Understanding Coordinated Communities through the Lens of Protest-Centric Narratives: A Case Study on #CAA Protest

Kumari Neha[1], Vibhu Agrawal[1], Saurav Chhatani[2], Rajesh Sharma[1,3,4], Arun Balaji Buduru[1], Ponnurangam Kumaraguru [2]

[1] Indraprastha Institute of Information Technology, Delhi
[2] International Institute of Information Technology, Hyderabad
[3]Institute of Computer Science, Tartu, Estonia
[4]Center for Applied Research and Data Science (CARDS), IIT Ropar
nehak@iiitd.ac.in, vibhu18116@iiitd.ac.in,
saurav.chhatani@students.iiit.ac.in,rajesh.sharma@ut.ee,arunb@iiitd.ac.in,pk.guru@iiit.ac.in

Abstract

Social media platforms, particularly Twitter, have emerged as vital media for organizing online protests worldwide. During protests, users on social media share different narratives, often coordinated to share collective opinions and obtain widespread reach. In this paper, we focus on the communities formed during a protest and the collective narratives they share, using the protest on the enactment of the Citizenship Amendment Act (#CAA) by the Indian Government as a case study. Since #CAA protest led to divergent discourse in the country, we first classify the users into opposing stances, i.e., protesters (who opposed the Act) and counter-protesters (who supported it) in an unsupervised manner. Next, we identify the coordinated communities in the opposing stances and examine the collective narratives they shared. We use content-based metrics to identify user coordination, including hashtags, mentions, and retweets. Our results suggest mention as the strongest metric for coordination across the opposing stances. Next, we decipher the collective narratives in the opposing stances using an unsupervised narrative detection framework and found call-to-action, on-ground activity, grievances sharing, questioning, and skepticism narratives in the protest tweets. We analyze the strength of the different coordinated communities using network measures, and perform inauthentic activity analysis on the most coordinated communities on both sides. Our findings suggest that coordinated communities, which were highly inauthentic, showed the highest clustering coefficient towards a greater extent of coordination.

Introduction

Twitter has emerged as one of the leading platforms for organizing and participating in online protests (Theocharis et al. 2015; Wei, Lin, and Yan 2020; Goel and Sharma 2020). During a protest, Twitter provides a platform for users to create and share various narratives collectively (Wang and Zhou 2021). Narratives are verbal, graphic, or written arguments of interconnected actors and events over time (Ranade et al. 2022). On social media such as Twitter, narratives are often fragmented, consisting of chained posts that link events across multiple sources. Examples of shared narratives during a protest include sharing of personal grievance (Sinpeng 2021), call for participation (Rogers, Kovaleva, and Rumshisky 2019) or reporting of on-ground activities (Varol et al. 2014). During protest participation, various actors have been found to coordinate different narratives for the widespread reach of the protest. For example, during Arab Spring, participants collectively posted a tweet with the narrative "FREEDOM LOADING" along with an image of a progress bar to represent their solidarity (Starbird and Palen 2012). Although social media protests are inherently coordinated in nature (Starbird and Palen 2012), coordinated groups of users have recently been found to work together to manipulate online discourse (Ng and Carley 2022). Coordination between a set of users can be defined as an exceptional similarity leading to suspicious behavior traces in content (hashtag, n-gram, etc.), activity (timestamp, geolocation), identity (username, description), or a combination of multiple metrics(Nizzoli et al. 2021). The behavior of coordinated malicious actors with manipulative intentions might seem innocuous at an individual level and require a deeper analysis on a network level (Hristakieva et al. 2022). The coordinated actions of malicious actors may amplify the dissemination of deliberate content containing propaganda, biased news, or polarized content, thereby intensifying protest and increased disharmony in the society (Nizzoli et al. 2021; Brannen, Haig, and Schmidt 2020).

In this paper, we study the coordinated behavior and the narratives shared during the online debate surrounding the Indian Government's enactment of the Citizenship Amendment Act (#CAA) on December 12, 2019, on Twitter (Web 2019). The Act sparked a polarized discourse on social media, with users divided into two groups: users who opposed the Act (Protesters) and users who supported the Act (Counter-Protesters) (Gallagher et al. 2018). We represent the users who opposed the Act as P, while supporters of the Act are represented as CP. Previous research found that users with similar stances during a discourse show strong coordination (Pacheco et al. 2021). To advance the previous work and delve deeply into the coordinated communities during the protest, we propose the study of the coordinated communities with respect to shared narratives by users belonging to the opposing stances, i.e., protesters (P) and counter-

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

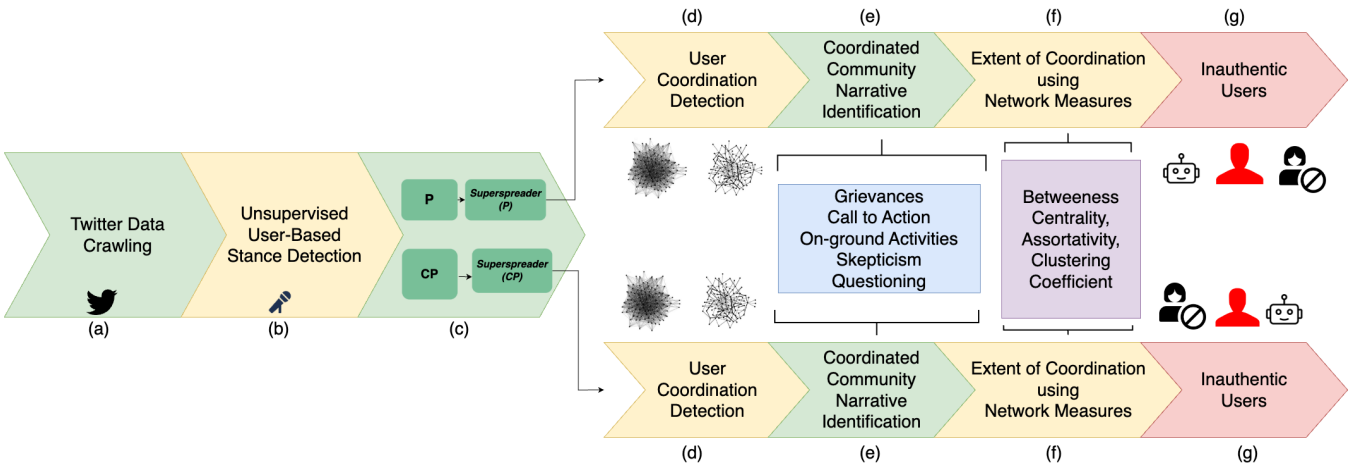


Figure 1: Overview of the proposed framework for studying narrative-based coordinated communities.

protesters (CP).

