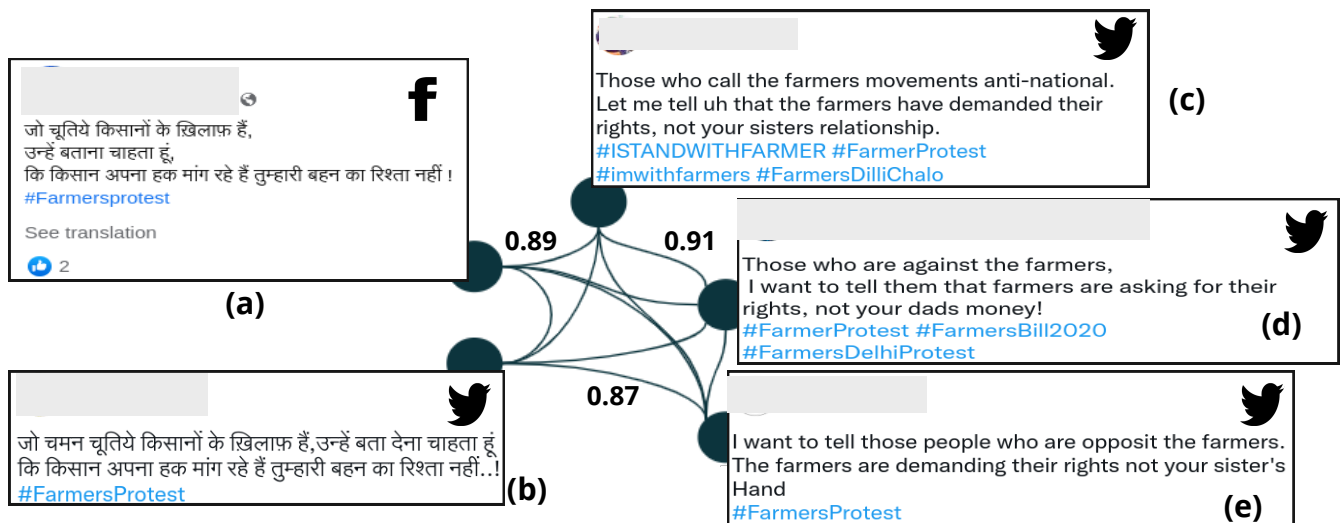


Figure 3: Examples of lexical mutants in an amplification campaign. Two messages (nodes) are connected together if they have high cosine similarity. A clique of nodes connected like this represents an amplification campaign with lexical mutants.

RQ1: Identifying Political Campaigns With Lexical Mutations

A key to identifying amplification campaigns beyond copy-pasta is to identify messages that are similar to each other in terms of core content but are not explicit re-posts or re-tweets (Pacheco, Flammini, and Menczer 2020). Unlike retweets or reposts, messages with similar content are treated as different by the platform and the link between the original and the copy is hard to detect (Pacheco, Flammini, and Menczer 2020). Consider, for example, the messages in Figure 3. All of the messages have similar content with slight lexical variations, posted by different users on different social media platforms. Below, we outline our methodology for identifying groups of messages with lexical mutations across multiple languages on Indian social media.

Characterizing Post Similarity Across Languages

Extracting Multilingual Embeddings: A usual approach to identifying messages with slight lexical mutations may be to compare the edit distance (Leskovec, Backstrom, and Kleinberg 2009), message keywords (Schoch et al. 2022) or to use sentence embeddings with cosine similarity (Pacheco, Flammini, and Menczer 2020). However, posts on Indian social media contain multiple languages. For example, in Figure 3, all the messages have similar content but some are in different languages. In this case, simple token-based word embeddings trained in English, or other fuzzy matching methods will not work.

Instead, we use multilingual sentence embeddings to represent texts across multiple Indian languages and also capture the semantic similarity between paraphrased or re-framed texts. There are several pre-trained multi-lingual models available, such as sentence-BERT (Reimers and Gurevych 2020), Language-Agnostic Sentence Representations (LASER) (Artetxe and Schwenk 2019) and multi-lingual BERT (Devlin et al. 2018). LASER was found to

outperform multilingual BERT for Hindi text classification (Joshi, Goel, and Joshi 2020). Since a significant portion of our text data is in Hindi, we use the LASER model which works with more than 90 languages containing more than 28 different kinds of alphabets. LASER is optimized and evaluated for parallel sentence extraction and translations across different languages, making it suitable for characterizing cross-lingual similarities (Artetxe and Schwenk 2019).

Identifying Similarity Threshold for Lexical Mutation:

Similarity between two texts can be calculated with the cosine similarity between the multilingual sentence embeddings described above. A cosine similarity of 1 indicates perfect similarity between the two texts. Examples in Figure 3 all contain a similar underlying message with variations in languages and phrases. To extract groups of texts such as this, we need to determine a threshold value of cosine similarity above which we can consider two texts to be lexical mutants of each other. We determine this threshold empirically by manually analyzing pairs of texts with different cosine similarity scores. Specifically, we randomly sampled 20 pairs of sentences for cosine similarity values each, starting from 0.5, with increments of 0.05. Some of the sampled pairs contained texts from the same language whereas other pairs contained texts from different languages. We labeled each pair as either 1—to indicate whether a lexical mutation still resulted in preserving the underlying message—or 0. The number of pairs scored as 1 in the sample of 20 naturally kept increasing with the cosine similarity score. All samples with a score of 0.85 were labeled as 1. Hence, we chose 0.85 as the cosine similarity threshold. In other words, for the rest of the downstream analysis, “lexical mutant” posts will mean posts with cosine similarity between multilingual embeddings equal to or above 0.85.

