A Tale of Two Sides: Study of Protesters and Counter-protesters on #CitizenshipAmendmentAct Campaign on Twitter

Kumari Neha
Indraprastha Institute of Information Technology, Delhi, India
nehak@iiitd.ac.in

Vibhu Agrawal
Indraprastha Institute of Information Technology, Delhi, India
vibhu18116@iiitd.ac.in

Vishwesh Kumar
Indraprastha Institute of Information Technology, Delhi, India
vishwesh18119@iiitd.ac.in

Tushar Mohan
Indraprastha Institute of Information Technology, Delhi, India
tushar19393@iiitd.ac.in

Abhishek Chopra
Netaji Subhas University of Technology, Delhi, India
abhishekchopra0907@gmail.com

Arun Balaji Buduru
Indraprastha Institute of Information Technology, Delhi, India
arunb@iiitd.ac.in

Rajesh Sharma
Institute of Computer Science, University of Tartu, Estonia
rajesh.sharma@ut.ee

Ponnurangam Kumaraguru
International Institute of Information Technology, Hyderabad, India
pk.guru@iiit.ac.in

ABSTRACT

Online social media platforms have evolved into a significant place for debate around socio-political phenomena such as government policies and bills. Studying online debates on such topics can help infer people's perception and acceptance of the happenings. At the same time, various inauthentic users that often pollute the democratic discussion of the subject need to be weeded out from the debate. The characterization of a campaign keeping in mind various forms of involved actors thus becomes very important. On December 12, 2019, Citizenship Amendment Act (CAA) was enacted by the Indian Government, triggering a debate on whether the act was unfair. In this work, we investigate the user's perception of the #CitizenshipAmendmentAct on Twitter, as the campaign unrolled with divergent discourse in the country. Keeping the campaign participants as the prime focus, we study 9,947,814 tweets produced by 275,111 users during the starting 3 months of protest. Our study includes the analysis of user engagement, content, and network properties with online accounts divided into authentic (genuine users) and inauthentic (bots, suspended, and deleted) users. Our findings show different themes in shared tweets among protesters and counter-protesters. We find presence of inauthentic users on both side of discourse, with counter-protesters having more inauthentic users than protesters. The follow network of the users suggests homophily among users on the same side of discourse and connection between various inauthentic and authentic users. This work contributes to filling the gap of understanding the role of users (from both sides) in a less studied geo-location, India.

WebSci'22, Barcelona, Spain,

ACM ISBN 978-1-4503-9191-7/22/06…$15.00
https://doi.org/10.1145/3501247.3531584

CCS CONCEPTS

• **Information systems** → **Social networks**.

KEYWORDS

Social media manipulation, Twitter, Stance, Bots, Network Analysis, Protest in India

ACM Reference Format:
Kumari Neha, Vibhu Agrawal, Vishwesh Kumar, Tushar Mohan, Abhishek Chopra, Arun Balaji Buduru, Rajesh Sharma, and Ponnurangam Kumaraguru. 2022. A Tale of Two Sides: Study of Protesters and Counter-protesters on #CitizenshipAmendmentAct Campaign on Twitter. In *14th ACM Web Science Conference 2022 (WebSci '22), June 26–29, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3501247.3531584

1 INTRODUCTION

Social media platforms are used as a primary source of information and opinion sharing in recent times [18, 23, 26, 38, 64]. Often heated debate on controversial topics leads to users divided into protesters and counter-protesters on the social media [28, 37, 44]. However, the divergent opinions present on the platform is not just coming from authentic users, but a mix of other content polluters or inauthentic users who also participate in the debate [10]. Previous research has found that the inauthentic users spread the content of a specific stance [60] or conduct influence operations [48] including interaction with authentic users [43]. Due to the presence of different inauthentic users, understanding of authentic participant's stance and perception in a discourse requires weeding out the inauthentic users and their manipulative content.

In this work, we study the online debate about Citizenship Amendment Act (CAA), enacted by the Indian Government on December 12, 2019. The enactment led to a divergent discourse on social media, with users divided in their opinion on the Act. Among the users who participated in the debate, one cohort rejected the Act, while another supported it. We define the users who reject the Act as protesters. The protesters were contested

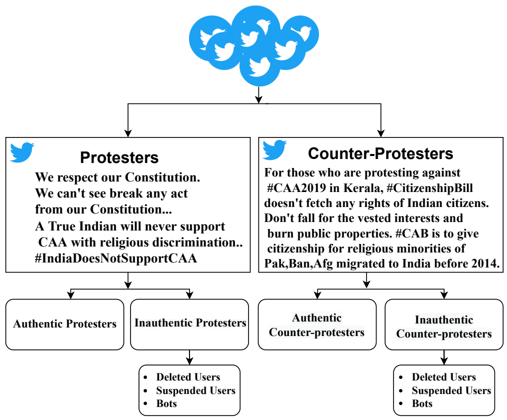


Figure 1: The users considered under study divided into 4 sets.

by a counter-protest campaign that questioned the protest and favored the Act. We define the users who were in favor of the Act as counter-protesters [28]. While the campaign gained traction on both Twitter and the offline world, the prevalence of manipulation of the campaign was found to be evident [32]. Given that the forms of manipulation of a discourse keep on innovating, it becomes crucial to filter the influence created by the inauthentic users in an online campaign. We define bots [60], suspended and deleted users (who tend to disseminate malicious content [1]) who participated in the discourse as *Inauthentic users*. In contrast, *Authentic users* are defined as the users who were not identified as bots, neither were suspended nor deleted. We thus study the online debate on the #CitizenshipAmendmentAct on Twitter with the participants divided into authentic and inauthentic users for both protesters and counter-protesters forming 4 set of users as shown in Figure 1.

Twitter has been the focus of various characterization studies involving online campaigns [21, 28, 51]. However, the characterization of a campaign concerning various sorts of authentic and inauthentic actors in discourse is limited [17]. To the best of our knowledge, we are the first to conduct a characterization study of a campaign with various users (Figure 1) in a less investigated setting, i.e., India. Our analysis contributes to a few recent preliminary studies on the CAA [32, 40] which provide a very coarse-grained analysis of the Act. We focus on a broader study of the Act, covering a larger dataset, multi-lingual tweets, and a richer analysis set.

To this end, we analyze 275, 111 users who post about topics relevant to CAA during the initial three months of the debate from December, 2019 to February, 2020. We seek to understand the interplay of authentic / inauthentic users and pro- / against stance on CAA and investigate the presence and participation of inauthentic users on both sides of the discourse. For the characterization study, we first identify the stance of the participants using unsupervised stance detection approach [20]. We further study the 4 set of participants from the user, content, and network perspective, to obtain a fine-grained analysis of the discourse. Broadly, we aim to answer the following research questions (RQs) through the characterization study of CAA.

RQ 1: *How are the protesters and counter-protesters involved in conducting the online campaign with respect to authentic and inauthentic users?*

The prevalence of inauthentic users has been studied in online campaigns, including elections [13], and more recently, the coronavirus [22]. In the CAA debate, we found the prevalence of inauthentic activity in both side of the debate, with the online protest being highly mediated by the inauthentic users.

RQ 2: *What did the users in the discourse discuss about?*

The discourse analysis helps identify various themes in the discussion to help understand the user's perception [37]. While the themes for protesters / counter-protesters varies in CAA, we also found difference in themes for authentic and inauthentic users in both sides, with inauthentic users posting lesser emotional content than authentic counterpart.

