Jurnal Komunikasi: Malaysian Journal of Communication
https://ejournal.ukm.my/mjc
Volume 41(3) 2025: 1-20

A Qualitative Exploration of Hate Speech Manifestation in Afghanistan's Twitter-Sphere

QURBAN HUSSAIN PAMIRZAD*
Xi'an Jiaotong University, China

ABSTRACT

The proliferation of hate speech on social media in recent years has become a public concern. This infodemic not only tears apart intergroup relationships and poses adverse psychological consequences but can also instigate violence and hate crimes. Due to hate speech perplexity and multi-dimensionality across different cultures and languages, researchers have called for worldwide investigations into its cross-language identification and manifestation. Employing a corpus of 536,080 characters and 68,975 tokens in Persian and Pashtu languages, a dataset from Twitter, this study explored the causes and manifestations of hate speech among Twitter users in Afghanistan through a qualitative investigation. The findings showed that real-life conflictual events are the primary source of hate speech among Afghanistan's Twitter users. Political instability and sectarianism, coupled with the traumatic experience of war, violence, and injustice in the past, resulted in profound public grievances being reflected on social media. The study also noted that hate speech is manifested in four main ways: gender-based, racial/ethnic, political, and religious hate speech. The results highlight the need for governments to tackle the underlying real-world hate-fueling problems to lessen their adverse effects on individuals and society online. The research implications and relevant suggestions are also discussed at the end of the article.

Keywords: *Afghanistan, hate speech, media affordance, national identity, Twitter.*

INTRODUCTION

The advent of the internet and social media as critical components of the fourth industrial revolution (4IR), along with big data, artificial intelligence (AI), supercomputing, and artificial intelligence-generated content (AIGC), significantly changed how people live and have made the global village idea a reality. Social networking sites (SNS) facilitate global connectivity among users, bridging the distance between nations and cultures. Conversely, its technological affordances, such as accessibility, virality (Gallacher & Bright, 2021), and anonymity, have dramatically increased hate speech and incivility (Castaño-Pulgarín et al., 2021; Lingam & Aripin, 2017). While the positive impacts of modern communication technology on the quality of human beings' lives are enormous, the dark sides of these tools are also concerning, particularly their negativity in the spread of hate speech and related hate crimes (Gorenc, 2022; Lingam & Aripin, 2017).

Studies suggest that, along with the growth of SNS, hate speech promoting violence and hatred, having psychological and societal consequences, has exponentially increased (Castaño-Pulgarín et al., 2021; Lingam & Aripin, 2017; Matamoros-Fernández & Farkas, 2021; Mathew et al., 2019; Ștefăniță & Buf, 2021; Sugarman et al., 2018; Williams et al., 2020). For instance, studies show that around 80% of European citizens have encountered hate speech on social media

*Corresponding author: pamirzad2023@stu.xjtu.edu.cn
E-ISSN: 2289-1528
https://doi.org/10.17576/JKMJC-2025-4103-01
Received: 27 June 2024 | Accepted: 8 December 2024 | Published: 30 September 2025

(Castaño-Pulgarín et al., 2021), as well as research has found associations between social media hate speech and real-life hate crimes (Bilewicz & Soral, 2020; Mathew et al., 2019). Given the seriousness of the issue and its negative social and personal harms, hate speech has captured the attention of policymakers, governments, private companies, and researchers alike (Bilewicz & Soral, 2020; Waseem & Hovy, 2016), with some countries, e.g., Germany and Britain, adopting new policies and regulatory mechanisms to curb this phenomenon (Butler & Turenne, 2022). Developed countries with robust data infrastructure and economic and political power can mitigate the negative impact of these tools on their citizens; however, many third-world nations, particularly those grappling with conflict and turmoil, such as Afghanistan, which lack necessary data infrastructures and the capacity to launch local alternative SNS, bitterly remain vulnerable to these tools' adverse effects.

Afghanistan has already been plunged into decades of bloody war coupled with religious fundamentalism, deep sociocultural divide, and ethnocentrism (Ibrahimi, 2023; Wafayezada, 2023), susceptible to social chasm and hatred. Malicious users can leverage social media affordances to spread hate speech based on existing social cleavages. Due to the lack of national cohesion and deep-rooted social disparity, almost any topic, from religion to language, culture, ethnicity, and even sport, can easily be contested in Afghanistan's social media landscape, leading to barrages of hate speech. However, except for limited literature associated with the issue of hate speech toward Afghanistan's immigrants in Turkey and Iran, domestically, it has not been explored (Parvaresh, 2023; Subay & Gökalp, 2023). Bridging this noticeable research gap, this study explores the potential causes and manifestations of hate speech among Afghanistan's Twitter users. The next section reviews relevant literature, and research questions are proposed. The following section introduces the research methodology and procedure. The research findings, discussion, and conclusion, along with recommendations and suggestions, will be shared in the final section.

LITERATURE REVIEW

What is Hate speech

Despite being a buzzword, hate speech still has no agreed-upon definition (Bilewicz & Soral, 2020). Due to it being a vague concept, the definition of hate speech lies between minor abnormal language and law-infringing behaviors (Butler & Turenne, 2022). Based on previous research, this concept can be defined from the following aspects:

a. the negativity associated with the message itself,
b. the negative impact it inflicts on the targeted social groups,
c. the forms in which these negative meanings are being communicated,
d. and the normative infringement associated with these types of content.

For instance, scholars defined it as the use of swearing, insulting, abusing, and hateful derogatory words (Lingam & Aripin, 2017). Other studies highlighted the negativity of the message targeting a specific group. Cohen-Almagor (2018) defined it as biased, hostile, and malicious speech directed toward an individual or a group based on their actual or perceived inherent traits. Fortuna and Nunes (2019) extended this definition and defined hate speech as any form of language that disparages or attacks a group of people based on their physical characteristics, religion, race, national or ethnic origin, sexual orientation, gender identity, or any

other characteristic (Schmid et al., 2022). Other scholars have focused on its communication form, such as verbal, non-verbal, visual, and textual communicative actions attacking a social group (Chetty & Alathur, 2019; Ștefăniță & Buf, 2021). Also, it has been characterized as the explicit and implicit use of derogatory words, including blatant hate speech, rhetorical questions, and even subtly expressed humor against an individual or group (Nascimento et al., 2023; Parvaresh, 2023; Schmid et al., 2022).

Furthermore, hate speech has been defined from a normative perspective as activities that violate social and ethical norms. For instance, Castaño-Pulgarín et al. (2021) described hate speech as a form of social deviance violating social and ethical norms in a society. Content that goes against the offline standard norms (Gorenc, 2022). These anti-normative activities can include defamation, calling for violence, agitation, and provoking statements for political and social issues, displaying discriminatory views and rumors, using sexist language, denial of human rights, ad hominem attacks, negative stereotypes, irony, manipulation, and slurs (Castaño-Pulgarín et al., 2021). Moreover, SNS companies have their own definitions. For instance, according to Ben-David Fernández (2016) and Mathew et al. (2019), Facebook and Twitter define hate speech as any post or tweet that advocates for the use of violence against individuals based on their race, ethnicity, national origin, gender, sexual orientation, age, disability, or serious illness, or that directly attacks such individuals. Synthesizing these definitions, aligned with Nascimento et al. (2023) and Schmid et al., 2022), this study defines hate speech as the implicit or explicit use of anti-normative, derogatory, hateful, and abusive language to target an individual or group of people based on their social group membership, be it religion, race, gender, socioeconomic class or political affiliation.