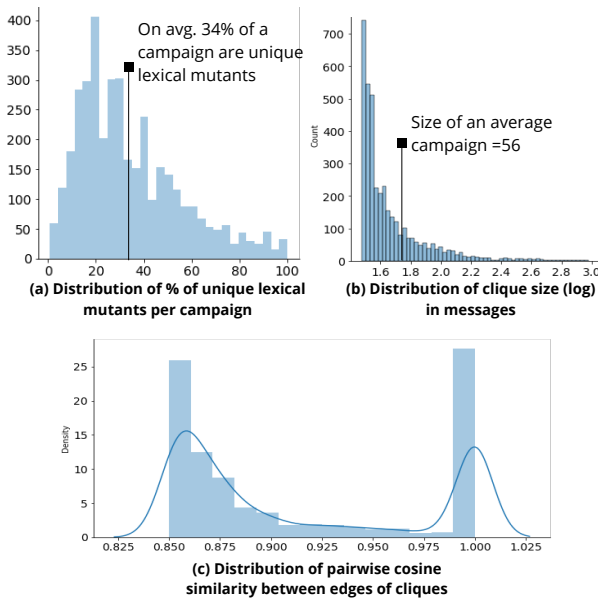


Figure 4: (a) Represents the distribution of the percentage of unique lexical variants in each campaign, calculated by removing all duplicated texts. (b) Shows the distribution of clique sizes in log scale. The average size of the clique is 56 messages. (c) Displays the KDE plot for pairwise cosine similarities between pairs of texts included in cliques. The two peaks of bimodal distribution represent our empirically determined cutoff of cosine similarity (left peak) and also the duplicated texts without lexical variations with the perfect cosine similarity (right peak).

Finding Amplification Campaigns with Lexical Mutations

After identifying lexical mutants through pairwise similarity of multilingual embeddings, to analyze amplification campaigns on a larger scale, it is important to determine large groups of similar messages. For example, there are several other messages similar to the examples in Figure 3 with high semantic similarity to each other. How can we find large groups of lexical mutants? We next take a network-based approach to identify groups of lexical mutants.

Network of Texts: Two messages are connected with each other through an edge if they have a cosine similarity equal to or more than 0.85 between their multilingual embeddings. In a network such as this, finding groups of lexical mutants will be equivalent to finding clusters of nodes that are all connected to each other through high cosine similarity. In other words, clusters of completely connected nodes will represent texts which are all lexical mutants of each other. This is commonly referred to as finding cliques (Clark and Holton 1991). A clique is a sub-graph in which all nodes are connected to each other through an edge (example Figure 3). In a network of messages connected with high cosine similarity, all similar messages will form a clique.

Finding Lexical Mutants Through Cliques: Computing cliques in large networks is a computationally expensive task. However, the method in which our network is constructed—drawing edges between nodes only when there is high cosine similarity—allows for a large number of connected components. For example, in the Farmers’ protest dataset, we had a total of 799K messages (nodes). Out of which 301,342 nodes had at least one edge. The resulting network had around 2.1K connected components. Connected components make for completely disjoint subgraphs and provide a much more computationally affordable space to find cliques. In fact 1,857 of the total components were also perfect cliques, suggesting that connected components could be a good approximation for finding cliques in a network such as this. A similar approach has proven successful in finding coordination networks in White Helmets (Pacheco, Flammini, and Menczer 2020). We consider each clique found with this method as a single campaign with lexical mutants.

Results: Lexical Mutant Amplification Campaigns

In total, we find 2,558 amplification campaigns relevant to Farmers’ protests spanning over 231,896 messages and 1,268 campaigns in the CAA dataset spanning over 146,465 messages. In the results, we only include campaigns with at least 10 messages with no two messages shared by the same Twitter handle, Facebook group, or page. While it is easier to identify and exclude multiple similar messages posted by a Twitter handle, it is challenging to establish this in the Facebook data, given that we don’t have user-level information on Facebook groups or page posts. For every clique, we only consider one message per Facebook group or page. It is possible that we are still considering messages posted by the same user in different Facebook groups, inflating the sizes of the cliques. To find out the extent of this, we manually analyze the authorship distribution for Facebook data points in a random sample of 1000 cliques. We found that 78% of the Facebook messages in a clique come from different users.

An average amplification campaign has 56 messages posted by different users (Figure 4 (b)). We found 219 campaigns with at least 100 messages from different user accounts and the largest campaign in the dataset contains 1,232 messages. We find that 29% of all messages in the Farmers’ protests dataset and 21% of all messages in the CAA dataset were part of an amplification campaign. Moreover, on average, 34% of the messages in the amplification campaigns are unique lexical mutations (Figure 4 (a)). The rest of the messages are copy-pastas of different lexical mutants, indicating that including lexical mutants helps in identifying larger amplification campaigns than simple copy-pastas.