RQ 3: *What was the network structure of the users?*

The analysis of the network structure helps examine issue alignment [69], and polarization around a controversial topic [29]. The follow network of users show homophily, where users with similar stance follow each other more than users with opposing stance. The analysis of the follow network shows edges between authentic and inauthentic users, showing risk of exposure of content from inauthentic users to the authentic users.

Our findings reveal the interplay of inauthentic and authentic users in the online discourse around CAA. Prevalence of inauthentic activity was found on both sides of the debate. However, user characterization reveals that inauthentic users are more prevalent in the counter-protesters than protesters. The content analysis of the 4 set of users shows that the inauthentic users highly mediated the online protest. Emotional analysis of the content posted by the 4 set of users shows that the inauthentic users use less emotional tweets than their authentic counterparts. Through follow network of the users, we found evidence of homophily in the network. However, the edges between various inauthentic and authentic users shows their connectedness, indicating risk of manipulating authentic users.

Background: In India, the first Citizenship Act was enacted in 1955, which enlisted the routes to obtain citizenship in India which includes birth, descent, registration, naturalization, and acquisition of a foreign territory. The amendment in the Act in 2019 (CAA 2019), allows the minority communities to apply for citizenship via registration or naturalization [16], with the caveat that migrants who have faced religious persecution in Afghanistan, Bangladesh or Pakistan, can seek citizenship in India if they have entered India on or before December 31, 2014 [16]. The debate on the non-secular roots of the Act were rooted in the exclusion of other religious minority communities like Rohingya Muslims, Jews, Bahais, Zoroastrians to seek citizenship. The protesters deemed it unconstitutional for being discriminatory on religious grounds, as only certain persecuted illegal immigrants benefited from the Act. While the supporters / counter-protesters based their argument on the presumption that refugees of particular minority religious communities are more in need of asylum [16].

[1]https://help.twitter.com/en/rules-and-policies/enforcement-options

2 RELATED WORK

Protests are a form of collective sociopolitical action in which members with similar beliefs express their objection to a cause or situation [7]. Time and again, the world witnesses protests over a government policy, bill [56, 72], or the government itself [64]. In online discussions related to societal issues, users in one group may show hatred for users with opposing views [72]. The *"no ban, no wall"* and *"day without immigrants"* protests are examples of people's divide on social media in their opinion to resist the punitive immigration policy [72]. #BlackLivesMatter (#BLM) is another campaign where the people on social media were divided into two groups [28]. Researchers studying online protests and campaigns on micro-blogging websites have used various stance detection techniques [45] and news articles [11, 57] to identify opposing views automatically. More recently, researchers have focused on opinion modelling, which reflects and justifies the belief or judgment of a person towards a target entity, irrespective of having the same stance [31]. The previous literature studied the contrasting opinions through computing topic models followed by Jensen-Shannon divergence among the individual topic opinions [24]. The different perspectives or viewpoints have also been explored using a graph partitioning method that exploits the social interaction between the users [52]. Previous research has also shown almost 75% of the protests are planned in advance [12]. There has been a lot of interest in the social media domain to predict the on-ground activity through the social media platform [21, 47, 53, 58]. The authors in [72] used protest as an intervention to reduce online prejudice, with focus on manual annotation for understanding prejudice in the tweets [72]. The study of protests have also been studied in regards to the volume of the status messages relating to the protest event [21, 28].

Table 1: Manually identified protest and counter-protest hashtags from trending topics during the period of data collection used for data collection.

Protest #tags	#CABProtest, #IndiaRejectsCAB, #HindusAgainstCAB, #SCSTOBC_Against_CAB, #IndiansAgainstCAB, #IndiaAgainstCAA, #CAA_NRC_Protest, #CAAprotests, #CAA_NRCProtests
Counter-protest #tags	#IsupportCAB2019, #HindusSupportCAB, #IndiaSupportsCAB, #ISupportCAA_NRC, #MuslimsWithNRC, #CAA_NRC_support, #ISupportCAA
Ambiguous #tags	#CAB, #CABBill, #cab, #CAB2019, #CitizenshipAmendmentAct, #caa, #CABPolitics, #CitizenshipAmmendmentAct

While social media has been used to share opinions and debate on current happenings [62], the involvement of inauthentic users is becoming more prevalent on the platform [25]. The manipulation of the debate are studied with regards to bots [13, 17, 60, 67], pre-defined campaign toolkit users [36], co-ordinated accounts [49, 50, 61], or trolls [30, 39]. Social media manipulation has been extensively studied with respect to election campaigns [13, 67]. In social media, bots refer to fully automated and semi-automated accounts that contribute to disinformation campaigns [25]. Uyheng and Carley [67] studied how bots propagate misinformation during electoral campaigns and found that bots participate in online discourse in high numbers and interact with humans via the use of mentions. The bots also share partisan or irrelevant content to pollute the discourse [22, 67]. While bot accounts that use abusive language are more likely to be suspended by Twitter [22, 67], social media manipulation might involve propaganda [30], or campaign toolkits [36], which do not necessarily use abusive language. Russian trolls' involvement during the 2016 US presidential elections are evidence of campaign manipulation through social media accounts that were not necessarily humans and were controlled by certain intelligence agencies [39].

In this paper, we contribute to the use of social media manipulation in other than western context during an online protest and study the online debate with different users involvement in India, a country in Asia-pacific.

3 DATA COLLECTION

Using the official Twitter API, we collect tweets around CAA between December 07, 2019, and February 27, 2020, through daily trending hashtags around the topic. The list of hashtags used for data collection is shown in Table 1. Our collected data consists of 11,350,276 tweets, with 1,543,805 unique tweets and 9,806,471 retweets from 931,175 users. We first collate all the tweets from a given user to identify users actively tweeting about the topic. Hence, we consider users who have at least five tweets during the period of data collection. The total number of users after the filtration process came down to 276,149.

Table 2: On-ground activities coincident with peak tweet days.

Date	Tweets	On-ground activities
December 11	158, 134.33	CAB passed by the upper house of parliament [19].
December 16	376, 788.00	Student protests in Delhi [71].
December 17	379, 699.00	Protest turns violent in Uttar Pradesh, Delhi, West Bengal and relaxed in Guwhati [9, 34].
December 20	436, 616.33	Protesters turn violent with stone pelting in Gujarat, police vehicle burnt in UP, journalists detained in Kerala [4].
December 22	783, 662.33	Protesters arrested, Women protest in Guwhati [59].
December 24	503, 779.00	Protesters die due to bullet injury in UP [2].
December 30	276, 724.33	Counter-protest rally in Madhya Pradesh, Indian-American protests in Washington [3, 33].
December 31	312569.66	Nation wide protests [5, 35].

Data Pre-processing: Twitter users often use various emoticons, emojis, media links, hashtags, and other non-alphabetic characters. The informal nature of Twitter often leads to spelling and grammatical errors or incomplete sentences. Thus, we follow the below list of pre-processing steps for the tweets before further analysis.

(1) Removal of all links and mentions from the tweets
(2) Removal of "RT" keyword from the beginning of retweets
(3) Split of the camel case words into distinct words
(4) Removal of punctuation marks
(5) Removal of extra spaces
(6) Replacement of digits with the word <number>
(7) Case-folding where we lower-cased letters
(8) Desertion of tweet if it had lesser than three terms left after all the above steps

After the pre-processing steps, 1,038 users were disregarded for further analysis. The study conducted in the paper was thus on the 275,111 users, who were most active during the campaign and their tweets contained substantial information for further analysis. For further division of the users into authentic / inauthentic, as shown in Figure 1, we query the Twitter API and botometer [73] on the user IDs obtained from tweets.