Figure 1 presents the analysis framework followed in this paper. To this end, we crawl 11,350,276 tweets from 931,175 users on Twitter around the CAA protest, using the trending protest-centric hashtags between December 07, 2019, and February 27, 2020 (Neha et al. 2022b). We first detect the stance of users using an unsupervised clustering-based approach (Rashed et al. 2021). Next, we identify the coordinated communities in the opposing stances using a network-based approach (Hristakieva et al. 2022). After this, we detect the narratives that the communities with different coordination strengths have shared in opposing stances, using an unsupervised narrative detection technique. We further analyze the presence of inauthentic behavior in the coordinated communities, defined by suspension on the Twitter platform or having a high bot score (Yang et al. 2020). More precisely, we ask the following research questions:

- How did the users coordinate during the protest?
  We use content-based metrics to identify user coordination, including hashtags, mentions, and retweets. Our findings suggest mentions as the strongest metric for coordination behavior in CAA, measured by the highest percentage of users retained on higher coordination threshold.
- What collective narratives are shared by the different communities from opposing stances?
  Our finding suggests *call-to-action*, *grievances*, *on-ground activities*, *skepticism*, and *questioning* were present across opposing stances. However, the presence of different narratives changed in frequency of occurrence across two opposing stances. For example, call-to-action was more common in counter-protesters, while questioning was dominant in P.
- What narratives prevailed in top coordinated communities, and did they indicate inauthentic activity?
  We found *skepticism* with *grievances* to be the most dominant narratives among counter-protesters communities, also exerting high inauthenticity. Among protesters, we found questioning with on-ground activity i.e., the second most dominant community exerted the highest inauthenticity.

Major highlights of the paper are:

- To the best of our knowledge, this work is the first that carries out the community analysis from opposing stances perspective, as compared to previous research that considered complete network analysis in a discourse (Pacheco et al. 2021).
- Next, we carried out the analysis of collective narratives shared by the coordinated communities during discourse.
- We employed three metrics: i) clustering coefficient, ii) betweenness centrality, and iii) assortativity to identify inauthentic behaviors. We found the clustering coefficient metric as a primary factor in identifying the inauthentic behaviors exerted by the users across opposing stances.

Ethics Statement

In conducting this research, we uphold ethical considerations paramount to the responsible analysis of online discourse. Our study centered on the examination of coordinated communities within the context of the Citizenship Amendment Act discourse on Twitter. It is essential to underscore that our data collection exclusively relied upon publicly available and anonymized Twitter content, thereby obviating the need for obtaining user consent. Moreover, we adhered to unsupervised methods for stance identification and narrative detection, ensuring that individual user privacy remained intact and no deliberate efforts were made to reveal personal demographic information. In addition, we conscientiously scrutinized the presence of inauthentic coordinated communities within the opposing stances, shedding light on the potential manipulation of online discourse. We agree that this kind of research is vital for understanding and protecting democratic conversations, as mentioned by Fiske (2022). In the ethical annotation of clusters within a protest context, we strongly emphasize fairness and impartiality. To achieve this, we take measures to mitigate location bias by engaging

annotators from diverse demographic backgrounds. This approach ensures that our analysis benefits from a broad range of perspectives and minimizes the risk of introducing geographic or regional biases into the annotation process. By drawing on annotators with varied backgrounds, we aim to uphold the ethical principles of inclusivity and fairness in our research.

Related Work

One of the first instances of coordination in protest participation was witnessed during the political uprising in Egypt in 2011, where the "Uninstalling dictator" with progress bar tweet was used with different variations by the participants towards a common goal (Starbird and Palen 2012). Studying individual perpetrators may overlook collective influence operations and fail to identify their inauthentic or problematic nature (Pacheco et al. 2021). The study of inauthentic coordinated activity also brings the challenge of distinguishing authentic activity from inauthentic activity, grassroots initiatives or deliberate malicious activity; and the narratives they share.

Coordinated Behavior

(Pacheco, Flammini, and Menczer 2020) proposed to use a binary distinction for coordinated inauthentic groups through retweets and narrative duplication as a metric. More precisely, the definition of coordination adopted by (Pacheco, Flammini, and Menczer 2020) was the immediacy of systematic retweets by accounts. In their approach, (Pacheco et al. 2021) first created suspicious behavior traces from content (hashtag, n-gram, etc.), activity (timestamp, geolocation), identity (username, description), or a combination of multiple dimensions. (Nizzoli et al. 2021) proposed the definition of coordination as an exceptional similarity between a group of users and chose a network-based approach for inauthentic coordination detection. (Vargas, Emami, and Traynor 2020) evaluated the effectiveness of the existing coordination detection approaches by building a binary classifier based on the statistical features extracted from the network for disinformation campaigns and legitimate Twitter communities. The major takeaway from the binary classifier-based approach was that the type of coordination and behavior based on it differ from campaign to campaign (Vargas, Emami, and Traynor 2020). (Sharma et al. 2021) propose a generative model to capture inherent coordination characteristics, leveraging Russia's Internet Research Agency dataset that targeted the 2016 U.S. Presidential Elections. (Hristakieva et al. 2022) pursued identifying coordination activity combined with propaganda detection and found that the combined analysis revealed harmful coordinated communities that were previously not noticeable. Most previous literature focused on elections to study coordination activity (Sharma et al. 2021; Nizzoli et al. 2021; Vargas, Emami, and Traynor 2020). However, the study of coordination activity for protest is scarce (Pacheco et al. 2021).

