The Power of Social Networks: Information Elites and the Spread of Politically Sensitive Information under Media Censorship*

Xuebo Wang,[1] Hong Yang[2]

December 17, 2025

Abstract

This paper examines how politically sensitive information spreads through social networks under strict media censorship in China. Using a unique dataset from Sina Weibo, we map large-scale anonymized users' online social networks and identify key network nodes—referred to as "information elites"—who can circumvent censorship to access uncensored content. We analyze users' public posts to determine whether they align with the Chinese government's propaganda on three key issues, which may reflect their sources of information: (1) COVID-19: Omicron remains fatal; (2) the Russia-Ukraine war: Russia fights for justice; and (3) Japan's discharge of nuclear wastewater into the ocean: Japan is extremely selfish and irresponsible. Our findings show that: (i) users connected to information elites are significantly more likely to disagree with government propaganda than unconnected users; (ii) even users who initially agree with the propaganda are more likely to shift their beliefs over time if connected to information elites. These results suggest that information elites may significantly reshape their friends' beliefs by sharing updated, uncensored information. Our findings highlight the power of social networks in undermining the effectiveness of media censorship in the digital age, with information elites playing a pivotal role in disseminating uncensored content among citizens.

Keywords: Social Networks, Information Elites, Politically Sensitive Information, Media Censorship

JEL codes: D72, D83, L82, L86, L88, P36, Z13

* We express our gratitude to Sili Cao, Yuanyuan Chen, Jin Feng, Xiao Han, Zibin Huang, Shanjun Li, Lei Ning, Nathan Schiff, Hong Song, Hongliang Zhang, Lei Zhang, Zhijian Zhang, Kang Zhou, Ling Zhou, and Mohan Zhou for their insightful and helpful comments. All errors are our own.

[1] School of Economics, Shanghai University of Finance and Economics. E-mail: wang.xuebo@mail.shufe.edu.cn.

[2] School of Economics, Shanghai University of Finance and Economics. E-mail: yanghong@stu.sufe.edu.cn.

1. Introduction

Social networks are powerful conduits for the dissemination of information, opinions, and behaviors (Golub and Jackson, 2010). Many individuals rely on their friends as key sources of both information and opinions (Banerjee et al., 2013). Extensive research has demonstrated that social networks shape beliefs and behaviors through information flows, significantly impacting key economic and political decisions such as job seeking, migration, investment, technology adoption, and political participation (e.g., Granovetter, 1973; Beaman, 2012; Bakshy et al., 2012; Bond et al., 2012; Burchardi and Hassan, 2013; Jackson, Rogers, and Zenou, 2017; Bailey et al., 2018; Barwick et al., 2023; Blumenstock, Chi, and Tan, 2025).

In authoritarian regimes, however, the flow of information is often tightly controlled. Governments devote substantial resources to media censorship systems that suppress regime-threatening content and shape public opinion to sustain regime stability (MacKinnon, 2012; King, Pan, and Roberts, 2013; 2014; Qin, Strömberg, and Wu, 2017; Roberts, 2018; 2020; Chen and Yang, 2019).

This raises several critical questions: How do social network—powerful channels for information dissemination—interact with media censorship designed to prevent the spread of politically sensitive content? Do social networks retain their power to disseminate such information, thereby undermining censorship efforts? Addressing these questions requires a comprehensive empirical investigation, which remains limited due to challenges related to data availability and identification difficulties.

China provides an ideal context for studying this issue. As the country with the largest number of active internet users, it generates millions of posts each day on vibrant domestic social networking platforms, yielding rich data for analysis. At the same time, the Chinese government operates one of the world's most sophisticated media censorship systems, using blocking, filtering, and censorship mechanisms to prevent citizens from accessing and sharing politically sensitive information.

Yet in the information age, censorship cannot fully block alternative sources. Internet users residing abroad, as well as domestic residents using virtual private networks (VPNs), can bypass these controls and access uncensored content. We refer to these individuals as "information elites." They are likely to share uncensored content with their friends, potentially shaping their beliefs. In this context, China's media censorship creates a quasi-natural experiment: individuals connected to information elites face a radically different information environment than those without such connections. This setting allows us to empirically examine how politically sensitive information diffuses through social networks and the pivotal role of information elites in this process.

Using a unique dataset from Sina Weibo (the China's version of Twitter), we map large-scale users' online social networks based on friendship links. Within these networks, we identify key nodes—"information elites"—who can circumvent censorship to access uncensored content. Specifically, we classify users with IP addresses originating outside mainland China as "information elites," indicating that they either reside abroad or use VPNs to bypass censorship. We then group users into those connected to information elites (the treatment group) and those not connected (the control group), which allows us to examine whether information elites significantly influence their friends' beliefs on politically sensitive issues.

We employ natural language processing (NLP) techniques to analyze users' public posts and determine their stance on key politically sensitive issues. Our analysis focuses on three topics heavily promoted by the Chinese government: (1) COVID-19: Omicron remains fatal; (2) the Russia-Ukraine war: Russia fights for justice; (3) Japan's discharge of nuclear wastewater into the ocean: Japan is extremely selfish and irresponsible. Given the extensive censorship and propaganda surrounding these topics, users' beliefs on these issues can shed light on their potential sources of information. Those who unconditionally support the government's propaganda are likely to rely primarily on state-controlled outlets, while users who express disagreement with government narratives may have access to uncensored information that challenges the official stance.

Our baseline results show that users connected to information elites are more likely to disagree with government propaganda than users without such connections. However, endogeneity poses significant challenges for a causal interpretation. Social links are typically formed endogenously through homophily—individuals' tendency to connect with similar others (McPherson, Smith-Lovin, and Cook, 2001; Jackson and López-Pintado, 2013). Consequently, users connected to information elites may, by nature, have higher levels of education, broader access to information, and greater independence in judgment, making them less susceptible to government propaganda. As such, the observed correlation may reflect shared characteristics or beliefs rather than a causal effect of exposure to information elites.

To address this concern, we exploit plausible exogenous cross-regional variation in the intensity of China's 2022 lockdowns. Our analysis shows that lockdown intensity is primarily driven by local pandemic control pressures—measured by the number of confirmed Omicron cases per 10,000 people—and is uncorrelated with preexisting regional characteristics. Prior work suggests that strict lockdowns prompted residents to seek external information via VPNs (Chang et al., 2022). For instance, VPN usage surged among internet users in Shanghai during the city's 2022 lockdown.

Consequently, the lockdowns likely increased the probability that some individuals became information elites, generating an exogenous shock to their contacts' networks.

Leveraging the lockdowns as a natural experiment, we construct an individual-level instrument for social networks by interacting each user's network composition—the proportion of friends residing in different regions—with those regions' lockdown intensities. Higher values of this instrument indicate a greater fraction of friends in high-intensity regions and thus greater exposure to lockdown effects. The IV estimates show that greater exposure increases the likelihood of being connected to elites (first stage) and, in turn, increases disagreement with government propaganda (second stage).

We further examine how users' beliefs evolve over time. While prior beliefs often reflect inherent traits, belief changes are more likely driven by new information. For example, if a user initially agrees with government propaganda but later comes to disagree, this shift is likely attributable to newly acquired information from alternative sources. This interpretation aligns with existing literature, which highlights that individuals tend to update their beliefs when presented with new information (Eil and Rao, 2011; Chen and Yang, 2019; Levy, 2021; Bursztyn et al., 2023). We find that users connected to information elites are more likely to shift from agreement to disagreement with government propaganda over time. This pattern suggests that information elites share relevant, uncensored content with their friends and thereby persuade them to deviate from government narratives.

Additionally, we exploit variation in the countries of residence of overseas information elites to identify their impact on the beliefs of their friends in China. We hypothesize that elites in countries where the mainstream viewpoints contradict Chinese government propaganda are more likely to oppose these narratives and share dissenting information, thereby shaping their friends' beliefs. Conversely, elites in countries where the mainstream perspectives align with Chinese propaganda—or where local media largely ignore these issues—tend to support government narratives or lack pertinent information, resulting in minimal influence on their friends' beliefs. Consistent with this hypothesis, we find that elites in dissenting countries exert a significantly stronger persuasive effect on their friends' beliefs than elites in aligned or indifferent countries.

We also examine heterogeneous treatment effects by users' proximity to information elites. We find that users who are geographically closer to, or have a more intimate relationship with, information elites are more likely to be influenced by them, highlighting the critical role of social interactions in the dissemination of politically sensitive information through social networks.

We further study spillover effects. The influence of information elites may extend beyond their direct contacts to indirectly connected users. To identify such network spillovers, we compare the

beliefs of users at a social distance of 2 from information elites to those at a social distance greater than 2. Compared to the latter group, users at distance 2 are significantly more likely to disagree with government propaganda and are to shift from agreement to disagreement over time. These patterns suggest that treated users further disseminate uncensored information to their other friends, allowing politically sensitive information to spread widely through social networks and substantially reshape public opinion.

In summary, our findings show that social networks play a critical role in undermining the efficacy of media censorship by facilitating the spread of uncensored information among citizens. We do not, however, claim that censorship is entirely ineffective in the information age. Censorship continues to shape public opinion, particularly in the short term. For instance, in our data, more than 70 percent of users express agreement with Chinese government propaganda on the three selected issues.

While social networks may not immediately dismantle censorship systems, their long-term effects can be far more profound. Our results indicate that social networks can gradually and deeply reshape public beliefs, leading to lasting societal impacts. As more citizens gain access to uncensored information and begin to question official narratives, censorship systems and authoritarian regimes may face growing challenges. A single spark can start a prairie fire: the spread of uncensored information can gradually erode the grip of censorship and state control over public opinion.

This fragility of censorship in the information age is particularly striking. While stringent controls over information and thought may be feasible in nearly closed regimes such as North Korea, they are far harder to sustain in countries with extensive interactions with the outside world. In nations like China, which rely heavily on international trade, the cross-border flow of people, goods, and information can significantly reshape citizens' beliefs, despite the government's efforts to control the media. Consequently, managing information flows in a highly interconnected and open economy is increasingly difficult. While censorship systems are likely to eventually collapse, the precise timing of such a collapse remains uncertain.

Our work directly contributes to the extensive literature on social learning in networks, where individuals obtain information and update their beliefs through interactions with their friends (Ellison and Fudenberg, 1993; 1995; Foster and Rosenzweig, 1995; 2010; Murphy and Shleifer, 2004; Golub and Jackson, 2010; Acemoglu, Ozdaglar, and ParandehGheibi, 2010; Acemoglu and Ozdaglar, 2011; Banerjee et al., 2013; Pogorelskiy and Shum, 2019; Azzimonti, 2023). This body of work, primarily theoretical and experimental, highlights the importance of social networks in shaping beliefs, opinions, and decisions. Recent empirical studies using data from popular social networking platforms, such as

Facebook, have shown that individuals' beliefs and economic decisions are significantly influenced by their friends' experiences (e.g., Bailey et al., 2018; Bailey et al., 2024). Our study provides a comprehensive empirical analysis of this issue in the context of media censorship in China, demonstrating how individuals update their beliefs and deviate from government propaganda by accessing uncensored content shared by their friends. Our findings indicate that, in a tightly controlled information environment, the flow of information through social networks—especially from trusted sources like information elite friends—becomes a powerful mechanism for belief updating.

The literature on information diffusion within social networks underscores the pivotal role of influential network members—often referred to as opinion leaders or central figures—in spreading information (Lazarsfeld, Berelson, and Gaudet, 1944; Katz and Lazarsfeld, 1955; Rogers, 1962; Jackson and Yariv, 2011; Banerjee et al., 2013; 2019; Beaman et al., 2021). This study explores this concept within the context of media censorship, highlighting the critical role of information elites—individuals who can bypass censorship—in disseminating politically sensitive information and shaping their friends' beliefs. Unlike traditional opinion leaders, who influence public opinion through open expressions of their views, information elites under media censorship often operate in more subtle, covert ways, exerting influence privately within their networks.

Furthermore, this study complements the literature on how information consumption shapes individual beliefs (DellaVigna and Kaplan, 2007; Allcott and Gentzkow, 2017; Zhuravskaya, Petrova, and Enikolopov, 2020; Fong, Guo, and Rao, 2024). Recent studies have also explored how individuals update their beliefs in distorted information environments (e.g., Bai et al., 2015; Huang and Yeh, 2019; Chen and Yang, 2019; Enke, 2020). We extend this literature by highlighting how information shared by individuals' friends can significantly reshape their beliefs, particularly in biased information environments under media censorship. This aligns with the theoretical predictions of Bowen, Dmitriev, and Galperti (2023), who argue that in low-quality information environments, information shared by friends can significantly influence individuals' beliefs about the state of the world.

Finally, this study contributes to the literature on the efficacy of information and thought control in authoritarian regimes. Previous studies have demonstrated that state-driven indoctrination and mass persuasion were highly effective in shaping public beliefs during the mass media era (Alesina and Fuchs-Schündeln, 2007; Voigtländer and Voth, 2015; Adena et al., 2015; Cantoni et al., 2017; Ou and Xiong, 2021). In the digital age, some scholars argue that governments' attempts to control the internet are largely futile, as censorship can often be easily circumvented (Diamond, 2010). Others contend that governments still possess substantial power to manipulate the information environment (Morozov, 2011; MacKinnon, 2012; Roberts, 2018; 2020).

Despite these debates, direct empirical evidence remains limited. A notable exception is Chen and Yang's (2019) seminal study, which demonstrates that China's censorship system remains robust, making citizens more supportive of the regime. While they observe the social transmission of uncensored information among college roommates, the effects appear to be of small magnitude. They note that their study captures transmission only within the specific context of roommates, which may underestimate the overall social transmission of information. By leveraging a large-scale dataset of internet users, this study explores the spread of uncensored information within a much broader social network context, highlighting the critical role of information elites in facilitating this process. Our results suggest that, despite the Chinese government's extensive media control, politically sensitive information can still seep through these controls and significantly reshape public opinion.

2. Background

China has the largest number of internet users in the world, with over one billion netizens actively engaging on popular social media platforms such as WeChat, Sina Weibo, and Douyin. At the same time, the country's media landscape is tightly regulated and censored. The Chinese government enforces tight control over online spaces to prevent citizens from accessing and disseminating potentially regime-threatening information. In this section, we begin by briefly outlining the operational mechanisms of China's internet censorship system. We then discuss how information-seeking netizens manage to bypass these restrictions. Finally, we explore three key issues heavily promoted by the government and analyze how users' stances on these issues can reflect their potential sources of information.

