Dissecting the Advocacy Discourse Behind the #StopAsianHate Movement on X/Twitter

Yuze Sha
Dept. of Linguistics & English Language
Lancaster University
Lancaster, UK
y.sha2@lancaster.ac.uk

Nicholas Micallef
Dept. of Computer Science
Swansea University
Swansea, Wales, UK
nicholas.micallef@swansea.ac.uk

Yan Wu
Dept. of Literature, Media & Language
Swansea University
Swansea, Wales, UK
y.wu@swansea.ac.uk

Abstract*—This paper takes a multidisciplinary approach and conducts a three-dimensional analysis of #StopAsianHate tweets from January 1, 2021, to December 31, 2022 by combining computer science, applied linguistics and cultural studies. Employing a 'funnel approach', this paper focuses progressively from a broad examination to specific sentimental and linguistic dimensions within the top 10% most engaged tweets. The analysis reveals that the #StopAsianHate hashtag is primarily used for counter-discourse against Anti-Asian hate crime, expressing collective ingroup indentity and inclusionary outgroup solidarity against racism. A key finding is the cultural study of the representation of Asian people as the 'model minority', derived from combined analyses of sentiments, politeness, toxicity, and Corpus-Assisted Critical Discourse Analysis of the tweets. The #StopAsianHate movement is characterised as moderate, evidenced by the large number of tweets with positive sentiment scores and frequent relational identification, which refers to anti-racism supporters as 'friends', 'folks', and 'family'. Though negative sentiment scores are also prevalent, they are found non-toxic and can be explained by tweet genre's rare use of polite expressions, as well as the prominence of #StopAsianHate thematic words such as 'hate', 'racism', and 'crime', serving not as negative sentiments but as tools to challenge racism. Additionally, the moderate tone is further highlighted through the discourse's strategic avoidance of direct condemnation of individuals, opting instead to highlight violent events like shootings and killings without directly attacking the perpetrators. The analysis also highlights a bottom-up self-positioning of Asian individuals within the movement as lacking power and being vulnerable, rather than being proactive campaigners. Nonetheless, the study provides fresh insights into the growing self-reflective collective awareness of the negative impacts of 'model minority' stereotypes within the Asian communities and discusses ongoing opportunities and challenges in #StopAsianHate movement.

Keywords*— #StopAsianHate, model minority, social media discourse, sentiment analysis, corpus-assisted critical discourse analysis.

I. Introduction

The utilization of X/Twitter as a political tool has become increasingly prevalent among various societal actors, including social elites, advocacy groups, and grassroots activists, for advancing political agendas. Notably, in recent years, there has been a significant surge in the use of X/Twitter to advocate social justice concerning gender, sexuality, and race related inequalities. While existing research is predominantly grounded in computational techniques and data mining, it often overlooks the importance of qualitative appraisal, particularly concerning the discursive nuances embedded within Tweets through cultural studies lens.

Taking a triangulation approach, we use both quantitative and qualitative methods in analysing X/Twitter activism via a case study of #StopAsianHate tweets during the COVID-19 global pandemic. This study features three variations of knowledge production that extend across computer science, applied linguistics, and cultural studies while combining methods from Natural Language Processing (NLP) and Corpus-Assisted Critical Discourse Analysis (CACDA) to explore the discursive features of the #StopAsianHate movement on X/Twitter.

While NLP facilitates the processing of large-scale data, traditionally utilising sentiment analysis (e.g., [1]) and topic modelling (e.g., [2]) to reveal general lexicon and semantic patterns, CACDA offers a framework for integrating quantitative and qualitative analyses at textual level and interprets discursive features within social contexts. By integrating NLP with CACDA, our study not only explores anti-Asian racism during COVID-19 global pandemic as an individual phenomenon, but also investigate how such discourse is supported by a dominant social structure that might be expressed in overt or covert terms.

Employing a 'funnel approach', this paper follows a data-inductive strategy, moving from broad patterns to specific features, thereby providing a comprehensive and in-depth analysis of the general characteristic, sentimental landscapes, and discourse practice associated with #StopAsianHate discourse on X/Twitter. Such an approach allows us to provide a comprehensive analysis of the discursive essence of the movement and its implications for and limits to causing social changes. Through the lens of cultural racism, we also investigate how anti-Asian hate online reflects the problematic 'model minority' ideologies and how such ideologies shape the counter discourse to anti-Asian racism.

II. Related Work

In this section, we firstly assess the concept of 'model minority', a particular form of cultural racism against Asian communities in the US. We then review related work on the technological affordances of X/Twitter in mobilizing people for racial justice. Both the Black Lives Matter (BLM) movement and the #StopAsianHate movement utilize hashtags and real-time engagement to raise awareness, challenge stereotypes, and advocate for social justice. Identifying current research gaps, we propose a multidisciplinary approach integrating natural language processing (NLP) and computational analysis of communication dynamics (CACDA) to analyse the nascent form of digital resistance to anti-Asian racism.

A. *'Model Minority' and 'Medical Scapegoating' Racism*

'Model Minority' as collective identity label has been attached to ethnic minority communities with Chinese, and other East Asian (such as Japanese, Korean) and sometimes other Southeast Asian heritages in the US. This term was firstly used by American press in the 1960s to laud the academic and economical success of such minority communities and their propensity of not causing "problems" [3], [4], [5]. Despite its seemingly positive connotation, the model minority stereotypes usually imply such community's lack of visibility and subordination in the society where 'the model minority has become integrated, modernised, and civilized' [6]. Model minority stereotype promotes the myth of the US as a colourblind post-racial society and become a hegemonic discourse to silence Asian Americans from voicing their grievance and isolate them from standing in solidarity with other oppressed minority communities in striving for racial equality and social justice. As Frank Wu argues, 'Asian Americans have been excluded by the very terms used to conceptualise race. People speak of "Americans" as if it means "white", and "minority" as if it means "black.". In that semantic formula, Asian Americans, neither black nor white, consequently are neither American nor minority' [7].