Protest Narratives

Understanding the narratives exchanged during protests is a crucial aspect that intersects with comprehending both protest participation and protest growth (Slobozhan, Brik, and Sharma 2023, 2022). Narratives are verbal, graphic, or written arguments of interconnected actors and events over time (Ranade et al. 2022). On social media, narratives are often fragmented, consisting of chained posts that link events across multiple sources. Protest studies have focused on shared grievances, revealing people's determination and hardships (Costa et al. 2015). Social media protests bring social justice and aid marginalized groups (Wu et al. 2023). Early research on social media protests focused on achieving critical mass and collective mobilization through network analysis of participants (González-Bailón et al. 2011; Barberá et al. 2015). Our work builds on the previous literature on the narratives present in the protests, including grievance (Sinpeng 2021), call-to-action (Rogers, Kovaleva, and Rumshisky 2019; González-Bailón et al. 2011), and reporting of on-ground activity (Lotan et al. 2011).

CP	#IsupportCAB2019, #HindusSupportCAB, #IndiaSupportsCAB, #ISupportCAA_NRC, #MuslimsWithNRC, #CAA_NRC_support, #ISupportCAA
P	#SCSTOBC_Against_CAB, #HindusAgainstCAB, #IndiansAgainstCAB, #CAA_NRC_Protest, #IndiaAgainstCAA, #CAA_NRCProtests, #CAAProtest, #CABProtest
Amb	#CAB, #CABBill, #cab, #CAB2019, #CitizenshipAmendmentAct, #caa, #CABPolitics, #CitizenshipAmmendmentAct

Table 1: Table showing initial set of hashtags used for data collection. The hashtags are manually identified as counter-protesters (CP), protesters (P), and Ambiguous (Amb).

Data

The dataset used in the study is created in two steps. Firstly, we use Twitter API to collect tweets, by tracking the trending hashtags around CAA. Next, we use an unsupervised stance detection technique to divide users into protesters (P) and counter-protesters (CP) based on their tweets. The anonymized version of our data is available at https://precog.iiit.ac.in/resources.html.

Twitter Data Crawling

We collect 11,350,276 tweets from 931,175 users, related to CAA between December 07, 2019, and February 27, 2020. The initial list of hashtags used for the data collection is shown in Table 1. The collected tweets contain 1,543,805 unique tweets and 9,806,471 retweets. We first map the tweets to their respective users and perform various preprocessing steps to filter out noise. First, we remove users with less than 5 tweets, reducing the total number of users to

276,149. Next, we removed links, emojis, emoticons, punctuation, and extra spaces from the tweets and performed case-folding. The tweets with less than three words were further removed. Another round of user filtration was performed to remove users with less than 5 tweets after tweet filtration step, removing 1,038 more users. The final set of users after all the pre-processing steps was 275,111 users.

We perform user-based unsupervised stance detection on the final set of 275,111 users to classify them as either P or CP. For coordination analysis, we further identify *superspreaders*, i.e., the top 1% of users from both P and CP.

Unsupervised Stance Detection: We build upon the unsupervised user-based stance detection technique proposed by (Rashed et al. 2021) to divide the users into P and CP. We identified 27 CP and 48 P hashtags by manually labeling the trending hashtags on CAA. The unsupervised detection of user's stance is carried out in 6 steps: (i) *Hashtag-based labeling*: We identify the users who only tweet with hashtags from either P or CP side, resulting in detection of 106,605 CP and 79,493 P users, (ii) *Label propagation*: We include re-tweeters of users identified in step (i), who retweeted P or CP users at least k-times (k = 15) (Rashed et al. 2021) respectively, resulting in an additional 114,977 CP and 79,613 P users, (iii) *Embedding creation*: We create 1024-dimensional user embedding obtained from taking the average of the vector of the filtered tweets for each user, using LASER (Language-Agnostic Sentence Representations)[1], (iv) *Dimensionality reduction*: We use Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes, Healy, and Melville 2018) to project users in 2-dimensional space, (v) *Clustering*: we cluster the 2-dimensional embedding using density-based approach, i.e., Hierarchical Density-Based Clustering(HDBSCAN) (McInnes and Healy 2017), and obtain 5 clusters for 270,889 users, (vi) *Cluster purity*: we use the identified stance produced from label propagation of step (i) to label the stance of the cluster if the cluster is pure (i.e., contains at least 30% labeled users obtained via label propagation and has at least 80% purity of labels). On performing purity analysis, we found 4 clusters have more than 80% purity of labels, with 2 belonging to P and 2 belonging to CP. The 4 clusters were used for further analysis and comprised 263,869 users, divided into 142,839 CP and 121,030 P.

	CP	**P**	**Total**
Users	7,480	5,383	12,863
Tweets	515	173	688
Retweet	732,035	434,611	1,166,646
Total Tweets	732,550	434,784	1,167,334

Table 2: Statistics of the total engagement produced by CP and P *superspreaders* during the online protest.

Superspreaders Identification: *Superspreaders* are defined as users who most actively participate in a campaign (Wang and Zhou 2021; Nizzoli et al. 2021). Here, *superspreaders* are defined as top 10% users from both P and CP sides, respectively, who authored the most popular tweets in the discourse. For most popular tweets, we filter tweets with 1,000 or more occurrences, identified through simple string-matching of content in the post (tweet-/retweet). Instead of retweeting posts, using the content of the post as an original tweet has recently gained attention in the research community as a means for a widespread reach of the post content (Jakesch et al. 2021). We select most popular tweets as a metric for *superspreader* identification as sharing similar tweets has been considered an essential means of coordination (Hristakieva et al. 2022). The *superspreaders* were responsible for 42% of total engagement (tweets/retweets) in the #CAA protest. Table 2 shows the statistics of engagement produced by *superspreaders* for the opposing stances. Our final dataset comprises 12,863 *superspreaders* who authored 1,67,334 tweets/retweets.

Coordination Detection

As a starting point for coordination detection, we need to identify the traces of suspicious behavior (Pacheco et al. 2021). An exceptional similarity between accounts' posts may be seen as evidence of suspicious coordinated behavior (Pacheco et al. 2021). We follow a content-based trace mining of the *superspreaders*, that includes hashtags, retweets, and mentions as potential metrics for coordination (Ng and Carley 2022). Once the traces of the *superspreaders* have been identified, we construct a network of *superspreaders* based on the similar behavior found in traces. We build upon the network-based coordination detection from previous literature, which considers coordination as a complex, non-binary concept (Nizzoli et al. 2021).