Inauthentic vs Authentic Users: The inauthentic users that we consider for the study include suspended users, deleted users and bots. Table 4 shows the total number of deleted and suspended users identified through querying the official Twitter API. We further collect the follower network using the official Twitter API for the users who were not deleted/ suspended/ private. We use Botometer [73], a tool used to identify a Twitter user as being automated (partially or fully) or not. Due to botometer API constraint, we collect the bot score for randomly selected 26,110 users (roughly equal to the total number of suspended/ deleted accounts in our dataset). We use the *Cumulative Automation Score* (CAP) score metric provided by the API to identify a user as a bot account.

On-ground activity: To identify the impact of on-ground activities on opinion sharing around CAA, we manually curate the on-ground activities of the peak tweeting days, as shown in Table 2. The first online tweet peak was seen on December 11, 2019, which coincided with the bill passed as Act by the Rajya Sabha (upper house) of the Indian parliament [70]. However, the highest peak was found on December 20, 2019, 9 days after the bill became an Act. On December 20, 2019, protesters around the country turned violent. A major protest was witnessed about the CAA bill in Guwahati (north-east state of India) on December 10, 2019, which was the beginning of the chain of protests in certain parts of the country.

The anonymized version of our data is available at https://precog.iiit.ac.in/resources.html

4 USER CHARACTERIZATION

To capture the fine-grained divergence among the users, we build on the previous work by Rashed et al. [54] that uses text-feature for identification of user's stance during a political campaign. We further identify the themes in shared tweets and discuss the presence of inauthentic users in the discourse.

4.1 Understanding the discourse through unsupervised stance detection

Based on the online discourse on the Act, we identify two cohorts of users. We call the users who opposed CAA as protesters. While users who share tweets in support of CAA are called counter-protesters. Rashed et al. [55] proposed unsupervised stance detection techniques based on the text of the tweets. Another reason for the choice of algorithm is to surpass the manual annotation required in a supervised setting.

The ground truth labelling process for the seed set of users constitutes of two steps:

(1) Manual Labelling: First, we manually identify a set of hashtags indicating stance, as shown in Table 1. We identified 27 hashtags as counter-protest hashtags on manual inspection, which occurred in over 1.3 million tweets. The count of protest hashtags were 48, which accounted for around 1.04 million tweets. In the first step of labelling, if a user used only counter-protest hashtags and never used protest hashtags, we label the user as counter-protester. Similarly, if a user used only protest hashtags, we classify the user as a protester. In the first level of manual labelling, we identified 106,605 users as counter-protesters and 79,493 users as protesters.

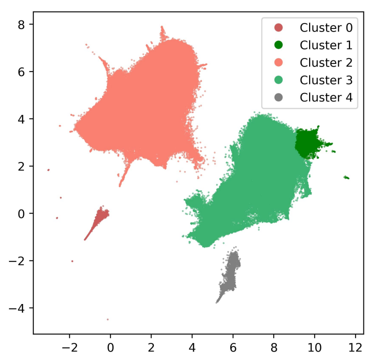


Figure 2: Here Clusters 0 and 2 represent counter-protest users and Clusters 1 and 3 represent protest users. Cluster 4 had a purity below 80% and hence was not considered.

(2) Label Propagation: Around 86% of the tweets in our dataset were retweets. Based on the tweets that a user retweets, users were further labelled such that a user with at least 15 retweets from protest and none from counter-protest side belongs to protesters. The intuition behind this approach is that the users retweet a given tweet if it aligns with their stance. We conduct this approach for two rounds. After the two rounds of label propagation, 114,977 users were identified as counter-protesters, while 79,613 were identified as protesters. The tweets of identified users were further pre-processed and users with less than five tweets were disregarded. The final set of users after the pre-processing is 270,889.

Embedding-based Stance Detection: The word-based embedding can capture fine-grained divergence between two sets of cohorts [55]. We apply LASER (Language-Agnostic Sentence Representations)[2] to obtain 1024-dimensional embeddings of users based on their tweets. LASER is a sentence encoder trained on 93 languages, including many Indian regional languages. To obtain user-level embedding, we use the average of the vector for the filtered tweets. The users are then projected in a 2-dimensional space using Uniform Manifold Approximation and Projection (UMAP) algorithm [42]. The projection of users on lower dimension helps overcome the curse of dimensionality [68]. UMAP projects the data elements closer if they are similar, while dissimilar data elements

[2] https://github.com/facebookresearch/LASER

are placed far apart. The projected user vectors are further clustered using hierarchical density-based clustering (HDBSCAN) [41]. Using the HDBSCAN algorithm, 5 clusters were formed, with the 270,889 users.

We consider clusters pure if they contain at least 30% of labelled users obtained via label propagation. We found 4 clusters have more than 80% purity of labels, as shown in Figure 2. Clusters 0 and 2 were identified as counter-protesters, while clusters 1 and 3 were identified as protesters' clusters according to the labelled users. The number of users identified in the 4 clusters was 263,869 users, with 142,839 counter-protesters and 121,030 protesters.

Topics discussed by the users in the different clusters:

Among the 4 clusters with high purity, the protesters are represented with shades of green, and counter-protesters are represented with shades of red, as shown in Figure 2. The two major clusters of opposing views (cluster 2 and cluster 3) shows rich discourse on the topic. For manual inspection of assigned clusters, we randomly picked 4 sets of 10 users from each cluster, and annotated all tweets for these users. We found the users in the clusters were indeed on the protester and counter-protester side, as identified through label propagation. To understand the theme of the 2 protester's clusters and 2 counter-protesters clusters, we go through all the tweets form the 4 sets. The topics discussed by the two cohorts in the 4 clusters shown in Figure 2 follow different themes as follows:

Cluster 0: (Counter-protesters) On a more thematic side, we found that the topics discussed by the users in Cluster 0 are mostly informative, with users sharing opinions on why CAA should be implemented.

Cluster 2: (Counter-protesters) The primary topic discussed by the users of this cluster includes questioning the protester about their actions and reasons for their disagreement with the implementation of CAA.

Cluster 1: (Protesters) The users in this cluster were tweeting about the on-ground activity of the protest, including public demonstrations, stone pelting, etc.

Cluster 3: (Protesters) The users in the cluster were posting informative tweets about CAA in the protest context.

4.2 Presence of authentic and inauthentic users in the discourse

Next, we identify users based on their authentic behaviour to study the role of inauthentic users in the mobilization of protests and counter-protests. As shown in Table 3, among the 263,869 users considered for the analysis, we found 13,871 users were suspended by Twitter, while 13,251 users were not found (referred to as deleted users) when queried for follower network. The number of non-authorized (private users) was 5,844. We were unable to retrieve information of 11,091 users using Twitter API. The Inauthentic users obtained so far is 27,122, including suspended and deleted users. Next, we use botometer API [73] to identify bot users. Given the limitation of botometer API, we randomly pick 27,122 users from the rest of the users to query botometer for bot scores. We could retrieve bot scores for 26,110 users out of which 14,970 were counter-protesters and 11,140 are protesters. Table 4 shows the complete set of users considered for the analysis.