2.1 Internet Censorship in China

China has invested substantial resources in building one of the most sophisticated internet censorship systems in the world. This system operates through three main mechanisms: content-control regulations, technical censorship, and proactive manipulation of online debates (OpenNet Initiative, 2012).[1]

First, at the policy level, the Chinese government enforces strict media regulations to control online information. A series of laws and regulations establish a legal framework for managing internet content, supported by penalty mechanisms to ensure compliance (Roberts, 2018). Tens of thousands of "internet police" (*wang jing*) patrol cyberspace to detect and remove "harmful content" that may

[1] Source: http://access.opennet.net/wp-content/uploads/2011/12/accesscontested-china.pdf.

threaten regime stability before it spreads widely. Domestic internet providers are required to implement self-censorship, promptly identifying and removing potential harmful content on their platforms. Additionally, the government mandates that users register for internet services using real identity information, which facilitates the monitoring of online activities.

Second, the Chinese government employs a wide range of techniques to restrict citizens' access to politically sensitive information, most notably through the "Great Firewall of China (GFW)." The GFW isolates Chinese netizens from the outside digital world by blocking specific foreign websites for users with IP addresses in mainland China. This significantly limits the information available to domestic users, leaving them with only content that aligns with state-approved narratives.

Finally, the Chinese government extends its control over online spaces to include active propaganda and information manipulation. It runs targeted propaganda campaigns on social media, disseminating messages that convey its beliefs or ideology to guide public opinion (Qin, Strömberg, and Wu, 2017; Roberts, 2020).[2] Additionally, the government has recruited a group of online commentators, known as the "50 Cent Party" (*wumao dang*), who actively post comments in favor of the state, strategically diverting public attention away from information that could threaten regime stability and manipulating public opinion (King, Pan, and Roberts, 2017).

Notably, while the Chinese government enforces the strictest media censorship in the world, this censorship is costly, and monitoring every internet user is impractical. As a result, the government focuses its efforts on controlling more influential users rather than ordinary individuals. This targeted repression allows the government to intimidate key figures without calling attention to censorship more broadly (Roberts, 2020). For example, during the 2022 Shanghai lockdown, a prominent Chinese infectious disease expert, Zhang Wenhong, was quickly silenced after suggesting that Omicron was mild and lockdowns were unnecessary. In contrast, many ordinary users were able to openly discuss the severity of Omicron and the rationale behind lockdown policies without facing significant restrictions.

As the literature suggests, this incomplete censorship—permitting some dissent, especially by ordinary individuals and on less sensitive issues[3]—can be attributed to several strategic considerations: (1) Fully monitoring and censoring millions of users is both challenging and economically costly, prompting the government to adopt the optimal censorship based on a cost-benefit analysis (Guriev and Treisman, 2019). (2) Tolerating some dissent on non-extremely sensitive issues helps maintain

[2] Qin, Strömberg, and Wu (2017) estimate that there are 600,000 government-affiliated accounts contributing 4% of all posts regarding political and economic issues on Sina Weibo.

[3] In China, certain issues such as the Tiananmen Square Massacre in 1989 are considered highly sensitive taboos and are strictly prohibited on social media. Any mention of them triggers automatic censorship mechanisms and leads to severe blocking.

public trust in the information environment and prevents backlash that might result from complete suppression (Shadmehr and Bernhardt, 2015; Hobbs and Roberts, 2018). (3) Strictly restricting content that could provoke collective action, while tolerating less threatening dissent, helps the government ensure regime stability (King, Pan, and Roberts, 2013; 2014; Qin, Strömberg, and Wu, 2017). (4) Allowing controlled dissent also enables the government to monitor public opinion, identify grievances, and address potential risks (Lorentzen, 2014; Qin, Strömberg, and Wu, 2017).

Taken together, the Chinese government makes significant efforts to exert ideological control over its citizens by restricting access to uncensored, regime-threatening information. However, such censorship faces growing challenges in the information age. The following section describes how information-seeking elites bypass censorship to access uncensored content.

2.2 Information Elites under Media Censorship

Two types of internet users are able to circumvent the GFW to access uncensored information. First, many active users on Chinese social media platforms actually reside outside mainland China (i.e., overseas users) and are therefore unaffected by the GFW. Second, domestic users can bypass the GFW by using Virtual Private Networks (VPNs), which utilize encryption technology to mask their real IP addresses and replace them with those from countries with unrestricted internet access. Most regular VPN users are tech-savvy individuals with high information demands, who are both aware of censorship and capable of bypassing it (Roberts, 2020). Approximately 30% of Chinese internet users were using VPNs in 2017 (Global World Index, Q2 2017), reflecting a strong demand for uncensored internet access among Chinese internet users.

Given that both overseas users and VPN users can access external, uncensored information, we designate them as "information elites" under media censorship. Many of these users actively engage in public discussions and disseminate information on Chinese social media platforms such as Sina Weibo. Despite the strict media controls, they retain the potential to effectively spread the uncensored information among the population.

As discussed above, the strategic and incomplete nature of censorship enables information elites to discuss and disseminate politically sensitive content on social media platforms despite restrictions. More importantly, censorship cannot prevent information elites from privately sharing uncensored information with their friends. These elites can communicate with friends through face-to-face conversations, phone calls, WeChat, or other private messaging apps—channels largely beyond the reach of censorship. This highlights the power of social networks to disseminate politically sensitive information among the public, even under strict media censorship.

Therefore, China's internet censorship creates a divided information environment. Users who primarily rely on government sources are mostly exposed to official narratives, while those connected to information elites have access to uncensored content, potentially leading their beliefs to diverge from the official narrative.

2.3 Three Politically Sensitive Issues

This section provides a brief overview of the three key issues central to our study.

Threat Posed by Omicron (a variant of COVID-19). The Omicron variant began spreading rapidly in late 2021 and became the dominant strain globally by early 2022. Compared to earlier strains of COVID-19, Omicron exhibited higher transmissibility but much lower fatality rates. In response, many countries shifted toward strategies aimed at coexisting with the virus, gradually relaxing pandemic control measures. However, the Chinese government maintained its strict control policies. Starting in late March 2022, Shanghai, China's largest city, endured a months-long lockdown. During this period, as reported and criticized by foreign media, the Chinese government consistently exaggerated the threat posed by Omicron to justify the continuation of its stringent measures.

In this context, statements contradicting the official narrative were suppressed, particularly those from influential figures. While ordinary users could still discuss the lethality of Omicron online, such discussions had limited impact, as they were drowned out by the overwhelming flood of government propaganda. Given that the lethality of Omicron directly affected citizens' health and well-being, and the strict lockdown measures severely disrupted daily life, many domestic citizens began to question government narratives and seek alternative sources of information.

Therefore, users' beliefs about the lethality of Omicron often reflect their sources of information. Those who view Omicron as highly fatal and strongly support government propaganda are likely loyal government followers who rely primarily on state-controlled media. In contrast, users who disagree with government narratives—seeing Omicron as mild and citing countries that have successfully coexisted with the virus—are likely exposed to alternative information sources and are more aware of events outside China.

Russia-Ukraine War. On February 24, 2022, Russian President Vladimir Putin announced a "special military action" against Ukraine, which quickly escalated into a full-scale war. The international community widely condemned Russia as the aggressor responsible for war crimes and responded with a range of sanctions. While the Chinese government officially claims to maintain a neutral stance, its domestic propaganda reveals sympathy for, and even support of, Russia's actions.

Chinese official narratives avoid condemning Russia for initiating the war, instead using neutral terms like "conflict" or "action" rather than "war" or "invasion," which could portray Russia negatively. Additionally, the Chinese government has accused the North Atlantic Treaty Organization (NATO) of employing Cold War strategies to contain Russia, framing Russia's actions as a defense of its sovereignty. Through this propaganda, the Chinese government has fostered the belief that Russia is fighting for justice, thereby cultivating public support for a strong and friendly China-Russia relationship.

Japan's Fukushima Nuclear Wastewater Discharge Plan. Since August 24, 2023, Japan has begun systematically discharging treated nuclear wastewater from the Fukushima power plant into the ocean. The wastewater, processed through the Advanced Liquid Processing System (ALPS) to remove most radioactive substances except tritium, is diluted to meet international safety standards before being released. Discharging treated radioactive wastewater is a routine practice in the nuclear industry and aligns with international norms. Japan's plan, supervised by the International Atomic Energy Agency (IAEA), has been confirmed to comply with international safety standards and is scientifically considered safe for both humans and the environment.

However, the plan sparked intense criticism from the Chinese government. Following Japan's announcement, Chinese officials accused Japan of acting selfishly and irresponsibly, claiming it was shifting the risk of nuclear contamination onto the entire world. Amplified by official propaganda, domestic media extensively reported the potentially catastrophic consequences of Japan's wastewater discharge. As a result, individuals' perceptions of this issue are largely shaped by the sources of information available to them.

As noted earlier, online dissent on the three issues has been allowed due to the strategic and incomplete censorship adopted by the Chinese government. This incomplete censorship in China enables internet users—particularly ordinary individuals—to express opinions on topics that are not highly sensitive, such as those selected for this study, thereby providing an opportunity to identify their beliefs on these issues.

3. Data

This section provides details on the data used for this study, which is sourced from Sina Weibo, a major Chinese social media platform with over 605 million monthly active users (Sina Weibo, 2023). We describe how we determine users' beliefs on selected politically sensitive issues based on their

public posts, measure users' online social networks, and identify key figures — referred to as "information elites"—within these networks.

3.1 Outcome Variable: Beliefs on Politically Sensitive Issues

We collect and analyze users' public posts to determine whether they agree with the Chinese government's propaganda on three politically sensitive issues: (1) on COVID-19: Omicron remains deadly; (2) on the Russia-Ukraine War: Russia is fighting for justice; and (3) on Japan's discharge of nuclear wastewater: Japan is extremely selfish and irresponsible.

To determine users' beliefs, we employ advanced natural language processing (NLP) techniques to analyze their posts. This process is carried out in four main stages: corpus collection, human coding, model training, and automated prediction. A brief explanation of each stage is provided below. For more detailed information, please refer to Appendix A.1.

Corpus Collection. We devise a crawler system to collect Weibo posts and comments related to the three selected issues. For each issue, we define a set of relevant keywords and a specific time window. The system then gathers all posts containing these keywords within the specified time frame, along with the corresponding comments. A complete list of keywords and time windows for each issue is provided in Table A1. This procedure enables us to collect nearly 34 million posts and comments from millions of users, forming the corpus for subsequent analysis.

Human Coding. In this stage, we manually annotate a large sample of text drawn randomly from the corpus. To ensure consistency and accuracy, we establish clear and detailed annotation criteria for each issue, as outlined in Appendix A.1.2.1, A.1.2.2, and A.1.2.3. These criteria guide annotators in categorizing posts into three belief types: *Disagree* (indicating the user disagrees with government propaganda), *Agree* (indicating the user agrees with government propaganda), or *Unidentifiable* (indicating insufficient information to determine the user's belief). For example, a post stating "Omicron is mild, not as scary as the government claims" is labeled *Disagree*, while one saying "Omicron is deadly, I'm afraid to go outside" is labeled *Agree*. A neutral post, such as "When will the pandemic end?" is classified as *Unidentifiable*.

Model Training and Automatic Prediction. We use Bidirectional Encoder Representations from Transformers (BERT), a model renowned for its strong performance in NLP tasks (Devlin et al., 2018), to automatically classify the universe of posts and comments within our dataset. Specifically, we train the BERT on our manually annotated texts and then apply it to the full corpus. The trained model preforms well on unseen data, achieving an accuracy of approximately 0.75 and an Area Under

the Curve (AUC) of 0.90.[4] Ultimately, the model identified more than 3 million posts made by over 800,000 users as either *Agree* or *Disagree*, with the remaining posts classified as *Unidentifiable* and treated as noise. The users whose beliefs were identified constitute the valid sample for further analysis.

Outcomes of Interest: Beliefs on the Three Issues. Based on the model's predictions, we construct two key outcome variables: (i) whether a user agrees with government propaganda on a given issue (labeled as $Belief_i$), and (ii) whether a user shifts from agreeing to disagreeing with government propaganda (labeled as $Belief\ Shift_i$).[5] In our data, some users posted multiple times about the same issue, and some exhibited changes in beliefs over time. We encode these variables as follows: if all of user i's posts on an issue are predicted as *Disagree*, we code $Belief_i$ as 1 (indicating disagreement with government propaganda). Conversely, if all posts are predicted as *Agree*, we code $Belief_i$ as 0 (indicating agreement).[6] If user i's belief shifts from *Agree* to *Disagree*, we code $Belief\ Shift_i$ as 1; otherwise, it is coded as 0. Specifically, if a user has published a total of N posts with identifiable beliefs, and if the labels of the first M (where M<N) posts are all *Agree* while the labels of all subsequent posts are *Disagree*, we define this as a shift in the user's belief.

Users' Prior Beliefs. We define a user's prior belief as the stance expressed in their first post that clearly indicates agreement or disagreement with government propaganda. To determine this, we first sort the user's posts in ascending order of posting time. Then, we apply our trained model to determine the belief expressed in each post while filtering out noise. The belief in the first valid post is then identified as the user's prior belief.

One concern is that our sample may include irrelevant users, such as online commentators employed by the Chinese government who actively post comments in favor of the state. Indeed, as mentioned earlier, the government has recruited a group known as the "50 Cent Party", whose members are tasked with posting pro-government comments. However, King, Pan, and Roberts (2017) provide compelling evidence that members of the 50c party rarely engage in debates in defending the regime, its leaders, or its policies. Instead, their posts primarily serve to promote national pride and discuss non-controversial topics in an attempt to divert attention from sensitive issues. As a result, these posts are unlikely to be captured by our crawling system, ensuring that our analysis reflects the genuine opinions of ordinary people rather than those of government-affiliated commentators.

[4] For further details on the training process and model performance, please refer to Appendix A.1.3.

