Leveraging AI To Understand Protests and Foster Secure Societies during Protests

By

Kumari Neha

Under the supervision of

Dr. Arun Balaji Buduru, IIIT Delhi
Dr. Ponnurangam Kumaraguru, IIIT Hyderabad

Indraprastha Institute of Information Technology Delhi

October 08, 2022

Leveraging AI To Understand Protests and Foster Secure Societies during Protests

By

Kumari Neha

Submitted

Comprehensive report for the degree of Doctor of Philosophy

to

Indraprastha Institute of Information Technology Delhi

October 08, 2022

Certificate

This is to certify that the comprehensive report titled **"Leveraging AI To Understand Protests and Foster Secure Societies during Protests"** being submitted by **Kumari Neha** to Indraprastha Institute of Information Technology Delhi, for the award of the Ph.D. degree, is an original research work carried out by her under my supervision. In my opinion, the comprehensive report has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this report have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

October 2022

Dr. Arun Balaji Buduru
Department of Computer Science & Engineering
Indraprastha Institute of Information Technology Delhi
India

October 2022

Dr. Ponnurangam Kumaraguru
International Institute of Information Technology Hyderabad
India

Acknowledgements

I want to thank my thesis advisors Dr. Arun Balaji Buduru, IIIT Delhi, and Dr. Ponnurangam Kumaraguru, IIIT Hyderabad, for their constant support and supervision. I am immensely grateful to them for providing me with their valuable guidance and insights. I also want to thank my colleagues at the Precog Lab, IIIT Delhi, who volunteered for any support or help I needed for this research project. Finally, I want to express my gratitude to my family for providing me with unfailing support and continuous encouragement throughout the journey.

October 2022

Kumari Neha
Computer Science and Engineering
Indraprastha Institute of Information Technology Delhi
New Delhi, India

Abstract

In recent years, we have encountered a mutual boost in mobile communication capabilities and the diffusion of Online Social Media (OSM). As an outcome, a new socio-technical convergence has been established that features a resilient network of humans who generate a continuous flow of information across online and offline environments. The information loop between the online and offline environments tends to provide a feedback effect, where the offline ecosystem may affect the online ecosystem and vice versa. The impact of the feedback ecosystem intensifies during times of protest (or movement), as information flow around the protest might affect a person's judgment followed by their action. On the bright side, the socio-technical convergence enables AI-powered applications to use social media to reach a critical mass during the protest, demystify people's opinions and address the concerns of the protest. Hence, our first research objective is to address how social media enables the advancement of a social movement's goal and to demystify opinions shared during the protest. On the dark side, the socio-technical convergence unveils unparalleled opportunities to manipulate and deceive users on social media leading to the manipulation of public opinions, the polarization of society, and violent protests in the offline ecosystem. The second objective of our research is to weed out the possible threats present during the protest to foster a secure online ecosystem and dilute violent on-ground activities.

Contents

List of Figures

List of Tables

Acronyms

OSM Online Social Media

ML Machine Learning

DL Deep Learning

SNA Social Network Analysis

SSR Shushant Singh Rajput

CAA Citizenship Amendment Act

AI Artificial Intelligence

FP Farmer Protest

KTB Kill the Bill protest

Chapter 1

Course Work

1. $BIO571$, Network Science, Registration Type: Open Elective, (Final Grade: B)
2. $CSE563$, Multimedia Computing and Applications, Registration Type: Department Elective, (Final Grade: A-)
3. $CSE543$, Machine Learning, Registration Type: Department Elective, (Final Grade: B-)
4. $CSE556$, Natural Language Processing, Registration Type: Department Elective, (Final Grade: B-)
5. $ENG599s$, Research Methods, Registration Type: Open Elective, (Final Grade: A)
6. $CSE799/PTH799$, PhD Thesis, Registration Type - Thesis, (92 credits, CGPA: 8.00)

Chapter 2

Introduction

The recent technological advancement has transformed the Online Social Media (OSM) platforms into a significant place for debate around socio-political phenomena, expressing opinions, and mediating social interactions [Pond and Lewis, 2019]. The abundance of communication capabilities has produced a resilient public network responsible for a continuous flow of information between the offline and the online ecosystem [Conti et al., 2012]. With the help of communication capabilities, social media helps people identify like-minded people who boost their belief system [Garimella et al., 2018b]. While identifying with a group of people on social media gives a sense of belongingness and helps fight for a cause [Bittner et al., 2020], it sometimes leads to a polarized information flow between users who are ignorant of the other side [Horawalavithana et al., 2021]. The tendency of users to adjust interests, opinions, and actions according to the recent observations introduces a feedback effect, where the offline and the online ecosystem might affect each other [Ramakrishnan et al., 2014]. The public posts on social media provide valuable information about the ongoing events in real-time [Muthiah et al., 2016b, Goode et al., 2015], and the anger-fueled discussion on social media can also give rise to an agitated society and mark the beginning of social movements or uprisings [Poltrock et al., 2012, De Choudhury et al., 2016a].

Protests and social movements are scarce; however, they may lead to dramatic outcomes when they occur. Social media, such as Twitter, has become a central point

for organizing and developing collective action, such as online protests worldwide. The manifestation of collective identity (for example, #wearethe99percent launched by the Occupy Wall Street movement) is accompanied by a set of goals that provides users with a collective sense of self and what they stand for [Gerbaudo and Treré, 2015]. The human feedback loop of socio-technical convergence has helped throw light on significant societal issues, including environmental change [Weart, 2015], breaking gender stereotypes [Bittner et al., 2020], and voicing marginalized social groups [Liu et al., 2017a] among others. We can extract actionable knowledge about diverse aspects of the current ongoing phenomena. In particular, the social-technical convergence has paved the way for social sensing, where humans act as data sensors that continuously post about the ongoing phenomena [Wang et al., 2014]. The collected information from the human sensors can provide data-driven decision support to policy-makers and stakeholders for making an informed decision and adjusting any interventions according to the needs of the people [Alonso et al., 2018, Alonso et al., 2018].

Due to the world-scale capabilities of social media for enabling communication between users and distribution and aggregation of information, OSM was initially considered a great opportunity to promote a diversified point of view, positively impacting individual critical thinking and democratic discussions. However, as social media became the main outlet for information dissemination and consumption, the socio-technical convergence started posing severe threats to society. As humans are prone to "confirmation bias", which induces them to consume information that confirms their pre-existing beliefs, the benefit of being exposed to different point-of-view is highly limited [Yardi and Boyd, 2010]. On the other hand, the news feed algorithms and social network dynamics also lead to reinforcement of selective exposure mechanism [Cinelli et al., 2020]. As a result, the democratic discussion on a given topic has formed the so-called echo chambers, where users tend to mutually reinforce their opinion and biases on a topic [Dash et al., 2021]. Such an ecosystem becomes a perfect breeding ground for malicious activities ranging from promoting terrorist activities [Badawy and Ferrara, 2018], disruption of foreign campaigns [Badawy et al., 2019], and inducing fear among fragile audience [Akhtar et al., 2021]. Hence, the

threat to the secure society can range from genuine users involved in occasional harassing fragile people [Fast and Horvitz, 2016] to more profound inauthentic actors who purposefully become part of an online discussion with the ill-intention to create polarization [Gorrell et al., 2019a], spread propaganda [Sree Hari et al., 2021], among other intentions. Despite the efforts of the platform to remove malicious content, the posts made by the accounts may reach a wider audience before the malicious content or account is suspended from the platform [Santini et al., 2021]. The malicious accounts also innovate themselves to deceive the platform's regulations. Hence, the malicious accounts often hide under the stream of benign OSM content and become viral before soliciting any intervention. Due to the feedback loop formed between the online and offline ecosystems, the malicious content spread in the online world might affect the offline ecosystem gravely. For example, the recent debate on vaccines on social media has not only led to an infodemic on social media, but it has also led to the slowing of the process of vaccinations [Germani and Biller-Andorno, 2021].

In summary, from a secure society perspective, there is a need to understand social movements mediated by social media and counter threats that pollute the online and offline ecosystem and might unfold grave consequences. For the above two objectives, a research endeavor must design approaches and apply suitable techniques for the challenges.

2.1 Understanding Protest Strategy and Objectives

In the past decade, the most effective approach to understanding social movements was grounded in the assumptions that shared grievances and potential means of reducing them are essential preconditions for the emergence of collective actions. However, recently the strong hypothesis about the centrality of deprivation and grievances has been pivoted to a weaker one. The current assumption is that any society always has sufficient discontent to provide the breeding ground for a movement, given that the campaign is organized efficiently [Mccarthy and Zald, 1977]. From the Arab spring witnessed during the start of the decade, sustainable protests

are distinguishing between two logics interplay: the formal association of organizational resources, i.e., the logic of collective action, and users' interest in sharing personalized content on social media, i.e., the logic of connective action [Bennett and Segerberg, 2012b]. Social media has become a prime site where protests are created, channeled, and contested [Gerbaudo and Treré, 2015]. According to the global protest tracker from Carnegie Endowment for International Peace [1], since 2017, over 230 anti-government protests have erupted worldwide, in more than 110 countries. Over 25 significant protests have been directly related to the coronavirus pandemic. Since social media has become one of the center places for organizations and sharing protest-related posts, social media can help in many ways. For example, the study of protesters' posts on social media on the 'no ball, no wall' protest was done to reduce the prejudice towards a given section of society [Wei et al., 2020a]. In another instance, the study of social media posts was used to understand the dogmatic mindset of the users of a marginalized community [Fast and Horvitz, 2016]. The new direction of social movement research has attracted a lot of attention in two directions: the movement-media relationship and social movement strategy.

2.1.1 Challenges

The understanding of the major objective and strategy adopted for a sustainable socio-technical protest requires (i) extracting actionable and concise knowledge from the online ecosystem; (ii) identifying and characterizing prime advocates involved in the online social movement; and (iii) designing suitable techniques to demystify online strategies used by activist for sustaining the movement online. At the same time, there is a range of work for protests that consider the western context, the work done on social movements in non-western countries are scarce [Gerbaudo and Treré, 2015]. Gathering user-generated content from OSM comes with its fair share of challenges. It included incompleteness, information overload, and multidimensional information (text, images, videos). One of the major challenges concerning study protests in non-western contexts remains the barriers by content shared in low-resource languages [Haider et al., 2020]. The debate on movement are both single-

[1] https://carnegieendowment.org/publications/interactive/protest-tracker

sided [Wang and Zhou, 2021a], as well as rich is discourse [Gallagher et al., 2018a]. Since the protests are unique and subjective, posing another major challenge in studying online social media protests. When understanding the strategies of the protest, the major challenges are understanding the most influential users (activists) and how to sustain the movement online [Wang and Zhou, 2021a].

2.2 Understanding online and offline threats

Posts on OSM are prone to subjectivity, informality, propaganda, and disinformation [Gorrell et al., 2019b]. The unreliable content shared by various inauthentic users gets attention from unaware users, who fall prey to propaganda or deception [Stella et al., 2018]. Apart from propaganda and disinformation, the content on social media also constitutes hate and fear speech that might affect users and debate on social media [Saha et al., 2021]. During a social media protest that unfolds into discourse, there is a risk of inauthentic or disrespectful content on either side of the discourse [Gallagher et al., 2018a]. To fully understand the interplay of inauthentic activity within the discourse, we need to focus on both sides of the discourse. We focus on the threats against inauthentic users and their content on one end. On the other side of a protest, we focus on the hateful content during a social media protest.

2.2.1 Challenges

The major challenge with these online threats is the barrier of low-resource language posts made during the protest. While protests are very prominent phenomena in different parts of the world, due to a shortage of different language representations, the protests of non-western countries remain understudied. Recently, we have seen a rise in the study of protests in Brazil [Costa et al., 2015] and various election campaigns in Asia-pacific [Uyheng and Carley, 2021a]. The study of protests is yet a long way to go [Wang and Zhou, 2021a]. Another major challenge with the study of protests is the awareness of the different political and moral values of the country or community under study [Haidt, 2011, Rezapour et al., 2019]. As delineated in

the previous Section, a common issue while dealing with OSM data is the need to automate the extraction of information from a large variety of inputs (such as texts, images, videos, etc.). This involves complex machine tasks, including Natural Language Processing (NLP), speech recognition, and computer vision which are naturally associated with human intelligence.

2.3 Solutions

Social Network Analysis enables modeling of information flow between the users in the OSNs, identifies the most relevant actors, and helps understand people's perceptions when combined with various AI techniques [Liu et al., 2018]. Hence, extracting relevant topics from large OSM discussions requires a combination of AI and SNA. We can classify a huge collection of data, understand hidden patterns of information in the data, and use the network of users and content to understand the user's perception and beliefs of the topic of discussion. While using AI and SNA to understand protest-related activities, we can understand the emotional take on the protest [Costa et al., 2015], the stance of the user on a particular debate [Gallagher et al., 2018a], understand their discontent with a political change [Wang and Zhou, 2021a], among other knowledgeable insights. One of our primary goals is to enhance the various techniques used to understand the protests in low-resource countries. For the next goal, i.e., to identify the online and offline threats in conjugation with the above methods, we need expertise in various political, psychological, and management-related domains. We also need to keep track of emergent new topics and keep pace with threats in the online and offline world. Collecting user-generated content in OSM and extracting actionable knowledge for decision support can provide enormous advantages for understanding our societies. However, those actions raise serious concerns about their probable adverse effect on other vital public goods, such as personal privacy or the freedom of speech. The next section elaborates on such legal and ethical considerations.

2.4 Legal and ethical concerns

Although users' profile data in OSM are 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. All the data collected in our study is from publicly available information, and no attempt to explore any user-level demographic information has been made. The opinion shared by the users on the campaign is broadly studied to understand public perception of various protests and not on an individual level to maintain users' privacy. While sharing tweet IDs is a common practice in such studies, there is a risk to sharing the Tweet IDs due to the sensitive nature of the campaign. For example, if we share the tweet IDs, we risk obtaining all the user-level information from the tweet ID. Hence, sharing the tweet IDs used in our various studies were not undertaken. Instead, tweet and user-level features without revealing personal information such as profile name, profile description, username, etc., are shared.

2.5 Targets and contributions

We first examine how to enhance AI techniques enabling essential applications. Secondly, we develop complex approaches targeting specific online and offline threats.

2.5.1 Understanding Protest narratives

As outlined in Section 2.1, OSM provides a ground for understanding the participants' protest narratives and prime opinions. Hence, adopting methods for automatic understanding of protest-related narratives and underlining topics becomes crucial for a secure society. Understanding activists and the content shared in the OSM not only helps users make an informed decision on their stance but the information available in OSM also helps policymakers and stakeholders make decisions that support the need of common people. Since each protest are unique, and no fixed set of labels can be applied to all protests under study, we propose an unsupervised framework for understanding major protest narratives.

Contributions. We contribute to the understanding of narratives in various non-western protests. In particular, we use various deep learning and unsupervised techniques to identify strategies and narratives in various recent protests.

#ShushantSinghRajput: Strategies by Counterpublics. Twitter has emerged as a prominent social media platform for activism and counterpublic narratives. Counterpublics [Jackson and Banaszczyk, 2016] are defined as marginalized communities that distribute messages to diverse social groups, raise awareness, and challenge dominant narratives. The counterpublics leverage hashtags to build a diverse support network and share content on a global platform that counters the dominant narrative. Our first work applies the framework of connective action to the counter-narrative campaign over the cause of death of #SushantSinghRajput. We combine descriptive network, modularity, and hashtag-based topical analysis to identify the campaign's three major mechanisms: generative role-taking, hashtag-based narratives, and forming an alignment network toward a common cause. Using the case study of #SushantSinghRajput, we highlight how the connective action framework can be used to identify different strategies adopted by counterpublics for the emergence of connective action.