A particular form of anti-Asian racism is what Trauner (1978) terms 'medical scapegoating', which descripts how the authorities blamed the ethnic Chinese during the public health crisis in the late 19th Century to early 20th Century to cover up their health governance failure [8]. The 'medical scapegoating' strategies have been similarly adopted during the COVID-19 global pandemic, which spurred a wave of Sinophobic violence in the US and worldwide. Asian descent worldwide had been scapegoated for the public health crisis and suffered from heightened hostility, exacerbated stereotyping, verbal harassment, physical attacks, and online hatred ranging from blatant racist postings to racism memes [9], affecting both their safety and mental wellbeing [10]. Meanwhile, politicians and media played significant roles in activating and intensifying anti-Asian hatred [11]. In the US, 'rhetorical labelling of the virus as "Chinese virus," "Wuhan virus," and "Kung flu" by influential and prominent individuals sharing their racism openly via both traditional media and social media platforms to an audience of millions strongly suggest[ed] that the virus' harms and dangers emanate from a specific community and location' [12].

B. Social Media for Racial Justice: from #BLM to #StopAsianHate

Information Communication Technologies has forever revolutionised the way in which social movement is initiated, coordinated, organized, and escalated [13], [14]. In particular, social networking services (SNSs) such as X/Twitter, play a key role in digital social movement for social justice in recent years. The "affordances of technology" of SNSs which includes searchability; replicability; scalability, the ability to capture and archive content; and the broad range of dissemination (boyd p. 7 in [15]) make them ideal tool for social movement. X/Twitter can mobilise and widely engage users 'by incorporating the use of hashtags, or conversation streams to label the meanings they express' [16].

SNSs played a crucial role during the Black Lives Matter (BLM) movement [17]. 'Black Twitter' convenes public discourse centring around the black identity construction and racial inequities felt by the American black community [18], [19]. Through the use of Twitter for real-time engagement, users experience 'heightened temporality' that creates a sense of belonging and solidarity among Twitter users and movement participants, contributing to movement mobilization and escalation [17]. Meanwhile, Yang contends that the temporal unfolding of the postings under a common hashtagged word or phrase develops a narrative form and agency, creating a powerful sociopolitical agenda for the BLM movement [20]. Worth noticing is also the use of emotions and sentiment in postings in building a following on social media. Keib et al. (2018) observe that tweets that express emotion, or tweets include content about policy or action or social actor are more often retweeted than neutral tweets [21]. Tweets contain all these components, i.e. emotion content about a policy or action or established group, were more often retweeted.

The anti-Asian hatred during the COVID-19 global pandemic reached its peak on 16th March 2021 when 21-year-old Robert Aaron Long carried out a shooting spree at two spas and a massage parlour in Atlanta, Georgia. Eight people were killed and a ninth was wounded and out of the eight death, six were women of Asian descent. The shooting triggered a wave of protest against anti-Asian crime and racism across the country. #StopAsianHate, #StopAAPIHate, and other hashtags trended online [22], calling law enforcement officials to tackle hate crime and advocating for social justice for Asian communities. Such hashtags help to raise awareness, break stereotypes, and challenge racial injustice and discrimination [23]. A study of 46,058 Tweets featuring #StopAsianHate and #StopAAPIHate reveals that the movement attracts more participation from women, younger adults, Asian and Black communities. About half of the Twitter users show direct support, but 5.43% show a negative attitude towards the movement. The Black and White communities blame each other for the anti-Asian hate crimes [1]. In recent work, Wheeler et al. (2022) studied both negative (e.g., #kungflu) and positive (e.g., #stopAAPIhate) hashtags and keywords related to anti-Asian prejudice [24]. Through descriptive analyses, they observed differences in the frequency of negative and positive keywords based on geographic location [24]. Using burst detection, they identified distinct increases in negative and positive content in response to key political tweets and events [24].

C. Research Gaps

It is evident that by delivering real-time streaming communication, X/Twitter is instrumental in activating, mobilising, escalating social movement for social justice. However, there are still debates as to how this plays out in particular contexts, political practice and activist organizations (see [25], [26], [27], [28], [29], [30]). Comparing to #BLM, #StopAsianHate is a less 'impactful' digital social movement. We therefore aim to examine the assembly of technical system in the specific contexts which led to its differentiating forces and outcomes. We are particularly interested in investigating the linguistic strategies and rhetorical devices used by #StopAsianHate activists on X/Twitter and consider in what ways, the sentiment, politeness, and toxicity features within #StopAsianHate Tweets facilitate or constrain the computational expression of the advocacy for social justice.

Methodologically, previous NLP studies examining sentiments in the #StopAsianHate movement often focused narrowly on a singular sentimental perspective [1], [24], which can result in contextual blindness and oversimplification. This limitation is particularly significant in social movement datasets, where the presence of extreme lexicons such as *hate* and *racism* may lead to false positives. Additionally, these studies typically concentrate solely on statistical patterns, neglecting the contextualisation of these features or overlooking the deeper, specific patterns of interaction that are crucial for understanding social movements. To bridge such research gaps, this study integrates NLP and CACDA methods, takes a multidisciplinary approach and offers a comprehensive analysis that connects statistical data with broader discourse dynamics in the study of digital resistance to a particular form of cultural racism.

III. CORPUS CONSTRUCTION AND ANALYTICAL PROCEDURES

The research design of this study is influenced by Fairclough's theoretical position of treating language as discourse and social practice. Employing a multidisciplinary approach, we utilized a three-dimensional strategy for data collection and corpus construction.

A. Data Collection and Corpus Construction

We collected data on #StopAsianHate, which was the most popular hashtag used to advocate racial equality and social justice for Asian communities on social media platforms [2], [31], [32], [33]. Similar to previous research that collected Tweets [34], [35], [36], we utilized a keyword-based method to collect tweets that mention the #StopAsianHate term. Since the streaming API has data collection limitations [37], we used the X Search API [38] to retrieve and collect data from January 1st, 2021 to December 31st, 2022. This method returned 721,531 tweets.

We adopt Fairclough (2013)'s framework of analysis of language texts, analysis of discourse practice and analysis of discursive events in data selection. The three analytical dimensions each required distinct datasets. After outlining the general characteristics in the first dimension, the second dimension (i.e., sentimental landscapes) analysed the English-language tweets, yielding a subset of 424,346 tweets. We further isolated the 10% most popular tweets (42,435 in total) according to retweets and favourites to compare features between these and the general dataset. Finally, the third dimension (discourse practices) focused on these popular English-language tweets for a more detailed analysis. Each tweet was manually cleaned, and its multimodal textual content was extracted. They were then uploaded to Sketch Engine, an online platform for corpus construction and analysis [39]. The finalised corpus comprises 591,583 tokens.

