Kulichkina et al. *EPJ Data Science* (2026) 15:39
https://doi.org/10.1140/epjds/s13688-026-00637-2
EPJ Data Science

RESEARCH

Open Access

Connective action and digital repression during China's COVID-19 protests: a computational analysis of multilingual coordinated activity on Twitter

Aytalina Kulichkina[1*], Paul Balluff[1,2], Nicola Righetti[3] and Annie Waldherr[1]

Handling Editor: Emilio Ferrara

*Correspondence:
aytalina.kulichkina@univie.ac.at
[1]Department of Communication, University of Vienna, Kolingasse 14-16, 1090, Vienna, Austria
Full list of author information is available at the end of the article

Abstract

In authoritarian contexts, social media serve as critical platforms for coordinating both protest and repression. This study centers on the unprecedented COVID-19 protests in the People's Republic of China, which were extensively tweeted and suppressed through contesting narratives. We explore prominent themes, temporal dynamics, and linguistic patterns of coordinated communication during these events. Using a coordination detection algorithm, we identified 13,557 Twitter accounts involved in 739,819 instances of coordinated sharing during the protests. We then applied topic modeling to categorize the coordinated tweets into topics supporting either the protests or repression. Drawing on the theory of authoritarian publics, we classified protest-supporting topics into three categories: *leadership-critical*, *policy-critical*, and *descriptive*. Similarly, building on the digital repression typology, we categorized repression-supporting topics into *government propaganda*, *distracting information*, and *demoralizing content*. Within protest-supporting content, policy-critical tweets were the most widely shared across three analyzed languages. Leadership-critical tweets were more prominent in traditional Chinese, while descriptive tweets were more common in simplified Chinese. Repression-supporting content was most prevalent in English, followed by simplified Chinese, with demoralizing and distracting information dominating discourse. Government propaganda was the least frequent and appeared primarily in simplified Chinese. Community detection revealed that 85.4% of coordinated tweets were amplified by ten major communities, each organized around a single language and goal—either supporting protests or promoting repression. By combining multiple computational approaches, this study offers a comprehensive framework for content-centered analysis of online protest-repression dynamics and contributes to our understanding of connective action and digital repression in authoritarian contexts.

Keywords: Connective action; Digital repression; Coordinated behavior; Twitter; China; COVID-19; Protest; Topic modeling; Community detection

1 Introduction

In authoritarian contexts, political protests rarely succeed in achieving immediate change. However, they play a crucial role in challenging oppressive systems and preventing further shifts toward authoritarianism. For over a decade, social media have empowered the public in non-democratic societies to mobilize and protest against authoritarian rule [54]. At the same time, they have enabled regimes to suppress political expression and increase the costs for digital activism [12]. On social media, protest and repression unfold in a dynamic interplay of action and response, much like in offline contexts [26]. Protesters aim to voice their claims, raise awareness, and mobilize supporters. By contrast, repressors seek to contain dissent through disinformation, propaganda, and distracting information, among other tactics [11]. The potential for online activism to spill into the streets continually pressures both sides to adapt and innovate their strategies. Literature shows that *coordinated behavior* on social media—organized publishing or sharing of content by multiple actors within a narrow time frame to increase influence [15]—is a viable strategy that can serve both protesters and those who wish to suppress them [34].

In late November 2022, citizens of the People's Republic of China (PRC), frustrated with the government's handling of the pandemic, took to the streets, staging the largest public demonstrations since the 1989 Tiananmen Square protests [56]. Outraged by the deadly apartment fire in Urumqi that killed ten people, protesters demanded an end to the stringent "zero-COVID" policy, Xi Jinping's resignation, and greater political freedoms. However, the government's strict control over domestic communication channels led to the suppression of online discussions related to the protests. In response, citizens used virtual private networks (VPNs) to circumvent blocked social media sites, particularly Twitter, to voice their dissent, document the protests, and attract international attention [38]. At the same time, their opponents adapted by flooding Twitter with counter-narratives and spam to crowd out protest content and dominate users' feeds [36].

The online protest and repression transcended borders, mobilizing allies and attracting onlookers worldwide [10]. Twitter discussions occurred in multiple languages, including three key written forms relevant to our study: simplified Chinese, traditional Chinese, and English. Simplified Chinese is primarily used in the PRC, while traditional Chinese is used in the historically autonomous regions of Hong Kong and Macau, as well as in democratic Taiwan. The protests were also widely tweeted in English to attract international support.

In this study, we use computational methods to examine the content published during the protests by highly coordinated Twitter accounts in simplified Chinese, traditional Chinese, and English. We further identify the discourse that either amplifies or suppresses protest visibility. Our findings show that leadership-critical, policy-critical, and descriptive tweets were most prominent in simplified and traditional Chinese, while demoralizing and distracting information dominated English tweets. Government propaganda appeared primarily in simplified Chinese, followed by English and traditional Chinese. Finally, ten distinct user communities were responsible for the coordinated sharing of most tweets during the protests, each focused on one language and one goal—either supporting protests or promoting repression. We conclude that while protest amplification was most efficiently coordinated in simplified Chinese, digital repression predominantly targeted global English-speaking audiences.

This research contributes to our understanding of online protest and repression dynamics in authoritarian contexts and beyond, emphasizing the role of coordinated communication and discursive practices in mobilizing or silencing the public. We contribute to the literature on digital repression (for an overview, see [11]) by identifying an additional form of information control and manipulation aimed at demoralizing and delegitimizing protests. We also contribute to the study of online protest-repression dynamics by introducing a methodological approach that identifies protest- and repression-related content through the combination of coordinated behavior detection and topic modeling. Our findings indicate that analyzing the discursive practices of coordinated social media posts reveals novel patterns and tactics used by activists and repressors to shape public opinion during social movements. The key takeaway is that digital repression extends beyond authoritarian publics, yet there remains insufficient coordination to amplify protests and counter repression online on a global scale.

2 Authoritarian publics on Chinese and overseas social media

In this study, we adopt Toepfl's [51] concept of *authoritarian publics*, defined as "specific constellations of participants, environments, and discursive practices" (p. 109). According to this theory, the PRC, as an authoritarian public-at-large, mainly consists of *uncritical* and *policy-critical publics*, in which criticism can be directed at "lower-level officials, policies, and institutions of the authoritarian regime" [51, p. 111]. However, it does not accommodate a significant number of *leadership-critical publics*, in which "criticism lashes out even at the country's highest-ranking political leadership" [51, p. 114]. Under Xi Jinping's leadership, the PRC's internet governance underwent centralization to harness digital technologies for propaganda and public opinion control while restricting public deliberation channels to diminish the autonomy of the online sphere [9]. For example, the government uses the infamous *Great Firewall* to block major international social media platforms such as Facebook, Twitter, and Instagram while promoting domestic alternatives like WeChat, Weibo, and Douyin. The digital architecture of such domestic platforms is highly influenced by the state. For example, messages published by the official party and state media accounts are algorithmically privileged, while the ability of regular users to challenge official narratives is diminished by the algorithmic design [5]. Therefore, China's social media allow only a constrained and censored environment for meaningful discussion, limiting the public's ability to contest the official discourse or express criticism [13].

Despite these challenges, social media users in the PRC navigate limited spaces for expression as long as their criticism is not deemed harmful to the regime [17, 23]. Notably, the Chinese Communist Party (CCP) focuses more on monitoring content creators rather than content itself, permitting limited open discussion among ordinary users while strategically repressing or co-opting influential accounts that could threaten its leadership [14]. Unlike traditional media in the PRC, social media are more likely to openly report on minor and nonviolent protests, as the platforms tend to censor them less than commonly assumed [16]. The government can also be responsive to public demands, especially when they are prominently voiced online [20]. Thus, despite stringent state control, Chinese social media accommodate both uncritical and policy-critical publics, offering space for various perspectives. This setup ultimately serves the state's interests by enabling authorities to gather information about social issues, monitor lower-level bureaucracy, and project an image of responsive governance [17, 51].

While allowing limited virtual space for critical voices can help autocrats maintain a façade of openness, it also introduces vulnerabilities for their regime. Leadership-critical discourse may be largely absent from Chinese social media, but this does not mean such sentiments do not exist; they may surface when perceived risks are lower. For example, the Chinese diaspora abroad, shielded from direct authoritarian control, has played a key role in transmitting and localizing social movements in the PRC via social media [57]. Additionally, even policy criticism, which is generally tolerated, can intensify during crises and erode regime legitimacy, escalating into broader dissent [51]. Hence, despite strict information control and the hegemony of pro-regime discourse online, it remains plausible, under favorable conditions, for typically silent leadership-critical publics to find avenues to voice their grievances and for protests to emerge in an authoritarian country like the PRC.

In authoritarian settings, crises may open temporary opportunities for leadership critics to benefit from increased information flow, public dissatisfaction, and a relaxation of government control due to the focus on crisis management. The COVID-19 crisis allowed Chinese citizens to engage with abundant information and accumulate discontent with the harsh zero-COVID policy. This motivated them to use VPNs to seek crisis-related information across the *Great Firewall*, which exposed them to unrelated and regime-damaging content [8]. Given the domestic social media limitations, critics and passive onlookers turned to banned platforms like Twitter to express their protest and observe the ongoing discourse. As a result, Twitter saw a tremendous increase in downloads right before and during the COVID-19 protests, soaring from 150th to ninth place among free iOS apps in the country [31]. Before this surge of mainland users, Twitter had already hosted polarized Chinese diasporic communities, pro-CCP individuals, and state entities engaged in various forms of discourse [58]. Thus, Twitter became a space where all authoritarian publics—uncritical, policy-critical, and leadership-critical—coexisted relatively freely during the COVID-19 protests.