Hate Speech Manifestations, Threats, and Social Media Affordance

Given that hate speech is targeted at an individual or a group based on their distinct property associated with group membership (Chetty & Alathur, 2019; Nascimento et al., 2023; Schmid et al., 2022; Ștefăniță & Buf, 2021), previous studies divided the manifestation of hate speech into religious, online racism, political, and gender-based categories (Castaño-Pulgarín et al., 2021; Sugarman et al., 2018). Religious hate speech refers to the use of inflammatory and sectarian language promoting hatred and violence toward the followers of a religion. Online racial hate speech involves the use of derogatory and hateful speech about a specific ethnicity and race; similarly, political hate speech involves the use of hateful, dehumanizing, and demonizing language against political opponents. Lastly, gender-based hate speech includes the use of negative and abusive words toward an individual or social group based on their gender or sexual orientation (Chetty & Alathur, 2019; Castaño-Pulgarín et al., 2021; Ștefăniță & Buf, 2021).

Furthermore, scholars suggest that hate speech has different degrees of intensity, which can be similar to Allport's (1954) discrimination continuum from using prejudiced language at the low to the intention of extermination at the highest level. Hate speech intensity can also lead to avoidance, discrimination, physical attack, and extermination (Bilewicz & Soral, 2020). For instance, curating social ties and selective network building based on avoidance of outgroups on social media reinforces the discrimination between social groups and leads to hostility (Lingam & Aripin, 2017). Similarly, cascading high-intensity hate speech inciting violence and genocide in real-life settings can contribute to hate speech-related crimes, such as the attack in Pittsburgh's

synagogue and the explosion at Christchurch Mosque in New Zealand (Bilewicz & Soral, 2020; Gallacher & Bright, 2021; Mathew et al., 2019)

Studies argue that social media technical features and affordances such as interactivity, accessibility, virality, and anonymity contributed to the spread of hate speech in social media feeds and comment sections (Ben-David & Fernández, 2016; Castaño-Pulgarín et al., 2021; Gorenc, 2022). The interactive feature of social media allows malicious users involved in spreading hate speech to aggregate clusters of online hate disseminators (Schmid et al., 2022). Similarly, scholars argue that social media anonymity affordance – providing a safe zone – enables malevolent users to spread violent, hostile, and uncivil content, known as cyber hate (Castaño-Pulgarín et al., 2021; Gorenc, 2022). However, other studies argue that not social media affordances per se but their misuse is the main reason for instigating hate speech (Ben-David & Fernández, 2016; Nascimento et al., 2023; Schmid et al., 2022). Malicious actors, commonly known as trolls, provoke heated debate among social media users and, through fueling the discussion with negatively charged content and weaponized information, upset the overall network, posing significant personal and societal harm (Gorenc, 2022; Lingam & Aripin, 2017; Nascimento et al., 2023).

According to previous research findings, hate speech led to victims' depression and anxiety at the individual level (Schmid et al., 2022; Ștefăniță & Buf, 2021), reinforcing discrimination and hostility in their worldview toward outgroups. Similarly, at the social level, hate speech as an iceberg of discrimination and injustice instigates violence and is susceptible to destabilizing countries, particularly those in post-conflict situations (Asimovic et al., 2021; Bilewicz & Soral, 2020; Sugarman et al., 2018). Furthermore, studies found that repeated exposure to hate speech on social media can lead to desensitization of the victims, which could have perilous cognitive, affective, and behavioral personal damages with severe social consequences for the intergroup relationship (Bilewicz & Soral, 2020; Schmid et al., 2022). Considering Afghanistan's situation, which suffers from a lack of internal cohesion and unity, coupled with long decades of hostility and violence, exploring the manifestations of hate speech whereby mitigatory policy recommendations can be proposed is of extraordinary significance. Such measures contribute to mitigating hate speech's impact on the masses. Against this backdrop, this study aims to answer the following questions:

RQ1: Which factors incite hate speech in Afghanistan's Twitter-sphere landscape?
RQ2: What are the dominant patterns and manifestations of hate speech among Afghanistan's Twitter users?

METHOD

Case Selection and Sampling

This study uses a qualitative content analysis method to investigate the potential causes and manifestations of hate speech in Afghanistan's Twitter-sphere. Hate speech has become ubiquitous on social media, and its virality on Twitter has made this platform the most studied regarding hate speech (Matamoros-Fernández & Farkas, 2021). At the same time, Twitter has become a hotbed of contentious political talks in Afghanistan; hence, it was chosen as the target platform. Using a keyword inquiry approach, a list of the most common ethnic-racial, political, and social hate speech-associated terms, considering the socio-political context of Afghanistan,

was searched. This is a common method for data extraction from social media (Waseem & Hovy, 2016). Additionally, to specifically trace the hate speech in Afghanistan's Twitter ecosystem, certain recent conflictual cases were used. The online campaign called "من افغان نیستم [I am not Afghan]," Afshar "افشار" [a tragic massacre event occurred in 1993 which is remembered every February], and "a pan-Persianate poem recitation advocating for the unity of Persian-speaking countries in March 2024" have recently been extensively debated and incited barrages of hate speech. These cases, carrying socio-political chasm and historical implications, helped to derive a corpus encompassing the comments and opinions of different groups. Choosing specific cases to study hate speech and social media polarization is a method already used in previous studies (Asimovic et al., 2021).

To narrow down the sample, this study selected the posts with over 20 comments. The reason for setting this criterion is that, based on previous studies, comments are one of the quantitative metrics of social media, which underscore the profundity of a discussion (Konovalova et al., 2023); thus, posts with less than 20 comments may look superficial, and their presence in the sample would not make a big difference. As a result, 62 posts met the above criterion, and their relevant (N = 3518) comments using the Twitter replies extractor plug-in software were extracted. The data for this study was collected in June of 2024, with comments time-frame associated with the selected posts on the subject matter from 4/29/ 2019 to 3/4/2024. The data collected for this research is located in the public domain, which does not violate users' privacy (Harrington, 2024). Besides, to avoid the potential harm in the context of hate speech to research subjects, the author applied two strategies: the comments were used in the study's general context without mentioning any personal information about the person who posted them. Additionally, the linguistic structure of comments was modified through keyword manipulation that keeps the meaning intact but prevents potential accessibility to searching the comment. By following these cautionary measures, the researcher discarded any ethical issues associated with the current study. The cleaned corpus included 536,080 characters and 68,975 words formatted in plain text for thematic analysis, and NVivo 1.7.2 was used for the qualitative data analysis.