Detection of Objectives and Narratives across Protests. Mass mobilization and protests are uncommon, but they could have unexpectedly dramatic results when they do happen. Twitter and other social media platforms have emerged as hubs for the planning and development of online protests all across the world. Interpreting the many narratives shared during an online protest is essential to grasp people's perspectives. In our next work, we introduce a methodology based on unsupervised clustering for comprehending the narratives present in a given online protest. We offer insights into the narratives expressed during an online protest through a comparative study of tweet clusters from three protests against laws affecting government policy. In all three of the protests under consideration, we discovered narrative clusters containing both reports of on-the-ground activity and calls

for people to participate. We also discovered protest-centric narratives in several protests, such as cynicism regarding the subject. The outcomes of our investigation can be used to comprehend and contrast how individuals view potential mass mobilizations in the future.

2.5.2 Understanding Online threat during protests

The threats in the online ecosystem may range from bots, and semi-automated accounts, to extremist accounts, which might get suspended or deleted later by the platform. We investigate how various inauthentic actors participate in the protest and what threats they impose on society.

#CitizenshipAmendmentAct. 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 the presence of inauthentic users on both sides of the discourse, with counter-protesters having more inauthentic users than protesters. The following network of users suggests homophily among users on the same side of discourse and a connection between various inauthentic and authentic users. This work contributes to filling the gap in understanding the role of users (from both sides) in a less studied geo-location, India.

2.6 Ph.d. Thesis outline

The research plan can be summarized in Figure 2.1. We focus on the 3 different objectives in detail in our work, understanding protest objectives, tackling the online threats, and tackling offline threats in the research objective.

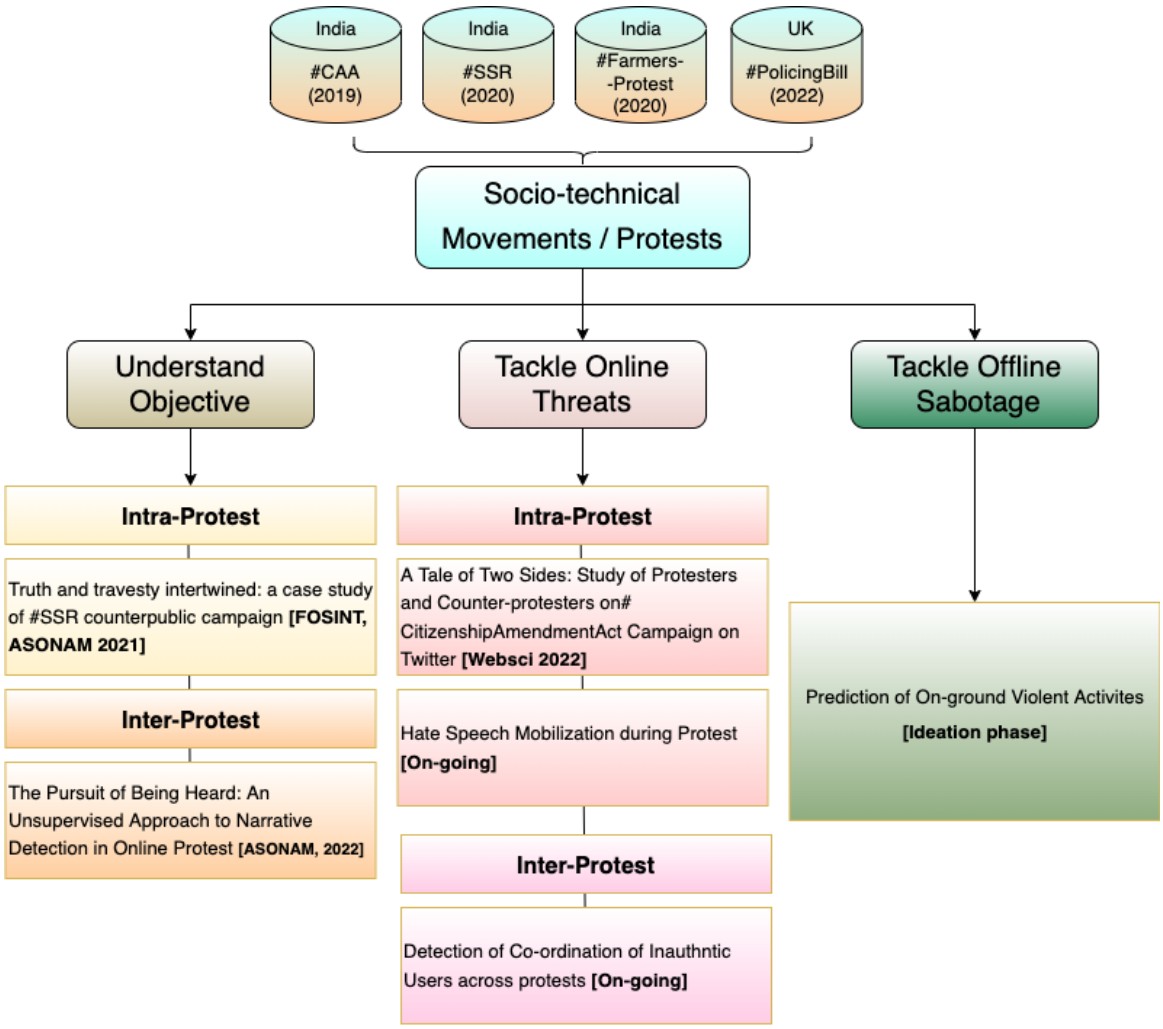

Figure 2.1: A depiction of Ph.D. Thesis outline and vision.

2.6.1 System Requirements

For our experiments, we used a Linux-based system with Xeon(R), an $x86$ microprocessor developed by Intel with a system memory of $62GB$. We ran our machine learning models using NVIDIA-SMI GPU with a driver version of 440.33.01 and installed Cuda version 10.2. Another server used for training our deep learning models was Nvidia RTX 3090 GPU system with an installed Cuda version of 11.3.

Chapter 3

Literature Review

From the start of the decade, the use of Twitter for mass mobilization has been very evident [González-Bailón et al., 2011]. During the onset of a social media-mediated movement, the content shared is usually distinguished between more influential generators and other users who become part of the cascade network [González-Bailón et al., 2011]. The study of protests has been broadly divided into two categories (i) when will the protest take place [Muthiah et al., 2016b, Goode et al., 2015], and (ii) whether and how the social media contributes to the explosion during the protest. What makes the studies on various protests challenging are the inherent nature and cause of protests. Over the last decade, several political protests have erupted in different parts of the world [Lotan et al., 2011, Xiong et al., 2019, Wang and Zhou, 2021b, González-Bailón et al., 2011]. What makes every study interesting is that all these protests show some similarity concerning the use of the same platform for conducting online protest; however, the narratives and expected outcomes remain subjective from one protest to another.

3.1 Understanding Protest narratives

Financial Crisis protest of Spain, 2011 The mobilization in Spain in 2011 emerged due to the political response to the financial crisis, which resulted in the demand for new forms of democratic representation. The main target of the protest

was to organize a mass mobilization on May 15, bringing people out on the streets in 59 cities. From May 15 to May 22, the participants camped in the city square, which was the date of the regional and national elections. Slowly after this, the movement lost its strength and faded.

In the work [González-Bailón et al., 2011], researchers study the dynamics of protest recruitment during the protest to identify whether the dynamics of mass mobilization depend on broad-casting links (i.e., weak links) or stronger connections. The study spanned a 30-day duration on the tweets done by 87,569 users and their 581,750 protest-related tweets. The authors formed a network of followers and retweets for the most active users during the protest and used threshold-based metrics to identify recruitment patterns in the protest. Their result shows that it is important to have multiple exposures rather than repeated exposure from the same individual to form a social contagion for a user to join the protest.

Egyptian uprising, 2011 In 2011, a string of political uprisings was witnessed around the Arab world. This also led to an uprising in Egypt following Tunisia's successful demonstrations. The protest was conducted to overthrow the authoritarian regime in Egypt. The uprising in Egypt started on January 25, 2011, and continued for 18 days until Egyptian president Mubarak resigned on February 11, 2011. The protest was seen as a peaceful demonstration at the start; however, on 2nd of February marked a significant shift to violent protest due to clashes between the pro-Mubarak and anti-Mubarak groups being formed. The pro-Mubarak users acted like 'Thugs' and attacked the Anti-Mubarak activists.

In the work [Starbird and Palen, 2012], the authors studied the interplay of the users involved in online activism and the users present on the ground during the protest. The study on the topic involved the study of tweets and the active users during the protest. The data collection included hashtag-based collection and user information from Twitter API. The authors studied the diffusion of the most popular tweets during the protest. The protest tweets included the bar-chart structure and on-ground activity of the users who were present at the location of the protest. The retweet was a prominent feature for the propagation of the tweet during the protest. The study of tweets also showed coordination between tweets, as various variations

of the "Uninstalling dictator" with progress bar tweet appear 19,836 times in the dataset.

Brazil Summer Protest, 2011 The protest in Brazil in the summer of 2011 was initially disrupted due to the rise in public transport fares. However, as the protest moved forward, the protest included corruption in politics and police brutality against the people conducting the protest.

The authors in the work [Costa et al., 2015] analyzed tweets shared during the protests in Brazil to find the emotional dynamics of the posts. They found that the peak in the tweets coincided with days with substantial online activity. They also found that the protest's tweets showed negative and positive emotions. To identify tweets that were protesting relevant, the authors used an SVM classifier on the initially collected tweets. They trained a multi-nominal naive Bayes classifier with 9003 tweets manually annotated as positive, neutral, and negative emotions.

Gezi Park protest, 2013 The Gezi park protest began quietly in Turkey, which was already politically divided at the time. On May 28, 2013, about 0–100 environmental activists gathered for a sit-in at Gezi Park in Taksim Square, Istanbul. They were there to demonstrate against the destruction of one of the last public green spaces in central Istanbul. The government had planned to make the construction of malls and luxurious residences in the park. The protesters were attacked by police with tear gas, and water cannons, triggering clashes between authorities and the demonstrators that lasted until the end of the park occupation on June 15.

In their work [Varol et al., 2014a], the authors focused on the extraction of topics of conversation about the social uprising and identified the trending topics. The authors also studied the Spatio-temporal characteristics of the conversation, including were tweets about protests started and what locations shared the most identical trends and topics. The authors also reported that the online content shared was highly affected by the on-ground activities.

Brexit Refendrum, UK Brexit (or the UK EU membership referendum) was done on 23rd June 2016 in the UK and Gibraltar. The main goal was to gauge support for whether to remain a member of or leave the EU by the countrymen. In October 2015, a cross-party, formal group campaigning for Britain to Remain a

member called Britain Stronger in Europe. Two groups promoting exit sought to be the official Leave campaign: *Leave*. Most of the UKIP party supported the EU, led by Nigel Farage, and Conservative Party, Eurosceptics supported Vote Leave. On April 13, 2016, the Electoral Commission announced that Vote Leave was the official leave campaign. The UK government's official position was to support the remaining option. The referendum turnout was 71.8%, with more than 30 million people voting. Leave won with 51.9%, while Remain got 48.1% of the votes.

The work done by authors in [Grčar et al., 2017] addresses two main questions. The first is the mood of the users on the Brexit referendum and who are the most influential users in the pro- and anti- stances. The authors collected geo-tagged tweets related to the Referendum, and the results of their opinion mining from the Twitter data matched well with the opinion polls on the topic. This becomes a very important result, as it sheds light on the importance of sharing on social media, such as Twitter can be equated to what people's views are on a given opinion piece. The authors in the work [Howard and Kollanyi, 2016] show that the two most important accounts in the referendum were indeed bots, i.e., @$iVoteLeave$, @$ivotestay$. The purpose of the bots was to amplify the source simply by aggregating the content and then retweeting it. The authors in [Grčar et al., 2017] collect 4.5 million tweets from almost 1 million users about Brexit from May 12, 2016, to June 24, 2016. 35,000 tweets were randomly selected for manual annotation. The study uses a score metric that considers users' leave, remain, and neutral tweet counts to judge the user's stance on the topic. The analysis of users who joined leave vs. remain discourse shows that leave users gradually increased compared to remaining users who were persistently present and contributing to the debate. As for the Influence, the authors use retweets and the number of posts a user created to measure influence. The users were ranked for influence using the Hirsch index (h-index) metric. The metric is taken from the author-level bibliometric indicator that quantifies the scientific output of a scholar by a number. Given a scholar with an index of h, he has published h papers, with each one having been cited in other papers at least h-times. In the case of a Twitter adaptation, the authors provide a Twitter user with an index of h if he has posted h tweets, each of which has been

retweeted at least h times. The Leave group is found to be considerably more active in generation as well as retweeting of content, while the Remain side was found to be less active.

Venezuela political crisis, 2019 In [Horawalavithana et al., 2021], the authors used Venezuela's political crisis in early 2019 as a case study to gauge how the external and internal factors drive the related activities on social media. In Venezuela, the past decade has witnessed a sociopolitical fragmentation due to differences in interests, identities, and politics. There are two ideologies in Venezuela, i.e., Chavism, embraced by supporters of the political ideology of the late president Hugo Chavez, and Anti-Chavism, embraced by people who strongly oppose Chavez's legacy. Chavism, however, still controls the Venezuelan political system with Nicolas Maduro as the state's head. The re-election of Nicolas Maduro as the country's president on January 10, 2019, led to the beginning of a presidential crisis driven by claims of illegitimacy and reports of coercion and fraud. The crisis continued for a while and slowly faded after March 25 when the Russian aircraft were seen arriving at the Caracas airport guarded by the Venezuelan military. The work done by the authors [Horawalavithana et al., 2021] focuses on the content being shared on social media during the crisis as a response to external and internal factors. The external data for the analysis was taken from ACLED (Armed Conflict Location and Event Dataset) [Raleigh et al., 2010] and GDELT database [Leetaru and Schrodt, 2013]. The authors first divided the users' tweets into anti-Maduro and pro-Maduro tweets. The internal drivers were politicians, media outlets, and normal users. The 200 most influential users were identified from both pro- and anti- tweets to identify the influence. On performing clustering, the authors found that the clustering coefficient for the anti-Maduro community decreases if media accounts are removed. This gives evidence of the media's involvement in the anti-Maduro campaign. For the pro-Maduro community, the clustering community decreases if political accounts are removed. The clustering coefficient did not change much if the random users were removed from the discussion. To gauge external drivers of the crisis campaign, the authors calculate the correlation between the volume of anti-Maduro and pro-Maduro daily Twitter activities and the volume of offline events reported in the

ACLED and GDELT databases. The anti-Maduro community related more with ACLED, suggesting that online discussions from the anti-side tend to align well with reports about protests and violent clashes as documented by ACLED.

