*Corresponding Author's Email: wafayezada@gmail.com
Proceedings of the International Academic Conference on Research in Social Sciences
Vol. 1, Issue. 1, 2024, pp. 36-57
DOI: https://doi.org/10.33422/iacrss.v1i1.768

ISSN: 3030-1289 online

Diamond Scientific Publishing

Polarization in Online Public Discourse: Analyzing Twitter Conversations in Post-2021 Afghanistan

Mohammad Qasim Wafayezada[1]*, Abdul Razaq Vahidi[1], Mohammad Ali Hussiny[2]
[1]Kabul University, Afghanistan
[2]Oslo Metropolitan University

Abstract

Afghanistan has witnessed increasing polarization in the online public discourse in the post-2021 period, highlighting a widening division of perception over critical social and political issues. This research explores the evolving socio-political polarization trends and aims to analyze and categorize Twitter conversations using keyword frequency analysis, clustering algorithms, and trend analysis over time. We use topic modeling in Natural Language Processing (NLP) with Machine Learning (ML) to analyze positive and negative sentiments across predefined themes such as governance, gender, social justice, terrorism/security, and culture. The study reveals significant shifts in the focus of Twitter conversations with varying emphasis on different themes and the interrelation of discussions over time intervals from 15 August 2021 to 9 May 2023. As the first study to pinpoint central issues in Afghanistan's public discourse and polarization trends in the online public sphere, this study highlights key trends and thematic clusters. It offers a timely contribution to understanding the complex dynamics of political discourse in Afghanistan.

Keywords: Education, Public Discourse, Security, Sentiment Analysis, Social Justice

1. Introduction

Afghanistan's fragmented social structure and political culture, in the words of Almond and Verba, has been embedded in its sub-cultures of fragmentation, producing an inbuilt conflict [Almond and Verba, 1956]. Identity politics and persistent traditionalism have been the two major forces of continuity and change in Afghanistan's society and politics that have shaped inter-ethnic and state-society relations. Throughout its modern history, politics of domination and subjugation, coupled with endemic injustices, inequalities, and discrimination, have resulted in the activation of ethnic boundaries and their political mobilization. Over four

decades of continued conflict and violence have deepened the social divides and aggravation of grievances, resulting in increasing polarization across the social and political spectrum. In addition, urban-rural divides and relentless conflict of traditionalist forces with modernization and progressive social change have been an inbuilt feature of Afghan society, causing value polarization.

The rise of radical Islamism in the aftermath of the Soviet invasion gave rise to extremism that, to a certain degree, underpins the tribal social structure and traditional rural layers of society. Value polarization in issues such as women's rights and their access to education and employment have been at the core of social and political contentions, which Sima Samar, a well-known human rights activist, called the 'politicization of women' [Samar, 2019]. The coupling of identity politics and Islamic radicalism has been the major underlying factor in successive failures of state-building and nation-building, resulting in political discontinuity.

In the post-2001 period, the democratization and peacebuilding processes initiated by the Bonn Agreement facilitated the establishment of a more inclusive constitutional political order. The introduction of democratic values and norms highlighted issues of human rights, justice, and particularly women's rights, contributing to a more constructive political culture that transcended ethnolinguistic divisions. However, the failure to decouple Pashtun ethnonationalism from Islamic radicalism hindered efforts to overcome historical challenges and implement effective constitutional nation-building based on shared values and a democratic civic culture.

The collapse of the republican government on August 15, 2021, and the return of the Taliban to power had a profoundly regressive impact on the country's society and politics. The Taliban's ethnonationalism and Islamic radicalism, manifested in their exclusive and repressive rule, caused a cataclysmic shift in public discourse and precipitated a new wave of increasing affective polarization. The imposition of strict, self-interpreted Sharia law and the termination of the 2004 Constitution has been followed by a crackdown on media, freedom of expression, and associations, along with the banning of political parties and the imposition of gender apartheid. This sudden alteration of values and norms in political culture has prompted a significant contextual change, dictated a reconfiguration of public opinion and public discourse, and necessitated a reexamination of issues that were once considered normal and crosscutting values. New patterns of contention and confrontation over such topics have emerged, leading to increasing polarization in Afghanistan's society and politics.

Given the unprecedented restrictions on freedom of expression in Afghanistan, social media has gained more importance in reflecting and amplifying social and political divisions. Social media platforms, especially Twitter, have become the main battleground of opinion and ideas for various Afghan political groups and civic activists. However, access to the internet in Afghanistan is limited, at about 22.9 percent [Internet World Stats, 2023], and only 5.4 million people have accounts on social media platforms [StatCounter, 2022; ITU, 2022]. Despite these limitations, social media significantly shapes public opinion and drives polarization over critical issues. Facebook has been the leading platform, with over 4.4 million users [Kemp, 2023], but due to restrictions on Taliban accounts, Twitter has become the main battleground

for all parties upholding opposing ideas and conflicting values. Klinger aptly argued that social media are not suitable platforms for rational discourse, and if a society's political communication is shifted to social media, one should expect some collateral effect [Klinger, 2021].

As society and politics in Afghanistan remain mired in ethnic politics, politicization of issues pertaining to human rights, and women's rights, active sub-cultures of fragmentation are increasingly generating divisive online content. For example, top-trending hashtags such as #LetHerLearn, #LetHerWork, #StopHazaraGenocide, #FreeAfghanistan, #UnitedAfghanistan, #BanTaliban, #StandWithNRF, #TalibanAreTerrorists, #Peace, #DoNotRecognizeTaliban, etc., manifest a shift in the public discourse and a tense polarization in online conversations. Before the Taliban's return to power, public conversations such as #LetHerLearn indicating a campaign for girls' education would not have been a serious concern or part of the public debate, in the absence of hindering factors. Noting the increasing polarization and grasping the emerging shift in public discourse, this study attempts to explain polarization in political discourse at the intersection of ethnic politics and radicalism. We examine five themes—governance, gender, social justice, terrorism/security, and culture—highlighted in the most popular hashtags shared between August 2021 and May 2023.

2. Literature Review

Studying affective polarization, divisive online content and behavior, as well as their sociopolitical implications, has gained increasing importance in recent years. Affective polarization refers to an individual's emotional attachment and identification with a particular identity, including ethnicity, ideology, or political party, giving rise to favoritism, mistrust, and bias [Druckman, 2019]. While issue-based or opinion-based polarization is a constructive and integral part of democratic deliberation and an essential component of a vibrant and dynamic society [Wafayezada, unpacking], affective polarization evolves into a sharp in-group and out-group division, where individuals dislike and distrust members of another identity group [Iyengar, 2019]. In deeply divided societies, plagued with politicization of identities and social cleavages, ample grounds of discontent and conflict exist, and therefore, opinion-based polarization can metamorphose into affective polarization. Lack of trust in political institutions or the existence of injustices, horizontal inequalities, and discrimination precipitate divisive expression and practices. Social media platforms, providing more personalized features and affordances, contribute to translating the divisions into affective polarization in the online public sphere.

Online divisive content and resulting affective polarization do not happen in a vacuum or in isolation from the social and political context in the physical world. As Schwab [Schwab, 2017] argued, the digital and physical domains are connected, and a meaningful interaction exists between the online and offline domains. The expansion of online platforms and internet-based communications has resulted in the emergence of "virtual communities" [Rheingold, 1991], a term that gained currency in the early stages of internet use [Miller, 2020; Porter, 2004], which denotes a certain networking and interaction pattern in the virtual sphere as an extension of the

physical sphere, however, with virtual characteristics. Today, these virtual communities are an integral part of public spheres, described by Habermas [Habermas, 2019], as an online sphere for discussion and debate of often political and social issues. However, some scholars have expressed concerns over the merger of virtual and real worlds [Putri, 2023]; there is a consensus among scholars and politicians alike that online public spheres are a core component of the political process in every country [Dahlgren, 2005].