FINDINGS

Themes

Following the steps for thematic analysis (Naeem et al., 2023), the author first, through reviewing and observing the data, familiarized himself with the corpus. In the second step, using keyword analysis on NVivo, which is one of the most used approaches to explore the linguistic importance of words inside a corpus (Weismayer & Pezenka, 2017), dominant initial codes emerged. Then, keyword frequencies were extracted based on which initial themes and sub-themes were established (Parvaresh, 2023). To delve deeper into the data, the word tree of the keywords was inspected to refine the themes. Additionally, manual coding was carried out to connect the keywords with relevant themes based on the references in the corpus. In the coding process, both inductive and deductive approaches were followed (Williams & Moser, 2019). Based on the deductive coding approach, the hate speech classification proposed by previous research: ethnic-racial, religious, political, and gender-based hate speech (Castaño-Pulgarín et al., 2021), was followed as a coding framework. Besides, the corpus was inductively scrutinized for emerging

themes such as the contested topic, figures, and hate speech-language. The author shared and discussed the themes with colleagues proficient in the Persian and Pashtu languages to resolve inconsistencies and finalize the themes and subthemes in inductive coding. For deductive coding, an undergraduate student proficient in the two mentioned languages was trained, and the author and assistant independently coded 10% of the comments (N = 351), achieving an acceptable average over a 0.70 Cronbach alpha score (Perreault & Leigh, 1989). The coding scheme was further discussed to resolve inconsistencies and ensure inter-coder reliability. Tables 1 and 2 show the inductive and deductive approaches applied in the coding process and the final established themes and subthemes.

Table 1: Inductive coding of hate speech-related themes, subthemes, and keywords

Themes	Subthemes	Persian Keywords	Phonetics	English gloss
Contested topics	Identity/ethnicity	افغانیت/اوغانیت "هویت ملی"	Afghānīyat/ūghānīyat;	Being Afghan [national identity].
	History	افشار	afshār;	Afshar.
	Language	فارسی/دری	fārsī/dari	Persian/Dari
Contested Figures		مسعود، مزاری، حکمتیار، ربانی، سیاف، دوستم	mæsuːd, mæzɑːri, hɪkmætjɑːr, ræbɑːni, Sayyaf, Dostum	Massoud, Mazari, Hikmatyar, Rabbani Sayyaf, Dostum
Hate speech language	Derogatory terms.	لوده، چاپلوس	luːdeh, chāpūlus;	Stupid, flatterer.
	Abusive words.	دشنام های رکیک	dashnām-hā-ye rukīk;	Blatant, abusive words.
	Common accusations	وطنفروش، خایین، جاسوس، مهاجر	vatan-forūsh, khāyīn, d͡ʒɑːsuːs, muhājir	Traitor to the homeland, Traitor, spy, Immigrants

Table 2: Deductive coding of hate speech manifestations and the subcategories

Themes	Subthemes	Persian Keywords	Phonetics	English gloss
Racial hate speech	Pashtun	اوغان، غول، انتحاری، تروریست	Ūghān, ghūl, entehɑːri, terrorist.	Awghan, Ghul, bombers, terrorist.
	Tajik	رقاصه، لندغر، کولابی	ræɢsɑːseh, lændɣɑːr, kuːlɑːbi;	Dancers, vandals, Kollabies;
	Hazara	مغول، نسل چنگیز، موشخور	moghūl, nasl-e changīz, mūshkhūr	Moghuls, Changiz-generation, Mouse-eaters
Political hate speech		جاسوس، مزدور، بغاوتگر، تروریست، فراری	d͡ʒɒːsuːs, mæzduːr, baghāvatgar, terrorist, færɒːri	Spy, Hireling, Insurgent, Terrorist, Fugitive
Religious hate speech		رافضی، کافر، مرتد، داعشی، وهابی، یهودی	rɒːfiːzi, kɒːfer, mortæd, dɒːʔeʃi, væhæbi, jæhuːdi	Rafizi (derogatory words toward Shi'a Muslim), pagan, apostate, Wahhabi, Jewish
Gender-based hate speech		فاحشه، تن فروش، هم جنس باز	Fahisheh, tan furosh, hamjinsbaz	Whore, prostitute, homosexual [derogatory tone]

QUALITATIVE ANALYSIS

The Socio-Political Context for Hate Speech in Afghanistan's Twitter-Sphere

Observing the contested topics through thematic analysis, as shown in Table 1, in answer to our first question, it was found that socio-political and historical events are the prevalent source of hate speech among Afghanistan's Twitter users. National identity, unpleasant historical events, and ethno-lingual topics are the main themes and instigators based on which flaming occurs online, and clusters of hate speech promoters aggregate around. National identity as "Afghan" has been a topic of hot debate for a long time. Within the last two decades, resentment intensified and resulted in structured and semi-structured attempts for its reconsideration. Historically, the term Afghan refers to the Pashtun people, who are one of the four major ethnic groups in the country, in addition to many other minority ethnic groups. This concept, as the national identity, was extended to all citizens of Afghanistan in the 1964 constitution (Dinakhel, 2019; Shahrani, 2018), which now refers to every citizen regardless of their ethnic group. Except for the Pashtun elites, particularly nationalists who believe that they make up over 50% of the country's population – no general census has proved it – and this term should be generalized to all whether one agrees or not (Ahady, 1995; Lieven, 2023; Sungur, 2016), the rest of ethnic groups, particularly elits, consider it a sort of imposition of one ethnic identity to others (Afzal, 2022). As social media has provided a platform for people to discuss their thoughts, this topic has become a hotbed for online clashes and hate speech in Afghanistan.

Unpleasant historical events and involved figures associated with such inter-ethnic conflict are other topics inciting online violence and hate speech in Afghanistan's Twitter-sphere. The Hazara people's mass killing during Abdurrahman Khan in the late 19th century (Ibrahimi, 2019; Hakimi, 2023), the Pashtun settlement in the north of Afghanistan known as Pashtun colonization (Afzal, 2022; Sungur, 2016), the Afshar massacre and Mujaheddin civil war, and the Taliban atrocities against other ethnic groups in the 90s (Human Rights Watch, 2005) are some of these unpleasant events. Additionally, individuals involved in such events have been associated with ethnic groups, laying the foundation of inter-ethnic animosity. Since almost any issue is seen through an ethnic lens in Afghanistan, hate speech targeting these individuals extends to the ethnic level on social media, provoking hate speech.

Furthermore, the language issue is another contentious topic among Afghanistan's Twitter users, which incites hate speech. Persian language, renamed Dari in Afghanistan, in the 1964 constitution, has been the nation's lingua franca for centuries, and approximately 80% of the population from all ethnic groups speak this language (Dinakhel, 2019). However, in the process of building an "Afghan" national identity in the 20th century, it became a victim of oppressive rulers (Afzal, 2022; Ibrahimi, 2019). Since the 1930s, systematic efforts have been made to marginalize this language (Dinakhel, 2019). In 1937, the Pashtu Academy (Pashtu Tolana) was established to make this predominantly orally spoken language and its various dialects (i.e., Pashtu language) into a standard single language as the country's first language instead of Persian/Dari. Since then, Afghanistan's rulers have systematically promoted the Pashtu language at the cost of suppressing Persian (Dinakhel, 2019), including taking down and cleaning the banners of offices and universities written in this language in some provinces in recent years (Ahwar, 2023). These omissive language biases continue to date and have become one of the contentious topics among Afghanistan citizens on Twitter.