Evaluating Identified Amplification Campaigns: How accurately does the methodology described above identify political amplification with lexical mutations? To evaluate, we analyze a random sample of 200 cliques (5% of all the cliques found) in the dataset and label for the similarity of the messages in the clique. All randomly sampled cliques were labeled by two annotators. We use a conservative labeling scheme— 1 if *all* messages in the clique satisfy our

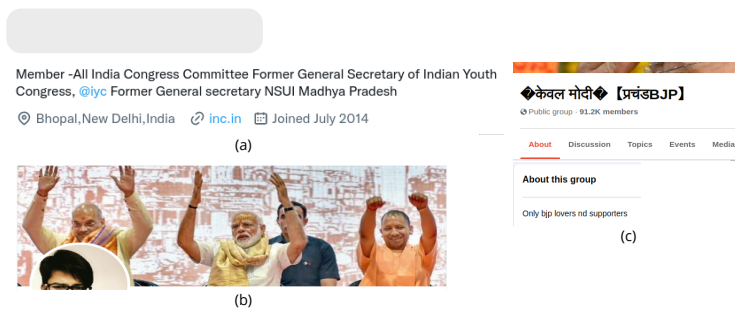


Figure 5: Examples of account bio, pictures, and descriptions in the dataset

definition of lexical mutations (see Introduction) and 0 if even one message is semantically dissimilar to the rest of the clique. Particularly, we labeled a clique as 1 only if *all* the messages in the clique contained the same underlying claim. Specifically, we looked for the similarity of topic, opinion, and argumentation while assessing the messages.

The annotators agreed on 197 out of 200 clique labels, resulting in 98.5% agreement. For final labels, we marked cliques as '1' where *both* annotators agreed on the label. We consider all disagreements as '0'. With this adjustment, we find that 191 cliques out of 200 sampled (95.5%) satisfy our criteria for lexical variants, indicating that our methods can identify amplification campaigns with lexical mutants with high confidence.

RQ2: Characterizing Hidden Amplification Campaigns

In RQ1, we identified 2,558 amplification campaigns relevant to Farmers' protests and 1,268 campaigns around Citizenship Amendment Act (CAA). While the previous literature has largely focused on studying political influence by the Indian right-wing, in this question we examine the extent of hidden amplification across the political spectrum in India. Moreover, we also analyze the expanse of political amplification across platforms and evaluate dominant narratives. Toward this goal, we first identify the political leaning of the accounts involved in the amplification campaign and measure the use of political amplification on Facebook and Twitter across different parties. We start by outlining our process for identifying the political leanings of social media accounts.

Labeling Accounts with Political Leanings

The amplification campaigns detected in the earlier research questions, spread over nearly 40K social media accounts. Currently, there is only one large-scale dataset—NivaDuck—by Panda (Panda et al. 2020) that records the political leanings of 18,500 Twitter accounts. However, only 368 accounts from NivaDuck overlap with the accounts in our dataset. Hence, we manually analyze the accounts in our dataset and record their political leanings. To better manage labeling resources, we focus only on repeat offenders—accounts that repeatedly participate in different amplification campaigns. In total, we recorded leanings of 493 Face-

book groups or pages and 1,631 Twitter accounts that were involved in 5 or more amplification campaigns. We want to note that for most cases, specifically for 97.2% Facebook accounts and 87.3% Twitter handles, the accounts had clear self-declarations of their political leaning, leaving little room for interpretation. For the rest of the accounts, we used a combination of manual and quantitative analysis to infer the political leanings. We describe both of these steps in detail next.

Labeling Accounts Based on Metadata:

- **Twitter:** We first cross-referenced the Twitter handles with the NivaDuck dataset (Panda et al. 2020) and borrowed the labels readily available. To label the remaining Twitter handles, we first look for cues in the Twitter bio, username, profile image, and background image for explicit political party affiliation. For example, the Twitter profile in Figure 5 (a) explicitly declares an association with the Indian National Congress (INC) political party. In some cases, the accounts also signal political affiliation through profile or Twitter background pictures, as shown in 5 (b). We were able to record definitive leanings of 87.3% of Twitter handles through the account metadata.
- **Facebook:** Assigning political leanings to Facebook groups of pages was significantly easier than Twitter. Our dataset contains 493 Facebook groups and pages that host messages used in at least 5 or more amplification campaigns. We label political leanings based on the group's name, description, rules, and recent posts. We find that in most cases, Facebook groups and pages had a clear political affiliation signaled either through group name or description. For example, the Facebook group in Figure 5 (C) is restricted only to the BJP supporters. Over 97.2% of Facebook, accounts had clear political affiliations represented in the metadata. Next, we explain our steps for assigning political leanings to the accounts with information beyond the account metadata.

Labeling Accounts Using Posts and Further Validation:

For the accounts that do not signal explicit political affiliation through metadata, we read through the 20 most recent tweets or Facebook posts and also consider the messages in the amplification campaigns associated with the account. Specifically, the first author of this paper looked at whether

the accounts posted content supporting specific political parties or politicians, policies, or events relevant to specific political parties. We were able to assign leanings to all of the remaining Facebook groups and 182 of the 207 remaining Twitter handles. The unlabeled Twitter handles are removed from the downstream analysis.

To further validate the manual labeling, we look at the accounts' network in amplification campaigns. In this network, two accounts are connected together if they ever participate in the same amplification campaign. We next use the label propagation algorithm (Cordasco and Gargano 2010) to infer the leanings of the accounts that did not have explicit affiliation in the metadata (See "Labeling accounts based on metadata"). We further compare the labels generated with label propagation and manual annotation and find 86% overlap. This can provide for external validation of the manual annotations efforts. We further discuss the challenges and limitations of manual labeling in the limitations section.

