On the Influence and Political Leaning of Overlap between Propaganda Communities

ANIRBAN SEN, Microsoft Research, India
SOUMYASIS GUN, IIIT Hyderabad, India
SOHAM DE, Microsoft Research, India
JOYOJEET PAL, University of Michigan, Ann Arbor, USA

Social media offers increasingly diverse mechanisms for the distribution of motivated information, with multiple propaganda communities exhibiting overlaps with respect to user, content, and network characteristics. This has particularly been an issue in the Global South, where recent work has shown various forms of strife related to polarizing speech online. It has also emerged that propagandist information, including fringe positions on issues, can find its way into the mainstream when sufficiently reinforced in tone and frequency, some of which often requires sophisticated organizing and information manipulation. In this study, we analyze the overlap between three events with varying degrees of propagandist messaging by analyzing the content and network characteristics of users leading to overlap between their users and discourse. We find that a significant fraction of users leading to overlap between the three event communities are influential in information spread across the three event networks, and political leaning is one of the factors that helps explain what brings the communities together. Our work sheds light on the importance of network characteristics of users, which can prove to be instrumental in establishing the role of political leaning on overlap between multiple propaganda communities.

CCS Concepts: • **Applied computing** → *Computers in other domains*; • **Information systems** → **Information retrieval**; **Social networks**.

Additional Key Words and Phrases: propaganda, misinformation, social media, social network analysis, community overlap

ACM Reference Format:
Anirban Sen, Soumyasis Gun, Soham De, and Joyojeet Pal. 2023. On the Influence and Political Leaning of Overlap between Propaganda Communities. In . ACM, New York, NY, USA, 23 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The notion of social media polluting the global information environment with misinformation, propaganda and polarizing speech is inextricable from any broad conversation on the technology-mediated communication [83]. There is a massive and ever-growing body of work on the scope and scale of misinformation and hate speech online [13, 54], including work that specifically examines the prevalence and impacts of extreme speech in various parts of the Global South [4, 9, 11, 12, 47, 57, 61, 63]. Politicians and state apparati have been complicit and often effective in running coordinated disinformation campaigns, rumor, and propaganda, fanning social schism, well before scholarly work on disinformation achieved prominence in Global North [79]. Propaganda and extreme speech often lead to both short term consequences such as one-off violent events, and longer-term divisions in society [43, 77, 78]. The adverse impact

Manuscript submitted to ACM

of propaganda on society is amplified by the power social media provides to mainstream and digital celebrities in shaping the views of the general public [19, 25, 58, 60].

While propaganda can both be positive and negative depending on the valence of messages, we focus on negative propaganda on Twitter (currently 'X'), i.e., posts that aim to negatively project an ostensible or constructed antagonist entity that have been theorized in the political science literature as an 'enemy' within a broader political spectacle [28], one whose discrediting or construction as problematic is fundamental to the legitimacy of the actor that does the propagandist work. Research has shown that online propaganda communities (i.e. event communities which have a high presence of propagandist messaging) exhibit various forms of overlap, including a cultural convergence on political leaning, as well as on specific worldviews such as distrust of mainstream news, the government, or science. At the same time, these overlaps may occur irrespective of the political or cultural positions of the actors engaged, building in turn new narratives and reframing alignments [56, 70].

Building on such work, we study the intersection between three propaganda communities in India, and analyze the network and content characteristics of the bridge nodes – the user accounts that sit at the intersection of the three community pairs. Analysis of such overlaps is instructive in understanding the inflection points of polarizing online discourses, which in turn lend insight into the capture of media institutions [86]. Such capture, already documented in recent years in several parts of the global south including India[30], Indonesia [34], Pakistan [1], is recognized widely as a significant threat to political and communal harmony, yet the mechanics of such capture in terms of the online discourse is not adequately understood.

Using Twitter data, we study two signals of community overlap, namely *retweet overlap* and *hashtag overlap*. The former is defined as the overlap that occurs due to a user highly active in one event community retweeting another active user from the other community. The latter is defined as overlap that occurs due to a single tweet containing hashtags belonging to both event communities in an overlap pair. We term the users who lead to these two types of inter-community overlaps as *Bridge Users*. We specifically focus on the aspect of political alignment in polarizing content. The research question that we address is: *Does political leaning have a significant correlation with the overlaps between fringe propagandist networks on social media?* There are two axes along which we investigate this question: (A) The retweet network characteristics around bridge users, and (B) Their tweet content characteristics. The analysis of network characteristics involves studying the influence of bridge users in information spread, their political leaning, and the political leaning of their retweet network neighborhood. As part of the content analysis of tweets/retweets posted by the bridge users, we study the presence of political keywords and political entities mentioned in them.

We conducted this study in India as it has the third-highest installed user base for Twitter and has also been in the news over social media use in large part due to news coverage of polarization and political media use [22]. We examine three events that saw significant and organized online activity and propagandist messaging in this paper – the conversations around the Men's Rights Activism (#MRA), the online attacks against the Chief Justice of India (#NotMyCJI denoted as CJI henceforth) D.Y. Chandrachud for taking a number of liberal positions, especially on women's rights, and the Boycott Bollywood (BB) movement that gained momentum in the wake of controversies around the death of a prominent actor and calls for boycotting films/actors/individuals related to Hindi films (#boycottbollywood). The events' positions and arguments were eschewed by public figures and mainstream media. However, unlike other viral news cycle events in which the online activity is mirrored to some comparable level in the mainstream media, the discussion for these events largely existed on social media. Hence, the social media communities around these events are considered to be "fringe communities". We discuss more about the origin, causes, and target groups of these propagandist events in the Appendix.

Our qualitative analysis reveals that a significant fraction of tweets contain negative propagandist content for all three events. Subsequently, we find that: (A) A significant fraction of accounts influential in terms of probabilistic diffusion of information across the MRA-CJI, BB-MRA, and BB-CJI community retweet networks are bridge nodes, (B) These bridge accounts are significantly more aligned towards the currently ruling alliance in India as inferred from their retweet network neighborhood, when compared to the opposition, and (C) While most of the content posted by the bridge users is politically neutral, their retweet neighborhood exhibits significant political leaning.

The primary finding of this work is that a significant fraction of bridge users leading to overlap between multiple propaganda communities (with varying target groups) are influential in terms of information diffusion and are driven by their political leaning. Additionally, this political leaning is not always observable from the content posted by them, but evident from their retweet network characteristics. These findings hint towards the possibility of carefully orchestrated online activism in fringe propaganda communities where users leading to community overlap carefully refrain from posting political content, keeping the discourse highly relevant to the issue. The methods used in this work could be generalized to the study of any pair of events, three of which we cover here. Our work opens directions for future work towards studying the correlation of political leaning with the overlap between various other propaganda communities, and emphasizes on the need to dig deeper into the network characteristics of users leading to these overlaps.

2 RELATED WORK

Here we describe the previous studies done in the area of propaganda analysis, political ideology analysis, and community overlap on social media.

2.1 Propaganda Analysis in Social Media

Several earlier studies have touched upon the topic of propaganda analysis. We describe in this section this body of work along two methodological directions, namely propaganda detection using text classification techniques and that using network analysis methods. Studies on text classification-based methods used supervised techniques to predict propaganda in a multi-class [72] and binary classification set up [6]. These approaches used word n-grams, formal text representations like readability and writing style, and lexicon based methods for propaganda prediction. Martino et al. [21] proposed an extractive propaganda detection method using deep neural networks, which unlike the other studies, detected the technique of propaganda spread along with the text fragment responsible. Network analysis techniques for propaganda detection initially focused on the problem of node detection, i.e., these techniques classified a user node as propagandist or not in isolation [44], using supervised techniques on various user level features including network features [84, 85], textual content used [71], and profile information [5, 35, 48, 80].

However, detecting a propagandist node or user in isolation may not be sufficient in many cases as most social media movements exhibit coordinated activity and collaboration among a group of users. Researchers in this area thus eventually focused on detecting coordinated behavior [36] among a set of user nodes, rather than considering a node in isolation, using semi-supervised and unsupervised techniques [10, 27]. In this direction, studies on detecting suspicious user connectivity patterns [17, 51, 64] and temporal tweeting/retweeting patterns [15, 26, 55] have been developed. Additionally, there are other works that target propaganda analysis on social media using both content and network analysis based methods [14, 29, 32].

Our work is motivated by these previous studies, and attempts to look at propaganda analysis through a different lens. Instead of focusing on detection/classification of propaganda or propagandist users, we study how different events consisting of some level of propagandist messaging overlap. Specifically, we focus on the relationship of propaganda community overlaps with political leaning around the content and network characteristics of the bridge users.

2.2 Political Ideology in Social Media

Social media has also emerged as a powerful tool in politics - often being used by politicians to maintain their online image [65], appeal to their core supporter bases [75] and influence larger political outcomes [41]. Prior research has shed light on several strategies adopted by political influencers to weaponize social media, such as polarising online audiences [53], associating with digital influencers [46] and other nationalistic entities such as defence veterans [23]. Given such a political online environment, detecting one's political ideology becomes a difficult yet important task. Popular methods try to infer political ideology from one's opinions on common topics using opinion-aware knowledge graphs [16], from one's digital trace of written texts or tweets using recurrent neural networks (RNNs) [38], using mass-media coverage [74] and even from social media images [82]. Stance detection [3] - the task of detecting whether one is for, against, or neutral with respect to any position has also been used in the context of political debates to infer one's political ideology [45]. Finally, political leaning, a more easily quantifiable attribute has also been actively studied using data such as tweets/retweets [81], information diffusion [40] and partisan engagement with news articles [88].

