Journal of Media Horizons
ISSN: 2710-4060 2710-4052 Volume 6, Issue 3, 2025

DIGITAL PUBLIC DIPLOMACY AND NATIONAL IMAGE-MAKING: CHINA'S TWITTER COMMUNICATION IN THE MIDDLE EAST AND NORTH AFRICA

Muhammad Khalil Khan[1], Zitong Qiu[*2], Muhammad Naseem Anwar[3], Sadaf Iqbal[4], Tahir Farid[5]

[1,2]Department of Journalism and Communication, School of Media and Law, NingboTech University, Ningbo, People's Republic of China
[1,2]Belt and Road International Communication Research Center, NingboTech University, Ningbo, People's Republic of China
[3]Department of Computer Arts, Hamdard University Islamabad Campus, Pakistan.
[4]Department of Psychology, Abdul Wali Khan University, Mardan, Pakistan

[*2]qiuzitong@nbt.edu.cn

DOI: https://doi.org/10.5281/zenodo.17035951

Keywords
digital public diplomacy, National Image, Twitter (X), sentiment analysis, China, MENA, Islamic World.

Article History
Received on 30 May 2025
Accepted on 05 August 2025
Published on 30 August 2025

Corresponding Author: * Zitong Qiu

Abstract
Twitter's global reach impacts countries' digital public diplomacy and image-building efforts. It provides a highly dynamic public space for government agencies to directly engage with a diverse international audience in near real-time, effectively promoting national interests and building a favorable image. This article explores China's digital public diplomacy and image-making efforts within the context of the Belt and Road Initiative (BRI) through Twitter communication in the Middle East and North Africa (MENA) region. It uses a census of tweets posted by Chinese diplomats or embassies/consulates on their official Twitter accounts and public responses to those tweets between August 1, 2019, and March 30, 2023. We used Python programming powered by machine learning algorithms and deep learning models to extract and analyze data to excerpt insights and knowledge from the large volume of unstructured textual-based data, shedding light on China's digital public diplomacy and image-making efforts in the MENA regions. The study identifies 12 key themes in China's Twitter communication in the MENA region. Among them, China's Foreign Relations (CFR) emerged as the most dominant theme. Secondly, we explore three aspects of public engagement: the frequency analysis, the sentiment analysis, and the geographical distribution of the sentiments.

INTRODUCTION

Digital public diplomacy is a multifaceted strategic communication process used primarily by governments, businesses, organizations, and individuals to engage with global audiences

effectively in the digital space. In recent years, social media platforms have emerged as a powerful and dynamic tool in the field of digital public diplomacy, enabling spontaneous, direct, continuous, and unrestricted engagement with global audiences without any bureaucratic and financial constraints (Graham et al., 2013; Strauß et al., 2015). Social media platforms like Facebook and Twitter's wide reach and ability to facilitate real-time communication, direct interaction, and dialogic communication with the wide array of the public have all contributed to their significant influence in the public diplomacy domain. Notably, Twitter, now referred to as X, stands as one of the most preferred social media platforms globally for conducting public diplomacy initiatives (Ittefaq, 2019; Kampf et al., 2015; Rybalko & Seltzer, 2010).

China is a relatively new player in this digital public diplomacy field, and the Ministry of Foreign Affairs of the PRC only registered on Twitter in August 2019. As well, in the Middle East and North Africa (hereafter MENA) regions, Twitter accounts associated with Chinese diplomats, embassies, or consulates were set up no earlier than August 2019. The year 2019 was considered the "inaugural year" for Chinese diplomats on Twitter, and BBC revealed the presence of 55 Twitter accounts associated with Chinese diplomats, embassies, or consulates by the conclusion of 2019. Among these, a notable count of 32 accounts was registered specifically during that year (Feng, 2019). Furthermore, the COVID-19 pandemic has been attributed to the substantial enhancement and extensive implementation of China's digital public diplomacy (Shi & Tong, 2020). This development has been particularly evident in the remarkable surge of tweets posted by the official Chinese Twitter accounts, making it the primary discourse field for China to counter Western criticism against China (Alden & Chan, 2021), hence indicating a heightened engagement in online diplomatic efforts. Similarly, scholarly interest has followed a similar trajectory, as Twitter and digital public diplomacy have transitioned from being topics with a scarcity of scholarly exploration (Huang & Wang, 2019) to becoming cutting-edge subjects that have garnered substantial academic attention in recent years (Shi & Tong, 2020).

Bearing the background in mind, this article, from the prism of strategic communication studies, attempts to assess Chinese digital public diplomatic efforts in MENA by looking at tweets, replies, and other forms of public engagement on these Twitter accounts. It aims to address the following research questions:

RQ1: What are the prominent themes and topics that have emerged in the Twitter accounts associated with Chinese diplomats, embassies, or consulates since their initial posts? How have these major themes evolved and transformed over time?

RQ2: How does the general public engage with the aforementioned themes? Specifically, what sentiments are expressed in response to the tweets classified under specific themes? Furthermore, what are the geographical distributions of the replies, along with their associated sentiments?

Method

We use Python programming to extract data available from various sources on Twitter - including tweets, replies, and public engagement metrics such as likes, retweets, and impressions - from the Twitter IDs (N=16) associated with the MENA region, as detailed in Table 1. Among these Twitter IDs, the first 11 countries with 12 Twitter IDs correspond to countries in the Middle East, while the remaining 3 countries with 4 accounts pertain to countries in North Africa. It is noteworthy that as of the conclusion of 2022, there are no Chinese diplomatic Twitter accounts associated with Syria and Palestine in the Middle East, as well as Libya and Morocco in North Africa. The data timeline encompasses the earliest post available up until March 30, 2023. Furthermore, the obtained data was subjected to several processing steps, including translation, sentence tokenization, paragraph segmentation, and other necessary data-cleaning procedures.

Table 1: Chinese Embassy Twitter Account Information of MENA Region

Country	Account name	Twitter IDs	Join Twitter	The first tweet available extracted

Bahrain	Embassy of China in Bahrain	@ChinaEmbBahrain	October 2021	October 26, 2021
Iran	Chang Hua (Ambassador of China to the Islamic Republic of Iran)	@AmbChangHua	October 2019	March 30, 2021
Iraq	China Embassy in Iraq	@ChinaIraq	January 2020	January 20, 2020
Kuwait	Embassy of China in Kuwait	@ChinaEmbKuwait	April 2020	April 7, 2020
Jordan	Embassy of China in Jordan	@ChineseembassyJ	September 2019	September 24, 2019
Lebanon	Qian Minjian (Chinese Ambassador to Lebanon)	@ChinainLebanon	February 2020	February 7, 2020
Oman	Embassy of China in Oman	@ChinaEmbOman	March 2020	August 2, 2022
Qatar	Zhou Jian (Chinese Ambassador to the State of Qatar)	@AmbZhouJian	February 2020	January 8, 2022
Saudi Arabia	Embassy of China in KSA	@ChinaEmbKSA	July 2019	August 9, 2019
United Arab Emirates	China Consulate Dubai	@CGPRCinDubai	December 2019	December 8, 2019
	Zhang Yiming (Ambassador of China to the United Arab Emirates)	@Amb_Yiming	September 2019	October 11, 2019
Yemen	Chinese Embassy to Yemen	@ChineseEmbtoYEM	September 2019	September 18, 2019
Algeria	Embassy of China in Algeria	@ChinaEmbAlgeria	June 202	June 15, 2020
Egypt	Liao Liqiang (Chinese Ambassador to Egypt)	@AmbLiaoLiqiang	December 2019	November 17, 2020
	Spokesperson of the Chinese Embassy in Egypt	@CHN_EGY	June 2020	June 30, 2020
Tunisia	Chinese Embassy in Tanzania	@ChineseEmbTZ	December 2019	February 26, 2020

Data Processing and Protocols

Recent years have seen a surge of research methods associated with computer science in social science. The techniques of machine learning, deep learning, and artificial intelligence are increasingly being used when doing research in this field (Bhagat & Singh, 2022), whether implemented through computer programming or pre-packaged software/application. Echoing this research method in social science, this paper uses text mining as the major research method facilitated by machine learning algorithms and deep learning models developed within the field of computer science. In particular, this article applies Large Language Models (LLMs) in the context of text

summarization, General Text Embedding (GTE), and the Vader algorithm of Natural Language Toolkit (NLTK) to perform tasks about theme identification, similarity comparison, and sentiment analysis.