[5] Under media censorship, changes in citizens' beliefs tend to be unidirectional, moving from agreeing to disagreeing with government propaganda. Initially, citizens may unconditionally trust government messages. However, once they acquire external information that contradicts the government narrative, they may begin to question and diverge from the official propaganda. Conversely, if citizens are already aware of the biases in government propaganda and do not trust it from the outset, they are unlikely to shift their beliefs and start supporting the government narrative. Since our goal is to identify the impact of acquiring uncensored information on users' belief updating, we focus on the first type of belief shift—from agreement to disagreement.

[6] When determining users' beliefs, we exclude samples that experienced belief shifts.

3.2 Measuring Users' Online Social Networks

Consistent with established practices in the literature (e.g., Bailey et al., 2018), we define two users as online friends if they follow each other. Mutual following, which requires consent from both parties, serves as a reliable indicator of social connections or friendships. Our sample includes over 800,000 Weibo users who engage in discussions on the three selected issues and have identifiable beliefs. From May to September 2023, we collected anonymized snapshots of these users' friendship links, allowing us to map their online social networks.[7]

Note that users' online social networks may not perfectly capture their real-world networks. For example, if users' real-world friends do not use Weibo, these friends will not appear in users' online networks. However, existing literature suggests that online social networks, as measured by popular social media platforms like Facebook, generally provide a reliable representation of individuals' real-world friendship networks (e.g., Bailey et al., 2018; Bailey et al., 2022). Therefore, while users' online social networks may not encompass all their real-word friends, as long as the majority of their Weibo-using friends are included in these networks, they still serve as a valid indicator of users' social networks.

For the users in our sample, we have access to both their online social network information and demographic details, including age, gender, education, and residence. Additionally, we capture users' network attributes, including the number of friends, followers, and followings, as well as account-specific attributes, including VIP status, account registration age, and whether they use an iPhone device. Furthermore, we assess users' areas of interest or information preferences based on the types of content they follow. Specifically, we categorize the content into five types: Entertainment and Culture, Lifestyle and Consumption, News and Current Affairs, Education and Knowledge, and Public Services and Social Responsibility. Detailed descriptions of each content type are provided in Appendix A.2.

3.3 Connection to Information Elites

We now determine whether users are connected to information elites. Among the over 800,000 users in our sample, who collectively have more than 11 million online friends, we aim to identify information elites within this network.

Sina Weibo's Move. In practice, we identify information elites by tracking users' IP addresses. Sina Weibo' move to disclose user IP addresses provides an opportunity for this analysis. Starting from April 28, 2022, Sina Weibo began displaying user locations based on their IP addresses when

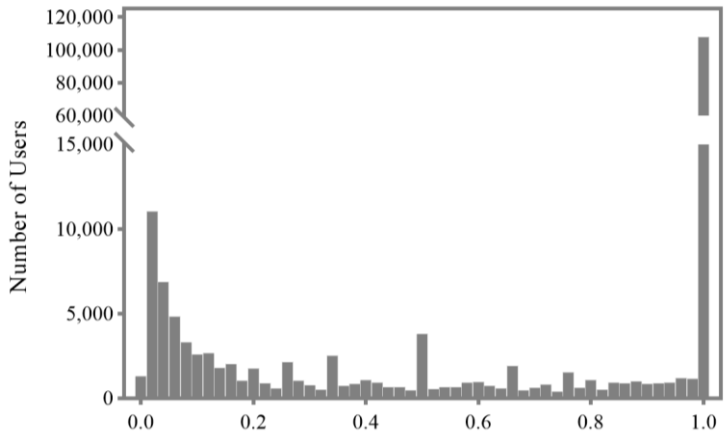
[7] The network snapshot is taken after the periods of interest for two of the issues, raising concerns that the network may have changed in the interim. We address this concern in Section S5 of the Supplementary Information.

posting content. For domestic users, the platform shows the province they reside in, while for overseas users, it shows the country or region of residence. Although the feature cannot be disabled, users can bypass it by using VPN tools to alter their IP addresses.

To identify information elites, we track the IP addresses of users' Weibo posts from May 1, 2022, to December 31, 2022. We focus on this period because the strict COVID-19 travel restrictions in China significantly limited cross-border travel, which allows us to more accurately identify information elites. For instance, if a user's IP address alternates between mainland China and overseas locations, it is more likely to be due to VPN usage rather than international travel. Based on this observation, users whose IP addresses are consistently located outside mainland China are classified as overseas information elites. Users whose IP addresses fluctuate between mainland China and foreign locations are classified as VPN information elites. Lastly, users whose IP addresses are consistently within mainland China are considered ordinary domestic users.[8]

Information Elite Set. In total, we collect over 1.2 billion original Weibo posts from our sample users and their friends, encompassing over 12 million users. For each post, we record details such as the posting time, IP address, text content, as well as associated comments and likes. By analyzing the IP addresses of these posts, we identify over 220,000 information elites. Of these, 58% are classified as overseas information elites, while 42% are categorized as VPN information elites.

Figure 1: Distribution of Information Elites by the Probability of IP Addresses Originating Outside Mainland China.



Notes. This figure illustrates the distribution of information elites based on the probability that their IP addresses are located outside mainland China. The x-axis represents the probability that users' posts are associated with IP addresses outside mainland China, while the y-axis represents the number of users within each probability range.

Figure 1 illustrates the distribution of information elites based on the probability that their IP addresses are located outside mainland China. An IP address with a probability of 1 indicates an

[8] This classification strategy may not be entirely accurate. A detailed discussion of this issue can be found in Section 4.7.

overseas elite, while lower probabilities correspond to VPN elites. Figure A3 in the Appendix shows the global distribution of overseas information elites, and Figure A4 presents the distribution of VPN elites across provinces in China.

Connection to Information Elites. We construct two indicators to measure a user's connection to information elites: (1) whether a user is connected to at least one information elite (labeled *Connected to Elites*), and (2) the proportion of information elites among a user's friends (labeled *Connection Intensity*). Users who are connected to at least one information elite are included in the treatment group, while those not connected to any information elite are placed in the control group. The second indicator measures the intensity of the treatment, based on the assumption that users with more connections to information elites are more likely to be influenced by them.

3.4 Summary Statistics

Table 1 presents the summary statistics for all sample users, categorized by the politically sensitive issues they discussed. Specifically, Columns (1)–(2), (3)–(4), and (5)–(6) show users discussing the lethality of Omicron, the Russia-Ukraine war, and Japan's discharge plan, respectively. Panel A displays descriptive statistics for the outcome variables of interest. Among our sample users, only a small proportion disagree with government propaganda: 23% of users perceive Omicron as mild, 19% view Russia as the aggressor, and only 2.6% believe Japan's discharge plan is scientifically safe. During the study period, a minority of users shifted their beliefs, moving from agreeing to disagreeing with government propaganda: 4.6% for the lethality of Omicron, 5.5% for the Russia-Ukraine war, and 0.9% for the Japan's discharge plan.

Panel B provides descriptive statistics for the treatment variables. Approximately 10% of users are connected to information elites, and on average, information elites constitute nearly 1% of a user's friends. Panel C displays statistics for the (provincial-level) lockdown intensity and the (individual-level) instrument. Lockdown intensity is measure by the reduction in human mobility during the lockdown period relative to normal times. The mean and standard deviation of lockdown intensity are 0.362 and 0.126, respectively, indicating that the lockdowns reduce human mobility by an average of 36.2 percent, with significant variations across regions. The instrument reflects the exposure of individual users' social networks to the lockdowns, with a mean ranging from 0.15 to 0.20 and similar standard deviations.

Panel D summarizes user characteristics. On average, users have around 10 online friends. We classify users based on the number of followers they have to capture their popularity, with those above the median level labeled as *High Followers*. Similarly, we divide users based on the number of users they follow to reflect their demand for information and networks, with those above the median level

labeled as *High Following*. Users' preference for information are measured based on the content they follow: 70% are interested in Entertainment and Culture, 50% in Lifestyle and Consumption, 60% in News and Current Affairs, 65% in Education and Knowledge, and 10% in Public Services and Social Responsibility. Overall, users exhibit diverse information demand. Panel D also reports users' account characteristics: users have been registered for an average of 8 years, hold a VIP level above 1, and approximately 30% access Weibo through iPhone devices.

Table 1: Summary Statistics

	Threat Posed by Omicron		**Russia-Ukraine War**		**Japan's Discharge Plan**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
Panel A: Outcome Variables						
Belief	0.230	0.421	0.194	0.396	0.026	0.160
Belief Shift	0.046	0.211	0.055	0.227	0.009	0.096
Panel B: Treatment Variables						
Connected to Elites	0.126	0.332	0.081	0.272	0.056	0.231
Connection Intensity	0.014	0.071	0.012	0.072	0.007	0.054
Connected to Overseas Elites	0.085	0.278	0.055	0.227	0.035	0.185
Connected to VPN Elites	0.069	0.253	0.038	0.192	0.027	0.163
Panel C: Instrumental Variable						
Lockdown Intensity	(N = 31, Mean = 0.362, S.D. = 0.126)					
Instrument	0.206	0.219	0.163	0.211	0.155	0.202
Panel D: User Characteristics						
Age	30.840	9.549	33.392	10.525	31.144	10.026
Male	0.454	0.498	0.676	0.468	0.474	0.499
Reported School	0.180	0.384	0.176	0.380	0.160	0.367
Number of Friends	12.397	24.406	8.293	20.744	11.589	28.086
High Followers	0.499	0.500	0.499	0.500	0.498	0.500
High Following	0.499	0.500	0.499	0.500	0.500	0.500
Account Age (years)	8.384	3.561	8.198	3.607	7.431	4.004
VIP Level	1.519	2.246	1.162	2.022	1.290	1.994
Has iPhone	0.340	0.474	0.242	0.428	0.250	0.433
Entertainment	0.694	0.461	0.634	0.482	0.731	0.443
Lifestyle	0.504	0.500	0.501	0.500	0.505	0.500
News	0.531	0.499	0.647	0.478	0.591	0.492
Knowledge	0.614	0.487	0.672	0.470	0.650	0.477
Responsibility	0.104	0.305	0.078	0.268	0.102	0.302
Number of Observations	403,966		299,382		121,136	

Notably, our Weibo user sample is not fully representative, as it excludes individuals who do not access the internet, use Weibo, or post about the three issues we focus on. However, this limitation is unlikely to undermine the validity of our analysis. If information elites within our sample are able to spread politically sensitive information through social networks and reshape their friends' beliefs, there is no reason to expect that this phenomenon would not occur more broadly.

4. Results

4.1 Baseline Results

Our baseline specification examines the relationship between users' connections to information elites and their beliefs on politically sensitive issues. Specifically, we estimate the following equation:

$$Belief_i = \alpha + \beta\ Connection_i + \gamma X_i + \varepsilon_i, \quad (1)$$

where $Belief_i$ denotes user i's belief on the three selected issues, taking a value of 1 if the user disagrees with government propaganda and 0 otherwise. $Connection_i$ represents a dummy variable, taking a value of 1 if user i is connected to information elites and 0 otherwise. Alternatively, it can be a continuous variable representing the proportion of information elites among user i's friends (i.e., *Connection Intensity*). X_i is a set of control variables, and ε_i is the error term. β is the parameter of interest, capturing the correlation between users' connection to information elites and their beliefs on government propaganda.

We estimate Equation (1) using ordinary least squares (OLS) and present the results in Table 2. Panels A–C report the results for the issues of Omicron, the Russia-Ukraine war, and Japan's discharge plan, respectively. Column (1) presents the results without any control variables, while Columns (2) to (6) gradually include control variables. For all three issues, the coefficients are significantly positive, indicating that users connected to information elites are more likely to disagree with government propaganda.

After including all control variables, Column (6) of Panel A shows that for the issue of Omicron, the coefficient is 0.025, indicating that users connected to information elites are 2.5% more likely to perceive Omicron as mild compared to those unconnected. This represents an approximately 11% increase relative to the sample mean (0.025/0.23). For the Russia-Ukraine war issue, users connected to information elites are 0.022 more likely to condemn Russia's actions, about an 11% increase relative to the sample mean (0.022/0.194). Regarding Japan's discharge plan, users connected to information elites are 0.013 more likely to consider the plan scientifically safe, a notable 50% increase relative to the sample mean (0.013/0.026).

Table 2: The Effects of Connections to Information Elites on User Beliefs Regarding Politically Sensitive Issues

Dependent Variable:	*Belief*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Threat Posed by Omicron							
Connected to Elites	0.024***	0.024***	0.035***	0.032***	0.030***	0.025***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Observations	403966	403966	403966	403966	403966	403966	331697
R-Square	0.000	0.000	0.001	0.006	0.009	0.012	0.013
Mean Y	0.230	0.230	0.230	0.230	0.230	0.230	0.231
Panel B: Russia-Ukraine War							
Connected to Elites	0.035***	0.040***	0.023***	0.024***	0.023***	0.022***	0.016***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	299382	299382	299382	299382	299382	299382	165538
R-Square	0.001	0.002	0.003	0.009	0.011	0.011	0.021
Mean Y	0.194	0.194	0.194	0.194	0.194	0.194	0.198
Panel C: Japan's Discharge Plan							
Connected to Elites	0.009***	0.012***	0.015***	0.014***	0.014***	0.013***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	121136	121136	121136	121136	121136	121136	57760
R-Square	0.000	0.005	0.005	0.006	0.006	0.008	0.011
Mean Y	0.026	0.026	0.026	0.026	0.026	0.026	0.025
Demographics	No	Yes	Yes	Yes	Yes	Yes	Yes
Network Characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Account Characteristics	No	No	No	Yes	Yes	Yes	Yes
Information Preference	No	No	No	No	Yes	Yes	Yes
Province FE	No	No	No	No	No	Yes	Yes
CEM Sample	No	No	No	No	No	No	Yes

Notes. This table reports the effect of connections to information elites on user beliefs regarding politically sensitive issues, specifically presenting the estimated coefficients (β) from Equation (1). Panels A–C display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. Columns (1)–(6) include the full sample, while Column (7) focuses on the matched samples obtained using the CEM procedure. The control variables include (i) demographic characteristics (gender and whether the user lists their school on Sina Weibo); (ii) network characteristics (number of friends, and high follower/followee indicators); (iii) account characteristics (Weibo account age, VIP level, and iPhone usage); (iv) information preference (five dummy variables indicating whether the user is interested in specific types of content: Entertainment, Lifestyle, News, Knowledge, and Responsibility); and (v) provincial fixed effects. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

These estimates are most pronounced for the issue of Japan's discharge plan. This may be because forming an objective belief on this issue requires a higher information threshold, with the acquisition

of external knowledge playing a more critical role in shaping beliefs. Specifically, the safety of Japan's wastewater discharge is fundamentally a scientific question, and understanding it requires certain scientific knowledge and background. Due to the Chinese government's extensive propaganda on this issue, the domestic public has limited access to relevant scientific information. As a result, when information elites share relevant scientific knowledge through social networks, it can significantly influence their friends' beliefs on the issue.