3 Digital repression in China and beyond

In this paper, we examine the interplay between critics and their opponents within authoritarian publics on Twitter. Toepfl's [51] original essay focuses on physical control as a tool autocrats use to manage authoritarian publics. However, in the context of social media, it is essential to consider not only physical but also the informational control wielded by autocrats and their supporters. Therefore, we adopt the concept of *digital repression*, as defined by Earl et al. [11], as "actions directed at a target to raise the target's costs for digital social movement activity and/or the use of digital or social media to raise the costs for social movement activity, wherever that contestation takes place" (p. 1). These actions expand conventional repression repertoires by including non-physical forms of informational control, such as content filtering, down-ranking, shadow-banning, disinformation, and misrepresentation, as well as posting distracting information or flooding online spaces and hashtags with irrelevant material [11]. The latter two are especially pertinent to digital repression on platforms beyond direct autocratic control, such as Twitter, where autocrats employ alternative tactics to interfere with, divert attention from, or manipulate critical discursive practices.

In the context of China, digital repression technologies are among the most sophisticated globally, driven primarily by the aim of preventing violent protests and maintaining social and political stability within the country [45]. Domestic social media repression combines targeted censorship with opinion management: content related to collective action is preferentially removed, while online discourse is flooded with pro-government or distracting narratives rather than engaging critics [24, 53, 55]. These tactics operate through what Roberts [44] terms *fear*, *friction*, and *flooding*—deterrence via surveillance and punishment, procedural obstacles that raise the cost of collective action, and high-volume propaganda and astroturfing that crowd out alternative narratives. By contrast, on overseas platforms such as Twitter—where Chinese authorities have limited formal leverage to coerce companies into censoring their platforms—the CCP employs a mixed strategy that combines public diplomacy [21] with state-linked influence operations aimed at wrangling the platforms' algorithms via paid or incentivized commentators (e.g., [19]).

During the COVID-19 protests, digital repression posed a major challenge for critical publics engaging in political activism online. However, the international scope of Twitter enabled allies abroad to join the movement and amplify attention to protests unfolding behind the *Great Firewall*. As a venue for digital public spheres, where local and nonlocal forms of communication coexist [52], Twitter also facilitated the Chinese diaspora's role as brokers in spreading information about the protests [58]. Hence, it is necessary to distinguish between critical publics operating within or near the sphere of authoritarian influence and those more distant from the potential consequences of repression. One indicator of this distance is the language users choose to tweet in, as it can signify ties to a certain geographic location, cultural identity, and the intended audience [25]. In this study, we focus on discursive practices strategically employed by and for users of simplified Chinese, traditional Chinese, and English in digital protest and repression.

Since the protests took place in mainland China, where simplified Chinese is the primary written language, analyzing discourse in this script is crucial for understanding communication tailored to local audiences. Traditional Chinese serves as the written form of communication in more liberal Taiwan, Hong Kong, and Macau. In these areas, citizens share a vested interest in protests against the regime within the PRC due to concerns over regional autonomy, human rights, and political freedoms. By examining discourse in traditional Chinese, we capture how users in these regions articulate their stances and engage with events in the mainland. Finally, English discourse on Twitter holds significance due to its accessibility to the international community. Content posted in English has the potential to raise international awareness and attract the attention of diverse audiences, including influential politicians, journalists, and international organizations. Therefore, examining discourse in English and traditional Chinese provides a broader perspective on the protests and their impact beyond the PRC's borders.

4 Coordinated behavior in connective action and digital repression

Language is not the only indicator of the underlying intent behind messages publicly shared during large-scale protests. In this paper, we adopt the theory of *connective action*, which emphasizes the role of digitally mediated communication in enabling decentralized protest mobilization [3]. According to this theory, the circulation of personalized action frames functions as a coordinating signal, enabling large-scale mobilization via digital networks [3]. In other words, effective social media movements involve coordination and personal action framing. The connective action lens has been increasingly extended to authoritarian contexts, for example, to examine the organizational dynamics of Russian opposition activists [50], to trace the development of the #MeToo movement in the PRC [57], and to evaluate coordination patterns in the pro-Navalny movement in Russia [26].

The latter case offers a helpful reference point due to its focus on coordinated behavior in connective action.

While much of the literature on coordination in social media has focused on malicious tactics involving "troll armies" (e.g., [24]), "political astroturfing" (e.g., [22]), "coordinated inauthentic behavior" (e.g., [15]), or "coordinated social media manipulation" [49], it is crucial to recognize that coordinated behavior is not inherently malicious or inauthentic. Recent research indicates that authentic protesters and activists also engage in coordinated behavior to amplify their voices online [6, 26, 34]. Accordingly, we define *coordination* as the organized sharing of content within narrow time windows by multiple social media accounts, repeated across multiple occasions, primarily for amplification. This definition responds to recent calls to disentangle coordination from the analytical lens of inauthentic behavior [33, 41, 49]. Thus, while coordination can be used as a malicious tactic for digital repression, it can also serve as an enabling tool for connective action that increases the visibility of protests on social media. In other words, coordination indicates that social media posts are shared to amplify content for specific goals, which can be further understood by analyzing the underlying discursive components.

Bennett and Segerberg [3] emphasize the importance of inclusiveness in discourses coordinated by crowds, noting that "people must show each other how they can appropriate, shape, and share themes" (p. 746). Identifying such themes shared by coordinated Twitter accounts can shed light on the key issues strategically amplified during protests, whether to support connective action or to promote digital repression. However, the circulation of themes is not independent of language use: even in the presence of automated translation, linguistic choices partition social media into partially overlapping publics, shaping how themes circulate, gain visibility, and align with specific audiences, and thereby indicate the intended scope and transnational reach of coordinated discourse [43]. To examine the discourses and dynamics of online protest and repression facilitated by coordinated tweeting during the COVID-19 protests in China, we therefore combine a content-oriented analysis of themes with a structural and linguistic perspective on coordinated networks.

First, we identify the dominant themes amplified by coordinated Twitter accounts through a computationally assisted, iterative engagement with the data, grounding the themes in the material while drawing on existing literature on connective action and digital repression. Second, we trace how these themes evolve over time to illuminate shifts in strategic emphasis across phases of contention. Third, we examine language use within the coordinated network and its communities, as language not only reflects thematic framing but also structures participation, audience targeting, and cross-border diffusion. By integrating these approaches, we address a gap in understanding the mechanisms through which coordination enables connective action or facilitates digital repression.

Accordingly, we pose the following research questions:

RQ1: What are the main themes of tweets published by coordinated Twitter accounts during the COVID-19 protests in China?

RQ2: How do the identified themes support digital protest and repression on Twitter during the COVID-19 protests in China over time?

RQ3: How does language use within the coordinated network and its communities relate to the thematic focus of digital protest and repression during the COVID-19 protests in China?

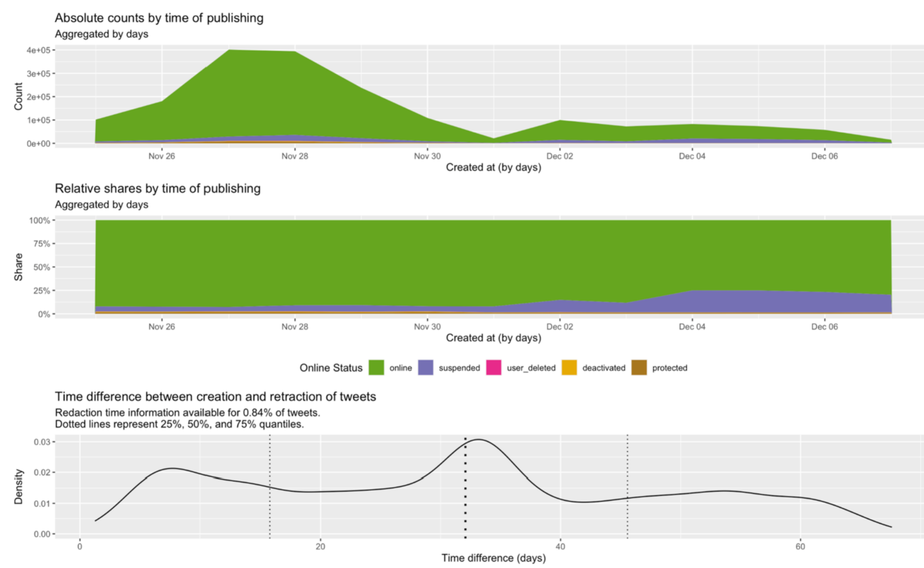

Figure 1 Tweet availability and unavailability status as of February 2, 2023

5 Methods

5.1 Data collection

We used Twitter API v2 for academic research (application programming interface, API) and the *academictwitteR* package [2] to collect over 12 million tweets related to the COVID-19 protests in China.[1] Our keyword-based data collection strategy, informed by real-time observations, included 61 protest-related terms in simplified Chinese, traditional Chinese, and English (see Appendix A). Data collection began on November 28 at 22:20 (GMT) and covered the period from November 25 to December 19, ultimately capturing a total of 12,409,946 tweets by the end of the collection on December 19. Of these, 7,276,728 tweets were in English, while 3,082,210 were in simplified and traditional Chinese.[2]

To assess the scale of the data unavailability problem, we evaluated the compliance status [46] of our dataset on February 2, 2023. The analysis revealed that 89.28% of all tweets were still available through the API. The unavailability of the remaining tweets was due to account suspension (8.41%), protected accounts (1.48%), deactivated accounts (0.69%), and deleted accounts (0.11%). Deleted tweets were, on average, made inaccessible 744 hours after their creation (MIN = 17 hours, MAX = 1621 hours). Figure 1 illustrates the availability of the collected tweets on February 2, 2023. Overall, the analysis indicates that the issue of deleted tweets had a minimal impact during the data collection period.