Table 3: Distribution of suspended and deleted accounts in protesters and counter-protesters in the dataset.

	Suspended Users	Deleted User
Counter-protesters	8655 (62.39%)	7440 (56.16%)
Protesters	5216 (37.60%)	5806 (43.83%)

Table 4: Distribution of authentic and inauthentic users in dataset.

Total Users	53, 227
Suspended Users	13, 871
Deleted Users	13, 246
Bots (CAP score>=0.8)	4, 664
Authentic Users	21, 446

Table 5: Distribution bots in the discourse with varying bot scores. P: protesters, CP: counter-protesters, T: total number of users for which botscore is known in our analysis.

Bot score (>=)	CP (% bots in CP)	Protesters (% bots in P)	Total (% bots in T)
0.8	2,589 (17.29%)	2,075 (18.62%)	4,664 (17.86%)
0.7	11,359 (75.87%)	8,214 (73.73%)	19,573 (74.96%)
0.6	12,706 (84.87%)	9,096 (81.65%)	21,802 (83.50%)
0.5	13,500 (90.18%)	9,688 (86.96%)	23,188 (88.80%)

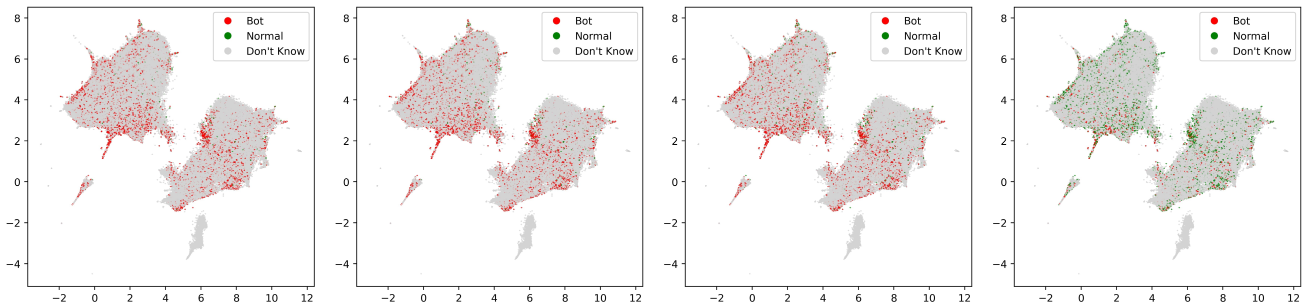
Findings: Through user characterization, we infer that both sides of the discourse had suspended, deleted users and bots. Counter-protesters had more than 50% suspended or deleted users on the platform, as shown in Table 3. Figure 4 shows the distribution of bots in the stance based cluster. We notice, as shown in Figure 3 and Table 5 that as the bot score varies from 0.8 to 0.5, there is a sharp decline of bots above 0.7. This shows the presence of semi-automated accounts in the discourse.

5 CONTENT CHARACTERIZATION

Through content characterization, we try to understand the interplay between the online and offline activities during the period of data collection and quantify the difference in opinion among the 4 set of users.

5.1 Online (Twitter) Vs. offline (on-ground) activity

Taking cues from previous works around planned protests [12, 47], we investigate the interplay of the online and on-ground activities during the CAA discourse, with respect to the 4 set of users in Table 4. Figure 5 shows the frequency of tweets by the 4 set of users during the 2 month of the protest period. The x-axis represents the days of protest taken as rolling average of 3 days (one day before the date and one day after). The on-ground activities corresponding to peaks in tweets are listed in the Table 2. The first peak in the



(a) bot score >= 0.5

(b) bot score >= 0.6

(c) bot score >= 0.7

(d) bot score >= 0.8

Figure 3: Distribution of the users with varying bot scores ranging from from 0.6-0.8.

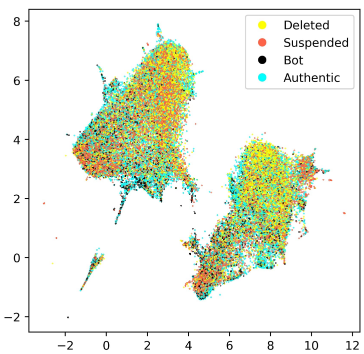


Figure 4: The presence of 4 set of users in the cluster.

dataset was on December 11, 2019, when the CAB (Citizenship Amendment Bill) was passed by the upper house of parliament and officially became an Act [19]. Students in Assam held protest opposing the Act [1] on this day. In the initial few days, authentic protesters were more active than inauthentic protesters. While there was almost an equal proportion of authentic vs inauthentic tweets during the initial days of passing of the bill. Another significant day was *December* 16, 2019, when students led the protest across the country, including Delhi, Maharashtra, and UP [71]. Anarchy was observed the same day in West Bengal, where people torched trains and staged sit-ins on the railway tracks [8]. Inauthentic counter-protesters made most tweets at this day, followed by authentic protesters. On *December* 17, 2019, several metro stations in Delhi [63] were closed and Section 144[3] was imposed in UP. The previous trend of high tweets from inauthentic counter-protesters followed by high tweets from authentic protesters continued.

December 20, 2019 witnessed nationwide protest eruption including states of Uttar Pradesh, Tamil Nadu, and Delhi [4]. The government opened to suggestions and reaching out to the protesters [4]. While the inauthentic counter-protesters were more active than inauthentic protester during the period, authentic counter-protesters made more tweets on around December 20 than authentic protesters.

December 22, 2019 had the largest peak in the dataset with on-ground counter-part of protesters being arrested and women leading the protest in Guwhati [59]. Both Inauthentic and authentic counter-protesters were more active around this day. *December* 24, 2019 showed the second largest peak in the dataset, which co-incided with protester's death in Uttar Pradesh, due to bullet injury [2]. The spikes on *December* 30, 2019 and *December* 30, 2019 found counter-protesters more actively posting than protesters. The on-ground activities for the day included continued protests in different parts of the country as well as abroad in Washington [3]. The counter-protesters started rallies on *December* 30, 2019 in support of CAA in different parts of the country [33]. One of the dip in tweets that we find was on *December* 19, 2019, when internet was shut down in many parts of the country [6].

The counter-protesters had more inauthentic activities during the start of the timeline, until the largest peak. After which both authentic and inauthentic protesters showed more activity than counter-protesters. While there was a mix of authentic and inauthentic activity found in both protesters and counter-protesters, the activities of inauthentic counter-protesters were always more than than the authentic protesters. While, in case of protesters, authentic users always dominated the conversation. A common pattern in all the peaks found was that more that 90% of the authors in the timeline during any peak were from inauthentic users.

5.2 Difference in opinion

We use LDA [14] for topic modeling and word shift graphs [27] to understand how diversified content were posted by the 4 set of users during the discourse. Table 6 shows the topics discussed among the

Table 6: Summary of topics for authentic and inauthentic protesters.