In the Supplementary Information (SI), we replace the dummy variable "connection to elites" with "connection intensity" (the proportion of information elites among a user's friends) and re-estimate Equation (1); the results are reported in Table S1. The findings remain consistent. We also examine heterogeneous effects by connection type (overseas elites and VPN elites) and find that users in both subgroups are significantly more likely to disagree with government propaganda. For additional results and discussion, see Section S1 in the SI.

Given the potential endogeneity of connections to information elites, our baseline results provide correlational rather than causal evidence. In the subsequent sections, we address this endogeneity concern by employing alternative strategies.

4.2 IV Estimation Results

A major identification concern is that social networks are formed endogenously, with individuals tending to connect with others who share similar characteristics. Therefore, the coefficients β in Equation (1) may reflect shared beliefs between users and their information elite friends, rather than a causal effect of information elites on their friends.

To address this endogeneity issue, we exploit plausibly exogenous cross-regional variation in the intensity of China's 2022 lockdowns. Prior work indicates that strict lockdowns drove residents to seek external information via VPNs (Chang et al., 2022). For example, during the Shanghai lockdown in April 2022, there was a notable 41% increase in visits to Twitter from Shanghai as users circumvented internet controls.[9] Given that the lockdown spurred efforts to access external information via VPNs and related tools, it likely increased the probability that some individuals became information elites, creating an exogenous shock to their contacts' networks.

Following prior work (Kraemer et al., 2020; Chang et al., 2022), we measure regional lockdown intensity in the first half of 2022—when restrictions peaked—by the reduction in human mobility relative to the first half of 2023, when lockdowns had ended and mobility returned to normal. Mobility data are publicly available from Baidu Migration Big Data (https://qianxi.baidu.com/#/), which uses

[9] For more information, please visit the following website: https://www.wired.com/story/shanghai-lockdown-china-censorship/. Section S2 in the SI discusses how the lockdowns prompted people to circumvent censorship to access uncensored content.

Baidu Maps location data to track real-time migration, including the daily national mobility index and daily provincial inflow/outflow indices. For data details and validation of this lockdown measure, see Section S3 in the SI.

We construct an individual-level instrument for social networks by interacting each user's network composition—the share of friends residing in different regions—with those regions' lockdown intensities. This instrument captures the exposure of users' social networks to the lockdowns. For user i, the exposure instrument is defined as the friendship-weighted average of provincial lockdown intensity:

$$Exposure_i = \sum_p s_{i,p} \times Lockdown\ intensity_p, \quad (2)$$

where $s_{i,p}$ is the fraction of user i's Weibo friends who reside in province p, and $Lockdown\ intensity_p$ is the lockdown intensity of province p.[10]

The first-stage and second-stage regressions of the IV estimation are specified as follows:

$$Connection_i = \beta^{FS} Exposure_i + \gamma_1 X_i + \varepsilon_{1i}, \quad (3)$$

$$Belief_i = \beta^{IV} \widehat{Connection}_i + \gamma_2 X_i + \varepsilon_{2i}. \quad (4)$$

A potential concern with this IV is that provincial lockdown intensity may be endogenous. For instance, stricter lockdowns could occur in more developed regions such as Beijing and Shanghai, where VPN use is more common. Users connected to residents in these regions would then be mechanically more likely to link to information elites, yielding a significant first-stage coefficient. In this case, β^{FS} could reflect a mechanical correlation between the instrument and the treatment rather than the causal effect of lockdowns on users' social networks.

To address this concern, we examine the factors influencing the cross-regional variations in lockdown intensity. Lockdown measures are primarily driven by local pandemic control pressures. Typically, when confirmed Omicron cases rise significantly in a region, local government are compelled to intensify their pandemic control measures. The number of confirmed Omicron cases is difficult to predict; it can be notably high in more developed regions such as Beijing and Shanghai, where interactions with the outside are frequent and migration flows are substantial. Additionally, it can also be elevated in less developed regions, like Northeastern provinces, where high human mobility occurs due to many residents migrating to Southern China for work and frequently returning to their hometowns.

[10] We exploit only cross-provincial variation in lockdown intensity because users' IP addresses contain only provincial information (without prefectural or county identifiers).

Figure S2 in the SI illustrates the lockdown intensities for all provinces or municipalities in the first half of 2022. The highest-intensity regions include developed areas such as Shanghai and Beijing, as well as less developed regions such as Liaoning, Heilongjiang, Jilin, and Inter Mongolia. In contrast, the lowest-intensity regions include developed regions such as Fujian and Guangdong, alongside less developed areas such as Tibet and Guizhou. Thus, lockdown intensity varies substantially across regions, with patterns that appear plausibly random.

For validation, we regress lockdown intensity on key provincial characteristics (see Column (1) of Table S3 in the SI). All coefficients are small and statistically insignificant, indicating no systematic correlation with provincial characteristics. As expected, when we include the number of confirmed Omicron cases per 10,000 people, its coefficient is significantly positive, while the others remain insignificant (see Column (2) of Table S3); the case count is also uncorrelated with other provincial characteristics (see Column (3) of Table S3). Together, these results suggest that lockdown intensity is driven primarily by local pandemic control pressures rather than by regional attributes such as economic development.

In summary, if regional lockdowns constitute an exogenous shock, they provide a natural experiment for analyzing the impact of users' social networks. Users in high-intensity regions are more likely to begin using VPNs to access uncensored content and become information elites. Consequently, users connected to those regions are more likely to link to information elites, thereby increasing their exposure to elite influence and their likelihood of disagreeing with government propaganda. β^{FS} and β^{IV} in Equations (3) and (4) capture these effects.

Table 3 reports the IV estimates for the three issues, with Panel B presenting the first-stage results. For the Omicron issue, the coefficient on the instrument is 0.236, indicating that a 0.01 increase in the instrument (a 4.8% increase relative to its sample mean) raises the probability of being connected to information elites by 0.00236, or approximately 1.03% relative to the sample mean—a non-negligible effect. The results are similar for the other two issues. Across all specifications, the Kleibergen-Paap F-statistics exceed 1,000, far above conventional thresholds for weak instrument concerns. Panel A presents the second-stage results, which consistently show a positive effect of connection to information elites on uses' belief.

A potential concern is that, although lockdowns are plausibly exogenous at the provincial level, the cross-provincial variation may not be large enough to ensure that they constitute an exogenous shock at the individual level. Indeed, Table S6 in the SI shows that the instrument is significantly correlated with some user characteristics, raising endogeneity concerns. In particular, if users with high values of the instrument happened to be better-educated individuals who are not only more likely

to be connected to information elites but also more adept at accessing uncensored information, this would confound the causal interpretation of our results.

Table 3: IV Estimates of the Effect of Connections to Information Elites on User Beliefs Regarding Politically Sensitive Issues

Dependent Variable:	*Belief*		
	Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
	(1)	(2)	(3)
Panel A. Second stage			
Connected to Elites	0.067***	0.077***	0.083***
	(0.014)	(0.020)	(0.016)
Observations	403966	299382	121136
Mean Y	0.230	0.194	0.026
Baseline Controls	Yes	Yes	Yes
Dependent Variable:	*Connected to Elites*		
Panel B. First stage			
Exposure	0.236***	0.191***	0.175***
	(0.002)	(0.003)	(0.004)
Observations	403966	299382	121136
R-Square	0.238	0.203	0.111
Mean Y	0.126	0.081	0.056
Kleibergen-Paap F-statistic	9574.451	4866.90	1965.471
Baseline Controls	Yes	Yes	Yes

Notes. This table reports IV estimates of the effect of connections to information elites on user beliefs regarding three politically sensitive issues. Panel B presents the first-stage results, reporting the estimated coefficients (β^{FS}) from Equation (3). Panel A presents the second-stage results, reporting the estimated coefficients (β^{IV}) from Equation (4). Columns (1)–(3) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

To address this concern, we use matching to balance observable characteristics between users with high and low exposures to the lockdowns. In the matched sample, the exposure instrument is uncorrelated with user observable characteristics. Conditional on observables, users who differ only in the extent of their lockdown exposure can therefore be viewed as having been quasi-randomly assigned to different exposure levels, plausibly creating a natural experiment. In this setting, if high-exposure users are more likely to be connected to information elites and more likely to disagree with government propaganda, this pattern would provide strong evidence of a treatment effect.

Table 4: IV Estimates of the Effect of Connections to Information Elites on User Beliefs Regarding Politically Sensitive Issues (Matched Sample)

Dependent Variable:	*Belief*		
	Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
	(1)	(2)	(3)
Panel A. Second stage			
Connected to Elites	0.057***	0.055**	0.078***
	(0.020)	(0.027)	(0.023)
Observations	155974	110880	41574
Mean Y	0.236	0.199	0.025
Dependent Variable:	*Belief Shift*		
Panel B. Second stage			
Connected to Elites	0.023**	0.035**	-0.001
	(0.010)	(0.016)	(0.014)
Observations	155974	110880	41574
Mean Y	0.045	0.054	0.009
Dependent Variable:	*Connected to Elites*		
Panel C. First stage			
Exposure	0.293***	0.240***	0.193***
	(0.004)	(0.004)	(0.006)
Observations	155974	110880	41574
R-Square	0.033	0.022	0.020
Mean Y	0.219	0.171	0.120
Kleibergen-Paap F-statistic	5116.729	2701.140	933.779

Notes. This table reports IV estimates of the effect of connections to information elites on user beliefs regarding three politically sensitive issues, using the matched samples obtained from the CEM procedure. Panel C presents the first-stage results, reporting the estimated coefficients (β^{FS}) from Equation (3). Panel A presents the second-stage results, reporting the estimated coefficients (β^{IV}) from Equation (4). Panel B estimates Equation (4) but replace the dependent variable with *Belief Shift*, with estimated coefficients (β^{IV}) presented in the table. Columns (1)–(3) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include province fixed effects. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

In practice, we implement Coarsened Exact Matching (CEM) (Iacus, King, and Porro, 2008; 2012) to pair each user with high exposure to lockdowns—defined as having an instrument value in the top quintile—with a user from the remaining sample on a set of covariates.[11] Unlike propensity-score methods, CEM is nonparametric and well-suited to our setting, where most covariates are categorical. This procedure yields a sample in which high- and low-exposure users exhibit no systematic

[11] Conceptually, we match each user with high lockdown exposure to a user with low lockdown exposure, and our results are not sensitive to the precise cutoff. We avoid very high cutoffs (e.g., the top 80 percent), however, because suitable matches become difficult to find in the remaining 20 percent of the sample.

differences in observable characteristics (see Appendix B for details). We then re-estimate Equations (3) and (4) using this matched sample and present the results in Panel C (first stage) and Panel A (second stage) of Table 4. The results remain similar to those in Table 3. Because the means of the dependent variables are nearly identical across the full and matched samples, the estimates from the two samples are directly comparable.

However, concerns remain even for this matched sample: high-exposure users may differ from low-exposure users in unobservable characteristics that affect both their connections to elites and their beliefs on politically sensitive issues. We address this concern from two angles. First, because lockdowns are essentially exogenous across regions, many otherwise similar users experience different degrees of lockdown. It is therefore highly plausible that we can identify such users in our sample through matching. Moreover, if users in the two groups do not differ in characteristics such as reported schooling and content interests, they are likely to be more comparable than users in the unmatched sample. If the estimates from the matched sample are similar to, or even larger than, those from the unmatched sample, this would suggest that remaining differences may not induce substantial bias in the estimated treatment effect in our setting.

Notably, all IV estimates presented in Tables 3 and 4 are substantially larger than the OLS estimates in Column (6) of Table 2. Although this may appear surprising—since common shocks or shared preferences are often thought to bias OLS upward—our results suggest that such unobservables induce only small bias in OLS.

In our setting, the larger IV estimates are more naturally interpreted as reflecting a difference between local average treatment effects (LATE) and population average treatment effects (ATEs). The instrument identifies the influence of users in regions with stricter lockdowns—who began using VPNs to access uncensored external information due to the lockdown—on their friends' beliefs. These complier users may exert stronger influence than the average information elites, perhaps because they were more engaged with politically sensitive topics during the lockdown and thus more motivated to share uncensored information. Accordingly, the larger IV estimates relative to OLS estimates may simply indicate that LATEs exceed ATEs in this context.

A final concern is the validity of the IV, which requires satisfaction of the exclusion restriction—that the instrument affects users' beliefs only through its impact on their social networks. This condition may be violated if lockdowns directly shape beliefs about politically sensitive issues. The concern is most salient for Omicron: lockdowns could directly influence residents' beliefs about its lethality and, in turn, their friends' beliefs. Stricter lockdowns might reasonably lead some residents to view Omicron as more dangerous, thereby increasing agreement with government propaganda. However, we find that greater exposure to high-intensity lockdown regions predicts more

disagreement with government propaganda, suggesting that this direct channel is unlikely to account for our results. For the other two issues—the Russia–Ukraine war and Japan's nuclear wastewater discharge plan—the exclusion restriction is more plausibly satisfied, as lockdowns are unlikely to directly affect beliefs on these topics.

4.3 Connections to Information Elites and Belief Updating

This section examines how users' beliefs evolve over time and the role of connections to information elites in that process. Because belief shifts are likely driven by new information, users connected to information elites may update their beliefs when exposed to elites' uncensored content. We therefore hypothesize that treated users (those connected to information elites) are more likely than control users to shift from agreeing to disagreeing with government propaganda over time. To test this prediction, we estimate the following equation:

$$Belief\ Shift_i = \alpha + \delta Connection_i + \gamma X_i + \varepsilon_i, \quad (5)$$

Equation (5) is similar to Equation (1), with the distinction that the dependent variable is now *Belief Shift*$_i$, a binary indicator that takes a value of 1 if user i changes their belief from agreeing to disagreeing with government propaganda, and 0 otherwise. The parameter of interest, δ, aims to capture the effect of connection to information elites on users' belief shift. A positive estimate of δ suggests that even for users who initially agree with government propaganda, those connected to information elites are more likely to change their beliefs over time. Thus, we interpret δ as the persuasive effect of information elites.