B. Analytical Framework and Procedures

Taking Fairclough's (2013) [40] perspective that views language as both discourse and social practice, we employed a three-dimensional approach to analyse the #StopAsianHate tweets, starting from broader perspectives and narrowing down to specific linguistic details.

The first dimension focuses on dataset characteristics. Specific focus is given to mapping the distribution of languages and geographical locations represented in the #StopAsianHate tweets. In addition, we calculate how many tweets share images, videos, and links to YouTube videos and the level of engagement (in terms of retweets and favourites) gathered by the tweets.

In the second layer, we used Natural Language Processing (NLP) methods to examine the sentimental landscapes in the dataset, analysing sentiment, politeness, and toxicity scores. We use state-of-the art classifiers which have been widely adopted by previous research [35], [41]. For sentiment study, we use the widely deployed SentimentIntensityAnalyzer python module, which generates a positive score to represent positive valence and a negative score to represent negative valence [42]. For politeness study, we use ConvoKit python module to calculate the probability that tweets are polite based on 'lexical and syntactic features operationalizing key components of politeness theory, such as indirection, deference, impersonalisation and modality' [43]. Higher scores indicate polite tweets, while scores close to 0 indicate impolite discourse. To measure toxicity, we use the detoxify python module [44] that is also a widely used classifier to predict harmful content within discourse [45]. Scores close to 1 indicate toxic discourse, while scores close to 0 indicate non-toxic discourse. Moreover, our analysis uses He et al. (2021) classifier to determine the percentage of tweets that can be considered hateful, counterspeech, or neutral [46].

In the last section, we narrowed our focus to the discourse practices of the 10% most popular English-language tweets. Utilising Sketch Engine [39], we began with Keyword Analysis to identify words that occur more frequently than expected by chance [47]. The reference corpus used for generating the keyword list was the English Web 2021 (enTenTen21), an English-language corpus of 52 billion words collected from the Internet. We semantically categorised the top 20 keywords, using extended concordances to delineate the discourses [48] present in the focused dataset.

IV. FINDINGS

Our findings reveal dataset characteristics, sentimental landscapes, and an investigation of the linguistic features and discourses in the 10% most popular English-language tweets in the dataset. Overall, the #StopAsianHate movement is characterised with positivity, politeness, a moderate tone, few toxic expressions, and an avoidance of condeming individual perpetrators of anti-Asian racism.

A. Dataset Characteristics

To determine locations, we extract the location configured by users in their X/Twitter profiles. Since this information is self-reported, we cannot verify the authenticity of this information and only analyse instances where such information is provided. It is evident that the technological affordance of X/Twitter enables the global reach of #StopAsianHate movement, and the use of hashtags enabled the focused public discussion about racially instigated hatred towards Chinese and Asian communities worldwide. However, the density of tweets containing this hashtag is biased towards the US. Most tweets (16%) are originated in the US while global cosmopolitan cities such as Los Angeles (1.2%), New York (1.1%), San Francisco (1%), Toronto (<1%), London (<1%) etc. that feature sizable Asian communities also are the major places where such virtual forms of public debate take place (see Figure 1).

As for the languages of the tweets we use the language attribute provided by the API and find that most tweets are in English (59%). Other popular languages of the rest of the tweets include Asian languages (Traditional Chinese - 12%, Korean – 2%, Thai -1%, Japanese 1%) and Latin languages (Spanish – 6%, Portuguese – 2%). The rest of the tweets have either an undetected language or a less popular language (<1%).

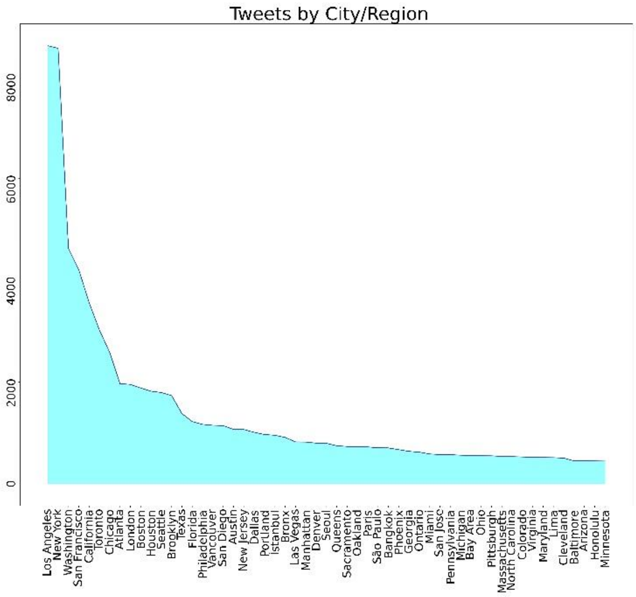

Fig. 1. Most popular cities/regions in our dataset (claimed by tweeters)

We calculated the number of tweets that contain an image, video, or link to a YouTube video. Interestingly, the most popular tweets are twice as much more likely to link to multimodal content (*All tweets* with images: 6%; *Most popular tweets* with images: 15%; *All tweets* with videos: 0.7%; *Most popuar tweets* with videos: 2.3%), which could be one of the reasons why these tweets are more popular. Surprisingly, a low proportion of tweets (less than 1%) share links to YouTube videos, which indicates that sharing YouTube video is not a common practice within the Asian community. In this case there is no difference between *All tweets* (1%) and the *Most popular tweets* (0.9%).

Most tweets do not recieve much engagement, with 77% of tweets not being retweeted and 58% of tweets not receiving any favourites. A considerable number of tweets receives moderate engagement, i.e., between 1 and 100 retweets or favourites (retweets: 22%, favourites: 41%), with only a very small percentage of tweets being widely diffused across the platform, i.e., receiving more than 100 retweets or favourites (retweets:0.63%, favourites: 1.42%). This finding shows that these tweets are receiving some traction within the platform, although they might not be receiving the widespread diffusion gathered by other social movements [19].

