

RESEARCH ARTICLE

Mapping Japan's China-related twitter discourse (2010–2024) using BERTopic

Shanshan Zhang^{1*}, Xi Chen²

1 Graduate School, Xi'an International Studies University, Xi'an, China, **2** School of Japanese Culture and Economics, Xi'an International Studies University, Xi'an, China

* zhangshanshan.xisu@gmail.com



Abstract

Given China's profound influence on Japan, Japanese public opinion toward China has been the focus of debate. Yet little is known about how such perceptions evolve over time within the digital public sphere. Drawing on a dataset of over one million China-related tweets from Japan (2010–2024), this study integrates BERTopic topic modeling with large language model-driven sentiment analysis to trace the dynamic evolution of Japanese perceptions of China. Empirical analyses find that: (1) Public attention to China has steadily increased, concentrating on four domains: diplomacy & security, environment & health, economy & trade, and culture & society; (2) overall sentiment is predominantly negative, with neutral and positive attitudes appearing in specific topics; and (3) topic-sentiment linkage analysis reveals divergent affective tendencies across topics, indicating that public opinion evolves in response not only to external events but also to the inherent characteristics of topics. By applying computational analysis to large-scale social media data, this study uncovers the dynamic structure of Japanese public opinion regarding China, offering insights into the mechanisms of opinion formation with implications for Sino-Japanese relations. Methodologically, it contributes innovative approaches to the analysis of transnational public discourse.

OPEN ACCESS

Citation: Zhang S, Chen X (2026) Mapping Japan's China-related twitter discourse (2010–2024) using BERTopic. PLoS One 21(2): e0343085. <https://doi.org/10.1371/journal.pone.0343085>

Editor: Pablo Henríquez, Universidad Diego Portales, CHILE

Received: September 25, 2025

Accepted: January 30, 2026

Published: February 18, 2026

Copyright: © 2026 Zhang, Chen. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The datasets analysed during the current study are available in the Zenodo repository, <https://doi.org/10.5281/zenodo.17170345>.

Funding: The Graduate Research Fund of Xi'an International Studies University (No. 2025BS007).

Introduction

China and Japan, the world's second- and third-largest economies, play a pivotal role in shaping regional security and global prosperity. Bilateral relations not only affect their domestic affairs but also exert significant influence on the stability of East Asia and beyond. Yet Japan's long-standing "U.S.-centred" foreign policy tradition, coupled with the rise of anti-China sentiment since the mid-2000s, has made the relationship increasingly complex [1]. More recently, the COVID-19 pandemic, the intensification of U.S.–China rivalry, and China's rapid rise have forced Japan to reconsider its geo-strategic positioning and pursue regional diplomacy to ease strategic anxieties [2].

Competing interests: The authors have declared that no competing interests exist.

Against this backdrop, public perceptions of China have become a crucial factor in shaping Japan's foreign policy calculus. In representative democracies, public opinion both constrains policy choices and provides critical cues for political decision-making [3–5]. Understanding how the Japanese perceive China is thus central to explaining the social foundations of Sino-Japanese relations and anticipating future policy orientations. Prior studies have examined the effects of state policy, territorial disputes, major international events, and media narratives on Japanese attitudes toward China [6–10]. Yet such research—largely reliant on survey data or short-term event-driven observations—tends to reduce public opinion to a binary “positive–negative” frame [11–13]. As a result, gaps remain in comprehensively analysing the nuances of public perceptions and their temporal evolution.

Social media platforms such as Twitter have become central arenas for public expression and political debate [14,15], providing unique opportunities to observe Japanese views of China [16,17]. Compared to traditional opinion polls, social media data are larger in scale, more dynamic, and more diverse [18]. Fine-grained analyses can thus capture real-time fluctuations in sentiment and reveal the thematic structures underlying public discourse. Despite these advantages, few studies have leveraged social media to systematically trace the long-term dynamics of Japanese public opinion toward China.

To address this gap, this study applies large language model-based topic modeling and sentiment analysis to more than one million China-related tweets posted in Japan between 2010 and 2024. Specifically, it seeks to answer two research questions:

RQ1: What thematic patterns characterize Japanese social media discourse on China, and how have they evolved over time?

RQ2: What sentiment orientations toward China emerge in these discussions, and how do they vary across topics and over time?

By combining methodological innovation with longitudinal scope, this study contributes in three ways. Empirically, it offers a fine-grained mapping of Japanese discourse on China across a fifteen-year period. Methodologically, it demonstrates the utility of large-scale computational approaches for integrating topic structures and sentiment dynamics. Theoretically, it deepens our understanding of how digital public opinion interacts with foreign policy, offering insights into the societal underpinnings of Sino-Japanese relations and, more broadly, the role of public sentiment in international politics. Together, these contributions highlight the value of computational social science approaches in capturing the complex, evolving relationship between public opinion and foreign policy.

Related work

The literature review is organized around three aspects. First, it systematically surveys existing scholarship on Japanese public opinion toward China. Second, it reviews the methodological advances in topic modeling within social media research. Third, it offers a comprehensive assessment of studies on sentiment analysis in social media, highlighting both their strengths and limitations.

Research on Japanese public opinion toward China

Japanese perceptions of China have profound implications for bilateral relations and policymaking [19]. Existing research has largely focused on elite-level dynamics, including diplomatic interactions, economic policy, and mainstream media narratives, examining how these actors shape public perceptions and issue framing. For example, Yoshimatsu finds that Japan's response to the Belt and Road Initiative combines normative frameworks such as "quality infrastructure" with narrative diplomacy, reflecting efforts to maintain international influence amid relative power shifts [2,20]. Media discourse studies further suggest that Japanese coverage of the Belt and Road Initiative exhibits caution and skepticism, which may exacerbate bilateral tensions and negatively influence public attitudes [10,21]. On specific policy issues, Shao emphasizes that discourse about China in Japan reflects competition among multiple political actors rather than a single national narrative [9], while Chen & Zhang highlight a strategic shift in Japan's China policy from traditional "hedging" toward a dual-track approach balancing economic engagement with security concerns [22].

At the public-perception level, studies have predominantly employed survey methods to examine impressions, attitudes, and policy preferences toward China. Research generally indicates that Japanese attitudes toward China have become increasingly negative [6,11,13,23], shaped by an interplay of factors such as geopolitical tensions, cultural proximity, historical memory, and perceived security threats [6,11,12]. Unlike the economically oriented evaluative patterns observed in many developing countries, the Japanese public tends to view China through an ideological lens [13]. Empirical analyses further suggest that Japanese attitudes toward China are influenced not only by economic expectations but are particularly sensitive to perceived political threats [7]. Moreover, China's promotion of the "peaceful rise" narrative has been identified as a major factor exacerbating negative perceptions [23].

Recently, a growing body of research has turned to social media data to investigate the dynamics of China-related public opinion during specific events. For instance, in international debates over Japan's nuclear wastewater discharge, China was framed as the primary opposing actor [24]. During the COVID-19 pandemic, China remained central to Japanese social media discourse, with public sentiment characterized predominantly by anxiety and concern, although emotional intensity moderated after April 2020 [8]. While these studies demonstrate the potential of social media for capturing public opinion dynamics, they remain event-specific and cross-sectional, offering limited insight into long-term or evolving patterns of sentiment. As a result, the longitudinal dynamics of Japan-related digital discourse on China remain underexplored.

Topic modeling in social media research

Social media content presents unique analytical challenges due to its fragmented nature, hierarchical topic structures, and multidimensional expressions [25]. For extracted Twitter data, Latent Dirichlet Allocation (LDA) is a commonly used topic modeling method that has long been applied to identifying public opinion topics and their temporal dynamics [26]. However, this approach typically suffers from issues such as semantic ambiguity of topics, limited ability to process multilingual and noisy text, and poor performance with short texts [27], making it difficult to extract valuable information from social media data and prone to overlooking niche yet significant topics [18].

With the rapid advances in machine learning, BERTopic-based topic modeling has emerged as a method capable of extracting dynamically evolving topic units with high semantic precision, automatically generating hierarchical topic structures, and offering flexible combinations of model components [28]. This approach not only accurately captures mainstream discourse topics but also highlights marginalized or overlooked voices, making it particularly suitable for handling unstructured, short-text, and multilingual social media data [29–33]. For instance, Murayama et al. analysed over 6.8 million tweets across 14 languages worldwide, demonstrating the method's adaptability for cross-hierarchical and multilingual analysis [29]. Veigel et al. tracked shifts in public cognition in German tweets surrounding flooding events since 2014, highlighting BERTopic's strengths in modeling dynamic topic evolution [30]. The widespread application of BERTopic to

large-scale multilingual datasets and event-specific social media analysis underscores its precision for large-scale language analytics. These advances indicate that BERTopic is a powerful tool for analysing and exploring long-term, dynamic patterns in public discourse.

Sentiment analysis

Sentiment analysis represents another key dimension in social media research, aiming to uncover the distribution and temporal evolution of public attitudes and emotions. Early lexicon-based methods inferred polarity and intensity from sentiment-laden words [34,35], performing adequately on small annotated corpora but struggling with contextual ambiguity and nuanced semantic interpretation. Subsequently, researchers employed traditional machine learning methods based on feature engineering, such as support vector machines and logistic regression, to improve classification performance, yet these approaches are constrained by heavy reliance on manually constructed features and limited generalizability [36].

In recent years, neural network-based pre-trained language models have demonstrated remarkable efficacy in tweet sentiment classification [37,38]. BERT-based sentiment analysis, in particular, substantially improves accuracy by capturing contextual dependencies and deep semantic representations [39,40]. Empirical evidence consistently shows that BERT and its variants outperform traditional methods in multilingual social media sentiment analysis [41–44]. However, most existing research has focused on English or Spanish, and sentiment studies on Japanese public social media discourse remain relatively scarce.

Moreover, recent research highlights the significant role of demographic characteristics in shaping emotional expression on social media. Gender, in particular, has been consistently shown to affect the intensity of emotional expression, engagement with toxic content, and persuasive outcomes [45,46]. This body of research underscores that online emotional expressions and interaction patterns are shaped not only by textual content and semantic structures but also by user attributes. These findings provide an important contextual foundation for understanding the distribution of emotions and interaction mechanisms in social media environments.