The challenge with most of these approaches is the requirement of a significant amount of ground truth data, i.e., documents/tweets apriori annotated for ideology or political leaning. For this reason, we infer the political leaning of bridge user accounts using their network neighborhood information on Twitter. The details of this approach is described in the Methodology section.

2.3 Community Overlap Analysis in Social Media

Studies such as [39] have shown that people with similar intentions tend to form communities within online social networks, such as Twitter. Even though Twitter wasn't originally designed to support the evolution of online communities, through analysis of one's Twitter networks, Gruzd et al.[31] found that real communities are able to thrive on Twitter despite the lack of in-person contact. Such online communities may also evolve over crisis events as studied in [52] with the particular example of natural disasters in Japan. Much of recent research has focused on detection such online communities, the most popular approach being via *modularity* [62] of the underlying Online Social Networks (OSNs) or the Retweet Overlap Networks (RONs) [33]. Consequently, a significant body of work [37, 50, 68] has tried to improve these detection measures by making them more efficient for use in real-time applications. Our study has a focus on the phenomenon of overlap between such communities - which has been shown to be crucial to their survival [89]. Using a novel model of overlapping communities, Cui et al.[20] proposed a method to infer overlapping communities a given vertex belongs to. Research along similar lines has also yielded in novel, efficient algorithms for detecting overlapping communities in both static [87] and dynamic networks [8].

Our work is motivated by the study by McQuillan et al. [56] who study the temporal overlap between apparently unrelated social media communities through common hashtags, topics of discussion, users, and URLs. We specifically focus on the political leaning of users enabling community overlap between three social media event communities based on their content and network characteristics.

3 DATA AND METHODS

We study three Indian events, namely MRA, CJI, and BB in this paper. While there exist several event communities with significant amount of propaganda on social media, we specifically chose these three events for specific reasons. MRA is

Movement	Set	Hashtags
CJI	S1	NotMyCJI, JudiciaryMustApologise
	$H1$	#notmycji, #genderbiasedlaws, #malegenocide, #chandrachud, #protestatun, #marriagestrike, #boycottmarriage, #498a, #mentoo, #fakecases, #feminismiscancer, #legalgenocide, #2ndclasscitizens, #legalterrorism, #legalextortion, #maritalrape, #wokejudge, #dowrylawmisuse, #feminism, #womenempowerment, #judiciarymustapologise
	H_{unique}	#notmycji, #chandrachud, #protestatun, #legalgenocide, #2ndclasscitizens, #wokejudge
MRA	S1	MaritalRape, boycottmarriage, marriagestrike, MenToo, FeminismIsCancer, womanisaburden, UnconstitutionalCRPC, MensRightsMatter, ABLANARI,AblaNaariSyndrome
	$H1$	#marriagestrike, #boycottmarriage, #mentoo, #maritalrape, #genderbiasedlaws, #fakecases, #498a, #ablanari, #feminismiscancer, #legalextortion, #legalterrorism, #dowrylawmisuse, #womenempowerment, #menscommission, #niruparoy, #saveinnocentmen, #fakecases_498a_dv_125_377_376_354, #femini, #malegenocide, #feminism #womanisaburden, #unconstitutionalcrpc, #mensrightsmatter
	H_{unique}	#ablanari, #menscommission, #niruparoy, #saveinnocentmen, #fakecases_498a_dv_125_377_376_354
Common to CJI - MRA	H_{common}	#boycottmarriage, #malegenocide, #marriagestrike, #498a, #legalextortion, #maritalrape, #feminism, #dowrylawmisuse, #feminismiscancer, #womenempowerment, #genderbiasedlaws, #fakecases, #legalterrorism, #mentoo
BB	S1	#BoycottBollywood, #BoycottBollywoodCompletely, #BoycottBollywoodForever
	H1	#boycottbollywood, #boycottvollywoodvompletely, #boycottbollywoodforever, #boycottalia, #boycottgangubai,#boycottpathan
	H_{unique}	#boycottbollywood, #boycottbollywoodcompletely, #boycottbollywoodforever, #boycottalia, #boycottgangubai,#boycottpathan

Table 1. Hashtags corresponding to each event. We did not find significant evidence of hashtag overlaps in the BB-MRA and BB-CJI pairs. *Note: Hashtags in* $H1$*,* H_{unique}*, and* H_{common} *are lower-cased.*

unique as an event since it is one of the very few events that has over time spawned several other propaganda events (e.g., CJI and #deppvsherd (discussions on a trial held in Virginia in 2022 related to allegations of defamation between formerly married prominent American actors)) with varying levels of online activism. CJI was selected for our study since it that spawned from MRA and has a natural overlap with it. Finally, BB is a long running social media event with known propagandist online discourse, which is the progenitor of several other events like #justiceforssr (a trending hashtag on the demise of a Bollywood actor and the conspiracy theories around it). While the target groups of the three events considered are not chiefly political – MRA talks about Men's Rights, CJI is an issue related to the judiciary and men's rights, and BB primarily targets the Hindi film fraternity – we investigate if political alignment has any relationship with the overlap between the three event pairs. In this section, we discuss the details of the data used and methodology followed for our work.

Property	MRA-CJI	BB-MRA	BB-CJI
No. of Nodes	1853	4140	2640
No. of Edges	86765	189785	100698

Table 2. Retweet Network statistics for the three event pairs, considering only the core users. The largest connected component was considered for the study.

3.1 Tweet Collection

We collected publicly available tweets for our study using the Twitter v2 API, in which we ensure anonymity of users before utilising them for exhibits/reproducibility of our study. For the three event communities, we defined a set of frequently occurring phrases and hashtags ($\mathcal{S}_1$) through a manual exploration of 100 tweets for each event, listed in Table 1, using which we collected all tweets for the events for the timelines considered. Next, we expanded this set by following two steps: A set of manually selected exemplar hashtags were selected from the tweets to augment the initial set of hashtags to obtain the expanded set, $H1$ (table 1). These new hashtags were discussed and shortlisted based on their frequency and contextual relevance to the events concerned. In doing so, we manually inspected every new hashtag and only added them to $H1$ if they were unambiguously event-specific. We then used the expanded keyword/hashtag set $H1$ to collect more tweets over the respective event timelines (mentioned below).

The objective behind this two-step data collection process was to optimise for recall, i.e., to ensure that our datasets adequately represent the universe of tweets for the events. We finally obtained 503540 tweets tweeted by 99355 unique users, 136893 tweets tweeted by 7004 unique users, and 643662 tweets tweeted by 64663 unique users for MRA, CJI, and BB respectively, including retweets. This data was used for all further analysis in this study.

During this process of tweet collection, we identified for each overlap pair (MRA-CJI, BB-MRA, and BB-CJI) a set of hashtags that are unique to each event, and another set of hashtags that are popular in both events in a pair, in terms of their frequency of occurrence (top 20 in order of frequency in our case). These sets are represented as H_{unique} and H_{common}, respectively (it must be noted that $H1 = H_{unique} \cup H_{common}$ for each event).

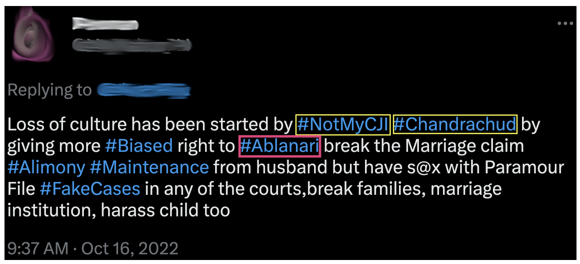
3.2 Identification of Core Users

We applied a filtering step to obtain the core users from the larger sets of unique users described in section 3.1 for the three events. We define *core users* as the set of users tweeting or retweeting on an event, who have a certain threshold of importance in the event community. In our case, 'importance' is defined by the cumulative number of retweets received by a user over the entire timeline of the event. Thus, the core users are users who have a higher degree of influence on the community narrative than their community peers, in terms of retweets received by their tweets. Based on the cumulative distributions for the retweets received by the users for the events, we considered the top 1 percentile of users as the core users. The events MRA, CJI, and BB were thus seen to carry 1694, 203, and 2446 core user accounts, respectively.

We next formed the retweet network from the core user sets of the three events. The retweet network is a directed graph G where an edge $(u, v) \epsilon E$ (E is the set of all edges in G) denotes that user account u has retweeted user account v at least once in the event timeline. We selected the largest connected component of the retweet networks for our study. Table 2 shows the retweet network statistics for the three event pairs. We use these networks for all of our further analyses.

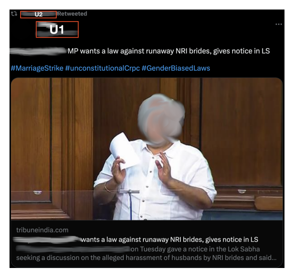
3.3 Identification of Bridge Users

Considering the three event pairs that we can form using the three event networks (MRA-CJI, BB-CJI, and BB-MRA), we define *bridge users* as core users belonging to either of the events in an event pair who satisfy at least one of the following two conditions: (A) They retweet authors (core users) from the other event to whose core set they themselves do not belong and (B) They post tweets or retweets containing at least one unique hashtag (ϵH_{unique}) from both of the events in an event pair simultaneously, or they post tweets/retweets with at least one common hashtag (ϵH_{common}) from both events. The core users satisfying the first condition are bridges that are said to enable *retweet overlap* between the events in an event pair, while those satisfying the second condition are said to enable *hashtag overlap* between the events. Figure 1 shows examples for hashtag overlap (Fig. 1a) & retweet overlap (Fig. 1b, 1c).



(a) *Hashtag* overlap in the tweet sample. The same tweet contains hashtags used by core members of CJI (*#NotMyCJI*, *#Chandrachud*, boxed in yellow) as well as hashtags used by core members of MRA (*#Ablanari*, boxed in pink).