Moreover, 27% of the users that used the #StopAsianHate hashtag have more than 1,000 followers with 44% having between 100 and 1,000 followers. This finding shows that although these tweets do not seem to receive much traction, they might still have been viewed by a considerable audience. Most users (98%) only posted between 1 to 10 times with accounts dedicated to addressing the racism topic posting several times (RespondToRacism having 2,789 posts and ActToChange having 294). This finding suggests that, with a few exceptions, posting about #StopAsianHate was a sporadic collective action.

B. Sentimental Landscapes

We conduct this analysis using only English tweets, since the classifiers used are most effective with English corpora. *All tweets* refer to all English tweets (424,346), while *Most popular tweets* refer to English tweets that received most retweets and favourites (42,435). The classifier developed by He et al. (2021) [46] found that 99% of our dataset is counterspeech discourse.

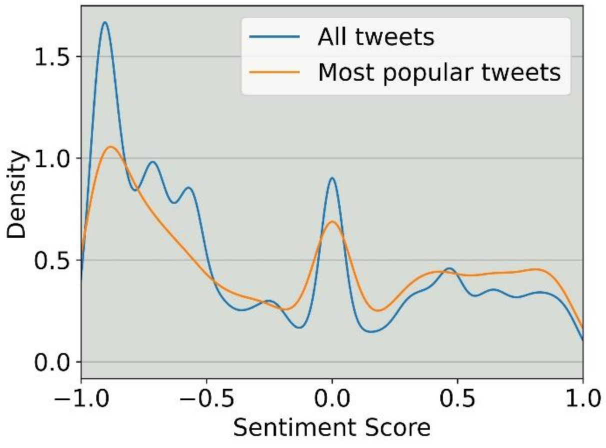

Fig. 2. Density distribution of Sentiment across our dataset.

TABLE 1. SENTIMENT, POLITENESS, AND TOXICITY WITHIN TWEETS

Tweet	Type
@XYZ Thank you Senator XYZ for standing up for the AAPI community and for fighting against hate. Please continue to spread the word in <location> (ps, thanks also for giving work opportunities to students in your office -awesome staff	Positive Sentiment
Hate needs to stop now Be kind to one another let's make the world a better place #stopasianhate <link>	Neutral Sentiment
😥 It's not just that Asian's are being attacked- it's elders! Fuck is wrong with people? #StopAsianHate <link>	Negative Sentiment
@XYZ @XYZ Great coverage of today's event and cute green dress thanks for sharing people's stories #StopAsianHate	Polite
I spoke with CNN to keep this important conversat going. #StopAsianHate #StopAAPIHate <link>	Moderately Polite
WOW WPs behaving badly. Really wild stuff downtown. A Stop Asian Hate rally is clashing with a Pro-Uighur drive by. The pro-Uighur group is shouting "F— China!" The Asian rally is responding by calling them "racist." <link>	Impolite
so it's #StopAsianHate.how about we make a change to that, just stop being a hate-filled fucker to others? ...no matter who they damn are. Sure I am an asshole and disagree with others but I at least do not hate nor disrespect ppl for petty shit such as skin color.	Toxic
Learn more from @XY in an hour than a lifetime in the US school system #StopAsianHate <link>	Moderately Toxic
and it's not just her story that makes me upset, it's the THOUSANDS of people sharing their stories. #StopAsianHate	Non-toxic

Figure 2 shows the density distribution of the tweet sentiment from negative (-1), to positive (1), with neutral (0) in the middle (*All tweets*: mean=-0.27, standard deviation= 0.60; *Most popular tweets*: mean=-0.16, standard deviation= 0.62). For both datasets the negative sentiment is the most prevalent (*All tweet*s: 66.4%, *Most popular tweets*: 53.4%), which could be explained by the negative terms, such as "hate", "racist", and "protest" that are generally used in such discourse (see row 3 in Table 1). There are still a considerable number of tweets that use positive terms (*All tweets*: 24.3%, *Most popular tweets*: 35.8%), meaning that despite the negative terms inherent to this topic people still find ways to spread positive messages to their followers (see row 1 in Table 1). The rest of the tweets are neutral (*All tweets*: 9.3%, *Most popular tweets*: 10.8%). Refer to Table 1 for examples of positive, neutral, and negative tweets. When comparing the two datasets, we find that the *Most popular tweets* have a significantly higher percentage of positive tweets. This finding could indicate that a positive sentiment might lead to more engagement, but also people/ organizations with more followers might be more careful in crafting positive messages. Building on Wheeler et al.'s (2022) analysis of sentiment in the #StopAsianHate dataset [24], our study addresses an important research gap, which has been previously unexplored, by delving deeper into the levels of politeness and toxicity within these tweets.

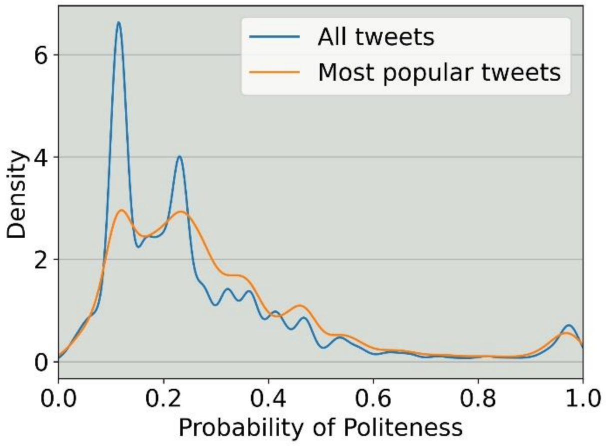

Fig. 3. Density distribution of Politeness across our dataset.