5.2 Data analysis

To address RQ1, we first identified coordinated accounts using the *CooRTweet* package [42]. Since the protests lasted from November 25 to December 7, we restricted coordination detection to this period.[3] The subset comprised 8,153,122 tweets, including 4,403,616

[1] The script for data collection is available in the supplemental materials: https://osf.io/egxrq.

[2] Twitter classification does not differentiate between traditional and simplified Chinese. We performed our own language classification following the coordination detection step (see Appendix D).

[3] For the robustness check covering the remaining time period, see Appendix B.

in English and 2,457,483 in simplified and traditional Chinese. We detected coordinated tweets that were co-retweeted at least twice within 120 seconds by accounts engaged in repetitive *co-retweeting*—defined as "retweeting within a short time period the same message from an account which may or may not be part of the campaign" [22, p. 265]. Accordingly, we operationalized coordination as repetitive *co-retweeting*. This operationalization is particularly relevant to Twitter, where retweeting serves as a primary mode of amplification and co-retweeting offers a minimal-effort, coordinated way to boost content visibility [30, 47]. The choice of temporal and repetition thresholds reflects both prior methodological work and practical constraints. Very short coordination windows (i.e., below 60 seconds) tend to disproportionately capture highly automated, bot-like activity and risk excluding forms of coordination observed among authentic accounts, such as organizational or movement-based amplification that unfolds over slightly longer temporal intervals. To balance precision with inclusivity, we adopted a 120-second window with a minimum repetition of 2 and a minimum edge weight of 3 (the 99.5th percentile), following Kulichkina et al. [26], who systematically evaluate alternative coordination thresholds and show that short time windows are effective in capturing highly synchronized, campaign-like yet human behavior.[4]

We adopted a Computational Grounded Theory (CGT; [37]) approach to guide the analysis, treating topic modeling and interpretation as an iterative mixed-methods process that bridges computational pattern detection with qualitative sensemaking. CGT integrates inductive exploration, theoretically informed interpretation, and computational validation, and has been increasingly applied and refined in the social sciences and related fields [1, 7]. In line with this approach, we engaged with the data through three iterative steps: (1) discovery, involving exploratory pattern detection; (2) interpretation and refinement, informed by prior literature while remaining open to emergent themes; and (3) classification and validation. Accordingly, we approached the output of computational models with informed openness, drawing on existing research on digital protest and repression without imposing a fully predefined coding scheme.

Within this CGT framework, we applied *BERTopic* [18] to computationally extract topics from the unique coordinated tweets. Unlike other topic modeling approaches, *BERTopic* can process multilingual text directly by loading transformer models that have been pre-trained on multilingual data.[5] The advantage of this approach is that the topics can be assigned across languages without applying machine translation or other techniques to harmonize the texts beforehand. The model clustered 30,886 unique coordinated tweets into 161 topics. Two authors, proficient in the target languages and familiar with the context of the COVID-19 protests in China, independently explored and tagged the topics by analyzing three representative tweets generated by *BERTopic* (for details, see [18]) along with the 20 tweets with the highest probability for each topic. As part of this exploratory step, we also considered additional materials extending beyond the textual content of the tweet. This included checking the account profile or inspecting attached images if they were still available.

---

[4] We conducted several exploratory tests with alternative coordination windows and repetition thresholds during preliminary analyses.

[5] The document embeddings were generated with the "paraphrase-multilingual-mpnet-base-v2" model from the Sentence-BERT framework [40].

To address RQ2, we moved from inductive topic discovery to theoretically informed categorization by developing six higher-level topic categories related to either protest or repression, based on the labeling results (Table 1). In line with the CGT approach, two authors fluent in Chinese engaged in iterative interpretation and refinement, relating the identified patterns back to existing theories of digital protest and repression where applicable. Disagreements between the two authors were resolved through discussion. Protest-related categories included leadership-critical [51], policy-critical [51], and descriptive content. Repression-related categories comprised government propaganda [11], distracting information [11, 24], and a newly emerged category: demoralizing content. Topics that did not fit any of these categories due to artifacts in the clustering algorithm were labeled as not classifiable.

As a final step in the CGT process, we conducted a classification and validation step to assess the robustness of topic categorization, given that topics were initially labeled based on the top 20 tweets per topic. We validated our findings by drawing a 5% random sample of all tweets with a probability value below 0.8. This procedure yielded 1289 tweets, which the two authors manually coded to assess whether each tweet was correctly assigned to its respective topic category. The *BERTopic* model achieved a weighted macro F1-score of 0.81 (for details, see Appendix C). Additionally, the two authors manually labeled all remaining tweets (n = 3575) that could not be clearly allocated to a specific topic.[6] Finally, we examined the temporal dynamics of digital protest and repression, along with the temporal characteristics of the coordinated accounts, tweets, and co-retweeting actions across the topic categories.

To address RQ3, we extended the content-based analysis with a structural perspective by examining how coordinated accounts cluster into communities and how language and thematic focus are distributed within these clusters. Language use provides an indicator of audience orientation and potential transnational reach, while community detection captures patterns of coordinated behavior that are not visible from tweet-level content alone. We applied the Louvain algorithm [4] to detect communities within the coordinated account dataset. First, we merged the labeled topics with the coordinated account dataset, assigning each co-retweet to the topic category of its original tweet. We then matched all tweets with their respective languages (for details, see Appendix D), which allowed us to identify language-specific communities within the coordinated network. Finally, we calculated the frequencies of shares by language and topic category, examined the proportions of topic categories and languages within communities, and conducted a chi-square test followed by an analysis of standardized residuals [48] to determine the relationship between topic categories and languages.[7]

6 Results

The coordination detection yielded a co-retweet network comprising 13,557 unique accounts, with the average account retweeting six tweets. Within this network, 1514 accounts repeatedly retweeted more than 100 tweets at least twice within a 120-second time frame throughout the entire duration of protests. Among the coordinated tweets, we identified 14,748 unique tweets in simplified Chinese, 2329 in traditional Chinese, and 13,553

[6] A random sample of n = 214 tweets was used to conduct an intercoder reliability test beforehand. The coders achieved a Krippendorff's Alpha of 0.76 with 7 different categories.

[7] The scripts used for the analyses are available in the supplemental materials: https://osf.io/egxrq.

in English. Our analysis shows that out of the 8,153,122 tweets captured through our data collection strategy during the protests, 739,819 (9%) were retweets amplified by accounts engaged in coordinated behavior. Notably, coordination was most extensive among accounts using simplified Chinese, followed by accounts tweeting in English.

RQ1 inquired about the main themes of tweets published by coordinated Twitter accounts during the COVID-19 protests in China. Among the 161 topics identified computationally and labeled manually, 48 protest-related topics explicitly criticized the regime and ruling elites. These topics primarily included demands for Xi Jinping and the Communist Party to step down, along with criticism of the regime and its tyranny, communist ideology, the "real foreign forces" (e.g., Marx, Engels, Lenin, or Hitler), and calls for freedom and revolution. Another 41 protest-related topics focused on criticism of the zero-COVID policy, lockdowns, nucleic acid tests, quarantine camps, COVID passports, domestic violence amid lockdowns, working conditions at Foxconn, censorship, police brutality during street protests, and Apple's alleged collaboration with the CCP. The remaining 19 protest-related topics did not contain criticism or take specific standpoints but instead provided live reports of street protests in various cities in China and abroad. These descriptions, often accompanied by photos and videos, factually documented protest events. Some descriptive topics also covered related developments explicitly mentioning the protests, such as oil price fluctuations and stock market drops, Jiang Zemin's death amid the protests, Apple's concerns about the unrest, military movements, and the gradual lifting of COVID-19 restrictions as a consequence of the protests.

Unlike themes that amplified the protests, some topics focused on unrelated issues while still incorporating protest-related keywords and hashtags. Such content does not articulate counter-frames or direct criticism of protesters; instead, it occupies discursive space associated with protest communication, thereby diluting attention and suppressing the visibility of protest-related tweets. Among these were 18 topics containing unrelated information, such as various religious content, the situation in Tigray, reports on Jiang Zemin's death, gambling advertisements, weather conditions in different cities, lives of celebrities (e.g., Zhang Zhehan), discussions on nuclear weapons, business-related content (e.g., Luis Vuitton), and miscellaneous advertisements. Another five topics explicitly praised the regime and its achievements while avoiding direct references to the protests. These included narratives on China's advancement in the space industry, praise for Xi Jinping's diplomacy, reports on flourishing trade between the European Union and China, and whataboutism (e.g., what about the events happening in the declining West while China thrives?). Finally, corroborating media reports, we identified 10 topics featuring adult content, which included suggestive texts, emojis, contact information (e.g., QQ numbers), as well as explicit videos and images. While such content is not inherently political, its systematic co-occurrence with protest-related hashtags situates it within protest-relevant communicative spaces. The remaining 20 topics were categorized as not classifiable due to their mixed content and ambiguity, as the tweets within them lacked a clear and consistent thematic focus.

RQ2 asked how the identified topics supported digital protest and repression on Twitter during the COVID-19 protests in China over time. As part of CGT, we first grouped the identified topics into six categories, as shown in Table 1, based on the results of topic modeling and the literature on authoritarian publics and digital repression.