Hate Speech Manifestations

To answer the second question, through deductive coding, the manifestation of hate speech was coded in previously established categories: Ethnic/racial, political, religious, and gender-based hate speech (Castaño-Pulgarín et al., 2021; Ștefăniță &Buf, 2021).

a. Ethnic/ Racial Hate Speech

Ethnic and racial hate speech involves using abusive, derogatory slurs against the members of an ethnic group (Chetty & Alathur, 2019; Castaño-Pulgarín et al., 2021; Ștefăniță &Buf, 2021). This type of hate speech has been prevalent in societies where ethnic and racial divisions exist. Afghanistan has been historically divided along ethnic lines, and due to inter-ethnic conflict and civil war, social cohesion has been torn apart (Ibrahimi, 2023). Among the fourteen ethnicities mentioned in the constitution enacted in 2004, Pashtun, Tajik, Hazara and Uzbek are the four major groups (Ibrahimi, 2023). However, except for temporary breakages, Pashtuns have been the country's rulers since the Durrani Kingdom's establishment in 1747 (Ahady, 1995; Ibrahimi, 2023). Their rule has been characterized by hegemonic power and subduing and marginalizing other groups (Ibrahimi, 2023; Lieven, 2023; Shahrani, 2018). Therefore, much of the hate speech directed at the Pashtun people is associated with their ruler's ethnocentric policy imposition (Wafayezada, 2023). They have been targeted using derogatory and abusive terms such as "terrorists, savages, bombers, Awghan Ghul (stupid)," coupled with "treachery and tribalism, as well as fascism" accusations. The term "Awghan" is a common colloquial noun for the Pashtun people, and nowadays, it carries a sort of stereotype equal to backwardness. Similarly, keywords such as "Awghanwali," a derogatory term used instead of Pashtunwali, which is the ethnic behavioral code of this ethnicity (Lieven, 2023), were also found in the corpus targeting this ethnic group.

> Persian phonetic: *"Afghan yani Awghan yani Taliban yani khayen yani entehari… ma Afghan Awghan nistim…"*
> Translation: "I am not Afghan. Afghan means Awghan, means Taliban, traitors, suicide bombers, and so forth…"

The above comment shows that the individual not only denied being known as "Afghan" but also made the stereotypical association between the term "Afghan" and derogatory and abusive words toward the Pashtun ethnic group. This Afghan identity denial stems partly from historical resentment and partly from the fact that this term has heavy ethnic implications being imposed on others without their consent (Afzal, 2022). During the last two decades, as ethnocentrism skyrocketed, public resentment against this term from marginal voices turned into mainstream debate. Expectedly, such debates have not remained unanswered and have attracted fierce responses from the Pashtun people.

> Persian phonetic: *"In malek az aghlīyat-hā-ye nāql delāl safat nist, in malek vāresin dāran, ān ham Ūghān."*
> Translation: "This land has legitimate heirs, and those are Awghans [Pashtuns]; it is not of dishonorable and treacherous minorities."

The above comments defending the term Afghan represent a typical group among the Pashtun nationalist movement, which treats other ethnic groups as migrants and only Pashtuns as the heirs of this land. The interactivity of Twitter reinforces this animosity chain in a more fierce and violent counter-argument as the debate goes by.

> Persian phonetic: *"Barāye hameh ma'loom dār ast, mohājer kist? Mohājer-e shomā Afghān-hā-ye terorist, ṭālibāni, Dā'eshi, 'Abd al-Raḥmāni, Larubar, mazdoor, ḡolām, khod forukhte, ma'āmelagor, khā'in, khāk-e jenāyat-kār, tajāvuzgar, luṭat-kār… hestid."*
> Translation: "It is evident to everyone who the immigrants are. You, immigrants, are the terrorists, Afghans, Talibanis, Daeshis, Abdul Rahmanis, Lar-aw-bars [the nationalist Pashtun slogan claiming the Khaybar Pakhtunkhwa of Pakistan to be Afghanistan's land], mercenaries, slaves, traitors, soil defilers, criminals, invaders, and looters…"

Similarly, the Tajik ethnic group has also been the target of hate speech on Twitter. They have been targeted partly because of challenging the ethnic hegemony of Pashtuns and partly because of Tajik Mujahideen involvement in events during the civil war in the 1990s. In 1929, a Tajik man, Habibullah Kalakani, with the help of religious clerics and tribal leaders, rebelled against the Amanullah Khan (i.e., Pashtun King) and became the first non-Pashtun ruler of the country (Ahady, 1995; Ibrahimi, 2023). His move challenged the two-century hegemony of the Pashtun people who had ruled the land back to 1747. Later, during the communist regime of Afghanistan (1978-1992), Babrak Karmal, another Tajik, became the second ruler of the country. Similarly, Burhanuddin Rabbani became the third Tajik ruler of Afghanistan after the collapse of the communist regime in 1992. Although not anti-Pashtun, these turning points have been bitterly challenging for the Pashtun nationalist who take ruling the country as their exclusive inherent right (see Ahady, 1995; Wafayezada, 2023). These power shifts prompt a large part of the derogatory words and hate speech targeted at the Tajiks. For instance, they are offensively called "Saqawi," a term that refers to Habibullah Kalakani, given that his father was a water carrier known as Saqaw in Persian (Lieven, 2023).

> Persian phonetic: *"Nasl-e khabeeseh, saqaav, churk, o Masoud ghadar, be haddi asliyat-e khod ra ashkaar kardin ke ba'd az in, hich kas shoma namus forush-ha ra kasi az Afghanistan hesab nemikond, o marg-e shoma namak-e haram-ha-ye tazi farz ast, dast-e Talib dard nakonad."*
> Translation: "You devious Masood [Ahmad Shah Masaud, the prominent anti-Soviet Tajik commander during the 90s in Afghanistan] and vile Saqaw generation, you have shown so much of your true selves that no one will ever think of you as being from Afghanistan again. It is required that you traitorous Tajiks die; may the Taliban be blessed when they execute you."

The above example targets Tajiks based on the context provided. Not only being called *Saqawi* (water carriers) but also using further abusive words toward Ahmad Shah Massoud and Burhanddin Rabanni, two other Tajik politicians who challenged Pashtuns' dominance in the 1990s. Besides, the above comment inciting violence and legitimizing the massacre against Tajiks (Human Rights Watch, 2022) can be considered as the extermination stage of the Allport racism continuum (Bilewicz & Soral, 2020). Another derogatory term against the Tajiks, and particularly to the former mujahideen, is "Landaghar." This term, which can be translated as vandalism, was originally used against the Jihadi commanders who held political and military power in the capital, and occasionally violated the law. Nonetheless, it has been politicized and, as a stereotype, was turned into an abusive word targeting the Tajik ethnic group as a whole. Furthermore, the Tajiks were targeted as Kullabies, implying that to were migrants from Tajikistan.

> Persian phonetic: *"Afghan budan nang, gheyrat o mardangi mikhwahad ke chenin chiz dar nasl-e Shingari o Kolabi nist. Talib be Shingari raftan, noqteh payan gozasht, va hali nasl-e mohajer bayad ke dobareh be Kolab beravand va Kolabi shavand."*
> Translation: "The Shingari and Kolabi generations [of Tajiks] lacked the honor, honesty, and bravery that it takes to be Afghan. The Taliban destroyed Kollabi's power, and now, the new generation of immigrants has to go back to Kolab and accept its ideals."