Political Leanings Labeling Scheme: The annotation process described above, resulted in the following labeling scheme. We focus on three political parties—BJP, INC, and AAP—that had the largest number of social media accounts in the NivaDuck dataset (Panda et al. 2020). We also include % accounts belonging to each leaning in brackets:

- **BJP affiliates (17%):** Accounts that explicitly signal affiliation with the ruling right-wing political party. Many of these accounts are created in support of the leading BJP politicians and the 2024 general elections. Widely protested Farmers' bills and Citizenship Amendment Act (CAA) were advocated under BJP leadership and support.
- **BJP supporters (35%):** Accounts that did not explicitly affiliate with the BJP but frequently posted content in support of BJP's mission and politicians.
- **INC affiliates (2%):** Accounts that openly affiliate with the leading opposition and Indian left-wing party—Indian National Congress (INC). INC joined the general opposition to Farmers' bills and CAA.
- **INC supporters (10%):** Accounts that do not clearly affiliate with INC but post content supporting INC politicians and mission.
- **AAP affiliates (0.6%):** Accounts affiliated with Aam Admi Party (AAP). AAP is often referred to as the "third opposition" with appeals to the common-man identity in India. AAP also joined in the opposition to Farmers' bills and CAA.
- **BJP opposition (32%):** A large chunk of the accounts that while not affiliating with any specific political party, explicitly oppose BJP politicians and policies.

Results RQ2: Characteristics of Amplification Campaigns

In the previous analysis, we identified political amplification campaigns around Farmers' protests and CAA and labeled the political leanings of the accounts participating in the campaigns. Here we discuss partisanship, social media

platform use, and narratives in the most widespread amplification campaigns. Figure 6 displays various views of the network of users involved in the amplification campaigns. Nodes are user accounts and edges connect two users that are involved in the same amplification campaign. Various network views are colored differently to represent user learning and social media platforms. For example, in (a), (b), and (e) node colors correspond to political leaning while in (c), node colors are based on the social media platform of the account.

RQ2a Political Amplification and Partisanship: Figure 6 (e) represents the network of users involved in political message amplification across both, Farmers' protests and CAA. Overall, 38% amplification campaigns spread through BJP accounts, 40% spread through BJP opposition accounts, and 22% amplification campaigns are propagated through accounts of other parties such as INC and AAP. More granular results across events are present in Table 2. Overall, we find that accounts all across the Indian political spectrum participate in political amplification in an equitable way.

After analyzing the networks of users that repeatedly participate in the amplification campaigns, we observe that BJP accounts are strongly clustered together without having edges to any other political leanings. On the other hand, non-BJP accounts (BJP opposition accounts with no specific party affiliation, INC accounts, and AAP accounts) all participate in common amplification campaigns.

RQ2b Political Amplification Across Platforms: Figure 6 (c) displays the network of user accounts involved in amplification with node color representing the social media platform (blue: Twitter, green: Facebook). In sum, 65% of the detected campaigns reside only inside Twitter, while 5% are confined to Facebook. 30% of the amplification campaigns spread across both platforms.

RQ2c Prominent Political Claims in Amplification Campaigns: We further analyze the dominant claims in amplification campaigns across different political leanings. Some examples of amplification narratives are described in Figure 6 (d) and Figure 6 (e). We manually analyzed the narratives in 63 large amplification campaigns—campaigns with more than 500 messages. Specifically, the first author of this paper manually noted the core message in each of the 63 amplification campaigns by analyzing the contents of the lexical mutants inside the campaign. During this process, we noticed that the largest amplification campaigns by BJP and non-BJP accounts could be considered direct antitheses of each other. For example, the largest anti-CAA campaign by non-BJP accounts claims that BJP is oppressing minorities through CAA. Whereas, the largest amplification campaign by BJP accounts claims that India is continuing their historical legacy of providing refuge to minorities through CAA. Similarly, non-BJP accounts attempt to depict Farm Bill supporters as traitors whereas BJP accounts criticize the opposition to Farm Bill for its political opportunism.

To further systematically analyze these contrasting political campaigns, we manually paired each amplification campaign with its counterargument, wherever possible. Out of the 63 large campaigns, we found counterarguments for 20