Table 1 Overview of topic categories

Topic category	Description
Protest	
Leadership-critical	Explicit criticism of the country's political leadership, including calls for the resignation of leaders, revolution, and regime change
Policy-critical	Explicit criticism of policies, lower-level institutions and officials, including grievances over policies, demands for policy change, and support for protests
Descriptive	Informative accounts of events, including images and videos documenting protests, presented without explicitly stated opinions
Repression	
Government propaganda	Content that portrays the country's political leadership, regime, policies, and actions in a favorable light
Distracting information	Seemingly inoffensive, irrelevant content or spam unrelated to protests, including generic advertisements, news, or updates that do not engage with the protest issues
Demoralizing content	Offensive and harmful content injected into protest-related discourse to degrade the public discussion space and delegitimize protests and hosting platforms
Not classifiable	Potentially protest-related content with ambiguous stances, including unclear humor or sarcasm, references to deleted materials, or cryptic messaging

We distinguished three categories aimed at supporting and amplifying protests: leadership-critical [51], policy-critical [51], and descriptive. The leadership-critical category is characterized by direct criticism of the government and ruling elites, as well as calls for revolution and regime change. Policy-critical discourse is directed toward specific policies and actions of local governments and law enforcement officers. Descriptive content avoids direct criticism and provides an informative account of protest-related events.

We further distinguished three categories aimed at suppressing protests and supporting repression: government propaganda [11], distracting information [11, 44], and demoralizing content. Government propaganda involves narratives in favor of the government's image and objectives. Consistent with research on Chinese state communication, such content emphasizes positive portrayals of China—such as economic progress, technological advancement, or diplomatic success—thereby redirecting attention away from contentious or destabilizing events rather than engaging them directly [24, 35]. Distracting information contains irrelevant materials that can shift attention from the protest-related discourse. Rather than countering protest claims, this content diverts visibility and engagement through volume and topical displacement. Demoralizing content is characterized by harmful and offensive materials that can disengage users exposed to it (e.g., adult content, harassment, or graphic content). Unlike direct criticism of protest activities, the themes identified in this category suggest a systematic pattern of discouraging engagement with protest-related hashtags and keywords, potentially deterring users from searching for or participating in protest-related discourse on Twitter.

The remaining topics lacked sufficient information to be assigned to any category and were labeled as not classifiable (the seventh category). Table 2 provides a summary of all classified tweets in the dataset.

Next, we plotted the time dynamics of each topic category for all coordinated accounts engaged in co-retweeting behavior, as shown in Figures 2 and 3. The largest group consisted of policy-critical tweets, reflecting coordinated efforts to amplify dissatisfaction

Table 2 Number of retweets per topic category shared from 2022-11-24 to 2022-12-08

Topic category	Co-retweets	Retweets by coordinated accounts	Total retweets
descriptive	49,643	56,008	134,467
policy-critical	147,148	317,887	907,260
leadership-critical	30,402	248,873	757,172
government propaganda	590	5367	50,067
distracting information	18,419	40,722	60,233
demoralizing content	37,938	68,244	136,369
not classifiable	527	2718	5738
Total	284,667	739,819	2,051,306

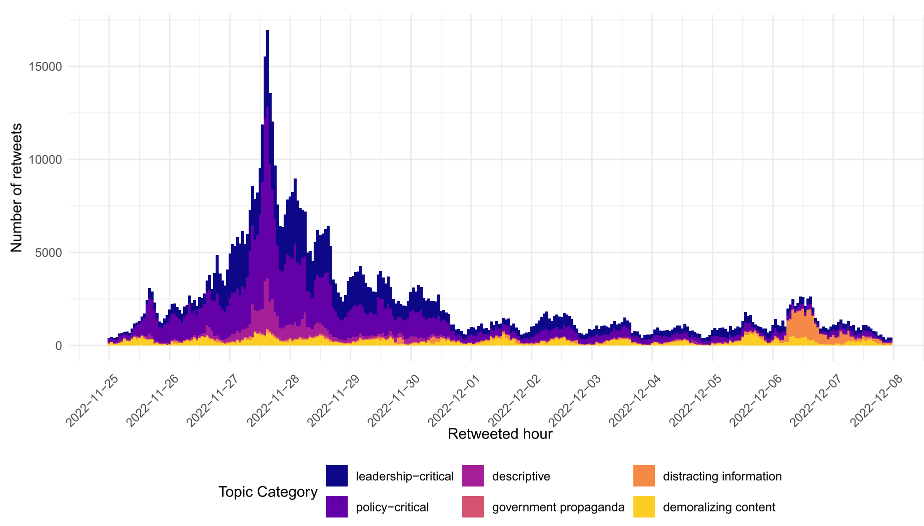

Figure 2 Hourly retweet counts by coordinated accounts across topic categories. N = 739,819; x-axis in UTC; no overlaps on y-axis

with the COVID-19 policies and their consequences. Leadership-critical tweets gained traction during the country-wide protests on November 26-28, with a steep increase on November 27. Descriptive tweets, primarily reporting on protest events, increased on November 26 following the surge in leadership-critical tweets. Notably, spikes in demoralizing content resemble those of government propaganda during the active phase of protests, suggesting heightened activity from repressive accounts. Nevertheless, while government propaganda sharply declined on December 1, demoralizing content continued to be retweeted regularly, though at a reduced frequency. Finally, distracting information showed less consistency, with spikes of coordinated activity on November 30 and December 6.

We further inspected the age of accounts engaged in coordinated behavior across the six topic categories. Figure 4 shows that accounts posting demoralizing content and distracting information tend to be significantly younger than those associated with the remaining topics.[8]

Subsequently, we investigated whether there were significant differences among topic categories based on the average delay of retweeting at the *level of individual tweets*. We

[8] A negative binomial regression confirmed that the difference was statistically significant.

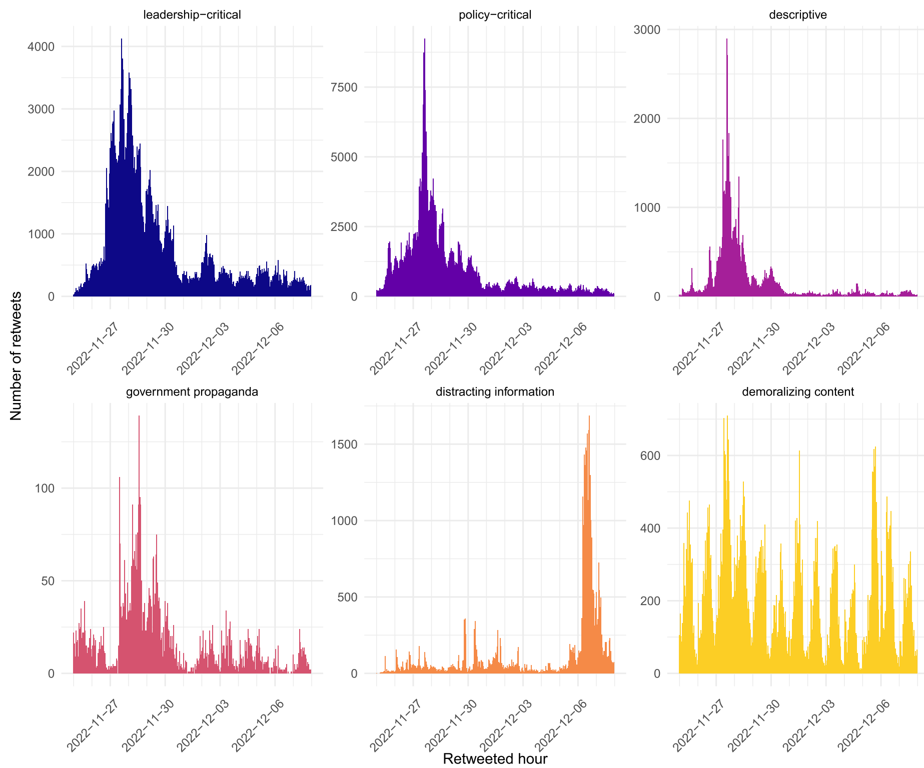

Figure 3 Hourly retweet counts by coordinated accounts separated by topic category. N = 739,819; x-axis in UTC, varying scales on the y-axis

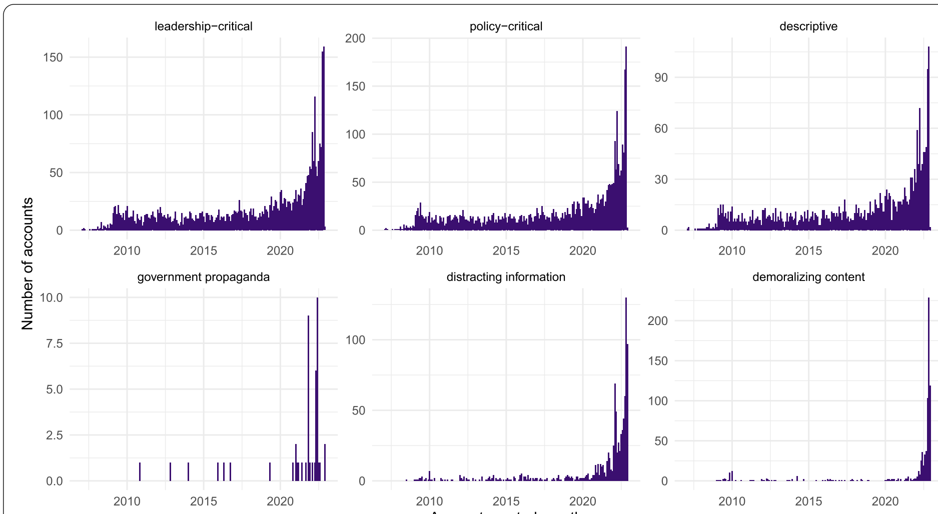

Figure 4 Monthly creation date counts of coordinated accounts by topic category. Accounts posting ambiguous content are excluded. N = 5,016

specified a negative binomial regression model, with retweet delay (measured in seconds) as the dependent variable and topic association for a tweet as the independent variable. Our analysis revealed that retweet delays were shortest for distracting information and longest for demoralizing content. In other words, *new* distracting information

Table 3 Number of co-retweets in English, simplified Chinese, and traditional Chinese by topic category

Topic category	English	Simp. Chinese	Trad. Chinese	N
leadership-critical	6579	22,931	892	30,402
policy-critical	9936	136,225	987	147,148
descriptive	1615	47,876	152	49,643
government propaganda	164	404	22	590
distracting information	11,614	6746	59	18,419
demoralizing content	24,730	13,161	47	37,938
not classifiable	329	189	9	527
Total	54,967	227,532	2168	284,667

was retweeted instantly, while *old* demoralizing content was retweeted afresh during the protests. This pattern contrasts with typical commercial spam dynamics, which tend to rely on the rapid promotion of newly created content and increasing hashtag activity to maximize visibility. The observed reactivation of older demoralizing content suggests a different propagation logic, raising questions about whether such activity reflects genuine advertising efforts or a more strategic, demand-driven use of content during moments of heightened protest-related attention. Additionally, policy-critical content was retweeted slightly slower than descriptive content.