In summary, previous research on Japanese public opinion toward China mainly fell short in three aspects. First, the longitudinal evolution of Japanese attitudes toward China remains insufficiently studied, particularly in digital discourse. Second, social media studies have often been limited to single events or cross-sectional snapshots. Third, methodological tools such as topic modeling and sentiment analysis, while promising, have seen limited application to Japanese data and are rarely integrated to capture dynamic, multi-dimensional patterns of discourse. Building on a large-scale dataset of Japan-related tweets, this study offers an in-depth examination of topic evolution and sentiment polarization, providing novel theoretical perspectives and empirical insights for understanding Japanese public opinion toward China.

Data and methodology

This section introduces the data and research methods, including the establishment of the dataset, construction of hierarchical topic structures, the analysis of topic and sentiment evolution. Fig 1 presents the overall research framework, illustrating the process from Twitter data to topic and sentiment evolution analysis.

Data construction

Building on the function of hashtags in delineating discursive communities on social media [47], this study uses “#中国” and “#中華” as search keywords to systematically capture the contours of Japanese public discussions concerning China. Prior studies suggest that, in public discourse on foreign-related issues, country names frequently serve as salient linguistic markers that explicitly signal the focal object of discussion and structure discursive meaning [48]. Building on this insight, this study employs highly recognizable country-name labels as the primary entry point for large-scale data collection. This strategy allows for a clear operational delineation of the research object while capturing the overwhelming

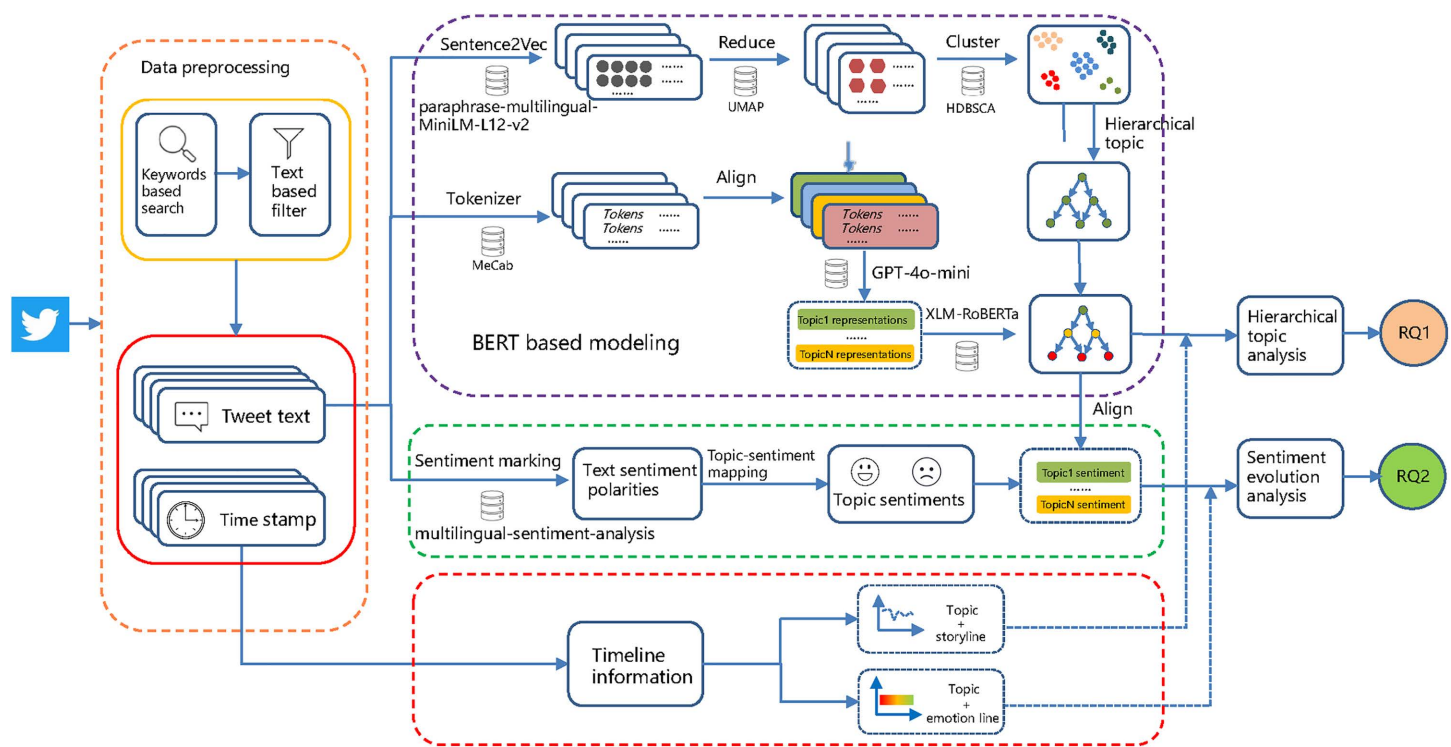


Fig 1. Overview of research framework from twitter data collection to the analysis of topic and sentiment dynamics.

<https://doi.org/10.1371/journal.pone.0343085.g001>

majority of China-related discussions among Japanese users, and is therefore unlikely to introduce systematic bias into the results. Using the Twitter API, all Japanese-language tweets containing these hashtags between January 1, 2010, and December 31, 2024, were collected, resulting in a 15-year corpus comprising over one million entries (N = 1,310,786). The decision to select 2010 as the starting point is based on the incident in which Japan detained a Chinese fishing vessel near the Diaoyu/Senkaku Islands, a watershed moment that marked the onset of a new phase of strategic competition and frequent tensions in Sino-Japanese relations. This turning point inevitably exerted a profound influence on public discourse practices. Accordingly, the present study adopts 2010 as the baseline, analysing 15 years of public discussions to capture the thematic and emotional dynamics of Japanese discourse on China.

The retrieved tweets included user ID, posting time, user nickname, user name, tweet content, number of likes, and retweet counts. Tweets were classified into four main types: original tweets, retweets, quote tweets, and reply tweets. All original tweets were retained as independent entries in the corpus. Retweets, being semantically identical to the original tweets, can create an “amplification effect” that disproportionately affects topic distributions; to maintain the stability of topic representation, retweets were removed. Quote tweets, by contrast, often include user-added commentary when referencing the original tweet, contributing additional semantic content; such commentary was extracted and treated as a separate entry. Reply tweets, given their interactive and often brief, colloquial nature, were also excluded from the corpus.

It should be noted that this study aims to identify China-related discussion topics among the Japanese public. However, as Twitter has continuously strengthened user privacy protections, accurate location information is unavailable for the vast majority of tweets. Accordingly, following a commonly adopted proxy approach in social media research, we define the analytical object as “China-related tweets written in Japanese.” This approach may introduce a limited number of tweets

authored in Japanese by non-Japanese users; however, the proportion of non-native users tweeting in Japanese is relatively small and unlikely to bias the results of this study.

For the raw data, duplicate tweets were merged and removed based on user ID and posting time, and completely identical tweets were deleted, resulting in a final dataset of 1,099,786 valid tweets.

The preprocessing procedure consisted of three steps. First, data cleaning was conducted to remove HTML tags, URLs, emojis, and extraneous whitespace, while normalizing full-width and half-width characters. Second, the tweet text and corresponding timestamps were retained, with key information systematically extracted. Third, Japanese text was segmented using the MeCab tokenizer, which integrates authoritative dictionaries including IPA, UniDic, and Juman. Owing to its extensive dictionary coverage and support for user-defined lexicons, MeCab delivers precise tokenization even in complex linguistic contexts.

Methodology

To address the two research questions, we conducted hierarchical topic extraction, topic evolution analysis, and sentiment evolution analysis.

Hierarchical topic extraction. This study employed BERTopic for constructing a hierarchical topic model. The detailed procedure is outlined as follows:

- (1) Pre-processed tweets were first embedded using the paraphrase-multilingual-MiniLM-L12-v2 model [49], a sentence-transformer that generates embeddings at the sentence level to preserve semantic integrity. In this study, we set the `max_length` parameter of the paraphrase-multilingual-MiniLM-L12-v2 model to 256;
- (2) The resulting high-dimensional vectors were reduced via UMAP [50]. UMAP is based on manifold learning and is well suited for handling complex, nonlinear high-dimensional manifold data, while preserving both local neighborhood structure and global structure. The main hyperparameters of the UMAP model are set as follows: `n_neighbors=20`, `n_components=5`, `min_dist=0.05`, `metric='cosine'`;
- (3) The reduced embeddings were clustered using HDBSCAN [51], a density-based clustering method that also provides a hierarchical structure. The key hyperparameters of the HDBSCAN model were configured as follows: `min_cluster_size=50`, `min_samples=15`, and `metric='euclidean'`;
- (4) Topic labels were extracted using the c-TF-IDF algorithm [28]. First, for the tweet vocabulary obtained through tokenization with the MeCab tool, we applied a standard Japanese stop-word list to remove common stop words. The remaining terms were subsequently ranked by frequency, then three Japanese language experts conducted an iterative manual screening of high-frequency terms in light of corpus characteristics, removing functional words and items without substantive semantic contribution to construct the final vocabulary. This final vocabulary was subsequently used as input for c-TF-IDF to extract topic labels. To reduce noise and prevent high-frequency terms shared across topics from dominating topic representations, we set `min_df=10`, `max_df=0.7`, and adopted an `ngram_range` of (1,2);
- (5) GPT-4o-mini was employed to generate concise and semantically precise topic summaries. The prompt used for GPT-4o-mini was: "You are an academic assistant. Give a brief topic description based strictly on the provided keywords. Do not introduce new concepts or interpretations";
- (6) These summaries, aligned with the original tweets, were then processed through the XLM-RoBERTa model [52] to map them to the lowest hierarchical layer, resulting in 62 fine-grained topics. This hierarchical mapping links each topic level to the original tweets, enabling a comprehensive transition from raw text to structured hierarchical labels and facilitating detailed analysis of topic evolution.

Topic evolution analysis. Topic evolution analysis aims to reveal the dynamic changes of topics over time, events, or environmental shifts. We adopted the method proposed by Wei et al.[18], which analyses topic evolution based on semantic similarity, to conduct an integrative study of China-related topics on Japanese social media.