We perform the coordination detection on the user separated by stance (i.e., P and CP *superpreaders* separately) to unravel the intricate coordination dynamics within a particular stance. Similar behavior in the traces of CP and P users are identified through TF-IDF-weighted vector of three metrics, Tweet-IDs [2] for retweets, mentions and hashtags. Computing TF-IDF-weighted vector helps discount popular tweets and emphasize relevant tweets. For traces in (i) retweets, we compute the TF-IDF vector of the Tweet-IDs the user has tweeted, (ii) hashtags, we compute the TF-IDF vector on the user's hashtags throughout the protest, (iii) mentions, we compute the TF-IDF vector on all user mentions done by the user in their tweets. Next, we perform cosine similarity on traces on 3 metrics in P and CP *superspreaders* respectively. The pair of *superspreaders* and their cosine similarity result in an undirected weighted user-similarity network, where edge weights correspond to coordination strength.

To identify the best metric for coordination, we gauge the strength of coordination among the *superspreaders* within the community. We utilize network dismantling and remove nodes and edges iteratively based on the moving edge weight

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

[2]One of the response fields when crawling Twitter data, which represents a unique identifier of a Tweet. https://developer.twitter.com/en/docs/twitter-api/tweets/lookup/api-reference/get-tweets-id

threshold, such that after every iteration, we retain edges whose edge weights are larger or equal to the current threshold. We remove the edges, with weights lesser than the threshold at each iteration, till we have exhausted all the nodes in the network. Each subsequent network represents a different extent of coordination, measured by the corresponding value of the moving threshold. Coordination score for a user corresponds to the threshold value at which the node gets disconnected from the rest of the network. Among the 3 metrics under consideration, we found mention showed the strongest coordination behavior in CAA (both P and CP), measured by the highest percentage of users retained on higher coordination threshold values (0.8 to 1). We choose *mentions* as a metric for further coordination study. The first level of community detection performed on the mention metric produced 9 communities for P and 8 communities for CP. Figure 2 shows the communities formed for P and CP *superspreaders* using mention metric.

Narrative Detection

In this section, we interpret the collective narrative shared by P and CP *superspreader* communities. The collective narrative understanding can help shed light on the intention of the tweets posted by coordinated communities. We use an unsupervised collective narrative detection technique to understand narratives shared by opposing stances. Since narratives shared during a protest are subjective, and a fixed set of labels might not encapsulate all the collective narratives, we propose an unsupervised method. To perform narrative detection on *superspreader's* tweets, we follow the below steps: (i) we first identify tweets with the highest semantic duplicates. Given a threshold value, we use a density-based clustering separation and identify the best threshold values for duplicate tweets (Moulavi et al. 2014). Our best separation was achieved using a threshold of 30, which obtained 36,109 tweets that corresponded to 7,878,996 duplicated tweets, (ii) we project the tweets onto a two-dimensional plane using UMAP, (iii) we cluster the projected tweet vectors using HDBSCAN, which resulted in the formation of 6 clusters±(Neha et al. 2022a). To understand the clusters of collective narratives, we perform manual annotation. Previous research has found that Collective narratives during protests include call-to-action (Rogers, Kovaleva, and Rumshisky 2019; González-Bailón et al. 2011), personal grievance (Sinpeng 2021), and on-ground activity reporting (Lotan et al. 2011). Taking cues from the previous narratives present in protests, two groups (consisting of 2 students each) of annotators annotated randomly selected 2 sets of 10 tweets in each cluster. We calculate the inter-annotation agreement using Cohen's Kappa (Artstein and Poesio 2008) and found a strong agreement between annotators (0.95) for 5 clusters. We discard clusters for which we were unable to obtain strong inter-annotators from further analysis.

Among the 5 clusters, the largest cluster was annotated as skepticism (SKEP), identified by having tweets with a doubt-like attitude towards CAA. The second largest cluster was labeled questioning (QUEST), identified by tweets that questioned protests, protesters, etc. The other three clusters were annotated with grievances (GRV), call-to-action (CTA), and on-ground activities (OGA). Figure 2 shows the narrative tweets example of the 5 clusters. Collective narrative tweets are further mapped to the *superspreaders*, where each user is involved in a multitude of narrative posts. Mapping the communities of users to narratives is done in a multi-label format, such that each community constitutes a gradient of narratives.

Table 3 shows the gradient of narratives in different coordinated communities of opposing stances. We found that both P and CP *superspreader's* communities contained *skepticim (SKEP)* or *questioning (QUEST)* as the dominant narrative. Hence, as a naming convention, we start the community name with S (for *SKEP*) or Q (for *QUEST*), based on which narrative had more tweets. Next, we compare the number of tweets from non-dominant narratives (grievances (*grv*), on-ground activities (*oga*), or call-to-action (*cta*)), and the majority of the non-dominant narrative is chosen to complete the community name. The largest coordinated community formed for CP contained 36.08% users and showed *skepticism* and shared *grievances*, leading to the name (S-GRV1). Similar naming convention is used for all the communities, as shown in Table 3.

One of the major findings from our narrative analysis across P and CP communities suggests that *skepticism* along with *grievances* (Community P-1, P-2, CP-1, and CP-3 in Table 3 respectively) were the most dominant narrative. The second most dominant narrative across P and CP superspreaders was *questioning* with *on-ground activities* (P-3 and CP-2 in Table 3). We also found that *call-to-action* narrative was more dominant in CP as compared to P *superspreaders*. The offline counterpart of the protests (P) showed major sit-in protests in the country [3]. However, dominant narrative in P did not show *call-to-action* tweets, unlike CP *superspreaders*, who were more active on the online platform.

Extent of Coordination in Narrative Sharing

So far, we have determined the most coordinated communities in P and CP *superspreaders* and identified the gradient of collective narratives they shared. This section analyzes the extent of coordination in the P and CP communities. We use network characteristics to measure the extent of coordination and analyze narratives shared by communities with diverse coordination strengths. For every user, the coordination score corresponds to the threshold value at which the node gets disconnected from the rest of the network.