Day Without Immigrants & No ball, no wall protest, 2020 The "Day Without Immigrants" and the "NoBan, No Wall" protests were the most recent nationwide protests in the US that aimed to show the important contributions of immigration and to resist punitive immigration policies. The "Day Without Immigrants" was held on February 16, 2017, in response to Donald Trump's plans to build a border wall, deport potentially millions of undocumented immigrants, and strip sanctuary cities of federal funding. The main aim of the protest was to show the importance of immigrants in the US economy. The "No Ban, No Wall" protest took place on January 28, 2017, in response to President Donald Trump's plan to ban citizens of certain Muslim countries from entering the US and suspend the admission of all refugees. Both protests used social media to disseminate information and aided the online protests that were going on at the time. The authors in the work [Wei et al., 2020b] performed a control focus group-based study to identify and reduce online prejudice towards a given part of the community. The work focuses on identifying a focal event that impacts people's behavior. Prejudices are a very mild form of hate or predefined mindset that a person has towards another community or people. The authors used the two protests as an intervention to reduce online prejudice. The results show positive and negative changes in people's prejudice after the protest. The authors also identified features of users who are more likely to change (or resist) their mindset after a protest. The findings of the work can be used to design targeted interventions during a protest-like situation.

3.2 Understanding Online Threats

With the rise in use of social media use for conducting protest activity, social media started becoming the target of various radicalization groups [Spiro and Ahn, 2016], inauthentic actors [Luceri et al., 2019] who started to use social media for nefarious reasons.

3.2.1 Accounts identified as ISIS groups

The authors in the work [Spiro and Ahn, 2016] used the pre-identified 25,538 ISIS accounts. They conducted a forecasting task to identify extremist users, estimating whether regular users will adopt their content and whether users will reciprocate contacts created by the extremists. The authors detected the extremist users with 93% AUC, while adoption of extremist content was forecaster with 80% AUC. The users were predicted to reciprocate interaction with extremist users with 72% AUC. The datasets the authors collected included 3,395,901 tweets by ISIS group accounts, 9,193,267 tweets generated by users exposed to the ISIS content from the ISIS account followers data, which was taken for 25,538 random users from the set of followers. The authors curated several feature sets for their prediction purpose and implemented several machine learning models for the classification task. The models included Logistic Regression with LASSO regularization and Random Forest with k-fold cross validation with the value of k set as 5. The authors used the greedy method to select the best set of features for conducting their prediction problem. The exposure to the content of the ISIS account is determined by the Retweet mechanism in Twitter while reciprocating the user's reply to the tweet as an alibi. As for the static prediction task, the model doesn't take advantage of the timeline of the activity sequence, while a dynamic model looks into the time while making the prediction. The Random Forest takes advantage of the temporal data dependency for real-time prediction. The work done by the authors in [Spiro and Ahn, 2016] is one of the few early works that shed light on the beginning of a new era of social media, where extremists groups and content manipulators started to co-exist in the digital ecosystem along with the other naive users.

3.2.2 Russian Trolls on Twitter

By 2016, researchers warned about trolls and other forms of online manipulations. The elections in a country are the breeding ground for manipulation. Bots have been introduced into the social media world. However, the researchers defined the trolls used in the 2016 US elections as semi-automated accounts with humans in

their blackened [Badawy et al.,]. The authors could accurately identify the Russian trolls with AUC 96% using 10-fold cross-validation. The most important features for their classification task were bot-like activity, account-level features, and political ideology. The authors collected 43 million tweets from 5.7 million users between September 16, 2016, and November 9, 2016. The dataset also contained 221 Russian trolls-produced tweets. The best algorithm for their case came as the Gradient boosting algorithm, whereas, in features, political ideology came as the most important feature in the task. The work analyzes how the users on social media are susceptible to the content they are exposed to and how easily target people can be made.

3.2.3 Bots

While there are accounts that are purposefully created for deceiving humans on social media, the automated accounts have drawn a lot of traction on social media [Uyheng and Carley, 2021a, Chang et al., 2021a]. The bots try to create content that may be polarized [Luceri et al., 2019], talking highly of one side or even helping spread propaganda on social media [Howard and Kollanyi, 2016]. The involvement of bots has led to discourse and tension in the online world, which are very much prevalent in Elections [Shevtsov et al., 2022]. However, the bots have most recently invaded any discussion space on the social media platform [Ferrara,]. The threat of automated and semi-automated accounts has been rising in social media and needs to be tackled for a safer society.

3.2.4 Hateful Users

Apart from trolls and bots, polarized and hateful users pose another threat to secure society. They tend to pollute online discussions irrespective of their knowledge of wrongdoing. Hate speech is *any content that promotes violence against the opposing stance cohort, directly or indirectly threatens the people based on their race, ethnicity, national origin, religious affiliation, political ideology, and political affiliation.* [Schmidt and Wiegand, 2019]. Few studies have been on hate speech detection of low-resource languages [Mathur et al., 2019]. The early work on hate and offensive

tweet detection in code-mixed language argues that the translation of code-mixed or low-resource language might alter the meaning and context of hate speech [Mathur et al., 2019].

3.2.5 Co-ordinated Campaigns

The manipulation of social media users has two important characteristics. The first one is the use of propaganda and the second one is the coordination of the inauthentic users to provide for the widespread reach of the propaganda. The work done by authors [Hristakieva et al., 2022] shows the interplay between the spread of propaganda and coordinated activities carried on the spread of propaganda which helps provides a better insight into the malicious behavior leading to a better understanding of coordinated inauthentic behavior. The authors collected 11,264,820 tweets about the 2019 UK general election, published by 1,179,659 users between 12 2019 and December 12, 2019 (coincides with Election day). For identifying coordination, the authors used network-based approach, with the extraordinary similarities between user's post as a proxy for coordinating communities. The analysis starts with selecting of top 1% users called the superspreaders. For each super-spreader, the TF-IDF vector of the tweet ids they have retweeted was created. The similarity between all the users was conducted using the cosine similarity between their corresponding vectors, thus obtaining a weighted undirected user-similarity metric. The network was filtered by calculating its multi-scale backbone, which allows for the statistically significant network structure to be kept. After we had a filtered network, the Louvain community detection algorithm was applied to group users into network communities. Finally, network dismantling was applied, which assigns a coordination score to each user in the network by iteratively removing network edges and nodes based on a moving edge weight threshold. For propaganda detection, the authors used Proppy, which performs best in the detection of propaganda. The results show that different parties can be identified using a coordination mechanism. While propaganda level for the different political communities varies.

Chapter 4

Understanding Counterpublic Campaign

Social media platforms are used as a primary source of information and opinion sharing in recent times [Liu et al., 2017b,ElSherief et al., 2017,Field et al., 2019,Starbird and Palen, 2012, Contractor et al., 2015]. A Twitter user involved in activism activities such as organizing online petitions and building a counter-public campaign narrative through hashtags is defined as a Twitter activist [Wang and Chu, 2019]. Often heated debate on controversial topics leads to users divided into protesters and counter-protesters on social media [Gallagher et al., 2018b,Khatua and Khatua, 2016,Mitra et al., 2016]. As the online movement involve multiple users and their interactions, the different studies have focused on understanding social media protests concerning different heterogeneous user data, including user profile information [Liu et al., 2017b], network of users involved in the protest [Wang and Zhou, 2021b] as well as the content of the tweet [Gallagher et al., 2018b].

4.0.1 #ShushantSinghRajput

Sushant Singh Rajput (SSR) was a Bollywood actor and celebrity who was found dead in his Mumbai apartment on June 14, 2020 [India Today, 2020]. The death of the 34-year old actor was reported as a case of suicide. However numerous dark conspiracies are triggered on social media, including debates of nepotism [Times, 2020], and the possibility of framing [Cohen et al., 2020] or murder [Contributors, 2020]. This led to rising of a social media movement, which was sustained on the social media and gave rise to a very connected and dedicated community of online users who identified themselves as *SSRians* [1].

Related Literature

A combined study of prominent news channels and politicians over the SSR controversy revealed that the commentators on the topic were rewarded with higher retweet rates, which can be attributed to the widespread discourse engagement [Akbar et al., 2020].

In our work [Neha et al., 2021], we study the social media users' narratives that followed after the actor's death broke on news and social media. The narrative included counterpublics [Jackson and Banaszczyk, 2016], defined as marginalized communities that distribute messages to diverse social groups, raise awareness, and challenge dominant narratives. Our study aims to reveal the strategies adopted by Twitter activists (i.e., counterpublics) to share, spread, and mobilize the support of the counterpublic campaign about the untimely death of the Bollywood actor.

Theory of Connective Action: The logic of collective action answers the general question of why people get involved in collaboration with one another by explaining that people act collectively to achieve a common goal [Marwell and Oliver, 1993]. Traditionally, collective action refers to loosely connected groups of individuals, usually led by certain organizers or influential users [Bimber et al., 2012]. In contrast, the logic of connective action is based on the idea of digital media functioning as organizing agents, whereas traditional organizations are either not present or

[1] `https://www.urbandictionary.com/define.php?term=SSRians&defid=15832257`

are loosely responsible for providing coordination [Bennett and Segerberg, 2012b]. In that sense, connective action leverages the weaker ties present in social media, where users are self-motivated to post about the topic or share them. The interpersonal network hence formed can be similar to collective action sans any formal organizations. There are underlining economic and psychological logic driving the connective action, i.e., co-production and personalized sharing of expression, respectively. The two prominent indicators of a connective action are (i) a large number of participants in a movement, and (ii) a very small number of users staging the connective action through the creation of content. To enrich the knowledge of how social media is deployed during social movements and how a movement is carried differently in the online world than the offline counterpart, we need to understand (i) who participates in a given movement, and (ii) how people create a narrative in the social media around the protest.

Connective action comprises networked and decentralized actions of mobilization in contrast to the traditional collective action characterized by centralized resource mobilization or led by a formal organization [Bennett and Segerberg, 2012a]. The most crucial aspect of the emergence of connective action is the rise of self-claimed activists who co-ordinate themselves, challenge the formal organization, and conduct a campaign [Bimber et al., 2012]. Counterpublics have been found to form retweet networks on social media to gain legitimacy [Lotan et al., 2011] and recommend relevant messages to the supporters of the campaign [Starbird and Palen, 2012]. Connective action holds an assumption of a decentralized network since the activists who participate in the campaign are self-motivated to participate [Marwell and Oliver, 1993]. The user retweet network can therefore be used to analyze the organizational structure of the campaign [Wang and Zhou, 2021b].

Methodology

We adopt a network perspective to unpack the three major mechanisms of the connective action framework. We focus on the activists and their content posted to understand the first mechanism (i.e., *generative role-taking*) underlying the connective action. When users on social media use common hashtags, it creates a context

for like-minded people. The connection of like-minded individuals thus gives rise to a networked public [Xiong et al., 2019, Xu, 2020, Wang and Zhou, 2021b]. We divide the networked public into two categories, information generators and information drivers. The information generators work on content creation, while the drivers engage in driving the discussion by retweeting the content. To inspect the second mechanism (i.e., *hashtag-based storytelling*), we perform an evolutionary analysis of hashtags used in the campaign. We divide the hashtags into buckets based on their mutually exclusive appearance in the tweets and use topic modeling on the content shared among the buckets to identify topics focused on in the different buckets. The third mechanism (i.e., *formation of alignment network*) focuses on how the activists use social media for issue alignment and achieve virality. Identifying fellow activists supporting the cause is crucial to achieving a collective goal (i.e., virality) [Bimber et al., 2012]. We thus use community detection to identify sub-communities within the activist community to account for the diversity of users involved in the campaign. We also focus on how the narratives differ among sub-communities and examine any pattern within and among sub-communities. This study thus expands the literature on connective action framework and counterpublic campaigns and asks the below research questions:

- RQ1: What is the organizational structure of the social media counterpublic campaign around the death of Singh Rajput (SSR)?
- RQ2: How did hashtag-based storytelling evolve during the counterpublic campaign?
- RQ3: How did the campaign activists with different perspectives achieve issue alignment on the topic?

Data

The time duration of data collection coincided with an increase in media coverage and counterpublic narratives on Twitter. We used the Twitter search API to collect the tweets about the topic through trending hashtags which included #candle4ssr, #justice4ssr, #ssr, #sushantsingrajput. We curated a total of $1,027,213$

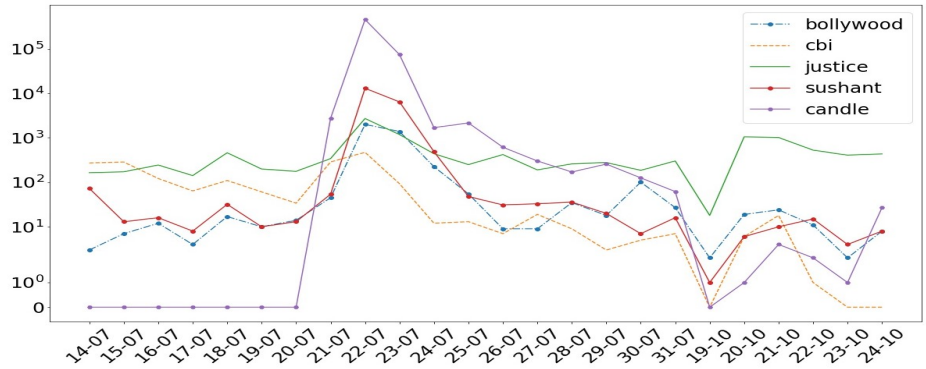

Figure 4.1: Evolution of counterpublic campaign over the period of three months with respect to hashtag buckets as presented in Table 4.1

tweets from 67,822 users using the official Twitter API. The duration of data collection spanned approximately 102 days from July 17, 2020, to October 21, 2020. The tweets consisted of 76,781 original tweets and 950,432 retweets. Any random tweet, on average, consists of 14.9 words, giving a good window for analysis of the user's thoughts around the campaign. counterpublics [Jackson and Banaszczyk, 2016], defined as marginalized communities that distribute messages to diverse social groups, raise awareness, and challenge dominant narratives. **Pre-processing** Before performing any analysis on the collected tweets, we converted all the tweets into lower-case, removed stop-words, and removed any occurrence of URL from the tweets. We removed any tweet with less than 3 words to keep informative tweets for further analysis. We also removed tweets with hashtags with a frequency less than 100 in our dataset. The selection of the most frequent hashtags served to identify the narratives that became popular. The hashtags belonging to a bucket were identified based on a common theme (e.g., Bollywood and media cover hashtags with movie actors or journalists) or a different variation of the same keyword (e.g., candle4SSR written as candleforssr or candle4shushant written as candleforsushant) as shown in Table 4.1. Tweets that used hashtags from more than one bucket were excluded from the analysis due to limitation of intention understanding that may require looking beyond the hashtag usage.

We construct a retweet network from the person who posted the message to

Table 4.1: Table showing the bucket of hashtags in the counterpublics campaign against the dominant narrative.

Hashtag bucket	**Hashtag variants**	**Tweet count**
#candleforssr	#candle4ssr, #candleforsushant, #candle4sushant, #candles4s	543, 897
#justiceforssr	#justiceforsushantsinghrajput, #ssrkoinsaafdo (give justice to SSR), #arrestculpritsofssr	11, 622
#sushantsinghrajput	#sushantsinghrajpoot, #sushantinourheartsforever, #ssrians, #sushanthsinghraj, #shushant	20, 486
#bollywood / #media	#akshaykumar, #salmankhan, #kanganaranaut, #bollywoodpakisilink, #rheachakraborty, #ankitalokhande, #boycottkhans	4, 064
#cbiforssr	#cbienquiryforsushantsinghrajput, #cbiivestigationforsushant, #cbicantbedeniedforssr, #cbienquiryforssr	1, 904

the user who retweeted the message to capture information-sharing activities for message-motivated communication. The retweet network is directed and weighted, where the direction indicates the flow of information, and the weight indicates the number of retweets between the two users.

We use descriptive network analysis coupled with modularity analysis and hashtag-based topical analysis to examine strategies used by Twitter users to build collective agendas and mobilize attention. We first make a user retweet network that consists of 79, 170 nodes and 490, 910 directed and weighted edges.