Specifically, topic keywords were initially extracted using the c-TF-IDF algorithm. Keywords from 2010 were selected as the “starting words,” and for each subsequent year, the semantic similarity between the current year’s keywords and the previous year’s keywords was calculated. The five keywords most similar to the previous year were selected to represent the evolution of the topic during that period. In this study, topic labels were converted into vectors using the paraphrase-multilingual-MiniLM-L12-v2 model, and cosine similarity was used to measure the semantic relationships between keywords, as expressed by the following formula:

$$\text{Cosine Similarity } (A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

Here, A and B represent the vector embeddings of two topic keywords, \cdot denotes the dot product, and $\|A\| \|B\|$ represents the norms of vectors A and B. Cosine similarity is used to measure the similarity between two topics.

By tracing the temporal evolution of topics and quantifying their discussion intensity, we quantified topic fluctuations and integrated dispersed information into a coherent narrative, thereby elucidating the evolution and underlying drivers of Japanese public discourse on China.

Sentiment evolution analysis. To investigate the distribution and evolution of sentiment in Japanese discussions on China, we employed the multilingual-sentiment-analysis model [53] to assess the sentiment of each tweet. This model is based on the mBERT architecture and pre-trained on large-scale multilingual corpora, enabling it to capture fine-grained sentiment features with high accuracy. The model classifies each tweet into five categories: very negative, negative, neutral, positive, or very positive, and provides a confidence score. Using these classifications, we analysed the sentiment distribution for each topic, thereby capturing the evolution and fluctuation of public emotions within each topic.

To visualize the temporal dynamics of Japanese public discourse on China, we employed sentiment heatmaps alongside stacked charts of topic-sentiment evolution. By integrating time, topic distribution, and sentiment polarity, our analysis provides a dynamic perspective on sentiment, revealing fluctuations within the same topic over different periods and comparisons of sentiment across topics at the same point in time. Through sentiment evolution analysis, we aim to capture subtle trajectories in Japanese attitudes toward China, providing robust empirical evidence for understanding the dynamic evolution of public opinion underpinning Sino-Japanese relations.

Results

The study results include: (1) an overview of hierarchical topic structure; (2) a topic evolution view integrating fragmented information; and (3) sentiment distribution charts and topic-sentiment linkage analysis.

Overview of hierarchical topic structure

To address our first research question (RQ1), the distribution and temporal evolution of China-related topics on Japanese social media, we analysed all relevant Japanese tweets from 2010 to 2024, counted the annual number of tweets (Fig 2), and clustered them into a hierarchical topic structure (Fig 3). As shown in Fig 2, the total number of China-related tweets increased steadily over time. Fig 3 presents a sunflower chart of the hierarchical topic architecture, displaying a three-layered structure. The top layer consists of 10 core topics: “Diplomacy & Security,” “Culture & Society,” “Environment & Health,” “Economy & Trade,” “Policy & Governance,” “Technology & Industry,” “Education,” “Law & Human Rights,” “Entertainment & Sports,” and “Historical and National Identity.” The bottom layer clearly presents 62 fine-grained topics, covering mainstream public discourse as well as smaller but critical topics such as AI achievements, traditional Chinese

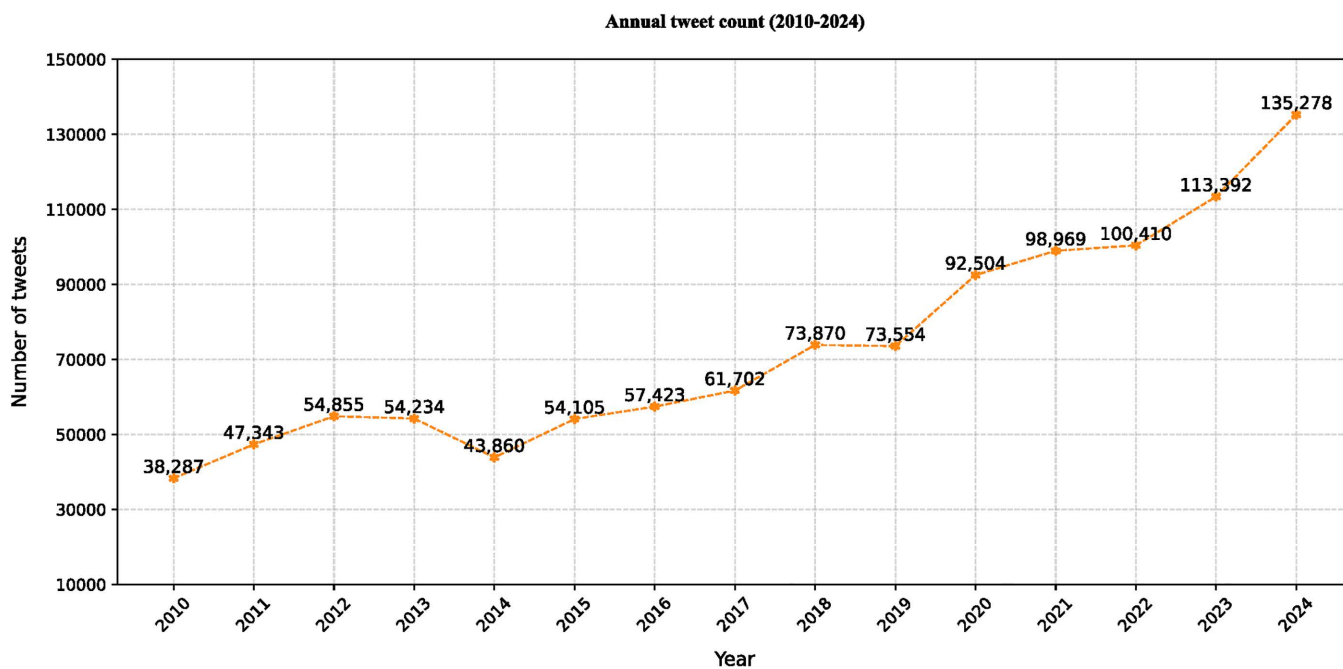


Fig 2. Annual number of China-related tweets.

<https://doi.org/10.1371/journal.pone.0343085.g002>

medicine, and nuclear energy. This figure illustrates the hierarchical structure of China-related discourse on Japanese social media, visually conveying the categories and nested relationships among topics.

[Table 1](#) presents the codebook for the ten core China-related topics on Japanese social media. It describes the topics within each top-level category and provides example tweets for each category. A detailed analysis of the topics within the hierarchical structure is presented in the Discussion section.

In addition to the fine-grained hierarchical analysis, the temporal trends in topic discussion intensity provide rich insights. As shown in the stream and bar charts ([Fig 4](#)), the annual number of tweets related to the ten core China-related topics on Japanese social media is unevenly distributed over time, exhibiting distinct peaks. [Fig 4a](#) shows that the discussion intensity of the “Diplomacy & Security” topic increased annually, peaking in 2019; the “Environment & Health” topic experienced a significant surge in 2020 due to the pandemic. “Culture & Society”, “Economy & Trade” and “Technology & Industry” showed a fluctuating upward trend since 2021, reflecting the gradual increase in the visibility of these topics in Japanese public opinion alongside China’s rapid economic development.

[Fig 4b](#) illustrates a pronounced long-tail distribution across topic categories [54]. The four topics “Diplomacy & Security”, “Environment & Health”, “Economy & Trade”, and “Culture & Society” dominate discussions, together comprising 69.87% of all tweets. The remaining six topics show comparatively balanced tweet volumes, with “Historical and National Identity” being the least represented. This long-tail distribution of China-related topics on Japanese social media indicates that although public attention is primarily focused on the four major topics, niche topics such as “Entertainment & Sports” and “Historical and National Identity” should not be overlooked.

Topic evolution view

This study integrates fragmented public discourse data into a coherent and easily interpretable narrative view. [Fig 5a–5c](#) illustrate the evolution of China-related topics on Japanese social media from 2010 to 2024, contextualized with major China-related events for each year. The horizontal axis represents time. Bubbles of different colors denote different

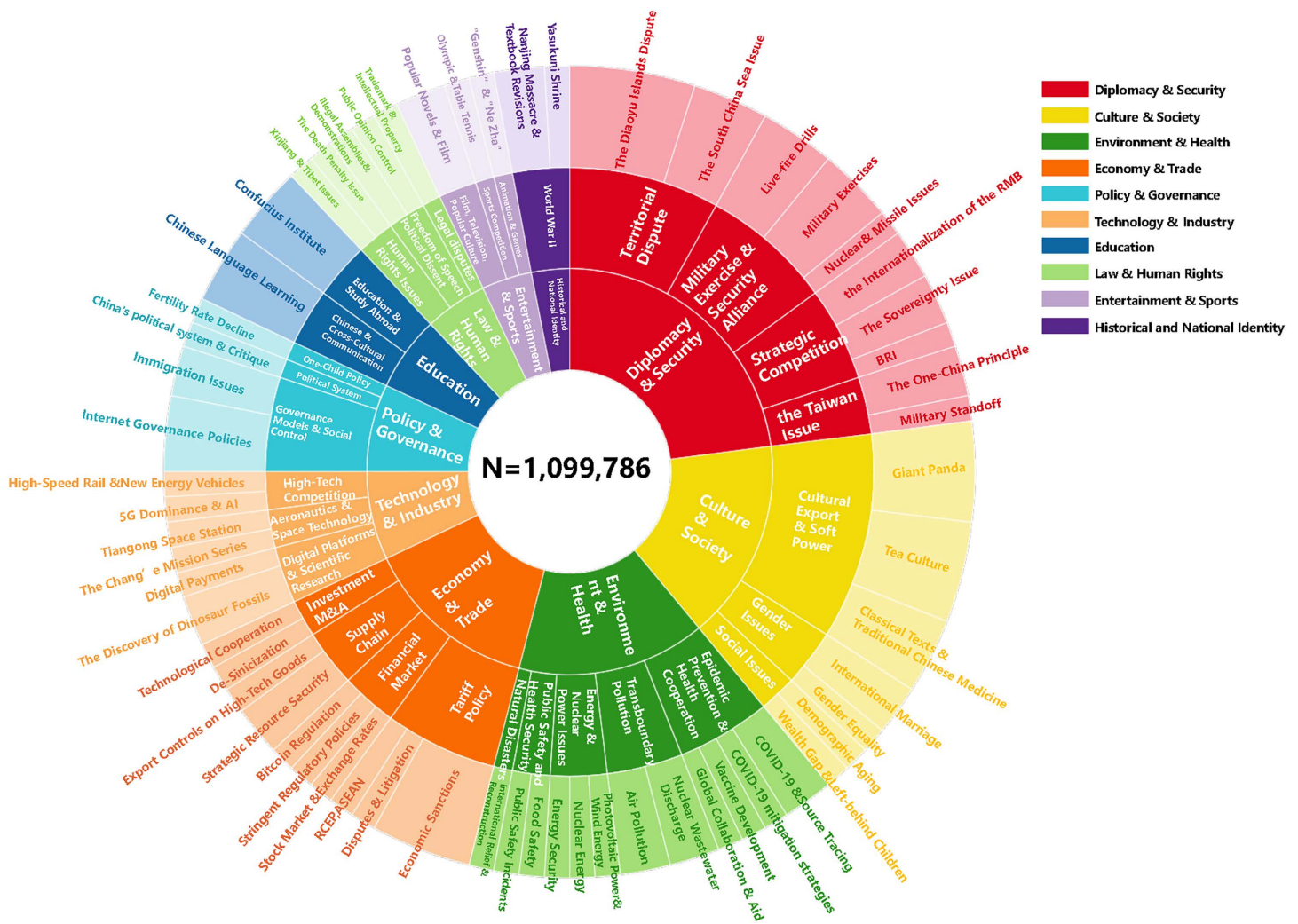


Fig 3. A hierarchical thematic structure of China-related discourse on Japanese twitter.