In divided societies, online public opinion gets fragmented into 'micro-publics' or public sphericules [Bradley, 2022], which are either emanating from real-world social polarization or amplify those polarizing subjects by projecting them into the social context at large. An important explanation that most of the scholarship on online affective polarization has focused on is linked to 'echo chambers,' where 'opinions are isolated and reinforced' [Garrett, 2009; Colleoni, 2014], as well as homophilic patterns of communication [Tornberg, 2021; Figeac, 2023]. However, this approach to explaining online affective polarization has been challenged by more recent sociopolitical literature, which argues that polarization is driven by ideological partisanship and identity politics [Bright, 2018]. Some experimental studies suggest that polarization is motivated by policy preference rather than identity [Iyengar, 2012]. Nevertheless, in divided societies, where ethnic identity is salient and politically relevant, ideological polarization shifts toward affective polarization, because power politics and economic opportunities and redistribution systems are structured based on such identity boundaries [Orr, 2023].

In addition, social media affordances, such as hashtag indexing, help amplify or signify certain themes and topics, which, in the words of Bruns and Burgess, 'aid in the formation of ad hoc publics' around these themes in the online public sphere [Bruns, 2015]. In this sense, Hashtags represent action/policy priorities or political values of a group that are confronted by an opposing hashtag, which Garimella and Weber called 'hashtag polarization' [Bruns, 2011]. Hashtags have been widely used to highlight issues of common concern, to draw focus and attention, and to turn an issue into a trend. Weber et al. suggested Political Hashtag Trends (PHT) to analyze political polarization [Garimella, 2017]. Another widely used method for the analysis of political polarization and its underlying sentiments is a keyword-based approach. In this method, commonly used words or a set of words on online platforms that are widely debated are identified to analyze the surrounding conversations and their underpinning sentiments [Putri, 2023].

Scholars have used different analytical and algorithmic approaches to analyze affective polarization in online platforms. Yarchi et al. used a combination of discourse content extraction with networked interactions analysis that would enable the tracking of changes in both content and interactions over time [Weber, 2013]. Discourse and networks are closely related in the online public sphere, where public discourses can be as diverse as the echo chambers and homophily groups that constitute numerous mini-networks. An important aspect of affective polarization analysis is its measurement within time and subject clusters. Different methods have been applied to measure affective polarization. Sentiment analysis or opinion mining is a method widely used in recent studies [Yarchi, 2020] for identification and extraction of sentiments embedded in the text using natural language processing and text

mining [Mao, 2024; Ignatow, 2016]. Sentiments can be extracted in simple positive/negative forms [Wankhade, 2022], or on a Likert scale ranging from extremely negative to extremely positive [Mentzer, 2020]. Generally, depending on the type of text and scope of the study, sentiments can be analyzed more inclusively by capturing feelings in a more detailed manner, such as surprise, trust, anticipation, anger, fear, sadness, disgust, joy [Bose, 2020], and other expressible ways of passion and affection.

Agarwal and Mittal noted three main approaches to sentiment analysis: semantic orientation-based approaches, knowledge-based approaches, and machine learning algorithms [Agarwal, 2016]. In the semantic approach, analysis is based on word-level sentiment scores that are extracted from a large data corpus or existing lexicons [Agarwal, 2016]. The knowledge-based approach is a concept-level approach to opinion mining and sentiment analysis that departs from unstructured data and merely textual information to structured and machine-processable data [Cambria, 2013]. Knowledge-based approaches leverage structured representations of information, such as ontologies, taxonomies, and rule-based systems, and, therefore, are more effective in handling complex sentences and domain-specific sentiment nuances. However, some studies have used a combination of both approaches to maximize the efficiency of analytical frameworks for sentiment analysis [Vizcarra, 2021]. In addition, machine learning approaches have also turned into a commonly used method in sentiment analysis. The machine learning approach relies on large datasets of labeled examples (i.e., text annotated with sentiment) to train models. These models learn patterns and correlations between text features and sentiment labels, in which accuracy is based on sufficient training data and proficient annotation and labeling processes. An important merit of the machine-learning approach is its adaptability to different domains and languages [Boiy, 2009].

Usually, researchers undergo a series of pre-processing steps to eliminate non-textual and irrelevant data, feature extraction, and sentiment classification [Liu, 2020]. For languages with substantial linguistic and technological resources, sentiment analysis can be done through existing NLP tools based on pre-developed lexicons and libraries and smoother machine-learning processes. However, for low-resource languages, this process is more complex and usually requires supervised machine learning to capture the proximate polarity of sentiments across units of text within large text corpora [Ignatow, 2016]. Therefore, developing independent and contextualized topic modeling word lists and sentiment polarization datasets can help overcome this challenge. Topic modeling is primarily a form of unsupervised machine learning and a text-mining methodology for identifying topics in documents [Gupta, 2021]. However, scholars tend to use semi-supervised [McAuliffe, 2007] or supervised topic modeling [Lu, 2008; Zhu, 2009; Sridhar, 2022] that combine elements of topic modeling with supervised learning. In the study, as elaborated further in the methodology and resources, we have adopted a hybrid approach, benefiting from the existing literature and considering the challenges in conducting research in Farsi/Dari as a low-resource language. Our focus when developing the topic modeling and polarization dataset was their capability to deal with the correlation between topics and the changes in topics over time, as highlighted by Vayanski and Kumar in their review of topic modeling methods [Vayanski, 2020].

3. Resource and Data

The data was collected from the X social network using an API with the Tweepy library and the Python programming language. The tweet collection focused on top trending hashtags related to various topics, including education, women's rights, social justice, government, security and terrorism, and culture, posted between August 2021 and May 2023. The introduction of new restrictive policies by Twitter in data-sharing hindered our efforts to extend the data collection period to mid-2024 in the absence of necessary financial resources. However, the collected data in terms of time span and coverage of thematic variety was sufficient and resulted in approximately 3.5 million tweets. The collected data was stored and filtered to remove duplicates, reducing the dataset to 2,166,535 unique tweets.

3.1 Data Preprocessing

We performed data filtering on the collected tweets to remove extraneous information such as URLs, external links, hashtags, mentions, and HTML tags. These elements were excluded because they do not contribute to a comprehensive textual analysis. Duplicate tweets were also removed to ensure the integrity and accuracy of the dataset.

3.2 Developing topics modeling Sets

A significant challenge for researchers working with low-resource languages like Dari is the inadequacy of existing lexicons or datasets in capturing the full spectrum of slang and expressions used to convey positive and negative sentiments. Furthermore, varying literacy levels among online users lead to the frequent incorporation of local slang, unique expressions, and the interchangeable use of Pashto and Dari within Afghanistan's bilingual online spaces. To address these challenges, this study developed two distinct datasets: one for topic modeling and another for polarity evaluation.

Our methodology for developing the dataset began by reviewing a random sample of at least 1,000 tweets in Farsi/Dari to ensure a comprehensive and representative analysis. Each tweet was carefully read and analyzed, allowing us to identify five key thematic categories: Governance, Gender, Social Justice, Terrorism/Security, and Culture. These categories were selected based on their prevalence and relevance to the data, ensuring that they align with critical socio-political issues in Farsi/Dari speaking regions.