To answer RQ1, we examine the overall network structure and information flow of the tweets among counterpublics. We also identify the most active hashtag activists from the collected dataset, defined by activists' in-degree and out-degree centrality scores. While the in-degree centrality captures the level of user initiative in information sharing, the out-degree centrality accounts for the influence and communication power of the activist.

For RQ2, we bucket the hashtags according to their mutually exclusive appearance. Social media users created numerous hashtags relating to the Bollywood actor. Selecting only the popular hashtags was to identify the narratives that went pop-

Table 4.2: Network descriptive statistics for the top information drivers and generators to understand the organizational structure of the counterpublic campaign. $^{*}p < 0.05$,** $p < 0.01$,*** $p < 0.001$ analyzed using unpaired Mann–Whitney U test. SD stands for Standard Deviation.

	Top Information Generator		Top Information Driver		
Metric	Mean	SD	Mean	SD	p
Active Days	7.65	20.19	12.05	24.94	$***$
Number of Followers	8024.8	107137.7	122.084	351.87	$***$
Number of Followees	479.54	3278.9	136.861	336.64	$***$
Number of Tweets	8225.29	22076.6	9204.433	14673.42	$***$
Indegree Centrality	8.37	0.0002	0.0013	0.0052	$*$
Outdegree Centrality	8.37	0.0018	0.0013	0.0042	$*$
Betweenness Centrality	4.86	1.50	1.29	0.00013	$***$
Closeness Centrality	0.003	0.0012	0.015	0.016	$***$
Eigenvector Centrality	0.0012	0.0035	0.0024	0.0097	$*$

ular during the campaign. The final set of hashtags' buckets used for the study is presented in Table 4.1. We further analyze the content of the tweets from different hashtag buckets to understand the dominant narratives around the hashtags.

To examine RQ3, we apply community detection on the retweet network to discuss how the counterpublic campaign narratives differ among the sub-communities. For community detection, we use CNM (Clauset-Newman-Moore) greedy modularity maximization algorithm [Clauset et al., 2004]. CNM is a bottom-up agglomerative clustering algorithm that maximizes the modularity [Newman and Girvan, 2004] of the community structure in a greedy manner. Once we have identified the sub-communities, we examine how the topics presented by the sub-communities differ for detecting alignment in the sub-communities.

Figure 4.2: #candleforssr

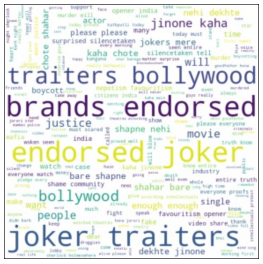

Figure 4.3: #bollywood

Figure 4.4: #cbiforssr

Figure 4.5: #justiceforssr

Figure 4.6: Word clouds for narrative hashtag bucket from Table 4.1.

Analysis

Network descriptive analysis The descriptive network analysis of a network can help identify the user dynamics and their clustering patterns during the online campaign. We present the descriptive analysis of the retweet network of the counter-public campaign in Table 4.3. The retweet network was found to be very sparse, with a network density of 0.000078. The sparseness in the network is expected given the large number of nodes and edges in the network. Usually, the retweet network tends to cluster rather than be evenly distributed, which can indicate the formation of an echo chamber around a topic [Shen et al., 2020]. The average in-degree and out-degree centrality for the activists were 7.83, which indicates that the average connection between activists for either retweeting or being retweeted is equal. The average clustering coefficient for the network is 0.060, which is very low. The low clustering coefficient indicates that all the activists are not well connected. Based on the out-degree centrality, a single user's highest number of retweets is $23,210$. While based on the in-degree centrality, the activist who retweeted the maximum number of times is $1,253$.

The in-degree centralization of the network is 0.0065, while the out-degree centralization is 0.29. A higher out-degree centralization indicates a set of users who were more frequently retweeted than others. Comparatively, a lower network in-degree suggests that the activists were more or less equally active while retweeting about the campaign. This result indicates the evidence towards slacktivism, defined

as actions requiring minimum effort and participation cost, like retweeting since it does not require the user to write their content [Bozarth and Budak, 2017]. Since the counterpublics were mostly slactivists, the campaign's main goal was to obtain momentum and raise awareness about the campaign.

Table 4.3: Descriptive statistics of the overall retweet network for SSR counterpublics campaign.

Metric	**Mean value**
Network Density	0.00078
In-degree Centrality	7.83
Out-degree Centrality	7.83
Clustering Co-efficient	0.060
In-degree Centralization	0.0065
Out-degree Centralization	0.29

To answer RQ1, we divide the activists involved in the counterpublic campaign into two parts based on their in-degree and out-degree centrality measures. We select the top 1000 activists in our dataset based on their in-degree and out-degree centrality. The top 1000 users with a high out-degree centrality are referred to as top information generators, and the top 1000 users with the highest in-degree centrality are referred to as top information drivers. We analyze the descriptive network statistics for the top information drivers and generators to understand the organizational structure of the counterpublic campaign. The descriptive network statistics for the top generators and drivers are listed in Table 4.2. Based on the descriptive statistic analysis summary of the activist's attributes, a typical information generator was active for 7.65 days, had about $8,024$ followers, followed 479 users, and tweeted $8,225$ times. While on the other hand, a typical information driver was active for 12 days and had a comparable follower-to-followee ratio. Mann–Whitney U tests were performed to examine whether the difference between information generators and information drivers is significant or not. We perform Mann–Whitney U tests since the test does not make any inherent assumption about the population distribution. We found that there is a significant difference between the active days, the number of followers, and the number of followers, as shown in Table 4.2.

Although the average number of days a user participated in the campaign is

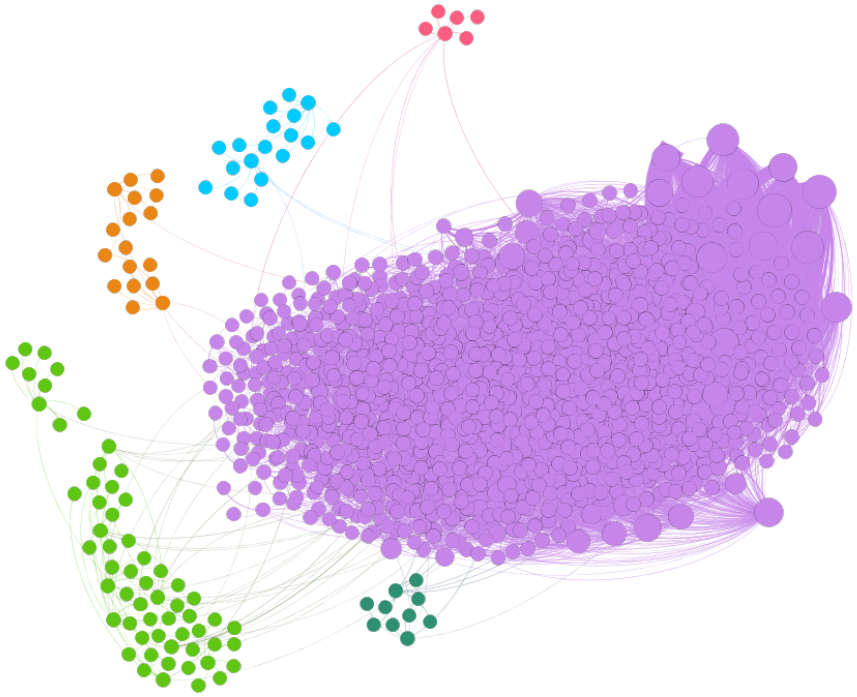

Figure 4.7: Figure showing the community formed among top information generators and their top drivers. Each color uniquely identifies a sub-community. Sub-community 1, shown in purple, constitutes 92.96% of the users. The second sub-community, shown in green, constitutes 4.15% of the users. While the blue sub-community includes 1.27%, orange comprises 1.2%, dark green comprises 0.7%, and pink sub-community comprises 0.42% of the users, respectively.

low for drivers and information generators, we found that the drivers were more active than the generators. From the eigenvector centrality score, we can conclude that since the information driver's score is more than the generator's score, drivers are more actively connected with other active campaign activists. However, the betweenness centrality for a generator is more than the driver, indicating generators are more likely to have a shorter path between two activists. The active retweeting of the campaign hashtags and a mix of centralized information aggregation and decentralized information generation are key to developing connective action.

Evolution of the counterpublic campaign narratives

To analyze how the counterpublic campaign evolved over the period, we plot the frequency of narratives' buckets identified through hashtags in Figure 4.1. The division of hashtags is presented in Table 4.1. We found that all the hashtags generally saw a spike between July, 20, 2020, and July, 24, 2020. The tweets with hashtags #cbiforssr and #justiceforssr were initially used more; however, during the period of highest frequency, #candleforssr was used most times. The use of #Bollywood hashtags also rises during the spike. #justiceforssr, however, was the most consistent hashtag bucket throughout the data collection.

To understand what narrative was spread in tweets within the hashtag buckets and how they differ, we plot the word cloud of the tweets from hashtag buckets as shown in Figure 4.6. The dominant narrative from #candleforssr was the declaration of online protest against the debate on the suicide of the late actor. The #candleforssr bucket revolves around demanding justice, mobilization through participation, and mention of debate and journalists (e.g., Arnab Goswami). The #justiceforssr bucket showed some narratives similar to #candleforssr, in addition to mentioning influential people, murder conspiracy, and shades at Mumbai police as shown in Figure 4.5. The #bollywood bucket in Figure 4.3, mainly included tweets mocking other Bollywood celebrities and despising nepotism. #cbiforssr, which was one of the first spikes in the dataset, consisted of tweets about inquiry, involvement of CBI (Central Bureau of Investigation), and topics of justice, protest, and nepotism as shown in Figure 4.4.

Issue alignment among the counterpublics

We used the top 1000 generators and their top 10 drivers to identify whether there is the formation of any sub-community within the network and whether different sub-communities share different narratives. The reason for selecting the top generators is to account for the most popular content in the campaign. We apply the CNM algorithm for community detection [Clauset et al., 2004] among the counterpublics. The number of iterations for the community detection algorithm was 100. The average clustering coefficient was 0.021, with an average degree of 14.075, modularity of 0.35, and network diameter of 9. We found 6 sub-communities in our user-retweet network as shown in Figure 4.7 with each community represented by a different color. The retweet network of top generators is densely connected, which shows evidence

Table 4.4: Table with topics discussed among top 1000 information generators and drivers respectively.

Justice	singh, world, justice, protest, digital
Candle	supporting, hope, smile, many, stand
Support$_T$	tweets, guys, digital, protest, million
Support$_C$	comment, below, million, reach, post
Media	arnab, goswami, debate, worldwide, live
Support	dead, watching, where, living, duty, suicide

of a connective campaign and a leaderless information-sharing framework. A few nodes with less connection indicate a centralized structure where information is shared from a few generators to many drivers. The formation of the dense cluster is evidence of connective action. We further perform LDA [Blei et al., 2002] on the combined tweets of top 1000 generators and top 1000 drivers to identify the major topics they share in the online environment.

Among the top 6 topics from the LDA as shown in Table 4.4, 3 dominant topics revolved around online mobilization represented as Support$_T$, Support$_C$, and Support. In the 3 mobilization topics, the social media users requested SSR fans and fellow social media users to retweet the content for widespread dissemination of information. While Support$_T$ is encouraged to tweet about the campaign, Support$_C$ suggests commenting on the posts to gain momentum on social media. On the topic

Table 4.5: Table with topics discussed among sub-communities.

Protest	protest, want, world, justice, digital, love, tweets
Media	arnab, know, rhea, raha, pagal (mad), aadmi (man), badla (revenge), will
Nepotism	money, huge, production, extract, houses, handle
Candle	light, candle, support, thank, fight, unity, hope, march

of Support, the counterpublics used words like duty and watching to encourage fellow campaigners and social media users to participate. The other 3 topics were identical to #justiceforssr and #candleforssr buckets, which were the two most prominent narratives in the overall campaign. The topic represented as Media included the debate led by news media on the investigation of the suicide.

To answer RQ3, we first run the LDA on the tweets from each sub-communities. Given that the people who were retweeting each other would belong to the same sub-community based on modularity analysis, the same set of tweets is expected from a given sub-community to remain connected. We set the number of topics as 3 with 10 words in each topic. To find the alignment among users from the 6 sub-communities, we identify the common topics in all the sub-communities. We found that users from sub-communities tweeted or retweeted more or less on the topics presented in Table 4.5 indicating an inter-connected community structure and issue alignment in sub-communities.

Conclusion

We apply the connective action framework to analyze the counterpublic campaign on online social media through a case study of the untimely death of Sushant Singh Rajput (SSR). We uncover the conditions under which hashtag activism can turn into connective action. With the help of a network-based approach, we investigate the users and their content simultaneously and identify three mechanisms of the connective action framework: generative role-taking, hashtag-based storytelling, and issue alignment among the different diverse groups of activists. To identify generative role-taking, we construct a user retweet network. We found that while top informa-

tion generators tend to have a shorter path than any fellow activist, the top drivers are more actively connected. The most consistent hashtag used for the counterpublic campaign was #justiceforssr, while #candle4ssr had the highest peak. Lastly, community detection indicates clique formation in the retweet network, where most of the top generators are densely connected, while a few have a sparse connection. The community of counterpublics thus indicates a mix of centralized and decentralized information aggregation with a strongly connected network with no standalone communities present.

Chapter 5

Understanding Common Narratives Across Protests

Mass mobilisation and protests are uncommon, but when they do happen, they could have unexpectedly dramatic results. Twitter and other social media platforms have emerged as hubs for the planning and development of online protests all across the world. To grasp people's perspectives, it becomes essential to interpret the many narratives shared during an online protest. In this paper, we suggest a methodology based on unsupervised clustering for comprehending the narratives present in a specific online protest. We offer new insights into the narratives expressed during an online protest through a comparative study of tweet clusters from three demonstrations against laws affecting government policy. We discovered well-known mass mobilisation narratives in case studies of government policy-related internet protests in India and the United Kingdom. We discovered accounts of local events and calls to action.

5.1 Introduction

Social media has become integral to various social movements and protests due to easy information dissemination and wider public reach [Korolov et al., 2016, De Choudhury et al., 2016b, Field et al., 2019, Wang and Zhou, 2021b, Lotan et al.,

2011]. Over 230 influential anti-government protests have erupted worldwide in the past six years, covering 110 countries [1]. Irrespective of the different socio-economic circumstances or political agendas, the various online protests share similar morphological features in using social media for self-organization and obtaining a more significant number of participants [González-Bailón et al., 2011]. Using a hashtag to build a collective narrative makes Twitter one of the prime spots for conducting protest [Wang and Chu, 2019]. While Twitter enables a broad reach of the protest, a fine-grained analysis of various narratives present within a protest setting may also help decipher the people's perception and shed light on people's will and social protest's overall focus.

Figure 5.1: Figure showing examples of different narratives expressed by people during online protests. CTA: Call-to-action, OGA: On-ground activities, GRV: personal grievances.