<https://doi.org/10.1371/journal.pone.0343085.g003>

topics, the bubble size represents the discussion intensity of the corresponding topic, with the radius determined by the normalized number of tweets contained within that topic. The distance between bubbles reflects the degree of correlation between topics. We measured topic correlations using the Spearman correlation coefficient [55]. Spearman's ρ was computed at a significance level of $\alpha=0.05$, and converted to distances using the mapping function $d_{ij} = 1 - \rho_{ij}$. These distances were then normalized and used to position the bubbles. The smaller the distance between two bubbles, the stronger the correlation between the corresponding topics.

We applied HDBSCAN (with parameters $\text{min_cluster_size} = 15$, $\text{min_samples} = 5$, and $\text{metric} = \text{'euclidean'}$) to the tweet embeddings within each topic for a second-stage clustering to identify event-level subclusters that are both semantically cohesive and temporally concentrated, thereby extracting specific events. Event intensity was then calculated based on the number of tweets and the temporal window as:

$$EventIntensity = \frac{N}{\Delta t} \tag{2}$$

Table 1. The codebook of China-related discourse on the ten topics.

Topics	Percentage	Definition	Example Tweets
Diplomacy & Security	23.68%	This topic refers to diplomatic interactions, military deployments, and geopolitical strategies between China and Japan, as well as within the broader regional security landscape.	“中国漁船団に捕獲してもらいましょう。RT @suka1975: 捕獲しなければどっかに流れて行くでしょう。”
Culture & Society	15.24%	This topic refers to perceptions of China's cultural image, social structure, and cross-cultural interactions.	“昨夜NHKで 中国に渡った上野生まれのパンダ シャンシャンを見て 今更ファンになる、”
Environment & Health	15.60%	This topic refers to China's environmental governance, pandemic control, and energy policies.	“中国の新型コロナウイルスの肺炎。来月あたり中国旧正月だよな?日本にもたくさん中国人くるよな?新型コロナウイルスって治療薬あんの?”
Economy & Trade	15.35%	This topic refers to discussions of Sino-Japanese bilateral economic and trade relations, regional economic interdependence, and global market dynamics.	“国際通貨基金(IMF)の見通しによると、中国の名目国内総生産(GDP)が今年には日本を抜いて世界第2位になる、とのこと。”
Policy & Governance	5.97%	This topic refers to China's governance model, policy implementation, and institutional structures, covering a wide range of issues from family planning to political systems and media control.	“一人っ子政策のおかげで人口激減が必至ですからね。”
Technology & Industry	7.08%	This topic centers on Japanese users' collective perceptions and discourses regarding China's technological development and industrial transformation.	“日本の生成AI技術が衰退したら任天堂と提携してる中国のテンセントやアリババの生成AI技術が参入してくるだけだと、反AIさん本当に理解しないな。”
Education	5.40%	This topic refers to China's education system, language promotion, and study-abroad exchanges, Confucius Institutes and the surge in Chinese language learning.	“孔子学院は立命館が第一号で、桜美林・山梨学院など全国に15校程ある模様。”
Law & Human Rights	4.43%	This topic focuses on China's legal system, human rights conditions, reflecting Japanese users' critical concerns regarding China's human rights governance and judicial independence.	“香港デモの途中まで、俺が多少香港警察に同情的というか、キッズに舐めた真似されてるのによくガマンしてるなあ、デモ隊側の被害者アピールもひどいなあとと思った一因に「日本の警察のほうの本気出したら暴力的」というのはあった。”
Entertainment & Sports	4.24%	This topic focuses on China's film and television, idols and popular culture exchanges, as well as sports competitions.	“原神を最初に知ったのはブレワイのバクリ記事からで、開発運営が中国だったので、かなり警戒してたけど、メーカーの事を調べたら、かなりオタク寄りの考えを持っていたので、萌え至上主義の私としては良し良しである(-_-)。”
Historical and National Identity	3.01%	This topic refers to historical events and memories related to China and Japan, encompassing issues such as war responsibility and cultural heritage.	“南京大虐殺を巡る日中間の共同研究。結局、30万人以上の中国側と、20万人までの日本側と、学問上の論争に決着つかず。まあ、無かったと言ってる人もいますけどね。”

<https://doi.org/10.1371/journal.pone.0343085.t001>

where N denotes the number of tweets in a given event and Δt represents its duration. Based on this metric, one or two high-intensity representative events were selected for each year as major events.

Using the 2010 keyword “Google exits China” (「グーグル撤退」) as the initial point of the evolution pathway, we employed semantic similarity clustering to identify the five most semantically related keywords, such as “Taobao” (「タオバオ」) and “Apple launch” (「アップル発売」). Keywords from the latest timestamp in the previous year subsequently evolve into the five keywords for the following year, continuing iteratively.

As shown in Fig 5, “Diplomacy & Security” (red) consistently appears and maintains a prominent position from 2010 to 2024. Key issues within this topic include the Taiwan issue, North Korea's nuclear program, China–Vietnam disputes in the South China Sea, the Sino–Japanese Diaoyu/Senkaku Islands dispute, and Sino–U.S. confrontation. These discussions are typically triggered by major diplomatic events and historical issues, such as the 2010 detention of a Chinese fishing boat by Japan, which sparked large-scale anti-Japan protests in China; North Korea's missile launch and Japan's “nationalization” of the Diaoyu/Senkaku Islands in 2012; China's intensified East China Sea patrols in 2013; and the escalation of Sino–U.S. strategic competition in 2018.

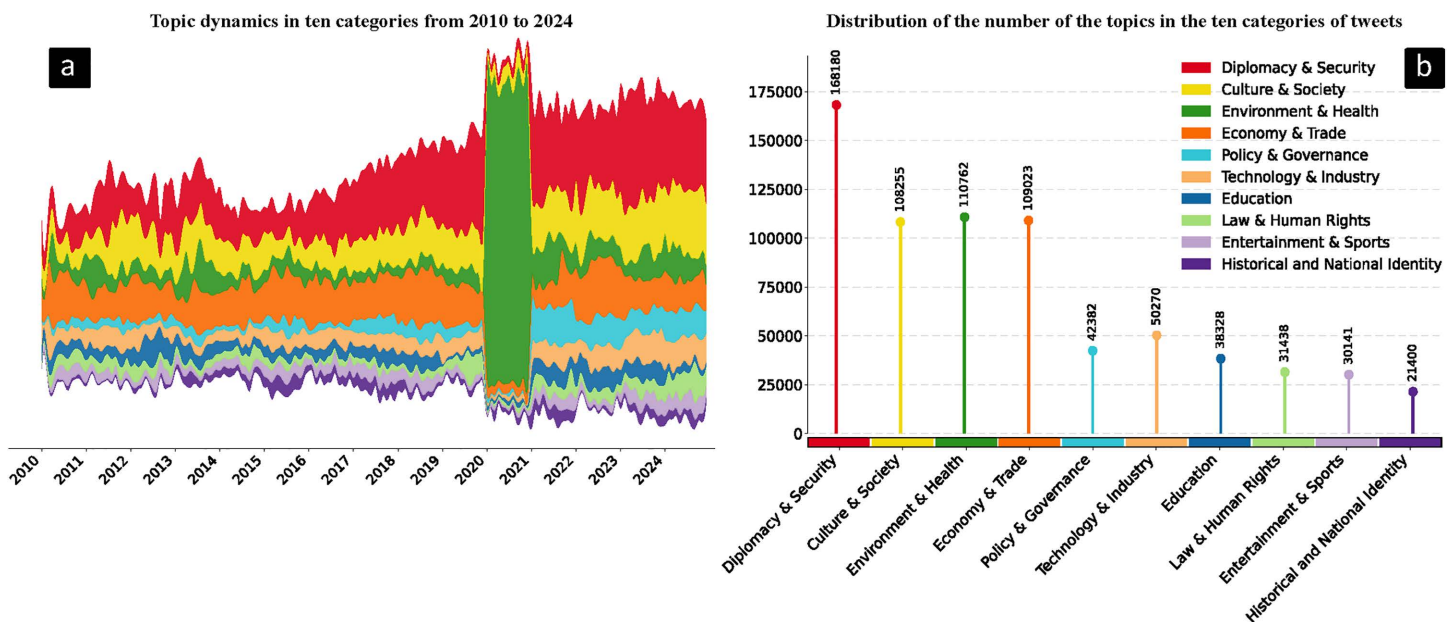


Fig 4. Temporal trends and topic distribution of China-related discourse on Japanese twitter. **a** Temporal trends showing the volume of tweets across the ten core topics over time; **b** Distribution of the ten topics ranked by tweet volume from highest to lowest.

<https://doi.org/10.1371/journal.pone.0343085.g004>

The “Economy & Trade” (orange-red) and “Technology & Industry” (creamy orange) topics exhibit strong associations with “Diplomacy & Security”. Early discussions focused on economic scale and market opportunities, including events such as China surpassing Japan in nominal GDP in 2010, Taobao reaching 370 million members, and the launch of the iPhone 4 in China. Subsequently, technological innovation and outbound investment became focal points, including the establishment of the Asian Infrastructure Investment Bank (2015), China’s high-speed rail expansion abroad (2016), cryptocurrency regulation in China (2017), Huawei and BYD’s overseas developments (2018, 2023), and breakthroughs in AI technology (2024).