The above comment further praises the Taliban, predominately a Pashtun group (Kaltenthaler et al., 2022), for the inhalation of the Tajik's power and denigrates their Afghanistan citizenship. Calling each other migrants, however, exists among all ethnic groups. While Pashtun nationalists consider the Tajiks to be migrants from Tajikistan, Uzbeks from Uzbekistan, and Hazaras from Mongolia, the Pashtuns are also considered Israeli migrants by others. That is why, different from other cases in the world, ethnic hate speech in Afghanistan is not against a specific group (Chetty & Alathur, 2019; Castaño-Pulgarín et al., 2021; Ștefăniță &Buf, 2021) but a full-fledged inter-group hate exchange.

Another ethnic group that has been subject to historical cruelty and even genocide is the Hazaras (Ibrahimi, 2023; Hakimi, 2023). This ethnic group, which is mainly the followers of the Shia sect of Islam, is not only subjected to racial/ethnic hate speech but also to religious hate speech, "hybrid hate speech" (Chetty & Alathur, 2019). They are being targeted using offensive and harmful terms such as the Changiz generation, Mughuls (Mongols), Mouse-eaters, etc.

> Persian phonetic: *"Valla in nasl-e harami-hā-ye Chingiz mazlum beguyad, inha mazlum-namā hastand. Dast-e shān berasad be Khodā, qatl-e 'ām mikonand, mānand bābā-ye Chingiz shān…"*
> Translation: "This Changiz generation [Hazara] is acting to be innocent. By God, if they have the power, they commit mass murder, just like their father, Chingiz."

Calling Hazar people the Changiz generation, regardless of being a topic of research (He et al., 2019), in this context, is used negatively to deny their legitimate rights as native citizens of Afghanistan. Another abusive term targeting Hazara people is calling them Mouse-eaters,

coupled with historical claims regarding the Hazara's Mujaheddin complacency in the civil war era.

> Persian phonetic: *"Sineh boridan o raqs-e morde miras-e shomā, qaum-e mush khor ast."*
> Translation: "Cutting women's breasts and dancing on the dead is the legacy of you mouse-eater people."

The above comment demonstrates the hateful inter-ethnic conflict in Afghanistan, which has been shaped by political animus in the aftermath of the communist regime in 1992. Reports demonstrate that almost all political jihadist factions were responsible for the unpleasant, cruel atrocities of that era (Human Rights Watch, 2005); however, now every ethnic group tries to exonerate itself and blame others.

b. Political Hate Speech

Political hate speech refers to using demonizing and degrading language against political opponents (Chetty & Alathur, 2019; Castaño-Pulgarín et al., 2021). In this study, except for the past political hatred intertwined with ethnic animus, three main groups were the target of political hate speech: the Taliban, previous government officials or affiliated individuals, and the armed resistance fronts opposing the Taliban. The Taliban were targeted with political hate speech, being referred to as the "Panjabi slaves and terrorists". However, these derogatory terms extended beyond the Taliban and also targeted the Pashtun ethnic group, "hybrid hate speech" (Chetty & Alathur, 2019), given that the Taliban are predominantly a Pashtun group (Lieven, 2023). Targeting the Taliban as minions of Pakistan has been a common way of addressing this group, with the Taliban denying any relationship with this country. However, reports and studies show that the Taliban has overtly and covertly deep-rooted relationships with Pakistan (Miller, 2021; Lieven, 2023). Therefore, whether it can be considered hate speech or a grounded claim about this group is a matter of different perspectives. Previous studies have also noticed the challenging nature of the demarcation of hate speech (Bilewicz & Soral, 2020).

> Persian phonetic: *"Bist sāl mazdoori o ghulāmi-ye Panjāb rā kardid, hanuz ham sharm nadārid. Lāshehā-ye Mollā Mansoor o Mollā 'Omar koor dar kojā be jahannam ravān shodand?"*
> Translation: "You spent twenty years as mercenaries and slaves of Punjab, yet you still feel no shame. Where did the corpses of Mullah Mansoor and Mullah Omar the Blind go when they departed to hell?"

On the contrary, the Taliban targeted their armed opposition in derogatory terms such as "Baghawatgar" (insurgents). In the Islamic political context, specifically based on the Taliban interpretation, anyone rebelling against the Islamic state is sentenced to death; hence, it connotes a strong negative meaning in Persian and Islamic contexts.

> Persian phonetic: *"In javān be esm-e Amīr Ḥosayn tavassot-e khānevāde-ye Khoshrash be qatl raside, 'enāṣer-e fathneh-gar, boghāvat-gar, monāfeq, qatl o-rā rabṭ bar mojāhedin-e Emārat-e Eslāmi midahand."*
> Translation: "The father-in-law's family murdered this young man named Amir Hossein…The elements of sedition, rebellion [Baghawatgar], hypocrisy, and murder are linked to the Islamic Emirate of Afghanistan's mujahideen."

Similarly, the Taliban targeted the people in the previous government and those who challenged their government as fugitive stooges (Muzdooran-e-Farary). The Taliban hate speech has dominantly used this labeling against their opponents.

> Persian phonetic: *"Herdiwuth murtad laloo panju jāsoos ham-jensbāz farāri Londgar Afghān shode nemitaaneh."*
> Translation: "No apostle, spy, fugitive, or homosexual can be Afghan."

Likewise, the Taliban Twitter users targeted their opponents as spies (Jasoos), blaming them for being involved in the United States and Western countries' invasion of Afghanistan.

> Pashtu phonetic: *"Hagha wakht che tay dashghāl-garwājarti qātil wei, hara warz pah las hāwūdāsē zwānān deta pah dastūr-aw jāsūsī, taror, udbamūno lāndē qurbānēdil arman che tāsu bēwajdāneh udshaytān ūlād demlat demhukamē na watshetēd, ūmajāzāt na shawē."*
> Translation: "In an era marked by your reign as a perpetrator of the innocent, the city witnesses the daily loss of young souls as a consequence of your commands and cunning. Regrettably, as Stan's descendant, lacking any semblance of a conscience, you escaped and went without facing justice."

Afghanistan's political landscape is so murky that every group holding power, from their perspective, they are the only side having the truth, while the rest are all wrongdoers.

c. Religious Hate Speech

Religious hate speech refers to using derogatory and offensive words to spread hatred toward a person or a group based on their religious beliefs (Castaño-Pulgarín et al., 2021; Ștefăniță & Buf, 2021). Although Islam is the predominant religion in Afghanistan [previously, there were some Hindus and Jews in the country], in addition to the predominant Hanafi sect, other sects of Islam, including the Shia (Twelvers) and Ismailia, Salafi, and some Sufis, also exist in the country. However, discrepancies in religious interpretations, even within one sect, have made religious belief a matter of conflict and a factor inciting hate speech online. Derogatory terms such as "Rafizi", which is an abusive word, attacking the Shia muslims, "Kafir" (pagan), "Munafiq" (hypocrite), and "Murtad" (apostate), "Yahudi" (Jewish) are the terms which are used to attack specific individuals or a social group based on religion.

Persian phonetic: *"Khāk be sar-e in khabeese mazdoor o chāpolus, ābro-ye Afghānhā, rah pish yek sag-e Rāfizi, sefr sākht. Mardom-e Afghānestān az saghā-ye chon nābārur o to kasif bīzār ast."*
Translation: "May this wicked, mercenary, and sycophantic individual be covered in dust. He has compromised the honor of Afghans by groveling before a Rafidi [Rafizi] dog. The people of Afghanistan despise dogs like him, who are not only sterile but also filthy."