Considering the behavior at the *account level*, we also examined the average speed of co-retweeting. We excluded accounts that had retweeted both protest- and repression-supporting content, as their behavior was ambiguous and potentially erratic. This resulted in a subset of 5016 accounts with consistent content patterns. Using linear regression, with the account's average retweet delay as the dependent variable and topic category as the independent variable, we found that accounts retweeting distracting information were the fastest, followed by accounts retweeting demoralizing content. Differences between the other topic categories were not significant.

RQ3 asked about the languages used to support protest and repression within the coordinated network and its communities. Table 3 shows the breakdown of co-retweets across three languages for protest-supporting and repression-supporting topic categories. Protest-related content was primarily co-retweeted in simplified Chinese, with a focus on policy-critical and descriptive content. Co-retweets in traditional Chinese and English were more focused on leadership and policy criticism. Among repression-supporting topic categories, English-language content comprised a larger share, suggesting that distracting and demoralizing information primarily targeted international English-speaking audiences. Government propaganda remained more prominent in simplified Chinese, indicating an effort to influence public opinion within the PRC. These findings were confirmed by a chi-square test (for details, see Appendix E).

A chi-square test further revealed significant differences in the language distribution across content categories (χ^2 (12, N = 284,667) = 100,110.44, $p < .0001$). The standardized residuals indicate distinct patterns: demoralizing content was strongly overrepresented in English (std. res. = 243.16) and underrepresented in simplified Chinese (std. res. = –236.30) and traditional Chinese (std. res. = –15.35). Similarly, distracting information appeared more frequently in English (std. res. = 155.52) and less in simplified Chinese (std. res. = –151.72). Leadership-critical content was predominantly shared in traditional

Chinese (std. res. = 46.10), while descriptive content was overrepresented in simplified Chinese (std. res. = 101.09) and underrepresented in English (std. res. = −99.74).

The Louvain algorithm identified 1534 communities, with the ten largest communities collectively accounting for approximately 85.4% of all co-retweets (see Appendix F). This indicates that a small number of large, coordinated communities operated along many smaller ones. For our analysis, we focus on the ten largest communities and their topic and language compositions. As shown in Fig. 5, these communities largely centered on one language and one primary goal—either supporting protests or promoting repression. Four distinct communities shared repression-supporting topics, primarily focusing on demoralizing content or distracting information in simplified Chinese and English (communities 08, 06, 05, and 09). In contrast, the largest community (community 02) and five others (communities 01, 03, 04, 07, and 10) were primarily engaged in sharing protest-supporting tweets across all three categories, each dominated by a single language. Traditional Chinese tweets mainly contained leadership-critical and policy-critical content, accompanied by posts in simplified Chinese (communities 01 and 03).

It is also evident that protest-supporting communities contained both policy-critical and leadership-critical publics, with the majority prioritizing policy criticism. Among these communities, simplified Chinese tweets placed less emphasis on descriptive content, while English-language communities were more actively engaged in documenting the protests through descriptive posts. By contrast, repression-supporting communities had a clearer focus on a single repression tactic and a single language. Interestingly, smaller communities (labeled as "other") were predominantly focused on demoralizing content in English, suggesting that this form of digital repression was coordinated through dispersed groups of accounts. Some communities showed discrepancies, containing a small number of tweets from the opposing camp, which may be due to artifacts in the topic modeling or the behavior of automated accounts.

7 Discussion

This study explored coordinated behavior on Twitter during the COVID-19 protests in China, focusing on accounts tweeting in simplified Chinese, traditional Chinese, and English. We applied computational methods to identify coordinated tweets, analyze their themes, and examine how they supported protest and repression through organized retweeting within coordinated communities across the three languages. We found that 9% of the tweets in our comprehensive sample were actively amplified in a coordinated manner. Extensive coordination was predominantly carried out among communities tweeting in simplified Chinese, followed by accounts tweeting in English. These findings suggest that coordinated efforts on Twitter during the COVID-19 protests in China targeted primarily audiences proficient in simplified Chinese and the global English-speaking community.

In simplified Chinese discourse, coordinated topics ranged from policy-critical conversations, protest-related reporting, and leadership-critical appeals to government propaganda, distracting information, and demoralizing content, in descending order. The themes discussed in traditional Chinese contained predominantly policy-critical and leadership-critical grievances. In contrast, coordinated tweets in English featured more demoralizing content, followed by distracting information. They also included policy-critical and leadership-critical posts, along with some descriptive reports on protests,

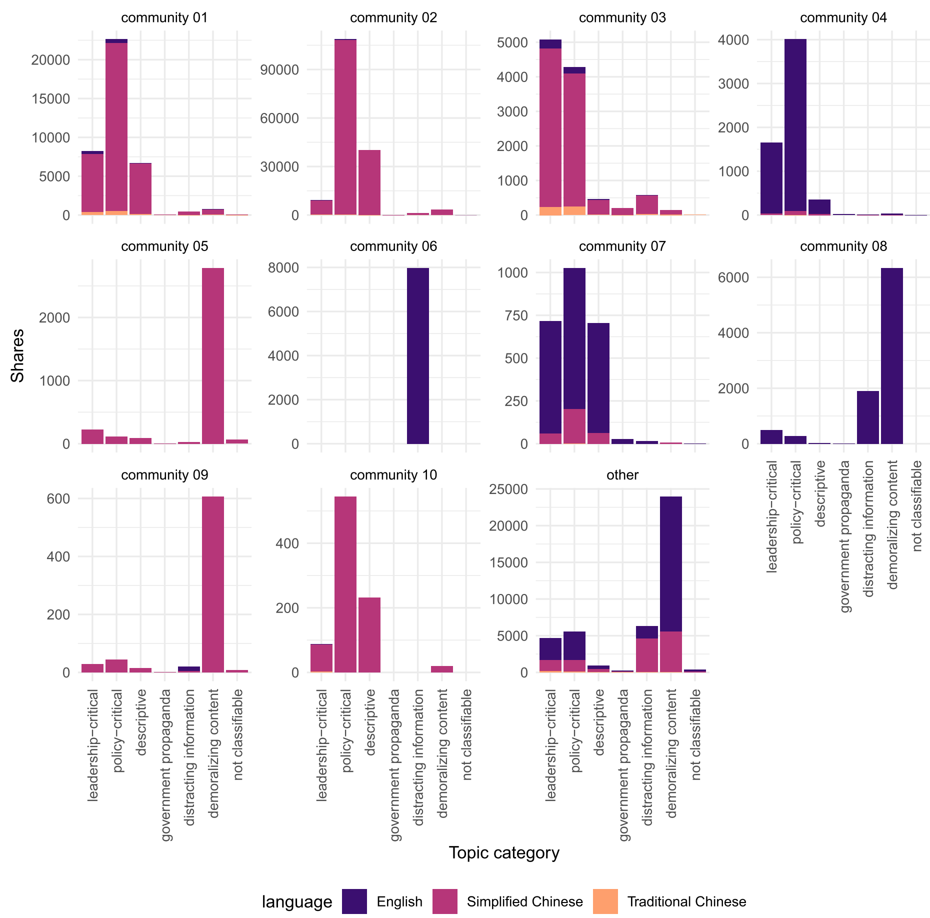

Figure 5 Topic categories and language distribution in the ten largest communities

though significantly fewer than coordinated tweets in simplified Chinese. Additionally, when aggregating the identified topics and examining their temporal distribution, we found that coordinated repression generally mirrored the trends seen in protest-related tweets during the active phase of protests.

The language-based differences indicate that coordinated distracting and demoralizing content interfered with protest discourse on Twitter in simplified Chinese, and even more so in English. In other words, ordinary users searching for protest-related information in these languages would inadvertently encounter numerous unrelated and inappropriate tweets shared in a coordinated manner. Given that these tweets incorporated protest-related hashtags or keywords and were shared in a coordinated way, it is plausible that they were intended to divert attention from protest-related content on the platform. This conclusion aligns with the findings of King et al. [24], who uncovered a secretive operation by the Chinese government aimed at distracting the public and shifting social media conversations rather than confronting regime critics. In the framework of digital repression [11], this strategy falls under the category of *covert information channeling*, a tactic used by state or private actors to redirect attention away from social mobilization. We argue that *demoralizing information*, mainly containing adult content in this case, represents another form of covert information channeling, with a different objective and effect.

Rather than simply redirecting attention to another topic, demoralizing content seeks to undermine the credibility, morale, and integrity of protests and the platform itself. It may discourage users from searching for protest-related content through hashtags and keywords and deter them from using the platform where such content is circulated. Notably, within the context of the COVID-19 protests in China, this strategy appears more targeted toward international audiences rather than domestic ones, suggesting that digital repression is used to reduce foreign support and prevent international pressure on the regime. It also shows that digital repression can easily transcend borders on a global platform like Twitter.