As a fundamental component of Natural Language Processing (NLP), text summarization compresses the large-scale textual content while preserving its most essential information.

Recent advancements in NLP have been marked in various LLMs, exemplified by Chat-GPT (Generative Pre-trained Transformer), an MML developed by OpenAI. LLMs in the context of text summarization have been studied mainly in the field of computer science. The theory of LLMs' powerful summarization capacity of text, to make it simple, lies in its ability to capture patterns, relationships, and semantic information present in the text, which it acquires through training on extensive textual data. Several studies have extensively investigated various approaches and evaluated the effectiveness of the LLMs, including Chat-GPT, for text summarization tasks (Basyal & Sanghvi, 2023; Bharathi Mohan et al., 2023). The scope of these studies of the field also extends to the news domain. For example, GPT-2, GPT-3, BERT (Bidirectional Encoder Representations from Transformers), and Transformer models are commonly examined and evaluated for news summarization, and their capabilities and performance have been assessed to provide concise and accurate summaries of news articles (Goyal et al., 2022; Gupta et al., 2022). Furthermore, scholars have shown comparable performance between text summarization facilitated by LLMs and human-authored summaries (Zhang et al., 2024). In this paper, we identify key themes from the tweets by the GPT3.5-turbo model and user-prompts formation. We first stored the tweet data in a text file and provided this file as input to ChatGPT. By using the prompt: "As a professional in journalism, can you help me summarize some key themes from these tweets?", we were able to extract 12 key themes. As well, we use text summarization of GPT to help extract key terms and topics of each subtheme.

However, despite ChatGPT's capability to achieve high performance and accuracy, it is considered a "black-box method" where the internal processes remain hidden from the users, thereby rendering it challenging to validate and evaluate. Hence, ChatGPT is unable to cluster exact tweets based on different themes. As a result, we introduce an additional approach by incorporating a "General Text Embeddings" model known as GTE (Li et al., 2023), which has been proposed by Alibaba, to facilitate theme clustering of the tweets. We conduct a thorough analysis of the relevance between each tweet and each theme, considering a tweet to be associated with a specific theme if its correlation is deemed significant, and excluding it otherwise. The model takes both tweets and themes as input and generates a similarity score ranging from 0 to 100. Tweets exhibiting a similarity score above 80 are considered semantically related, allowing us to assign them to their respective themes. Through this process, we ultimately derive the tweet count and corresponding weights for each theme through computational procedures.

GTE is also used to facilitate the identification of examples of tweets that exhibit the highest similarity to each respective subtheme. Through key terms summarized by GPT and key examples provided by GTE, we can ascertain the topics associated with each subtheme.

For text sentiment analysis, we utilize the NLTK, a Python library widely employed for NLP tasks[1]. NLTK provides a range of methods for conducting sentiment analysis, and one of the most popular algorithms attuned to social media is Vader (Valence Aware Dictionary for Sentiment Reasoning). Vader combines a sentiment lexicon approach with grammatical and syntactical aspects of text for expressing sentiment polarity and intensity (Hutto & Gilbert, 2014). In news and communication, NLTK is widely used to analyze sentiments associated with news headlines, news articles, and social media content (Khemani & Adgaonkar, 2021; Sah et al., 2022). Therefore, by harnessing the capabilities of the built-in machine learning algorithms of NLTK, we can identify positive, negative, or neutral sentiments from the Replies. This further enables us to gain valuable insights into the global audience's sentiments regarding specific themes.

Results and Discussion

Finding on RQ1

To address RQ1, we investigate the themes from two perspectives. On the one hand, we examine the characteristics of the 12 main themes and their subthemes. On the other hand, we observe the temporal evolution of these themes.

Major themes and subthemes

Using the tweets from the sixteen diplomatic Twitter accounts, ChatGPT analyzed and summarized twelve major themes and their corresponding subthemes, as shown in Table 2, and we created Figure 1 by arranging the themes in descending order based on their weights. Among these themes, China's Foreign Relations (CFR) emerged as the most dominant, accounting for 17.98% of the total analyzed tweets.

CFR has eight subthemes: (1) China-Pakistan Cooperation (15.32%) (e.g. "agreement on Asian Cultural Heritage Protection", joint projects, regional cooperation) ; (2) China-Arab Cooperation (15.32%)(e.g. economic prospect, diplomatic relations, joint actions); (3) China's Full Respect for the Internal Affairs of Middle Eastern Countries(13.63%)(e.g. internal affairs, Monroe doctrine, strategic coordination) ; (4) China-India Cooperation (13.35%) (e.g. "promote mutual respect and learning", economic collaboration, diplomatic ties, cultural exchange); (5) China's Relations with Middle Eastern Countries(12.15%)(e.g. Monroe doctrine, national development, strategic coordination),(6) China-Russia Strategic Cooperation(10.65%)(e.g. "comprehensive strategic partner of coordination", peace talks, settlement of the Ukraine crisis), (7) China-Africa Cooperation (10.92%)(e.g. strengthening strategic coordination, peaceful integration, trade summit, poverty alleviation); (8) China-Ukraine Peace Dialogue(8.67%)(e.g. Ukraine crisis, peace talks, Russia, US, Monroe doctrine).

One can conclude that within this theme, nearly half of the subthemes (comprising 47.99%) map out China's relationship with MENA nations. The remaining half of the subthemes (52.02%) revolve around regional conflicts such as the China-Ukraine Peace Dialogue and China's collaborations with major powers in Asia, including Pakistan, India, and Russia. These dynamics significantly impact the geopolitical landscape of the MENA regions.

It is noteworthy that China's cooperation with the Middle East and North Africa has garnered varying degrees of attention. Despite collecting data from four accounts associated with North Africa and twelve accounts related to the Middle East, the theme of China-Africa Cooperation (CAC), however, emerges as a more significant component, accounting for a substantial weight of 11.35% in the overall theme analysis. Within the theme of CAC, out of the six identified subthemes, significant attention is devoted to (1) People-to-People Exchange (34.8%) (e.g. "China's promotion of Tanzanian products", investment); (2) China-Africa Relations (31.24%) (e.g. "build a high-level China-Africa community", infrastructure, development); (3) China's Aid to Africa (23.91%) (e.g. "Juncao technology to Tanzania", international cooperation, aid). Additionally, there is a small number of tweets that focus on (4) Humanitarian Assistance in Africa (4.13%) (e.g. assistance, poverty alleviation); (5) Infrastructure Construction in Africa (3.33%) (e.g. "first electric LRT project", infrastructure construction, sustainable development); (6) Economic Development in Africa (2.58%) (e.g. "promotion of connectivity and connectivity", "guarantee of industrial and supply chains in Africa", "cooperation in the areas of the digital economy, smart cities, clean energy", "development and renaissance in Africa", trade).

In contrast, the theme of Cooperation and Diplomacy in the Middle East Region (CDMER) holds a comparatively lower weight of 6.38% in the total analysis. Six subthemes were identified: (1) Relations with Middle Eastern Countries (30.44%) (e.g. international integration, multilateralism, solidarity); (2) Economic Cooperation in the Middle East (26.33%) (e.g. peaceful integration, BRI, multilateralism, trade summit.); (3) Middle East Peace Process (17.1%)(e.g. multilateralism, peace process);(4) Security Situation in the Middle East (9.6%)(e.g. security, peace, stability);(5) Investment in the Middle East (9.09%) (e.g. economic cooperation, BRI, multilateralism, trade summit); (6) Palestinian Issue (7.44%) (e.g. Palestine-Israel, Arab countries, Arab League, two-state solution). Within the theme of CDMER, apart from the subtheme Relations with

Middle Eastern Countries, considerable emphasis is placed on economy and security.