The “Environment & Health” (green) topic encompasses narratives related to accidents, epidemics, and food safety, with discussion peaks largely triggered by public safety incidents, including the Wenzhou train collision (2011), the H7N9 avian influenza outbreak (2013), the Fuxi food contamination scandal (2014), and the COVID-19 pandemic (2020). By contrast, topics such as “Culture & Society”, “Law & Human Rights”, “Education”, and “Entertainment & Sports” maintained relatively stable discussion levels, centering on cultural exchange, Chinese language learning, sporting events, and select societal incidents.

Sentiment distribution

To explore the sentiment distribution and evolution patterns in China-related discourse on Japanese social media (RQ2), this study employed the multilingual- sentiment-analysis pre-trained model to perform fine-grained sentiment classification on the collected China-related tweets, categorizing them into five classes: very negative, negative, neutral, positive, and very positive.

This study found that the overall sentiment distribution of China-related discourse on Japanese social media is presented in [Table 2](#):

[Table 2](#) indicates that the sentiment of Japan’s social media discourse on China is predominantly negative, with very negative and negative sentiments together representing 42.5%, compared to just 30.3% for very positive and positive sentiments, highlighting a clear predominance of unfavourable evaluations over favourable ones.

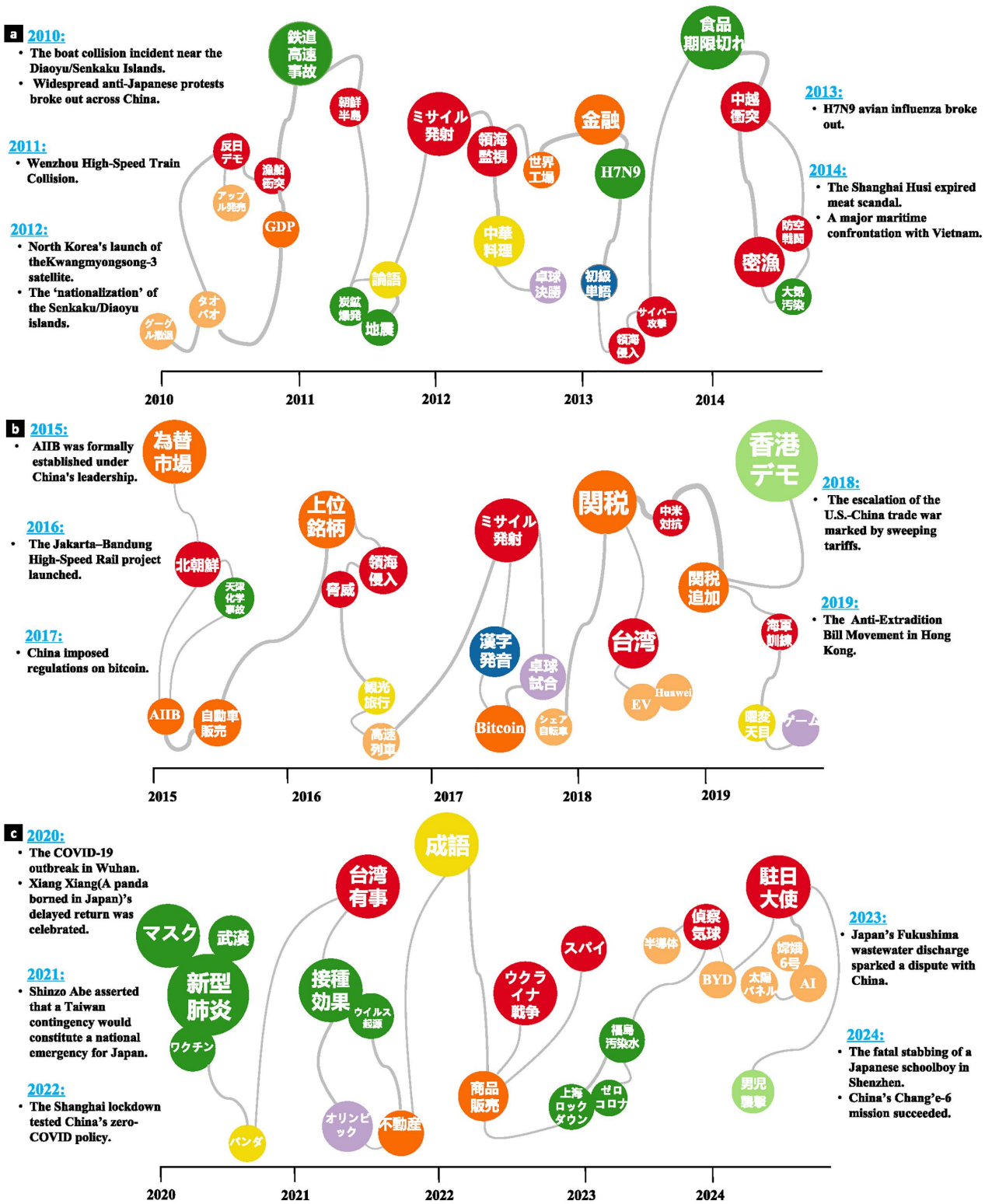


Fig 5. Topic evolution view of China-related discourse on Japanese twitter (2010-2024). a shows period from 2010-2014, using the topic "Google's exit from the Chinese market" as an example. b shows period from 2015-2019, While c shows period from 2020-2024. The same color represents topics belonging to the same category within the ten topics of China-related discourse. The x-axis in the figure represents the timeline, with the proximity of topic bubbles indicating the degree of semantic relatedness between topics. The size of the bubble represents the number of topics. The bigger the

bubble representing a topic is distributed in the graph, the larger its relative quantity. Note that this graph is a schematic, and the time and quantity do not represent precise values.

<https://doi.org/10.1371/journal.pone.0343085.g005>

Table 2. Overall sentiment distribution of China-related discourse on Japanese twitter.

Very Negative	Negative	Neutral	Positive	Very Positive
35.5%	7.0%	27.2%	8.3%	22.0%

<https://doi.org/10.1371/journal.pone.0343085.t002>

Fig 6 illustrates the sentiment distribution across the ten core topics. As shown, the proportion of neutral sentiment in each topic is roughly similar, ranging from 20% to 30%, indicating that approximately one quarter of China-related discussions maintain a relatively neutral tone. In addition, the share of positive and negative sentiments is relatively small across all topics, whereas the proportions of very positive and very negative sentiments are comparatively larger.

As shown in Fig 6, the proportion of very negative sentiment is notably high in five domains: “Diplomacy & Security” (39.2%), “Policy & Governance” (42.2%), “Environment & Health” (42.8%), “Law & Human Rights” (53.9%), and “Historical and National Identity” (40.1%). In contrast, the proportion of very positive sentiment is markedly higher in five topics: “Culture & Society” (31.2%), “Economy & Trade” (28.3%), “Technology & Industry” (27.0%), “Education” (26.4%), and “Entertainment & Sports” (35.6%) compared to other domains.

Topic-sentiment linkage analysis

Drawing on the sentiment heatmap (Fig 7) and the stacked charts of topic-sentiment evolution (Fig 8), this study uncovers the temporal dynamics of sentiment across different topics within Japan’s social media discourse on China.

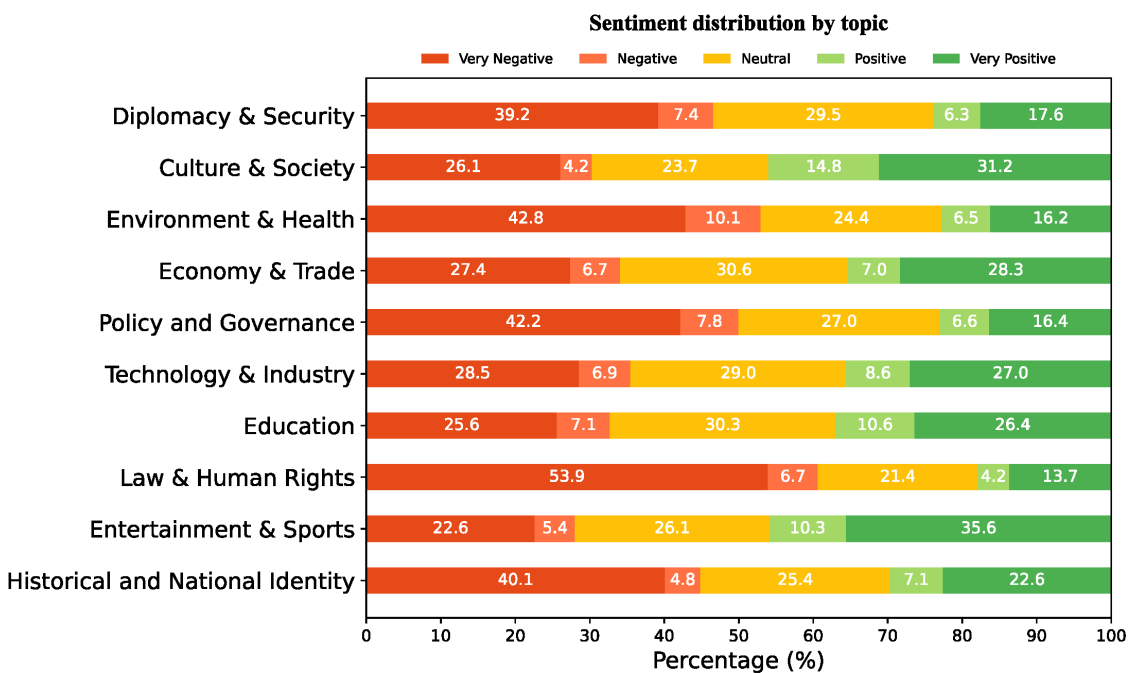


Fig 6. Sentiment distribution across the ten topics. The figures indicate the proportion of tweets in each of the five sentiment categories (very negative, negative, neutral, positive, very positive) as a percentage of the total tweets under each topic.