(b) *Retweet* overlap in the tweet sample. Our analysis identifies that username *U1* is a core MRA user and username *U2* is a core CJI user.



(c) *Retweet* overlap in the tweet sample. Our analysis identifies that username *U1* is a core MRA user and username *U2* is a core CJI user.

Fig. 1. Examples of *hashtag* and *retweet* overlaps between MRA & CJI community. We did not find significant hashtag overlaps for the BB-MRA and BB-CJI pairs. Note : *User profile information, named entities and images of persons in context have been blurred to ensure anonymity.*

To identify bridge users, we first removed common users from the two core user sets for the event pairs considered. The users who appear as core users for both events in a pair or the *transitioning users* are not considered in this study, since we wanted to ensure that we captured the event community overlap characteristics with respect to the activities of users, who were dedicated to one event, but also sporadically participated in the other event's discourse through retweet and hashtag overlaps. Thus, we were interested in capturing the role that users (bridges), fundamentally different in their preference towards the two social media events in an event pair, played in the overlap between them. We finally obtained a set of 325, 239, and 54 bridge users after following this process for the MRA-CJI, BB-MRA, and BB-CJI pairs (the BB-MRA and BB-CJI pairs consisting of only retweet overlaps).

3.4 Political Leaning of Users

The political leaning of a user account on Twitter can be inferred in multiple ways – based on the content posted by the user, their network neighborhood, or their profile descriptions. In this paper, we considered this problem to be a binary classification problem, i.e., a user could either have a pro-ruling alliance or *pro-ruling* leaning (i.e., they are aligned towards the currently ruling Bharatiya Janata Party (BJP) or its allied parties) or a *pro-opposition* leaning (i.e., they are aligned towards one or many of the opposition parties). A major challenge in inferring the political leaning from the content posted by the users was that many of the users active in the event discourse had posted tweets with keywords only relevant to the events considered. Several of them, in fact, had joined or started being active on Twitter primarily to contribute to the social media movements around the events. As a result, their tweets did not have a significant number of political keywords/entities from which their political leaning could be inferred (based on manual analysis of 100 tweets selected randomly from the bridge users' tweets for each event). Additionally, many of their profile descriptions too did not have enough hints that acted as proxy to their political leaning (based on TF-IDF analysis of the keywords of their profile descriptions).

Therefore, we used the network characteristics of the users to infer their political alignment. We say that a user account on Twitter is pro-ruling, if the account itself or majority of their followings (friends) on Twitter are political entities belonging to the ruling alliance. On the other hand, a user is pro-opposition leaning if the account or majority of their followings are pro-opposition entities. We use Nivaduck [69] a dataset of Indian political entities for this purpose. This simple technique of detecting the political leaning indeed provided us near accurate results. For the purpose of a sanity check, we randomly selected a few (around 10) users from the events who were active on Twitter even outside the events, i.e., they also posted content that was not relevant to the three events considered. A manual check of their latest 20 tweets (not belonging to MRA, CJI, and BB) revealed that the users classified as pro-ruling ideologists by our method had posted politically polar content aligned to the ruling alliance (or against the opposition) frequently, and vice versa (the accuracy of our method on the set of 200 tweets manually checked was above 98% for this task).

3.5 Annotation Method

3.5.1 Annotation of Propagandist and Non-Propagandist Tweets. To understand the degree of propagandist messaging for the three events, we categorized tweets corresponding to the events into the categories of *Propaganda* (P) and *Non-Propaganda* (NP). Three annotators outside the group of authors of this study manually annotated the top 500[1] most retweeted messages corresponding to each event into the two categories. The political affiliations as well as perspectives on these issues of the three annotators are unknown to us, two identified as male and one as female. All three were Indian and regular consumers of news, thus familiar with the issues they were annotating for.

As part of this process, the annotators first had a group discussion and created a coding scheme with descriptions for inclusion and exclusion criteria (Table 4, 5). The discussions included conversations on the events and metaphor around them, which were important since there is significant use of innuendo and contextual information. The coding

[1]We considered the top 500 retweets based on the cumulative distribution of retweets received for each event.

Event	Primary Target Groups	Propaganda (% of tweets)
MRA	Women, political entities	65%
CJI	Women, political entities, judiciary	67%
BB	Hindi film fraternity (actors, directors, and producers)	98%

Table 3. Categorization of events based on qualitative analysis: Three annotators independently annotated 500 top retweeted tweets for each event as propaganda/non-propaganda

scheme contained a set of example tweets for each event, their gold standard labels (P/NP), and the reason behind their categorization into one of the two categories.

The tweets were labeled based on the presence of 18 features of propagandist messaging listed in Table 7 in the Appendix, as identified by Martino et al. [21]. We considered tweets to be propagandist if: (A) They contained at least one of the 18 features as part of their textual content, and (B) They directly vilified or discredited a particular group or individual with an explicit agenda to demonstrate support towards a predetermined belief (in this case, the target group being the women, judiciary, Bollywood, or politicians).

Next, we evaluated the annotators' understanding of the exercise by presenting a set of 20 tweets (10 propaganda and 10 non-propaganda tweets) to them and asking them to categorize the tweets. This exercise had an agreement of above 86% for each event among the annotators, i.e., for around 14% of the tweets, at least one annotation differed from the others. For these tweets, the annotators decided on a gold standard annotation along with the rationale behind it after discussion.

Finally, the annotators were provided a set of 500 most retweeted tweets for each event as the main annotation task. To measure the inter-annotator agreement, we calculated the Cohen's Kappa statistic which provided a K value of 0.90, indicating a significantly high degree of agreement. For the tweets carrying disagreement, majority label was considered as the final annotation. We present the categorization of the three events in table 3. We find that the three events involve significant amount of propagandist messaging from the users based on our definition of propaganda. While CJI is a fringe extension of the more mainstream MRA and the messaging style for both events is to some extent similar, CJI explicitly targets the judiciary (especially Justice Chandrachud) and uses significantly more legal terms in its community narrative. BB on the other hand is a highly propagandist event with a different root cause and target groups.

We annotated the top 500 most retweeted messages for the events, since the top retweeted messages provide us a fairly accurate idea of what is "successful" as far as the outreach goes, and in turn, defines the nature of dominant Twitter discourse around the event communities considered. If a significant fraction of the most retweeted messages around an event are propagandist, we can consider the event a propaganda event as these messages carry the maximum impact on the community.

3.5.2 Annotation of Political Stance of Tweets. The annotators (described in the previous section) also annotated a set of tweets mentioning political accounts, to categorize them into three classes of positive, negative, and neutral based on the stance of the tweets towards the accounts. In this step again, the annotators had an initial group discussion and populated the coding scheme with descriptions for inclusion and exclusion criteria (Tables 4, 5). The scheme contained example tweets for each event, their gold standard annotation (positive/negative/neutral), and an explanation of why they belonged to the corresponding category.

The tweets were supposed to be labeled as *positive* if they praised the political entities mentioned (politician's account or the political party's account) directly. They were labeled as *negative* if they criticized or berated political entities for their inaction or statements on the issues, either directly or sarcastically. Finally, tweets labeled *neutral* neither praised nor criticized the political entities mentioned, and merely mentioned them as a form of appeal for action on a particular issue or to bring certain matters to their attention (e.g., images or videos on the issue displaying a certain incident).

The annotators first labeled a set of 20 tweets for each event based on the coding scheme. We obtained an agreement of 95% for this annotation task. For the tweets on which the annotators had a disagreement, another round of discussion was done and their gold standard labels were finalized. Finally, the annotators annotated all tweets/retweets made by the bridge accounts that mentioned political entities for the events considered. Cohen's Kappa for the exercise revealed a K-value of 0.86. We found that most of the annotations where the annotators disagreed were tweets that were ambiguous in terms of their stance towards the political entities mentioned (e.g., tweets that contained sarcasm or rhetorical questions). In such cases, we considered the majority annotation (label) as the final label for the tweet.

We did not find any example of a tweet with positive stance towards a political entity mentioned for MRA-CJI. This is because the event pair MRA-CJI is inherently critical of the status quo around men's rights. Thus, the users mostly uphold the cause of the movements by being either critical of the political entities, or by bringing their attention to particular cases of atrocities against men. We found no examples of political mentions for BB among the top retweeted 500 tweets, since BB generally has a chiefly non-political target group and cause.

4 RESULTS

In this section, we analyze the influence that the bridge nodes carry in terms of the information diffusion they enable in the MRA-CJI, BB-MRA, and BB-CJI networks, and report their tweet content and network neighborhood characteristics.

4.1 Role of Bridge Nodes in Propaganda Spread

We find that the three overlap pairs show a steady increase in the number of bridge nodes over time resulting in increasing overlap between the two communities in a pair. To understand the importance of these bridge accounts in guiding the community discourse, we tried to understand if they belonged to some of the most influential nodes in the retweet network in terms of the reach of the messages posted by them. For this purpose, we ran influence maximization on the retweet network. Influence maximization (IM) selects a small subset of nodes (also called seed nodes) in a network that could maximize the spread of influence. IM models information spread in the network using various models of information diffusion. In our case, we considered the *Independent Cascade* model, a stochastic information diffusion model where the information is assumed to flow over the network through cascade [42]. We ran the CELF algorithm [49] for IM (parameters p=0.1 and mc=50) over multiple iterations, considering different cardinalities of the seed set (5 to 100) and checked the fraction of bridge nodes present in the seed set. If a bridge node belongs to the seed set, it would indicate that the node has a significant influence in terms of the reach it can achieve for information generated (tweeted) by it.