Figure 3 shows the density distribution of the probabilities of the tweets being polite (*All tweets*: mean=0.27, standard deviation=0.21; *Most popuar tweets*: mean=0.31, standard deviation=0.21). Most tweets are scored as impolite because they were classified with a score less than 0.4 (see row 6 in Table 1). This finding confirms the previously described finding that most tweets contain a negative sentiment. However, a good number of tweets contain some polite discourse (between 0.4 and 0.6) and only very few tweets are polite (scores more than 0.8). Refer to Table 1 for examples of impolite, neutral, and polite tweets. Our in-depth analysis of which terms are used by polite and impolite tweets revealed that the most common terms used in impolite tweets are: “discrimination”, “racial”, “racist”, “unprofessional”, “hate”, “condemn”, “stand”, “violence”, “stop”, “apologize”. Some of these terms are also common in neutral tweets, which in addition contained terms such as “support”, “together”, “respect”, “community”, “live”, “represent”, “comment”. Polite tweets consisted of the usual terms such as “racism”, “hate”, “stop”, “condemn”, but also have a prevalence of polite and positive terms, mainly: “thank”, “together”, “respected”, “support”, “voice”, “proud”. Moreover, although the distributions show similar patterns for *All tweets* and *Most popular tweets*, Figure 3 shows that *Most popular tweets* contain slightly more polite discourse. This finding could be explained by these tweets being written by professionals who manage accounts on social media platforms for organizations and public figures.

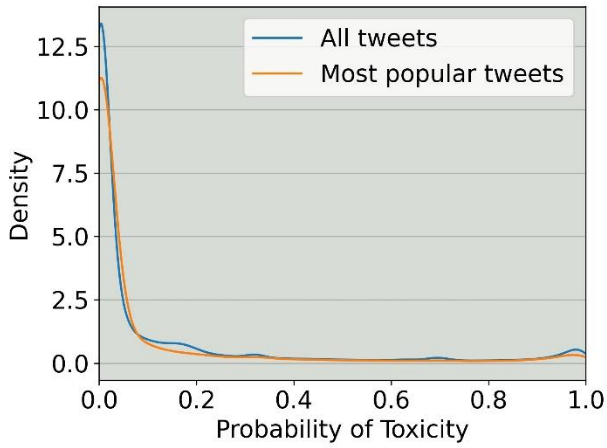

Fig. 4. Density distribution of Toxicity across our dataset.

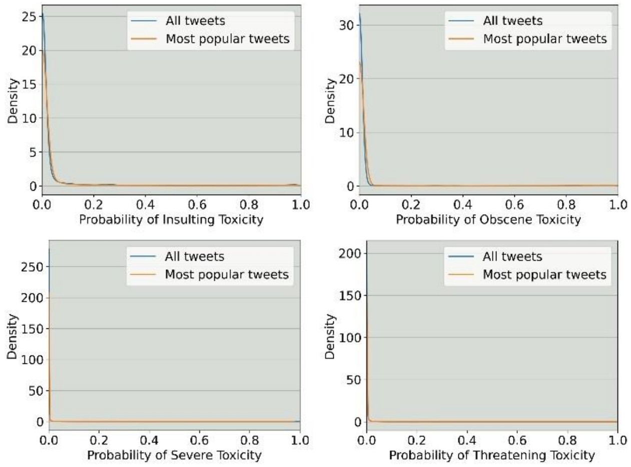

Fig. 5. Density distribution of Insulting, Obscene, Severe, and Threatening Toxicity.

Figure 4 shows the density distribution of the probabilities of the tweets being toxic, meaning having harmful terms (*All tweets*: mean=0.12, standard deviation= 0.25; *Most popular tweets*: mean=0.09, standard deviation=0.22). Interestingly, we find that most tweets are not toxic. Despite most tweets being negative and impolite, further analysis (see Figure 5) shows that in general the discourse around #StopAsianHate does not contain insulting, obscene, severe, or threatening terminology. Our in-depth analysis of which terms are used in toxic tweets revelead that profane terms are common in these tweets, e.g., “f**k”, “f**king”, “*ss”, “sh*t”, “wtf”. This finding shows that the discourse around this topic is mostly moderate and does not get out of control and radicalized. Refer to Table 1 for examples of toxic, moderate, and non-toxic tweets.

C. Textual Level

The textual level analysis focuses on the linguistic features and discourses in the 10% most popular English-language tweets in our dataset. Table 2 displays the top 20 keywords. Based on the keyword and concordance analyses, the top 10% most popular English-language tweets in our dataset can be interpreted from four perspectives: (1) counter-discourse against anti-Asian hate, (2) self-positioning and attack dynamics, (3) tweet dissemination, and (4) expressions of attitudes and sentiments.

TABLE 2. TOP 20 KEYWORDS ORGANISED BY SEMANTIC GROUP

No	Top 20 Keywords (Lemma)
(1)	(1) anti-asian, (2) aapi, (3) asians, (4) asian, (7/13) asian-american(s), (18) islander;
	(5) racism, (9) xenophobia, (11) hate;
	(12) xiaojie, (16) daoyou, (17) delaina, (20) yaun;
(2)	(10) supremacy;
(3)	(6) carrd; (8) bts, (15) rt, (19) tacha;
(4)	(14) solidariy;

TABLE 3. TWEET EXAMPLES

No	Examples
1	Racism is NOT an opinion.
	Stand against racism.
	We condemn racism and xenophobia in all forms.
2	He was one of the victims of violence against Asian people, especially the elderly ones.
3	....it's a reality now 😭😭 BTS PAVED THE WAY YOU HAVE OUR BACK SO PROUD OF YOU BTS BTS IN WHITE HOUSE BTS SOUTH KOREA'S PRIDE …
4	.. the racism and violence against asians (especially recently) is disgusting and i hate that i live in constant fear. please let us be heard

1) *Counter-Discourse against Anti-Asian Hate*

The first group of keywords is related to the counter-discourse against anti-Asian hate. These keywords frequently appear in the form of slogans within the tweets (see Example 1 in Table 3), articulating the demands of the #StopAsianHate movement.

Asians, particularly Asian Americans, are positioned as central social actors within the target corpus, as evidenced by the top keywords like anti-Asian, AAPI, Asian(s), and Asian-American(s). Notably, keywords such as AAPI and Asian-American(s) indicate a strong connection to the US context in these most popular English-language tweets.