After establishing the five categories, the next step involved a detailed analysis of the tweets. By incorporating Llama3, we leveraged its capabilities to assist in categorizing the tweets. Specifically, Llama3 was tasked with assigning words from the tweets into the relevant categories—Governance, Gender, Social Justice, Terrorism/Security, and Culture—based on the predefined categories.

Following categorization based on Llama3, we cleaned the developed dataset to ensure accuracy. This process involved removing irrelevant or redundant information, such as tweet replies that included descriptions or unrelated content. To ensure consistency and accuracy in the annotation process, we developed a comprehensive set of instructions for the human

annotators. These guidelines were carefully designed to cover all aspects of the five thematic categories, ensuring that the annotators understood the nuances of each category. The annotators were then tasked with verifying and refining the categorization by reviewing the set of words or phrases for each category. This word set was constructed to reflect the dimensions and subtleties of each category, ensuring that the dataset would effectively capture the thematic content relevant to governance, gender, social justice, terrorism/security, and culture.

An overall 18,000 randomly picked tweets were used to create the word list in our topic modeling for all five thematic categories. As depicted in **Table 1** on sample words, the five thematic categories in our topic modeling included a list of words from the user-generated data and online discourse and were evaluated as relevant to each category by annotators.

Table 1: Sample words for thematic categories with Persian and English translations

Thematic Categories	**Total word list**	**English translation of the Samples**	**Samples in Persian language**
Governance	3005	Hegemony, executive, judiciary, policy, law-enforcement, parliament, constitution, regulations, order, corruption, development, reform, obligatory, restriction, lawlessness, central	هژمونی، اجرایی، قضائیه، پالیسی، قانون گذاری، پارلمان، قانون اساسی، مقررات، نظم، فساد اداری، توسعه، اصلاحات، اجباری، محدودیت، بی قانونی، مرکزی
Gender	580	Women, rights, equality, right to work, right to education, girls, education, sex, paternalism, misogyny, deprivation	زنان، حقوق، برابری، حق کار، حق آموزش، دختران، تحصیل، جنسیتی، مردسالاری، زن ستیزی، محرومیت
Social Justice	3686	Apartheid, freedom, humanitarian, fairness, justice, ethnic groups, minorities, equality, rights, oppressor, oppressed, deprived, adjudication	آپارتاید، آزادی، انسانیت، انصاف، عدالت، اقوام، اقلیت های قومی، برابری، حقوق، ظالم، ستمگر، مظلوم، محروم، احقاق
Terrorism/ Security	2170	Suicidal, explosion, murder, imprisoned, incident, missing, security, Talib, terror, armed, force, intimidation, threatening, police, security forces	انتحاری، انفجار، کشتن، زندانی، ناپدید، حادثه، امنیت، ترور، طالب، مسلح، زور، تهدید، ارعاب، پولیس، قوای امنیتی
Culture	3999	Farsi, Dari, Pashtu, Uzbeki, language, national, history, past, cultural, heritage, dance, music, singer, science	فارسی، پشتو، ازبیکی، زبان، ملی، تاریخ، پیشینه، فرهنگی، آثار، رقص، موسیقی، آوازخوان، دانش

3.3 Polarization Dataset

To identify polarization in tweets and due to the lack of available sentiment analysis datasets with binary labels, we developed a specialized sentiment dataset with two classes: positive and negative.

Due to time constraints and limited human resources, we randomly selected a sample of 8,650 records for annotation. The annotation was conducted by two native Persian speakers, who diligently classified each tweet as either positive or negative in sentiment. The dataset consists of 4,771 negative and 3,879 positive labels, providing a balanced and reliable resource for training and evaluating sentiment analysis models.

Table 2: Sample of the Sentences with Negative and Positive Sentiments

Sentiment	Original Sentence (Persian)	English Translation
N1	دیروز دختر بچه های هرات گریه کردند جهان نگاه کرد	Yesterday, Herati girls cried, but the world only observed!
N2	اینجا در دردها مشترک نیست هم وطن تنها یک کلمه است و بقول آن شاعر اینجا درد هر کس بخودش مربوط است	Here, the pains are not shared; 'compatriot' is just a word. And as that poet said, here, everyone's pain is their own concern.
P1	افغانستان یک کشور اسلامی است تاجک پشتون هزاره ازبک و همه اقوام باهم برادر هستیم لطفا از قتل و خونریزی جلوگیری کنید	Afghanistan is an Islamic country. Tajiks, Pashtuns, Hazaras, Uzbeks, and all ethnic groups are brothers. Please prevent killing and bloodshed.
P2	یک دست صدا نداره باید همه باهم جلوی ای وحشی های ایستاد شویم وتعصب را کنار بگذاریم	One hand alone doesn't make a sound; we must all stand together against these savages and set aside prejudice.

4. Methodology

As explained above, we developed two datasets in this study: 1) the topic modeling dataset and 2) the polarization dataset. These datasets aid in thoroughly analyzing the data and capture sentiments in Dari, a low-resource language with limited computational analysis studies. K-means clustering was also utilized to categorize the user-generated data into thematic and time clusters.

The topic model was used to analyze tweet data and extract keywords related to various topics. Initially, raw tweet data was collected and carefully preprocessed to ensure accuracy and consistency [Blei, 2003]. Using the topic model, keywords related to each category were identified, and the occurrences of these keywords in each tweet were counted. This information was added as new columns to the tweet data. The process involved defining appropriate functions for counting keywords and using regular expressions to increase counting accuracy, as Sasaki has aptly noted [Sakaki, 2010].

4.1. K-Means Clustering

This study applied K-means clustering to analyze and categorize the tweets into different clusters, as explained by [Pedregosa, 2011]. The process was as follows:

4.1.1. Feature Selection: Columns related to the keyword counts for each category were selected.

4.1.2. Data Standardization: The data was standardized to improve clustering.

$$Z = \frac{X - \mu}{\sigma}$$

where:

- Z is the standardized value
- X is the original value
- μ is the mean of the data
- σ is the standard deviation of the data

4.1.3. K-Means Clustering: The standardized data was divided into 5 clusters using the K-means algorithm.

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} \left\| x_j^{(i)} - \mu_i \right\|^2$$

where:

- J is the objective function
- k is the number of clusters
- n is the number of data points
- $x_j^{(i)}$ is the jjj-th data point in the iii-th cluster
- μ_i is the centroid of the iii-th cluster

4.1.4. Principal Component Analysis (PCA): To visualize the clusters, the data was reduced to two principal components.

$$Z = XW$$

where:

- Z is the matrix of principal components
- X is the standardized data matrix
- W is the matrix of eigenvectors of the covariance matrix of X