We explore the structural properties of the coordinated communities with the help of network measures of the network formed after each iteration of network dismantling. The network dismantling is done in 10 steps for each community formed in P and CP, leading to formation of 10 networks. We compute the structural properties of the 10 network, with communities intact as the first step, such that at each step the users retained are more strongly coordinated than in the previous step. We consider assortativity, clustering coefficients, and betweenness centrality to measure net-

[3] https://www.freepressjournal.in/delhi/delhi-metro-closes-4-stations-on-magenta-line-following-violent-protests-against-caa

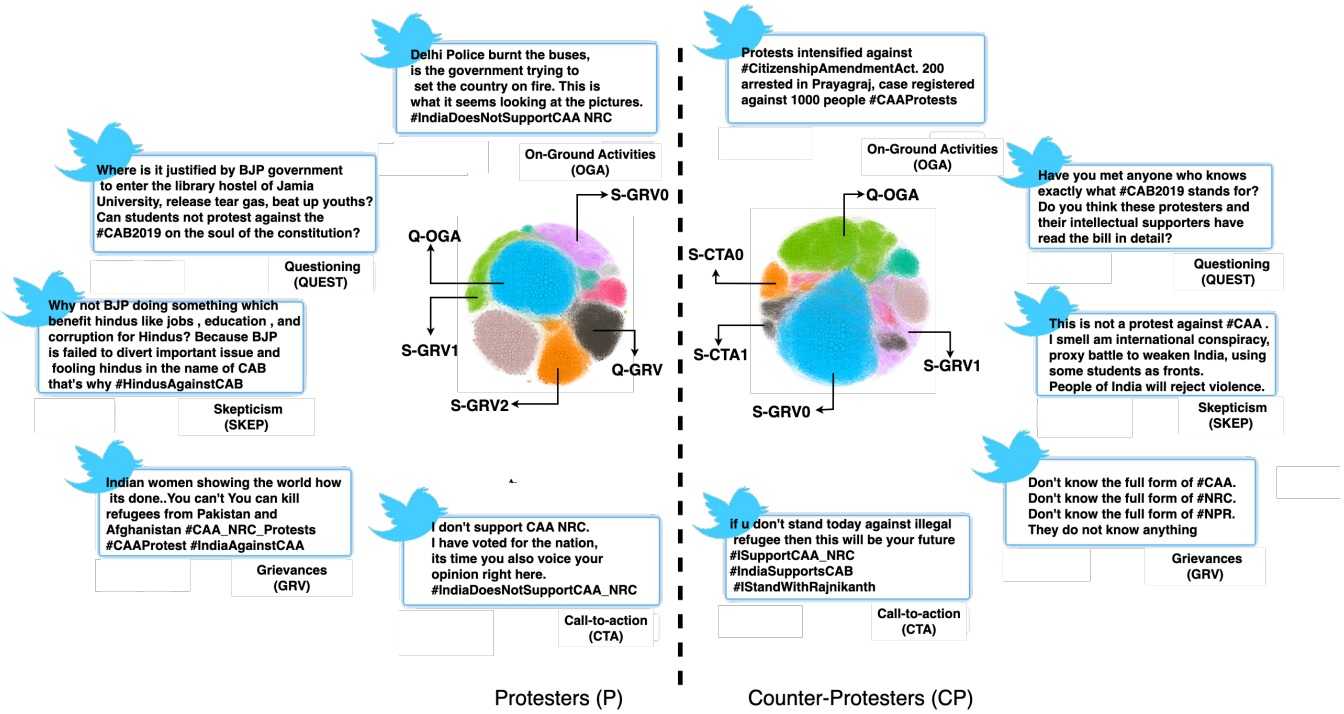
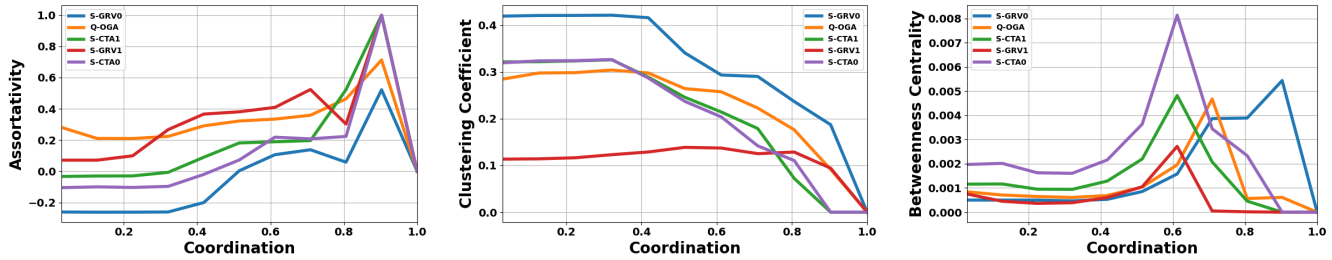
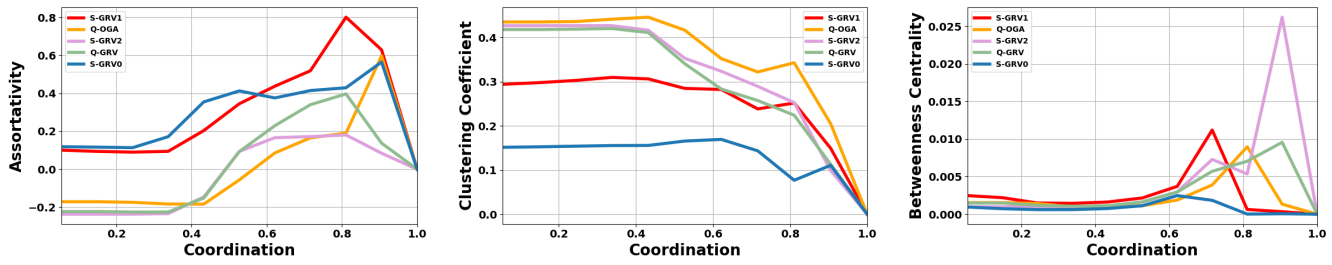


Figure 2: Communities obtained on the user-similarity network from mention metric for *superspreaders*. A total of 9 communities were formed in P, and 8 communities were formed in CP. Narrative labels are written for the top 5 communities in opposing stances, with P narratives on the left side and CP narratives written on the right side.