In the above comment, the term "Rafizi" is used as a religious offensive word that targets the Shia Muslims in general. Other comments using hybrid hate speech (Chetty & Alathur, 2019) targeted a person or social group based on multiple factors. E.g.:

Persian phonetic: *"Unsel yahud o dokhtar forush-hā-ye bi farhang mazdoor oghan khar, shomā nasl-e kāfer vatanshūsh khod-e bā qahramān-e meli-ye keshvar barābar mikonid."*
Translation: "You, the Jews and uncultured prostitutes, mercenaries, traitors, and donkeys, equate yourselves with the national hero of the country."

This hateful comment not only contains religious hate speech but also carries strong, harmful ethnic, racial, and even gender-based hate speech while defending a self-ethnic position. Overall, religious intolerance is the main reason for spreading religious hate speech in Afghanistan.

Pashtu phonetic: *"Daghay murtad har che bayalli da de doy gheyrat de."*
Translation: "This apostate is nothing but a shameless person lacking honor."

d. Gender-Based Hate Speech

Gender-based hate speech refers to the use of derogatory and abusive language toward an individual or a social group based on their sexual orientation (Chetty & Alathur, 2019; Castaño-Pulgarín et al., 2021). Studies stated that women and LGBTQ+ individuals are victims of gender-based hate speech, and probably, this type of hate speech is the most vicious and blatant abusive language (Chetty & Alathur, 2019; Ștefăniță & Buf, 2021).

Persian phonetic: *"To wari fāheshah o bi haweyat hich vaght dar jameh Afghān hā nist. [Afghān nist.]"*
Translation: "You, the strumpet faceless, have never been among Afghans."

The above abusive and misogynistic comment is against a woman who has joined the "I am not Afghan" campaign. A large part of gender-based hate speech includes vulgar words not only to the person involved in the debate but also to their female family members – as, unfortunately, it is common to abuse female members of one's family in a heated conflict. In the other types of hate speech discussed above, the use of abusive gender-based words towards one's family, such as wife, mother, daughter, and sister, was also evident.

Persian phonetic: *"Fetneh-e akhir pirovān-e Mas'ud hamin ast ke chand zan-e fāheshah rā peydā mikonand tā ed'ā-ye dorughin-e bi nāmusi rā be pish babarand."*
Translation: "The recent sedition of Masoud's followers is precisely this: they find a few prostitute women to present false claims of dishonor."

This example demonstrates the use of gender-based and ethnic hate speech. It has an extremely accusative and judgmental tone toward women who, based on international organization reports, were kidnapped from Kabul streets (UN News, 2022). Those involved in the hate speech not only deny the reality but further dehumanize the subjects of injustice.

Figure 1: Word cloud of hate speech manifestation

The most dominant words shown in Figure 1 are related to derogatory terms for ethnic hate speech [Awghan, Mughl, Landaghar, Kollabi, Heramandawai, Tribal, Mouse-eater], political [Terrorist, Talib, Baghawatgar, hireling], religious [apostate, Jewish, Pagan], and gender-based [whore], hate speech. Vulgar words were also the most frequently used, which were filtered from the word cloud.

DISCUSSION AND CONCLUSION

The concern associated with the exponential proliferation of hate speech on social media needs stakeholders' intervention to mitigate its harm (Chetty & Alathur, 2019; Castaño-Pulgarín et al., 2021; Ștefăniță &Buf, 2021). Studies have noticed that the Internet and social media affordances are susceptible to being misused for spreading hate speech (Ben-David & Fernández, 2016; Schmit et al., 2022), which results in the downgrading, belittling, and dehumanization of an individual or a group, causing harmful psychological and societal impacts (Asimovic et al., 2021; Bilewicz & Soral, 2020; Schmid et al., 2022; Ștefăniță & Buf, 2021; Sugarman et al., 2018). Additionally, hate speech online has been found to lead to hate crimes such as the Pittsburgh synagogue attack and the Christchurch Mosque explosion in New Zealand (Bilewicz & Soral, 2020; Gallacher & Bright, 2021; Mathew et al., 2019). As such, detecting hate speech on social networks

remains an important research topic, as it will help stakeholders initiate mitigative approaches to tackle this ever-growing challenge (Sugarman et al., 2018). Furthermore, unpacking the multidimensional and language-specific features of hate speech will help companies design computer-based algorithms to filter out hate speech content effectively and also enable governments to adopt preventive measures (Ștefăniță & Buf, 2021; Waseem & Hovy, 2016).

This study, in line with previous research in hate speech detection on social media, contributed to the literature in this field and can be considered one of the first of its kind in Afghanistan. Through thematic analysis of a Twitter corpus, this study found that offline contested topics and individuals or social groups involved are the sources of ethnic, political, religious, and gender-based hatred in Afghanistan's Twitter-sphere. The national identity, known as Afghan, regardless of being the official national identity, is one of these inciting hate speech themes. Nationalism in Afghanistan has been an offshoot of British colonialism in the region and the Great Game, which shaped the land named Afghanistan as a buffer zone between the British Indian Empire in the south and the Russian Tsars in the north (Ibrahimi, 2019). Nation-building in the country has been closely associated with ethnic hegemony and coercion without having any democratic foundation (Afzal, 2022; Ibrahimi, 2023; Ibrahimi, 2019). Therefore, this theme has become a source of exchanging hate speech between the proponents and opponents. Except for the Pashtun nationalists (Afzal, 2022; Wafayezada, 2023), who are stern proponents of the generalization of the term "Afghan" to all of Afghanistan's people, there have been voices that demanded a democratic referendum to reach a consensus on using this term or another as the national identity. However, since this term pronounces the specific ethnic group dominance, they consider it uncompromisable. This finding, consistent with previous studies, shows the interconnectedness of hate speech, racism, and ethnocentrism on social media (Matamoros-Fernández & Farkas, 2021).

Another set of contested and hatred-inciting topics is associated with past unpleasant political events and figures, or social groups involved. Events such as a public massacre known as Afshar, which happened in February 1993, and other civil war-era events have incited inter-ethnic conflicts and hate speech until now. Despite being historical facts and having enough evidence of complacency of all sides in the atrocities of the past (Human Rights Watch, 2005), these events have been transmitted through an ethnic-oriented narrative – exonerating self and demonizing others. For instance, every year in February, at the anniversary of Afshar, barrages of hate speech are exchanged on social media, and the users are divided based on ethnic lines, fueling hatred and hostility. Previous studies in post-conflict countries such as Bosnia and Herzegovina found that grievances about past unpleasant historical events have been the source of hatred and polarization (Asimovic et al., 2021).