<https://doi.org/10.1371/journal.pone.0343085.g006>

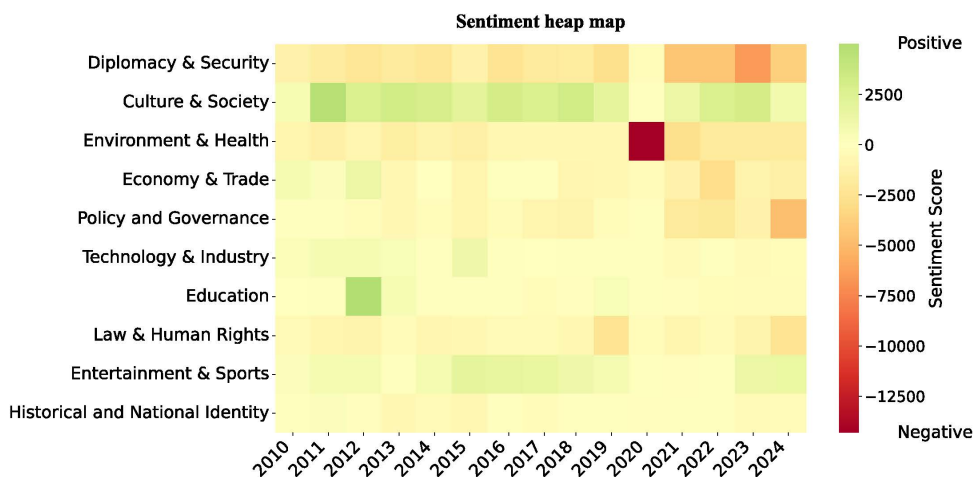


Fig 7. Overall sentiment evolution in China-related discourse on Japanese twitter (2010–2024). It presents a sentiment heatmap organized by year and topic. Color intensity reflects the sentiment score, calculated annually by summing the sentiment values of all tweets within each topic (very negative = -2, negative = -1, neutral = 0, positive = +1, very positive = +2). Darker colors indicate stronger sentiment polarity, whether positive or negative. The horizontal axis represents the year, while the vertical axis lists the ten topics.

<https://doi.org/10.1371/journal.pone.0343085.g007>

Fig 7 illustrates that sentiment trends diverge markedly across different topics. Specifically, “Diplomacy & Security”, “Law & Human Rights”, “Policy & Governance”, “Environment & Health”, and “Historical and National Identity” exhibit a clear negative shift overall. In contrast, “Culture & Society”, “Education”, and “Entertainment & Sports” maintain a predominantly positive shift, while “Economy & Trade” and “Technology & Industry” fluctuate between positive and negative sentiments.

Fig 8 traces the thematic trajectories in finer detail, corroborating the broader trends outlined above. In the aftermath of Japan’s 2012 “nationalization” of the Diaoyu/Senkaku Islands, negative and very negative sentiments in “Diplomacy & Security” spiked to 54.92%. Similarly, the proportion of negative sentiments within “Historical and National Identity” remained around 60% between 2014 and 2015, peaking at 63.51% in October 2014, highlighting the immediate mobilizing effect of territorial disputes on national identity and collective belonging [56]. In addition, “Environment & Health” experienced a steep surge in negative sentiment during the COVID-19 outbreak in 2020, while “Law & Human Rights” reached a negative sentiment peak during the 2019 Hong Kong anti-extradition bill movement.

At the same time, positive sentiment clustered most prominently within socially interactive topics. In “Entertainment & Sports”, positive and very positive sentiments consistently accounted for 20–40%, while “Culture & Society” maintained a higher range of 30–50%, reflecting a sustained polarization toward positive affect. Notably, during the COVID-19 pandemic in 2020, the share of positive sentiment in “Entertainment & Sports” stayed around 40%, and in October of the same year, following the success of the Chinese mobile game Genshin Impact at the top of Japan’s App Store rankings, positive sentiment rose to 55.41%.

Sentiments within the “Economy & Trade” and “Technology & Industry” fluctuated noticeably in response to policies and events. In “Economy & Trade”, negative sentiment climbed to 47.35% in 2013 amid regional economic tensions, and surged again to 47.86% during the 2018 US–China trade war. Following the signing of the Regional Comprehensive Economic Partnership (RCEP) in 2020, positive sentiment (positive and very positive) rebounded to 39.37%, underscoring the regulatory influence of policy signals on public attitudes. The “Technology & Industry”, though generally rated positively, displayed even sharper fluctuations: In 2018, the US imposed sanctions on Huawei, and Japan promptly removed the

Thematic sentiment evolution

Very Negative Negative Neutral Positive Very Positive

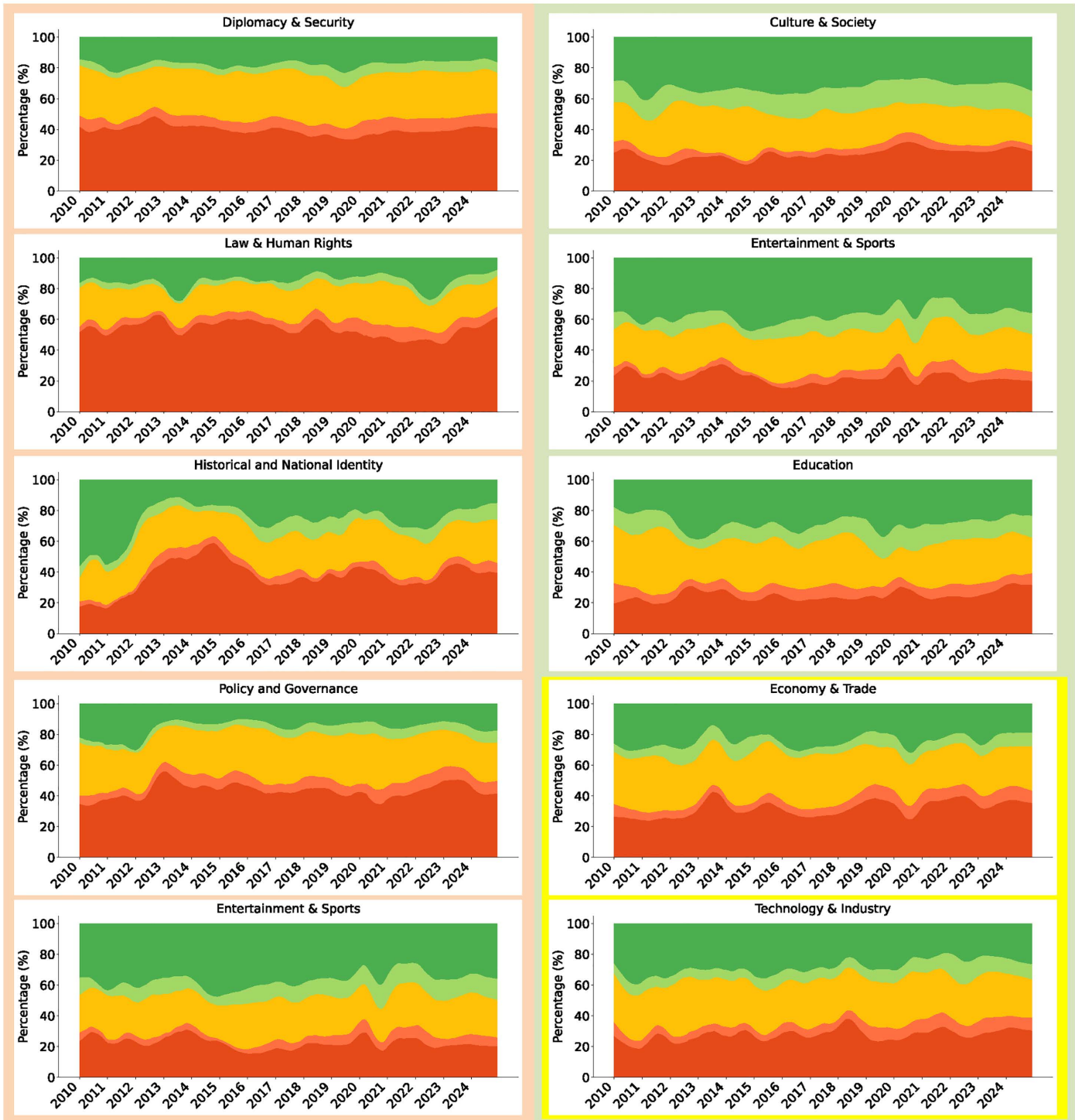


Fig 8. Stacked charts of topic-sentiment evolution (2010-2024). This figure consists of ten subplots, each representing a distinct topic. The ten subplots are divided into two groups: the five on the left represent topics in the domain of high politics, while the five on the right correspond to topics in the domain of low politics. Among the latter, the first three pertain to culture and education, and the final two relate to economics and technology. In each

subplot, the horizontal axis denotes time, while the vertical axis displays the stacked area segmented into five colors. Each color corresponds to the proportion of tweets expressing one of five sentiment categories—very negative, negative, neutral, positive, and very positive—relative to the total number of tweets within that topic.

<https://doi.org/10.1371/journal.pone.0343085.g008>

company from government procurement lists, driving negative sentiment within this topic to 43.82%. By 2019, however, sentiment had swiftly shifted back to positive.

Discussion

This study seeks to illuminate the multidimensional complexity and dynamic evolution of Japanese public perceptions and attitudes toward China. Unlike prior research largely dependent on conventional opinion polls, it offers a fine-grained “topic–sentiment” mapping framework. Our data analysis reveals that:

First, the topic modeling results indicate that public attention toward China has generally increased, with a core focus on “Diplomacy & Security”, followed by “Environment & Health”, “Economy & Trade”, and “Culture & Society”. This largely supports Gong & Nagayoshi’s observation that security concerns form the cornerstone of Japanese perceptions of China [7], while also uncovering a more nuanced picture: within “Diplomacy & Security”, the public’s focus extends beyond the Taiwan Strait crisis, military drills, and U.S.–China confrontations to incorporate the “Technology & Industry” topic through discourses of “technological hegemony” and “dual-use technologies.” Moreover, after the pandemic, public attention toward cultural, economic, and technological topics increased significantly, highlighting diversification in public perceptions of China.

Second, sentiment analysis reveals that Japanese attitudes toward China are overall skewed negative, yet neutral sentiment constitutes about one-quarter of responses. Positive and negative sentiments are roughly balanced, with strongly positive and strongly negative sentiments relatively pronounced. Compared to earlier studies [8], the higher prevalence of neutral sentiment here indicates that a significant segment of the public maintains a nuanced, non-polarized stance toward China, rather than being confined to sharply opposing views.

The topic–sentiment linkage analysis reveals that Japanese attitudes exhibit diverse patterns across topics. Highly politicized topics like “Diplomacy & Security” and “Historical and National Identity” are saturated with negative sentiment, acting as flashpoints for digital nationalism. In contrast, cultural and social domains remain largely depoliticized spaces for positive engagement. Economic and technological topics exhibit dynamic ambivalence, framed alternatively as threats or opportunities, while health crises trigger negative sentiments linked to perceptions of governance failure and “risk spillover”. This pattern strongly indicates that opinion formation is a dual process, driven by external events but mediated through topic-specific affective affordances.