5. Data Analysis

5.1 Cluster Analysis of Social and Political Tweets

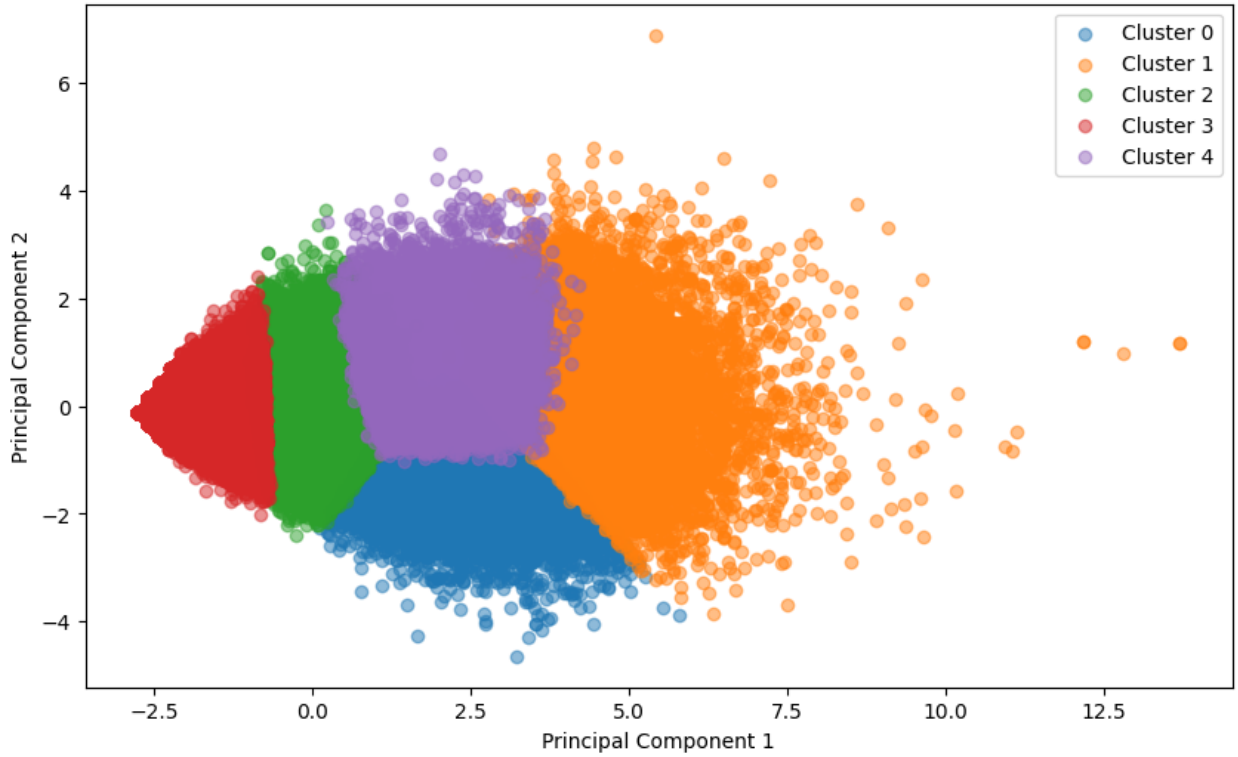
Table 3 displays the clustering outcomes of tweet data using the K-Means Clustering algorithm and Principal Component Analysis (PCA). Each column represents a cluster, and the rows show the mean, standard deviation, and count of tweets in each category (i.e., Governance, Gender, Social Justice, Terrorism/Security, and Culture).

Table 3: Corrected Cluster Analysis Results

Category	Metric	Cluster 0	Cluster1	Cluster2	Cluster3	Cluster4
	Count	156,779	94,363	486,009	672,981	202,522
Governance	Mean	10.96	12.99	5.45	1.97	9.17
	Std	3.25	3.71	2.39	1.52	3.19
Gender	Mean	3.23	7.61	2.76	1.06	5.91
	Std	1.5	2.58	1.46	1.02	1.93
Social Justice	Mean	9.04	13.47	5.27	2	9.27
	Std	2.66	3.23	1.91	1.39	2.55
Terrorism/Security	Mean	7.2	7.59	2.82	1.04	3.72
	Std	2.12	2.81	1.53	1.05	1.62
Culture	Mean	7.33	11.84	5.31	2.19	7.95
	Std	2.77	3.38	2.37	1.53	2.85

Figure 1, resulting from PCA and K-Means clustering, shows the tweets divided into five clusters. Each point in the graph represents a tweet, categorized into one of the clusters based on the occurrence of different keywords. Different colors (blue, orange, green, red, and purple) represent the five clusters. The horizontal axis (Principal Component 1) and the vertical axis (Principal Component 2) represent the two principal components that best visualize our data.

Figure 1: K-Means Clustering of Tweet Categories



After segmenting user tweets into five clusters, it becomes evident that **Cluster 3**, which contains the most data, records the lowest averages across various categories such as Governance, Gender, Social Justice, Terrorism/Security, and Culture. This cluster engages in expansive yet unfocused discussions, reflecting a broad discourse in which specific topics do not receive intense focus. Such a pattern could symbolize the intricate and diverse socio-political fabric of Afghanistan, where multiple issues vie for attention concurrently. Notably, the minimal averages of 1.97 for Governance and 1.04 for Terrorism/Security in **Cluster 3** suggest a widespread engagement with these vital subjects, crucial in the context of Afghanistan's persistent governance and security challenges. This broad focus might stem from a fragmented media landscape and varied public awareness or access to information on these issues.

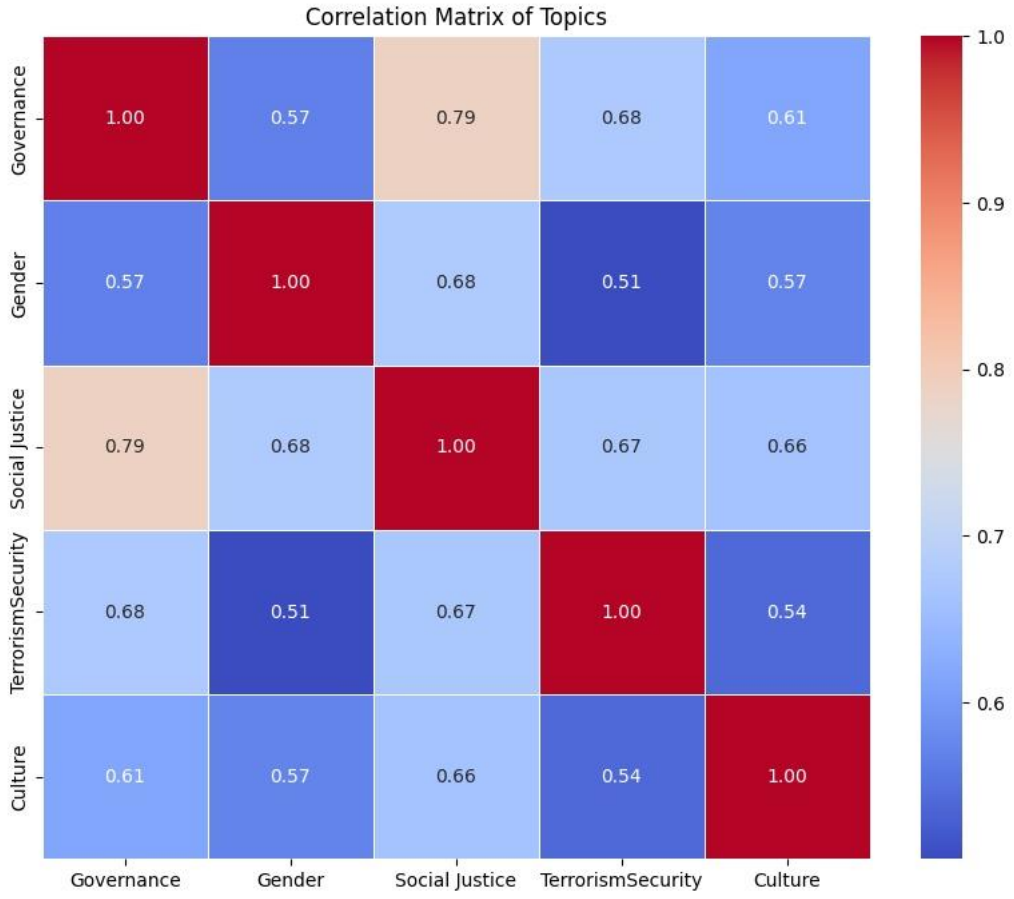
In contrast, users in **Clusters 1 and 4** demonstrate more significant engagement with Governance, Social Justice, and Culture. For example, **Cluster 1** exhibits the highest averages for Governance (12.99) and Social Justice (13.47), indicating a more concentrated dialogue among possibly more knowledgeable or active demographics. These interactions may center on critical discussions about reforms, civil rights, or cultural preservation, which are vital for nation-building and enhancing social cohesion within Afghanistan. Additionally, **Cluster 4**, with notable averages for Culture (7.95) and Governance (9.17), suggests focused discussions on cultural identity and governance issues, highlighting the societal drive towards preserving cultural heritage and enhancing governance frameworks.