(a) Assortivity: CP (b) Clustering coefficient: CP (c) Betweeness centrality: CP

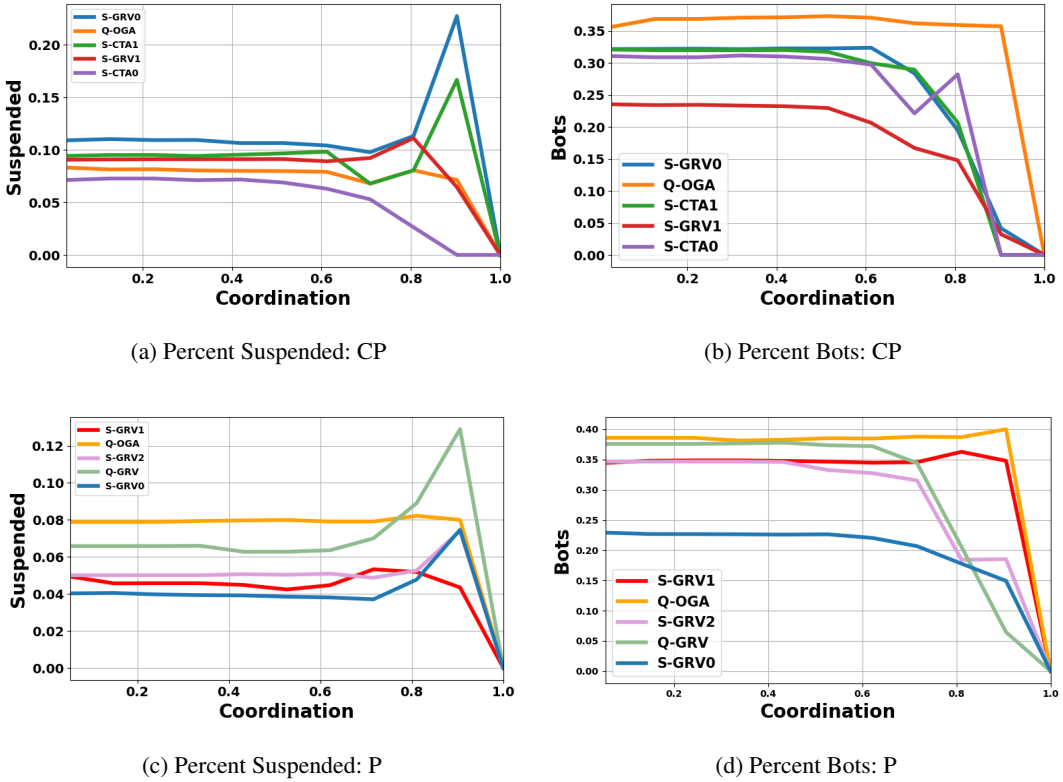


(d) Assortivity: P (e) Clustering: P (f) Betweeness centrality: P

Figure 3: Figure showing the relationship between computed network measures for each coordinated community as a function of the extent of coordination. Threshold and retained *superspreader* percentage in (i) CP: 0.7 - 50%, 0.8 - 40%, 0.9 - 1.5% and (ii) P: 0.7 - 69%, 0.8 - 28%, 0.9 - 6.3%.

	Name	Users	SKEP	QUE	GRV	CTA	OGA
			P				
P-1	S-GRV0	37.03%	1826	159	1252	203	154
P-2	S-GRV1	13.09%	430	318	317	72	260
P-3	Q-OGA	10.41%	266	290	202	44	239
P-4	Q-GRV	9.06%	460	24	322	33	23
P-5	S-GRV2	8.91%	437	40	282	43	48
			CP				
CP-1	S-GRV1	36.08%	1,963	684	1,209	895	318
CP-2	Q-OGA	19.67%	392	1,062	450	442	513
CP-3	S-GRV0	12.47%	623	292	362	354	153
CP-4	S-CTA1	11%	776	39	327	368	42
CP-5	S-CTA0	6.28%	266	195	164	206	164

Table 3: Distribution of the different narratives in the P and CP coordinated communities.



(a) Percent Suspended: CP

(b) Percent Bots: CP

(c) Percent Suspended: P

(d) Percent Bots: P

Figure 4: Figure showing the relationship between the percent suspended users, percent bots, and mean hate for each coordinated community as a function of the extent of coordination for the opposing stances.

work structure (Freeman et al. 2002). Figure 3 shows the coordination threshold versus the network measures to analyze how the network structure changes from least to strongest coordinated users in the community. For each coordination threshold, the community retains all the edges whose weight is equal to or more than the threshold. The findings from the plot are as follows:

Figure 3(a) and Figure 3(d) show the assortativity of the coordinated communities for CP and P *superspreaders* respectively. Assortativity measures the tendency of nodes to be connected to similar nodes in the network. The communities appear strongly assortative for both CP and P, especially as we move towards a higher coordination extent. This shows that the users were connected with similar users, forming a clique of coordinated users. Among the communities in CP, elevated assortativity towards a higher coordination extent was found for S-CTA1 and S-CTA0, while in P, it was S-GRV1 and S-GRV0. Hence, *skepticism* and *call-to-action* were majorly tweeted by most coordinated assortative nodes in CP. While in P, *skepticism* and *grievances* were majorly tweeted by most coordinated assortative nodes.

Figure 3(b) and Figure 3(e) show the clustering coeffi-

cient vs. extent of coordination for CP and P communities, respectively. The clustering coefficient measures the triads in the network and is one of the ways to capture the tightness of the network Both CP and P communities show organization and clustering in the network, given by the decreasing trend of the communities toward higher coordination. In CP, the highest clustering is exerted by S-GRV0, and Q-OGA respectively, whereas for P they are Q-OGA and S-GRV2, respectively. This result indicated that the community most clustered in P discussed questioning and on-ground activity, while most clustered CP users discussed skepticism and grievances tweets.

Figure 3(c) and Figure 3 (f) show the average betweenness centrality vs. the extent of coordination of the CP and P communities, respectively. Betweenness centrality measures how much influence a particular node has on the flow of information in the graph. We witness that towards the greater extent of coordination, the average betweenness centrality of network shows a falling trend for both CP and P communities. For CP, the community with higher average betweenness centrality towards stronger coordination was S-GRV0 (coordination $\simeq 0.9$), while one exceptional average betweenness centrality community for **P** was S-GRV2 (coordination $\simeq 0.9$). This indicated a set of highly coordinated users who participated in the flow of information shared *skepticism* and *grievances* narratives in both CP and P.