We first estimate Equation (5) on the full sample for all three issues and report the results in Columns (1), (4), and (7) of Table 5. We then apply CEM to balance observable characteristics between connected and unconnected users and re-estimate the same specification on the matched sample; these results appear in Columns (2), (5), and (8) (see Appendix B for details on the matching procedure). Finally, we restrict the matched sample to users who initially agreed with government propaganda and report the estimates for this subsample in Columns (3), (6), and (9). In this most restricted subsample, treated and control users are balanced on observables and share identical prior beliefs, making them highly comparable.

We find that as we progressively reduce differences between the treatment and control groups by restricting the sample, the coefficients increase in magnitude and become more statistically significant. This suggests that unobservable differences between the groups may induce a downward bias in our estimates. While some unobservable differences may remain—even in the most comparable

Table 5: The Effects of Connections to Information Elites on Shifts in User Beliefs Regarding Politically Sensitive Issues

Dependent Variable:	*Belief Shift*								
	Threat Posed by Omicron			Russia-Ukraine War			Japan's Discharge Plan		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Connected to Elites	0.006***	0.007***	0.013***	0.001	0.002	0.006**	0.002	0.0021	0.0025*
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.0013)	(0.0014)	(0.0014)
CEM Sample	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Prior: Agree Gov	No	No	Yes	No	No	Yes	No	No	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	403966	331697	233725	299382	165538	119417	121136	57760	55513
R-Square	0.005	0.006	0.011	0.003	0.005	0.008	0.003	0.004	0.005
Mean Y	0.046	0.046	0.065	0.055	0.055	0.077	0.009	0.009	0.009

Notes. This table reports the effect of connections to information elites on shifts in user beliefs regarding politically sensitive issues, specifically presenting the estimated coefficients (δ) from Equation (5). Columns (1)–(3), (4)–(6), and (7)–(9) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. For each issue, we sequentially present the results for the full sample, the matched sample, and restricted subsample where users' prior beliefs align with government propaganda. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

subsamples reported in Columns (3), (6), and (9) of Table 5—further reducing these differences would likely yield even larger coefficients. Accordingly, the estimates derived from the above steps likely represent a lower bound on the persuasive effect of information elites on their friends' belief shifts.

Our results provide strong evidence of persuasive effects from information elites. Specifically, Columns (1)–(3) of Table 5 show that for the issue of Omicron, the coefficients for the full sample, matched sample, and restricted subsample are 0.006, 0.007, and 0.013, respectively. This indicates that connection to information elites increases users' likelihood of belief shift by 0.006, 0.007, and 0.013 or approximately 13% (0.006/0.046), 15% (0.007/0.046), and 20% (0.013/0.065) relative to the sample means.

Columns (4)–(6) show the corresponding estimates for the Russia-Ukraine war issue, which are 0.001, 0.002, and 0.006, respectively. Although the estimates for the full and matched samples are small and statistically insignificant, restricting the sample to users with same prior beliefs leads to a significant coefficient of 0.006. This implies that, for this restricted subsample, connection to information elites increases the probability of belief shifts by 0.006, or approximately 8% (0.006/0.077) relative to the sample mean.

Finally, Columns (7)–(9) of Table 5 show that while the coefficients are around 0.002 for the issue of Japan's discharge plan, they grow larger and become more statistically significant as we further restrict the sample. Our results suggest that, for this issue, connection to information elites increases the probability of belief shift by 0.002, or approximately 22% (0.002/0.009) relative to the sample mean.

To further address endogeneity, we follow the previous section and instrument connections to information elites with users' social network exposure to lockdowns. The IV estimates for the full sample (Table A5 in the Appendix) show that only the coefficient for the Russia–Ukraine war is statistically significant, whereas the coefficients for the other two issues are small and insignificant.

Panel B of Table 4 reports the corresponding IV estimates for the matched sample. The coefficients for the Omicron and Russia–Ukraine war issues are statistically significant and substantially larger than those obtained from the full sample. The coefficient for the Japan nuclear wastewater discharge issue, however, remains small and statistically insignificant.

One likely explanation is the timing and concentration of discourse on Japan's discharge plan, which was largely confined to the one- to two-week period surrounding the start of the discharge, when the Chinese government strongly condemned Japan's actions and both domestic propaganda and public attention peaked (see Figure A1 in the Appendix). Within this short window, only 0.9 percent of users in our dataset exhibited belief shifts. As a result, the statistical power of our instrument may be insufficient to detect such subtle changes with precision.

In summary, our findings in this section show that users connected to information elites are more likely to experience belief shifts, transitioning from agreeing to disagreeing with Chinese government propaganda. This suggests that information elites may play a critical role in reshaping their friends' beliefs by providing updated, uncensored information that challenges official narratives.

4.4 Heterogenous Treatment Effects by the Residing Country of Overseas Information Elites

Because information elites can bypass censorship and access uncensored information that may contradict government narratives, they are likely to disagree with government propaganda. However, not all information elites do so. Some Chinese citizens with uncensored internet access may still support government propaganda on certain issues due to limited political interest or distrust of foreign media (Chen and Yang, 2019). We therefore hypothesize that information elites with different beliefs exert different impacts on their friends' beliefs. Specifically, elites who support government propaganda are more likely to share content that reinforces official narratives—already widely known—yielding minimal influence. By contrast, elites who disagree with government propaganda are more likely to share uncensored information, prompting their friends to deviate from official narratives. This hypothesis aligns with selective sharing: individuals preferentially share information that aligns with their preexisting beliefs while dismissing or ignoring contradictory information (Bakshy, Messing, and Adamic, 2015; Mocanu et al., 2015; Shin and Thorson, 2017; Bowen, Dmitriev, and Galperti, 2023).

Unfortunately, most information elites did not engage in discussions on the three selected issues, making it difficult to determine their relevant beliefs. However, we can infer the potential beliefs of overseas elites based on their countries or regions of residence. Not all countries' stances on these three issues contradict Chinese government propaganda. For instance, during the Omicron pandemic, while many countries relaxed controls and adopted a strategy of coexisting with the virus, some, like China, continued with strict measures. We expect that overseas information elites residing in countries maintaining strict pandemic control measures are more likely to view Omicron as deadly. In contrast, elites in countries that adopted a "coexist with the virus" approach are more likely to view it as mild.

We hypothesize that overseas information elites residing in countries where the mainstream views contradict Chinese government propaganda are more likely to oppose these narratives, sharing relevant, uncensored information with their friends in China and thereby influencing their beliefs. Conversely, elites residing in countries where the prevailing opinions align with Chinese government propaganda—or where local media largely ignore these issues—tend to either endorse government narratives or lack pertinent information, resulting in limited influence on their friends' beliefs.

To test this hypothesis, we categorize countries based on their potential positions regarding each issue:

Omicron issue: Countries maintaining strict controls are likely to align with the Chinese government narratives, while those that relaxed measures are more likely to contradict them.

Russia-Ukraine war: Countries that condemned Russia or imposed sanctions are more inclined to oppose Chinese narratives, whereas those with neutral or supportive stances towards Russia are likely to align with China's position.

Japan's discharge plan: Countries that explicitly expressed understanding and support for Japan's discharge plan are likely to challenge Chinese narratives, while those criticizing Japan or maintaining unclear positions are less likely to strongly oppose them.

Appendix A.3 provides detailed classification criteria for each issue, and Table A4 lists the categorization of countries and regions according to their potential stances on the three issues.

Based on this classification, we categorize the sample into three subgroups: (1) *Group TC:* Treatment users connected to overseas information elites residing in countries where the prevailing viewpoint contradicts Chinese government propaganda, (2) *Group TA:* Treatment users connected to overseas information elites residing in countries where the prevailing viewpoint aligns with Chinese government propaganda, and (3) *Group C:* Control users not connected to any information elites, serving as the reference group.

We estimate the following equations to assess how the treatment effects vary across these treatment groups:

$$Belief_i = \alpha + \sum_{g \neq C} \beta^g Group\ g_i + \gamma X_i + \varepsilon_i, \tag{6}$$

$$Belief\ Shift_i = \alpha + \sum_{g \neq C} \delta^g Group\ g_i + \gamma X_i + \varepsilon_i. \tag{7}$$

These mirror Equations (1) and (5), with *Group C* (no connection to information elites) as the reference. The coefficients β^g and δ^g capture, respectively, the deviation in beliefs and belief shifts for each treatment group (*Groups TC and TA*) relative to *Group C.*

We first estimate Equation (6) for each issue and present the results in Table 6. Columns (1)–(3) show that, for all three issues, the coefficients for *Group TC* are large and statistically significant, while the coefficients for *Group TA* are smaller and less significant. This suggests that information elites in countries where mainstream views contradict Chinese government propaganda have a significant influence on the beliefs of their friends in China, while those in countries with aligned views have minimal impact.

Table 6: The Heterogenous Effects of Connections to Information Elites on User Beliefs, by the Country of Residence of Information Elites

Dependent Variable:	*Belief*		
	Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
	(1)	(2)	(3)
Group TC	0.019***	0.035***	0.021***
	(0.003)	(0.005)	(0.005)
Group TA	0.001	-0.017**	0.002
	(0.010)	(0.007)	(0.003)
Observations	376279	287925	117841
R-Square	0.011	0.011	0.007
Mean Y	0.228	0.193	0.026
p-value (*Group TC*==*Group TA*)	0.095	0.000	0.003
Baseline Controls	Yes	Yes	Yes

Notes. This table examines heterogenous effects of connections to information elites on user beliefs based on the residence of overseas information elites, specifically presenting the estimated coefficients (β^g) from Equation (6). Group TC consists of users connected to overseas information elites residing in countries where the prevailing viewpoints contradict Chinese government propaganda, while Group TA consists of users connected to overseas information elites residing in countries where the mainstream viewpoints align with Chinese government propaganda. Columns (1)–(3) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. The p-values test for statistically significant differences between the coefficients for Group TC and Group TA. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

However, endogeneity remains a concern: variations in the countries of residence of overseas information elites—categorized as either aligning with or contradicting Chinese propaganda—may correlate with characteristics of their friends in China. For instance, if users connected to overseas elites in dissenting countries (*Group TC*) are generally more educated or open-minded than those connected to elites in aligned countries (*Group TA*), the observed effects might stem from these traits rather than the elites' influence.

To address this concern, we follow Bailey et al. (2018) to assess the extent to which users' observable characteristics explain variations in their elite friends' countries of residence. We define a dummy variable equal to 1 for users in *Group TC* and 0 for those in *Group TA*, then regress this indicator on users' observable characteristics. The resulting R^2 can be as low as 0.004, suggesting that observable characteristics account for only a minor fraction of these variations.

Furthermore, we adopt an approach similar to Section 4.3 to examine how observable and unobservable differences between users in these groups affect our findings. Specifically, we estimate Equation (7) to examine heterogeneous effects on belief shifts across treatment groups, and report the results in Table 7. For each issue, we sequentially report the estimates for the full sample, the matched

sample, and the restricted subsample of users whose prior beliefs align with government propaganda. The pattern closely parallels Table 6: coefficients for *Group TC* are consistently positive and significant, whereas those for *Group TA* are notably smaller and statistically insignificant.

Moreover, as we progressively reduce differences between the treatment and control groups by restricting the sample, the coefficients for *Group TC* increase in magnitude and statistical significance. By contrast, the coefficients for *Group TA* exhibit no such trend and remain small and insignificant. This pattern strengthens our interpretation that the results reflect the persuasive influence of information elites: overseas elites in countries with dissenting mainstream views exert sizable persuasion effects on their friends, whereas those in aligned countries have minimal impact on their friends' belief shifts.

A particularly compelling piece of evidence comes from the contrasting positions of the USA and Canada on Japan's discharge plan. The USA explicitly endorsed Japan's plan, expressing support and understanding, whereas Canada took no clear stance, with its local media largely ignoring the issue. Table A6 in the Appendix shows that users connected to information elites in the USA exhibit minimal differences in most observable characteristics compared to those connected to elites in Canada, while Column (1) in Table A7 reveals that both groups share similar beliefs regarding Japan's discharge plan. Nevertheless, Column (2) in Table A7 shows that the former are significantly more likely to shift their beliefs from agreement to disagreement with Chinese government propaganda over time than the latter. These findings suggest that U.S.-based elites may have significantly shaped their friends' beliefs by sharing relevant, uncensored information.

Taken together, these results support our hypothesis that information elites with greater exposure to politically sensitive information are more likely to share relevant, uncensored content with their friends in China, thereby influencing their beliefs.

Geographic proximity and relational closeness between information elites and their friends may play a critical role in the dissemination of uncensored information through social networks. In the SI, we examine how these factors affect the treatment effects based on users' proximity to information elites. Our findings indicate that users who are geographically closer to, or have a more intimate relationship with, information elites are more likely to be influenced by them, highlighting the critical role of social interactions in the dissemination of politically sensitive information through social networks. See Section S4 in the SI for details.

4.5 Potential Concern: Belief-Driven Sorting through Social Interactions

A potential concern is that users' connections with information elites may arise from pre-existing, similar beliefs. For instance, users may perceive shared beliefs with information elites on the three

issues through online interactions, which could motivate them to form friendship links. This belief-driven sorting could threaten the causal interpretation of our results, suggesting that the connection between users and information elites could be the result of shared beliefs, rather than the cause of belief formation.

However, this concern is unlikely to undermine our findings. As mentioned earlier, the majority of Weibo users' online friends are also their real-world friends, making it unlikely that their friendship links are formed solely through online interactions. Additionally, we randomly selected 300,000 posts on the three propagandized topics, along with all the comments following these posts, capturing online interactions on these issues. Our analysis reveals that nearly all (99.4%) of the authors of the comments have no friendship links with the authors of the posts they are commenting on. This indicates that online discussions on politically sensitive issues overwhelmingly occur with strangers and rarely result in the formation of new friendships.

Moreover, the results from the previous section suggest that belief-driven sorting is unlikely to be the primary factor driving our findings. In our strongest specification, where both treatment and control users initially agree with government propaganda and show no observable differences, we find that users connected to information elites are more likely to change their beliefs. Since information elites have access to external uncensored information, they are more likely to disagree with government propaganda. Therefore, their online friendships with treatment users—who initially agree with government propaganda—are unlikely to have been formed due to shared beliefs during interactions. Therefore, a more plausible explanation for our results is that information elites gradually persuaded their friends to shift from agreeing to disagreeing with government propaganda by sharing uncensored information that contradicts official narratives.