Cluster 0 is predominantly focused on Governance (10.96) and Social Justice (9.04), possibly reflecting conversations among activists or policy influencers who place a high priority on these areas due to their substantial influence on the nation's legal and social structures.

Overall, the data distribution and thematic emphases within these clusters likely reflect Afghanistan's socio-political dynamics. Clusters with heightened attention to Governance and Social Justice (**Clusters 1 and 4**) may reflect the deep involvement of certain societal segments directly impacted by these issues. Meanwhile, the dispersed focus of **Cluster 3** signifies the general population's extensive but shallow engagement with key national concerns. This analysis provides nuanced insights into how different segments of the Afghan population interact with critical issues at varying levels of intensity.

The correlation matrix in **Figure 2** demonstrates the interconnections between various topics. The values range from 0 to 1, where 1 indicates a perfect correlation (topics are strongly related) and lower values indicate weaker correlations.

Figure 2: Correlation Matrix among various topics



Governance and Social Justice have the highest correlation (0.79), suggesting that discussions about governance often overlap with issues of social justice. Governance also shows a moderate correlation with Terrorism/Security (0.68) and Culture (0.61), indicating these topics are frequently discussed together. Gender has a moderate correlation with Social Justice (0.68)

and a lower correlation with Terrorism/Security (0.51), reflecting that gender issues are often discussed in the context of social justice, but less so with security concerns. These correlations provide insights into the interconnectedness of public discourse themes in the context of Afghanistan's socio-political environment, as discussed in your broader research.

This matrix is a valuable tool for understanding the dynamics of public discussions and how different themes interact within the discourse. It can help identify key areas where multiple issues converge, potentially guiding future research or policy-making efforts to address these interlinked topics effectively.

5.2 Trend Analysis of Topics

The next step involves analyzing trends in various categories (Governance, Gender, Social Justice, Terrorism, Security, and Culture) over time. Tweets spanning from August 2021 to May 2023 are filtered monthly. For each month, the percentage of tweets related to each category is calculated by summing the keyword counts for each category and dividing by the total number of tweets.

Table 4: Tweet Counts from August 2021 to May 2023

Period	Aug-21	Sep-21	Oct-21	Nov-21	Dec-21	Jan-22	Feb-22	Mar-22	Apr-22	May-22	Jun-22	Jul-22	Aug-22	Sep-22	Oct-22	Nov-22	Dec-22	Jan-23	Feb-23	Mar-23	Apr-23	May-23
Tweet Counts	169,923	153,878	50,855	38,104	41,500	40,642	34,454	34,605	47,761	40,745	33,710	31,643	37,415	35,729	336,220	170,823	107,815	65,946	40,134	40,962	45,634	14,146

In general, the **average score (AS)** represents the overall relevance (sentiment) of data points connected to a particular topic. Specifically, in our research, the average score (AS) shows how positively or negatively a topic is perceived by users according to the analyzed tweets.

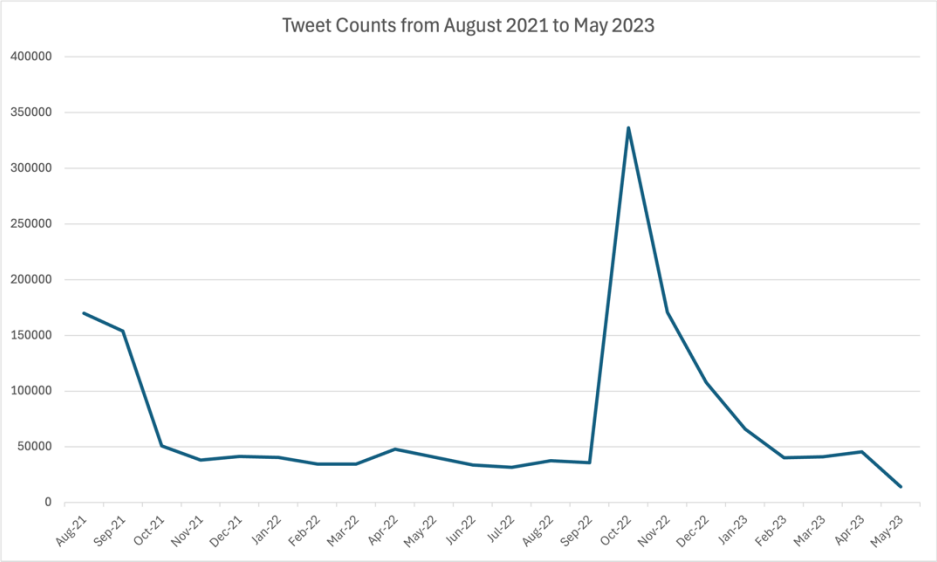
The Table 4 shows the average scores for five categories across the specified periods. Graphs show that in October 2022, there was a notable surge in the volume of tweets, yet the AS for various topics remained low. If a large number of tweets are published on various topics, but none of these topics are particularly prominent, it can result in a lower AS. In other words, attention to multiple and scattered topics can keep the AS low. Specific events may have occurred, generating a large number of tweets, but these tweets do not relate to the topic models. In October 2022, the trending hashtag "#StopHazaraGenocide" significantly influenced the topic models [BBC, 2022; AFINTL, 2022]. This indicates that highly sensitive and important issues can dominate online discussions, shape the direction of topic analysis, and reflect the collective response to a humanitarian crisis.