Figure 2 shows the fraction of seed nodes (influential nodes) that are bridges for different iterations of the run. We can see that above 60% of the top 5 most influential nodes are bridge nodes for the MRA-CJI and BB-MRA pairs. As the size of the seed set is increased, the fraction of seed nodes that are bridges reduces and stagnates between 30-35% for these two overlaps. This clearly indicates that a significant fraction (above 30%) of the most influential nodes in the network – in terms of information spread – are bridge nodes. Given that a large proportion of the tweets corresponding to the events considered are propagandist in nature, this also suggests that such bridge nodes aiding in overlap between

Event	Example Tweet	Stance Label	Propaganda Label	Reason
MRA	None Found	Positive	None	None
	Dear @narendramodi Ji, Happy birthday Sir Just make fake #AblaNari Aatmanirbhar, lots of people will give there blessings #StopLegalTerrorism #GenderBiasedLaws #HelpMeMyNation	Negative	P	The tweet blames the Prime Minister for not taking action against women who make fake claims of being weak and dependent (#AblaNari), albeit in a sarcastic fashion (negative stance). The tweet suggests the PM to take the necessary action of making the so called 'dependent women' independent to get blessings from the citizens. The use of #AblaNari and #StopLegalTerrorism introduce the propaganda features of **name calling and exaggeration** in the tweet (P).
	Men are victims of violence too be it physical, mental, emotional or financial.@narendramodi @PMOIndia @myogiadityanath Please work on making gender neutral laws. #NyayPrayaas4Men #GenderBiasedLaws #NoViolenceOnMen #IndianArmy	Neutral	NP	The tweet brings attention of the political entities to the fact that men could be victims too (NP). It does not explicitly accuse them of inaction (neutral stance).
CJI	None Found	Positive	None	None
	@PMOIndia @narendramodi @RSSorg How long Hindu men, husband have to suffer? Is this what you all do for them? People like #NotMyCJI are openly abusing Hindu family system and you do nothing but expect votes from very men? @realsiff helping thousands men..	Negative	P	The tweet blames the Prime Minister of inaction towards protecting men belonging to a particular religion, thus **exaggerating** the issue by connecting it to religious atrocities and **flag bearing** (P). Additionally, it charges him of working only towards the needs of his vote bank (negative stance).
	<URL> ++ @KirenRijiju @narendramodi @PMOIndia we have to get rid of Collegium soon. #NotMyCJI	Neutral	NP	The tweet simply carries a suggestion (or a personal opinion) of removing the collegium (group of judges who form the primary decision making body for recruitment of judges) in the Indian Supreme Court.

Table 4. Coding scheme (snapshot) for annotation and categorization of bridge tweets into (A) Propaganda and Non-Propaganda, and (B) Those that mention accounts of political entities, into three classes of positive, negative, and neutral based on their stance towards the entities. We did not find any example of tweets with positive stance towards the political entities in our dataset. The propaganda features evident in the tweets are marked in bold.

Event	Example Tweet	Stance Label	Propaganda Label	Reason
BB	SSR Sacrificed his life to Expose Gutterwood, Charsiwood Bullyweed #BoycottBollywood Sushant Conquered BWood URL	None	P	The tweet contains negatively targets a group, and contains **exaggeration** and **name calling**.
	A 2015 Story is viral, where #ShahRukhKhan helped a #KashmiriPandit family in need of financial aid. A filmmaker, who is currently trending #BoycottBollywood in every news debate, then thanked #SRK for his big ... @srk never claimed such good deeds, even I have seen him doing so URL	None	NP	The tweet narrates an incident and praises an individual. No trace of vilification of any group or individual is present.

Table 5. Coding scheme (snapshot) for annotation and categorization of bridge tweets (BB) into (A) Propaganda and Non-Propaganda, and (B) Those that mention accounts of political entities, into three classes of positive, negative, and neutral based on their stance towards the entities. We did not find any example of tweets with mention of political entities among top 500 tweets for BB (thus, the Stance Label for political mentions is not shown).

three event communities also have a significant influence in propaganda spread. For BB-CJI, the fraction of bridges among the influential seed nodes are comparatively smaller (settles at around 13%). This might be because the BB-CJI overlap contains very few bridges overall (54) and hence, they exhibit little intersection with the seeds. These findings (especially those for MRA-CJI and BB-MRA) motivate us to further explore the content and network characteristics of bridges that enable information diffusion across multiple event communities.

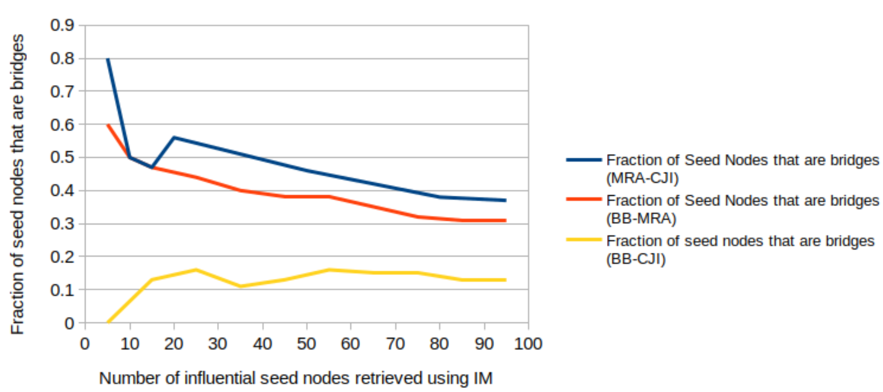


Fig. 2. Fraction of influential seed nodes retrieved using Influence Maximization that are bridges in the three overlap pairs: A significant fraction of seed nodes are bridges in all of the pairs, indicative of the influence of bridges in information spread across the retweet network

4.2 Political Leaning of Bridges

4.2.1 User and Network Characteristics of Bridge Nodes. The political leaning of a Twitter user account was obtained from Nivaduck, provided that the user was present in the dataset. In case the user account was not present in Nivaduck, we derived the political leaning from the percentage of the user's followings (or friends) who were present in Nivaduck as discussed in section 3.4. However, we found that several bridge user accounts (above 18% for all pairs) did not have any friends who were influential politicians, i.e., entities that belonged to Nivaduck. To work around this problem we used the following method: We first extracted the retweet network neighborhood [2] of bridge users, i.e., the users who retweet the bridges and the users who are retweeted by the bridges at least once during the event timeline. Mathematically, the retweet neighborhood can be represented as follows:

$$S_m^{out} = \{v\}, \text{ where } v \epsilon V \text{ and } (m, v) \epsilon E$$
$$S_m^{in} = \{u\}, \text{ where } u \epsilon V \text{ and } (u, m) \epsilon E$$

Here, V and E are the set of vertices (user accounts) and edges respectively in the retweeet network. Any edge (a, b) represents a retweet edge from account a to account b, indicating the "a retweets b" relationship. S_m^{in} and S_m^{out} are sets of user accounts retweeting and retweeted by a bridge account m, respectively. Next, we inferred the political leaning of this neighborhood (political leaning of users retweeting the bridges and users retweeted by the bridges) using their friends as described in section 3.4. In other words, if majority of the nodes in the retweet network neighborhood of a user were found to be aligned to the ruling alliance (or opposition), we considered the leaning of the neighborhood to be pro-ruling (or pro-opposition).

Thus, based on the political leaning of the three node sets – those retweeting the bridge, the bridge itself, and those retweeted by the bridge – we can form retweet triplets as shown in figure 3. In these triplets, R represents a pro-ruling political leaning of the set, L represents a pro-opposition leaning of the set, and X represents unidentified leaning (i.e., cases where there were no politician friends as identified using Nivaduck for the user).

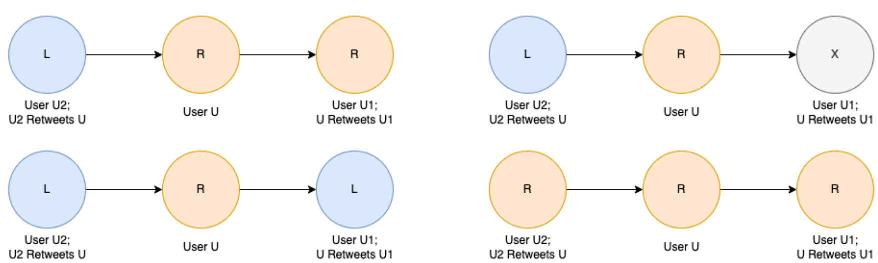


Fig. 3. Examples of political leaning of bridge nodes (User U) and their retweet network neighborhood (A->B shows the A retweets B relationship): R represents pro-ruling leaning, L represents a pro-opposition leaning, and X represents unidentified leaning. For example, the retweet triplet L->R->R represents a pro-ruling bridge node being retweeted by a pro-opposition leaning user and the bridge node retweeting a pro-ruling user.

In table 6, we show the percentage of bridge nodes involved in each type of retweet triplet, for each of the three event pairs. We find that the percentage of bridges belonging to the triplets with pro-ruling neighborhood (RRR, RXR, and RLR) significantly exceeds those belonging to the triplets with pro-opposition neighborhood (LLL, LXL, and LRL),

[2]We only considered the retweets and not quoted tweets in this study. This is because retweets can be considered as endorsement of the views of the retweeted user. However, quoted tweets may express an opinion opposing the views expressed in the original tweet.

indicating that in all of the overlaps, the bridges retweet or are retweeted by pro-ruling accounts much more than pro-opposition accounts. Interestingly, we also find that for BB-MRA and BB-CJI overlaps, a large percentage of bridges are retweeted by users with a different polarity than the users whom the bridges retweet (e.g., the mixed neighborhood types RRL, LRR, RLL, and so on). This behavior is indicative of the event overlaps cutting across political lines where irrespective of the polarity of the bridge, both pro-ruling and pro-opposition users come together to retweet or get retweeted by the bridge.