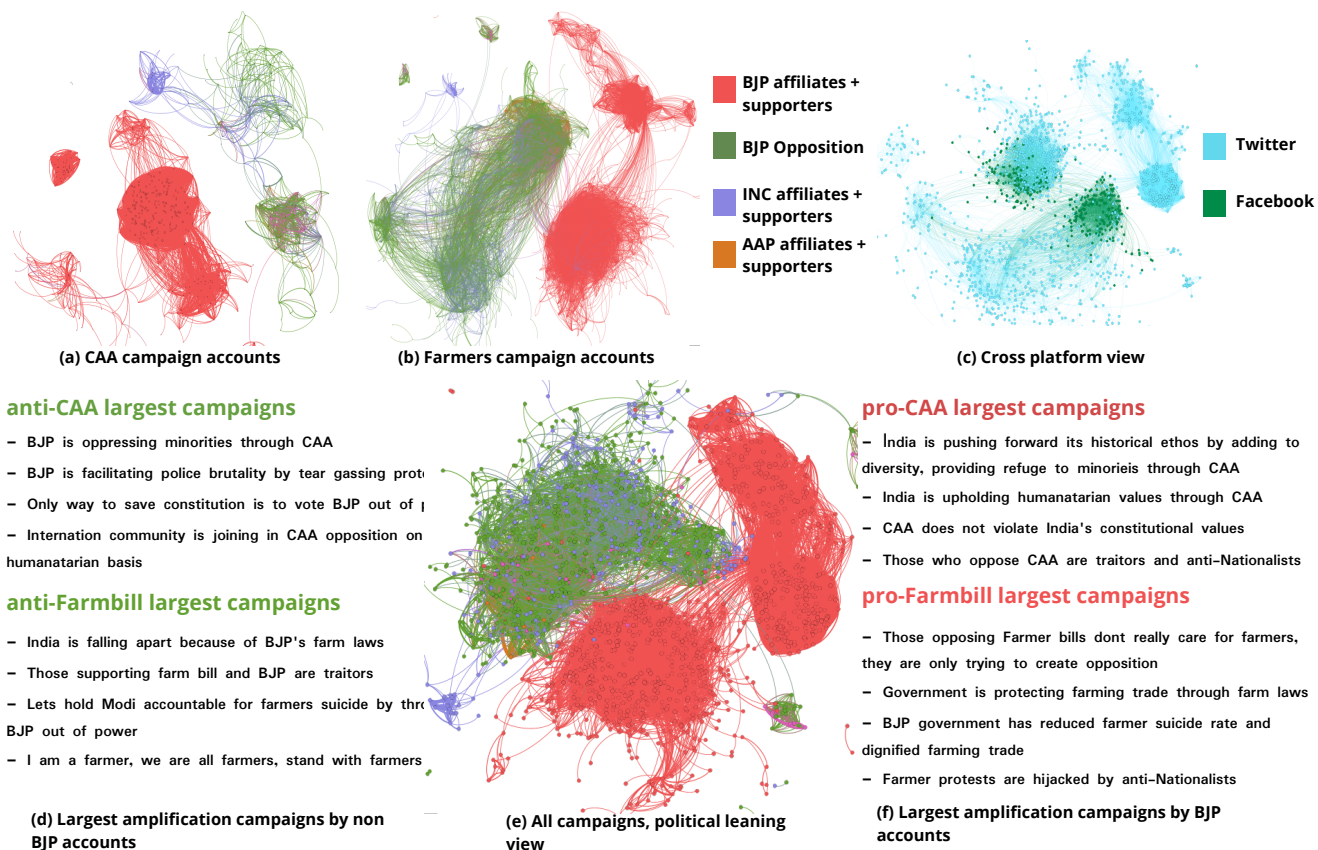


Figure 6: (a) and (b) represent the network of users involved in political amplification relevant to CAA and Farmers’ protests respectively. (e) is the overall user network across both events. Nodes are the user accounts and colors represent political leaning. The precise percentage of accounts belonging to different political leanings is reported in Table 2. (c) represents the combined network of users across Facebook and Twitter. The node color distinguishes Facebook users (green) from Twitter users (blue). (d) and (f) list messages from the largest amplification campaigns from (d) non BJP accounts and (e) BJP accounts

using the core campaign message recorded before. Table 1 records the core messages of contrasting political amplification campaigns.

We further analyze the timelines of contrasting campaigns. We find that the claims and counter-claims follow precise temporal orders. For example, the bulk of messages suggesting that BJP is threatening minorities through the CAA (Figure 7) was posted by non-BJP accounts around December 15th, 2019. Amplification campaigns by BJP contesting this claim were spread during the subsequent days claiming that CAA promotes diversity and the historical legacy of providing refuge to minorities. Similar trends followed in the 17 other contrasting campaigns. Overall, the temporal order political messaging suggests that amplification campaigns are possibly being used as devices to counter the narratives from both sides of the Indian political spectrum.

Discussion

In this paper, we analyzed how political parties leverage amplification campaigns with lexical mutations to popularize contrasting political stances on Indian social media. Focus-

ing on two key recent events —Farmers’ protests and the introduction of the Citizenship Amendment Act (CAA)—we first identified over 3.8K amplification campaigns across Facebook and Twitter. Next, we characterize the use of amplification campaigns across multiple political parties in India and multiple and social media platforms. Our results provide an updated understanding of political amplification by looking beyond the popularly studied platform Twitter and by considering the multi-party political landscape of India. Below we discuss some of our findings in detail.

Amplification With Lexical Mutants: Moving Beyond Copy-Pasta

Previous studies have revealed the presence of copy-pasta political campaigns both in India (Jakesch et al. 2021) and the West (François, Nimmo, and Eib 2019). We focus on surfacing large-scale campaigns that not only include copy-pastes but also detect slight changes in phrasing and language—lexical mutations. The analysis of large amplification campaigns detected in RQ1, clearly indicates that the messages have a similar underlying template. Here we discuss two key questions raised by our findings: do users

Core arguments by BJP affiliated accounts (number of large campaigns)	Core arguments by non-BJP affiliated accounts (number of large campaigns)
BJP is fostering diversity and humanitarian values and historical legacy (2)	Modi's policies are anti-humanitarian and shameful for Indians (1)
Government is promoting diversity through CAA (1)	BJP is threatening minorities through CAA (1)
CAA is designed respecting the Indian constitution (1)	CAA is unconstitutional (1)
Government is preserving the culture of northeast India with CAA (1)	Modi is isolating people from northeast India (1)
It is the duty of every Indian to support the government we voted for (1)	We should stand up to Modi and his bullying (1)
Farmers' protests are disrupting daily businesses and peace (1)	Government is terrorizing peaceful farmers' protests (1)
Opposition parties don't care about farmers. They want to oppose Modi (1)	Anyone who is a patriot should oppose farmers' bills (1)
Farmers' protests are hijacked by paid protestors (1)	Farmers don't have IT cells or paid supporters (1)
Modi effect: Modi has made farmers' lives better (2)	All kinds of farmers are opposing farm bills