Furthermore, Chinese diplomatic Twitter accounts actively engaged in promoting its cultural traditions, articulating the contemporary political theory of the Chinese Communist Party, and highlighting China's Economy and its role in global collaboration and development in various areas. These corresponding themes, namely Chinese Cultural Traditions (CCT) (10.69%), the Chinese Communist Party Education Campaign (CPEC) (10.52%), China's Economy and Global Growth (CEGG) (10.06%), and International Cooperation and Development (ICD) (9.86%), each account for approximately 10 percent of the total weight.

To articulate further, theme CCT encompasses five distinct subthemes, offering insights into various aspects of China's rich cultural heritage. (1) Chinese Intangible Cultural Heritage (27.51%) (e.g. intangible culture exhibition, Expo City Dubai); (2) Chinese New Year Celebrations (22.65%) (e.g. Chinese New Year, modernization); (3) Chinese Traditional Art (18.18%) (e.g. traditional tea processing techniques), (4) Chinese Music Culture (17.78%) (e.g. "Chinese music with traditional Chinese music machine"), (5) Chinese-style Band Performances (13.92) (e.g. Chinese universities, Ma Bang Band, Hala China).

The CPEC theme predominantly focuses on the subtheme of Xi Jinping's Thought (94.17%) (e.g. CPC education campaign, 20th CPC National Congress), leaving tiny room for the subtheme Education Activities (5.83%) (e.g. Confucius, Chinese School Dubai, Chinese curriculum school).

The CEGG theme comprises seven subthemes that center around China's economic activities and investments within and beyond its borders. These include (1) China's Contribution to Global Growth (27.91) (e.g. modernization, strategic partnership, trade resilience, democracy, economic future); (2) China's Economic Development Model (19.25%) (e.g. "The Chinese Model of Modernization", foreign investment); (3) China's Outward Investment (19.27%) (e.g. "China's foreign trade", outward investment, economic recovery); (4) China's Economic Outlook (17.09%) (e.g. "China's long-term economic prospect", globalization, poverty alleviation, opening-up); (5) China's Economic Recovery (13.34%) (e.g. global impact); (6) Domestic Demand (1.97%) (e.g. domestic demand, COVID-hit economy, gradual recovery); (7) Consumption and Investment (1.17%) (e.g. stability, foreign direct investment, economic indicators, fixed assets investment, global investors).

The subthemes of ICD comprehensively encompass a range of cooperation and development initiatives across various domains. These include (1) Global Security Initiatives (31.79%) (e.g. Socrates, Confucius, comprehensive strategic partnership, BRI cooperation); (2) Africa Centers for Disease Control and Prevention (17.93%) (e.g. UN General Assembly, virus spread, multiparty cooperation, epidemic cooperation); (3) Global Economic Growth (15.12%)(e.g. China's development, world opportunities, global economic recovery); (4) Global Biodiversity Framework (12.37%) (e.g. global biodiversity framework,); (5) Digital Development (11.43%) (e.g. community with a shared future, opportunities, BRI); (6) International Patent Applications (11.36%) (e.g. China-Indonesia cooperation, BRI, foreign investment, open China, firearms control, services trade fair).

Lastly, these diplomatic accounts also covered various international affairs, although they were not as prominent in terms of weightage. The themes of International Security and Peace (ISP) (8.4%), International Trade and Economic Cooperation (ITEC) (5.8%), Human Rights Issues (HRI) (3.5%), Environmental Protection and Sustainable Development (EPSD) (3.32%), and Technological Development and Innovation (TDI) (2.14%) held weights that were below 10 percent of the total, indicating their relatively lesser significance within the overall analysis.

To elaborate, six subthemes were identified within the theme ISP: (1) Global Security Situation (28.33%) (e.g. violent spirit, regional security, territorial integrity); (2) International Military Cooperation (25.29%) (Peaceful integration, confrontation, peaceful coexistence, win-win cooperation); (3) Middle East Peace Process (14.78%) (e.g. dialogue, peace, diplomatic efforts); (4) Arms Control (12%) (e.g. firearms control, national security law, pandemic control); (5) Counterterrorism Efforts (11.46%) (e.g. counterterrorism, human wrong, national

sovereignty, territorial integrity); (6) Ukraine Crisis (8.15%) (e.g. China-Russia relations, United States, peaceful settlement, Cold War mentality).

Theme ITEC has five subthemes: (1) International Economic Cooperation (35.99%) (e.g. China, international cooperation, economic growth, trade, BRI); (2) Protection of World Intellectual Property Rights (19.42%) (e.g. WTO, intellectual property rights, protection); (3) Resilience in Trade (16.83%) (e.g.); (4) Optimism in the Global Market (14.57%) (e.g.); (5) Export and Growth Expectations (13.2%) (e.g. resilient exports, trade, and growth, economic stability, WTO reform).

In regard to HRI, it consists of six subthemes. They are (1) US Human Rights Records (38.14%) (e.g. Report on Human Rights Violations, politicizing human rights, crimes, violations, Xinjiang, interference); (2) Comparison of Human Rights between China and the US (23.27%) (e.g. violations, global governance, Chinese FM); (3) China's Global Human Rights Governance (22.85%) (e.g. human rights violations, Chinese FM, Xinjiang, politicizing human rights); (4) Human Rights Situation in the Middle East (14.93%) (e.g. politicizing, internal affairs, UN Human Rights Council, Xinjiang, Hong Kong, Tibet, sovereignty, territorial integrity); (5) Assistance to Afghan Refugees (8.1%) (e.g. humanitarian, humanitarian crisis, United Nations Relief and Works Agency for Palestine Refugees (UNRWA)).

ESPD is comprised of six subthemes, and the weight of each subtheme is relatively evenly distributed: (1) Climate Change Mitigation (17.98%) (e.g. combating climate change, greening campaign, grass planting); (2) Green Development (17.69%) (e.g. Green Development, COP15, global economic recovery, afforestation); (3) Global Biodiversity (16.98%) (e.g. biodiversity loss, halt and reverse, a community of life); (4) Environmental Conservation Initiatives (16.76) (e.g. global biodiversity framework); (5) Environmental Sustainability (16.73%) (e.g. Fighting climate change, Greening campaign, Ecological protection, Green agenda); (6) New Energy Vehicles (13.38) (e.g. sustainable transport, low-emission development, technological innovation.).

Finally theme TDI focuses on six subthemes: (1) Technological Cooperation (19.17) (e.g. Chinese modernization, digital economy, technological innovation, financial cooperation); (2) Digital Development (17.76%) (e.g. global vision, advanced technology, digital technology, technological innovation, high-quality development, fifth-generation technology); (3) International Patent Applications (16.28%) (e.g. intellectual property, global influence, scientific achievements, open global economy, BRI); (4) Applications of Artificial Intelligence (15.74%) (e.g. intelligent development, World Congress of Artificial Intelligence, smart transport, robotics, fifth generation telecommunications, digital transformation); (5) Electric Vehicle Charging Infrastructure (15.57%) (e.g. smart, green, low-emission, new energy vehicles, BRI, renewable energy, infrastructure construction); (6) Space Science Experiments (15.49%) (e.g. Space science, microgravity, space technology, astronaut, space station, celestial bodies.).