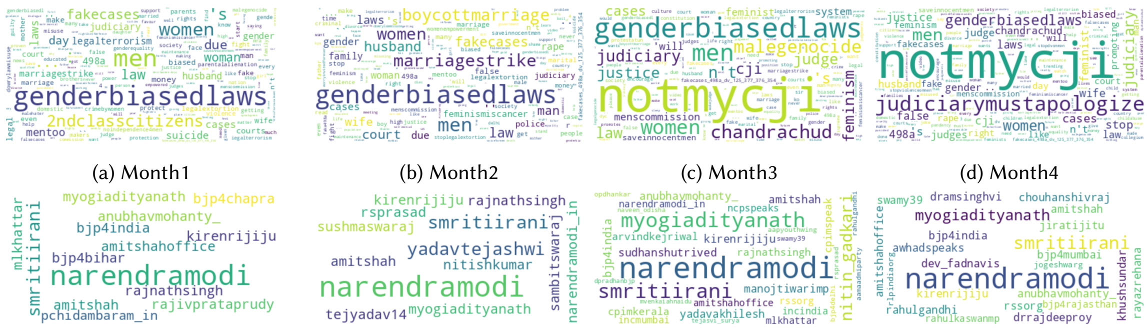
Retweet Triplet	Count (MRA-CJI)	Count (BB-MRA)	Count (BB-CJI)	Neighborhood Type
RRR	**163** (50.20%)	21 (19.09%)	**13 (36.11%)**	pro-ruling (R)
RLR	64 (19.70%)	11 (10.00%)	3 (8.33%)	
RXR	49(15.10%)	6 (5.45%)	1 (2.78%)	
RRL	12 (3.70%)	**24 (21.79%)**	7 (19.44%)	Mixed
LRR	7 (2.20%)	3 (2.73%)	1 (2.78%)	
RLL	6 (1.90%)	10 (9.09%)	3 (8.33%)	
RXL	5 (1.50%)	9 (8.18%)	1 (2.78%)	
RLX	_	6 (5.45%)	6 (16.67%)	
LLL	3 (0.90%)	7 (6.36%)	_	pro-opposition (L)
LXL	2 (0.60%)	9 (8.18%)	_	
LRL	9 (2.80%)	4 (3.63%)	1 (2.78%)	

Table 6. Percentage of bridge nodes for each retweet triplet type: The triplet configurations not shown in the table have no bridge nodes belonging to them for most event pairs. Additionally, the XXX triplet (when neither the political leaning of the bridge, nor its neighborhood could be found) is not shown in the table.

Therefore, we find that political leaning, especially pro-ruling leaning, carries a significantly higher correlation with the community overlaps of the three event pairs, when compared to pro-opposition leaning. This is most pronounced when we consider the retweet network neighborhood for the MRA-CJI overlap, which is pro-ruling for majority of the bridges. However, there also exist examples of bridge accounts with pro-opposition leaning connecting with a pro-ruling neighborhood for the community discourse, and vice versa. Thus, MRA, CJI, and BB being fringe movements that are relatively less political in nature compared to mainstream political events, also attract overlaps through users from a wide political spectrum.

4.2.2 Content Characteristics. We also studied the textual content of tweets posted by the bridge user accounts (both retweet and hashtag overlaps) over time, to understand if keywords/hashtags inherently political in nature were present in them. For this purpose, we first analyzed the unigrams present in the tweet text for each month of tweeting activity of the bridge users. The corresponding word-clouds are shown in figure 4 (a, b, c, d). As evident from these plots, not many political keywords are present among the top keywords present in the tweet text. We also see that most of the keywords for the event pairs are highly relevant to the movement considered – the prominent keywords for the MRA-CJI overlap are *genderbiasedlaws, boycottmarriage, notmycji*, etc. It should be noted that for the last two months *#notmycji* appears as the most frequent keyword in the narrative, indicating the transition of the community narrative towards the fringe #NotMyCJI movement. During these two months, the community narrative also exhibits usage of other, more legal terms by the bridge accounts (e.g., *chandrachud* – the Chief Justice of India around whom the *#notmycji* controversy was formed, *judge*, and *judiciary*).

We next analyzed the political accounts (accounts of politicians and accounts of regional and national political parties) mentioned by the bridge nodes during the timeline of the movements. The word-clouds for the frequently



(a) Month1 (b) Month2 (c) Month3 (d) Month4

(e) Month1 (f) Month2 (g) Month3 (h) Month4

Fig. 4. Word-clouds for top keywords and mentioned accounts for four consecutive months of tweeting by bridge accounts for MRA-CJI overlap (preprocessing to convert all into lowercase was done first)

mentioned political accounts are shown in figure 4 (e, f, g, h). We find that the bridge accounts do engage in mentioning multiple political accounts in some of their posts. However, these mentions occurred for only a few tweets/retweets during the event timeline (only 312 tweets considering the entire timeline for MRA-CJI). This finding corroborates our previous finding of an event discourse devoid of political messaging. We also found that a majority of these highest mentioned political accounts (above 90%) belong to the BJP, the currently ruling party. The most frequently mentioned account across the event timeline belongs to Narendra Modi, the current Prime Minister of India who is also a highly popular figure on Twitter. Presence of other high profile political accounts of BJP (in terms of their number of followers and their national political stature) include *@AmitShah* – the Home Minister and *@myogiadityanath* – the current Chief Minister of the state of Uttar Pradesh. Regional accounts of the political party (e.g. *@BJP4Bihar*) are also mentioned significantly in the posts as evident from the plots. We report the content characteristics for the two remaining event pairs (BB,MRA) and (BB,CJI) in the appendix, which also reveal similar findings.

To understand the context in which these political accounts had been mentioned, we qualitatively studied the tweets/retweets posted by the bridge accounts that mentioned at least one political account. As discussed in section 3.5, three annotators manually categorized the stance of these messages into three classes of *positive, neutral,* and *negative*. From the annotation results, we found that most of the messages (around 64% for MRA-CJI) mentioning BJP political accounts urged the ruling dispensation to take appropriate action, both at the policy and judicial levels, to protect men's rights and to make Indian laws gender neutral (such tweets are annotated as neutral). The tweets generally cited previous cases of harassment against men in specific regions and called for the government to take a stand in this respect. Narendra Modi, the face of BJP, was expectedly the most mentioned entity in these tweets. Surprisingly, we also found that a section of bridge accounts that exhibited a significantly pro-ruling network neighborhood (friends/followings), engaged in berating the ruling party and its politicians directly for inaction on the matter or indirectly through sarcasm (annotated as negative stance) for MRA-CJI. The mention of accounts belonging to the opposition was negligible, although a few important politicians like *@yadavtejashwi* and *@nitishkumar* (current deputy Chief Minister and Chief Minister of Bihar belonging to the opposition parties Janata Dal United and Rashtriya Janata Dal, respectively) were mentioned mostly with similar urges/criticisms around taking action against the alleged perpetrators (mostly women). We thus see that the content characteristics of bridge users do not reflect significant political leaning and that the constituent messaging is mostly focused on the cause of the movement. However, when mentioned, the political accounts are either appealed to take action for the movement, or castigated for their inaction, irrespective of the political leaning of the bridge user.

These findings paint an interesting picture – while the MRA-CJI[3] community overlap exhibits a pro-ruling leaning based on the retweet network characteristics of bridges, the bridges' usage of political keywords is negligible when it comes to tweeting about the movement. This could be an indication of the sincerity that these bridge accounts exhibit in their tweeting behavior, i.e., while the accounts themselves show significant political leaning, they avoid using political keywords and stick to pertinent messaging for the cause of the movement, to avoid diverting the mainstream audience attention to other political issues. Thus, while being inherently political, the community's apparent narrative is mostly apolitical, indicating the use of coordinated and disciplined messaging done by the bridges.

5 CONCLUSION AND DISCUSSION

There is little doubt that the quality of information on social media serves as an important marker of social schisms in multiple geographies. While there has been much discussion around this, there is more to do on the organization and spillover of manufactured campaigns in the global south. Specifically in the case of India, there is work to suggest that politicians and the state apparati engage actively on social media with regard to narrative building with the purpose of influencing public opinion at a large scale [66], just as there is work to show that what was the fringe of political and social thought is increasingly powerful in everyday life, due to its power on social media [76]. Our work here shows, through the examination of three fringe communities that are largely engaged in propaganda and hate speech, that their networks with relatively established communities on social media help build bridges that can lead to greater mainstreaming. Moreover, the overlaps between multiple such communities are dominantly seen to be driven more by users with a particular political alignment.

As we see in the examination of the content and network characteristics in the MRA-CJI, BB-MRA, and BB-CJI overlap pairs, the *bridge user accounts* through cross-event retweeting and hashtagging are significantly influential in terms of information diffusion in the MRA-CJI event network. Additionally, while the content of their tweets lack political keywords such as party slogans or nomenclature, the retweet network neighborhoods of these bridges show significant political alignment towards the ruling dispensation at the national level, when compared to the opposition. We thus find that the three fringe propaganda communities and the overlap between them are majorly driven by political leaning on propagandist discourse in social media. However, this political leaning is significantly observable only around the retweet network characteristics of bridge user accounts, instead of the content posted by them. These findings lead to several interesting conclusions. First, while not all bridge accounts are influential in terms of the probabilistic diffusion of information across the overlapping event networks, a significant percentage of them are. This paves the way for overlap or merger of such fringe event communities by way of spreading information (and propaganda) at a massive scale. Second, the negligible use of political keywords indicates the organization of bridge users' content around the primary cause of the movement. The collective use of issue specific keywords and hashtags also is a hint towards the coordination that the bridge users show amongst each other in the functioning of the community.