The associated concordances highlight a strong sense of community and solidarity. AAPI community appeared 831 times in the dataset, with the singular form being much more prevalent than the plural, suggesting a unified group solidarity. The keyword solidarity was prominent, appearing 930 times, with support, love, stand in solidarity, and rally frequently collocating with it. Additionally, the sense of community is reinforced with relational identification strategies [49] in portraying these Asian social actors. They are referred to as friends, folks, family, and groups, further emphasising collective in-group identity and inclusionary outgroup solidarity.

The second focus within Group 1 is racism and hate crimes, highlighted by the keywords racism, hate and xonophobia. A notable incident emerged from the keywords is the Atlanta Shooting on March 16th, 2021, where six of the eight victims were women of Asian descent, classifying the event as a hate crime. Many among the top 10% most popular tweets commemorate the victims (e.g., *Xiaojie*, *Daoyu*, *Delaina* & *Yaun*).

2) *Self-positioning & attack*

A bottom-up self-positioning is prominent in the discourse, condemning the problem of anti-Asian racism being ignored for a long time and portraying an unequal power relationship in the discourse. In the concordances, Asian individuals are frequently depicted as the patient of actions (i.e., social actors being the recipient of / affected by the actions), with 848 instances of the phrase *against Asians* (see Example 2 in Table 3). In terms of how racists are represented, notable keywords include *supremacy* (ranked 10th), appearing 447 times predominantly in *white supremacy*, stating the subordinating position of Asian in American society.

3) *Dissemination*

The third keyword group sheds light on the dissemination and promotion of these tweets. This can be found through providing *carrds* (ranked 6th) providing further information, sources and links, along with requests for retweets (*RT* & *retweet*) and adding *hashtags*.

Celebrity endorsements emerge among the keywords, notably with K-pop boyband BTS (ranked 8th). The concordances of BTS reveal expressions of support and acknowledgement for BTS's anti-racism speech at the White House (see Example 3 in Table 3). Additionally, Tacha (19th), Anita Natacha Akide, a prominent social media Key Opinion Leader (KOL) who expressed her support for the #StopAsianHate movement, also emerged as a keyword.

For the dissemination discourse, we further went down the keyword list to investigate potential associations with previous movements. We discovered that BLM (Black Lives Matter) (101st) appeared in 50 instances, which is not very frequent considering the size of our dataset. This suggests a relatively weak connection between BLM and the most popular #StopAsianHate tweets.

4) *Expressions of attitudes/sentiments*

The three aforementioned keyword groups cover most of the top 20 keywords, with *solidarity* notably expressing support towards #AntiAsianHate movement and condemnation of anti-Asian crimes. Extending our analysis to the top 50 keywords, we categorised them into two groups. The first group includes responses to hate crimes, such as *heartbroken* (ranked 31st), *heartbreaking* (45th), while the second group describes the crimes or the perpetrators using terms like *senseless* (37), *horrific* (41st), *disgusting* (42nd), and *sickening* (44th). However, an interesting pattern emerges when examining the concordances: the majority of these descriptors are applied with the abstraction strategy (i.e., a detachment from the agent of social actions), such as *shootings*, *acts*, *killings*, and *violence* (see Example 4 in Table 3), rather than to specific individuals such as killers, racists, perpetrators, or men. This echoes the linguistic trends observed in the first keyword group, where abstracted slogans rarely criticise or condemn individuals directly but the racist actions and behaviours. This prevalent use of abstraction suggests a deliberate linguistic practice to avoid direct attacks on individual racists, mirroring the patterns identified in the second layer of analysis.

V. DISCUSSION

Through an examination of the #StopAsianHate tweets from January 1, 2021, to December 31, 2022 on X/Twitter, this study consists of three analytical dimensions and employs a ‘funnel approach’, starting with a broad examination of the #StopAsianHate movement and progressively delving into the specific sentimental and linguistic dimensions within the top 10% most popular tweets. The analysis shows #StopAsianHate to be predominantly used in the US, with English being the mostly employed language. This suggests this movement is originated from English-speaking countries and targeting at a broader, English-speaking global audience to raise awareness and support. Overall, #StopAsianHate is predominantly used as counter-discourse.

A key insight from the study of #StopAsianHate discourse is how ‘model minority’ is represented as both a trope and a myth within the Asian community on X/Twitter. This conclusion is derived from combining NLP’s sentiment, politeness, and toxicity analyses with Corpus-Assisted Critical Discourse Analysis results. Firstly, the movement appears moderate from a sentimental perspective, as evidenced by many tweets receiving positive sentiment scores, highlighted by expressions of solidarity at the textual level. The popular tweets frequently use relational identification strategy, referring to those countering racism as ‘friends’, ‘folks’, and ‘family’. Although the politeness analysis indicates that many tweets may not seem polite (likely due to the brevity of tweet texts, which typically feature fewer polite lexicons than longer communication genres) and the qualitative analysis shows that tweets with negative sentiment scores often contain theme words like hate, racism, and crime, these are not harmful tweets. Instead, these tweets directly challenge racism and hate crime, which is evidenced by the extreme low probability of toxicity.

A significant linguistic feature further supporting the ‘model minority’ conclusion is the avoidance of direct condemnation or confrontation against individual racists. Even when negative attitudinal lexicons like ‘disgusting’ and ‘senseless’ are used, as shown in concordance examples, they target events or actions such as shootings and killings rather than the perpetrators. This prevalent use of abstraction suggests a deliberate linguistic strategy to focus on challenging the institutional racism (in this case, law enforcement agencies’ lack of understanding and prosecution of anti-Asian hate crime) while avoiding direct conflicts with individuals, further revealing the movement’s moderate tone.

Lastly, the discourse reveals #StopAsianHate activists’ self-positionings as lacking influential power and as being vulnerable in the movement. This is evident in the frequent use of keywords such as (white) supremacy and descriptions of Asians as passive recipients of actions. Such self-positionings expose the unequal power dynamics within the discourse, while indicating a growing self-reflecting collective awareness of the negative impacts of model minority stereotypes on Asian communities.