Overall, this study moves beyond the binary framing of “positive–negative” attitudes by conceptualizing Japanese perceptions of China as a communicative process unfolding within the digital public sphere [57,58]. The findings show that public opinion is not simply a reflection of geopolitical events but a discursive formation in which topics such as diplomacy, culture, health, and technology serve as sites of affective negotiation [59]. Social media thus constitutes a hybrid arena where digital nationalism and everyday discourse intersect, shaping, contesting, and reconfiguring China’s image over time. This perspective highlights digital media not merely as a channel of information exchange but as a constitutive space for public diplomacy and transnational identity-making in East Asia.

By combining fine-grained sentiment analysis with BERTopic, this study expands the theoretical boundaries of research on Japan’s China-related public opinion by introducing a dynamic socio-cognitive perspective. We demonstrate that Japanese public attitudes toward China are not merely static reflections of individual experience but evolve through the interplay of major events, user engagement, and discourse framing. This analytical framework elucidates the temporal dynamics of topic prominence and the associated emotional valences, thereby offering methodological and theoretical

guidance for examining China-related discourse on social media. On one hand, the framework enables researchers to extract events from large-scale social media data in a hierarchical manner and integrate the fragmented information into a coherent narrative. On the other hand, it allows researchers to identify critical points of emotional shift, examine connections between different topics, and anticipate emerging public opinion risks, thereby informing policy assessment and optimization.

Based on the above findings, this study offers multiple insights for the management of public opinion and public diplomacy in Sino–Japanese relations. First, given the thematic variation in public sentiment, policymakers should precisely identify and address issue-specific concerns, implementing tailored communication and engagement strategies. Second, security concerns constitute a central consideration in China’s proactive outreach to Japan. Integrating narratives of “cooperative security” and “mutual trust” into non-security domains (such as environmental governance or renewable energy cooperation) can function as a buffer to mediate between conflict and collaboration. Third, the complexity of Japanese perceptions toward China calls for coordinated, multi-platform and multi-actor initiatives to amplify positive narratives. For instance, Chinese residents in Japan, international students, online influencers, and Japanese creators could collaboratively produce short videos highlighting social interactions, thereby reinforcing the dissemination of positive sentiment. Fourth, in sudden public health or environmental crises, China should strengthen risk communication within public diplomacy, ensuring information is released in a timely, transparent, and credible manner to mitigate the sudden surge of negative sentiment.

Limitations

Although our study addresses the two research questions, several limitations remain. First, public opinions expressed on social media may be subject to sampling bias. According to a 2024 survey by Japan’s Ministry of Internal Affairs and Communications, Twitter users aged 20–29 constitute the largest share at 81.6%, followed by adolescents aged 10–19 (65.7%) and adults aged 30–39 (61.0%), whereas participation among the 40–60 age group remains comparatively low [60]. Consequently, the social media data used in this study may primarily reflect the perspectives of younger users, urban residents, or individuals with particular interests, whereas older adults, low-income individuals, and users in remote regions are underrepresented, potentially introducing sampling bias. This sampling bias may limit generalizability. Specifically, it may amplify the views and emotional expressions of digitally active users, thereby constraining a comprehensive understanding of China-related attitudes across different segments of Japanese society. Second, the collection of tweet data relied on pre-defined keyword searches. Although this approach is highly targeted and operationally feasible, it may omit content closely related to the research themes that does not explicitly contain the selected keywords. Future studies could combine multi-level keywords and semantic expansion dictionaries to further enhance data coverage. Third, although the sentiment analysis model and LLM-assisted topic labeling methods used in this study are well-established in current text-mining research, their pre-trained nature inevitably entails a “black-box” characteristic. This includes potential biases in the training corpus and the non-transparent reasoning mechanisms, which may introduce errors in sentiment recognition and topic labeling and pose fairness and ethical risks [61]. Nevertheless, these errors primarily affect model output precision without altering overall topic trends or study conclusions, which remain robust.

Conclusion

Drawing on tweets related to China on Japan’s Twitter platform from 2010 to 2024, this study provides a long-term, multi-level, and fine-grained account of Japanese perceptions and attitudes toward China. The research contributes to the field in three ways. First, by combining large-scale social media data with detailed “topic–sentiment” analysis, it transcends the conventional quantitative approach reliant on opinion surveys, uncovering differentiated patterns and evolution trajectories of Japanese attitudes across distinct topics, thereby providing a more explanatory framework for studying Japanese public opinion. Second, with its large sample size and extended temporal coverage, the study offers a systematic, dynamic

perspective spanning macro- to micro-level analyses. Third, it is the first to integrate BERTopic topic modeling with large-scale sentiment analysis, examining the coupled evolution of topics and sentiments in Japan-related social media discourse on China, and providing methodological tools and analytical frameworks that can guide future research on public opinion.

Future research could be expanded in several directions. First, Future research may incorporate demographic variables such as gender and age within permissible data constraints to examine differences in expression, emotional patterns, and interaction behaviors across groups, thereby providing a more comprehensive understanding of public attitude formation mechanisms. Second, cross-national comparative studies could examine how China-related discourse is constructed on social media platforms in Japan versus countries such as the UK and the US, shedding light on the cultural, social, and institutional mechanisms underlying these differences. Third, future research could build on the analytical framework of this study by incorporating domain expert knowledge to construct an ontology tailored to Japan-related Chinese discourse on social media, and integrate it with LLM-driven text classification and topic identification pipelines to enhance the robustness and interpretability of the analysis across contexts and events. Technically, the hyper-knowledge graph system introduced by Chen et al. offers a useful model for structuring complex text analysis outputs into a queryable and inferable knowledge system [62]. Following this approach, future work could map tweets, topics, sentiment orientations, and their temporal evolution into nodes and edges within a graph database, offering a potential framework for supporting policy analysis and public opinion research.

Acknowledgments

The authors have no acknowledgments to declare.

Author contributions

Conceptualization: Xi Chen.

Data curation: Shanshan Zhang.

Formal analysis: Shanshan Zhang.

Funding acquisition: Shanshan Zhang.

Investigation: Shanshan Zhang.

Methodology: Shanshan Zhang.

Project administration: Xi Chen.

Resources: Shanshan Zhang.

Supervision: Xi Chen.

Validation: Shanshan Zhang.

Visualization: Shanshan Zhang.

Writing – original draft: Shanshan Zhang.

Writing – review & editing: Shanshan Zhang.

References

1. Tanabe S. Nihon ni okeru Nashonarizumu no Jitenkanhikaku to Kiteiyuin. 2016.
2. Yoshimatsu H. Japan's strategic response to China's geo-economic presence: quality infrastructure as a diplomatic tool. *The Pacific Review*. 2021;36(1):148–76. <https://doi.org/10.1080/09512748.2021.1947356>
3. Key VO. *Public Opinion and American Democracy*. New York: Knopf. 1961.