The accounts mentioned in these tweets by bridges also include an overwhelming number of individuals or groups belonging to the ruling party. Call-outs are indeed a standard strategy aimed at getting attention, and it makes most sense to catch the attention of individuals from the party in power, but the style of appealing to sources of authority within the ruling establishment suggests a calibration with the rules of play. The act of using a popular politician's

[3]Tweets in BB do not exhibit a significant number of political mentions and we report these findings in the Appendix.

handle as part of an appeal of opposition (for instance, opposition towards the CJI) shows a willingness to use social media tweets as a means of petition to the powers in place, and signaling political alignment with them on this issue. As we see in the following text from a highly engaged tweet, the tone is generally conciliatory and worded as an appeal, positioned as a call to the Prime Minister to act on an issue that is generally aligned with his party. However, at the end of the tweet we also see a threat, much like what was seen on social media during the Trump era, when a number of his fringe supporters started proposing ideas which when rejected, were pushed back with the pejorative 'RINO' (Republican in Name Only), a term used to attack Republicans who did not take on some of the more extreme positions.

> *"Husband suicide has reached an alarming level as a result of lopsided #GenderBiasedLaws. Forget about #NotMyCJI; if the BJP does not wake up and act immediately, I am sorry to inform you Modi Ji, that you will be #NotMyPM in 2024."*

It is important to mention that while the retweet network neighborhood of the bridges for the three event pairs exhibit a much higher pro-ruling than pro-opposition leaning, there is also significant evidence of users with different political leanings coming together to enable the overlaps. This is especially evident for the BB-MRA and BB-CJI pairs. This phenomenon might occur due to the fact that the three events considered have primarily non-political target groups, thereby their overlaps often bringing new buyers to either side of the political spectrum. As part of future work, it will be interesting to examine the retweet and hashtag overlaps that bring users with opposite political leanings together, and the content posted specifically by such users with opposite leanings.

Our work opens several other directions for further research. While in this study we focused on the content and network characteristics of bridge accounts and their relationship with political leaning of bridge users, it would be interesting to see why and how certain nodes or user accounts get converted to bridges over time. There could be several reasons behind a user account leading to overlap between three event communities. Users originally belonging to a long sustained mainstream event might get interested in creating or participating in other fringe events to gain respect or a sense of importance in the community. In this direction, including *transitioning users* or users who tweet at different points in time on each event in an event pair will be necessary. Centrality based measures observed temporally can provide an idea of the reason behind such user movements.

This paper studies the overlap between two closely related events – MRA and CJI, where the latter is more of an extension of the former. However, previous work has shown that even seemingly unrelated event communities may come together or overlap on social media simply through sporadically similar topics of discussion. There also exist indications of users belonging to two disjoint event communities coming together owing to a third mainstream social media movement that they are supportive of. This is evident from the (BB-MRA) and (BB-CJI) overlaps. In both of these pairs, two social media movements with completely different causes exhibit overlap. While through this work, we find a close relationship of such overlaps with the political leaning of users, automatic methods to identify other factors that act as latent pillars for overlaps between apparently unrelated fringe events might be an important research direction to pursue.

A broader research problem might be to model the overlap between a set of social media events through observation of their user, content, and network characteristics. Early detection of events consisting of significant amount of propagandist messaging and misinformation is a well studied problem. It will also be important to develop methods that could predict the overlap between multiple event communities early in their timeline. Not only could such methods aid social media users to take preventive measures against merger of propaganda communities, they could also help social media regulatory groups in strategically avoiding such mergers through appropriate policy level interventions (similar to suspension of user accounts owing to mass reporting).

Our work has a few limitations. First, we have only considered overlaps caused by users involved in cross-retweeting and hashtagging on the three events. There are several other kinds of overlap that need to be studied, e.g., overlap through topics of discussion and temporal movement of common or transitioning users. We intend to work in this direction as an extension of this work. Second, our definition of 'propaganda' is heavily dependent on existing literature on the topic and the definition of this term is subjective. Thus, the focus of this paper is on studying the relationship of political leaning on event overlap, rather than detection of propagandist messages. In section 3.5, we merely commented on the degree of propagandist messaging evident in the events based on our definition of propaganda. Third, our method of political leaning detection using network characteristics (followings) of users needs refinement. In this direction, we could also consider unigram/n-gram/document representations of tweets (external to the event space considered) with opposite polarities to understand if they are positioned differently in a bipolar ideological space. Fourth, the process of propaganda detection followed in this paper is manual. We are currently in the process of developing a classifier specifically to classify Indian propaganda events. Finally, while this paper studies the overlap between three events, on a selected cross-section of data (Jan-Dec 2022), we are also planning to add more event pairs along with longitudinal analysis to understand if our findings could be generalized further. The methods used in this work are generalizable though, and could be used to study the overlap between any pair of social media event communities.

REFERENCES

[1] Muhammad Afzaal, Swaleha Bano Naqvi, and Gul Rukh Raees. 2022. Representations of Naya Pakistan: A corpus-based study of Pakistani media discourses. *Asian Journal of Comparative Politics* 7, 3 (2022), 521–538.

[2] Syeda Zainab Akbar, Ankur Sharma, Himani Negi, Anmol Panda, and Joyojeet Pal. 2020. Anatomy of a Rumour Social media and the suicide of Sushant Singh Rajput. *arXiv preprint arXiv:2009.11744* (2020).

[3] Abeer Aldayel and Walid Magdy. 2019. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 205 (nov 2019), 20 pages. https://doi.org/10.1145/3359307

[4] Safa Alsafari, Samira Sadaoui, and Malek Mouhoub. 2020. Hate and offensive speech detection on Arabic social media. *Online Social Networks and Media* 19 (2020), 100096.

[5] Amit A Amleshwaram, Narasimha Reddy, Sandeep Yadav, Guofei Gu, and Chao Yang. 2013. Cats: Characterizing automation of twitter spammers. In *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS)*. IEEE, 1–10.

[6] Alberto Barrón-Cedeno, Israa Jaradat, Giovanni Da San Martino, and Preslav Nakov. 2019. Proppy: Organizing the news based on their propagandistic content. *Information Processing & Management* 56, 5 (2019), 1849–1864.

[7] Srimati Basu. 2019. The Cinematic Pleasures of Indian Men's Rights Activists. *Critical Analysis of Law* 6, 1 (2019).

[8] Souâad Boudebza, Rémy Cazabet, Faiçal Azouaou, and Omar Nouali. 2018. OLCPM: An online framework for detecting overlapping communities in dynamic social networks. *Computer Communications* 123 (2018), 36–51.

[9] Doreen Busolo and Samuel Ngigi. 2018. Understanding hate speech in Kenya. *New Media and Mass Communication* 70 (2018), 43–49.

[10] Chiyu Cai, Linjing Li, and Daniel Zeng. 2017. Detecting social bots by jointly modeling deep behavior and content information. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1995–1998.

[11] Carlos Arcila Calderón, Gonzalo de la Vega, and David Blanco Herrero. 2020. Topic modeling and characterization of hate speech against immigrants on Twitter around the emergence of a far-right party in Spain. *Social Sciences* 9, 11 (2020), 188.

[12] Thomas Carothers and Andrew O'Donohue. 2019. *Democracies divided: The global challenge of political polarization.* Brookings Institution Press.

[13] Sergio Andrés Castaño-Pulgarín, Natalia Suárez-Betancur, Luz Magnolia Tilano Vega, and Harvey Mauricio Herrera López. 2021. Internet, social media and online hate speech. Systematic review. *Aggression and Violent Behavior* 58 (2021), 101608.

[14] Akemi Takeoka Chatfield, Christopher G Reddick, and Uuf Brajawidagda. 2015. Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks. In *Proceedings of the 16th annual international conference on digital government research.* 239–249.

[15] Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. 2016. Debot: Twitter bot detection via warped correlation.. In *Icdm*. 817–822.

[16] Wei Chen, Xiao Zhang, Tengjiao Wang, Bishan Yang, and Yi Li. 2017. Opinion-aware Knowledge Graph for Political Ideology Detection.. In *IJCAI*, Vol. 17. 3647–3653.

[17] Aditya Chetan, Brihi Joshi, Hridoy Sankar Dutta, and Tanmoy Chakraborty. 2019. Corerank: Ranking to detect users involved in blackmarket-based collusive retweeting activities. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 330–338.

[18] Romit Chowdhury. 2014. Conditions of emergence: The formation of men's rights groups in contemporary India. *Indian Journal of Gender Studies* 21, 1 (2014), 27–53.

[19] Hiu-Fung Chung and Edmund W Cheng. 2022. Constructing patriotic networked publics: conservative YouTube influencers in Hong Kong. *Chinese Journal of Communication* 15, 3 (2022), 415–430.

[20] Wanyun Cui, Yanghua Xiao, Haixun Wang, Yiqi Lu, and Wei Wang. 2013. Online Search of Overlapping Communities. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data* (New York, New York, USA) *(SIGMOD '13)*. Association for Computing Machinery, New York, NY, USA, 277–288. https://doi.org/10.1145/2463676.2463722

[21] Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeno, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*. 5636–5646.

[22] Saloni Dash, Dibyendu Mishra, Gazal Shekhawat, and Joyojeet Pal. 2022. Divided We Rule: Influencer Polarization on Twitter During Political Crises in India. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 135–146.

[23] Soham De, Agrima Seth, Arshia Arya, Steven Wilkinson, Sushant Singh, and Joyojeet Singh. 2022. Closed Ranks: The Discursive Value of Military Support for Indian Politicians on Social Media. *arXiv preprint arXiv:2204.03098* (2022).