Table 2: Themes, Subthemes and Weights

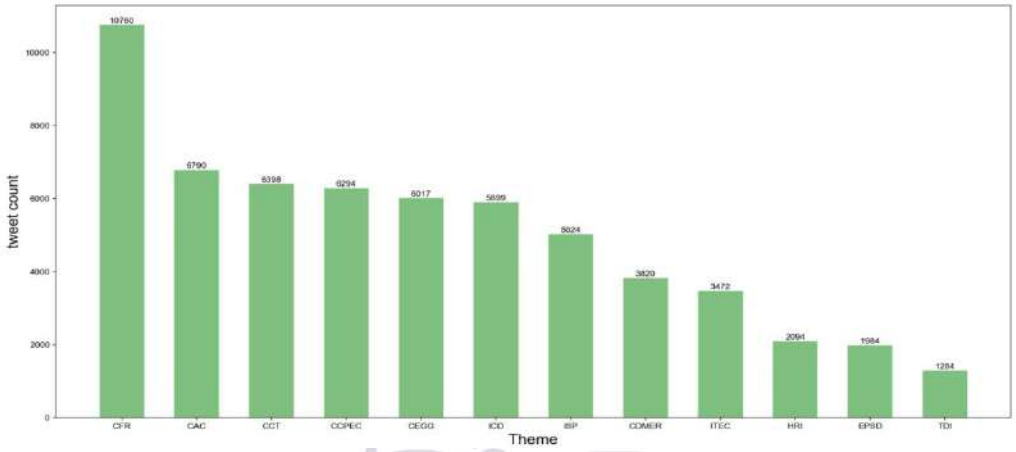
Theme	Tweet Count	Percentage (100% in total)	Subthemes	Tweet Count	Percentage (100% in each theme)
China's Foreign Relations (CFR)	10760	0.1798	China-Pakistan Cooperation	8194	0.1532
			China-Arab Cooperation	8193	0.1532
			China's Full Respect for the Internal Affairs of Middle Eastern Countries	7291	0.1363
			China-India Cooperation	7142	0.1335
			China's Relations with Middle	6500	0.1215

			Eastern Countries		
			China-Russia Strategic Cooperation	5695	0.1065
			China-Africa Cooperation	5840	0.1092
			China-Ukraine Peace Dialogue	4638	0.0867
China-Africa Cooperation (CAC)	6790	0.1135	People-to-People Exchange between China and Africa	5703	0.3480
			China-Africa Relations	5120	0.3124
			China's Aid to Africa	3918	0.2391
			Humanitarian Assistance in Africa	677	0.0413
			Infrastructure Construction in Africa	546	0.0333
			Economic Development in Africa	423	0.0258
Chinese Culture and Tradition (CCT)	6398	0.1069	Chinese Intangible Cultural Heritage	5320	0.2751
			Chinese New Year Celebrations	4374	0.2265
			Chinese Traditional Art	3516	0.1818
			Chinese Music Culture	3438	0.1778
			Chinese-style Band Performances	2692	0.1392
Chinese Communist Party Education Campaign (CCPEC)	6294	0.1052	Xi Jinping Thought	4148	0.9417
			Education Activities	257	0.0583
China's Economy and Global Growth (CEGG)	6017	0.1006	China's Contribution to Global Growth	5782	0.2791
			China's Economic Development Model	3988	0.1925
			China's Outward Investment	3991	0.1927
			China's Economic Outlook	3539	0.1709
			China's Economic Recovery	2763	0.1334
			Domestic Demand	408	0.0197
			Consumption and Investment	242	0.0117
International Cooperation and Development (ICD)	5899	0.0986	Global Security Initiatives	3333	0.3179
			Africa Centers for Disease Control and Prevention	1880	0.1793
			Global Economic Growth	1585	0.1512
			Global Biodiversity	1297	0.1237

			Framework		
			Digital Development	1198	0.1143
			International Patent Applications	1191	0.1136
International Security and Peace (ISP)	5024	0.0840	Global Security Situation	4250	0.2833
			International Military Cooperation	3795	0.2529
			Middle East Peace Process	2217	0.1478
			Arms Control	1800	0.1200
			Counterterrorism Efforts	1719	0.1146
			Ukraine Crisis	1223	0.0815
Cooperation and Diplomacy in the Middle East Region (CDMER)	3820	0.0638	Relations with Middle Eastern Countries	2974	0.3044
			Economic Cooperation in the Middle East	2573	0.2633
			Middle East Peace Process	1671	0.1710
			Security Situation in the Middle East	938	0.0960
			Investment in the Middle East	888	0.0909
			Palestinian Issue	727	0.0744
International Trade and Economic Cooperation (ITEC)	3472	0.0580	International Economic Cooperation	3229	0.3599
			Protection of World Intellectual Property Rights	1743	0.1942
			Resilience in Trade,	1510	0.1683
			Optimism in the Global Market	1307	0.1457
			Export and Growth Expectations	1184	0.1320
Human Rights Issues (HRI)	2094	0.0350	US Human Rights Records	1275	0.3814
			Comparison of Human Rights between China and the US	778	0.2327
			China's Global Human Rights Governance	764	0.2285
			Human Rights Situation in the Middle East	499	0.1493
			Assistance to Afghan Refugees	27	0.0081
Environmental Protection and Sustainable Development (EPSD)	1984	0.0332	Climate Change Mitigation	1314	0.1798
			Green Development	1293	0.1769
			Global Biodiversity	1241	0.1698
			Environmental Conservation Initiatives	1225	0.1676
			Environmental Sustainability	1223	0.1673

			New Energy Vehicles	1014	0.1387
Technological Development and Innovation (TDI)	1284	0.0215	Technological Cooperation,	1208	0.1917
			Digital Development	1119	0.1776
			International Patent Applications	1026	0.1628
			Applications of Artificial Intelligence	992	0.1574
			Electric Vehicle Charging Infrastructure	981	0.1557
			Space Science Experiments	976	0.1549

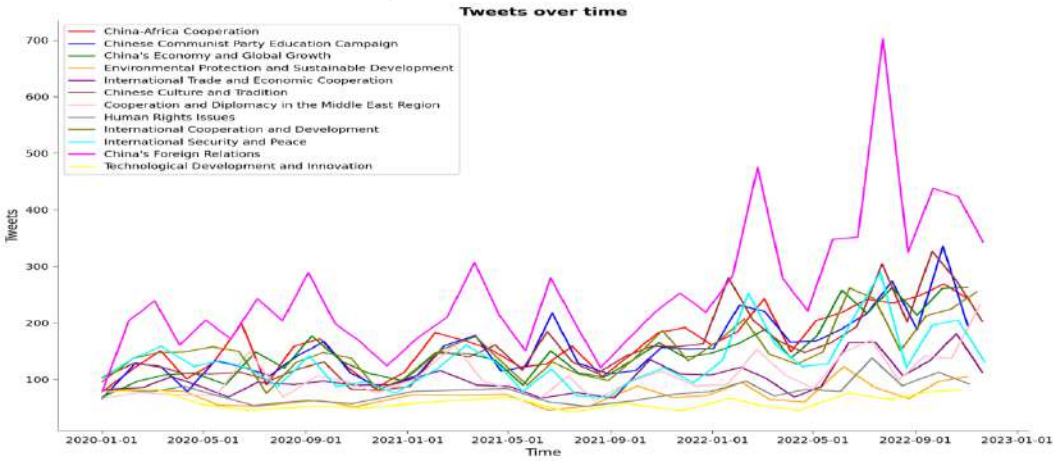
Figure1 Theme Weight



Transformation of the themes

Regarding the temporal evolution of the themes, the line graph showcases the dramatic fluctuations in the data. The peaks and valleys in the graph exhibit a similar rhythm, indicating that the timing of almost all the peaks across the various subjects is nearly identical. This intense volatility is caused by international diplomatic events. We can observe eight peaks, each corresponding to a diplomatic event. Furthermore, we can observe that despite significant oscillations, nearly all subject tweets, except for TDI, demonstrate an upward trend over the studied period

Figure 2 Themes over time



Finding on RQ2

We will elucidate three aspects related to public engagement: frequency analysis, sentiment analysis, and the geographical dimension of sentiments. Frequency analysis considers statistical measures such as Likes, Retweets, and Replies to explore the varying degrees of public engagement. Sentiment analysis, on the other hand, aims to uncover the positive, negative, or neutral sentiments expressed within the Replies. Furthermore, we employ maps to visually represent the geographical distribution of these sentiments.