VI. Conclusion

In conclusion, this paper takes an innovative approach by integrating NLP and CACDA methodologies in the study of X/Twitter activism via a case study of #StopAsianHate tweets during the COVID-19 global pandemic. NLP techniques identify statistical patterns related to the emotional and social appropriateness of tweets, a capability where CACDA methods are comparatively less effective. Conversely, CACDA provides a three-dimensional framework that incorporates textual features, discursive patterns, and social practices, thereby enriching NLP findings by situating them within the social context and interpreting them through linguistic features and strategies. Findings suggest that the moderate tone, communal sentiment and the vulnerability self-positionings reveal Asian communities’ general subordinate position in the political power structure, and their lack of visibility and experience in activism. However, data at the same time provides fresh insights into the growing self-reflective collective awareness within the Asian communities of the negative impacts of ‘model minority’ stereotypes. The use of #StopAsianHate as a response to Atlanta shooting hate crime demonstrates the growing knowledge and skillsets of digital activism within the Asian community. Celebrity endorsement contributes to the global reach, dissemination and solidarity of the movement, giving momentum to Asian communities’ ongoing efforts in fighting racism in all forms.

Acknowledgements

This work was supported by the Morgan Advanced Studies Institute, Swansea University under Grant number MID1001-104.

References

[1] H. Lyu, Y. Fan, Z. Xiong, M. Komisarchik, and J. Luo, “State-level Racially Motivated Hate Crimes Contrast Public Opinion on the #StopAsianHate and #StopAAPIHate Movement.” arXiv, Apr. 29, 2021. doi: 10.48550/arXiv.2104.14536.

[2] C. S. Lee and A. Jang, “Questing for Justice on Twitter: Topic Modeling of #StopAsianHate Discourses in the Wake of Atlanta Shooting,” Crime & Delinquency, vol. 69, no. 13–14, pp. 2874–2900, Dec. 2023, doi: 10.1177/00111287211057855.

[3] S. J. Lee, Unraveling the “Model Minority” Stereotype: Listening to Asian American Youth, 2nd Edition. Teachers College Press, 2015.

[4] R. S. Chou and J. R. Feagin, Myth of the Model Minority: Asian Americans Facing Racism, Second Edition, 2nd ed. New York: Routledge, 2015. doi: 10.4324/9781315636313.

[5] N. D. Hartlep, The Model Minority Stereotype: Demystifying Asian American Success (Second Edition). IAP, 2021.

[6] V. Bascara, Model-Minority Imperialism. U of Minnesota Press.

[7] F. H. Wu, Yellow: Race In America Beyond Black And White. Basic Books, 2002.

[8] J. B. Trauner, “The Chinese as Medical Scapegoats in San Francisco, 1870-1905,” California History, vol. 57, no. 1, pp. 70–87, 1978, doi: 10.2307/25157817.

[9] Y. Wu and M. Wall, “COVID-19 and viral anti-Asian racism: A multimodal critical discourse analysis of memes and the racialization of the COVID-19 pandemic,” Journal of Contemporary Chinese Art, vol. 8, no. 2, pp. 107–127, Nov. 2021, doi: 10.1386/jcca_00040_1.

[10] H.-L. Cheng, “Xenophobia and racism against Asian Americans during the COVID-19 pandemic: Mental health implications,” Journal of Interdisciplinary Perspectives and Scholarship, vol. 3, no. 1, p. 3, 2020.

[11] Y. Li and H. L. Nicholson Jr., “When ‘model minorities’ become ‘yellow peril’—Othering and the racialization of Asian Americans in the COVID-19 pandemic,” Sociology Compass, vol. 15, no. 2, p. e12849, 2021, doi: 10.1111/soc4.12849.

[12] M. Ittefaq, M. Abwao, A. Baines, G. Belmas, S. A. Kamboh, and E. J. Figueroa, “A pandemic of hate: Social representations of COVID-19 in the media,” Analyses of Social Issues and Public Policy, vol. 22, no. 1, pp. 225–252, 2022, doi: 10.1111/asap.12300.

[13] M. Castells, Networks of Outrage and Hope: Social Movements in the Internet Age. John Wiley & Sons, 2015.

[14] S. Hill, Digital revolutions: Activism in the internet age. New Internationalist, 2013. Accessed: Apr. 03, 2024. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=fMj0AgAAQBAJ&oi=fnd&pg=PA6&dq=Digital+revolutions:+Activism+in+the+internet+age&ots=9SK9ZeuYDd&sig=-TqLinFoYqLuN3xquA0kojww6Tg

[15] D. Boyd, “Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications,” 2010, Accessed: Apr. 03, 2024. [Online]. Available:

http://kristinarola.com/356/spring12/readings/unit2/boyd_SNSasNetworkedPublics.pdf

[16] M. Zappavigna, Discourse of Twitter and Social Media: How we use language to create affiliation on the web. London: Continuum International Publishing Group, 2012. Accessed: Apr. 03, 2024. [Online]. Available: https://www.academia.edu/18311721/Discourse_of_Twitter_and_Social_Media_How_we_use_language_to_create_affiliation_on_the_web

[17] Y. Bonilla and J. Rosa, "#Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States: #Ferguson," American Ethnologist, vol. 42, no. 1, pp. 4–17, Feb. 2015, doi: 10.1111/amet.12112.

[18] S. Florini, "Tweets, Tweeps, and Signifyin': Communication and Cultural Performance on 'Black Twitter,'" Television & New Media, vol. 15, no. 3, pp. 223–237, Mar. 2014, doi: 10.1177/1527476413480247.

[19] S. Sharma, "Black Twitter? Racial hashtags, networks and contagion," New formations, vol. 78, no. 78, pp. 46–64, 2013.

[20] G. Yang, "Narrative Agency in Hashtag Activism: The Case of #BlackLivesMatter," Media and Communication, vol. 4, no. 4, pp. 13–17, 2016, doi: https://doi.org/10.17645/mac.v4i4.692.

[21] K. Keib, I. Himelboim, and J.-Y. Han, "Important tweets matter: Predicting retweets in the #BlackLivesMatter talk on twitter," Computers in Human Behavior, vol. 85, pp. 106–115, Aug. 2018, doi: 10.1016/j.chb.2018.03.025.

[22] Associated Press, "#StopAsianHate trends as Asian-Americans grieve after Atlanta attack," South China Morning Post. Accessed: Apr. 14, 2024. [Online]. Available: https://www.scmp.com/news/world/united-states-canada/article/3126040/stopasianhate-trends-online-asian-americans-grieve

[23] A. Lloret-Pineda, Y. He, J. M. Haro, and P. Cristóbal-Narváez, "Types of Racism and Twitter Users' Responses Amid the COVID-19 Outbreak: Content Analysis," JMIR Formative Research, vol. 6, no. 5, p. e29183, 2022.