Within the framework of authoritarian publics, Twitter served as an environment where uncritical (including pro-regime), policy-critical, and leadership-critical publics coexisted during the COVID-19 protests in China. The issue-specific character of the protests was reflected in the overall predominance of policy-critical content, including criticism of lockdowns, nucleic acid tests, COVID passports, and quarantine camps. This pattern may indicate attempts to reach broader, more heterogeneous publics—ranging from uncritical or pro-regime users to transnational audiences—for whom policy-focused critique is potentially more accessible, resonant, and shareable than direct denunciations of the leadership. Policy criticism can also be perceived as relatively safer and more sustainable for coordination and amplification than leadership criticism. Even so, despite higher perceived risks for users with ties to China, leadership-critical and regime-challenging content did gain traction. This is a noteworthy case, particularly for an authoritarian context like the PRC, where local media environments typically lack the presence of leadership-critical publics [51]. Our study shows that under favorable circumstances, such publics can surface and amplify critical discourse. It is crucial for an authoritarian society, where policy change is often unlikely to occur without substantial pressure on the leadership itself. Historical reflections on Chinese political culture illuminate this point. Lu Xun famously remarked:

> By temperament the Chinese love compromise and a happy mean. For instance, if you say this room is too dark and a window should be made, everyone is sure to disagree. But if you propose taking off the roof, they will compromise and be glad to make a window. In the absence of more drastic proposals, they will never agree to the most inoffensive reforms. [32, p. 165]

Our study further highlights the importance of taking into account opposing publics and discursive practices aimed at suppressing or diverting attention from criticism, as well as demoralizing potential protest supporters. This is particularly relevant in interactive communication spaces such as social media. We contribute to the understanding of *authoritarian publics* [51] by integrating the concept and typology of *digital repression* [11]. Specifically, we propose examining authoritarian publics through the lens of digital repression techniques, particularly information channeling, which can influence the volume and forms of criticism within policy-critical or leadership-critical authoritarian publics.

We further expand the framework of digital repression by identifying another distinct covert information channeling technique—*demoralizing content*—aimed at undermining protests and demotivating users from engaging with protest-related discourse. Rather than replacing existing repression strategies, this technique appears to operate alongside other well-documented forms of online manipulation. As this technique primarily targeted English-speaking audiences on a platform hosting multiple publics, we can infer that democratic publics may be susceptible to authoritarian repressive techniques. Within China, digital repression typically focuses on spreading positive energy, guiding public opinion, and reducing emotional outpourings [28]. Our findings suggest that, beyond domestic platforms, digital repression likely combines overt and covert strategies: continuing to disseminate favorable narratives and positive representations, while simultaneously engaging in coordinated campaigns that flood protest-related hashtags and keywords with spam-like or inappropriate content.

According to the framework of digital repression outlined by Earl et al. [11], covert information channeling is not primarily designed to persuade; rather it aims to alter the informational environment in which contention unfolds. Our findings suggest that on global social media platforms, repression effects are diversified rather than replaced, continuing to include the promotion of positive energy or advanced socialist culture while incorporating additional covert tactics. The latter function to degrade the quality and usability of protest-related discourse by increasing noise, confusion, and affective discomfort, ultimately discouraging ordinary users' engagement with protest-related content.

The lack of moderation on Twitter during the Chinese protests enabled information channeling, whether carried out by government agents or private actors. This shows that while platform interference can impact activism [11], inaction can also have negative consequences. Therefore, digital repression strategies such as *information channeling* and *information coercion*, regardless of the actors behind them, can have detrimental consequences for social movement capacities worldwide. Covert forms of digital repression pose an even greater threat, as they are challenging to recognize, and the intent or identity of the perpetrators is usually difficult to establish. Studying the extent of coordinated digital repression in democratic publics can thus be an area for future research.

It is important to acknowledge the limitations of this study. While Twitter API v2 allowed us to promptly collect a comprehensive dataset of tweets published during the COVID-19 protests in China, our sample lacks deleted tweets, especially those removed during the first three days of the protests. However, since our data collection coincided with Twitter's suspension of its content moderation and policy teams [10], we likely captured a more comprehensive dataset than would typically be possible. Nevertheless, it is important to note that in authoritarian contexts, users may engage in self-censorship by deleting their tweets within the first hours of protests. Consequently, we may not have captured this data. In addition, our analysis is limited to a single platform which represents only one segment of a broader, highly fragmented communication ecology. Protest-related discourse and repression strategies may unfold differently across platforms with distinct affordances, governance regimes, and audience composition. A cross-platform comparative design would therefore be a valuable direction for future research, allowing scholars to examine how coordination and audience targeting vary across interconnected yet institutionally distinct platforms.

Additionally, our data collection relied on a comprehensive but not exhaustive list of protest-related keywords. As a result, there is a possibility that some protest-related tweets may have been excluded from our sample simply because they did not contain one of these keywords. This issue is particularly relevant for tweets containing images or videos with minimal textual content. However, this limitation also indicates that a broader audience may have encountered challenges in finding such posts using popular protest-related hashtags or keywords. Consequently, these posts might have had lower visibility and less relevance for studying coordinated behavior, which aims to amplify specific content. Another limitation concerns our operationalization of coordination. Because our focus was on the themes amplified during the protests, we intentionally opted for co-retweeting behavior—a well-established and robust form of synchronized amplification. However, this design choice means that we may overlook other forms of coordination, such as synchronized URL sharing, hashtag co-use, near-duplicate text production, simultaneous original or quote tweeting, or coordinated replies and mentions. Shifting the lens from themes and language to coordinated actors and actions, future research could apply a multimodal coordination detection to provide a more comprehensive account of coordinated activity.

Finally, it is essential to clarify that we do not make definitive claims regarding the identities of the entities responsible for organizing online protest and repression. We cannot rule out the possibility that distracting or demoralizing content may have hijacked protest-related hashtags and keywords for genuine purposes, such as drawing attention to a certain issue or attracting new customers for products or services. At the same time, our results indicate that accounts spreading demoralizing content exhibited behavioral patterns that differ systematically from those commonly associated with opportunistic commercial spam. In particular, these accounts primarily reactivated older content during key protest moments rather than promoting newly created material, and their activity was tightly synchronized with protest-related peaks (see Results section for RQ2). These patterns suggest a form of coordinated intervention into the protest-related information space that is difficult to explain solely by profit-seeking motives, yet also hard to attribute to pro-regime actors. Importantly, this activity affected the discourse composition on Twitter and ultimately benefited the regime by reducing the visibility of protests. In this sense, our findings speak to effects rather than attribution: even in the absence of conclusive evidence about organizers, the observed coordination aligns with mechanisms of covert information channeling described in the digital repression literature [11]. Finally, our analysis focuses primarily on coordination at the level of tweets and retweeting behavior, rather than on a systematic examination of account-level characteristics. While our analysis incorporates selected account-level characteristics, such as account creation dates and average retweet delay, to differentiate coordination patterns, we do not conduct a comprehensive account-level profiling of coordinated actors. Incorporating an investigative approach could enable a more fine-grained grouping of coordinated accounts into distinct cliques or roles, helping to distinguish, for example, accounts primarily engaged in opportunistic spam from those participating in more sustained and strategically timed coordination. Future research combining coordination detection with richer account-level diagnostics could therefore address questions about the organization, division of labor, and persistence of coordinated repression and protest activity.

Despite its limitations, our study shows how online discourse supported and suppressed China's COVID-19 protests on Twitter through coordinated sharing in simplified Chinese, traditional Chinese, and English. By conducting this study, we responded to calls for research on digital spaces that transcend China's borders and go beyond conventional tools of repression [39]. The tweets we analyzed were posted by users from all over the world, including those locked within the *Great Firewall*. They likely reflected the grievances of individuals who know how to get around the firewall, the relative "winners" of China's economic reform [20]. However, it is crucial to remember that the protests were sparked by the plight of less privileged citizens in Xinjiang, whose voices were inadequately represented in online discourse. While the protests led to the relaxation of COVID-19 measures, it remains unclear whether the underlying grievances of marginalized groups and individuals were addressed. Distracting and demoralizing content targeting English-speaking audiences further hindered global awareness and advocacy for these repressed voices. This highlights the need for more coordinated and inclusive efforts to amplify protests in authoritarian regimes and counteract repression on an international scale.

Appendix A

A.1 Keywords for data collection

The list of keywords includes terms such as "白纸运动", "白紙運動", "白纸革命", "白紙革命", "乌鲁木齐", "烏魯木齊", "共产党下台", "共產黨下台", "习近平下台", "習近平下台", "习近平", "習近平", "A4纸革命", "A4紙革命", "清华", "清華", "新疆", "白纸抗议", "白紙抗議", "中共下台", "女权", "女權", "颜色革命", "顏色革命", "白纸起义", "白紙起義", "大陆", "大陸", "境外势力", "境外勢力", "上海", "北京", "广州", "廣州", "西方主流媒体", "疫情", "A4Revolution", "FreeChina", "foxconn", "Urumqifire", "ChinaRevolution", "ChinaProtests", "ChinaUprising", "ChinaProtest2022", "A4 Revolution", "Free China", "China Revolution", "China Protests", "China Uprising", "China Unrest", "Urumqi fire", "Urumqi", "Ürümqi", "Wulumuqi", "Beijing", "Shanghai", "Guangzhou", "Xi Jinping step down", "CCP step down", "white paper movement", "Communist party step down".