Authentic protesters topics	
Topic 1	india, bjp, police, muslim, student
Topic 2	police, student, hindu, assam, people
Topic 3	muslim, jamia, anti, student, delhi
Inauthentic protesters topics	
Topic 1	muslim, protest, hindu, student, protest
Topic 2	country, display, protest, acceptance, together
Topic 3	people, india, protest, police, citizenship

[3]https://www.aninews.in/news/national/general-news/up-section-144-imposed-in-rampur-after-protest-against-caa20191217125542/

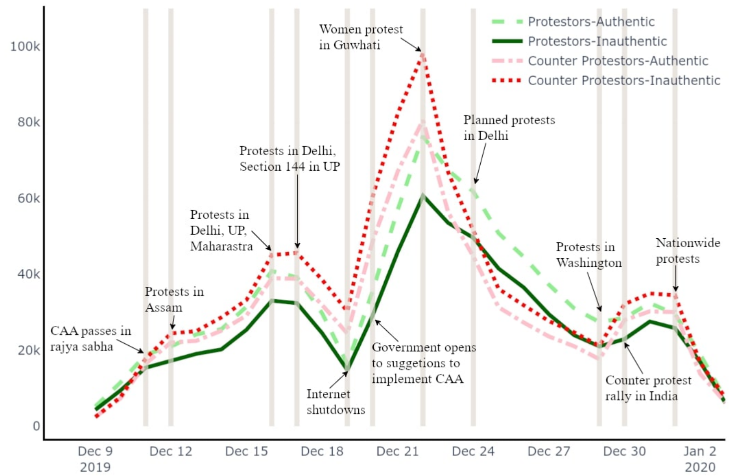


Figure 5: Timeline of counter-protest and protest vs on-ground activity

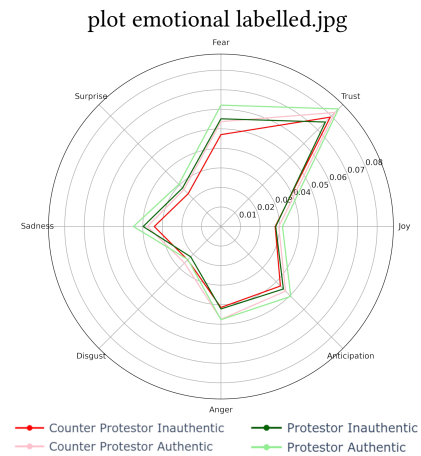


Figure 6: Radar plot to show the 4 set of users and their plutchik-8 emotions.

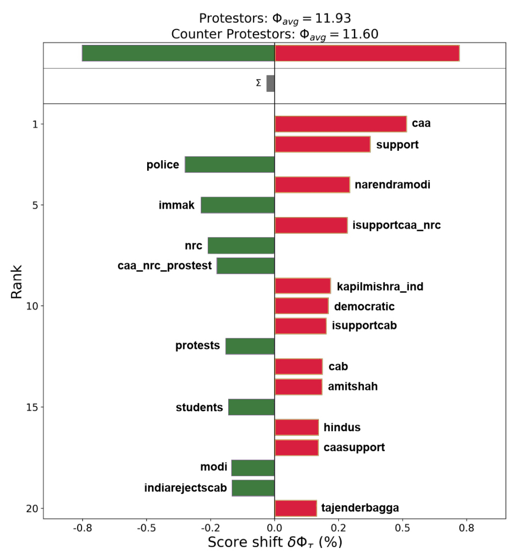


Figure 7: Application of word shift graphs for highlighting narratives that characterize protesters and counter-protesters. Protesters are shown in green, while counter-protesters are shown in red.

authentic and inauthentic protesters. Two of the dominant topics in authentic protesters had religious words, including hindu and muslim. The third topic included police and places of protest. While, the inauthentic protesters had one topic on religion, other 2 major topics included, citizenship, country and India as words. Table 7 shows the topics discussed by the authentic and inauthentic counter-protesters. While one major topic from authentic counter-protesters was support, hindu and caa, the second topic included politicians, and country. For authentic counter-protesters, best coherence score yields 2 topics. The inauthentic counter-protesters had one major topic including politicians, while another two dominant topics

Table 7: Summary of topics for authentic and inauthentic counter-protesters.

Authentic counter-protesters topics	
Topic 1	caa, support, people, anti, hindu
Topic 2	india, narendramodi, today, country
Inauthentic counter-protesters topics	
Topic 1	narendramodi, amitshah, kapilmishra_ind, delhi
Topic 2	hindu, support, indian, pakistan, citizenship
Topic 3	caa, support, democratic, india, humanitarian

included citizenship, India and democracy and support as narrative. We report the most significant topics from the 4 set of users due to limited space.

From the above analysis, we conclude that both protesters and counter-protesters discussed topics around religion, politician and the Act in general. However, the inauthentic users share content very similar to authentic counter-part, thus risking authentic users into believing them as authentic users.

Next, we gauge the frequency of usage of various topics by the 4 set of users through word-shift graphs [27]. We use Shannon's entropy as a measure of diversity, where high Shannon entropy implies the text is less predictable [28] implying more diverse content.

Figure 7 shows that protesters talked more about student, while counter-protesters talked more about hindus. We further study what do the authentic and inauthentic protesters / counter-protesters share more frequently. Figure 8 shows that inauthentic counter-protesters are more appealing (e.g: humanitarian, solidarity, secular), while inauthentic protesters more frequently use words that show mistrust in government. Authentic users on both side are more frequently talking about protest and violence.

5.3 Emotion Analysis

We use NRC lexicon [46] that consists of 8 emotions developed from crowd-sourced manual annotation to identify the emotions of the users in the 4 set of users considered in the study. The NRC

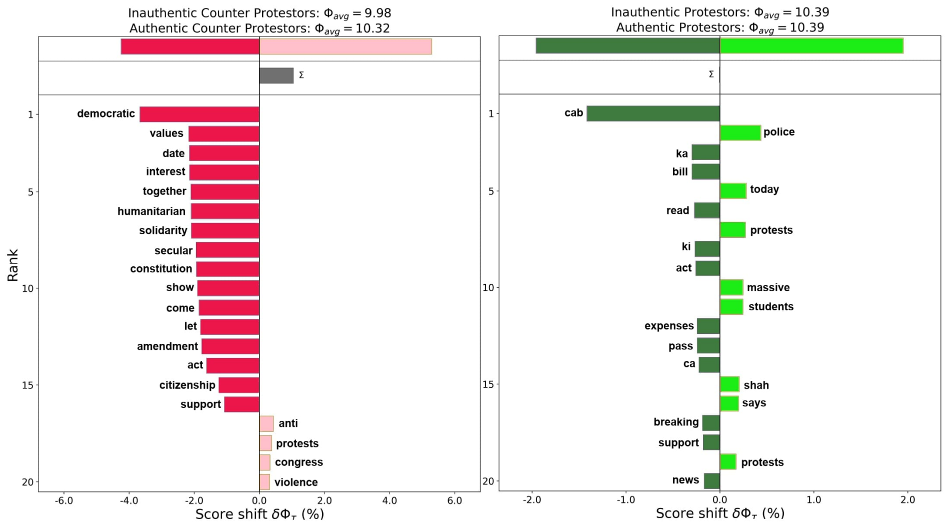


Figure 8: Application of word-shift graph for highligting narratives that charactrize 4 set of users.

lexicon uses the plutchik's 8 wheel of emotion for English, as well as other translated Indian languages. The 8 emotions that are used in the analysis include, anger, anticipation, disgust, fear, joy, sadness, surprise, trust. Figure 6 shows that the authentic protesters had most dominant emotions for all the 8 categories. The authentic counter-protesters and inauthentic protesters had almost similar emotions for fear, surprise, sadness. The inauthentic counter-protesters had least emotional content among the 4 set of users.