Previous studies on social media movements/protests have focused on different collective narratives in the campaign [Neha et al., 2021, Wang and Zhou, 2021b]. The narratives range from information dissemination (such as personal grievances) around the topic [Sinpeng, 2021, Field et al., 2019]; to call for participation [Rogers et al., 2019] or reporting of on-ground activities [Varol et al., 2014b], as shown

[1] https://carnegieendowment.org/publications/interactive/protest-tracker

in Figure 5.1. The grievance narrative might include personal stories of perceived injustice or other forms of hardships related to the cause. On-ground activities are narrative that either comes from people who are witnessing the offline protest or posts about current online activities related to the protest. The call for participation (call-to-action) narrative urges the users to participate in the cause by either being part of the physical protest or using social media to tweet protest-related posts. Although the different narratives during a protest have been studied individually, a unified discussion of various narratives present within a protest is scarce [Wang and Zhou, 2021b].

In this work, we focus on various narratives in recent instances of the Reform movement [DeFronzo and Gill, 2019] in India and the UK, where policies introduced by the government in power were deemed unjust and demanded to be repealed [desk, 2021,Damini Nath, 2019,desk, 2022]. According to Social Movement Theory, Reform movements [] is a subclass of movements that calls for change in a policy/behavior without alteration to the complete social institution. The reform movements studied in this work are as follows -

Citizenship Amendment Act, 2019 (CAA): The Citizenship amendment Act, 2019 was passed by the Indian Government on December 11, 2019. It allows the illegal immigrants who have faced religious persecution in Afghanistan, Bangladesh, or Pakistan to seek citizenship in India if they have entered India on or before December 31, 2014 [Chandrachud, 2020]. This led to a protest in the country with a debate on the non-secular roots of the Act. The protests were rooted in the exclusion of other religious minority communities like Rohingya Muslims, Jews, Bahais, and Zoroastrians from seeking citizenship.

Farmer's Protest, 2020 (FP): The Indian government proposed the Farmer's bill on September 20, 2020, which stirred the country. The country's farmers feared that the three laws introduced in the bill would result in the abolishment of the minimum support price (MSP), leaving farmers at the mercy of big corporations. Protests broke out in both the online and offline world due to the proposed bill, with people demanding it be repealed. The turn of events in the country led the Indian government to finally repeal the bill on November 09, 2021, ending the year-long

protest in the country [desk, 2021].

Kill the Bill Protest, 2022 (KTB): The Police, Crime, Sentencing, and Courts Bill (PCSC) introduced new police powers and reviewed the present rules around crime and protests in England and Wales. The activists opposed the law due to its ability to impose conditions on any protest deemed disruptive to the local community, leading to upto 10 years of jail. The punishable conditions included disruption of public properties, and statues, along with restricting access in and out of parliament [desk, 2022].

Since each protest is unique in its goals, we propose an unsupervised cluster-based framework to identify the different narratives of the protest. The primary motivation for using cluster-based analysis is to leverage the semantic difference between clustered texts and identify fine-grained separation between clusters as different narratives in a protest. We also focus on a comparative study of narratives in protests under study to examine converging narratives across the different protests. Using a clustering-based framework, we bridge the gap of unified narrative detection in social media protests and identify converging narratives across different protests. Broadly, we ask the following research questions:

RQ 1: What are the different narratives present in a protest?

RQ 2: What are the most prominent narratives present within a protest?

RQ3: Are there any converging narratives across protests?

The succeeding sections of the paper are organized as follows. We discuss the related work in Section II. Next, we discuss the Data and Methods in Section III, followed by Results in Section IV and the Conclusion in Section V.

5.2 Related work

The early work on social media protests focused on how a protest can reach critical mass for collective mobilization through network analysis of participants [González-Bailón et al., 2011, Barberá et al., 2015]. The analysis of textual features for understanding the sentiment of protest tweets shows the prevalence of negative sentiment [Costa et al., 2015] and specific psycho-linguistic lexicons over the oth-

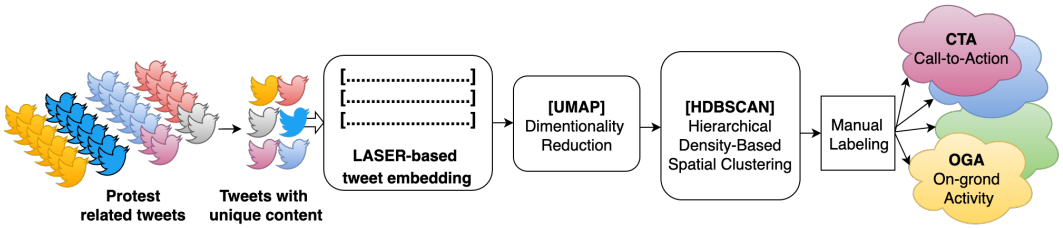

Figure 5.2: Framework to identify dominant narratives amid social media protest. The different color of tweet represents different narrative tweets present in the dataset.

ers [De Choudhury et al., 2016a]. A study of tweeting activity during a protest shows that social media activists plan the protest and share relevant tweets with a future date and time of offline protest conduct (call-to-action) to gain critical mass [Muthiah et al., 2015, Yaqub et al., 2017]. The call-to-action tweets have helped successfully predict future protests [Muthiah et al., 2016b, Goode et al., 2015, Korkmaz et al., 2016].

More recently, researchers have focused on advocates [Ranganath et al., 2016a] and extreme users [Zheng, 2016, Spiro and Ahn, 2016, Dash et al., 2021, Horawalavithana et al., 2021] who tend to spread the content of one particular side over the other, leading to the formation of echo chambers and biased opinions [Ingrams, 2017, Garimella et al., 2018a]. Moreover, the politicians use social media to create a "us vs. them" narrative leading to marginalization and polarization among the public at large [Karkın et al., 2015]. While some protests are accompanied by offline gatherings, which may turn violent [Lotan et al., 2011, Wang and Zhou, 2021a, Sinpeng, 2021], others are sustained on the online platform only [Mitra et al., 2016, Neha et al., 2021]. The use of collective action to conduct recent anti-government protests has shown how hashtag activism has helped reach mass mobilizaiton [Sinpeng, 2021, Wang and Zhou, 2021a].

Social media protests often tend to bring social justice and help marginalized social groups [Khatua et al., 2019]. On the other hand, posts shared during protest activity shed light on the people's will and hardships [Costa et al., 2015]. Protest tweets have been used to study and reduce online prejudice around certain social groups [Wei et al., 2020c]. The study of anti-vaccine infodemic helped to understand

the human perception around the topic [Germani and Biller-Andorno, 2021]. With twitter achieving the center position for most of the modern online activism and protests, manipulation of the campaigns has emerged as another topic of interest among various research [Jakesch et al., 2021, Badawy et al., , Luceri et al., 2019]. The study of social media-mediated protests have been done concerning protest prediction [Korolov et al., 2016], protest participation [González-Bailón et al., 2011], and study of protest growth [Barberá et al., 2015].

Our work builds on the previous literature on the ingredients present in the protests, including grievance [Sinpeng, 2021], call-to-action [Rogers et al., 2019, González-Bailón et al., 2011], and reporting of on-ground activity [Lotan et al., 2011]. However, to the best of our knowledge, we are the first to propose an unsupervised tweet clustering-based framework to identify the presence and relative abundance of all the narratives in an online protest.

5.3 Results

RQ1: Narratives present in a protest

Per RQ 1, we examine the clusters formed in each campaign using our framework. We leverage the semantic difference in the clusters to identify plausible narratives in the campaign. We have not reported the tweets clustered as noise for brevity. For annotation of protest clusters into different narratives, we leverage the previous literature on protest studies in different parts of the world [Rogers et al., 2019, Sinpeng, 2021, Lotan et al., 2011].

CAA: With the duplicate threshold set as 30, the number of unique tweets for clustering was 36, 109. As shown in Figure 5.3, 6 clusters of tweets were formed for CAA. For analysis of narratives, we manually annotate randomly selected two sets of 10 sample tweets from each cluster. Table 5.1 shows the 4 different narratives clusters in the campaign with highest engagement. The other two clusters belonged to personal grievances and location-specific tweets. In terms of engagement (i.e., tweet/retweet activity), the largest cluster showed skepticism towards the Act. On manual intervention, we found skepticism in both tweets that opposed the Act and

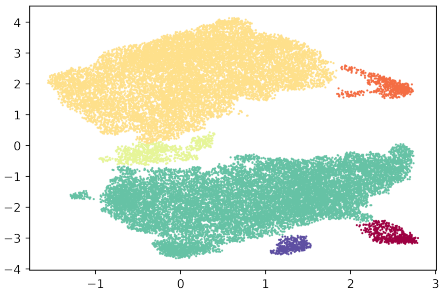

Figure 5.3: #CAA

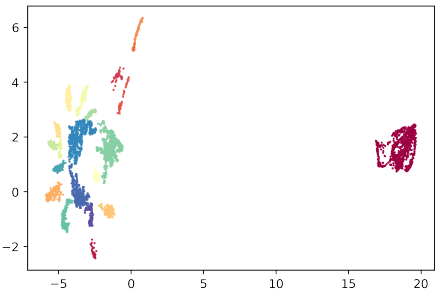

Figure 5.4: #FP

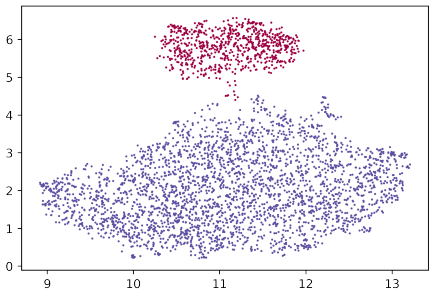

Figure 5.5: #KTB

Figure 5.6: Clusters of narratives for CAA, FP and KTB respectively.

those who opposed the protesters. The second dominant narrative for CAA was the Questioning cluster, where the tweets posed questions to the Act, politicians, and protesters for violent actions. The other two important narrative clusters included call-to-action and on-ground activities clusters. The example tweets for the 4 major narratives are presented in Table 5.1.

FP: The duplicate threshold to give the best clustering result for FP is 30. Unlike CAA, with the same framework for narrative clustering, we found 20 clusters for FP. However, we focused on the top 4 clusters for further analysis, constituting more than 500 unique tweet text each. As shown in Table 5.1, the most dominant narrative in FP was call-to-action, with $6,287$ (CTA-AP) and 845 (CTA-AP) unique tweets respectively . While the cluster (denoted as CTA) called for participation in support of farmers, the cluster CTA-AP (i.e., Call to action against politicians) contained tweets against the ruling government for their proposal of the bill.

KTB: The duplicate threshold for KTB was set to 5, as the data collected for the protest was small. With duplicate threshold as 5, KTB had $203,355$ total engagements, with $5,601$ original tweets and $197,754$ retweets. The UK protest on the policing formed 2 clusters using our framework as shown in Figure 5.5. Among the two clusters, more engagement was around call-to-action. Table 5.1 shows the example of tweets from both on-ground activities and call-to-action for the protest.

Table 5.1: Main narratives present in the protests under study. P stands for Protests

P	Narrative	Unique Tweets	#Tweets	#Users	Example
	Questioning	13,380	2,387,533	278,184	The police showed patience and did not shoot. Who fired at 56 policemen in Lucknow? Those who are saying that they do not have any paper, are they who are the end? Listen to the story of Pakistani Hindu.
CAA	Skepticism	15,274	3,911,679	466,139	Thousands on the street in support of CAA! I was not expecting this from Bhubaneshwar
	CTA	865	154,926	72,415	What ever way is there we oppose poisonous #CAA Rangoli is our tool
	OGA	647	98,221	48,276	The demonstration was held today at the Valluvar Fort in Chennai on behalf of the Tamil National Party and the Tamil National Alliance. Urged to withdraw the Citizenship Act
	CTA	6,287	13,734	464	Through violence, haarsh weather, beatings, & amp; Deaths of OurThers and Sisters, We Stand Tall And Adud! We Will Not Back Up Down Till Farm Laws ARE Repealed. #300deathsatProtest The war continues ... the war continues ...
FP	CTA-AP	845	26,897	9,470	We want humanity in our country We want a government who serve for nation/people not for corporations No more BJP
	OGA	683	66,660	2,538	Watch- On #HolikaDahan, Farmers in Rajasthan #BurnFarmLawsOnHoli amidst slogans for 300+ who have died in #FarmersProtest.
	OGA	742	20,557	9,431	Don't worry we are no longer being gaslighted @BorisJohnson @Conservatives @sajidjavid no trial needed you are as bad as each others. Lie after lie after lie #BorisJohnsonMustGo #ToriesDevoidOfShame #ToriesUnfitToGovern
KTB	CTA	2,958	178,499	56,079	The government are stripping away our fundamental rights with the #PolicingBill. It would: - Ban noisy protests - Criminalise the GRT community - Increase stop search powers - Jail protest organisers for up to 10 years. Join us at protests tomorrow to #KillTheBill

Chapter 6

Tackling Online Threats during Protest

6.1 #CitizenshipAmmendmentAct

In India, the first Citizenship Act was enacted in 1955, which enlisted the routes to obtain citizenship in India, which include 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 [Chandrachud, 2020], 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 [Chandrachud, 2020]. The debate on the non-secular roots of the Act was rooted in the exclusion of other religious minority communities like Rohingya Muslims, Jews, Bahais, and Zoroastrians from seeking citizenship sd. The protesters deemed it unconstitutional to discriminate on religious grounds, as only certain persecuted illegal immigrants benefited from the Act. At the same time, the supporters / counter-protesters based their argument on the presumption that refugees of particular minority religious communities are more in need of asylum [Chandrachud, 2020].

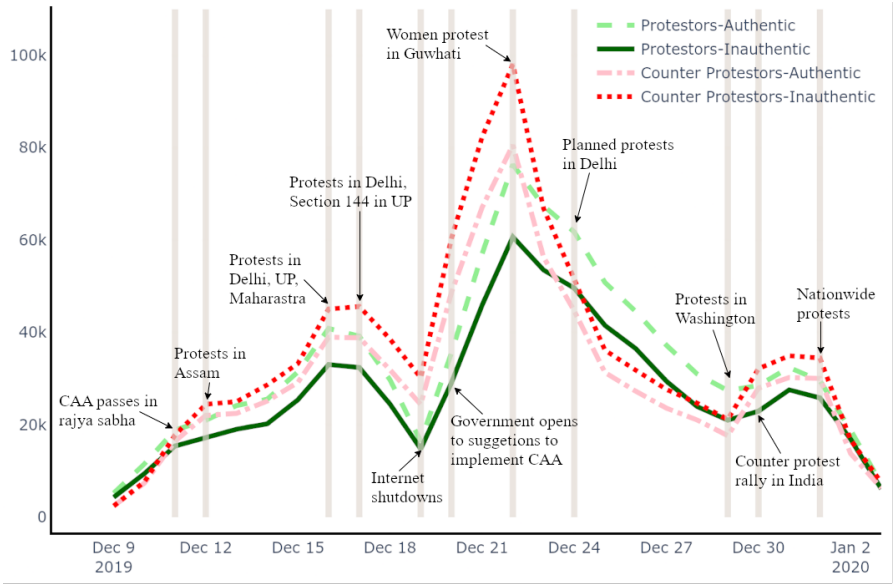

Figure 6.1: Timeline of counter-protest and protest vs on-ground activity

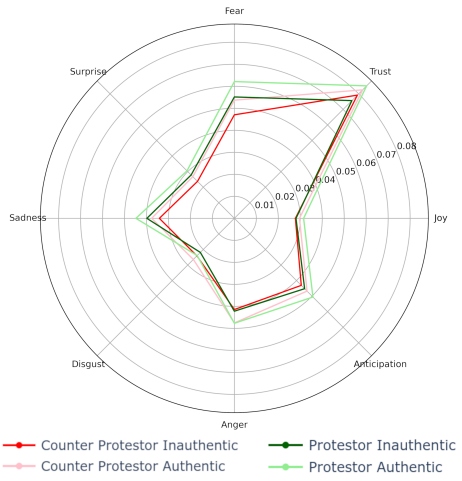

Figure 6.2: Radar plot to show the 4 set of users and their plutchik-8 emotions.