Table 1: Table recording some of the arguments and their counterparts identified through manual analysis of the large (>500 messages) amplification campaigns. We identified a total of 9 arguments and counterarguments, spanning 20 distinct amplification campaigns. The left column presents narratives amplified by BJP affiliated accounts and the right column presents the counter-narrative identified through manual analysis. The numbers in parenthesis represent the number of amplification campaigns with the specified core narrative.

accounts from	CAA	Farmers' protests	Overall	TW	FB
BJP	47%	35%	38%	37%	39%
BJP Opposition	35%	44%	40%	41%	40%
Other parties	18%	21%	22%	22%	21%

Table 2: Participation in political amplification by accounts of different political leanings. Here we report on the proportion of accounts with different leanings in different political events and platforms.

deliberately tweak the messages for amplification and why?

Hidden Amplification and Platform Manipulation Policies: In one of the previous studies (Jakesch et al. 2021), researchers found that local political leaders in India seek out participation through WhatsApp groups to spread messages from “Tweet banks”. One such message quoted in their paper explicitly instructs users to alter the wording in the template: “*Note-Please don't just copy-paste the sample tweets, please alter it a bit.*” (Jakesch et al. 2021) Pg. 9]. Our results indicate that users might be getting these types of instructions on a larger scale for campaigns of different political parties. Moreover, we argue that this kind of hidden amplification might be designed to bypass the platform manipulation and spam policies⁴. For example, on Twitter, copy-paste campaigns are subject to review, and the platform is committed to reducing the visibility of such content. In fact, repeatedly violating copy-paste rules is considered a severe violation of the platform manipulation policy. In our study, we identify thousands of user accounts that repeatedly participate in amplification campaigns (Figure 6). The fact that we were able to detect large-scale campaigns and repeat offender users, who are still active on the platforms, may suggest that hidden amplification may be proving effective

⁴<https://help.twitter.com/en/rules-and-policies/copy-paste-duplicate-content>

in evading platform governance policies. In the future, a more standardized and scalable approach such as the one suggested in RQ1 could be required to counter large-scale platform manipulation. For example, our entire pipeline to detect amplification campaigns with lexical mutants could be used in real-time to detect political amplification. Detecting messages involved in amplification could help platforms limit their visibility soon after they are published, and also keep track of repeat offenders across different political events.

Indian Political Amplification Campaigns and Reactionary Politics

Our study offers a unique look into the political influence exerted by accounts from various political leanings in India. By considering the multi-party political spectrum in India, we provide an essential context to previous research focusing on only one political party. More importantly, our results reveal the reactionary nature of political influence in India (Figure 7). Specifically, by investigating the temporal order of contrasting amplification campaigns, we observe that large-scale amplification campaigns arise in the form of arguments and counter-argument between different political viewpoints. Researchers studying the role of copy-paste in reactionary politics around neo-Nazism in the United States propose that such reactionary politics often gives rise to radicalization (Topinka 2022). Specifically, influencing public arguments and counter-arguments through amplification can harm democratic equality and can create an exaggerated sense of democratic divide in the public eye. Our methods can also be extended to study the temporal patterns in political amplification campaigns outside of India.

Ethical Considerations

Adhering to AAAI code of conduct and ethics guidelines, here we discuss the stakeholders, harm, privacy, and confidentiality in our research. First, we acknowledge that social media users and social media platforms are primary stakeholders in this research. With this work, we intend to contribute insight into both, the expanse of political amplification