6 NETWORK CHARACTERIZATION

To determine if protesters and counter-protesters are in homophily and how authentic and inauthentic users are connected, we study the follow network of users in our dataset. We build a follow graph induced by the users in the dataset for network characterisation. The users for whom the follow network was obtained from Twitter API exclude private accounts and accounts for which information was not obtained due to API constraints. The final follow network was obtained for $226,412$ users. First $5,000$ followers were retrieved from Twitter API for each user from the sample. We consider the graph of $226,412$ users as G. Directed edge from user x to user y exists if x follows y. We use this convention to ensure the network under study is campaign-specific, as participants in the online debate constrain the edges in the graph G. The graph G contains $21,495,449$ edges, and $226,412$ vertices. We found $33,278$ connected components in the network. The largest strongly connected component contains $192,903$ users with $89,377$ protesters and $103,526$ counter-protesters. Since a strongly connected component in a directed graph is its maximal strongly connected subgraphs, the presence of both protesters and counter-protesters in the largest strongly connected sub-graph indicates the path between the protesters and counter-protesters [65]. The betweenness centrality of the graph G is $9.80e^{-06}$ (SD $1.388e^{-07}$), which indicates how much a node appears in the shortest path between two nodes. Since the network has very low betweenness centrality, this implies that the users in the network do not occur in many shortest paths in the follow network. The average eigenvector centrality for the network is 0.00056 (SD $4.25e^{-06}$), which shows that the users in the network are connected to influential neighbours, i.e., user-nodes which themselves have high eigenvector centrality (or high in-degree). The network density is 0.0004 indicating a sparse follow network. Figure 9 shows the follower-followee graph of $10,000$ random users selected from $263,869$ users. We experimented with different random samples of $10,000$ users to check for consistency in network structure and observed a similar structure across various random sampled networks. In Figure 9, we can observe two distinct clusters of follow network, clearly showing homophily among the users. The analysis of the graph G shows that the CAA debate

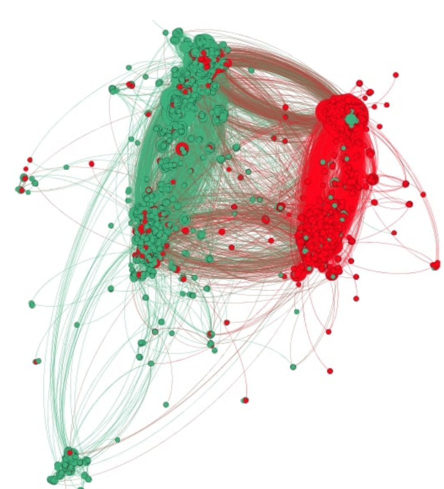


Figure 9: Overall follower-followee network of the protesters and counter-protesters. protesters are represented by green colour while counter-protesters by red colour.

Table 8: Network descriptive statistics for the authentic and bot accounts who participated in the discourse. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ **analyzed using unpaired Mann–Whitney U test. SD stands for Standard Deviation.**

	Authentic Users		Inauthentic Users (Bots)		
Metric	Mean	SD	Mean	SD	p
Number of Followers	22.91	43.84	27.57	46.49	$***$ ($5.5e^{-32}$)
Number of Followees	22.43	61.00	29.70	72.50	$***$ ($9.07e^{-09}$)
Eigenvector Centrality	0.002	0.006	0.003	0.007	$***$ ($2.55e^{-26}$)
Betweeness Centrality	0.00011	0.0004	0.0001	0.00038	$**$ (0.01)

on Twitter was conducted by campaigners who were connected to both sides of the debate; were not strongly connected among each other, forming a sparse network; were connected to many influential users on the platform.

Follow network for authentic and inauthentic users: In order to gauge presence of inauthentic users, we construct a graph H from a set of authentic and inauthentic users (bot scores ($>= 0.8$)). We see a mix of different set of users in the follow network, indicating that the inauthentic users are connected with the authentic users. Consequently, exposing authentic users to the content posted by the inauthentic users. We study the authentic and inauthentic users in the graph H and discuss the network descriptive statistics of authentic and inauthentic users. Table 8 shows the difference in the authentic and inauthentic users with respect to various network descriptive statistics measures. We see there is a very significant difference between the followers and followees of the authentic and inauthentic users. The inauthentic users tend to have a higher follower and followee than the authentic counterparts. The eigenvector centrality shows a significant difference among the authentic and inauthentic users, with bot being prominent in both the measures. As a result, inauthentic users are more reachable than authentic users and have a stronger influence in the network as compared to the authentic users.

7 CONCLUSION

In this work, we characterize the Citizenship Amendment Act (CAA) discourse on Twitter, with respect to various authentic and inauthentic users. We identify the participants' stance using unsupervised learning in a multilingual setup. Using the sampled cluster analysis, we were also able to identify major topics of the discourse from both protester's and counter-protester's standpoints. We further study the presence and perception of various authentic and inauthentic actors in the discourse. The inauthentic actors considered for the study are bots, suspended and deleted users. Users who were not deleted, suspended, or bots were considered Authentic users. To this end, we collected 9 million tweets revolving around CAA through trending hashtags in India. Our findings suggest the presence of inauthentic activities on both sides of the discourse. However, counter-protesters show more inauthentic activity than protesters. We observe through tweets frequency over the timeline that most of the discussion was driven by inauthentic users, who also post lesser emotional content than their authentic counterparts. The content shared by authentic users on both sides mainly revolved around violence and protest, while inauthentic user's posts were more appealing. The follower network of the participants reveals the presence of homophily, where users with similar stances tend to follow each other. One of the largest connected components in the follower network suggests the presence of a path between authentic and inauthentic users, suggesting rechability of inauthentic users to their authentic counterparts.

8 LIMITATIONS AND ETHICAL CONCERN

We understand that the study of a social media campaign is not without caveat. The first being hashtags covered in the data collection might not cover all the aspects of the campaign. The reliance on a single media platform and public APIs is another limitation of the work [15, 66]. The hashtags that we use for the analysis constitute the most recent discussion on the topic, which can not capture the long-term evolution of the debate. There are important ethical concerns while working with the user on social media.

We understand that although the profile data is publicly available, it is inherently sensitive. For example, users who post about the campaign might not anticipate the use of their data by anyone, especially around sensitive topics such as opinion on CAA or their follow network. We note that data collected in this study is from the publicly available set of information, and we do not attempt to explore any user-level demographic information. We emphasize that the opinion shared by the users on the campaign are broadly studied to understand public perception on #CitizenshipAmendmentAct, and not on an individual level. While sharing tweet IDs is a common practice in such studies, there is a risk to share the Tweet IDs due to sensitive nature of the campaign. For example, if we share the tweet IDs, there is a risk of obtaining all the user-level information from the tweet ID. Hence, we opt out of sharing the tweet IDs used in the study. Instead, we share tweet and user-level features without revealing personal information such as profile name, profile description, username etc.

ACKNOWLEDGMENTS

We thank Precog Research Group for stimulating discussions and giving us constructive feedback. Special thanks to Mallika Subramanian (IIIT Hyderabad) and Tanvi Karandikar (IIIT Hyderabad) for their help and discussions. Rajesh Sharma is supported by EU H2020 program under the SoBigData++ project (grant agreement No. 871042).

REFERENCES

[1] 2019. CAB protest: Students clash with police near Assam secretariat. *Economic Times* (2019).
[2] 2019. Anti-CAA Protests Highlights December 24: Rahul and Priyanka Gandhi stopped outside Meerut by police, returns Delhi. *jagran News Desk* (2019).
[3] 2019. Indian-Americans protest CAA third time in nine days, author=IANS. *freepressjournal* (2019).