Finally, the issue of language is another topic of hate speech among Afghanistan's Twitter users. Approximately 80% of Afghanistan's population, regardless of ethnic background, speaks the Persian language, known as Dari. However, in the nation-building process since the 1930s, the Pashtun ruler of the country tried to derail linguistic balance by deliberately engineering this dynamic, including renaming Persian to Dari (Dinakhel, 2019). In 1937, the Pashtu Academy Association was founded to standardize this sparsely spoken language (i.e., Pashtu language) and diverse dialects into a unified language as the country's first language. Since then, the governments of Afghanistan favorably promoted the Pashtu language (Dinakhel, 2019) while

systematically downgrading Persian. In addition, during the past 20 years, the language debate has been a source of intense animosity and hatred, with the government's ethnocentric policies occasionally leading to the removal of the Persian language from university billboards and official event banners (Ahwar, 2023). These ongoing ethnocentric marginalizing practices created deep rifts in the country, fueling hostility on social media. As suggested by previous studies, findings indicate that offline frustration and conflict in many cases and countries may be the driving force of online hate speech (Parvaresh, 2023; Williams et al., 2020). Similarly, social media hate speech can reinforce social fragmentation, leading to a cyclical chain of offline violence and hate crimes (Bilewicz & Soral, 2020; Gallacher & Bright, 2021).

Studies noted that, in countries where ethnic conflicts are prevalent, stigmatization, accusations, marginalization, and stereotyping language are the most common forms of hate speech (Gallacher & Bright, 2021; Jokanovic, 2018). Similarly, in the case of ethnic hate speech among Afghanistan Twitter users, it was found that all ethnic groups, to some extent, were involved in spreading hate speech in a blaming game approach – justifying themselves and blaming others. Findings in Pakistan and Indonesia show the same pattern, where different ethnic and political groups use stereotypical language against each other (Bajari et al., 2021; Batool et al., 2024). Additionally, political and religious hate speech was instigated because of untransparent sectarian politics and intolerant religious beliefs. However, in contrast to other countries in which a single group, mainly minorities, is targeted (Williams et al., 2020), in Afghanistan's Twitter-sphere, every group is the victim and perpetrator of hate speech at the same time. Lastly, gender-based hate speech was also noticeable in the sample of this study, and the most vulgar, abusive, and misogynistic words were used. Afghanistan women, suffering unprecedented social restrictions, have been the target of systematic repression, exclusion, and marginalization since 2021. Meanwhile, female civil society activists and political actors have been repeatedly silenced with vulgar hate speech online.

Recommendations, Research Limitations, and Future Directions

Studies in post-conflict societies have shown that hate speech can quickly plunge a country into chaos (Mathew et al., 2019). This study's findings show that most topics causing online hate speech in Afghanistan are real-world topics with proponents and opponents on the ground. Therefore, the following suggestions are proposed: First, to alleviate the ethno-racial hate speech in Afghanistan, its inciting roots and causes should be handled. Starting with addressing common grievances and reaching a consensus over conflictual issues, including national identity through a public referendum, significantly mitigates relevant hatred online and contributes to establishing a cohesive nation. Besides, the presence of nationalist ethnocentric organizations and individuals at the policy-making level should be restrictedly controlled. Otherwise, the intensity of this phenomenon inflicts unrepairable damage to the country's very existence (Ahady, 1995; Faqiri & Faqirzada, 2021). Second, the unpleasant past cannot be changed, but its repetition can be avoided. Involved parties, through sincere confession, which only needs moral courage, can significantly contribute to social healing and avoid the spread of animosity among the younger generations. Individual parties can use media and other public outreach approaches to communicate their position with the common people honestly. Additionally, through conflict resolution mechanisms, peace-building bodies can mediate and facilitate mutual understanding among involved parties, reaching an agreed-upon position. The resulting agreement can be

included in the educational curricula and discussed in the media, laying the foundation of an unbiased and agreed-upon narrative, breaking the cycle of hostility. On the platform level, companies should design automatic computer algorithms to detect and curb hate speech in other languages, besides English (Matamoros-Fernández & Farkas, 2021).

This study also has some shortcomings. First, the sample is limited to Twitter, which may not fully represent the overall manifestation of hate speech in Afghanistan's online sphere. Future research can adopt a cross-platform approach to enrich these findings further. On the methodological aspect, the current study on the exploratory level provides a general picture of hate speech in the country; however, due to being qualitative in nature as well as limited in scope and context, these findings may not be generalizable to other countries. Cross-language and cross-cultural analysis of hate speech is required to identify the commonalities and differences in its manifestation.

ACKNOWLEDGMENT

The author would like to thank the editor and the two anonymous reviewers for their constructive comments on the earlier draft of this article.

Disclaimer: There are some potentially unpleasant, rude, or upsetting words in this article. They are only included to analyze hate speech for scholarly purposes. The readers' discretion is advised.

BIODATA

Qurban Hussain Pamirzad is a PhD candidate in New Media and Social Governance at the School of Journalism and New Media, Xi'an Jiaotong University, Shaanxi 710049, China. His research interests include social media polarization, online public opinion, and the media landscape in Afghanistan. Email: pamirzad2023@stu.xjtu.edu.cn

REFERENCES

Afzal, O. M. (2022). Sub-nationalism and nationalism: The dilemma of power and ethnicity in Afghanistan. *Beijing Law Review*, *13*, 414–428. https://doi.org/10.4236/blr.2022.132026

Ahady, A. (1995). The decline of the Pashtuns in Afghanistan. *Asian Survey*, *35*(7), 621–634. https://doi.org/10.2307/2645419

Ahwar, J. (2023, August 8). Decoding the Taliban anti-Persianism. *The Diplomat*. https://tinyurl.com/yewavxuf

Asimovic, N., Nagler, J., Bonneau, R., & Tucker, J. A. (2021). Testing the effects of Facebook usage in an ethnically polarized setting. *Proceedings of the National Academy of Sciences*, *118*(25), e2022819118. https://doi.org/10.1073/pnas.2022819118

Bajari, A., Koswara, I., & Erlandia, D. R. (2021). Hatenography: An analysis of hate speech on Facebook in the 2019 Indonesian presidential campaign. *Jurnal Komunikasi: Malaysian Journal of Communication*, *37*(4), 122-141. https://doi.org/gpz562

Batool, S., Aslam, M., & Elahi, E. (2024). Hate speech on social media: A study of user's perspective. *Journal of Development and Social Sciences*, *5*(2), 303-314.

Ben-David, A., & Fernández, A. M. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, *10*, 27.

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, *41*, 3-33. https://doi.org/10.1111/pops.12670

Butler, O., & Turenne, S. (2022). The regulation of hate speech online and its enforcement - A comparative outlook. *Journal of Media Law*, *14*(1), 20-24. https://doi.org/prqm

Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T., & López, H. M. H. (2021). Internet, social media, and online hate speech. Systematic review. *Aggression and Violent Behavior*, *58*, 101608. https://doi.org/10.1016/j.avb.2021.101608

Chetty, N., & Alathur, S. (2019). Racism and social media: A study in the Indian context. *International Journal of Web-Based Communities*, *15*(1), 44-61. https://doi.org/prqn

Cohen-Almagor, R. (2018). Taking North American white supremacist groups seriously: The scope and the challenge of hate speech on the Internet. *International Journal of Crime, Justice, and Social Democracy*, *7*(2), 38-57. https://doi.org/10.5204/IJCJSD.V7I2.517

Dinakhel, M. A. (2019). An analysis of the conflict between Pashto and Dari languages of Afghanistan. *Central Asia*, *83*, 79-99. https://doi.org/10.54418/ca-83.30