Table 5: Average Scores from September 2021 to May 2023

Period	Governance	Gender	Social Justice	Terrorism/Security	Culture
Sep-21	5.96	2.92	5.55	3.29	5.14
Oct-21	6.37	2.91	5.74	3.33	5.07
Nov-21	6.86	3	6.04	3.43	5.17
Dec-21	6.79	2.87	5.86	3.1	5.05
Jan-22	6.66	3.08	5.89	3.04	5.16
Feb-22	6.81	2.98	6.06	3.23	5.32
Mar-22	6.9	3.21	6.26	3.12	5.65
Apr-22	6.52	3	6.13	3.65	5.58
May-22	6.83	3.28	6.29	3.81	5.78
Jun-22	6.82	3.02	6.16	3.48	5.79
Jul-22	6.38	2.95	5.62	3.27	5.28
Aug-22	6.75	3.06	6.13	3.69	5.48
Sep-22	6.16	3.08	5.89	3.21	5.29
Oct-22	3.85	2.29	4.21	2.28	4.17
Nov-22	3.83	2.38	4.32	2.19	4.5
Dec-22	5.11	3.26	5.45	2.66	4.98
Jan-23	5.48	2.95	5.39	2.72	5.01
Feb-23	6.29	2.95	5.68	3.03	5.32
Mar-23	6.48	2.96	5.77	2.98	5.36
Apr-23	6.61	3.16	5.92	3.11	5.41
May-23	6.4	3.24	6.03	3.12	5.33

This issue is also observed in months with high tweet counts, such as August 2021 (169,923 tweets), September 2021 (153,878 tweets), October 2022 (336,220 tweets), November 2022 (170,823 tweets), and December 2022 (107,815 tweets). Although the AS is independent of the number of tweets, it reflects how deeply or significantly specific topics are addressed in the tweets. These scores are calculated independently of the total number of tweets in a given period and depend more on the intensity of focus on specific topics.

Figure 3: Tweet Counts from August 2021 to May 2023

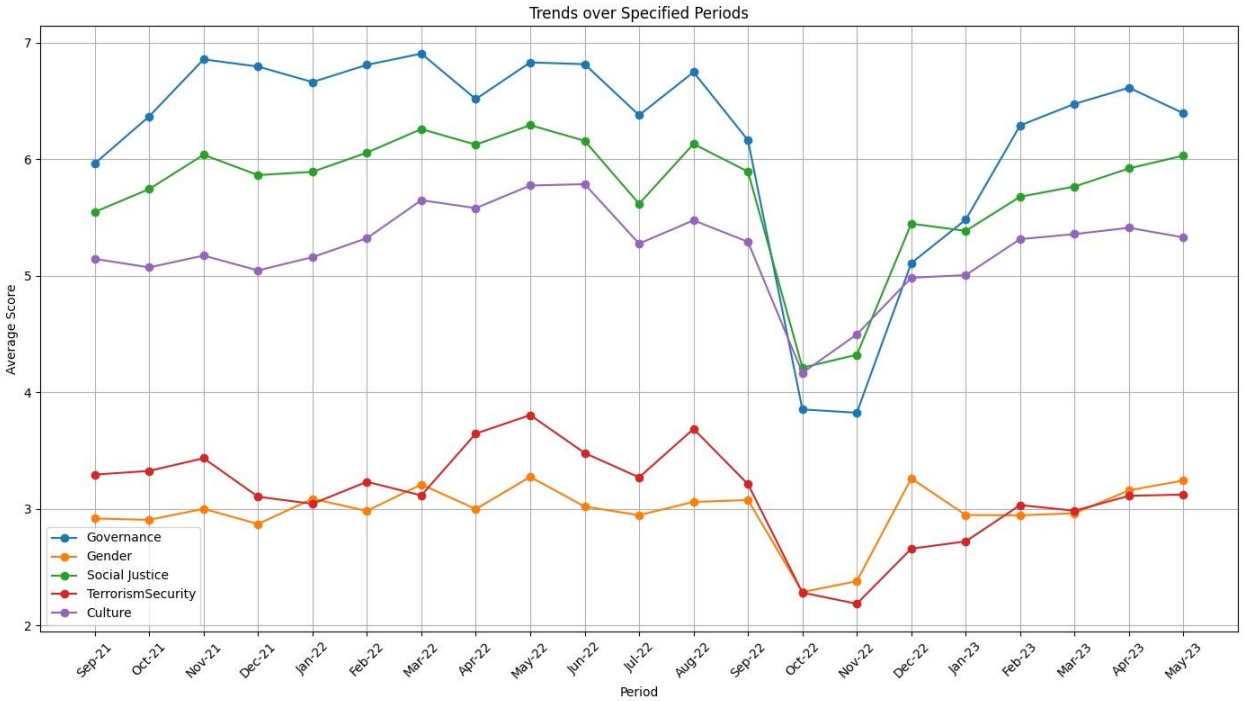


In months with more tweets, if significant new topics are not introduced, we may observe the following effects on the AS:

- **Attention Dispersion:** If the number of tweets increases but the focus remains on similar topics, this attention may be divided among the existing topics. This can cause the AS for each topic to decrease, as tweets may speak less about the deeper or newer aspects of each topic.
- **Increase in Noise:** An increase in the number of tweets without adding new topics might mean an increase in repetitive or less important content. This can lower the quality of the AS for existing topics, as a large amount of content may have low informational value.
- **Averaging Effect:** With the increase in the number of tweets, the presence of repetitive or low-impact individual scores, which might be part of the general discourse without specific emphasis on precise topics, may affect average scores and pull them down.

Ultimately, if the content of tweets in high-traffic months mainly includes superficial or general discussions without special emphasis on newer or deeper aspects of topics, the AS might remain low.

Figure 4: Trends over Specified Periods



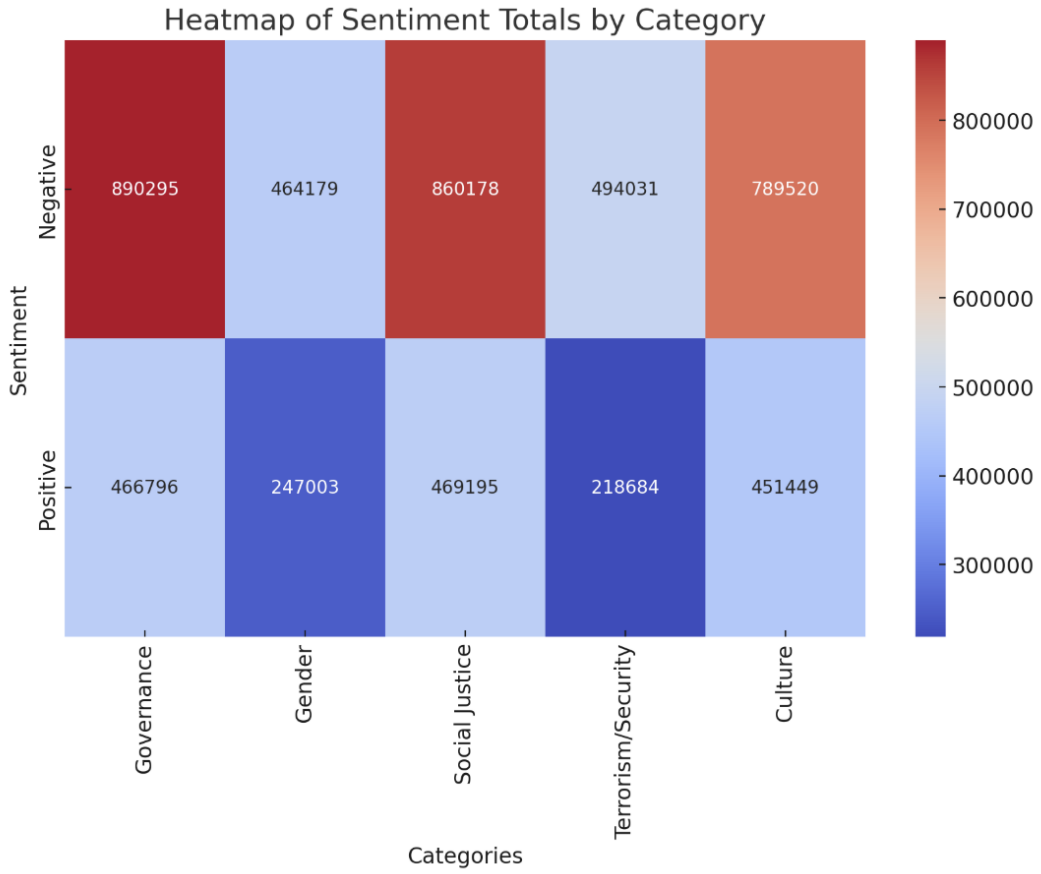
The comparison between the tweet counts and average scores highlights how significant political events, policy changes, and international reactions drive social media activity and public sentiment. The sharp fluctuations in tweet volumes and scores reflect the dynamic and often volatile situation in Afghanistan, with major events leading to spikes in discussions and shifts in public opinion across various categories.