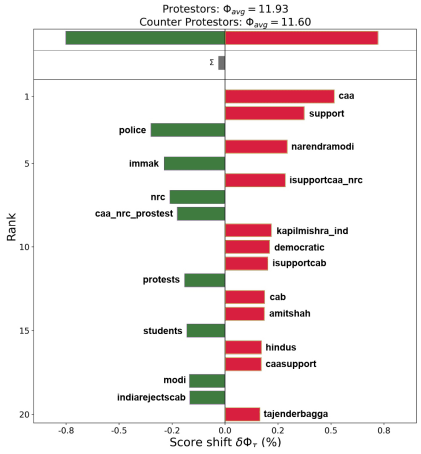

Figure 6.3: Application of word shift graphs for highlighting narratives that characterize protesters and counter-protesters. Protesters are shown in green, while counter-protesters are shown in red.

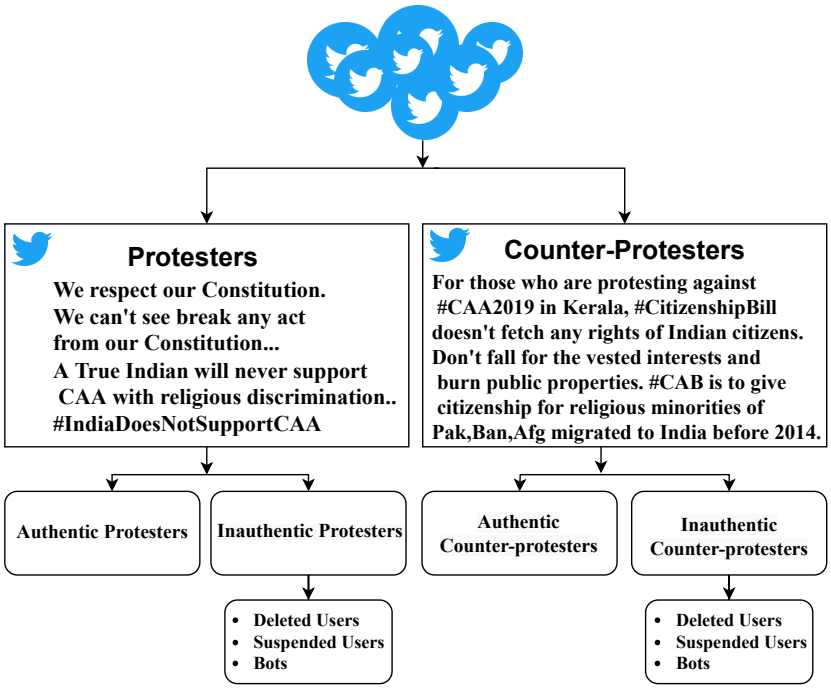

Figure 6.4: The users considered under study divided into 4 sets.

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 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 [Gallagher et al., 2018b]. While the campaign gained traction on both Twitter and the offline world, the prevalence of manipulation of the campaign was found to be evident [Hari et al., 2021]. 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 [Shao et al., 2018], suspended and deleted users (who tend to disseminate malicious content [1]) who participated in the discourse as *Inauthentic users*. In contrast, *Authentic users* are

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

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 6.4.

Twitter has been the focus of various characterization studies involving online campaigns [Gallagher et al., 2018b, De Choudhury et al., 2016b, Panda et al., 2020]. However, the characterization of a campaign concerning various sorts of authentic and inauthentic actors in discourse is limited [Chang et al., 2021b]. To the best of our knowledge, we are the first to conduct a characterization study of a campaign with various users (Figure 6.4) in a less investigated setting, i.e., India. Our analysis contributes to a few recent preliminary studies on the CAA [Mahapatra and Garimella, 2021,Hari et al., 2021] 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 [Rashed et al., 2021]. 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 [Bessi and Ferrara, 2016], and more recently, the coronavirus [Dunn et al., 2020]. 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 [Khatua and Khatua, 2016]. 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 [Wang and Zhou, 2021b], and polarization around a controversial topic [Garimella et al., 2018c]. 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 [Chandrachud, 2020], 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 [Chandrachud, 2020]. 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 sd. 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 [Chandrachud, 2020].

Related Work

Protests are a form of collective sociopolitical action in which members with similar beliefs express their objection to a cause or situation [Amenta and Young, 1999]. Time and again, the world witnesses protests over a government policy, bill [Raynauld et al., 2018, Wei et al., 2020d], or the government itself [Starbird and Palen, 2012]. In online discussions related to societal issues, users in one group may show hatred for users with opposing views [Wei et al., 2020d]. 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 [Wei et al., 2020d]. #BlackLivesMatter (#BLM) is another campaign where the people on social media were divided into two groups [Gallagher et al., 2018b]. Researchers studying online protests and campaigns on micro-blogging websites have used various stance detection techniques [Mohammad et al., 2016] and news articles [Riedel et al., 2017, Awadallah et al., 2012] to identify opposing views automatically. More recently, researchers have

Table 6.1: 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

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 [Gurukar

et al., 2020]. The previous literature studied the contrasting opinions through computing topic models followed by Jensen-Shannon divergence among the individual topic opinions [Fang et al., 2012]. The different perspectives or viewpoints have also been explored using a graph partitioning method that exploits the social interaction between the users [Quraishi et al., 2018]. Previous research has also shown almost 75% of the protests are planned in advance [Bahrami et al., 2018]. There has been a lot of interest in the social media domain to predict the on-ground activity through the social media platform [Ranganath et al., 2016b, Rogers et al., 2019, Muthiah et al., 2016a, De Choudhury et al., 2016b]. The authors in [Wei et al., 2020d] used protest as an intervention to reduce online prejudice, with focus on manual annotation for understanding prejudice in the tweets [Wei et al., 2020d]. The study of protests have also been studied in regards to the volume of the status messages relating to the protest event [De Choudhury et al., 2016b, Gallagher et al., 2018b].

While social media has been used to share opinions and debate on current happenings [Slobozhan et al., 2021], the involvement of inauthentic users is becoming more prevalent on the platform [Ferrara et al., 2016]. The manipulation of the debate are studied with regards to bots [Shao et al., 2018, Bessi and Ferrara, 2016, Uyheng and Carley, 2021b, Chang et al., 2021b], pre-defined campaign toolkit users [Jakesch et al., 2021], co-ordinated accounts [Pacheco et al., 2020,Pacheco et al., 2021,Sharma et al., 2021], or trolls [Luceri et al., 2020, Gorrell et al., 2019c]. Social media manipulation has been extensively studied with respect to election campaigns [Uyheng and Carley, 2021b,Bessi and Ferrara, 2016]. In social media, bots refer to fully automated and semi-automated accounts that contribute to disinformation campaigns [Ferrara et al., 2016]. [Uyheng and Carley, 2021b] 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 [Uyheng and Carley, 2021b, Dunn et al., 2020]. While bot accounts that use abusive language are more likely to be suspended by Twitter [Uyheng and Carley, 2021b, Dunn et al., 2020], social media manipulation might involve propaganda [Gorrell et al., 2019c], or campaign toolkits [Jakesch et al., 2021], 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 [Luceri et al., 2020].

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 user's involvement in India, a country in Asia-pacific.

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 6.1. Our collected data

Table 6.2: 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 [Damini Nath, 2019].
December 16	376, 788.00	Student protests in Delhi [Web, 2019].
December 17	379, 699.00	Protest turns violent in Uttar Pradesh, Delhi, West Bengal and relaxed in Guwhati [ANI, 2019b, IANS, 2019b].
December 20	436, 616.33	Protesters turn violent with stone pelting in Gujarat, police vehicle burnt in UP, journalists detained in Kerala [dec, 2019d].
December 22	783, 662.33	Protesters arrested, Women protest in Guwhati [Service, 2019].
December 24	503, 779.00	Protesters die due to bullet injury in UP [dec, 2019b].
December 30	276, 724.33	Counter-protest rally in Madhya Pradesh, Indian-American protests in Washington [dec, 2019c, IANS, 2019a].
December 31	312569.66	Nation wide protests [IANS, 2019c, dec, 2019e].

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.

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 6.4, we query the Twitter API and botometer [Yang et al., 2020] on the user IDs obtained from tweets.

The inauthentic users that we consider for the study include suspended users, deleted users and bots. Table 6.4 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 [Yang et al., 2020], 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 6.2. 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 [Watch, 2019]. 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`

User Characterization

To capture the fine-grained divergence among the users, we build on the previous work by [Rashed et al., 2021] 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.

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., 2021] 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 6.1. 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.

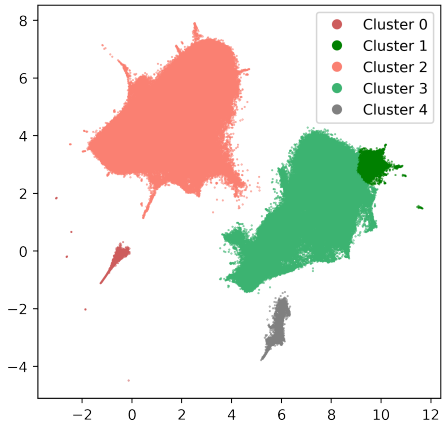

Figure 6.5: Here Clusters 0 and 2 represent counter-protest users and Clusters 1 and 3 represent protest users. Cluster 4 had a purity below 80% and hence was not considered.

(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 [Rashed et al., 2021]. We apply LASER (Language-Agnostic Sentence Representations)[2] to obtain 1024-dimensional

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

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 [McInnes et al., 2018]. The projection of users on lower dimension helps overcome the curse of dimensionality [Verleysen et al., 2003]. UMAP projects the data elements closer if they are similar, while dissimilar data elements are placed far apart. The projected user vectors are further clustered using hierarchical density-based clustering (HDBSCAN) [McInnes and Healy, 2017]. 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 6.5. 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 6.5. 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 6.5 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.

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.

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

Taking cues from previous works around planned protests [Bahrami et al., 2018, Muthiah et al., 2016a], 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 6.4. Figure 6.1 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 6.2. The first peak in the 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 [Damini Nath, 2019]. Students in Assam held protest opposing the Act [dec, 2019a] 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 [Web, 2019]. Anarchy was observed the same day in West Bengal, where people torched trains and staged sit-ins on the railway tracks [ANI, 2019a]. Inauthentic counter-protesters made most tweets at this day, followed by authentic protesters. On *December* 17*,* 2019, several metro

stations in Delhi [Soni, 2019] 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 [dec, 2019d]. The government opened to suggestions and reaching out to the protesters [dec, 2019d]. 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 [Service, 2019]. 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 [dec, 2019b]. 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 [dec, 2019c]. The counter-protesters started rallies on *December* 30*,* 2019 in support of CAA in different parts of the country [IANS, 2019a]. 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 [int, 2019].

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

6.1.1 Threat by Inauthentic Users

CAA: Presence of authentic and inauthentic users in the discourse we identify users based on their authentic behavior to study the role of inauthentic users in mobilizing protests and counter-protests. As shown in Table 6.3, among the 263,869 users considered for the analysis, we found Twitter suspended 13,871 users. In comparison, 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 are 27,122, including suspended and deleted users. Next, we use botometer API [Yang et al., 2020] 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 were protesters. Table 6.4 shows the complete set of users considered for the analysis.

Table 6.3: 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 6.4: 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

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 6.3. Figure 6.11 shows the distribution of bots in the stance-based cluster. We notice, as shown in Figure 6.10 and Table 6.5 that as the bot score varies from 0.8 to 0.5, there is a sharp

Table 6.5: Distribution bots in the discourse with varying bot scores. P: protesters, CP: counter-protesters, T: total number of users for which bot score 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%)

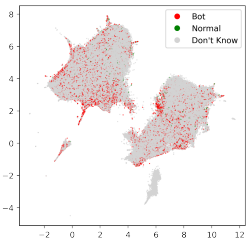

Figure 6.6: bot score >= 0.5

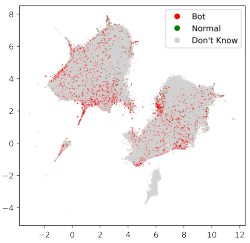

Figure 6.7: bot score >= 0.6

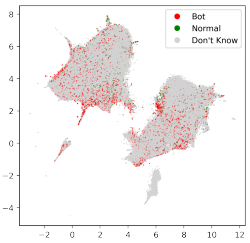

Figure 6.8: bot score>= 0.7

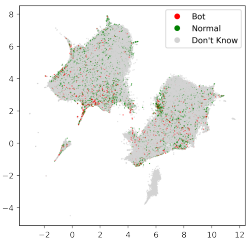

Figure 6.9: bot score >= 0.8

Figure 6.10: Distribution of the users with varying bot scores ranging from from 0.6-0.8.

decline of bots above 0.7. This shows the presence of semi-automated accounts in the discourse.

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

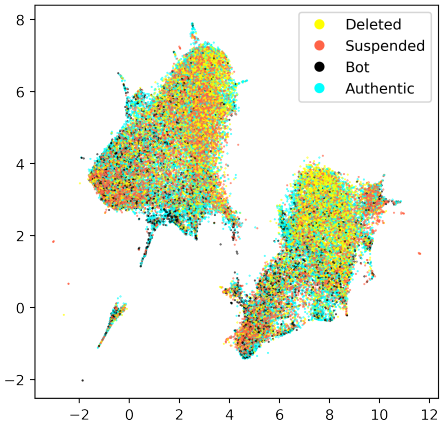

Figure 6.11: The presence of 4 set of users in the cluster.

Table 6.6: 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)

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 it's maximal strongly connected sub-graphs, the presence of both protesters and

counter-protesters in the largest strongly connected sub-graph indicates the path between the protesters and counter-protesters. 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 7.1 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 7.1, 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 on Twitter was conducted by campaigners who were connected to both sides of the debate; were not strongly

Figure 6.12: Overall follower-followee network of the protesters and counter-protesters. protesters are represented by green color while counter-protesters by red color.

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 the 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 a different sets 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 6.6 shows the difference between 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 followers 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.

Conclusion

In this work, we characterize the Citizenship Amendment Act (CAA) discourse on Twitter concerning 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 reachability of inauthentic users to their authentic counterparts.

Chapter 7

Timeline

		2022			2023				
S.No	Task	Oct	Nov	Dec	Jan	Feb	Mar	...	July
1	Co-ordination Detection across campaigns								
1.1	Build a co-ordination detection algorithm								
1.3	Result assesment								
1.4	Paper Submission								
2	Prediction of Offline Protest activities								
2.1	Creation of dataset from available protest								
2.2	Build a prediction classifier for protest prediction								
2.4	Paper Submission								
3	Journal Writing								
4	Thesis Writing								
5	Defense								

Figure 7.1: Overall Timeline for Ph.D

Chapter 8

Outline of Thesis

- CHAPTER 1: Introduction
  - Understanding Protest Strategies and Objectives
  - Understanding online threats
  - Understanding offline threats
  - Legal and Ethical Concerns
  - Targets and Contributions
- CHAPTER 2: Literature Review
  - Understanding Protest Strategy and Objectives
  - Understanding Online Threats during Protests
  - Understanding Offline Threats during Protests
- PART I : Understanding Project Strategies and Objectives

  CHAPTER 3: Understanding Counterpublic Campaign

  CHAPTER 4: Understanding Common Narratives Across Protests
- PART II: Understanding online threats

  CHAPTER 5: Understanding Threats by Inauthentic Actors

  CHAPTER 6: Understanding Threat by Hateful Actors

CHAPTER 7: Understanding Threat by Coordinated Actors

- PART III: Understanding Offline Threats

  CHAPTER 8: Detection of On-ground Protest Activity During Protests

Chapter 9

Publications

1. Neha, K., Mohan, T., Buduru, A. B., & Kumaraguru, P. (2021, November). Truth and travesty intertwined: a case study of #SSR counterpublic campaign. In Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 643-648).