Appendix B

B.1 Robustness of coordination detection

It could be argued that coordinated activity during protest events may capture routine spam and noise on Twitter, despite the use of protest-related hashtags and keywords. To verify that the coordinated activity we identified during the period of protests (November 25 to December 7, 2022) was not routine, we performed additional detection of coordination for the subsequent period extending to December 19. We found that coordinated activity has massively decreased, with only 802 accounts co-retweeting 1315 unique posts. Of the original 759 accounts co-retweeting demoralizing content only 75 remained engaged in coordination in the period immediately following the protests. Moreover, the number of accounts co-retweeting distracting information decreased from 949 accounts to merely 23. And finally, the 46 accounts spreading government propaganda completely desisted. Due to our continuous data collection during the protests and afterward, we can rule out the possibility that these accounts and their posts were deleted.

To assess whether the post-protest decline in coordinated activity was driven by one side of the network, we also investigate coordination by community affiliation. As shown in Table 4, both protest-supporting and repression-supporting communities exhibit clear declines after December 8, though the reduction is generally more pronounced among repression-aligned clusters. One repression community (community 08) ceased coordination entirely (–100%), and the broader repression-aligned "other" group dropped by 49%, whereas protest-supporting clusters show more moderate but still substantial reductions (ranging from –1.6% to –62.6%). This pattern indicates that the decrease in coordination

Table 4 Decline in the number of accounts within coordinated communities

Community	Affiliation	N accounts before Dec 8	N accounts after Dec 8	Decline (%)
community 01	Protest	1401	1137	−18.84
community 02	Protest	1217	752	−38.21
community 03	Protest	379	373	−1.58
community 04	Protest	637	238	−62.64
community 05	Repression	355	340	−4.23
community 06	Repression	255	160	−37.25
community 07	Protest	83	63	−24.1
community 08	Repression	242	0	−100.00
community 09	Repression	66	62	−6.06
community 10	Protest	46	26	−43.48
other	Repression	3144	1598	−49.17

is system-wide and reflects an event-specific surge rather than routine or automated activity.

Appendix C

C.1 Validation metrics of BERTopic

Table 5 Validation metrics of BERTopic

Topic category	Precision	Recall	F1
leadership-critical	0.84	0.87	0.85
policy-critical	0.86	0.82	0.84
descriptive	0.95	0.51	0.67
government propaganda	0.93	0.54	0.68
distracting information	0.83	0.86	0.84
demoralizing content	0.84	0.83	0.83
not classifiable	0.44	0.94	0.60
overall	0.84	0.80	0.81

Appendix D

D.1 Language detection method

The Twitter API returns a language classification for each tweet that only contains one label for Chinese (zh). Automatically distinguishing content posted in traditional Chinese from simplified Chinese can be challenging, because there are many cases where the tweet Authors may use both scripts together (e.g., writing in traditional Chinese, but using hashtags in simplified Chinese). Therefore, the character usage is arguably a continuous measurement, rather than discrete. In order to distinguish tweets composed in simplified Chinese from traditional Chinese, we resorted to a dictionary-based approach. The tmcn package [29] contains a dictionary of simplified and traditional Chinese that can be used to translate characters from one script to another. We used this reference to count all traditional and simplified Chinese characters in a tweet, then we divided each count by the total number of Chinese characters for each tweet. Finally, we subtracted the proportion of simplified Chinese characters from the proportion of traditional Chinese characters, yielding a score ranging from –1.0 to 1.0. Negative values indicate more simplified Chinese characters than traditional Chinese characters, while positive values indicate the opposite. We then assigned tweets to traditional Chinese if their score was greater than zero and to simplified Chinese if their score was zero or below.

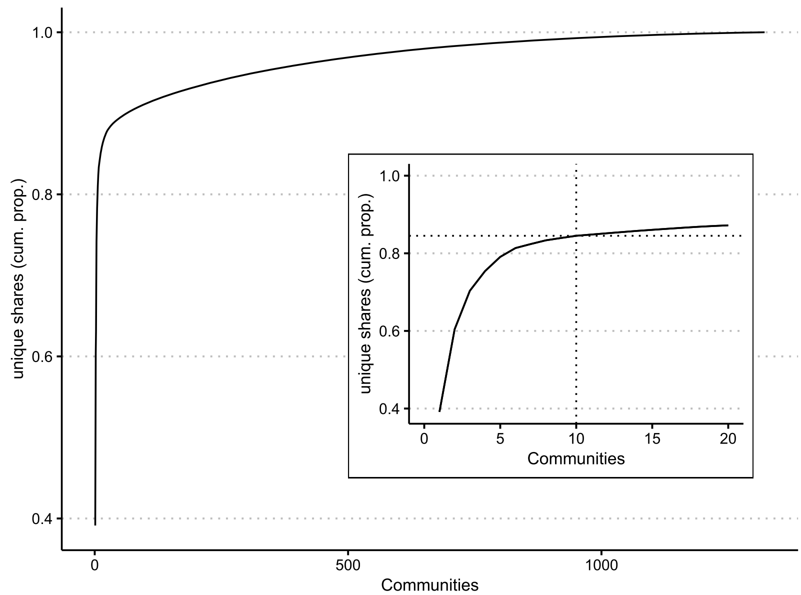

Figure 6 Cumulative shares for communities detected with the Louvain algorithm

Appendix E

E.1 Chi-square test

Protest-supporting content appears to be shared predominantly in Chinese, while repressive and, to a lesser extent, not classifiable content is more likely to be posted in English, according to a chi-square test, χ^2 (2, N = 284,667) = 92,699, p < .0001. This conclusion is supported by the standardized residuals [48], which show a significant overrepresentation of protest content in Chinese (std. res. = 304.46) and repressive content in English (std. res. = 302.82), with not classifiable content also showing a moderate tendency toward English (std. res. = 25.10).

Appendix F

To evaluate the robustness of the Louvain community partitions, we followed a common practice for stochastic community-detection algorithms and generated an ensemble of 100 independent Louvain runs [27]. Pairwise similarity between two randomly drawn partitions was quantified using both the Rand Index and the Adjusted Rand Index (ARI). The high mean Rand Index (0.976) and Adjusted Rand Index (0.812) show that the algorithm repeatedly returns highly similar community structures despite random initializations, confirming the robustness of the identified modular organization. Cumulative shares for communities detected with the Louvain algorithm are shown in Fig. 6.

Abbreviations

API, application programming interface; ***CCP***, Chinese Communist Party; ***CGT***, Computational Grounded Theory; ***COVID-19***, coronavirus disease 2019; ***PRC***, the People's Republic of China; ***VPN***, virtual private network.

Acknowledgements

We thank the anonymous reviewers for their thoughtful comments and helpful suggestions.

Author contributions

AK contributed to the conception and design of the study, theoretical framework, methodology development, data collection, formal analysis, data interpretation, manual content coding, and manuscript writing. PB contributed to the study design, methodology development, formal analysis, data interpretation, manual content coding, and manuscript writing. NR contributed to the study design, methodology development, formal analysis, data interpretation, and manuscript writing. AW contributed to the study design, theoretical framework, critical feedback, and manuscript writing. All authors have read and approved the final manuscript.

Authors' information
AK and PB have lived in mainland China and Hong Kong for extended periods and are proficient in Chinese. Their contextual familiarity and linguistic proficiency have informed their understanding of the data and ensured its accurate interpretation.

Funding information
Open access funding provided by University of Vienna. No funding was received for this study.

Data availability
The dataset used in this study was collected from Twitter (now X) using the Twitter API v2. While the raw data cannot be shared due to the terms of service and ethical considerations, the scripts for data collection are provided in the supplemental materials to enable replication: https://osf.io/egxrq.

Declarations

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Communication, University of Vienna, Kolingasse 14-16, 1090, Vienna, Austria. [2]GESIS – Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667, Cologne, Germany. [3]Department of Communication Science, Humanities and International Studies, University of Urbino Carlo Bo, Via Aurelio Saffi 2, 61029, Urbino, Italy.

Received: 22 September 2025 Accepted: 2 March 2026 Published online: 13 March 2026

References

1. Alqazlan L, Fang Z, Castelle M, Procter R (2025) A novel, human-in-the-loop computational grounded theory framework for big social data. Big Data Soc 12(2). https://doi.org/10.1177/20539517251347598
2. Barrie C, Ho JC (2021) AcademictwitteR: an R package to access the Twitter academic research product track v2 API endpoint. J Open Source Softw 6(62):3272. https://doi.org/10.21105/joss.03272
3. Bennett WL, Segerberg A (2012) The logic of connective action: digital media and the personalization of contentious politics. Inf Commun Soc 15(5):739–768. https://doi.org/10.1080/1369118X.2012.670661
4. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008
5. Brussee V (2022) Authoritarian design: how the digital architecture on China's Sina Weibo facilitate information control. Asiascape Digit Asia 9(3):207–241. https://doi.org/10.1163/22142312-bja10033
6. Canevez RN, Maikovska K, Zwarun L (2024) Tactics and affordances in the mediatization of war: pro-Ukrainian cyber resistance on Telegram. Digit War 5:167–180. https://doi.org/10.1057/s42984-024-00096-x
7. Carlsen HB, Ralund S (2022) Computational grounded theory revisited: from computer-led to computer-assisted text analysis. Big Data Soc 9(1). https://doi.org/10.1177/20539517221080146
8. Chang KC, Hobbs WR, Roberts ME, Steinert-Threlkeld ZC (2022) COVID-19 increased censorship circumvention and access to sensitive topics in China. Proc Natl Acad Sci USA 119(4). https://doi.org/10.1073/pnas.2102818119
9. Creemers R (2017) Cyber China: upgrading propaganda, public opinion work and social management for the twenty-first century. J Contemp China 26(103):85–100. https://doi.org/10.1080/10670564.2016.1206281
10. Culpan T (2022) China protests give Musk a true taste of running Twitter. Bloomberg. https://web.archive.org/web/20221201221940/. https://www.bloomberg.com/opinion/articles/2022-11-30/china-protests-give-musk-a-true-taste-of-running-twitter. Accessed 30 November 2022
11. Earl J, Maher TV, Pan J (2022) The digital repression of social movements, protest, and activism: a synthetic review. Sci Adv 8(10). https://doi.org/10.1126/sciadv.abl8198
12. Feldstein S (2021) The rise of digital repression: how technology is reshaping power, politics, and resistance. Oxford University Press, London
13. Fu D, Göbel C (2024) Exposing state repression: digital discursive contention by Chinese protestors. Stud Comp Int Dev. https://doi.org/10.1007/s12116-024-09428-0
14. Gallagher M, Miller B (2021) Who not what: the logic of China's information control strategy. The China Q 248(1):1011–1036. https://doi.org/10.1017/S0305741021000345
15. Giglietto F, Righetti N, Rossi L, Marino G (2020) It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 Italian elections. Inf Commun Soc 23(6):867–891. https://doi.org/10.1080/1369118X.2020.1739732
16. Göbel C, Steinhardt HC (2022) Protest event analysis meets autocracy: comparing the coverage of Chinese protests on social media, dissident websites, and in the news. Mobilization 27(3):277–295. https://doi.org/10.17813/1086-671X-27-3-277
17. Göbel C (2020) The political logic of protest repression in China. J Contemp China 30(128):1–17. https://doi.org/10.1080/10670564.2020.1790897
18. Grootendorst M (2022) BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794
19. Han R (2015) Manufacturing consent in cyberspace: China's "Fifty-Cent Army". J Curr Chin Aff 44(2):105–134. https://doi.org/10.1177/186810261504400205