Faqiri, S.M., & Faqirzada, A.S. (2021). Social barriers to nation building in Afghanistan. *Shanlax International Journal of Arts, Science and Humanities, 9*(1). https://doi.org/prqp

Fortuna, P., da Silva, J. R., Wanner, L., & Nunes, S. (2019). A hierarchically labeled Portuguese hate speech dataset. *Proceedings of the Third Workshop on Abusive Language Online* (pp. 94–104). https://doi.org/10.18653/v1/W19-3510

Gorenc, N. (2022). Hate speech or free speech: An ethical dilemma? *International Review of Sociology*, *32*(3), 413–425. https://doi.org/10.1080/03906701.2022.2133406

Gallacher, J., & Bright, J. (2021). Hate contagion: Measuring the spread and trajectory of hate on social media. *PsyArXiv*. https://doi.org/10.31234/osf.io/b9qhd

Hakimi, M. J. (2023). The Afghan state and the Hazara genocide. *Harvard Human Rights Journal*, *37*. https://ssrn.com/abstract=4516797

Harrington, C. (2024). Making ethical judgement calls about qualitative social media research on sensitive issues. *International Journal of Social Research Methodology*, 1–13. https://doi.org/10.1080/13645579.2024.2393796

He, G., Adnan, A., Rakha, A., Yeh, H. Y., Wang, M., Zou, X., ... & Wang, C. C. (2019). A comprehensive exploration of the genetic legacy and forensic features of Afghanistan and Pakistan Mongolian-descent Hazara. *Forensic Science International: Genetics*, *42*. https://doi.org/10.1016/j.fsigen.2019.06.018

Human Rights Watch. (2005). Blood-stained hands past atrocities in Kabul and Afghanistan's legacy of impunity. https://www.hrw.org/reports/2005/afghanistan0605/afghanistan0605.pdf

Human Rights Watch. (2022, June 10). Afghanistan: Taliban Torture civilians in Panjshir. https://tinyurl.com/2dp4n3nn

Ibrahimi, S. Y. (2019). Afghanistan's political development dilemma: The centralist state versus a centrifugal society. *Journal of South Asian Development*, *14*(1), 40-61.

Ibrahimi, N. (2023). The state, identity politics and ethnic boundaries in Afghanistan: The case of Sunni Hazaras. *Nations and Nationalism*, *29*(2), 669–685. https://doi.org/prqq

Jokanovic, J. (2018). Hate crime victims in Serbia: A case study of context and social perceptions. *International Journal for Crime, Justice, and Social Democracy*, *7*(2), 21-37. https://doi.org/10.5204/ijcjsd.v7i2.518

Kaltenthaler, K., Kruglanski, A. W., & Knuppe, A. J. (2022). The paradox of the heavy-handed insurgent: Public support for the Taliban among Afghan Pashtuns. *Studies in Conflict & Terrorism*, 1-25. https://doi.org/10.31235/osf.io/2fkvt

Konovalova, E., Le Mens, G., & Schöll, N. (2023). Social media feedback and extreme opinion expression. *PloS One, 18*(11). https://doi.org/10.1371/journal.pone.0293805

Lieven, A. (2023). An Afghan tragedy: The Pashtuns, the Taliban and the State. *Global Politics and Strategy, 63*(3), 7-36. https://doi.org/10.1080/00396338.2021.1930403

Lingam, R. A., & Aripin, N. (2017). Comments on fire! Classifying flaming comments on YouTube videos in Malaysia. *Jurnal Komunikasi: Malaysian Journal of Communication*, *33*(4), 104-118. https://doi.org/10.17576/JKMJC-2017-3304-07

Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media*, *22*(2), 205-224. https://doi.org/10.1177/1527476420982230

Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019, June). Spread of hate speech in online social media. *Proceedings of the 10th ACM Conference on Web Science* (pp. 173–182). https://doi.org/10.1145/3292522.3326034

Miller, M. C. (August 2021). Pakistan's support for the Taliban: What to know. *Council on Foreign Relations*. https://www.cfr.org/article/pakistans-support-taliban-what-know

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods*, *22*. https://doi.org/10.1177/16094069231205789

Nascimento, F. R., Cavalcanti, G. D., & Da Costa-Abreu, M. (2023). Exploring automatic hate speech detection on social media: A focus on content-based analysis. *SAGE Open*, *13*(2).

Parvaresh, V. (2023). Covertly communicated hate speech: A corpus-assisted pragmatic study. *Journal of Pragmatics*, *205*, 63-77. https://doi.org/10.1016/j.pragma.2022.12.009

Perreault Jr, W. D., & Leigh, L. E. (1989). Reliability of nominal data based on qualitative judgements. *Journal of Marketing Research*, *26*(2), 135-148. https://doi.org/gk458t

Schmid, U. K., Kümpel, A. S., & Rieger, D. (2022). How social media users perceive different forms of online hate speech: A qualitative multi-method study. *New Media & Society*, https://doi.org/10.1177/14614448221091185

Shahrani, N. M. (2018). Conflict and peace in Afghanistan: A northern, non-Pashtun perspective. *Accord: An International Review of Peace Initiatives*, *27*, 41-47. https://www.c-r.org/accord/afghanistan/conflict-and-peace-afghanistan-northern-non-pashtun-perspective

Ștefăniță, O., & Buf, D. M. (2021). Hate speech in social media and its effects on the LGBT community: A review of the current research. *Romanian Journal of Communication and Public Relations*, *23*(1), 47-55. https://doi.org/10.21018/RJCPR.2021.1.322

Subay, Ö. Ö., & Gökalp, B. (2023). Hate speech in the context of digital violence in Türkiye: Perceptions and biases on Afghan identity. *Liberal Düşünce Dergisi*, *111*, 93-120. https://doi.org/10.36484/liberal.1314063

Sugarman, D. B., Nation, M., Yuan, N. P., Kuperminc, G. P., Hassoun Ayoub, L., & Hamby, S. (2018). Hate and violence: Addressing discrimination based on race, ethnicity, religion, sexual orientation, and gender identity. *Psychology of Violence*, *8*(6), 649-656. https://doi.org/10.1037/vio0000222

Sungur, Z. T. (2016). Early modern state formation in Afghanistan in relation to Pashtun Tribalism. *Studies in Ethnicity and Nationalism*, *16*(3), 437-455. https://doi.org/10.1111/sena.12211

UN News. (2022, February 1). Six women's rights activists still missing in Afghanistan. https://news.un.org/en/story/2022/02/1111012

Wafayezada, M. Q. (2023). Hybrid extremism: Ethnonationalism and territorialized Islamic fundamentalism in Afghanistan. *The Review of Faith & International Affairs*, *21*(3), 7-21. https://doi.org/10.1080/15570274.2023.2235834

Waseem, Z., & Hovy, D. (2016, June). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. *Proceedings of the NAACL Student Research Workshop* (pp. 88–93). https://doi.org/10.18653/v1/N16-2013

Weismayer, C., & Pezenka, I. (2017). Identifying emerging research fields: A longitudinal latent semantic keyword analysis. *Scientometrics*, *113*(3), 1757-1785. https://doi.org/gcmkrq

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the Machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, *60*(1), 93–117. https://doi.org/c9qh

Williams, M., & Moser, T. (2019). The art of coding and thematic exploration in qualitative research. *International Management Review*, *15*(1), 45-55.