2. Neha, K., Agrawal, V., Kumar, V., Mohan, T., Chopra, A., Buduru, A. B., ... & Kumaraguru, P. (2022, June). A Tale of Two Sides: Study of Protesters and Counter-protesters on #CitizenshipAmendmentAct Campaign on Twitter. In 14th ACM Web Science Conference 2022 (pp. 279-289).

3. Neha, K., Agrawal, V., Buduru, A. B., ... & Kumaraguru, P. The Pursuit of Being Heard: An Unsupervised Approach to Narrative Detection in Online Protest. In Proceedings of the 2022 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining.

Chapter 10

Acknowledgements

1. I want to extend my thanks to the Precog Research Lab and IIIT Hyderabad members, for providing insightful comments on the work.

2. I want to acknowledge and thank all my co-authors.

3. I want to thank the IT support team at IIIT Delhi and IIIT Hyderabad for being so active in solving all concerns.

4. Lastly, thanks to the incredible family I have, and their tireless effort to keep me sane and going through this journey.

Bibliography

[dec, 2019a] (2019a). CAB protest: Students clash with police near Assam secretariat. Economic Times.

[dec, 2019b] (2019b). Anti-CAA Protests Highlights December 24: Rahul and Priyanka Gandhi stopped outside Meerut by police, returns Delhi. jagran News Desk.

[dec, 2019c] (2019c). Indian-Americans protest CAA third time in nine days, author=IANS. freepressjournal.

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

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

[int, 2019] (2019). Internet shutdowns.

[Akbar et al., 2020] Akbar, S. Z., Sharma, A., Negi, H., Panda, A., and Pal, J. (2020). Anatomy of a rumour: Social media and the suicide of sushant singh rajput. arXiv preprint arXiv:2009.11744.

[Akhtar et al., 2021] Akhtar, S., Basile, V., and Patti, V. (2021). Whose Opinions Matter? Perspective-aware Models to Identify Opinions of Hate Speech Victims in Abusive Language Detection.

[Alonso et al., 2018] Alonso, O., Kandylas, V., and Tremblay, S. E. (2018). How it Happened: Discovering and Archiving the Evolution of a Story Using Social

Signals. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, (1):193–202.

[Amenta and Young, 1999] Amenta, E. and Young, M. P. (1999). Democratic states and social movements: Theoretical arguments and hypotheses. Social Problems, 46(2):153–168.

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

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

[Awadallah et al., 2012] Awadallah, R., Ramanath, M., and Weikum, G. (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, pages 523–532.

[Badawy et al., 2019] Badawy, A., Addawood, A., Lerman, K., and Ferrara, E. (2019). Characterizing the 2016 Russian IRA influence campaign. Social Network Analysis and Mining, 9(1).

[Badawy and Ferrara, 2018] Badawy, A. and Ferrara, E. (2018). The rise of jihadist propaganda on social networks. Journal of Computational Social Science, 1(2):453–470.

[Badawy et al.,] Badawy, A., Lerman, K., and Ferrara, E. Who Falls for Online Political Manipulation?

[Bahrami et al., 2018] Bahrami, M., Findik, Y., Bozkaya, B., and Balcisoy, S. (2018). Twitter reveals: Using twitter analytics to predict public protests. CoRR, abs/1805.00358.

[Barberá et al., 2015] Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., and González-Bailón, S. (2015). The critical periphery in the growth of social protests. PLoS ONE, 10(11).

[Bennett and Segerberg, 2012a] Bennett, W. L. and Segerberg, A. (2012a). The logic of connective action: Digital media and the personalization of contentious politics. Information, communication & society, 15(5):739–768.

[Bennett and Segerberg, 2012b] Bennett, W. L. and Segerberg, A. (2012b). The logic of connective action: Digital media and the personalization of contentious politics. Cambridge University Press.

[Bessi and Ferrara, 2016] Bessi, A. and Ferrara, E. (2016). Social bots distort the 2016 us presidential election online discussion. First monday, 21(11-7).

[Bimber et al., 2012] Bimber, B., Flanagin, A., and Stohl, C. (2012). Collective action in organizations: Interaction and engagement in an era of technological change. Cambridge University Press.

[Bittner et al., 2020] Bittner, M., Dettmar, D., Morejon Jaramillo, D., and Valta, M. J. (2020). Virtual tribes: Analyzing attitudes toward the LGBT movement by applying machine learning on Twitter data. Springer Proceedings in Complexity, pages 157–175.

[Blei et al., 2002] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2002). Latent dirichlet allocation. In Advances in neural information processing systems, pages 601–608.

[Bozarth and Budak, 2017] Bozarth, L. and Budak, C. (2017). Is slacktivism underrated? measuring the value of slacktivists for online social movements. In Proceedings of the International AAAI Conference on Web and Social Media, volume 11.

[Chandrachud, 2020] Chandrachud, A. (2020). Secularism and the citizenship amendment act. Indian Law Review, 4(2):138–162.

[Chang et al., 2021a] Chang, H.-C. H., Chen, E., Zhang, M., Muric, G., and Ferrara, E. (2021a). Social Bots and Social Media Manipulation in 2020: The Year in Review.

[Chang et al., 2021b] Chang, H.-C. H., Chen, E., Zhang, M., Muric, G., and Ferrara, E. (2021b). Social bots and social media manipulation in 2020: The year in review. arXiv preprint arXiv:2102.08436.

[Cinelli et al., 2020] Cinelli, M., Brugnoli, E., Schmidt, A. L., Zollo, F., Quattrociocchi, W., and Scala, A. (2020). Selective exposure shapes the facebook news diet. PloS one, 15(3):e0229129.

[Clauset et al., 2004] Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. Physical review E, 70(6):066111.

[Cohen et al., 2020] Cohen, S., Nutt, W., and Sagic, Y. (2020). Sushant Singh Rajput's father files FIR against actor's friend for abetting suicide.

[Conti et al., 2012] Conti, M., Das, S. K., Bisdikian, C., Kumar, M., Ni, L. M., Passarella, A., Roussos, G., Tröster, G., Tsudik, G., and Zambonelli, F. (2012). Looking ahead in pervasive computing: Challenges and opportunities in the era of cyber–physical convergence. Pervasive and mobile computing, 8(1):2–21.

[Contractor et al., 2015] Contractor, D., Chawda, B., Mehta, S., Subramaniam, L., and Faruquie, T. A. (2015). Tracking political elections on social media: Applications and experience. In IJCAI.

[Contributors, 2020] Contributors, I. T. (2020). SSR death case: Study decodes how BJP politicians pushed 'murder' narrative. [Online; accessed 06-October-2020].

[Costa et al., 2015] Costa, J. M., Rotabi, R., Murnane, E. L., and Choudhury, T. (2015). It is not only about grievances: Emotional dynamics in social media during the brazilian protests.

[Damini Nath, 2019] Damini Nath, V. S. (2019). After a heated debate, Rajya Sabha clears Citizenship (Amendment) Bill. The Hindu.

[Dash et al., 2021] Dash, S., Mishra, D., Shekhawat, G., and Pal, J. (2021). Divided We Rule: Influencer Polarization on Twitter During Political Crises in India.

[De Choudhury et al., 2016a] De Choudhury, M., Jhaver, S., Sugar, B., and Weber, I. (2016a). Social Media Participation in an Activist Movement for Racial Equality. Technical report.

[De Choudhury et al., 2016b] De Choudhury, M., Jhaver, S., Sugar, B., and Weber, I. (2016b). 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, volume 2016, page 92. NIH Public Access.

[DeFronzo and Gill, 2019] DeFronzo, J. and Gill, J. (2019). Social problems and social movements. Rowman & Littlefield.

[desk, 2021] desk, E. W. (2021). Farmers end year-long protest: A timeline of how it unfolded.

[desk, 2022] desk, T. B. I. W. (2022). What are the Kill the Bill protests? The Big Issue.

[Dunn et al., 2020] Dunn, A. G., Surian, D., Dalmazzo, J., Rezazadegan, D., Steffens, M., Dyda, A., Leask, J., Coiera, E., Dey, A., and Mandl, K. D. (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):S319–S325.

[ElSherief et al., 2017] ElSherief, M., Belding, E., and Nguyen, D. (2017). #notokay: Understanding gender-based violence in social media. In Eleventh International AAAI Conference on Web and Social Media.

[Fang et al., 2012] Fang, Y., Si, L., Somasundaram, N., and Yu, Z. (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, pages 63–72.

[Fast and Horvitz, 2016] Fast, E. and Horvitz, E. (2016). Identifying Dogmatism in Social Media: Signals and Models. Technical report.

[Ferrara,] Ferrara, E. WHAT TYPES OF COVID-19 CONSPIRACIES ARE POPULATED BY TWITTER BOTS? Technical report.

[Ferrara et al., 2016] Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7):96–104.

[Field et al., 2019] Field, A., Bhat, G., and Tsvetkov, Y. (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, 13:158–169.

[Gallagher et al., 2018a] Gallagher, R. J., Reagan, A. J., Danforth, C. M., and Dodds, P. S. (2018a). Divergent discourse between protests and counter-protests: #BlackLivesMatter and #AllLivesMatter. PLoS ONE, 13(4).

[Gallagher et al., 2018b] Gallagher, R. J., Reagan, A. J., Danforth, C. M., and Dodds, P. S. (2018b). Divergent discourse between protests and counter-protests: #blacklivesmatter and #alllivesmatter. PLOS ONE, 13(4):1–23.

[Garimella et al., 2018a] Garimella, K., De Francisci Morales, G., Gionis, A., and Mathioudakis, M. (2018a). Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018, 2:913–922.

[Garimella et al., 2018b] Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. (2018b). Quantifying Controversy on Social Media. ACM Transactions on Social Computing, 1(1):1–27.

[Garimella et al., 2018c] Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. (2018c). Quantifying controversy on social media. Trans. Soc. Comput., 1(1).

[Gerbaudo and Treré, 2015] Gerbaudo, P. and Treré, E. (2015). In search of the 'we' of social media activism: introduction to the special issue on social media and protest identities. Information Communication and Society, 18(8):865–871.

[Germani and Biller-Andorno, 2021] Germani, F. and Biller-Andorno, N. (2021). The anti-vaccination infodemic on social media: A behavioral analysis. PLoS ONE, 16(3 March):1–14.

[González-Bailón et al., 2011] González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. (2011). The dynamics of protest recruitment through an online network. Scientific Reports, 1:1–7.

[Goode et al., 2015] Goode, B. J., Krishnan, S., Roan, M., and Ramakrishnan, N. (2015). Pricing a protest: Forecasting the dynamics of civil unrest activity in social media. PLoS ONE, 10(10):1–25.

[Gorrell et al., 2019a] Gorrell, G., Bakir, M. E., Roberts, I., Greenwood, M. A., Iavarone, B., and Bontcheva, K. (2019a). Partisanship, Propaganda and Post-Truth Politics: Quantifying Impact in Online Debate.

[Gorrell et al., 2019b] Gorrell, G., Bakir, M. E., Roberts, I., Greenwood, M. A., Iavarone, B., and Bontcheva, K. (2019b). Partisanship, Propaganda and Post-Truth Politics: Quantifying Impact in Online Debate. The Journal of Web Science, page 7.

[Gorrell et al., 2019c] Gorrell, G., Bakir, M. E., Roberts, I., Greenwood, M. A., Iavarone, B., and Bontcheva, K. (2019c). Partisanship, propaganda and post-truth politics: Quantifying impact in online debate. The Journal of Web Science, 7.

[Grčar et al., 2017] Grčar, M., Cherepnalkoski, D., Mozetič, I., and Kralj Novak, P. (2017). Stance and influence of twitter users regarding the brexit referendum. Computational social networks, 4(1):1–25.

[Gurukar et al., 2020] Gurukar, S., Ajwani, D., Dutta, S., Lauri, J., Parthasarathy, S., and Sala, A. (2020). Towards quantifying the distance between opinions. In ICWSM.

[Haider et al., 2020] Haider, S., Luceri, L., Deb, A., Badawy, A., Peng, N., and Ferrara, E. (2020). Detecting Social Media Manipulation in Low-Resource Languages.

[Haidt, 2011] Haidt, J. (2011). Moral Psychology and the Misunderstanding of Religion1. In The Believing Primate: Scientific, Philosophical, and Theological Reflections on the Origin of Religion.

[Hari et al., 2021] Hari, K. S., Aravind, D., Singh, A., and Das, B. (2021). Detecting propaganda in trending twitter topics in india—a metric driven approach. In Emerging Technologies in Data Mining and Information Security, pages 657–671. Springer.

[Horawalavithana et al., 2021] Horawalavithana, S., Ng, K. W., and Iamnitchi, A. (2021). Drivers of Polarized Discussions on Twitter during Venezuela Political Crisis. In ACM International Conference Proceeding Series, pages 205–214. Association for Computing Machinery.

[Howard and Kollanyi, 2016] Howard, P. N. and Kollanyi, B. (2016). Bots,# strongerin, and# brexit: computational propaganda during the uk-eu referendum. arXiv preprint arXiv:1606.06356.

[Hristakieva et al., 2022] Hristakieva, K., Cresci, S., Da San Martino, G., Conti, M., and Nakov, P. (2022). The spread of propaganda by coordinated communities on social media. In 14th ACM Web Science Conference 2022, WebSci '22, page 191–201, New York, NY, USA. Association for Computing Machinery.

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

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

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

[India Today, 2020] India Today (2020). Sushant Singh Rajput dies by suicide at 34 in Mumbai. [Online; accessed 14-June-2020].

[Ingrams, 2017] Ingrams, A. (2017). Connective action and the echo chamber of ideology: Testing a model of social media use and attitudes toward the role of government. Journal of Information Technology and Politics, 14(1):1–15.

[Jackson and Banaszczyk, 2016] Jackson, S. J. and Banaszczyk, S. (2016). Digital standpoints: Debating gendered violence and racial exclusions in the feminist counterpublic. Journal of Communication Inquiry, 40(4):391–407.

[Jakesch et al., 2021] Jakesch, M., Garimella, K., Eckles, D., and Naaman, M. (2021). Trend alert: A cross-platform organization manipulated twitter trends in the indian general election. Proc. ACM Hum.-Comput. Interact., 5(CSCW2).

[Karkın et al., 2015] Karkın, N., Yavuz, N., Parlak, İ., and İkiz, Ö. Ö. (2015). Twitter use by politicians during social uprisings: An analysis of gezi park protests in turkey. In Proceedings of the 16th Annual International Conference on Digital Government Research, pages 20–28.

[Khatua et al., 2019] Khatua, A., Cambria, E., Ghosh, K., Chaki, N., and Khatua, A. (2019). Tweeting in support of LGBT? A deep learning approach. In ACM International Conference Proceeding Series, pages 342–345. Association for Computing Machinery.

[Khatua and Khatua, 2016] Khatua, A. and Khatua, A. (2016). Leave or remain? deciphering brexit deliberations on twitter. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pages 428–433.