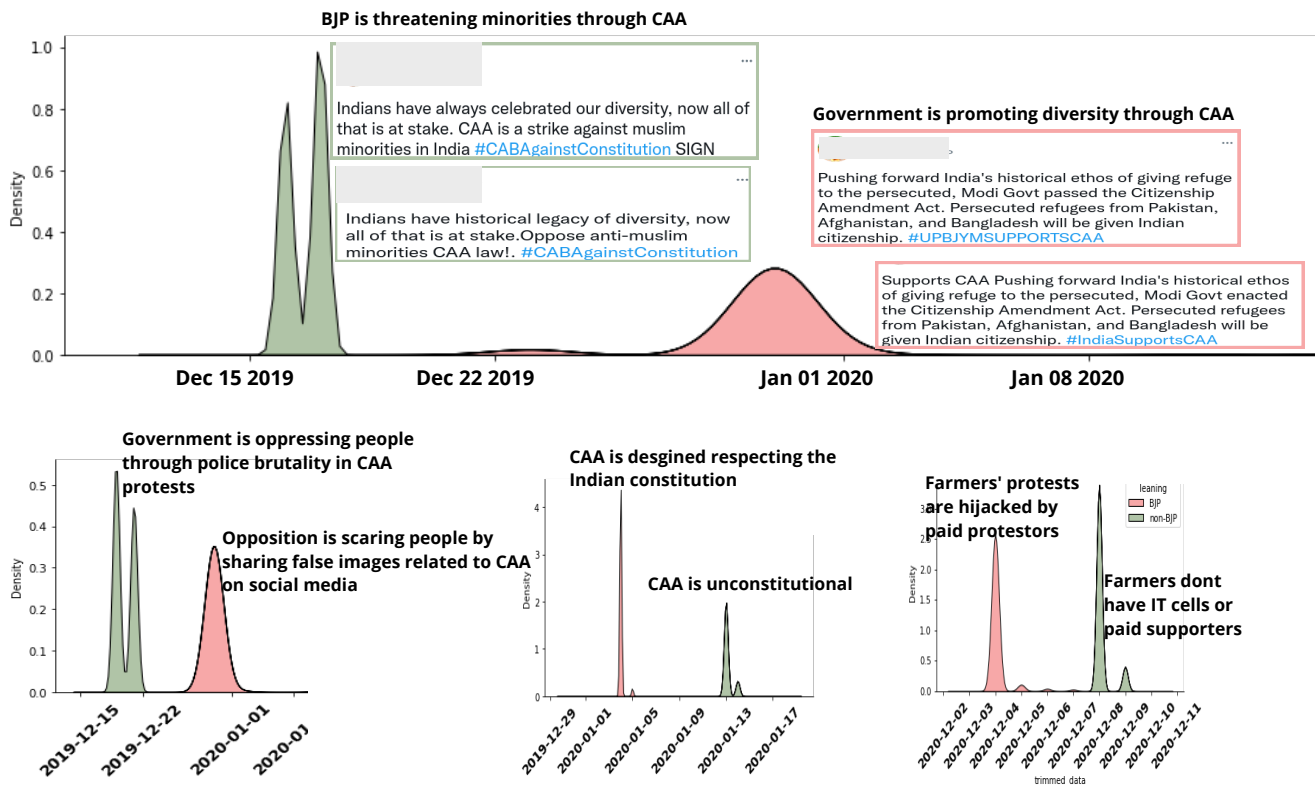


Figure 7: Timelines of largest amplification campaigns. In each plot, x-axis represents the actual calendar dates, and y-axis displays the proportion of messages in the amplification campaign posted on that date. We specifically analyze the timelines of the largest campaigns on the opposite sides of the political spectrum that make contrasting claims.

tion on social media as well as computational ways to detect political amplification which could be useful in platform governance. This research is retrospective and involves no interaction with the users studied. Moreover, we do not include any user information in the paper and blur out user handles from shared figures. This limits the harm and maintains user privacy. Moreover, all the data used in this paper is publicly available preserving the data confidentiality on social media.

Limitations

Our work primarily looks at political amplification campaigns with lexical mutants, without including retweets or quote tweets. Incorporating retweets and other signals, along with lexical variations, or image or video coordination can provide a more comprehensive picture of political amplification in India. Moreover, as mentioned while discussing RQ1 results, we can not establish the authorship of each and every post made in Facebook groups or pages. While only 5% of amplification campaigns strictly reside inside Facebook, and while our manual robustness check reveals that 78% of messages in the campaign come from different users, we acknowledge that having authorship information could enhance our results. Currently, there is no way to obtain this information through any official API, and web crawling for author information on Facebook posts could pose serious ethi-

cal challenges. While calculating multi-lingual embeddings we used the LASER model, which is trained and evaluated on translated pairs of texts. However, our current methodology does not account for transliterations, which can be incorporated in the future to find larger campaigns. Moreover, our process for identifying the narratives and counter-narratives involved only one annotator with proficiency in multiple Indian languages and familiarity with the Indian political landscape. The annotation performance could be made better by involving more expert annotators. Finally, while it is easier to detect which messages are part of a political amplification campaign, we still cannot determine whether the message reflects the user’s genuine political opinion.

Conclusion

In this work, we analyze Indian political amplification campaigns with lexical mutations. By modeling messages in national political events, we uncover over 3.8K political amplification campaigns across languages, social media platforms, and political parties in India. Our findings enable future research that can explore the repeat offender accounts in-depth and assess their legitimacy.

References

Ahmed, S.; Jaidka, K.; and Cho, J. 2016. The 2014 Indian elections on Twitter: A comparison of campaign strategies of