Our findings suggest that S-GRV0 in CP and S-GRV2 in P are the two communities with high clustering coefficients and betweenness centrality. This indicates a balance between local community cohesion and influential connectors bridging different groups, shedding light on the network's resilience between local cohesion and global influence, facilitating a continuous flow of information around the narrative. S-CTA1 in CP and S-GRV1 in P are the most assortative coordinated communities, indicating that *superspreaders* in these communities were tightly interconnected and similar to each other.

Inauthentic Behavior in Coordinated Communities

Although protests are inherently coordinated in nature (Starbird and Palen 2012), we investigate increased coordination in the user's tweeting behavior in relation to inauthentic behavior during protests. In this section, we discuss the presence of inauthentic users in the coordinated communities and whether specific coordinated communities showed higher inauthentic behavior. We also analyze the narratives shared by the highest inauthentic coordinated community to gauge whether coordinated communities with high inauthentic behavior preferred to share certain narratives over others. We define Inauthentic users as those who were either suspended by Twitter [4] or were identified as bot accounts using Botometer API (Yang et al. 2020). Due to rate-limit, we randomly sample 50% *superspreaders* from each opposing stance (4,262 users from **CP** and 3,463 users from **P**) and used the universal scores value ≥ 0.7 to label a user as bot. We plot the extent of coordination vs. the percentage suspended and bots, respectively for the sampled set of P and CP *superspreaders* in Figure 4.

Figure 4(a),(c) shows the percentage of suspended users present at different levels of coordination in the CP and P communities, respectively. In CP, the suspended users show the most strongly coordinated behavior for S-GRV0 and S-CTA1, indicating clique formation of inauthentic actors (coordination value $\simeq 0.9$). In P, the suspended users show the most strongly coordinated behavior for Q-GRV, while S-GRV0 also exerted an exceptionally increased percentage of suspended users on higher coordination threshold value. In summary, one of the narrative communities with skepticism and grievances in CP was inauthentic, while in P, they were questioning and grievances.

Next, we analyze the presence of bot activity in the coordinated communities. In CP, out of 4,262 users, 2,630 were identified as bot accounts, while 2,153 out of 3,463 P users were identified as bots. Figure 4(b),(d) shows the percentage of bots present at different levels of coordination in the CP and P communities, respectively. For both CP and P, we found that Q-OGA exerted the strongest coordination among the bots, evidenced by the plateaux structure, with a percent bot showing indifference to the changing coordination threshold. Communities that showed higher coordination among suspended users, i.e., S-GRV0 in CP and Q-OGA in P also showed high bot activity, indicative of the inauthenticity of the communities.

In summary, S-GRV0 community in CP (with 12.47% *superspreaders*), which showed a high clustering coefficient and high betweenness centrality, also showed high bot and suspended users, indicating inauthentic user involvement in the network. Whereas the Q-OGA community in P (with 10.41% *superspreaders*), which showed a high clustering coefficient, also showed high bot activity, indicating inauthentic user involvement in sharing the narrative.

Conclusion

We conducted the first combined analysis of narratives present in coordinated communities during online discourse, on the granularity of user's stance. Specifically, we applied our methodology to Twitter dataset of discourse around Citizenship Amendment Act, 2019 in India. We follow an unsupervised method to identify user's stance on CAA. We identify coordinated communities in the opposing stances, through a network-based approach on the various metrics used for coordination. Next, we apply an unsupervised narrative detection technique to identify user narratives, followed by analysis of coordinated communities. Our analysis reveals (i) user's mention in tweets led to the strongest coordinated network in CAA, (i) skepticism and questioning narratives were the two most dominant narratives across the opposing stances, (ii) call-to-action narrative, was more dominant narrative in counter-protesters, i.e., users who supported CAA, (iii) most assortative coordinated community in counter-protesters discussed skepticism with call-to-action narrative, while in protesters (users who opposed CAA) shared skepticism with grievances narrative, (iv) we

[4] https://help.twitter.com/en/managing-your-account/suspended-twitter-accounts

identify inauthentic coordinated communities in the opposing stances, were the narrative focus of inauthentic community in counter-protesters was skepticism with grievances, while in protesters was questioning and on-ground activities. Our findings also suggest the coordinated communities, which were highly inauthentic, showed the highest clustering coefficient towards a higher extent of coordination.

Limitations

Our study of CAA protests is subject to several important considerations and limitations. Firstly, the reliance on hashtags for data collection may not provide a comprehensive view of the entire discussion surrounding CAA. While hashtags are valuable for organizing and categorizing content, not all relevant discussions may use the designated hashtags. Users may employ alternative keywords or phrases to engage in conversations about CAA, potentially leading to an incomplete dataset. Secondly, the reliance on a single social media platform, such as Twitter, and public APIs for data collection pose constraints. Different social media platforms have distinct user demographics, cultures, and engagement behaviors. Relying solely on one platform may introduce selection bias and limit the generalizability of our findings. Furthermore, the accessibility and features of public APIs can change over time, potentially affecting data collection efforts.

Acknowledgements

This research is supported by EU H2020 program under the SoBigData++ project (grant agreement No. 871042), by the CHIST-ERA grant No. CHIST-ERA-19-XAI-010 (ETAg grant No. SLTAT21096), and partially funded by CHIST-ERA project HAMISON.

References

Artstein, R.; and Poesio, M. 2008. Inter-coder agreement for computational linguistics. *Computational linguistics*, 34(4): 555–596.

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).

Brannen, S.; Haig, C.; and Schmidt, K. 2020. The age of mass protests: understanding an escalating global trend.

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.

Fiske, S. T. 2022. Twitter manipulates your feed: Ethical considerations.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Freeman, L. C.; et al. 2002. Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge*, 1: 238–263.

Gallagher, R. J.; Reagan, A. J.; Danforth, C. M.; and Dodds, P. S. 2018. Divergent discourse between protests and counter-protests: #BlackLivesMatter and #AllLivesMatter. *PLOS ONE*, 13(4): 1–23.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Goel, R.; and Sharma, R. 2020. Understanding the metoo movement through the lens of the twitter. In *Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6–9, 2020, Proceedings 12*, 67–80. Springer.