[Korkmaz et al., 2016] Korkmaz, G., Cadena, J., Kuhlman, C. J., Marathe, A., Vullikanti, A., and Ramakrishnan, N. (2016). Multi-source models for civil unrest forecasting. Social Network Analysis and Mining, 6(1):1–25.

[Korolov et al., 2016] Korolov, R., Lu, D., Wang, J., Zhou, G., Bonial, C., Voss, C., Kaplan, L., Wallace, W., Han, J., and Ji, H. (2016). On predicting social unrest

using social media. In 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pages 89–95. IEEE.

[Leetaru and Schrodt, 2013] Leetaru, K. and Schrodt, P. A. (2013). Gdelt: Global data on events, location, and tone, 1979–2012. In ISA annual convention, volume 2, pages 1–49. Citeseer.

[Liu et al., 2017a] Liu, F., Ford, D., Parnin, C., and Dabbish, L. (2017a). Selfies as social movements: Influences on participation and perceived impact on stereotypes. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW).

[Liu et al., 2017b] Liu, F., Ford, D., Parnin, C., and Dabbish, L. (2017b). Selfies as social movements: Influences on participation and perceived impact on stereotypes. volume 1, New York, NY, USA. Association for Computing Machinery.

[Liu et al., 2018] Liu, Q. H., Lü, F. M., Zhang, Q., Tang, M., and Zhou, T. (2018). Impacts of opinion leaders on social contagions. Chaos, 28(5).

[Lotan et al., 2011] Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., et al. (2011). The Arab Spring— the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. International journal of communication, 5:31.

[Luceri et al., 2019] Luceri, L., Badawy, A., Deb, A., and Ferrara, E. (2019). Red bots do it better: Comparative analysis of social bot partisan behavior. In The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019, pages 1007–1012. Association for Computing Machinery, Inc.

[Luceri et al., 2020] Luceri, L., Giordano, S., and Ferrara, E. (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):417–427.

[Mahapatra and Garimella, 2021] Mahapatra, S. and Garimella, K. (2021). Digital public activism and the redefinition of citizenship: The movement against the cit-

izenship (amendment) act of india. In Weizenbaum Conference 2021: Democracy in Flux–Order, Dynamics and Voices in Digital Public Spheres, page 3. DEU.

[Marwell and Oliver, 1993] Marwell, G. and Oliver, P. (1993). The critical mass in collective action. Cambridge University Press.

[Mathur et al., 2019] Mathur, P., Sawhney, R., Ayyar, M., and Shah, R. (2019). Did you offend me? Classification of Offensive Tweets in Hinglish Language. pages 138–148.

[Mccarthy and Zald, 1977] Mccarthy, J. D. and Zald, M. N. (1977). Resource Mobilization and Social Movements: A Partial Theory. Technical Report 6.

[McInnes and Healy, 2017] McInnes, L. and Healy, J. (2017). Accelerated hierarchical density based clustering. In 2017 IEEE International Conference on Data Mining Workshops (ICDMW), pages 33–42. IEEE.

[McInnes et al., 2018] McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

[Mitra et al., 2016] Mitra, T., Counts, S., and Pennebaker, J. W. (2016). Understanding anti-vaccination attitudes in social media. In Tenth International AAAI Conference on Web and Social Media.

[Mohammad et al., 2016] Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C. (2016). Semeval-2016 task 6: Detecting stance in tweets. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pages 31–41.

[Muthiah et al., 2016a] Muthiah, S., Butler, P., Khandpur, R. P., Saraf, P., Self, N., Rozovskaya, A., Zhao, L., Cadena, J., Lu, C.-T., Vullikanti, A., et al. (2016a). 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, pages 205–214.

[Muthiah et al., 2016b] Muthiah, S., Butler, P., Khandpur, R. P., Saraf, P., Self, N., Rozovskaya, A., Zhao, L., Cadena, J., Lu, C. T., Vullikanti, A., Marathe, A., Summers, K., Katz, G., Doyle, A., Arredondo, J., Gupta, D. K., Mares, D., and Ramakrishnan, N. (2016b). EMBERS at 4 years: Experiences operating an open source indicators forecasting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-Augu:205–214.

[Muthiah et al., 2015] Muthiah, S., Huang, B., Arredondo, J., Mares, D., Getoor, L., Katz, G., and Ramakrishnan, N. (2015). Planned protest modeling in news and social media. Proceedings of the National Conference on Artificial Intelligence, 5:3920–3927.

[Neha et al., 2021] Neha, K., Mohan, T., Buduru, A. B., and Kumaraguru, P. (2021). Truth and travesty intertwined: a case study of# ssr counterpublic campaign. In Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pages 643–648.

[Newman and Girvan, 2004] Newman, M. E. and Girvan, M. (2004). Finding and evaluating community structure in networks. Physical review E, 69(2):026113.

[Pacheco et al., 2020] Pacheco, D., Flammini, A., and Menczer, F. (2020). Unveiling coordinated groups behind white helmets disinformation. In Companion Proceedings of the Web Conference 2020, pages 611–616.

[Pacheco et al., 2021] Pacheco, D., Hui, P.-M., Torres-Lugo, C., Truong, B. T., Flammini, A., and Menczer, F. (2021). Uncovering coordinated networks on social media: Methods and case studies. In Proceedings of the International AAAI Conference on Web and Social Media, volume 15, pages 455–466.

[Panda et al., 2020] Panda, A., Kommiya Mothilal, R., Choudhury, M., Bali, K., and Pal, J. (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):1–14.

[Poltrock et al., 2012] Poltrock, S., ACM Digital Library., and ACM Special Interest Group on Computer-Human Interaction. (2012). (How) Will the Revolution be Retweeted? Information Diffusion and the 2011 Egyptian Uprising. CSCW, page 1434.

[Pond and Lewis, 2019] Pond, P. and Lewis, J. (2019). Riots and Twitter: connective politics, social media and framing discourses in the digital public sphere. Information Communication and Society, 22(2):213–231.

[Quraishi et al., 2018] Quraishi, M., Fafalios, P., and Herder, E. (2018). Viewpoint discovery and understanding in social networks. In Proceedings of the 10th ACM Conference on Web Science, pages 47–56.

[Raleigh et al., 2010] Raleigh, C., Linke, A., Hegre, H., and Karlsen, J. (2010). Introducing acled: an armed conflict location and event dataset: special data feature. Journal of peace research, 47(5):651–660.

[Ramakrishnan et al., 2014] Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R., Saraf, P., Wang, W., Cadena, J., Vullikanti, A., Korkmaz, G., Kuhlman, C., Marathe, A., Zhao, L., Hua, T., Chen, F., Lu, C. T., Huang, B., Srinivasan, A., Trinh, K., Getoor, L., Katz, G., Doyle, A., Ackermann, C., Zavorin, I., Ford, J., Summers, K., Fayed, Y., Arredondo, J., Gupta, D., and Mares, D. (2014). 'Beating the news' with EMBERS: Forecasting civil unrest using open source indicators. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (July 2015):1799–1808.

[Ranganath et al., 2016a] Ranganath, S., Hu, X., Tang, J., and Liu, H. (2016a). Understanding and identifying advocates for political campaigns on social media. WSDM 2016 - Proceedings of the 9th ACM International Conference on Web Search and Data Mining, pages 43–52.

[Ranganath et al., 2016b] Ranganath, S., Morstatter, F., Hu, X., Tang, J., Wang, S., and Liu, H. (2016b). Predicting online protest participation of social media users. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16, page 208–214. AAAI Press.

[Rashed et al., 2021] Rashed, A., Kutlu, M., Darwish, K., Elsayed, T., and Bayrak, C. (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, volume 15, pages 537–548.

[Raynauld et al., 2018] Raynauld, V., Richez, E., and Morris, K. B. (2018). Canada is #idlenomore: exploring dynamics of indigenous political and civic protest in the twitterverse. Information, Communication & Society, 21(4):626–642.

[Rezapour et al., 2019] Rezapour, R., Ferronato, P., and Diesner, J. (2019). How do moral values difer in Tweets on social movements? In Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, pages 347–351. Association for Computing Machinery.

[Riedel et al., 2017] Riedel, B., Augenstein, I., Spithourakis, G. P., and Riedel, S. (2017). A simple but tough-to-beat baseline for the fake news challenge stance detection task. ArXiv, abs/1707.03264.

[Rogers et al., 2019] Rogers, A., Kovaleva, O., and Rumshisky, A. (2019). Calls to action on social media: Potential for censorship and social impact. EMNLP-IJCNLP 2019, page 36.

[Saha et al., 2021] Saha, P., Mathew, B., Garimella, K., and Mukherjee, A. (2021). "short is the road that leads from fear to hate": Fear speech in indian whatsapp groups. In Proceedings of the Web Conference 2021, WWW '21, page 1110–1121, New York, NY, USA. Association for Computing Machinery.

[Santini et al., 2021] Santini, R. M., Salles, D., and Tucci, G. (2021). When machine behavior targets future voters : The use of social bots to test narratives for political campaigns in Brazil. International Journal of Communication, 15:1220–1243.

[Schmidt and Wiegand, 2019] Schmidt, A. and Wiegand, M. (2019). A survey on hate speech detection using natural language processing. In Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media,

April 3, 2017, Valencia, Spain, pages 1–10. Association for Computational Linguistics.

[Service, 2019] Service, E. N. (2019). Government open to suggestions to implement Citizenship Act, says MHA amid raging protests. newindianexpress.

[Shao et al., 2018] Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., and Menczer, F. (2018). The spread of low-credibility content by social bots. Nature communications, 9(1):1–9.

[Sharma et al., 2021] Sharma, K., Zhang, Y., Ferrara, E., and Liu, Y. (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, pages 1441–1451.

[Shen et al., 2020] Shen, F., Xia, C., and Skoric, M. (2020). Examining the roles of social media and alternative media in social movement participation: A study of hong kong's umbrella movement. Telematics and Informatics, 47:101303.

[Shevtsov et al., 2022] Shevtsov, A., Tzagkarakis, C., Antonakaki, D., and Ioannidis, S. (2022). Identification of Twitter Bots Based on an Explainable Machine Learning Framework: The US 2020 Elections Case Study. Technical report.

[Sinpeng, 2021] Sinpeng, A. (2021). Hashtag activism: social media and the #freeyouth protests in thailand. Critical Asian Studies, 53(2):192–205.

[Slobozhan et al., 2021] Slobozhan, I., Brik, T., and Sharma, R. (2021). Longitudinal change in language behaviour during protests: A case study of euromaidan in ukraine. arXiv preprint arXiv:2109.11623.

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

[Spiro and Ahn, 2016] Spiro, E. and Ahn, Y.-Y., editors (2016). Predicting Online Extremism,Content Adopters, and Interaction Reciprocity,

volume 10047 of Lecture Notes in Computer Science. Springer International Publishing, Cham.

[Sree Hari et al., 2021] Sree Hari, K., Aravind, D., Singh, A., and Das, B. (2021). Detecting Propaganda in Trending Twitter Topics in India—A Metric Driven Approach. pages 657–671.

[Starbird and Palen, 2012] Starbird, K. and Palen, L. (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, CSCW '12, page 7–16, New York, NY, USA. Association for Computing Machinery.

[Stella et al., 2018] Stella, M., Ferrara, E., and De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems.

[Times, 2020] Times, H. (2020). Sushant's death embroiled in a web of theories. [Online; accessed 03-August-2020].

[Uyheng and Carley, 2021a] Uyheng, J. and Carley, K. M. (2021a). Computational Analysis of Bot Activity in the Asia-Pacific: A Comparative Study of Four National Elections. Technical report.

[Uyheng and Carley, 2021b] Uyheng, J. and Carley, K. M. (2021b). 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, volume 15, pages 727–738.

[Varol et al., 2014a] Varol, O., Ferrara, E., Ogan, C. L., Menczer, F., and Flammini, A. (2014a). Evolution of online user behavior during a social upheaval. WebSci 2014 - Proceedings of the 2014 ACM Web Science Conference, (i):81–90.

[Varol et al., 2014b] Varol, O., Ferrara, E., Ogan, C. L., Menczer, F., and Flammini, A. (2014b). Evolution of online user behavior during a social upheaval. In Proceedings of the 2014 ACM conference on Web science, pages 81–90.

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

[Wang et al., 2014] Wang, D., Amin, M. T., Li, S., Abdelzaher, T., Kaplan, L., Gu, S., Pan, C., Liu, H., Aggarwal, C. C., Ganti, R., et al. (2014). Using humans as sensors: an estimation-theoretic perspective. In IPSN-14 proceedings of the 13th international symposium on information processing in sensor networks, pages 35–46. IEEE.

[Wang and Chu, 2019] Wang, R. and Chu, K.-H. (2019). Networked publics and the organizing of collective action on Twitter: Examining the #Freebassel campaign. Convergence, 25(3):393–408.

[Wang and Zhou, 2021a] Wang, R. and Zhou, A. (2021a). Hashtag activism and connective action: A case study of #HongKongPoliceBrutality. Telematics and Informatics, 61.

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

[Watch, 2019] Watch, H. R. (2019). India: Citizenship Bill Discriminates Against Muslims. Human Right Watch.

[Weart, 2015] Weart, S. (2015). Global warming: How skepticism became denia.

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

[Wei et al., 2020a] Wei, K., Lin, Y.-R., and Yan, M. (2020a). Examining Protest as An Intervention to Reduce Online Prejudice: A Case Study of Prejudice Against Immigrants. pages 2443–2454. Association for Computing Machinery (ACM).

[Wei et al., 2020b] Wei, K., Lin, Y. R., and Yan, M. (2020b). Examining Protest as An Intervention to Reduce Online Prejudice: A Case Study of Prejudice Against Immigrants. In The Web Conference 2020 - Proceedings of the World Wide Web

Conference, WWW 2020, pages 2443–2454. Association for Computing Machinery, Inc.

[Wei et al., 2020c] Wei, K., Lin, Y. R., and Yan, M. (2020c). Examining Protest as An Intervention to Reduce Online Prejudice: A Case Study of Prejudice Against Immigrants. In The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020, pages 2443–2454. Association for Computing Machinery, Inc.

[Wei et al., 2020d] Wei, K., Lin, Y.-R., and Yan, M. (2020d). Examining protest as an intervention to reduce online prejudice: A case study of prejudice against immigrants. In Proceedings of The Web Conference 2020, WWW '20, page 2443–2454, New York, NY, USA. Association for Computing Machinery.

[Xiong et al., 2019] Xiong, Y., Cho, M., and Boatwright, B. (2019). Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement. Public Relations Review, 45(1):10–23.

[Xu, 2020] Xu, W. W. (2020). Mapping Connective Actions in the Global Alt-Right and Antifa Counterpublics. International Journal of Communication, 14(0).

[Yang et al., 2020] Yang, K.-C., Varol, O., Hui, P.-M., and Menczer, F. (2020). Scalable and generalizable social bot detection through data selection. Proceedings of the AAAI Conference on Artificial Intelligence, 34(01):1096–1103.

[Yaqub et al., 2017] Yaqub, U., Chun, S. A., Atluri, V., and Vaidya, J. (2017). Analysis of political discourse on twitter in the context of the 2016 US presidential elections. Government Information Quarterly, 34(4):613–626.

[Yardi and Boyd, 2010] Yardi, S. and Boyd, D. (2010). Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. Bulletin of Science, Technology & Society, 30(5):316–327.

[Zheng, 2016] Zheng, X. (2016). Social Network of Extreme Tweeters : A Case Study.