4.6 Network Spillover Effect

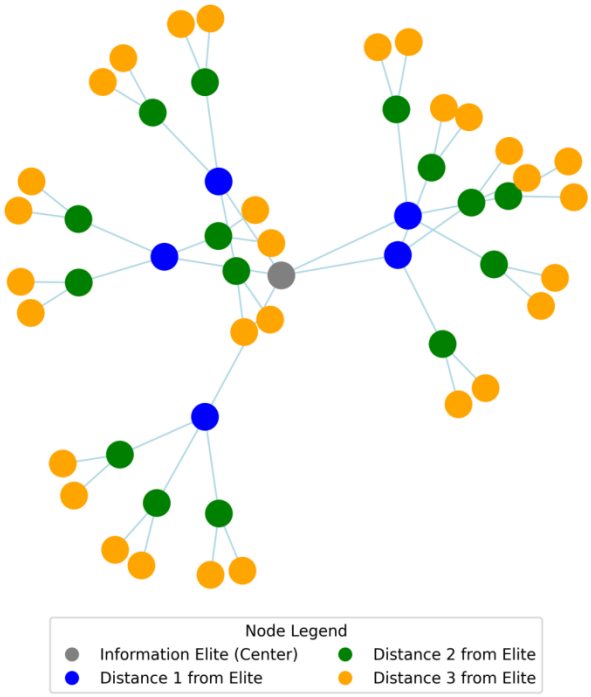
Our previous results suggest that information elites can shape their friends' beliefs on politically sensitive issues by sharing relevant uncensored information. Once users receive this information, they may further disseminate it to others, leading to a rapid diffusion of uncensored content across social networks. Therefore, we hypothesize that users' connections to information elites not only directly influence their own beliefs but also have an impact on their broader networks, including friends of friends. In this section, we explore this network spillover effect.

Table 7: The Heterogenous Effects of Connections to Information Elites on Shifts in User Beliefs, by the Country of Residence of Information Elites

Dependent Variable:	*Belief Shift*								
	Threat Posed by Omicron			Russia-Ukraine War			Japan's Discharge Plan		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Group TC	0.008***	0.009***	0.015***	0.002	0.002	0.009**	0.005	0.006*	0.007**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Group TA	0.004	0.005	0.008	0.001	0.002	0.001	-0.003*	-0.002	-0.002
	(0.005)	(0.005)	(0.007)	(0.004)	(0.004)	(0.006)	(0.002)	(0.002)	(0.002)
CEM Sample	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Prior: Agree Gov	No	No	Yes	No	No	Yes	No	No	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	376279	304349	215071	287925	154278	111571	117841	54586	52504
R-Square	0.005	0.006	0.011	0.003	0.005	0.008	0.003	0.005	0.005
Mean Y	0.047	0.046	0.065	0.056	0.055	0.077	0.009	0.009	0.009
p-value (*Group TC==Group TA*)	0.488	0.510	0.335	0.928	0.892	0.275	0.018	0.019	0.017

Notes. This table examines heterogenous effects of connections to information elites on shifts in user beliefs based on the residence of overseas information elites, specifically presenting the estimated coefficients (δ^g) from Equations (7). Group TC consists of users connected to overseas information elites residing in countries where the prevailing viewpoints contradict Chinese government propaganda, while Group TA consists of users connected to overseas information elites residing in countries where the mainstream viewpoints align with Chinese government propaganda. Columns (1)–(3), (4)–(6), and (7)–(9) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. For each issue, we sequentially present the results for the full sample, the matched sample, and restricted subsample where users' prior beliefs align with government propaganda. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Figure 2: An Example of a Typical Social Network Structure



Notes. This figure illustrates a typical social network, with the gray node at the center representing an information elite. The other nodes are categorized by their distance from the elite: distance 1 (blue), distance 2 (green), and distance 3 (orange). The connections between nodes represent both direct and indirect ties, with closer proximity to the elite indicating a higher potential for influence or access to information.

Figure 2 illustrates an example of a typical social network structure, where nodes represent users and edges denote friendship links. The grey dot at the center of the network represents an information elite. Users are divided into three groups based on their social distance from the elite: those directly connected to the elite (blue dots), those two steps away (green dots), and those more than two steps away (yellow dots). Users with a closer social distance to the elite are more likely to receive politically sensitive information directly. Even users not directly connected to the elite may still be exposed to uncensored information through their friends or friends of friends. Thus, if network spillovers exist, we expect that users indirectly connected to information elites will also be affected, with those closer to elites experiencing a stronger impact.

To test this hypothesis, we compare the beliefs and belief shifts of users who are not directly connected to information elites but are closer to them, versus those who are more distant. Specifically, we categorize users into two group: the treatment group, which includes users at a social distance of 2 from information elites, and the control group, which includes users who are farther than 2 steps away. We then examine whether users in the treatment group, who are closer to information elites, are more

likely to be influenced by their connections and, ultimately leading to greater disagreement with government propaganda.

Table 8: The Spillover Effects of Connections to Information Elites on User Beliefs Regarding Politically Sensitive Issues

Dependent Variable:	*Belief*		
	Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
	(1)	(2)	(3)
Distance=2	0.006**	0.022***	0.002*
	(0.003)	(0.004)	(0.001)
Observations	94733	63671	43204
R-Square	0.015	0.015	0.008
Mean Y	0.235	0.250	0.017
Baseline Controls	Yes	Yes	Yes

Notes. This table examines the spillover effects of connections to information elites on user beliefs regarding politically sensitive issues, presenting the estimated coefficients (β) from Equation (1). Specifically, the treatment group consists of users who are at a social distance of 2 from information elites, while the control group comprises users are farther than 2 steps away. The independent variable, *Distance=2*, is a dummy variable that equals one if the user belongs to the treatment group and 0 if the user belongs to the control group. Column (1)–(3) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

In practice, we create a dummy variable (labeled *Distance=2*), which takes the value of 1 if the user belongs to the treatment group and 0 if the user belongs to the control group. We use this variable as the treatment in Equation (1), with results reported in Columns (1)–(3) of Table 8. The coefficients for all three issues are all positive and statistically significant. These results suggest that, compared to users in the control group, those in the treatment group are more likely to disagree with government propaganda. While these effects are smaller than the direct influence of connections to information elites presented in Table 2, they remain non-negligible.

Next, we further examine whether users in the treatment group are more likely to experience belief shifts over time, compared to users in the control group. Following a similar approach, we re-estimate Equation (5) and present the results in Table 9. For each issue, we sequentially present the results for the full sample, the matched sample, and the restricted subsample where users' prior beliefs align with government propaganda.

The first three columns of Table 9 show that, for the Omicron issue, the coefficients for the three samples are 0.001, 0.003, and 0.004, respectively. This suggests that, compared to users in the control group, users in the treatment group are significantly more likely to change their beliefs, shifting from

Table 9: The Spillover Effects of Connections to Information Elites on Shifts in User Beliefs on Politically Sensitive Issues

Dependent Variable:	*Belief Shift*								
	Threat Posed by Omicron			Russia-Ukraine War			Japan's Discharge Plan		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Distance=2	0.001	0.003*	0.004**	-0.002	-0.001	0.001	0.002**	0.003***	0.003***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)
CEM Sample	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Prior: Agree Gov	No	No	Yes	No	No	Yes	No	No	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	94733	89995	63771	63671	58478	39068	43204	38640	37633
R-Square	0.006	0.007	0.011	0.005	0.007	0.011	0.006	0.007	0.008
Mean Y	0.038	0.038	0.053	0.051	0.051	0.077	0.006	0.006	0.006

Notes. This table examines the spillover effects of connections to information elites on user belief shifts regarding politically sensitive issues, presenting the estimated coefficients (δ) from Equation (5). Specifically, the treatment group consists of users who are at a social distance of 2 from information elites, while the control group comprises users are farther than 2 steps away. The independent variable, *Distance=2*, is a dummy variable that equals one if the user belongs to the treatment group and 0 if the user belongs to the control group. Columns (1)–(3), (4)–(6), and (7)–(9) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. For each issue, we sequentially present the results for the full sample, the matched sample, and restricted subsample where users' prior beliefs align with government propaganda. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

perceiving Omicron as fatal to viewing it as mild. In our strongest specification in Column (3), the estimate indicates that treatment users are 0.004 more likely to experience a belief shift, representing a 7.6% (0.004/0.053) increase relative to the sample mean. The results for Japan's discharge plan show similar patterns. For the Russia-Ukraine war issue, although the coefficients for the three samples are not statistically significant, there is a trend of increasing coefficients, and the strongest specification in Column (6) yields a positive estimate, though statistically insignificant.

The presence of network spillovers implies that our baseline results likely underestimate the full impact of information elites across social networks. While only a small portion (about 10% in our sample) are directly connected to information elites, nearly all users are indirectly connected. These indirect connections are crucial for enabling the broader public to access politically sensitive information, which in turn shape public opinion. More importantly, these spillovers ensure that politically sensitive information does not remain confined to a small group, but spreads to a wider population through social transmission, ultimately undermining the effectiveness of media censorship.

4.7 The Issue of Measurement Errors

This section addresses potential measurement errors in identifying users' connections to information elites and their beliefs.

We determine information elites by tracking users' IP addresses when they post. However, this strategy may not capture all information elites, leading to possible measurement errors in identifying users' connections to information elites. For instance, if an information elite regularly uses a VPN to browse foreign websites but never activates the VPN service when posting on the Sina Weibo platform, our method would fail to classify them as an information elite. This would result in a smaller set of identified elites and lead to misclassification of some users who are actually connected to information elites as part of the control group.

However, such measurement errors are unlikely to seriously threaten our identification. Treatment users are generally less likely to believe government propaganda and are more likely to experience belief shifts. Therefore, misclassifying treatment users as control users would make it harder to detect significant treatment effects. In other words, this type of measurement error is likely to lead to an underestimation of the true impact of information elites on their friends' beliefs.

Furthermore, measurement errors may also arise in identifying users' beliefs on the selected issues. Although our NLP model performs well in predicting user beliefs, its accuracy is not perfect. However, if these prediction errors are random and not systematically correlated with user characteristics, they should not pose a significant problem. To further assess the

robustness of our results, we define user beliefs more flexibly, assuming that a user disagrees with propaganda if more than 90% or 75% of his posts are predicted as Disagree. As shown in Table A8 in the Appendix, our main results remain robust across these different thresholds for defining user beliefs.

5. Conclusions

In the digital age, social networks have become a powerful mechanism for information dissemination, even under strict media censorship. Using a unique dataset from Sina Weibo, this study empirically examines how politically sensitive information spreads through social networks in China, where censorship is pervasive. Our results reveal that information elites—individuals who can bypass censorship to access uncensored content—significantly influence their friends' beliefs, making them more likely to deviate from Chinese government propaganda. Moreover, we observe considerable network spillover effects, where the impact extends to users indirectly connected to information elites. This suggests that politically sensitive information, which the government aims to suppress, can spread to a broader audience through social networks, ultimately reshaping public opinion. Our findings indicate that the power of social networks in information dissemination poses a significant challenge to the effectiveness of media censorship in authoritarian regimes.

As highlighted in the literature, the demand for uncensored information among citizens is a key determinant of the effectiveness of media censorship (Chen and Yang, 2019). Similarly, the potential of social networks to undermine the effectiveness of censorship depends, in part, on citizens' demand for uncensored content. Only when citizens become aware of censorship and develop a demand for uncensored content do they have incentives to leverage resources such as social networks to access it (Roberts, 2020). As Chen and Yang (2019) demonstrate, China's censorship apparatus remains robust due to the low demand for uncensored information among its citizens, largely because they are unaware of its value. Consequently, censorship systems remain effective, even when the cost of circumventing them is relatively low (Roberts, 2020).

However, as the literature suggests, citizens' demand for uncensored information is not always low. As Chen and Yang (2019) note, even when initial demand is low, once citizens are incentivized to access uncensored content and recognize its value, their demand increases and persists. Furthermore, in authoritarian regimes, while demand for uncensored content may remain low during normal times, it tends to surge during times of crisis. This heightened demand prompts citizens to bypass censorship barriers in search of relevant information (Ball-Rokeach and DeFleur, 1976; Loveless, 2008; Weidmann and Rød, 2019; Chang et al., 2022).

More generally, as education and income levels rise, citizens become more adept at recognizing and bypassing censorship, further increasing their demand for uncensored content (Roberts, 2020).

As observed recently in China, when economic situation worsens and government credibility declines, citizens may begin to question official narratives and seek alternative sources of information. As this demand for uncensored content grows, both censorship systems and autocratic regimes face increasing challenges. How autocratic regimes respond to these new challenges—and how the dynamics of demand for uncensored information interact with resources such as social networks to bypass censorship in the context of potential regime responses—remain crucial issues for future research.

References

Acemoglu, Daron, and Asuman Ozdaglar, "Opinion Dynamics and Learning in Social Networks," *Dynamic Games and Applications*, 1 (2011), 3–49.

Acemoglu, Daron, Asuman Ozdaglar, and Ali ParandehGheibi, "Spread of (Mis)information in Social Networks," *Games and Economic Behavior*, 70 (2010), 194–227.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya, "Radio and the Rise of the Nazis in Prewar Germany," *The Quarterly Journal of Economics*, 130 (2015), 1885–1939.

Alesina, Alberto, and Nicola Fuchs-Schündeln, "Good-Bye Lenin (or Not?): The Effect of Communism on People's Preferences," *American Economic Review*, 97 (2007), 1507–1528.

Allcott, Hunt, and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 31 (2017), 211–236.

Azzimonti, Marina, "Social Media Networks, Fake News, and Polarization," *European Journal of Political Economy*, 76 (2023). https://doi.org/10.1016/j.ejpoleco.2022.102256.

Bai, Jie, Mikhail Golosov, Nancy Qian, and Yan Kai, "Understanding the Influence of Government-Owned Media: Evidence from Air Pollution in China," *Unpublished*, 2015.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, "The Economic Effects of Social Networks: Evidence from the Housing Market," *Journal of Political Economy*, 126 (2018), 2224–2276.