Table 3 Sentiment Analysis of the Public Engagement

Theme	Likes	Retweets	Replies	Impressions	Positive count	Negative count	Neutral count
China's Foreign Relations (CFR)	243451	356962	15325	1267942	3337	685	11303
China-Africa Cooperation (CAC)	204471	158853	11365	1063410	2576	430	8359
Chinese Culture and Tradition (CCT)	144042	128871	9745	991370	2063	316	7366
Chinese Communist Party Education Campaign (CCPEC)	122716	179389	7937	573155	1793	337	5807
China's Economy and Global Growth (CEGG)	78556	131114	7126	635134	1782	347	4997
International Cooperation and Development (ICD)	145171	93172	9982	679870	2227	351	7404
International Security and Peace (ISP)	159675	104515	10514	549374	2149	428	7937
Cooperation and Diplomacy in the Middle East Region (CDMER)	193740	85663	10392	836574	1587	208	8597

International Trade and Economic Cooperation (ITEC)	56838	35406	4848	310091	1244	179	3425
Human Rights Issues (HRI)	23957	24766	2952	131842	725	179	2048
Environmental Protection and Sustainable Development (EPSD)	24985	14718	2388	140577	622	89	1677
Technological Development and Innovation (TDI)	16686	6261	1464	112381	284	33	1147

Public engagement： frequency analysis

Twitter has its own mechanism that facilitates interactive communication, employing features such as likes, retweets, replies, and impressions, thereby serving as channels for public engagement. Scholars have extensively examined how engagement possibilities, along with enhanced visibility and direct communication with the public facilitated by social media, have redefined the landscape of public diplomacy and have transformed the role of diplomats as public relations practitioners in the global arena (Khan et al, 2021). We draw on Khan and Pratt's (2022) research on the public engagement of Facebook, which is operationalized at three levels - low, moderate, and high levels of engagement. In the same vein, we consider Likes on Twitter represent a low level of engagement where users express general interest; retweets indicate a moderate level of engagement where users not only display interest but also commit to sharing the post on their profiles; replies, representing the high level of engagement, reflect users' active involvement by expressing their stance, which can include agreement, resistance, or negotiation with the message conveyed in the tweet. In addition, it is worth noting that Twitter Impression gauges the extent of exposure of tweets to the public, quantifying the frequency with which a tweet has been viewed by users on the platform.

By dividing the counts of Impressions, Likes, Retweets, and Replies for each of the 12 themes by their sum, we obtained the percentage data, which is displayed in Figure 3. From the figure, we can derive some interesting findings. Firstly, despite some fluctuations, there is an overall downward trend in three levels of public engagement across all themes, which correlates with the general trend of decreasing weights observed in the 12 themes. However, exceptions arise from the low and high levels of public engagement in the themes ICD (Like 10.26%, Reply 10.61%), ISP (11.29%,11.18%), and CDMER (13.7%,11.05%). In other words, despite having lower theme weights, the percentage of Likes and Replies in these three themes is higher than in the preceding themes CCT (10.18%,10.36), CCEPC (8.68%, 8.44%), and CEGG (5.55%,7.58%), which have higher theme weights. Secondly, in terms of low-level public engagement, the proportion of likes only exhibits a clear advantage in two themes, namely CAC (14.46%) and CDMER (13.7%). Interestingly, these two themes happen to be the ones that are directly related to the MENA region. Thirdly, regarding to moderate level of public engagement, the Retweet metric holds significant importance in themes CFR (27.05%), CCPEC (13.59%), and CEGG (9.94%), surpassing the other two levels of public engagement. Finally, in four themes, namely ITEC (5.16%), HRI (3.14%), ESPD (2.54%), and TDI (1.56%), although the weight of each level of public engagement is relatively small, it is noteworthy that the proportion of Replies is much higher compared to the other two forms of public engagement.

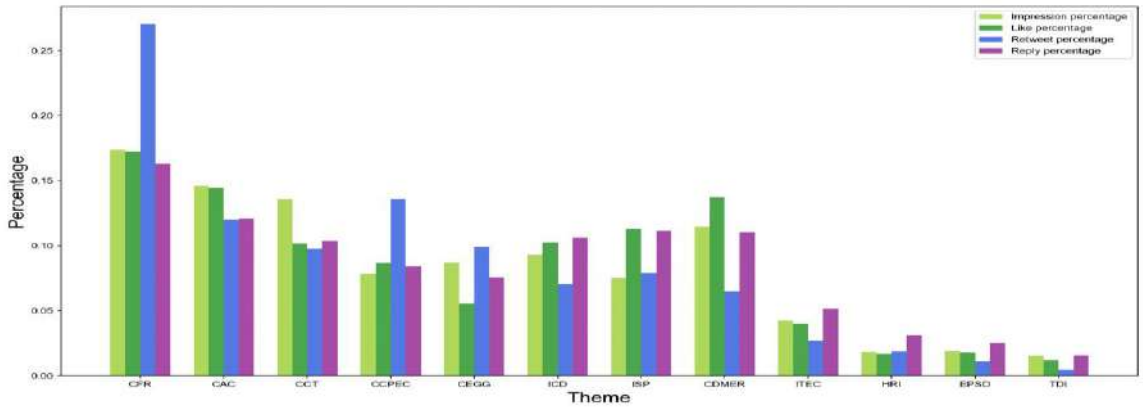


Figure 3 (left)

Public engagement – sentiment analysis

Now we examine the sentiment components within the replies. NLTK's Vader analysis is utilized to categorize the text into three types of sentiments, such as neutral, positive, or negative. Positive sentiment refers to the identification of text or language that conveys a favorable sentiment, emotion, or opinion. Negative sentiment, on the other hand, represents the opposite scenario. In other cases, replies are classified as neutral.

Upon analyzing numerical numbers associated with the three sentiments for each theme, it becomes evident that the neutral sentiment garners the highest number of replies, ranging from 69.38% to 82.73% of the total replies. Following this, positive sentiments account for approximately 15.27% to 26.05% across each theme's replies, while negative sentiments comprise approximately 2% to 6.06%. By dividing the replies expressing positive, negative, and neutral sentiments for each theme by the respective total counts of each sentiment, we can obtain percentage values, which reveal some additional intriguing findings.