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): 1–7.

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.

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, 191–201. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391917.

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).

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.

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

McInnes, L.; Healy, J.; and Melville, J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Moulavi, D.; Jaskowiak, P. A.; Campello, R. J.; Zimek, A.; and Sander, J. 2014. Density-based clustering validation. In *Proceedings of the 2014 SIAM international conference on data mining*, 839–847. SIAM.

Neha, K.; Agrawal, V.; Buduru, A. B.; and Kumaraguru, P. 2022a. The Pursuit of Being Heard: An Unsupervised Approach to Narrative Detection in Online Protest. In *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 256–260.

Neha, K.; Agrawal, V.; Kumar, V.; Mohan, T.; Chopra, A.; Buduru, A. B.; Sharma, R.; and Kumaraguru, P. 2022b. A tale of two sides: Study of protesters and counter-protesters on# citizenshipamendmentact campaign on twitter. In *Proceedings of the 14th ACM Web Science Conference 2022*, 279–289.

Ng, L. H. X.; and Carley, K. M. 2022. Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In *14th ACM Web Science Conference 2022*, 12–21.

Nizzoli, L.; Tardelli, S.; Avvenuti, M.; Cresci, S.; and Tesconi, M. 2021. Coordinated behavior on social media

in 2019 UK general election. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 443–454.

Pacheco, D.; Flammini, A.; and Menczer, F. 2020. Unveiling Coordinated Groups behind White Helmets Disinformation. In *The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020*, 611–616. Association for Computing Machinery. ISBN 9781450370240.

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, 455–466.

Ranade, P.; Dey, S.; Joshi, A.; and Finin, T. 2022. Computational Understanding of Narratives: A Survey.

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, 537–548.

Rogers, A.; Kovaleva, O.; and Rumshisky, A. 2019. Calls to Action on Social Media: Potential for Censorship and Social Impact. *EMNLP-IJCNLP 2019*, 36.

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*, 1441–1451.

Sinpeng, A. 2021. Hashtag activism: social media and the #FreeYouth protests in Thailand. *Critical Asian Studies*, 53(2): 192–205.

Slobozhan, I.; Brik, T.; and Sharma, R. 2022. Longitudinal change in language behaviour during protests: a case study of Euromaidan in Ukraine. *Social Network Analysis and Mining*, 12(1): 107.

Slobozhan, I.; Brik, T.; and Sharma, R. 2023. Differentiable characteristics of Telegram mediums during protests in Belarus 2020. *Social Network Analysis and Mining*, 13(1): 19.

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, 7–16. New York, NY, USA: Association for Computing Machinery. ISBN 9781450310864.

Theocharis, Y.; Lowe, W.; Van Deth, J. W.; and García-Albacete, G. 2015. Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, 18(2): 202–220.

Vargas, L.; Emami, P.; and Traynor, P. 2020. On the detection of disinformation campaign activity with network analysis. In *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop*, 133–146.

Varol, O.; Ferrara, E.; Ogan, C. L.; Menczer, F.; and Flammini, A. 2014. Evolution of online user behavior during a social upheaval. In *Proceedings of the 2014 ACM conference on Web science*, 81–90.

Wang, R.; and Zhou, A. 2021. Hashtag activism and connective action: A case study of #HongKongPoliceBrutality. *Telematics and Informatics*, 61.

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

Wei, K.; Lin, Y.-R.; and Yan, M. 2020. 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, 2443–2454. New York, NY, USA: Association for Computing Machinery. ISBN 9781450370233.

Wu, H. H.; Gallagher, R. J.; Alshaabi, T.; Adams, J. L.; Minot, J. R.; Arnold, M. V.; Welles, B. F.; Harp, R.; Dodds, P. S.; and Danforth, C. M. 2023. Say their names: Resurgence in the collective attention toward Black victims of fatal police violence following the death of George Floyd. *PLOS ONE*, 18(1): 1–26.

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.

Ethics checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, this work follows the required principles.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes all contributions listed in the abstract and introduction accurately reflect the paper's contribution. See Introduction Section.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, we explain the aims and how our proposed methodological tools helps us reach those aims.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, we clarify that our data is limited to English.
   (e) Did you describe the limitations of your work? Yes.
   (f) Did you discuss any potential negative societal impacts of your work? We do not foresee any potential negative societal impacts of this work.
   (g) Did you discuss any potential misuse of your work? We do not foresee any potential misuse of this work as we will be releasing the deanonymised version of the dataset.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Not Applicable.

(i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, we have read the ethics review guidelines and believe our paper conforms to them.

2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Not applicable as we were conducting an exploratory work.
   (b) Have you provided justifications for all theoretical results? Not applicable as we were conducting an exploratory work.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Not applicable as we were conducting an exploratory work.
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Not applicable as we were conducting an exploratory work.
   (e) Did you address potential biases or limitations in your theoretical framework? Not applicable as we were conducting an exploratory work.
   (f) Have you related your theoretical results to the existing literature in social science? Not applicable as we were conducting an exploratory work.
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Not applicable as we were conducting an exploratory work.
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? Not applicable as we did not have any theoretical proofs.
   (b) Did you include complete proofs of all theoretical results? Not applicable as we did not have any theoretical proofs.
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Anonymised dataset, Code instructions will be released after the double blind review.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? We are using existing tools and techniques.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Not Applicable.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No, as we were able to do our anlaysis and experiments on a laptop CPU could handle it, we did not see the need to specify these details.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Not Applicable.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? Yes.
   (b) Did you mention the license of the assets? Not Applicable.
   (c) Did you include any new assets in the supplemental material or as a URL? Yes. Please see the ethical consideration in the Introduction section for more details.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Not Applicable.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Dataset is anonymised before conducting any experiments.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11. (2020.))? Not applicable as we did not release a new dataset.
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? Not applicable as we did not release a new dataset.
6. Additionally, if you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots? Not applicable as we did not directly conduct research with human subjects
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Not applicable as we did not directly conduct research with human subjects
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Not applicable as we did not directly conduct research with human subjects
   (d) Did you discuss how data is stored, shared, and deidentified? Not applicable as we did not directly conduct research with human subjects