Bailey, Michael, Drew Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel, "Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing during the COVID-19 Pandemic," *Journal of Political Economy Microeconomics*, 2 (2024), 463–494.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic, "Exposure to Ideologically Diverse News and Opinion on Facebook," *Science*, 348 (2015), 1130–1132.

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada Adamic, "The Role of Social Networks in Information Diffusion," in *Proceedings of the 21st international conference on World Wide Web*. (Lyon, France: ACM, 2012), 519–528.

Ball-Rokeach, S.J., and M.L. DeFleur, "A Dependency Model of Mass-Media Effects," *Communication Research*, 3 (1976), 3–21.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson, "The Diffusion of Microfinance," *Science*, 341 (2013). https://doi.org/10.1126/science.1236498.

——, "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials," *Review of Economic Studies*, 86 (2019), 2453–2490.

Barwick, Panle Jia, Yanyan Liu, Eleonora Patacchini, and Qi Wu, "Information, Mobile Communication, and Referral Effects," *American Economic Review*, 113 (2023), 1170–1207.

Beaman, Lori A., "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S.," *The Review of Economic Studies*, 79 (2012), 128–161.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak, "Can Network Theory-Based Targeting Increase Technology Adoption?," *American Economic Review*, 111 (2021), 1918–1943.

Blumenstock, Joshua E, Guanghua Chi, and Xu Tan, "Migration and the Value of Social Networks," *The Review of Economic Studies*, 92 (2025), 97–128.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler, "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature*, 489 (2012), 295–298.

Bowen, T. Renee, Danil Dmitriev, and Simone Galperti, "Learning from Shared News: When Abundant Information Leads to Belief Polarization," *Quarterly Journal of Economics*, 138 (2023), 955–1000.

Burchardi, Konrad B., and Tarek A. Hassan, "The Economic Impact of Social Ties: Evidence from German Reunification*," *The Quarterly Journal of Economics*, 128 (2013), 1219–1271.

Bursztyn, Leonardo, Aakaash Rao, Christopher Roth, and David Yanagizawa-Drott, "Opinions as Facts," *The Review of Economic Studies*, 90 (2023), 1832–1864.

Cantoni, Davide, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Curriculum and Ideology," *Journal of Political Economy*, 125 (2017), 338–392.

Chang, Keng-Chi, William R. Hobbs, Margaret E. Roberts, and Zachary C. Steinert-Threlkeld, "COVID-19 Increased Censorship Circumvention and Access to Sensitive Topics in China," *Proceedings of the National Academy of Sciences*, 119 (2022). https://doi.org/10.1073/pnas.2102818119.

Chen, Yuyu, and David Y. Yang, "The Impact of Media Censorship: 1984 or Brave New World?," *American Economic Review*, 109 (2019), 2294–2332.

DellaVigna, Stefano, and Ethan Kaplan, "The Fox News Effect: Media Bias and Voting," *The Quarterly Journal of Economics*, 122 (2007), 1187–1234.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv Working Paper, 2018. https://doi.org/10.48550/arXiv.1810.04805.

Diamond, Larry, "Liberation Technology," *Journal of Democracy*, 21 (2010), 69–83.

Eil, David, and Justin M. Rao, "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 3 (2011), 114–138.

Ellison, Glenn, and Drew Fudenberg, "Rules of Thumb for Social Learning," *Journal of Political Economy*, 101 (1993), 612–643.

——, "Word-of-Mouth Communication and Social Learning," *Quarterly Journal of Economics*, 110 (1995), 93–125.

Enke, Benjamin, "What You See Is All There Is," *Quarterly Journal of Economics*, 135 (2020), 1363–1398.

Fong, Jessica, Tong Guo, and Anita Rao, "Debunking Misinformation About Consumer Products: Effects on Beliefs and Purchase Behavior," *Journal of Marketing Research*, 61 (2024), 659–681.

Foster, Andrew D., and Mark R. Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103 (1995), 1176–1209.
——, "Microeconomics of Technology Adoption," *Annual Review of Economics*, 2 (2010), 395–424.
Golub, Benjamin, and Matthew O. Jackson, "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2 (2010), 112–149.
Granovetter, Mark S., "The Strength of Weak Ties," *American Journal of Sociology*, 78 (1973), 1360–1380.
Guriev, Sergei, and Daniel Treisman, "Informational Autocrats," *Journal of Economic Perspectives*, 33 (2019), 100–127.
Hobbs, William R., and Margaret E. Roberts, "How Sudden Censorship Can Increase Access to Information," *American Political Science Review*, 112 (2018), 621–636.
Huang, Haifeng, and Yao-Yuan Yeh, "Information from Abroad: Foreign Media, Selective Exposure and Political Support in China," *British Journal of Political Science*, 49 (2019), 611–636.
Iacus, Stefano Maria, Gary King, and Giuseppe Porro, "Matching for Causal Inference Without Balance Checking," available at SSRN (2008): http://dx.doi.org/10.2139/ssrn.1152391.
——, "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis*, 20 (2012), 1–24.
Jackson, Matthew O., and Dunia López-Pintado, "Diffusion and Contagion in Networks with Heterogeneous Agents and Homophily," *Network Science*, 1 (2013), 49–67.
Jackson, Matthew O., Brian W. Rogers, and Yves Zenou, "The Economic Consequences of Social-Network Structure," *Journal of Economic Literature*, 55 (2017), 49–95.
Jackson, Matthew O., and Leeat Yariv, "Diffusion, Strategic Interaction, and Social Structure," in *Handbook of Social Economics*, Benhabib, Bisin and Jackson, eds. (North-Holland, 2011), 645–678.
Katz, Elihu, and Paul F. Lazarsfeld, Personal Influence: The Part Played by People in the Flow of Mass Communication (Glencoe, IL: Free Press, 1955).
King, Gary, Jennifer Pan, and Margaret E. Roberts, "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review*, 107 (2013), 326–343.
——, "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation," *Science*, 345 (2014). https://doi.org/10.1126/science.1251722.
——, "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument," *American Political Science Review*, 111 (2017), 484–501.
Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., ... & Scarpino, S. V., "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, 368(2020), 493-497.
Lazarsfeld, P. F., B. Berelson, and H. Gaudet, The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign (New York: Duell, Sloan and Pearce, 1944).
Levy, Ro'ee, "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment," *American Economic Review*, 111 (2021), 831–870.
Lorentzen, Peter, "China's Strategic Censorship," *American Journal of Political Science*, 58 (2014), 402–414.
Loveless, Matthew, "Media Dependency: Mass Media as Sources of Information in the Democratizing Countries of Central and Eastern Europe," *Democratization*, 15 (2008), 162–183.
MacKinnon, Rebecca, Consent of the Networked: The Worldwide Struggle for Internet Freedom (New York: Basic Books, 2012).

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook, "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27 (2001), 415–444.
Mocanu, Delia, Luca Rossi, Qian Zhang, Marton Karsai, and Walter Quattrociocchi, "Collective Attention in the Age of (Mis)information," *Computers in Human Behavior*, 51 (2015), 1198–1204.
Morozov, Evgeny, The Net Delusion: The Dark Side of Internet Freedom (*New York: PublicAffairs*, 2011).
Murphy, Kevin M., and Andrei Shleifer. "Persuasion in Politics," *American Economic Review*, 94 (2004), 435–439.
Ou, Susan, and Heyu Xiong, "Mass Persuasion and the Ideological Origins of the Chinese Cultural Revolution," *Journal of Development Economics*, 153 (2021). https://doi.org/10.1016/j.jdeveco.2021.102732.
Pogorelskiy, Kirill, and Matthew Shum, "News We Like to Share: How News Sharing on Social Networks Influences Voting Outcomes," available at SSRN (2019): http://dx.doi.org/10.2139/ssrn.2972231.
Qin, Bei, David Strömberg, and Yanhui Wu, "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda," *Journal of Economic Perspectives*, 31 (2017), 117–140.
Roberts, Margaret E., Censored: Distraction and Diversion Inside China's Great Firewall (Princeton, NJ: Princeton University Press, 2018).
——, "Resilience to Online Censorship," *Annual Review of Political Science*, 23 (2020), 401–419.
Rogers, Everett M., Diffusion of Innovations (New York: Free Press, 1962).
Shadmehr, Mehdi, and Dan Bernhardt, "State Censorship," *American Economic Journal: Microeconomics*, 7 (2015), 280–307.
Shin, Jieun, and Kjerstin Thorson, "Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media," *Journal of Communication*, 67 (2017), 233–255.
Voigtländer, Nico, and Hans-Joachim Voth, "Nazi Indoctrination and Anti-Semitic Beliefs in Germany," *Proceedings of the National Academy of Sciences*, 112 (2015), 7931–7936.
Weidmann, Nils B., and Espen Geelmuyden Rød, The Internet and Political Protest in Autocracies (Oxford, UK: Oxford University Press, 2019).
Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political Effects of the Internet and Social Media," *Annual Review of Economics*, 12 (2020), 415–438.

Online Appendix for

The Power of Social Networks: Information Elites and the Spread of Politically Sensitive Information under Media Censorship

Xuebo Wang,[1] Hong Yang[2]

December 17, 2025

[1] School of Economics, Shanghai University of Finance and Economics. E-mail: wang.xuebo@mail.shufe.edu.cn.
[2] School of Economics, Shanghai University of Finance and Economics. E-mail: yanghong@stu.sufe.edu.cn.

A. Data Collection and Processing

A.1 Determining Users' Beliefs on the Three Issues Using NLP Models

We collect data on Sina Weibo users' posts related to the three selected issues and determine their stance on these issues using advanced natural language processing (NLP) techniques. This process consists of four main stages: data collection and corpus construction, human coding, model training, and automated prediction. Below, we provide a detailed explanation of the tasks involved in each stage.

A.1.1 Data Collection and Corpus Construction

We begin by collecting a large volume of posts related to the three selected issues from the Sina Weibo platform. To do so, we develop a crawler system that extracts Weibo posts and comments on these issues. For each issue, we analyze a substantial set of online discussions and identify frequently occurring keywords. Based on these observations, we define a comprehensive set of relevant keywords and specify a time window for each topic, as outlined in Table A1. The crawler then retrieves all posts containing these keywords within the designated time frames, along with the associated comments. This process generates nearly 34 million posts and comments from millions of users. Figure A1 illustrates the temporal trends in the volume of Sina Weibo discussions for the three selected topics.

The inherent noise and irrelevant information commonly present in social media posts can undermine the accuracy, efficiency and overall performance of subsequent model training and prediction tasks (Tabassum and Patil, 2020). To ensure high data quality, and in line with established practices in the literature (Ainslie et al., 2020; Egger and Gokce, 2022; Yu et al., 2023), we implement the following preprocessing steps:

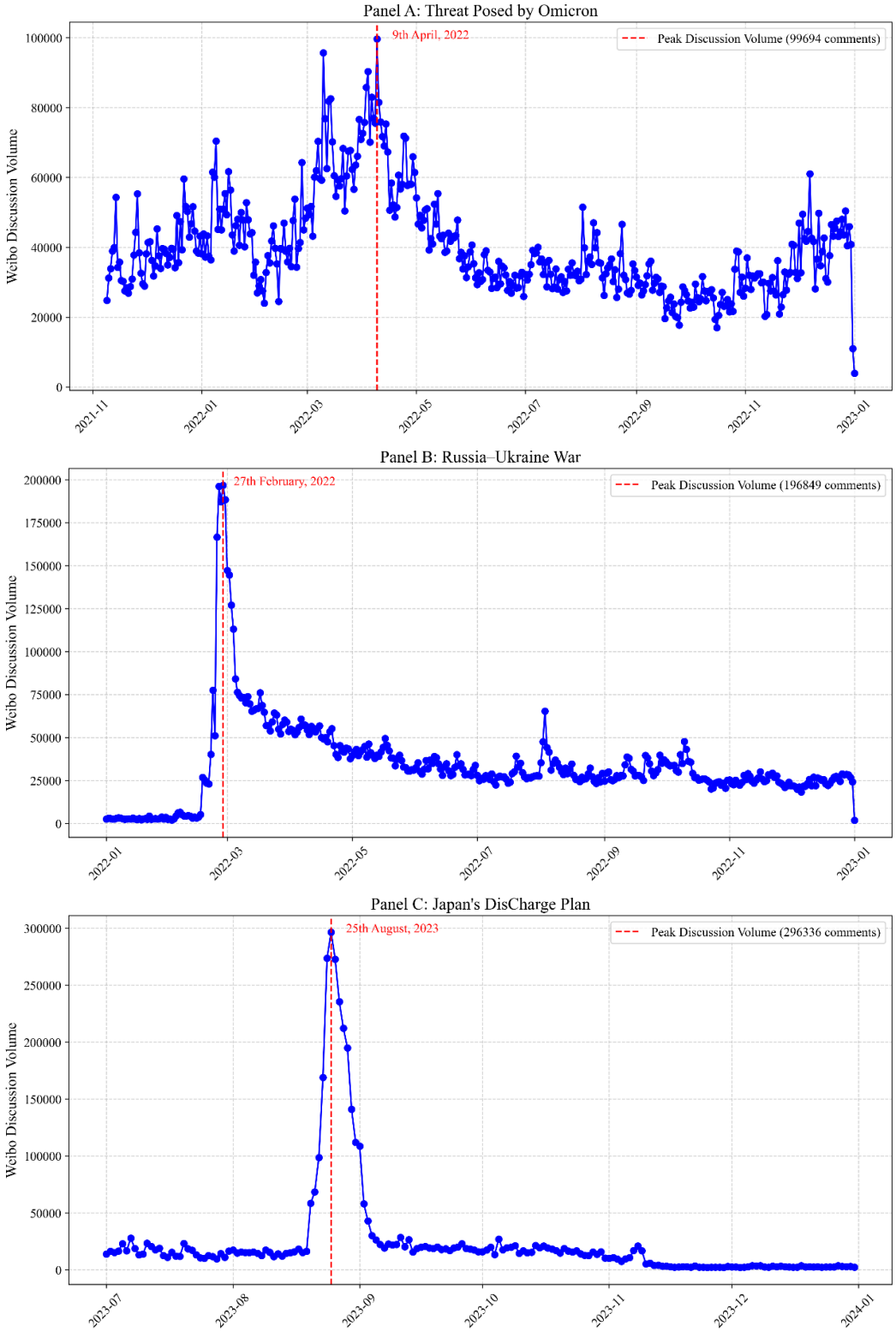
- **Removing irrelevant symbols and noise:** We remove noise elements such as HTML tags, URLs, and emojis, as they do not contain meaningful semantic information and may interfere with subsequent model training and prediction.
- **Removing stopwords:** Stopwords refer to common words that frequently appear in text but contribute little to semantic understanding (e.g., "the," "is," "in"). Removing them reduces redundancy, improves training efficiency, and enhances the model's focus on meaningful content, thereby boosting prediction accuracy.
- **Controlling text length:** We delete texts longer than 1000 characters, which constitute less than 0.5% of the dataset. Long texts often contain redundant information that hinders the model's ability to capture key content, thus negatively impacting both training efficiency and prediction performance.