- political parties. *Telematics and Informatics*, 33(4): 1071–1087.
- Akbar, S. Z.; Panda, A.; Kukreti, D.; Meena, A.; and Pal, J. 2021. Misinformation as a Window into Prejudice: COVID-19 and the Information Environment in India. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3): 1–28.
- Artetxe, M.; and Schwenk, H. 2019. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7: 597–610.
- Bureau, O. W. 2021. 'Manipulated Media' Controversy: Government Vs Twitter. <https://www.outlookindia.com/website/story/india-news-manipulated-media-controversy-government-vs-twitter/383352>. (Accessed on 01/14/2023).
- Cinelli, M.; Cresci, S.; Quattrociocchi, W.; Tesconi, M.; and Zola, P. 2022. Coordinated inauthentic behavior and information spreading on twitter. *Decision Support Systems*, 113819.
- Clark, J.; and Holton, D. A. 1991. *A first look at graph theory*. World Scientific.
- Cordasco, G.; and Gargano, L. 2010. Community detection via semi-synchronous label propagation algorithms. In *2010 IEEE international workshop on: business applications of social network analysis (BASNA)*, 1–8. IEEE.
- Das, A.; and Schroeder, R. 2021. Online disinformation in the run-up to the Indian 2019 election. *Information, Communication & Society*, 24(12): 1762–1778.
- Dash, S.; Arya, A.; Kaur, S.; and Pal, J. 2022. Narrative Building in Propaganda Networks on Indian Twitter. In *14th ACM Web Science Conference 2022*, 239–244.
- Dash, S.; Mishra, D.; Shekhawat, G.; and Pal, J. 2021. Divided We Rule: Influencer Polarization on Twitter During Political Crises in India. *arXiv preprint arXiv:2105.08361*.
- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Edingo, D. B. 2021. Social media, public sphere and counterpublics: An exploratory analysis of the networked use of Twitter during the protests against the Citizenship Amendment Act in India. *The Journal of Social Media in Society*, 10(2): 76–101.
- François, C.; Nimmo, B.; and Eib, C. S. 2019. The IRA copy-paste campaign. *Graphika*, okt.
- Frenkel, S.; and Alba, D. 2021. In India, Facebook Struggles to Combat Misinformation and Hate Speech - The New York Times. <https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html>. (Accessed on 04/21/2022).
- Gallagher, R. J.; Doroshenko, L.; Shugars, S.; Lazer, D.; and Foucault Welles, B. 2021. Sustained online amplification of COVID-19 elites in the United States. *Social Media + Society*, 7(2): 20563051211024957.
- Hristakieva, K.; Cresci, S.; Da San Martino, G.; Conti, M.; and Nakov, P. 2022. The spread of propaganda by coordinated communities on social media. In *14th ACM Web Science Conference 2022*, 191–201.
- Jaffrelot, C. 2015. The Modi-centric BJP 2014 election campaign: New techniques and old tactics. *Contemporary South Asia*, 23(2): 151–166.
- Jakesch, M.; Garimella, K.; Eckles, D.; and Naaman, M. 2021. # Trend Alert: How a Cross-Platform Organization Manipulated Twitter Trends in the Indian General Election. *arXiv preprint arXiv:2104.13259*.
- Joshi, R.; Goel, P.; and Joshi, R. 2020. Deep learning for hindi text classification: A comparison. In *International Conference on Intelligent Human Computer Interaction*, 94–101. Springer.
- Leskovec, J.; Backstrom, L.; and Kleinberg, J. 2009. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 497–506.
- Pacheco, D.; Flammini, A.; and Menczer, F. 2020. Unveiling coordinated groups behind white helmets disinformation. In *Companion Proceedings of the Web Conference 2020*, 611–616.
- Panda, A.; Gonawela, A.; Acharyya, S.; Mishra, D.; Mohapatra, M.; Chandrasekaran, R.; and Pal, J. 2020. Nivaduck-a scalable pipeline to build a database of political twitter handles for india and the united states. In *International Conference on Social Media and Society*, 200–209.
- Rajgarhia, S. 2020. *Media Manipulation in the Indian Context: An Analysis of Kashmir-related Discourse on Twitter*. Ph.D. thesis, Harvard University.
- Reimers, N.; and Gurevych, I. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. 4512–4525.
- Sambhav, K.; and Ranganathan, N. 2022. How a Reliance-funded firm boosts BJP's campaigns on Facebook — Business and Economy News — Al Jazeera. <https://www.aljazeera.com/economy/2022/3/14/how-a-reliance-funded-company-boosts-bjps-campaigns-on-facebook>. (Accessed on 04/23/2022).
- Schoch, D.; Keller, F. B.; Stier, S.; and Yang, J. 2022. Co-ordination patterns reveal online political astroturfing across the world. *Scientific reports*, 12(1): 1–10.
- Simmons, M.; Adamic, L.; and Adar, E. 2011. Memes online: Extracted, subtracted, injected, and recollected. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 353–360.
- Singh, M. 2021. India objects to 'manipulated' label on politicians tweets; asks removal of reference to 'Indian variant' of coronavirus — TechCrunch. <https://techcrunch.com/2021/05/21/india-twitter-politicians-tweets-manipulated/>. (Accessed on 01/14/2023).
- Sohal, S.; and Kaur, H. 2018. A content analysis of YouTube political advertisements: evidence from Indian parliamentary elections. *Journal of creative communications*, 13(2): 133–156.
- Tiwari, S. 2019. Elections 2019: 45 million new, young voters could play a key role in 2019 – and they want jobs. <https://scroll.in/article/913411/data-check-45-million-new-young-voters-could-play-a-key-role-in-2019-elections>. (Accessed on 03/09/2022).

Topinka, R. 2022. The politics of anti-discourse: Copy-paste, the Alt-Right, and the rhetoric of form. *Theory & Event*, 25(2): 392–418.

Weber, D.; and Neumann, F. 2021. Amplifying influence through coordinated behaviour in social networks. *Social Network Analysis and Mining*, 11(1): 1–42.

Woolley, J. K.; Limperos, A. M.; and Oliver, M. B. 2010. The 2008 presidential election, 2.0: A content analysis of user-generated political Facebook groups. *Mass Communication and Society*, 13(5): 631–652.

Paper Checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper’s contributions and scope? Yes
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
 - (e) Did you describe the limitations of your work? Yes
 - (f) Did you discuss any potential negative societal impacts of your work? Yes
 - (g) Did you discuss any potential misuse of your work? Yes
 - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? NA
 - (b) Have you provided justifications for all theoretical results? NA
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
 - (e) Did you address potential biases or limitations in your theoretical framework? NA
 - (f) Have you related your theoretical results to the existing literature in social science? NA
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
 - (a) Did you state the full set of assumptions of all theoretical results? NA ■
 - (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA ■
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
 - (f) Do you discuss what is “the cost” of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
 - (a) If your work uses existing assets, did you cite the creators? Yes
 - (b) Did you mention the license of the assets? NA
 - (c) Did you include any new assets in the supplemental material or as a URL? NA
 - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? NA
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see ?)? NA
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see ?)? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
 - (a) Did you include the full text of instructions given to participants and screenshots? NA
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
 - (d) Did you discuss how data is stored, shared, and de-identified? NA