[24] B. Wheeler, S. Jung, M. C. N. Barioni, M. Purohit, D. L. Hall, and Y. N. Silva, "#WashTheHate: Understanding the Prevalence of Anti-Asian Prejudice on Twitter During the COVID-19 Pandemic," in 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Nov. 2022, pp. 484–491. doi: 10.1109/ASONAM55673.2022.10068578.

[25] D. Murthy, Twitter: social communication in the Twitter age. in Digital media and society. Cambridge: Polity, 2013.

[26] E. Morozov, To save everything, click here: The folly of technological solutionism. PublicAffairs, 2013.

[27] M. Castells, The Rise of the Network Society: The Information Age: Economy, Society, and Culture. Wiley, 1996.

[28] M. Castells, Communication Power, Second Edition, Second Edition. Oxford, New York: Oxford University Press, 2013.

[29] M. Castells, The Network Society: A Cross-Cultural Perspective. Edward Elgar Publishing, Incorporated, 2004.

[30] C. Tilly, Social Movements, 1768-2004. New York: Routledge, 2019. doi: 10.4324/9781315632063.

[31] J. Cao, C. Lee, W. Sun, and J. C. De Gagne, "The #StopAsianHate Movement on Twitter: A Qualitative Descriptive Study," International Journal of Environmental Research and Public Health, vol. 19, no. 7, Art. no. 7, Jan. 2022, doi: 10.3390/ijerph19073757.

[32] J. J. Lee and J. Lee, "#StopAsianHate on TikTok: Asian/American Women's Space-Making for Spearheading Counter-Narratives and Forming an Ad Hoc Asian Community," Social Media + Society, vol. 9, no. 1, p. 205630512311575, Jan. 2023, doi: 10.1177/20563051231157598.

[33] X. Tong, Y. Li, J. Li, R. Bei, and L. Zhang, "What are People Talking about in #BackLivesMatter and #StopAsianHate? Exploring and Categorizing Twitter Topics Emerged in Online Social Movements through the Latent Dirichlet Allocation Model," in Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, in AIES '22. New York, NY, USA: Association for Computing Machinery, Jul. 2022, pp. 723–738. doi: 10.1145/3514094.3534202.

[34] E. Chen, K. Lerman, and E. Ferrara, "Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set," JMIR public health and surveillance, vol. 6, no. 2, p. e19273, 2020.

[35] N. Micallef, B. He, S. Kumar, M. Ahamad, and N. Memon, "The role of the crowd in countering misinformation: A case study of the COVID-19 infodemic," in 2020 IEEE international Conference on big data (big data), IEEE, 2020, pp. 748–757. Accessed: Apr. 03, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9377956/

[36] N. Micallef, M. Sandoval-Castañeda, A. Cohen, M. Ahamad, S. Kumar, and N. Memon, "Cross-Platform Multimodal Misinformation: Taxonomy, Characteristics and Detection for Textual Posts and Videos," Proceedings of the International AAAI Conference on Web and Social Media, vol. 16, pp. 651–662, May 2022, doi: 10.1609/icwsm.v16i1.19323.

[37] "Understanding X limits | X Help." Accessed: Apr. 06, 2024. [Online]. Available: https://help.twitter.com/en/rules-and-policies/x-limits

[38] "Standard search API," X Development Platform. Accessed: Apr. 06, 2024. [Online]. Available: https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets

[39] A. Kilgarriff et al., "The Sketch Engine: ten years on," Lexicography ASIALEX, vol. 1, no. 1, pp. 7–36, Jul. 2014, doi: 10.1007/s40607-014-0009-9.

[40] N. Fairclough, "Critical discourse analysis," in The Routledge handbook of discourse analysis, Routledge, 2013, pp. 9–20. Accessed: Apr. 06, 2024. [Online]. Available: https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.4324/9780203809068-3&type=chapterpdf

[41] O. Ajao, D. Bhowmik, and S. Zargari, "Sentiment Aware Fake News Detection on Online Social Networks," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2019, pp. 2507–2511. doi: 10.1109/ICASSP.2019.8683170.

[42] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, Art. no. 1, May 2014, doi: 10.1609/icwsm.v8i1.14550.

[43] C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts, "A Computational Approach to Politeness with Application to Social Factors." arXiv, Jun. 25, 2013. doi: 10.48550/arXiv.1306.6078.

[44] L. Hanu and team Unitary, "Detoxify." Nov. 2020. doi: 10.5281/zenodo.7925667.

[45] L. H. Haco James Thewlis, Sasha, "How AI Is Learning to Identify Toxic Online Content," Scientific American. Accessed: Apr. 09, 2024. [Online]. Available: https://www.scientificamerican.com/article/can-ai-identify-toxic-online-content/

[46] B. He, C. Ziems, S. Soni, N. Ramakrishnan, D. Yang, and S. Kumar, "Racism is a virus: anti-asian hate and counterspeech in social media during the COVID-19 crisis," in Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, in ASONAM '21. New York, NY, USA: Association for Computing Machinery, Jan. 2022, pp. 90–94. doi: 10.1145/3487351.3488324.

[47] T. McEnery, Swearing in English: Bad Language, Purity and Power from 1586 to the Present. London: Routledge, 2005. doi: 10.4324/9780203501443.

[48] P. Baker, C. Gabrielatos, and T. McEnery, "Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press," in Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press, C. Gabrielatos, P. Baker, and T. McEnery, Eds., Cambridge: Cambridge University Press, 2013, pp. iii–iii. Accessed: Apr. 08, 2024. [Online]. Available: https://www.cambridge.org/core/books/discourse-analysis-and-media-attitudes/discourse-analysis-and-media-attitudes/FEF234ED4D4B09E58E1753E9D2413951

[49] T. van Leeuwen, Discourse and Practice: New Tools for Critical Analysis. Oxford University Press, 2008. doi: 10.1093/acprof:oso/9780195323306.001.0001.