[4] 2019. Breaking news on December 20, author=India TV News Desk. *newindianexpress* (2019).

[5] 2019. Flash protest in Hyderabad again CAA, NRC on New Year's eve: Six detained. *The News Minute*. (2019).

[6] 2019. Internet shutdowns. (May 2019). https://internetshutdowns.in/static-page/caa-protest/

[7] Edwin Amenta and Michael P. Young. 1999. Democratic States and Social Movements: Theoretical Arguments and Hypotheses. *Social Problems* 46, 2 (1999), 153–168. http://www.jstor.org/stable/3097250

[8] ANI. 2019. Amid protests, Mamata Banerjee announces mega rally in Kolkata to protest against CAA, NRC today. *freepressjournal* (2019).

[9] ANI. 2019. CAA protest row: Section 144 imposed in UP Mau's Hajipura Chowk area. *freepressjournal* (2019).

[10] Ahmer Arif, Leo Graiden Stewart, and Kate Starbird. 2018. Acting the part: Examining information operations within# BlackLivesMatter discourse. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–27.

[11] Rawia Awadallah, Maya Ramanath, and Gerhard Weikum. 2012. Harmony and dissonance: organizing the people's voices on political controversies. In *Proceedings of the fifth ACM international conference on Web search and data mining*. 523–532.

[12] Mohsen Bahrami, Yasin Findik, Burçin Bozkaya, and Selim Balcisoy. 2018. Twitter Reveals: Using Twitter Analytics to Predict Public Protests. *CoRR* abs/1805.00358 (2018). arXiv:1805.00358 http://arxiv.org/abs/1805.00358

[13] Alessandro Bessi and Emilio Ferrara. 2016. Social bots distort the 2016 US Presidential election online discussion. *First monday* 21, 11-7 (2016).

[14] David M Blei, Andrew Y Ng, and Michael I Jordan. 2002. Latent dirichlet allocation. In *Advances in neural information processing systems*. 601–608.

[15] Danah Boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5 (2012), 662–679.

[16] Abhinav Chandrachud. 2020. Secularism and the Citizenship Amendment Act. *Indian Law Review* 4, 2 (2020), 138–162. https://doi.org/10.1080/24730580.2020.1757927 arXiv:https://doi.org/10.1080/24730580.2020.1757927

[17] Ho-Chun Herbert Chang, Emily Chen, Meiqing Zhang, Goran Muric, and Emilio Ferrara. 2021. Social Bots and Social Media Manipulation in 2020: The Year in Review. *arXiv preprint arXiv:2102.08436* (2021).

[18] Danish Contractor, Bhupesh Chawda, Sameep Mehta, L. Subramaniam, and Tanveer A. Faruquie. 2015. Tracking Political Elections on Social Media: Applications and Experience. In *IJCAI*.

[19] Vijaita Singh Damini Nath. 2019. After a heated debate, Rajya Sabha clears Citizenship (Amendment) Bill. *The Hindu* (2019).

[20] Kareem Darwish, Peter Stefanov, Michaël Aupetit, and Preslav Nakov. 2020. Unsupervised User Stance Detection on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* 14, 1 (May 2020), 141–152. https://ojs.aaai.org/index.php/ICWSM/article/view/7286

[21] Munmun De Choudhury, Shagun Jhaver, Benjamin Sugar, and Ingmar Weber. 2016. Social media participation in an activist movement for racial equality. In *Proceedings of the... International AAAI Conference on Weblogs and Social Media. International AAAI Conference on Weblogs and Social Media*, Vol. 2016. NIH Public Access, 92.

[22] Adam G Dunn, Didi Surian, Jason Dalmazzo, Dana Rezazadegan, Maryke Steffens, Amalie Dyda, Julie Leask, Enrico Coiera, Aditi Dey, and Kenneth D Mandl. 2020. Limited role of bots in spreading vaccine-critical information among active twitter users in the United States: 2017–2019. *American Journal of Public Health* 110, S3 (2020), S319–S325.

[23] Mai ElSherief, Elizabeth Belding, and Dana Nguyen. 2017. #notokay: Understanding gender-based violence in social media. In *Eleventh International AAAI Conference on Web and Social Media*.

[24] Yi Fang, Luo Si, Naveen Somasundaram, and Zhengtao Yu. 2012. Mining contrastive opinions on political texts using cross-perspective topic model. In *Proceedings of the fifth ACM international conference on Web search and data mining*. 63–72.

[25] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.

[26] Anjalie Field, Gayatri Bhat, and Yulia Tsvetkov. 2019. Contextual Affective Analysis: A Case Study of People Portrayals in Online #MeToo Stories. *Proceedings of the International AAAI Conference on Web and Social Media* 13 (Jul. 2019), 158–169. https://www.aaai.org/ojs/index.php/ICWSM/article/view/3358

[27] Ryan J. Gallagher, Morgan R. Frank, Lewis Mitchell, Aaron J. Schwartz, Andrew J. Reagan, Christopher M. Danforth, and Peter Sheridan Dodds. 2020. Generalized Word Shift Graphs: A Method for Visualizing and Explaining Pairwise Comparisons Between Texts. *CoRR* abs/2008.02250 (2020). arXiv:2008.02250 https://arxiv.org/abs/2008.02250

[28] Ryan J. Gallagher, Andrew J. Reagan, Christopher M. Danforth, and Peter Sheridan Dodds. 2018. Divergent discourse between protests and counter-protests: #BlackLivesMatter and #AllLivesMatter. *PLOS ONE* 13, 4 (04 2018), 1–23. https://doi.org/10.1371/journal.pone.0195644

[29] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying Controversy on Social Media. *Trans. Soc. Comput.* 1, 1, Article 3 (Jan. 2018), 27 pages. https://doi.org/10.1145/3140565

[30] Genevieve Gorrell, Mehmet Emin Bakir, Ian Roberts, Mark Anthony Greenwood, Benedetta Iavarone, and Kalina Bontcheva. 2019. Partisanship, Propaganda and Post-Truth Politics: Quantifying Impact in Online Debate. *The Journal of Web Science* 7 (2019).

[31] Saket Gurukar, Deepak Ajwani, Sourav Dutta, J. Lauri, S. Parthasarathy, and A. Sala. 2020. Towards Quantifying the Distance between Opinions. In *ICWSM*.

[32] K Sree Hari, Disha Aravind, Ashish Singh, and Bhaskarjyoti Das. 2021. Detecting Propaganda in Trending Twitter Topics in India—A Metric Driven Approach. In *Emerging Technologies in Data Mining and Information Security*. Springer, 657–671.

[33] IANS. 2019. Bhopal: Hundreds join BJP's pro CAA-NRC rally. *freepressjournal* (2019).

[34] IANS. 2019. Anti-CAA protests continue to hit rail, road traffic in West Bengal. *freepressjournal* (2019).

[35] IANS. 2019. Delhi's Shaheen Bagh rings in new year with anti-Citizenship Act slogans. *The Hindu* (2019).