20. Hassid J (2015) China's responsiveness to Internet opinion: a double-edged sword. J Curr Chin Aff 44(2):39–68. https://doi.org/10.1177/186810261504400203
21. Huang ZA, Wang R (2021) Exploring China's digitalization of public diplomacy on Weibo and Twitter: a case study of the U.S.–China trade war. Int J Commun 15:1912–1939. https://ijoc.org/index.php/ijoc/article/view/15105
22. Keller FB, Schoch D, Stier S, Yang J (2020) Political astroturfing on Twitter: how to coordinate a disinformation campaign. Polit Commun 37(2):256–280. https://doi.org/10.1080/10584609.2019.1661888
23. King G, Pan J, Roberts ME (2013) How censorship in China allows government criticism but silences collective expression. Am Polit Sci Rev 107(2):326–343. https://doi.org/10.1017/S0003055413000014
24. King G, Pan J, Roberts ME (2017) How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. Am Polit Sci Rev 111(3):484–501. https://doi.org/10.1017/S0003055417000144
25. Kobilke L, Kulichkina A, Baghumyan A, Pipal C (2023) Blaming it on NATO? Framing the role of NATO in the full-scale Russian invasion of Ukraine on Twitter. Front Polit Sci 5, 1122439. https://doi.org/10.3389/fpos.2023.1122439
26. Kulichkina A, Righetti N, Waldherr A (2025) Protest and repression on social media: pro-Navalny and pro-government mobilization dynamics and coordination patterns on Russian Twitter. New Media Soc 27(9):5433–5454. https://doi.org/10.1177/14614448241254126
27. Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. Sci Rep 2:336. https://doi.org/10.1038/srep00336
28. Li AK (2015) Towards a more proactive method. China Perspectiv 2015(4):15–23. https://doi.org/10.4000/CHINAPERSPECTIVES.6840
29. Li J (2019) tmcn: a text mining toolkit for Chinese. R package version 0.2-13. https://CRAN.R-project.org/package=tmcn
30. Lindgren S (2019) Movement mobilization in the age of hashtag activism: examining the challenge of noise, hate, and disengagement in the #MeToo campaign. Policy Internet 11:418–438. https://doi.org/10.1002/poi3.212
31. Liao R (2022) Despite ban, Twitter downloads surge in China amid COVID protests. TechCrunch. https://techcrunch.com/2022/11/28/despite-ban-twitter-downloads-surge-in-china-amid-covid-protests/. Accessed, 29 November 2022
32. Lu X, Yang G (1973) Silent China: selected writings of Lu Xun. Oxford University Press, London
33. Mannocci L, Cresci S, Magnani M, Monreale A, Tesconi M (2026) Multimodal coordinated online behavior: trade-offs and strategies. Inf Sci 737:123125. https://doi.org/10.1016/j.ins.2026.123125
34. Mannocci L, Mazza M, Monreale A, Tesconi M, Cresci S (2024) Detection and characterization of coordinated online behavior: a survey. arXiv preprint. https://doi.org/10.48550/arXiv.2408.01257
35. Mattingly D, Incerti T, Ju C, Moreshead C, Tanaka S, Yamagishi H (2025) Chinese state media persuades a global audience that the "China model" is superior: evidence from a 19-country experiment. Am J Polit Sci 69:1029–1046. https://doi.org/10.1111/ajps.12887
36. Menn J (2022) Twitter grapples with Chinese spam obscuring news of protest. The Washington Post. https://web.archive.org/web/20221128004939/. https://www.washingtonpost.com/technology/2022/11/27/twitter-china-spam-protests/. Accessed 28 November 2022
37. Nelson LK (2020) Computational grounded theory: a methodological framework. Sociol Methods Res 49(1):3–42. https://doi.org/10.1177/0049124117729703
38. Purnell N, Cheng S (2022) Twitter becomes stage for China protests despite ban by Beijing. Wall St J. https://web.archive.org/web/20221130173946/. https://www.wsj.com/articles/twitter-becomes-stage-for-china-protests-despite-ban-by-beijing-11669829265. Accessed, 30 November 2022
39. Reese SD, Chen W, Pan Z (2023) Revisiting networked China: challenges for the study of digital media and civic engagement. Inf Commun Soc 26(2):239–252. https://doi.org/10.1080/1369118X.2022.2161831
40. Reimers N, Gurevych I (2019) Sentence-BERT: sentence embeddings using Siamese BERT-networks. In: Proceedings of the 2019 conference on empirical methods in natural language processing. arXiv:1908.10084
41. Rogers R, Righetti N (2025) Coordinated inauthentic behaviour on Facebook? A typology of manufactured attention. Platf Soc 2. https://doi.org/10.1177/29768624251369784
42. Righetti N, Balluff P (2025) CooRTweet: a generalized R software for coordinated network detection. Comput Commun Res 7(1). https://doi.org/10.5117/CCR2025.1.7.RIGH
43. Righetti N, Kulichkina A, Almeida Paroni B, Cseri ZF, Aguirre SI, Maikovska K (2025) Mainstreaming and transnationalization of anti-gender ideas through social media: the case of CitizenGO. Inf Commun Soc 28(15):2658–2681. https://doi.org/10.1080/1369118X.2025.2470229
44. Roberts ME (2018) Censored: distraction and diversion inside China's great firewall. Princeton University Press, Princeton
45. Rothschild V (2024) Protest and repression in China's digital surveillance state. J Inf Technol Polit, 1–16. https://doi.org/10.1080/19331681.2024.2326464
46. Schatto-Eckrodt T (2022) Hidden biases – the effects of unavailable content on Twitter on sampling quality. In: Jünger J, Gochermann U, Peter C, et al (eds) Grenzen, Probleme und Lösungen bei der Stichprobenziehung. Herbert von Halem Verlag, pp 178–195
47. Schoch D, Keller FB, Stier S, Yang J (2022) Coordination patterns reveal online political astroturfing across the world. Sci Rep 12:4572. https://doi.org/10.1038/s41598-022-08404-9
48. Sharpe D (2015) Your chi-square test is statistically significant: now what? Pract Assess Res Eval 20(8). https://eric.ed.gov/?id=EJ1059772
49. Thiele D, Milzner M, Heft A, Gong B, Pfetsch B (2025) Attributing coordinated social media manipulation: a theoretical model and typology. New Media Soc. https://doi.org/10.1177/14614448251350100
50. Toepfl F (2018) From connective to collective action: Internet elections as a digital tool to centralize and formalize protest in Russia. Inf Commun Soc 21(4):531–547. https://doi.org/10.1080/1369118X.2017.1290127
51. Toepfl F (2020) Comparing authoritarian publics: the benefits and risks of three types of publics for autocrats. Commun Theory 30(2):105–125. https://doi.org/10.1093/ct/qtz015
52. Waldherr A, Stoltenberg D, Maier D, Keinert A, Pfetsch B (2023) Translocal networked public spheres: spatial arrangements of metropolitan Twitter. New Media Soc 26(11):6636–6657. https://doi.org/10.1177/14614448231156579

53. Wang W, Tan X (2023) Demoralizing Internet contention: affective publics and emotional mobilization on China's social media during the Covid-19 pandemic. Crit Crim 31:417–431. https://doi.org/10.1007/s10612-023-09708-6
54. Weidmann NB, Rød EG (2019) The Internet and political protest in autocracies. Oxford studies in digital politics
55. Yang G (2009) China since Tiananmen: online activism. J Democr 20(3):33–36. https://doi.org/10.1353/jod.0.0094
56. Yuan L (2022). China's protest prophet. NY Times. https://www.nytimes.com/2022/12/07/briefing/china-protest-peng-lifa.html. Accessed 7 December
57. Zeng J (2020) #MeToo as connective action: a study of the anti-sexual violence and anti-sexual harassment campaign on Chinese social media in 2018. Journal Pract 14(2):171–190. https://doi.org/10.1080/17512786.2019.1706622
58. Zeng J, Cheng CY (2024) Diasporic citizen journalism: exploring the discussion on the 2022 blank paper protests in the Chinese Twitter community. Journalism 26(5):1047–1066. https://doi.org/10.1177/14648849241250191

Publisher's note