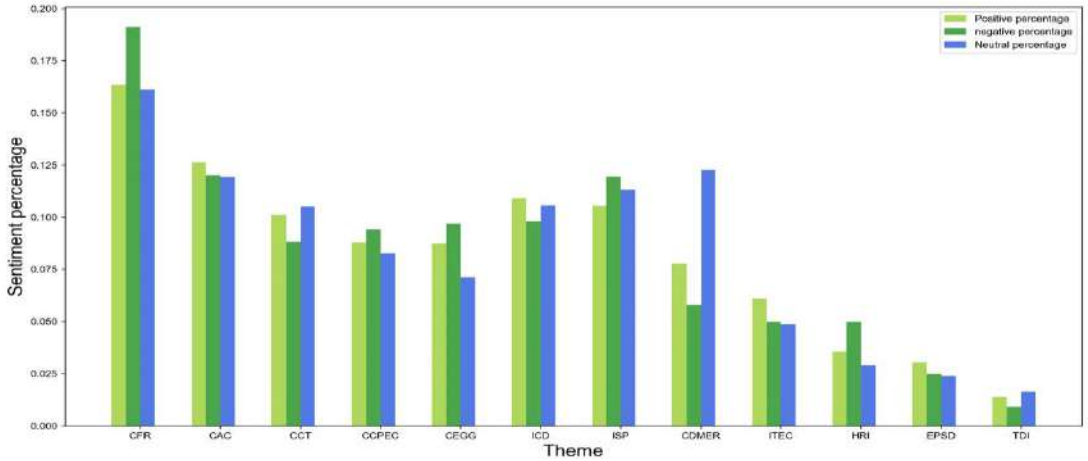
With the normalized data presented in Figure 4, firstly, we can observe that the distribution of the three kinds of sentiments across the 12 themes follows a similar pattern to the theme weight distribution, showing a downward trend. However, there are exceptions in the case of ICD, ISP, and CDMER. Both ICD (positive 10.92%, negative 9.8%) and ISP (10.54%, 11.95%) exhibit higher levels of positive and negative sentiments compared to CCT (10.12%, 8.82%), CCPEC (8.79%, 9.41%), and CEGG (8.74%, 9.69%), which rank higher in terms of theme weights. Additionally, CDMER, ranked eighth in theme weights, displays a higher proportion of neutral sentiment (12.27%) compared to all themes preceding it, except for CFR (16.13%).

Secondly, we proceed to examine the weight distribution of different sentiments within each theme. Notably, in the case of CAC (positive 12.63%, negative 12%, neutral 11.93), ICD (10.92%, 9.8%, 10.57%), ITEC (6.1%, 5%, 4.89%), and EPSD (3.05%, 2.48%,2.39%), the weight attributed to positive sentiment surpasses that of the other two forms of sentiments, albeit without exhibiting a pronounced superiority.

Additionally, Among the following five themes, the proportion of negative sentiment stands out prominently. These themes include CFR, CCPEC (9.41%), CEGG (6.69%), ISP (11.95), and HRI. Particularly noteworthy are CFR (positive 16.37% negative 19.12%, neutral 16.13%) and HRI (3.56%, 5%, 2.92%), as these two themes exhibit a much higher proportion of negative sentiment compared to the other two sentiments.

Finally, let us turn our attention to the prevalence of neutral sentiment. In the themes of CCT, CDMER, and TDI, neutral sentiment emerges as the dominant sentiment. Notably, in the CDMER theme (positive 7.78%, negative 5.81, neutral 12.27%), the proportion of neutral sentiment far exceeds that of the other two sentiments.

Figure 4: Sentimental percetage



Public engagement – geographical dimensions

We further explored the geographical dimension of the sentiment analysis of the Replies. To do so, we track the Twitter profiles of the users who post replies with either positive or negative sentiments and record their country settings information. If the country set by the user on their profile exists, we record it. However, if the user randomly sets a location that does not belong to any country, we filter out such information.

Table 4 lists the top 20 countries with the highest number of marked positive and negative sentiment replies sorted in descending order based on positive comments. We observed that Qatar had an overwhelming number of both positive and negative sentiment comments, surpassing nearly 10000, which significantly outstripped other countries in absolute numbers. Following closely were Tanzania, with over 1500 comments, and the United States, with nearly 1000 comments.

Table 4 Top 20 Countries Sentiment Analysis

Rank	Country	Positive count	Negative count	Sum
1	Qatar	7750	1946	9696
2	Tanzania	1292	215	1507
3	United States	868	103	971
4	Lebanon	560	47	607
5	Indonesia	515	9	524
6	Kuwait	257	17	274
7	Saudi Arabia	256	36	292
8	China	214	24	238
9	United Kingdom	96	53	149
10	Namibia	95	7	102
11	Pakistan	75	3	78
12	Egypt	50	3	53
13	Canada	37	4	41
14	India	33	31	64
15	Yemen	24	10	34
16	Oman	18	10	28
17	France	16	15	31
18	German	14	5	19
19	United Arab Emirates	13	6	19
20	Australia	12	11	23

Figure 5 Positive and negative sentiment percentage of the top 20 countries

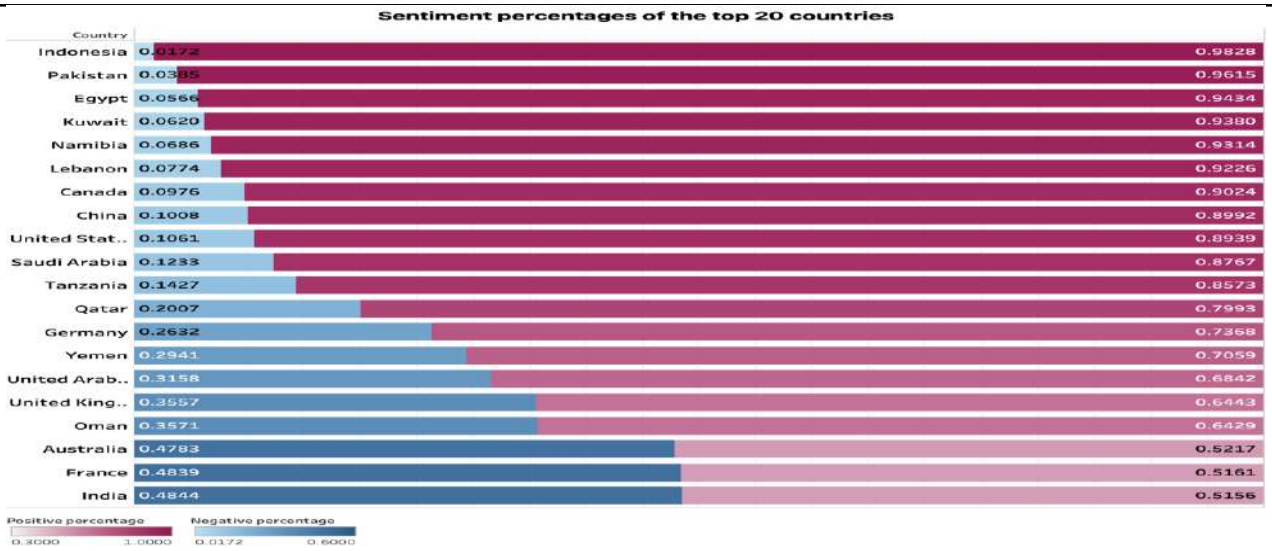
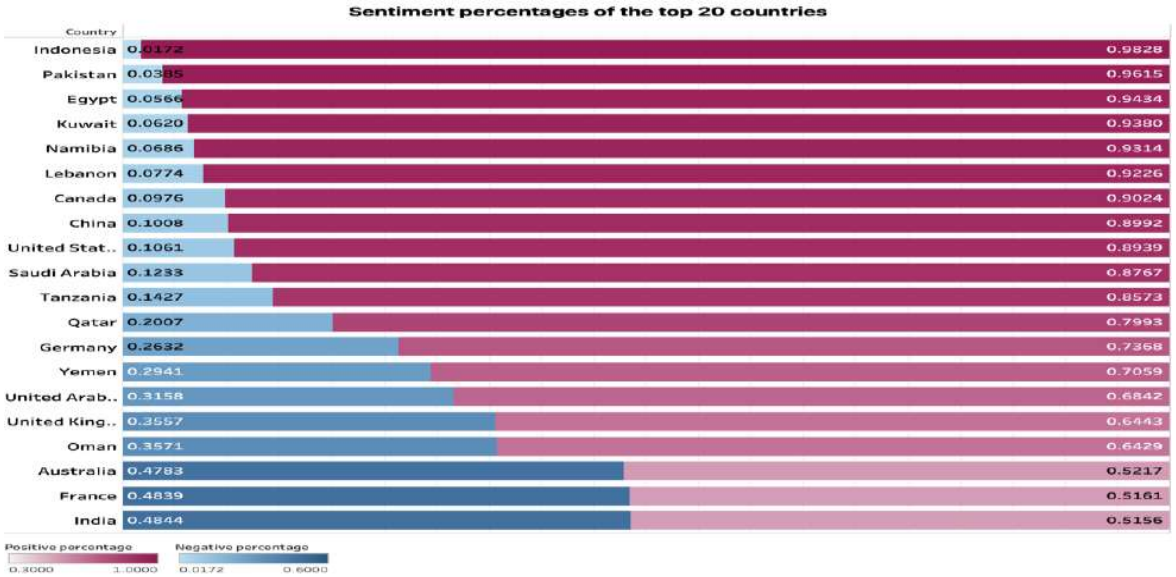