- **High Tweet Volume in August 2021:** The high number of tweets in August 2021 corresponds with the Taliban takeover of Kabul, leading to extensive social media activity discussing the political and humanitarian implications. This period shows high scores in Governance and Social Justice, reflecting the intense global focus on these issues during the power transition.
- **Steady Tweet Volumes with High Scores:** From October 2021 to August 2022, the number of tweets remains relatively stable, while the average scores for Governance, Social Justice, and other categories stay high. This suggests ongoing discussions and concerns about the new Taliban regime's policies and their impact on Afghan society.
- **The spike in tweet volume in October 2022 (~336,000 tweets) correlates with significant events such as the trending hashtag "#StopHazaraGenocide."** This period shows a sharp decline in scores across all categories, indicating widespread negative sentiment and criticism.
- **Decline and Recovery (November 2022 - May 2023):** After the spike, tweet volumes decline, and average scores start to recover. This trend could be attributed to the initial shock and reaction settling down, with ongoing international humanitarian efforts and partial stabilization under the new regime contributing to the recovery in public discourse and sentiment.

5.3 Sentiment Polarity Evaluation

Sentiment polarization analysis depicts a highly polarized social and political landscape across almost all thematic categories, such as governance, gender, social justice, terrorism/security, and culture. As shown in Figure 5, negative sentiments significantly outweigh positive sentiments in governance and social justice, followed by culture, gender, and security. However, spikes in certain months induce higher levels of negative sentiments. Generally, negative sentiment persists with varying fluctuation levels throughout the data collection period (September 2021 to May 2023), highlighting the public's reaction to the sociopolitical dynamics in a period marked by significant transitional events.

Figure 5: Heatmap of sentiment totals by category



5.3.1 Prominent Spikes in Activity:

As mentioned in subsection 4.4, October 2022 experienced a notable surge in tweet activity, especially in the Governance and Culture categories. This likely reflects a significant event or public discussion that attracted widespread attention in these areas.

5.3.2 Sentiment Polarity Evaluation Across Topics

Sentiment Distribution: During this peak, the sentiment was predominantly negative, suggesting that the events or topics discussed might have been controversial or unfavorable among the public.

- **Governance:** Negative sentiments in governance peaked around October 2022, possibly correlating with specific governmental actions or international diplomatic events affecting Afghanistan.
- **Gender:** Discourse around gender shows persistent negativity, reflecting ongoing concerns about gender rights and the societal position of women under changing regimes.
- **Social Justice:** Social justice tweets spiked in negativity during legislative or executive decisions of the Taliban's *de facto* authority, indicating public concern over justice and equality.
- **Terrorism/Security:** This theme saw heightened negative sentiments during incidents of violence or security threats.
- **Culture:** Cultural discussions also exhibited a negativity bias, potentially due to the imposition of restrictive cultural norms and regulations by the ruling authorities.

The pronounced negativity across all themes suggests a general dissatisfaction among Twitter users concerning the trajectory of sociopolitical developments in Afghanistan. This digital discourse analysis reflects the immediate reactions to political events and serves as a barometer for more profound societal and cultural shifts.

6. Conclusion

The collapse of the republic government on 15 August 2021 marked the failure of state-building and the democratization process, leading to the return of the Taliban and exposing deep-rooted tensions within the country. Polarization in political discourse occurs along the lines of Talibanism versus opposing ideas that support a more progressive, democratic, constitutional political order and inclusive politics. These polarized discourses are not limited to the opposing political parties but extend deep into society in the online sphere. Such politics of polarization play on the social fault lines of ethnolinguistic partisanship and identity issues. Social media and the online public sphere negatively influence the width and breadth of polarization built on these sub-cultures of fragmentation in Afghanistan's society and politics. Online platforms have gained more significance and importance in the post-2021 period, mainly due to the repressive policies of the Taliban that curtailed freedom of expression and civic and political activism within the country, driving Afghan citizens to the digital realm to voice their opinions and share ideas.

We developed two datasets, including the topic modeling dataset and a polarization dataset, through a meticulous annotation process. Since Dari is considered a low-resource language, and existing lexicons or datasets for polarization do not adequately capture the range of slang and expressions used to convey positive and negative sentiments, the two datasets also contribute to future studies on sentiment analysis and studying discourse polarization in Afghanistan.

As this study highlights, political polarization overwhelmingly extends to all five thematic areas of governance, social justice, gender, security and terrorism, and culture. The nature of

polarization is often sentimental and affective. For example, polarization over girls' education and women's right to work, as universal values enshrined and protected by the 2004 Afghanistan constitution, manifests not only value polarization but also the sentimental nature of opposition to that right. Our overtime trend analysis shows spikes that coincide with some major online activism sparked by bitter ground realities. The first spike is the heated debates in the immediate aftermath of the government collapse, followed by a major spike resulting from the trending hashtag #StopHazaraGenocide. Affective polarization is more explicit in topics involving ethnicity and identity issues.

This study reveals a stark reality: negative sentiments far outweigh positive sentiments across all five categories. This overwhelming negativity is a clear manifestation of the increasing frustration, despair, and dissatisfaction with the status quo and the ongoing situation in Afghanistan. It underscores the gravity of the situation and how social media users sense the urgent need for meaningful political change.

The study reveals that polarization over all five categories has consistently continued over the time span of our data collection. Observable spikes in the time cluster indicate the relationship between online and offline public spheres and how ground realities and developments lead to increased polarization on Twitter. Banning girls' schooling and women's employment, increasing targeting of former government officials and security personnel, continued deliberate attacks on Hazaras, etc., are reflected in the increased online polarization and spikes in Twitter.

As the first step in delving into the nature and expanse of polarization in Afghanistan's online public discourse, this research contributes to understanding the post-2021 sociopolitical context. However, resource limitations such as NLTK tools, tokenization capabilities within the Persian data, and the lack of access to granular metadata restricted our ability to analyze temporal trends, contextual relationships, and region-specific patterns. Therefore, we recognize the necessity of conducting a network analysis of the online public sphere and discourse polarization in future studies.

References

Almond, G. A. (1956). Comparative political systems. The Journal of Politics, 18(3), 391–409. https://doi.org/10.2307/2127255

Samar, S. (2019). Feminism, peace, and Afghanistan. Journal of International Affairs, 72(2), 145–158.

Internet World Stats. (2023). Internet World Stats.

Statcounter. (2022). Social media stats Afghanistan: Nov 2021-Nov 2022.

International Telecommunication Union. (2022). Digital development dashboard, infrastructure, and access in Afghanistan.

sKemp, S. (2023). Digital 2023: Afghanistan.

Klinger, U. (2021). Digital democracy and public discourse: Dissonant, disrupted and unedited? Behind the Headlines, CIC, 69(26).