4. Baum MA, Potter PBK. The relationships between mass media, public opinion, and foreign policy: Toward a theoretical synthesis. *Annu Rev Polit Sci.* 2008;;39–65.
5. Manza J, Cook FL. A Democratic Polity?. *American Politics Research.* 2002;30(6):630–67. <https://doi.org/10.1177/153267302237231>
6. Chu Y, Kang L, Huang M. How East Asians View the Rise of China. *Journal of Contemporary China.* 2014;24(93):398–420. <https://doi.org/10.1080/10670564.2014.953810>
7. Gong S, Nagayoshi K. Japanese Attitudes Toward China and the United States: A Sociological Analysis. *Chinese Sociological Review.* 2019;51(3):251–70. <https://doi.org/10.1080/21620555.2019.1611374>
8. Lee H, Noh EB, Choi SH, Zhao B, Nam EW. Determining Public Opinion of the COVID-19 Pandemic in South Korea and Japan: Social Network Mining on Twitter. *Healthc Inform Res.* 2020;26(4):335–43. <https://doi.org/10.4258/hir.2020.26.4.335> PMID: 33190468
9. Shao J. Territorial dispute at home: strategic narratives contestation over the Diaoyu/Senkaku Islands nationalization policy in Japan. *International Relations of the Asia-Pacific.* 2021;23(1):61–92. <https://doi.org/10.1093/irap/lcab024>
10. Yang H, Van Gorp B. A frame analysis of political-media discourse on the Belt and Road Initiative: evidence from China, Australia, India, Japan, the United Kingdom, and the United States. *Cambridge Review of International Affairs.* 2021;36(5):625–51. <https://doi.org/10.1080/09557571.2021.1968794>
11. Chu Y, Chang Y. Xi's foreign-policy turn and Asian perceptions of a rising China. *Glob Asia.* 2017;12:104–11.
12. Huang M, Chu Y. The Sway of Geopolitics, Economic Interdependence and Cultural Identity: why are some Asians more favorable toward China's rise than others?. *Journal of Contemporary China.* 2014;24(93):421–41. <https://doi.org/10.1080/10670564.2014.953842>
13. Xie Y, Jin Y. Global Attitudes toward China: Trends and Correlates. *Journal of Contemporary China.* 2021;31(133):1–16. <https://doi.org/10.1080/10670564.2021.1926088>
14. Hänska M, Bauchowitz S. Can social media facilitate a European public sphere? Transnational communication and the Europeanization of Twitter during the Eurozone crisis. *Social Media + Society.* 2019;5(3). <https://doi.org/10.1177/2056305119854686>
15. Takikawa H, Nagayoshi K. Political polarization in social media: Analysis of the "Twitter political field" in Japan. In: 2017 IEEE International Conference on Big Data (Big Data), 2017. 3143–50. <https://doi.org/10.1109/bigdata.2017.8258291>
16. Yoshida M, Toriumi F. Information Diffusion Power of Political Party Twitter Accounts During Japan's 2017 Election. *Lecture Notes in Computer Science.* Springer International Publishing. 2018. 334–42. https://doi.org/10.1007/978-3-030-01159-8_32
17. Zhao Z, Zhao J, Sano Y, Levy O, Takayasu H, Takayasu M, et al. Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Sci.* 2020;9(1). <https://doi.org/10.1140/epjds/s13688-020-00224-z>
18. Wei M, Zhang P, Chen C, Chen D, Zuo C, Meng L. Mapping AI ethics narratives: evidence from Twitter discourse between 2015 and 2022. *Humanit Soc Sci Commun.* 2025;12(1). <https://doi.org/10.1057/s41599-025-04469-9>
19. Kingston J. The Emptiness of Japan's Values Diplomacy in Asia. *Asia-pac j.* 2020;18(3). <https://doi.org/10.1017/s1557466020030478>
20. Yoshimatsu H. Examining China and Japan's Narratives as Diplomatic Tools: Competitive Nature and International Influence. *Journal of Contemporary China.* 2024;34(153):367–82. <https://doi.org/10.1080/10670564.2024.2352760>
21. McNeill D, Kato N. Fanning the Flames: China vs. Japan in the Media. *Asia-Pac J.* 2024;22(10). <https://doi.org/10.1017/s1557466024029140>
22. Chen L, Zhang B. Hedging No More: Security Seeking and Japan's Dual-Track China Policy. *Journal of Contemporary China.* 2025;;1–20. <https://doi.org/10.1080/10670564.2025.2459621>
23. Sonoda S. Asian views of China in the age of China's rise: interpreting the results of pew survey and Asian student survey in chronological and comparative perspectives, 2002-2019. *Journal of Contemporary East Asia Studies.* 2021;10(2):262–79. <https://doi.org/10.1080/24761028.2021.1943116>
24. Cai Y, Qi H, Lu W, Zhang K, Shen Q. Whose story is prevailing on Twitter? The international discourse battle over Japan's nuclear wastewater discharge. *Asian Journal of Communication.* 2025;35(4):351–75. <https://doi.org/10.1080/01292986.2025.2522688>
25. Qasim R, Bangyal WH, Alqarni MA, Ali Almazroi A. A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification. *J Healthc Eng.* 2022;2022:3498123. <https://doi.org/10.1155/2022/3498123> PMID: 35013691
26. Blei DM, Jordan MI. Latent Dirichlet Allocation. *J Mach Learn Res.* 2003;;993–1022.
27. Laureate CDP, Buntine W, Linger H. A systematic review of the use of topic models for short text social media analysis. *Artif Intell Rev.* 2023;;1–33. <https://doi.org/10.1007/s10462-023-10471-x> PMID: 37362887
28. Grootendorst M. BerTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv.* 2022. <https://doi.org/10.48550/arXiv.2203.05794>
29. Murayama T, Miyazaki K, Matsubara Y, Sakurai Y. Linguistic landscape of generative AI perception: A global Twitter analysis across 14 languages. *arXiv.* 2025. <https://doi.org/10.48550/arXiv.2405.20037>
30. Veigel N, Kreibich H, de Bruijn JA, Aerts JCJH, Cominola A. Content analysis of multi-annual time series of flood-related Twitter (X) data. *Nat Hazards Earth Syst Sci.* 2025;25(2):879–91. <https://doi.org/10.5194/nhess-25-879-2025>
31. Dehghan S, Pir Mohammadiani R, Mohammadi S. The credibility assessment of Twitter/X users based organization objectives by heterogeneous resources in big data life cycle. *Computers in Human Behavior.* 2025;162:108428. <https://doi.org/10.1016/j.chb.2024.108428>
32. Ng QX, Teo YQJ, Kiew CY, Lim BP-Y, Lim YL, Liew TM. Examining the Prevailing Negative Sentiments Surrounding Measles Vaccination: Unsupervised Deep Learning of Twitter Posts from 2017 to 2022. *Cyberpsychol Behav Soc Netw.* 2023;26(8):621–30. <https://doi.org/10.1089/cyber.2023.0025> PMID: 37358808

33. Son H, Park YE. Agenda-setting effects for covid-19 vaccination: Insights from 10 million textual data from social media and news articles using BERTopic. *International Journal of Information Management*. 2025;83:102907. <https://doi.org/10.1016/j.ijinfomgt.2025.102907>
34. Saif H, He Y, Fernandez M, Alani H. Adapting Sentiment Lexicons Using Contextual Semantics for Sentiment Analysis of Twitter. *Lecture Notes in Computer Science*. Springer International Publishing. 2014. 54–63. https://doi.org/10.1007/978-3-319-11955-7_5
35. Tang D, Wei F, Qin B, Zhou M, Liu T. Building large-scale Twitter-specific sentiment lexicon: A representation learning approach. In: *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers*, 2014. 172–82.
36. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*. 2014;5(4):1093–113. <https://doi.org/10.1016/j.asej.2014.04.011>
37. Abas AR, Elhenawy I, Zidan M, Othman M. BERT-CNN: A Deep Learning Model for Detecting Emotions from Text. *Computers, Materials & Continua*. 2022;71(2):2943–61. <https://doi.org/10.32604/cmc.2022.021671>
38. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the NAACL-HLT*, 2019. 4171–86. <https://aclanthology.org/N19-1423>
39. Camacho-Vázquez LA, Camacho-Vázquez VA, Orantes-Jiménez SD, Sidorov G. Detection of Negative Emotions in Short Texts Using Deep Neural Networks. *Cyberpsychol Behav Soc Netw*. 2025;28(6):399–407. <https://doi.org/10.1089/cyber.2024.0457> PMID: 40331318
40. Wu Z, Ong DC. Context-Guided BERT for Targeted Aspect-Based Sentiment Analysis. *AAAI*. 2021;35(16):14094–102. <https://doi.org/10.1609/aaai.v35i16.17659>
41. Bedi J, Toshniwal D. CitEnergy : A BERT based model to analyse Citizens' Energy-Tweets. *Sustainable Cities and Society*. 2022;80:103706. <https://doi.org/10.1016/j.scs.2022.103706>
42. Choi S, Liu K, Hanaoka S. Perceived Responses of International Tourists to Transportation and Tourism Services During Typhoons Faxai and Hagibis in Japan. *Sustainability*. 2024;16(20):9114. <https://doi.org/10.3390/su16209114>
43. Leow EKW, Nguyen BP, Chua MCH. Robo-advisor using genetic algorithm and BERT sentiments from tweets for hybrid portfolio optimisation. *Expert Systems with Applications*. 2021;179:115060. <https://doi.org/10.1016/j.eswa.2021.115060>
44. Wan B, Wu P, Yeo CK, Li G. Emotion-cognitive reasoning integrated BERT for sentiment analysis of online public opinions on emergencies. *Information Processing & Management*. 2024;61(2):103609. <https://doi.org/10.1016/j.ipm.2023.103609>
45. Hudders L, De Jans S. Gender effects in influencer marketing: an experimental study on the efficacy of endorsements by same- vs. other-gender social media influencers on Instagram. *International Journal of Advertising*. 2021;41(1):128–49. <https://doi.org/10.1080/02650487.2021.1997455>
46. Thakur N, Cui S, Khanna K, Knieling V, Duggal YN, Shao M. Investigation of the Gender-Specific Discourse about Online Learning during COVID-19 on Twitter Using Sentiment Analysis, Subjectivity Analysis, and Toxicity Analysis. *Computers*. 2023;12(11):221. <https://doi.org/10.3390/computers12110221>
47. Chen K, Molder AL, Duan Z, Boulianne S, Eckart C, Mallari P, et al. How Climate Movement Actors and News Media Frame Climate Change and Strike: Evidence from Analyzing Twitter and News Media Discourse from 2018 to 2021. *The International Journal of Press/Politics*. 2022;28(2):384–413. <https://doi.org/10.1177/19401612221106405>
48. Wang Y, Xie Y. How country names legitimize economic nationalism: a critical discourse analysis. *Discov glob soc*. 2025;3(1). <https://doi.org/10.1007/s44282-025-00223-3>
49. Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019. 3980–90. <https://doi.org/10.18653/v1/d19-1410>
50. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2020. <https://doi.org/10.48550/arXiv.1802.03426>
51. Campello RJGB, Moulavi D, Sander J. Density-Based Clustering Based on Hierarchical Density Estimates. In: Pei J, Tseng VS, Cao L, Motoda H, Xu G. *Advances in Knowledge Discovery and Data Mining*. Berlin, Heidelberg: Springer Berlin Heidelberg. 2013. 160–72. https://doi.org/10.1007/978-3-642-37456-2_14
52. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, et al. Unsupervised Cross-lingual Representation Learning at Scale. 2020. <https://doi.org/10.48550/arXiv.1911.02116>
53. Liu J, Li K, Zhu A, Hong B, Zhao P, Dai S, et al. Application of Deep Learning-Based Natural Language Processing in Multilingual Sentiment Analysis. *MJBAS*. 2024;08(02):243–60. <https://doi.org/10.46382/mjbas.2024.8219>
54. Anderson C. The long tail. In: Mandiberg M. *The social media reader*. New York University Press. 2012. 137–54.
55. Arsov N, Dukovski M, Evkoski B, Cvetkovski S. A Measure of Similarity in Textual Data Using Spearman's Rank Correlation Coefficient. *arXiv*. 2019. <https://doi.org/10.48550/arXiv.1911.11750>
56. Kim NK. Territorial disputes and individual willingness to fight. *Journal of Peace Research*. 2019;57(3):406–21. <https://doi.org/10.1177/0022343319880952>
57. Castells M. *The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance*. *The ANNALS of the American Academy of Political and Social Science*. 2008;616(1):78–93. <https://doi.org/10.1177/0002716207311877>

58. Habermas J. Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research. *Commun Theory*. 2006;16(4):411–26. <https://doi.org/10.1111/j.1468-2885.2006.00280.x>
59. Papacharissi Z. *Affective publics: sentiment, technology and politics*. New York: Oxford University Press. 2015.
60. Ministry of Internal Affairs and Communications J. Survey report on media usage and information behavior. Ministry of Internal Affairs and Communications, Japan. 2024.
61. Chen C, Luo P, Zhao H, Wei M, Zhang P, Liu Z, et al. A unified ontological and explainable framework for decoding AI risks from news data. *Sci Rep*. 2025;15(1):25621. <https://doi.org/10.1038/s41598-025-10675-x> PMID: [40664767](https://pubmed.ncbi.nlm.nih.gov/40664767/)
62. Chen C, Feng Y, Wei M, Liu Z, Luo P, Wang S, et al. A hyper-knowledge graph system for research on AI ethics cases. *Heliyon*. 2024;10(7):e29048. <https://doi.org/10.1016/j.heliyon.2024.e29048> PMID: [38601681](https://pubmed.ncbi.nlm.nih.gov/38601681/)