[36] Maurice Jakesch, Kiran Garimella, Dean Eckles, and Mor Naaman. 2021. Trend Alert: A Cross-Platform Organization Manipulated Twitter Trends in the Indian General Election. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 379 (oct 2021), 19 pages. https://doi.org/10.1145/3479523

[37] Apalak Khatua and Aparup Khatua. 2016. Leave or Remain? Deciphering Brexit Deliberations on Twitter. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. 428–433. https://doi.org/10.1109/ICDMW.2016.0067

[38] Fannie Liu, Denae Ford, Chris Parnin, and Laura Dabbish. 2017. Selfies as Social Movements: Influences on Participation and Perceived Impact on Stereotypes. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 72, 21 pages. https://doi.org/10.1145/3134707

[39] Luca Luceri, Silvia Giordano, and Emilio Ferrara. 2020. Detecting Troll Behavior via Inverse Reinforcement Learning: A Case Study of Russian Trolls in the 2016 US Election. *Proceedings of the International AAAI Conference on Web and Social Media* 14, 1 (May 2020), 417–427. https://ojs.aaai.org/index.php/ICWSM/article/view/7311

[40] Sangeeta Mahapatra and Kiran Garimella. 2021. Digital Public Activism and the Redefinition of Citizenship: The movement against the citizenship (amendment) act of India. In *Weizenbaum Conference 2021: Democracy in Flux–Order, Dynamics and Voices in Digital Public Spheres*. DEU, 3.

[41] Leland McInnes and John Healy. 2017. Accelerated hierarchical density based clustering. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 33–42.

[42] Leland McInnes, John Healy, and James Melville. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426* (2018).

[43] Marcelo Mendoza, Maurizio Tesconi, and Stefano Cresci. 2020. Bots in social and interaction networks: detection and impact estimation. *ACM Transactions on Information Systems (TOIS)* 39, 1 (2020), 1–32.

[44] Tanushree Mitra, Scott Counts, and James W Pennebaker. 2016. Understanding anti-vaccination attitudes in social media. In *Tenth International AAAI Conference on Web and Social Media*.

[45] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. 31–41.

[46] Saif Mohammad and Peter Turney. 2010. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Association for Computational Linguistics, Los Angeles, CA, 26–34. https://aclanthology.org/W10-0204

[47] Sathappan Muthiah, Patrick Butler, Rupinder Paul Khandpur, Parang Saraf, Nathan Self, Alla Rozovskaya, Liang Zhao, Jose Cadena, Chang-Tien Lu, Anil Vullikanti, et al. 2016. Embers at 4 years: Experiences operating an open source indicators forecasting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 205–214.

[48] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, and Maurizio Tesconi. 2021. Coordinated Behavior on Social Media in 2019 UK General Election. In *ICWSM*.

[49] 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.

[50] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. 2021. Uncovering Coordinated Networks on Social Media: Methods and Case Studies. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 455–466.

[51] Anmol Panda, Ramaravind Kommiya Mothilal, Monojit Choudhury, Kalika Bali, and Joyojeet Pal. 2020. Topical Focus of Political Campaigns and its Impact: Findings from Politicians' Hashtag Use during the 2019 Indian Elections. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–14.

[52] Mainul Quraishi, Pavlos Fafalios, and Eelco Herder. 2018. Viewpoint discovery and understanding in social networks. In *Proceedings of the 10th ACM Conference on Web Science*. 47–56.

[53] Suhas Ranganath, Fred Morstatter, Xia Hu, Jiliang Tang, Suhang Wang, and Huan Liu. 2016. Predicting Online Protest Participation of Social Media Users. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (Phoenix, Arizona) *(AAAI'16)*. AAAI Press, 208–214.

[54] Ammar Rashed, Mucahid Kutlu, Kareem Darwish, Tamer Elsayed, and Cansın Bayrak. 2021. Embeddings-Based Clustering for Target Specific Stances: The Case of a Polarized Turkey. *Proceedings of the International AAAI Conference on Web and Social Media* 15, 1 (May 2021), 537–548. https://ojs.aaai.org/index.php/ICWSM/article/view/18082

[55] Ammar Rashed, Mucahid Kutlu, Kareem Darwish, Tamer Elsayed, and Cansın Bayrak. 2021. Embeddings-Based Clustering for Target Specific Stances: The Case of a Polarized Turkey. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 537–548.

[56] Vincent Raynauld, Emmanuelle Richez, and Katie Boudreau Morris. 2018. Canada is #IdleNoMore: exploring dynamics of Indigenous political and civic protest in the Twitterverse. *Information, Communication & Society* 21, 4 (2018), 626–642. https://doi.org/10.1080/1369118X.2017.1301522 arXiv:https://doi.org/10.1080/1369118X.2017.1301522

[57] B. Riedel, Isabelle Augenstein, Georgios P. Spithourakis, and S. Riedel. 2017. A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. *ArXiv* abs/1707.03264 (2017).

[58] Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2019. Calls to Action on Social Media: Potential for Censorship and Social Impact. *EMNLP-IJCNLP 2019* (2019), 36.

[59] Express News Service. 2019. Government open to suggestions to implement Citizenship Act, says MHA amid raging protests. *newindianexpress* (2019).

[60] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. *Nature communications* 9, 1 (2018), 1–9.

[61] Karishma Sharma, Yizhou Zhang, Emilio Ferrara, and Yan Liu. 2021. Identifying coordinated accounts on social media through hidden influence and group behaviours. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1441–1451.

[62] Ivan Slobozhan, Tymofii Brik, and Rajesh Sharma. 2021. Longitudinal change in language behaviour during protests: A case study of Euromaidan in Ukraine. *arXiv preprint arXiv:2109.11623* (2021).

[63] Jitendra Soni. 2019. CAA protest row: Section 144 imposed in UP Mau's Hajipura Chowk area. *freepressjournal* (2019).

[64] Kate Starbird and Leysia Palen. 2012. (How) Will the Revolution Be Retweeted? Information Diffusion and the 2011 Egyptian Uprising. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 7–16. https://doi.org/10.1145/2145204.2145212

[65] Abraham Ronel Martínez Teutle. 2010. Twitter: Network properties analysis. In *2010 20th International Conference on Electronics Communications and Computers (CONIELECOMP)*. IEEE, 180–186.

[66] Zeynep Tufekci. 2014. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Eighth international AAAI conference on weblogs and social media*.

[67] Joshua Uyheng and Kathleen M Carley. 2021. Computational Analysis of Bot Activity in the Asia-Pacific: A Comparative Study of Four National Elections. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 727–738.

[68] Michel Verleysen et al. 2003. Learning high-dimensional data. *Nato Science Series Sub Series III Computer And Systems Sciences* 186 (2003), 141–162.

[69] Rong Wang and Alvin Zhou. 2021. Hashtag activism and connective action: A case study of# HongKongPoliceBrutality. *Telematics and Informatics* 61 (2021), 101600.

[70] Human Right Watch. 2019. India: Citizenship Bill Discriminates Against Muslims. *Human Right Watch* (Dec. 2019).

[71] FPJ Web. 2019. CAA protests: Why are students protesting? Here's all you need to know. (2019).

[72] Kai Wei, Yu-Ru Lin, and Muheng Yan. 2020. Examining Protest as An Intervention to Reduce Online Prejudice: A Case Study of Prejudice Against Immigrants. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 2443–2454. https://doi.org/10.1145/3366423.3380307

[73] Kai-Cheng Yang, Onur Varol, Pik-Mai Hui, and Filippo Menczer. 2020. Scalable and Generalizable Social Bot Detection through Data Selection. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 01 (Apr. 2020), 1096–1103. https://doi.org/10.1609/aaai.v34i01.5460