Druckman, J. N., & Levendusky, M. S. (2019). What do we measure when we measure affective polarization? Public Opinion Quarterly, 83(1), 114–122. https://doi.org/10.1093/poq/nfz003

Wafayezada, M. Q. (2024). Unpacking affective polarization in Afghanistan: Ethnic politics, elite competition, and online divisive content. Policy Brief No. 193. Toda Peace Institute.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. Annual Review of Political Science, 22(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Schwab, K. (2017). The fourth industrial revolution. Crown Currency.

Rheingold, H. (1991). The great equalizer. Whole Earth Review, 71, 88–94.

Miller, D., & Slater, D. (2020). The internet: An ethnographic approach. Routledge.

Porter, C. E. (2004). A typology of virtual communities: A multi-disciplinary foundation for future research. Journal of Computer-Mediated Communication, 10(1), 00–00. https://doi.org/10.1111/j.1083-6101.2004.tb00228.x

Habermas, J. (2019). The public sphere: An encyclopedia article (1974). In Crime and media (pp. 11–19). Routledge.

Putri, N. A. D., Setiyono, B., Yuwono, T., & Erowati, D. (2023). Affective polarization: Not always between ingroup vs outgroup (Evidence from Twitter conversation with keywords Jokowi and PDIP). In E3S Web of Conferences (Vol. 440, p. 03013). EDP Sciences. https://doi.org/10.1051/e3sconf/202344003013

Dahlgren, P. (2005). The internet, public spheres, and political communication: Dispersion and deliberation. Political Communication, 22(2), 147–162. https://doi.org/10.1080/10584600590933160

Bradley, M., & Chauchard, S. (2022). The ethnic origins of affective polarization: Statistical evidence from cross-national data. Frontiers in Political Science, 4, 920615. https://doi.org/10.3389/fpos.2022.920615

Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among internet news users. Journal of Computer-Mediated Communication, 14(2), 265–285. https://doi.org/10.1111/j.1083-6101.2009.01440.x

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. Journal of Communication, 64(2), 317–332. https://doi.org/10.1111/jcom.12084

Tornberg, P., Andersson, C., Lindgren, K., & Banisch, S. (2021). Modeling the emergence of affective polarization in the social media society. PLOS ONE, 16(10). https://doi.org/10.1371/journal.pone.0258259

Figeac, J., & Favre, G. (2023). How behavioral homophily on social media influences the perception of tie-strengthening within young adults' personal networks. New Media & Society, 25(8), 1971–1990. https://doi.org/10.1177/14614448231152749

Bright, J. (2018). Explaining the emergence of political fragmentation on social media: The role of ideology and extremism. Journal of Computer-Mediated Communication, 23(1), 17–33. https://doi.org/10.1093/jcmc/zmy003

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. Public Opinion Quarterly, 76(3), 405–431. https://doi.org/10.1093/poq/nfs038

Orr, L. V., et al. (2023). Is affective polarization driven by identity, loyalty, or substance? American Journal of Political Science, 67(4), 948–962. https://doi.org/10.1111/ajps.12589

Bruns, A., & Highfield, T. (2015). Is Habermas on Twitter?: Social media and the public sphere. In A. Bruns et al. (Eds.), The Routledge companion to social media and politics (pp. 56–73). Routledge.

Bruns, A., & Burgess, J. (2011). The use of Twitter hashtags in the formation of ad hoc publics. In Proceedings of the 6th European Consortium for Political Research (pp. 1–9). Colchester: The European Consortium for Political Research.

Garimella, V. R. K., & Weber, I. (2017). A long-term analysis of polarization on Twitter. In Proceedings of the eleventh international AAAI conference on web and social media (ICWSM) (pp. 528–531). AAAI.

Weber, I., et al. (2013). Political hashtag trends. In ECIR 2013, LNCS 7814 (pp. 857–860). Springer.

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. Political Communication, 38(1–2), 98–139. https://doi.org/10.1080/10584609.2020.1811901

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University - Computer and Information Sciences, 36(4). https://doi.org/10.1016/j.jksuci.2023.07.019

Ignatow, G., Evangelopoulos, N., & Zougris, K. (2016). Sentiment analysis of polarizing topics in social media: News site readers' comments on the Trayvon Martin controversy. In J. Smith (Ed.), Communication and information technologies annual (Vol. 11, pp. 259–284). Emerald Group Publishing Limited.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55, 5731–5780. https://doi.org/10.1007/s10462-022-10185-x

Mentzer, K., et al. (2020). Measuring and unpacking affective polarization on Twitter: The role of party and gender in the 2018 senate races. Computer Information Systems Journal Articles, 30.

Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2020). Sentiment analysis on online product reviews. In Information and communication technology for sustainable development: Proceedings of ICT4SD 2018 (pp. 559–569). Springer.

Agarwal, B., & Mittal, N. (2016). Machine learning approach for sentiment analysis. Springer.

Agarwal, B., & Mittal, N. (2016). Semantic orientation-based approach for sentiment analysis. In Prominent feature extraction for sentiment analysis (pp. 77–88). Springer.

Cambria, E., Schuller, B., Liu, B., Wang, H., & Havasi, C. (2013). Knowledge-based approaches to concept-level sentiment analysis. IEEE Intelligent Systems, 28(2), 12–14. https://doi.org/10.1109/MIS.2013.40

Vizcarra, J., Kozaki, K., Ruiz, M. T., & Quintero, R. (2021). Knowledge-based sentiment analysis and visualization on social networks. New Generation Computing, 39, 199–229. https://doi.org/10.1007/s00354-021-00090-z

Boiy, E., & Moens, M.-F. (2009). A machine learning approach to sentiment analysis in multilingual web texts. Information Retrieval, 12, 526–558. https://doi.org/10.1007/s10791-008-9070-z

Liu, B. (2020). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.

Gupta, S., & Sharma, G. (2021). Topic modeling in natural language processing. International Journal of Engineering Research & Technology (IJERT), 10(6), 1–10.

Mcauliffe, J., & Blei, D. (2007). Supervised topic models. In Advances in neural information processing systems (Vol. 20). https://doi.org/10.7551/mitpress/7503.003.0081

Lu, Y., & Zhai, C. (2008). Opinion integration through semi-supervised topic modeling. In Proceedings of the 17th international conference on world wide web (pp. 121–130). https://doi.org/10.1145/1367497.1367514

Zhu, J., Ahmed, A., & Xing, E. P. (2009). MedLDA: Maximum margin supervised topic models for regression and classification. In Proceedings of the 26th annual international conference on machine learning (pp. 1257–1264). https://doi.org/10.1145/1553374.1553526

Sridhar, D., Daumé III, H., & Blei, D. (2022). Heterogeneous supervised topic models. Transactions of the Association for Computational Linguistics, 10, 732–745. https://doi.org/10.1162/tacl_a_00508

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. Information Systems, 94, 101582. https://doi.org/10.1016/j.is.2020.101582

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. In Proceedings of the 19th international conference on world wide web (pp. 851–860). https://doi.org/10.1145/1772690.1772777

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

BBC Persian. (2022). Hashtag against "Hazara genocide" surpasses three million. BBC. Retrieved August 7, 2024, from https://www.bbc.com/persian/articles/c72zlxxj3kro

Afghan International. (2022, October 2). Hashtag "Stop Hazara genocide" becomes Afghanistan's top trend on Twitter. Afghan International. Retrieved from https://www.afintl.com/202210026703