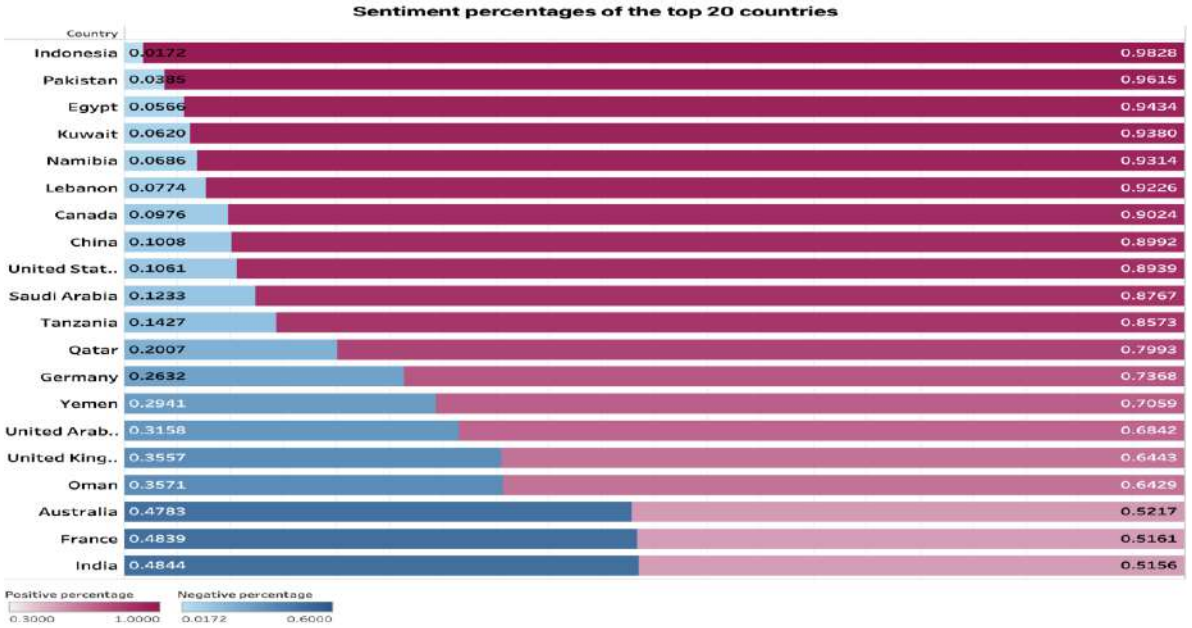
Figure 6 Positive and negative sentiment percentage of the top 20 countries



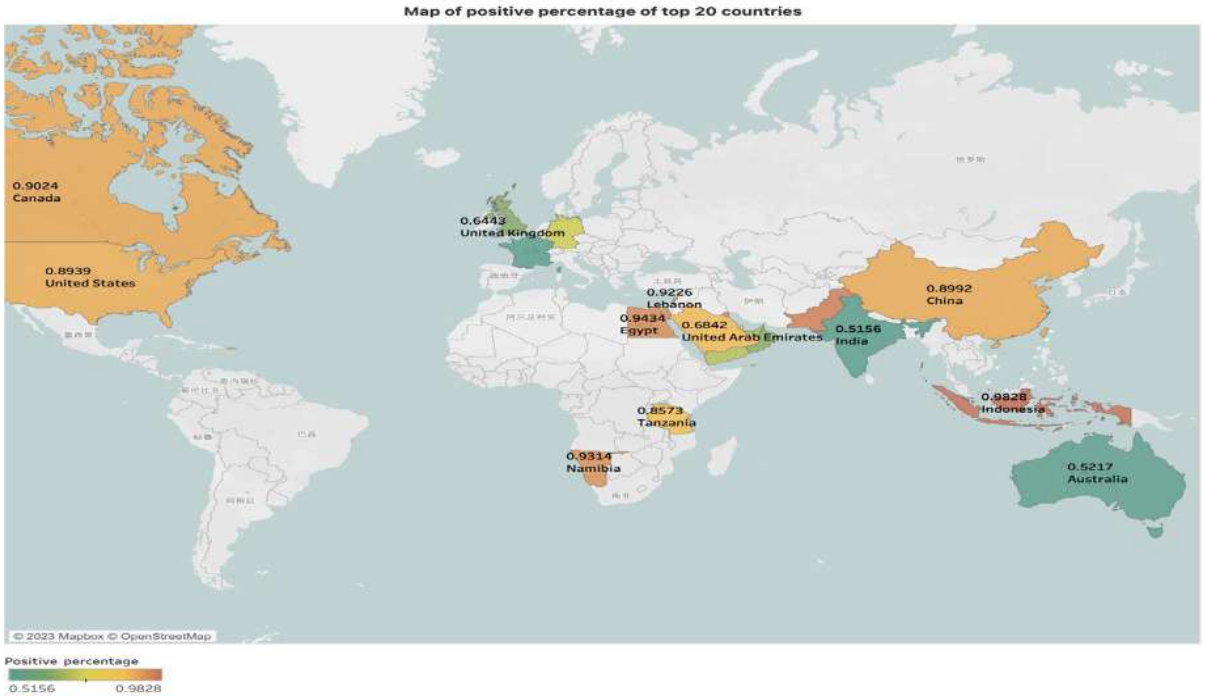
If we examine these data from a proportional perspective, we make different observations. By dividing the number of replies with positive or negative sentiments for each country by the sum of the two, we obtain the ratios shown in Figure 8. As depicted in Figure 8, audiences from five countries, namely Indonesia, Pakistan, Egypt, Kuwait, and Namibia, are more inclined to respond with positive sentiments. These countries have positive comment ratios exceeding 90%. Particularly in the case of Indonesia and Pakistan, negative sentiment accounts for less than 5% of the total. On the other hand, the three countries with the highest proportion of negative reviews are India, France, and Australia. In these nations, the positive and negative comment counts are almost evenly distributed. Following closely behind these three countries, Oman, the United Kingdom, and the United Arab Emirates they also have negative review rates exceeding 30%. Maps 1 and 2 showcase a more visually intuitive representation of the percentage of positive and negative replies for these 20 countries

.