[24] Sandipan Deb. August 2022. The politics of boycott Bollywood. https://www.moneycontrol.com/news/trends/entertainment/the-politics-of-boycott-bollywood-9058811.html.

[25] Joseph Downing and Richard Dron. 2022. Theorising the 'security influencer': Speaking security, terror and Muslims on social media during the Manchester bombings. *new media & society* 24, 5 (2022), 1234–1257.

[26] Andrej Duh, Marjan Slak Rupnik, and Dean Korošak. 2018. Collective behavior of social bots is encoded in their temporal twitter activity. *Big data* 6, 2 (2018), 113–123.

[27] Juan Echeverrï£¡ a, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Gianluca Stringhini, and Shi Zhou. 2018. LOBO: Evaluation of generalization deficiencies in Twitter bot classifiers. In *Proceedings of the 34th annual computer security applications conference*. 137–146.

[28] Murray Edelman. 1988. *Constructing the political spectacle*. University of Chicago Press.

[29] Emilio Ferrara. 2017. Contagion dynamics of extremist propaganda in social networks. *Information Sciences* 418 (2017), 1–12.

[30] Biswajit Ghosh. 2022. Politics of manufacturing consent in a post-truth society. *Journal of Developing Societies* 38, 1 (2022), 7–26.

[31] Anatoliy Gruzd, Barry Wellman, and Yuri Takhteyev. 2011. Imagining Twitter as an imagined community. *American Behavioral Scientist* 55, 10 (2011), 1294–1318.

[32] Stefano Guarino, Noemi Trino, Alessandro Celestini, Alessandro Chessa, and Gianni Riotta. 2020. Characterizing networks of propaganda on twitter: a case study. *Applied Network Science* 5, 1 (2020), 1–22.

[33] Frederic Guerrero-Solé. 2017. Community detection in political discussions on Twitter: An application of the retweet overlap network method to the Catalan process toward independence. *Social science computer review* 35, 2 (2017), 244–261.

[34] Michael Hatherell and Alistair Welsh. 2020. Populism and the risks of conceptual overreach: A case study from Indonesia. *Representation* 56, 1 (2020), 53–69.

[35] Maryam Heidari, James H Jones, and Ozlem Uzuner. 2020. Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter. In *2020 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 480–487.

[36] Brian Heredia, Joseph Prusa, and Taghi Khoshgoftaar. 2017. Exploring the effectiveness of twitter at polling the united states 2016 presidential election. In *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 283–290.

[37] Jianbin Huang, Heli Sun, Yaguang Liu, Qinbao Song, and Tim Weninger. 2011. Towards online multiresolution community detection in large-scale networks. *PloS one* 6, 8 (2011), e23829.

[38] Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1113–1122.

[39] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. Why We Twitter: Understanding Microblogging Usage and Communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis* (San Jose, California) *(WebKDD/SNA-KDD '07)*. Association for Computing Machinery, New York, NY, USA, 56–65. https://doi.org/10.1145/1348549.1348556

[40] Julie Jiang, Xiang Ren, and Emilio Ferrara. 2022. Retweet-BERT: Political Leaning Detection Using Language Features and Information Diffusion on Social Networks. *arXiv preprint arXiv:2207.08349* (2022).

[41] Andreas Jungherr. 2016. Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics* 13, 1 (2016), 72–91.

[42] David Kempe, Jon Kleinberg, and Eva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. 137–146.

[43] Emily Kubin and Christian von Sikorski. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* 45, 3 (2021), 188–206.

[44] Sneha Kudugunta and Emilio Ferrara. 2018. Deep neural networks for bot detection. *Information Sciences* 467 (2018), 312–322.

[45] Mirko Lai, Alessandra Teresa Cignarella, Delia Irazú Hernández Farías, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020. Multilingual stance detection in social media political debates. *Computer Speech & Language* 63 (2020), 101075.

[46] Faisal M Lalani, Ramaravind Kommiya Mothilal, and Joyojeet Pal. 2019. The appeal of influencers to the social media outreach of indian politicians. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*. 267–271.

[47] Zachary Laub. 2019. Hate speech on social media: Global comparisons. *Council on foreign relations* 7 (2019).

[48] Sangho Lee and Jong Kim. 2014. Early filtering of ephemeral malicious accounts on Twitter. *Computer communications* 54 (2014), 48–57.

[49] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. 420–429.

[50] Ian XY Leung, Pan Hui, Pietro Lio, and Jon Crowcroft. 2009. Towards real-time community detection in large networks. *Physical Review E* 79, 6 (2009), 066107.

[51] Shenghua Liu, Bryan Hooi, and Christos Faloutsos. 2017. Holoscope: Topology-and-spike aware fraud detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1539–1548.

[52] Xin Lu and Christa Brelsford. 2014. Network structure and community evolution on twitter: human behavior change in response to the 2011 Japanese earthquake and tsunami. *Scientific reports* 4, 1 (2014), 6773.

[53] Sangeeta Mahapatra and Johannes Plagemann. 2019. Polarisation and politicisation: the social media strategies of Indian political parties. (2019).

[54] Ariadna Matamoros-Fernández and Johan Farkas. 2021. Racism, hate speech, and social media: A systematic review and critique. *Television & New Media* 22, 2 (2021), 205–224.

[55] Michele Mazza, Stefano Cresci, Marco Avvenuti, Walter Quattrociocchi, and Maurizio Tesconi. 2019. Rtbust: Exploiting temporal patterns for botnet detection on twitter. In *Proceedings of the 10th ACM conference on web science*. 183–192.

[56] Liz McQuillan, Erin McAweeney, Alicia Bargar, and Alex Ruch. 2020. Cultural convergence: Insights into the behavior of misinformation networks on twitter. *arXiv preprint arXiv:2007.03443* (2020).

[57] Maya Mirchandani. 2018. Digital hatred, real violence: Majoritarian radicalisation and social media in India. *ORF Occasional Paper* 167 (2018), 1–30.

[58] Dibyendu Mishra, Syeda Zainab Akbar, Arshia Arya, Saloni Dash, Rynaa Grover, and Joyojeet Pal. 2021. Rihanna versus Bollywood: Twitter Influencers and the Indian Farmers' Protest. *arXiv preprint arXiv:2102.04031* (2021).

[59] Abir Misra. 2021. Can the Celebrity Speak? Controversies and the Eulogistic Fandom of Shah Rukh Khan. *American Behavioral Scientist* (2021), 00027642211042286.

[60] Rachel R Mourão and Weiyue Chen. 2020. Covering protests on Twitter: The influences on journalists' social media portrayals of left-and right-leaning demonstrations in Brazil. *The International Journal of Press/Politics* 25, 2 (2020), 260–280.

[61] Abdul Muhid, Moch Hadi, Aris Fanani, Ahmad Arifin, and Ahmad Hanif. 2019. The effect of hate speech exposure on religious intolerance among indonesian Muslim teenagers. In *2019 Ahmad Dahlan International Conference Series on Education & Learning, Social Science & Humanities (ADICS-ELSSH 2019)*. Atlantis Press, 39–44.

[62] Mark EJ Newman. 2006. Modularity and community structure in networks. *Proceedings of the national academy of sciences* 103, 23 (2006), 8577–8582.

[63] Oluwafemi Oriola and Eduan Kotzé. 2020. Evaluating machine learning techniques for detecting offensive and hate speech in South African tweets. *IEEE Access* 8 (2020), 21496–21509.

[64] Diogo Pacheco, Alessandro Flammini, and Filippo Menczer. 2020. Unveiling coordinated groups behind white helmets disinformation. In *Companion proceedings of the web conference 2020*. 611–616.

[65] Joyojeet Pal, Priyank Chandra, and VG Vinod Vydiswaran. 2016. Twitter and the rebranding of Narendra Modi. *Economic and Political Weekly* (2016), 52–60.

[66] Joyojeet Pal and Anmol Panda. 2019. Twitter in the 2019 Indian general elections: Trends of use across states and parties. *Economic and Political Weekly* 54, 51 (2019), 1–17.

[67] Zachary D Palmer and Mangala Subramaniam. 2017. Abstract egalitarianism and men as victims: strategic choice of frames by men's rights organisations. *International Social Science Journal* 67, 225-226 (2017), 97–108.

[68] Gang Pan, Wangsheng Zhang, Zhaohui Wu, and Shijian Li. 2014. Online community detection for large complex networks. *PloS one* 9, 7 (2014), e102799.

[69] Anmol Panda, A'ndre Gonawela, Sreangsu Acharyya, Dibyendu Mishra, Mugdha Mohapatra, Ramgopal Chandrasekaran, and Joyojeet Pal. 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.

[70] Josh Pasek, Tobias H Stark, Jon A Krosnick, and Trevor Tompson. 2015. What motivates a conspiracy theory? Birther beliefs, partisanship, liberal-conservative ideology, and anti-Black attitudes. *Electoral Studies* 40 (2015), 482–489.

[71] Francisco Rangel and Paolo Rosso. 2019. Overview of the 7th author profiling task at PAN 2019: bots and gender profiling in twitter. In *Proceedings of the CEUR Workshop, Lugano, Switzerland*. 1–36.

[72] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2931–2937.

[73] Ann-Kathrin Rothermel. 2020. "The other side": Assessing the polarization of gender knowledge through a feminist analysis of the affective-discursive in anti-feminist online communities. *Social Politics: International Studies in Gender, State & Society* 27, 4 (2020), 718–741.

[74] 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)*. 627–634. https://doi.org/10.1109/ASONAM49781.2020.9381344

[75] Sebastian Stier, Arnim Bleier, Haiko Lietz, and Markus Strohmaier. 2018. Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter. *Political communication* 35, 1 (2018), 50–74.

[76] Ravi Sundaram. 2021. The Fringe as Media Infrastructure. *Media and the Constitution of the Political* (2021), 261.

[77] Petter Törnberg, Claes Andersson, Kristian Lindgren, and Sven Banisch. 2021. Modeling the emergence of affective polarization in the social media society. *Plos one* 16, 10 (2021), e0258259.

[78] Jay J Van Bavel, Steve Rathje, Elizabeth Harris, Claire Robertson, and Anni Sternisko. 2021. How social media shapes polarization. *Trends in Cognitive Sciences* 25, 11 (2021), 913–916.

[79] Herman Wasserman and Dani Madrid-Morales. 2022. *Disinformation in the Global South.* Wiley Online Library.

[80] Nishan Chathuranga Wickramarathna, Thiruni D Jayasiriwardena, Malith Wijesekara, Pasindu Bawantha Munasinghe, and Gamage Upeksha Ganegoda. 2020. A framework to detect twitter platform manipulation and computational propaganda. In *2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer)*. IEEE, 214–219.

[81] Felix Ming Fai Wong, Chee Wei Tan, Soumya Sen, and Mung Chiang. 2016. Quantifying Political Leaning from Tweets, Retweets, and Retweeters. *IEEE Transactions on Knowledge and Data Engineering* 28, 8 (2016), 2158–2172. https://doi.org/10.1109/TKDE.2016.2553667

[82] Nan Xi, Di Ma, Marcus Liou, Zachary C Steinert-Threlkeld, Jason Anastasopoulos, and Jungseock Joo. 2020. Understanding the political ideology of legislators from social media images. In *Proceedings of the international aaai conference on web and social media*, Vol. 14. 726–737.

[83] Kai-Cheng Yang, Francesco Pierri, Pik-Mai Hui, David Axelrod, Christopher Torres-Lugo, John Bryden, and Filippo Menczer. 2021. The covid-19 infodemic: Twitter versus facebook. *Big Data & Society* 8, 1 (2021), 20539517211013861.

[84] Kai-Cheng Yang, Onur Varol, Clayton A Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies* 1, 1 (2019), 48–61.

[85] Zhi Yang, Jilong Xue, Xiaoyong Yang, Xiao Wang, and Yafei Dai. 2015. VoteTrust: Leveraging friend invitation graph to defend against social network sybils. *IEEE Transactions on dependable and secure computing* 13, 4 (2015), 488–501.

[86] Kerem Yıldırım, Lemi Baruh, and Ali Çarkoğlu. 2021. Dynamics of campaign reporting and press-party parallelism: Rise of competitive authoritarianism and the media system in Turkey. *Political Communication* 38, 3 (2021), 326–349.

[87] Xuewu Zhang, Huangbin You, William Zhu, Shaojie Qiao, Jianwu Li, Louis Alberto Gutierrez, Zhuo Zhang, and Xinnan Fan. 2015. Overlapping community identification approach in online social networks. *Physica A: Statistical Mechanics and its Applications* 421 (2015), 233–248.

[88] Daniel Xiaodan Zhou, Paul Resnick, and Qiaozhu Mei. 2011. Classifying the political leaning of news articles and users from user votes. In *Proceedings of the international AAAI conference on web and social media*, Vol. 5. 417–424.

[89] Haiyi Zhu, Robert E. Kraut, and Aniket Kittur. 2014. The Impact of Membership Overlap on the Survival of Online Communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 281–290. https://doi.org/10.1145/2556288.2557213

A APPENDIX

A.1 Description of Events

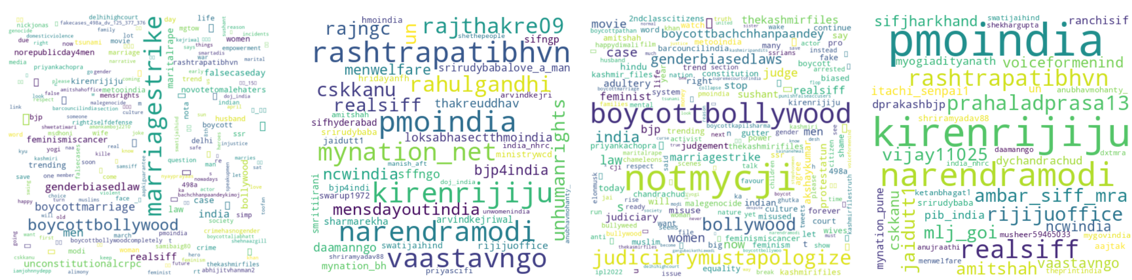
We describe here each event considered in this paper, to provide the reader an idea of the nature, cause, and target groups of their social media communities. **Men's Rights Activism or MRA** consists of different groups and individuals (activists or MRAs) who focus on social issues that adversely impact, or discriminate against men. Some of the commonly discussed topics within MRAs include laws and aspects on child custody, alimony, marital rape, marital property distribution, reproduction, suicides, domestic violence against men, and circumcision [18]. The social media presence of MRAs has elements of legal activism, but also has significant misogynistic and anti-feminist content, with discursive forms and network characteristics that are reflective of systematic propaganda [7, 67, 73]. Messages spread on the movement vary from posting images/text on various cases of violence against men, to insulting women and violence against women as part of retribution.

#NotMyCJI, denoted in the paper as CJI, was driven by organized messaging against the Chief Justice of India by Twitter users opposed to his liberal positions on various issues. However, a key event where he was considered to be directly opposing public comments was where he urged Indian law students students to "incorporate feminist thinking" into their practice of law. While this had sparse or largely positive coverage from mainstream media and commentators, on social media, this sparked a flood of activity, including calls for his removal. Since his feminist position was at the heart of the engagement, a significant driver of content against the CJI on this issue was the MRA community.

The third event which had a significant social media footprint was the **Boycott Bollywood** campaign that frequently encourages people to boycott films made by a certain cohort of the Hindi-language film actors' industry, referred to as *Bollywood* since it is largely run from the state of *Mumbai*, earlier known as *Bombay*. The campaign started gaining momentum starting in the early days of COVID-19, with the death of a popular Indian actor named *Sushant Singh Rajput (SSR)* in June 2020 [2]. His fans claimed that the actor's demise was either a pre-meditated murder or a suicide triggered by systemic bullying from a nepotistic clique in the Mumbai film industry. Although initial investigations into the matter concluded that the death was by suicide, the campaign to investigate various angles as well as to boycott the output of the industry was kept alive by a set of dedicated social media fans or micro-influencers [24]. While the actor's death was a precipitating event, Boycott Bollywood took a communal turn as the discourse turned its attention to Muslim actors in the Hindi film industry [59], some of who are the highest-paid and most successful stars.

A.2 Content Characteristics

Figure 5 shows the content characteristics of the two remaining overlapping pairs – (BB-MRA) and (BB-CJI). The content characteristics for the (MRA-CJI) overlap is discussed in Results section. We find that even for the (BB-MRA) and (BB-CJI) overlaps, the tweets by bridge users seldom mention any political keyword, indicative of the bridges' sincerity of sticking to the cause of the movement. This finding is similar to the result presented in this paper for the (MRA-CJI) overlap.



(a) BB-MRA Keywords (b) BB-MRA Mentions (c) BB-CJI Keywords (d) BB-CJI Mentions

Fig. 5. Word-clouds for top keywords and mentioned political accounts for BB-MRA and BB-CJI overlaps (preprocessing to convert all into lowercase was done first): We find minimal presence of political accounts owing to the presence of BB, an event with a chiefly non-political agenda. We do not show the temporal change in the word-clouds for BB-MRA and BB-CJI (unlike MRA-CJI) since the findings are similar across time.

We also find that the accounts related to the ruling party and its alliance members (like @NarendraModi, @PMOIndia, and @KirenRijiju) are most frequently mentioned by the bridges in BB-MRA and BB-CJI. On a closer analysis of the bridge tweets, we find that most of the political mentions are in the MRA and CJI tweets, rather than BB, which contains negligible tweets mentioning political accounts. This is primarily due to the fact that BB is a dominantly non-political event, which does not have significant mentions of political accounts and policy level interventions, unlike MRA and CJI that include appeals to the ruling party to formulate stronger policies and laws for men.

A.3 Features for Propaganda Detection

We show in table 7 the 18 features used for propagandist message detection as identified by previous work [21]. We used these features for manual annotation of tweets into two classes of propaganda (P) and non-propaganda (NP),

Propaganda Technique	Definition
Name Calling	Attack an object/subject of the propaganda with an insulting label
Repetition	Repeat the same message over and over
Slogans	Use a brief and memorable phrase
Appeal to Fear	Support an idea by instilling fear against other alternatives
Doubt	Questioning the credibility of someone/something
Exaggeration / Minimization	Exaggerate or minimize something
Flag-Waving	Appeal to patriotism or identity
Loaded Language	Appeal to emotions or stereotypes
Reduction ad hitlerum	Disapprove an idea suggesting it is popular with groups hated by the audience
Bandwagon	Appeal to the popularity of an idea
Casual Oversimplification	Assume a simple cause for a complex event
Obfuscation, Intentional Vagueness	Use deliberately unclear and obscure expressions to confuse the audience
Appeal to authority	Use authority's support as evidence
Black and White Fallacy	Present only two options among many
Thought terminating clichés	Phrases that discourage critical thought and meaningful discussions
Red herring	Introduce irrelevant material to distract
Straw men	Refute argument that was not presented
Whataboutism	Charging an opponent with hypocrisy

Table 7. Features used to determine if tweet is Propagandist or Non-Propagandist

corresponding to the three events considered. A tweet was labeled as propagandist if it contained at least one of these features, and if it vilified an individual or target group using a negative or sarcastic tone.