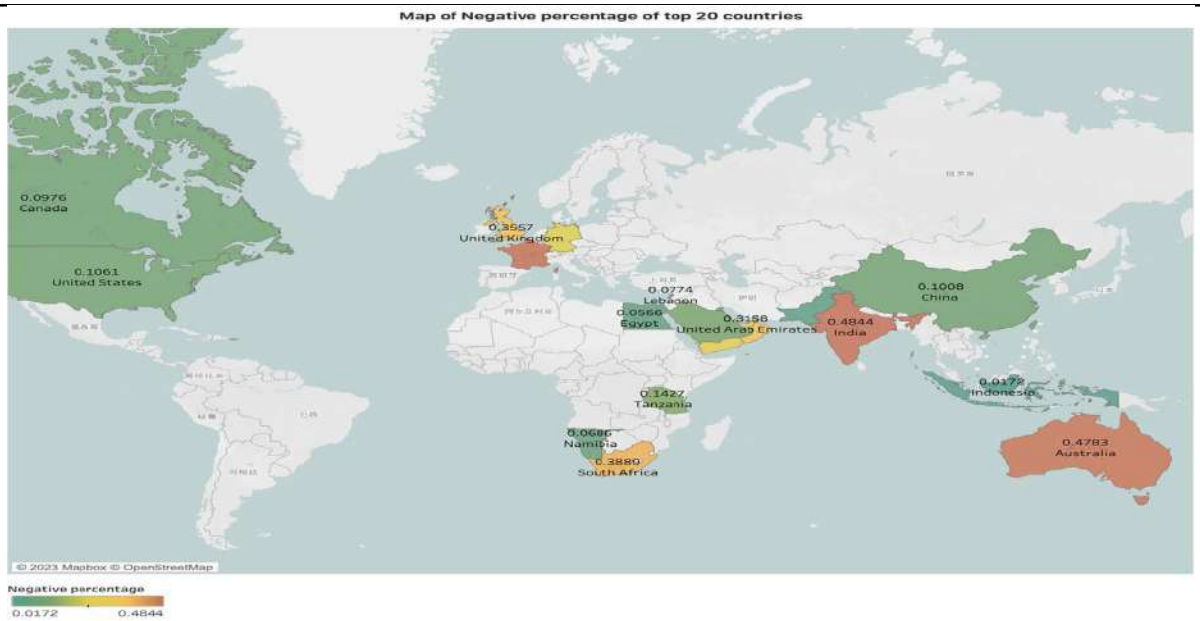
Figure 7 Positive and negative sentiment percentage of the top 20 countries



Map 8 Map of positive percentage of top 20 countries



Map 9 map of positive percentage of top 20 countries



Discussion and Conclusion

This study focuses on digital public diplomacy through Chinese diplomatic Twitter accounts from August 2019 to March 30, 2023. We employ innovative research methods from the field of computer science to conduct text mining on Twitter accounts and other content.

On the one hand, we utilize ChatGPT to generate 12 main themes along with their respective subthemes and their corresponding weights. On the other hand, the paper extensively explores public engagement with these 12 themes. This exploration primarily revolves around three aspects: frequency analysis, sentiment analysis, and geographical analysis of sentiments.

The frequency analysis examines the number of low, moderate, and high levels of public engagement by looking at the number of Likes, Retweets, and Replies. We discovered that the trends of the three types of public engagement largely align with the patterns observed in the tweet's numbers of the themes. Exceptional outliers are mainly from CEGG and CDMER. For the CEGG theme, all three types of public engagement fall below the theme weight, indicating a mismatch between the proportion of related tweets and corresponding public engagement. In contrast, the CDMER theme displays a remarkable anomaly with significantly higher levels of public involvement across all three types, particularly in terms of Likes and Replies. In addition, in absolute terms, the levels of public engagement for most themes exhibit a descending order of high, medium, and low; however, the normalized data still implies a significant presence of moderate to high-level public engagement for the majority of the themes.

Sentiment analysis shows that Across all twelve themes, there is a notable prevalence of neutral sentiment in Replies, and conversely, positive and negative sentiments exhibit relatively lower percentages. Moreover, there is a close relationship between positive and negative sentiments across the twelve themes, indicating a synchronized proportional tendency. In the same vein, there is a parallel pattern between positive and negative sentiments in terms of their proportional changes relative to the total number of replies for each theme.

There are two sets of interesting exceptions. For the CCT theme, negative sentiment is relatively less common compared to total and positive replies; by contrast, the CFR and HRI themes exhibit a higher prevalence of negative sentiment compared to both the overall number of replies and the positive sentiment counterparts. this may necessitate a more nuanced approach by Chinese diplomatic accounts in handling these two topics in the future, HRI in particular. Another interesting set of examples are

from CDMER and CEGG. Another intriguing set of examples can be observed within the CEGG and CDMER themes. In the case of the former, the proportion of positive and negative sentiment comments exceeds the overall proportion of comments. Conversely, for the latter, the proportion of positive and negative sentiment comments is significantly lower than the overall proportion of comments, suggesting that individuals are less likely to express explicit positive or negative emotions when commenting on this theme.

Lastly, we examine the distribution of sentiments on a world map. In terms of absolute numbers, we observed that Qatar had a significant number of both positive and negative sentiment replies, surpassing other countries by a considerable margin. Tanzania and the United States followed thereafter. However, if we examine the percentage of positive and negative sentiments in proportion, we find that Indonesia, Pakistan, Kuwait, Egypt, and Namibia have the highest proportions of positive sentiment evaluations. Conversely, the countries with the highest proportions of negative sentiment include India, France, and Australia, where the positive-to-negative ratio is nearly balanced.

References

Alden, C., & Chan, K. (2021). Twitter and Digital Diplomacy: China and Covid-19. *LSE Ideas*, *June*.

Basyal, L., & Sanghvi, M. (2023). Text Summarization Using Large Language Models: A Comparative Study of MPT-7b-instruct, Falcon-7b-instruct, and OpenAI Chat-GPT Models. *ArXiv Preprint ArXiv:2310.10449*.

Bhagat, H. V, & Singh, M. (2022). Machine-Learning and Deep-Learning Techniques in Social Sciences. *Machine Learning Algorithms for Signal and Image Processing*, 409–428.

Bharathi Mohan, G., Prasanna Kumar, R., Parathasarathy, S., Aravind, S., Hanish, K. B., & Pavithria, G. (2023). Text Summarization for Big Data Analytics: A Comprehensive Review of GPT 2 and BERT Approaches. *Data Analytics for Internet of Things Infrastructure*, 247–264.

Feng, Z. (2019). China and Twitter: The year China got louder on social media. *BBC*. https://www.bbc.com/news/world-asia-china-50832915

Goyal, T., Li, J. J., & Durrett, G. (2022). News summarization and evaluation in the era of gpt-3. *ArXiv Preprint ArXiv:2209.12356*.

Graham, T., Broersma, M., Hazelhoff, K., & Van't Haar, G. (2013). Between broadcasting political messages and interacting with voters: The use of Twitter during the 2010 UK general election campaign. *Information, Communication & Society*, *16*(5), 692–716.

Gupta, A., Chugh, D., Anjum, & Katarya, R. (2022). Automated news summarization using transformers. In *Sustainable Advanced Computing: Select Proceedings of ICSAC 2021* (pp. 249–259). Springer.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225.

Ittefaq, M. (2019). Digital diplomacy via social networks: A cross-national analysis of governmental usage of Facebook and Twitter for digital engagement. *Journal of Contemporary Eastern Asia*, *18*(1).

Kampf, R., Manor, I., & Segev, E. (2015). Digital diplomacy 2.0? A cross-national comparison of public engagement in Facebook and Twitter. *The Hague Journal of Diplomacy*, *10*(4), 331–362.

Khemani, B., & Adgaonkar, A. (2021). A review on reddit news headlines with nltk tool. *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*.

Li, Z., Zhang, X., Zhang, Y., Long, D., Xie, P., & Zhang, M. (2023). Towards general text embeddings with multi-stage contrastive learning. *ArXiv Preprint ArXiv:2308.03281*.

Rybalko, S., & Seltzer, T. (2010). Dialogic communication in 140 characters or less: How Fortune 500 companies engage stakeholders using Twitter. *Public Relations Review*, *36*(4), 336–341.

Sah, A., Chatterjee, R., & Gourisaria, M. K. (2022). Machine Learning Approaches to Assess Mood of the News Editorial. *2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 1–6.

Strauß, N., Kruikemeier, S., van der Meulen, H., & van Noort, G. (2015). Digital diplomacy in GCC countries: Strategic communication of Western embassies on Twitter. *Government Information Quarterly*, *32*(4), 369–379. https://doi.org/10.1016/j.giq.2015.08.001

Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B. (2024). Benchmarking large language models for news summarization. *Transactions of the Association for Computational Linguistics*, *12*, 39–57.