Table A1: Keywords and Time Windows for Corpus Collection

Topic & Time	Keyword List
Threat Posed by Omicron [2021/11/01 to 2022/12/31]	COVID-19 (新冠疫情), Omicron (奥密克戎), Positive Patients (阳性病例), High Fever (高烧), Asymptomatic (无症状), Mild Symptoms (轻症), Severe Symptoms (重症), Virus-Related Fatality Rate (病毒致死率), Long COVID (长新冠), Dynamic Zero-COVID Policy (动态清零政策), Temporary Hospitals (方舱医院), Stay-at-home Restrictions (居家隔离), Lockdowns (封城), Prevention and Control (防控), Nucleic Acid Test (核酸检测), Reopening (解封), Coexist with Virus (共存), Zero-COVID Faction (清零派), Opening-up Faction (放开派), Medical System Congestion (医疗挤兑), Upper Respiratory Infection (上呼吸道感染), Wenhong, Zhang (张文宏), Nanshan, Zhong (钟南山), Xijin, Hu (胡锡进), Lanjuan, Li (李兰娟), Zunyou, Wu (吴尊友), Tedros Adhanom Ghebreyesus (谭德赛), Big Flu (大号流感), Vaccination (疫苗接种), Herd Immunity (群体免疫), Antiviral Drugs (特效药), Booster Shot (加强针)
Russia-Ukraine War [2022/01/17 to 2022/12/31]	Russia-Ukraine War (俄乌战争), Russia-Ukraine Conflict (俄乌冲突), Special Military Action (特别军事行动), Vladimir Putin (普京), Volodymyr Zelensky (泽连斯基), NATO Eastward Expansion (北约东扩), War of Self-Defense (自卫战), Aggressive War (侵略战争), Genocide (种族灭绝), War Crimes (战争罪), Fascism (法西斯), Nazis (纳粹), Eastern Ukraine (乌东), Donbas (顿巴斯), Donetsk (顿涅兹克), Luhansk (卢甘斯克), Crimea (克里米亚), Kyiv (基辅), Bucha Massacre (布查惨案), Minsk Agreement (明克斯协议), Chechnya (车臣), Black Sea Fleet (黑海舰队), Referendum (公投), Separatism (分裂主义), Violation of Sovereignty (侵犯主权), Land Annexation (侵吞土地), Korean War (抗美援朝), Anti-War (反战), Military Support (军事支援), Military Sanctions (军事制裁), Geopolitics (地缘政治), Strategic Buffer (战略缓冲), Security Bottom Line (安全底线)
Japan's Discharge Plan [2023/07/01 to 2023/12/31]	Fukushima Nuclear Power Plant (福岛核电站), TEPCO (东京电力公司), Japan's Discharge Plan (日本排海计划), Nuclear Contaminated Water (核污水), Nuclear Wastewater (核废水), Cooling Water (冷却水), ALPS (多核素去除设备), Treated Water (处理水), Diluted Discharge (稀释排放), Chernobyl (切尔诺贝利), Daya Bay Nuclear Power Plant (大亚湾核电站), Qinshan Nuclear Power Plant (秦山核电站), Nuclear Pollution (核污染), Nuclear Radiation (核辐射), Background Radiation (本底辐射), Radioactive Substances (放射性物质), Radioisotopes (核素), Tritium (氚), Half-life (半衰期), Ocean Currents (洋流), Bioaccumulation Effect (富集效应), Boycott Japanese Goods (抵制日货), Seafood (海鲜), Emission Concentration (排放浓度), Emission Volume (排放量), Nuclear Experts (核专家), IAEA (国际原子能机构), Rafael Grossi (格罗西), Safety Assessment Report (安全评估报告), Compliant with International Safety Standards (符合国际安全标准), Independent Sampling (独立取样), Independent Testing (独立检测), 70 Billion Public Relations Expense (700 亿公关费), Ocean's Self-Cleaning Ability (海洋自净能力), Marine Environment (海洋环境), Marine Life (海洋生物)

Notes. This table presents the relevant keywords and corresponding time windows for each of the three selected topics. Our crawling system collects all posts containing these keywords that were published within the specified time frames.

Figure A1: Temporal Trends in Sina Weibo Discussion Volume



Notes. These figures illustrate the temporal trends in the volume of Weibo discussions for the three selected issues. Panels A–C shows the trends for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. The red dashed lines mark the peak discussion dates, along with the corresponding comment volumes.

- **Text segmentation:** Since Chinese text lacks clear delimiters like spaces, segmentation is essential for breaking sentences into units that NLP models can process. We use *Jieba*, a widely used Python module for Chinese word segmentation, to segment the text into meaningful words or phrases.

These preprocessing steps yield a high-quality dataset, which is used for subsequently manual annotation, model training, and prediction tasks.

A.1.2 Human Coding

We first manually annotate a large set of text samples randomly selected from the dataset, which will serve as training data for the machine learning models. To ensure consistency and accuracy throughout the annotation process, we establish clear and detailed criteria for each topic, which will be outlined in Sections A.1.2.1, A.1.2.2, and A.1.2.3. These guidelines offer annotators explicit instructions on how to categorize each text into one of three subgroups:

Agree (indicating the user agrees with government propaganda)

Disagree (indicating the user disagrees with government propaganda)

Unidentifiable (indicating insufficient information to determine the user's belief).

We recruit three research assistants (RAs) to carry out the annotation task. To ensure they fully grasp the criteria, we provide detailed explanations and practical examples to guide their application of the criteria. We also conduct a trial annotation period during which annotators receive feedback to enhance their precision.

During the actual annotation process, we implement a dual-annotation and arbitration mechanism (Li, Rubinstein, and Cohn, 2019): each data entry was independently annotated by two annotators. In cases of disagreement, a third annotator serves as an arbitrator to resolve discrepancies and ensure the final annotation was accurate and consistent. The manual annotation process took approximately three months, during which around 60,000 text samples were annotated.

A.1.2.1 The Threat Posed by Omicron: Annotation Guidelines

For the issue of Omicron, our goal is to determine whether a user agrees with the Chinese government's propaganda that "Omicron remains deadly". As noted earlier, annotators should classify each post into one of three categories: *Agree*, *Disagree*, and *Unidentifiable*. Below are the types of expressions in the posts that should be classified as *Agree*, *Disagree*, or *Unidentifiable*, respectively.

1. Posts to be labeled as *Agree*

If a user's post explicitly or implicitly supports the view that "Omicron remains deadly," it should be classified as *Agree*. Specifically, posts containing the following types of expressions should be labeled as *Agree*:

- Explicitly stating Omicron's lethality: e.g., "Omicron is really dangerous, and millions have died in the U.S. because of it," or "This variant is more deadly than previous ones."
- Emphasizing Omicron's severe threat to health: e.g., "Many individuals will suffer from the long-term effects of COVID-19, resulting in lifelong consequences," or "Infection with Omicron leads to serious long-term health damage."
- Citing data or reports supporting its lethality: e.g., "According to the latest research, Omicron could cause severe long-term health damage," or "News reports say that Omicron has caused millions of deaths."
- Opposing the strategy of coexistence with the virus: e.g., "Western countries' shift toward coexistence with the virus is catastrophic," or "Zhang Wenhong must have been bribed by the West for advocating a coexistence strategy."
- Supporting strict epidemic control measures: e.g., "Omicron poses a significant threat to the elderly and children; strict control is necessary to protect them," or "We should adhere to the Zero-COVID policy to eradicate Omicron, even if it comes at the cost of economic and personal freedom."

2. Posts to be labeled as *Disagree*

If a user's post explicitly or implicitly opposes the view that "Omicron remains deadly," it should be classified as *Disagree*. Specifically, the following types of expressions should be labeled as *Disagree*:

- Denying Omicron's lethality: e.g., "Omicron is mild and not deadly," or "Omicron infection is no worse than a mild cold."
- Believing that domestic media has exaggerated Omicron's lethality: e.g., "The media has overstated Omicron's deadliness, essentially defending the legitimacy of the dynamic Zero-COVID policy," or "Many countries have returned to normal life, and the risk from Omicron is far less than what the media claims."
- Supporting the strategy of coexistence with the virus: e.g., "We should follow the example of Western countries, which have chosen to coexist with the virus, allowing life to essentially return to normal," or "The virus will become milder over time; there is no need for continued lockdowns, and gradual reopening is the inevitable trend."
- Supporting lenient control measures or opposing strict ones: e.g., "The secondary disasters caused by strict controls far outweigh the actual threat posed by Omicron," or "Flu causes many deaths every

year, yet we've never implemented such harsh control measures; Omicron does not warrant these measures either."

3. Posts to be labeled as *Unidentifiable*

If a user's post does not clearly indicate support for or opposition to the view that "Omicron remains deadly," it should be classified as *Unidentifiable*. Specifically, the following types of expressions should be labeled as *Unidentifiable*:

- Neutral or ambiguous discussions: e.g., "I've heard that Omicron is more contagious than previous variants, but I'm not sure whether it is more deadly," or "The situation with Omicron is complex, and we need more research to fully understand its effects."
- Emotional or neutral expressions without a clear stance: e.g., "I've been anxious during the pandemic, not knowing when it will end."

4. Other Considerations

- Implied views: While users may not explicitly state their stance, their view can often be inferred from the context. For example, if a user questions the credibility of domestic reports or data, this may imply disagreement with the official narrative. In such cases, annotators can infer the user's position based on contextual cues.
- Sarcasm: Users may express their opinions in a euphemistic or sarcastic manner. For example, a user who exaggerates support for strict government measures (e.g., "Stick to the dynamic Zero-COVID policy for 100 years!") may, in fact, be expressing opposition. Annotators should carefully assess the context and wording to identify the user's true stance.

A.1.2.2 Russia-Ukraine War: Annotation Guidelines

For the issue of Russia-Ukraine War, our goal is to determine whether a user agrees with the Chinese government's propaganda that "Russia is fighting for justice". Below are the types of expressions in the posts that should be classified as *Agree*, *Disagree*, or *Unidentifiable*, respectively.

1. Posts to be labeled as *Agree*

If a user's post explicitly or implicitly supports the view that "Russia is fighting for justice," it should be classified as *Agree*. Specifically, the following types of expressions should be labeled as *Agree*:

- Explicit support for Russia's actions: e.g., "Support Russia, support Putin the Great," or "Russia is fighting a just war."
- Denying the war's aggressive nature: e.g., "Russia is not the aggressor; this conflict should not be defined as a war of aggression."

- Defending or justifying Russia's position: e.g., "Russia is protecting the people of the Donbas region from persecution by the Ukrainian government," or "Russia was forced into this war to safeguard its sovereignty."
- Criticizing Western interference and sanctions: e.g., "Western sanctions will only escalate tensions; I support Russia's countermeasures," or "Western countries use the so-called 'human rights' as a pretext for interfering in the internal affairs of other countries."
- Explicit opposition to Ukraine's position: e.g., "Ukraine is merely a pawn of the West, with its government run by Nazis," "Ukraine started this war and is now facing the consequences."
- Supporting China aligning with Russia: e.g., "China should stand with Russia to oppose Western hegemony," or "We should provide military support to Russia; if Russia falls, our country could be the next target."

2. Posts to be labeled as *Disagree*

If a user's post explicitly or implicitly opposes the view that "Russia is fighting for justice," it should be classified as *Disagree*. Specifically, the following types of expressions should be labeled as *Disagree*:

- Supporting Ukraine's resistance: e.g., "Ukraine has the right to defend its sovereignty and territorial integrity under international law," or "The Ukrainian people are fighting for their country and deserve recognition and respect."
- Condemning Russia's aggression: e.g., "Russia launched an unjust war; this is outright aggression," or "Russia's invasion of Ukraine is driven by territorial ambitions."
- Criticizing Russia's political decisions: e.g., "If Russia dares to use nuclear weapons, it will be acting against the entire world," "Putin's decisions have put Russia in a difficult situation; war is not the solution."
- Supporting sanctions against Russia: e.g., "Sanctions on Russia are not only about aiding Ukraine but also about maintaining global peace and stability."
- Criticizing domestic media narratives: e.g., "Why is our media always biased in favor of Russia? Clearly, it is invading another country," or "As a country historically invaded by others, China should not side with Russia; we should firmly oppose any invasion."

3. Posts to be labeled as *Unidentifiable*

If a user's post does not clearly indicate support for or opposition to the view that "Russia is fighting for justice," it should be classified as *Unidentifiable*. Specifically, the following types of expressions should be labeled as *Unidentifiable*:

- Neutral or ambiguous discussions: e.g., “I don’t know which side to support; this war is too complex,” or “This war has had a huge economic impact, and everyone’s life has been affected.”
- Emotional expressions without a clear stance: e.g., “Regardless of who is right or wrong, I hope the war ends soon and that people can live in peace,” or “There are no winners in war; a peaceful resolution must be found as soon as possible.”

4. Other Considerations

- Implied views: Users may not explicitly state their stance, but their view can be inferred from the context. For example, if a user criticizes domestic biased reporting, this may suggest a disagreement with the propaganda and, therefore, an opposition to the view that “Russia is fighting for justice.” In such cases, annotators can infer the user’s position based on contextual cues.
- Sarcasm: Users may express their opinions in a euphemistic or sarcastic manner. For example, comments like “Ukraine is so fortunate to have a ‘great neighbor’ like Russia. To the countries supporting Russia, I hope you also have such a ‘good neighbor.’” should be interpreted as sarcasm criticizing Russia’s actions. Annotators should carefully assess the context and wording to identify the user’s true stance.

A.1.2.3 Japan’s Discharge Plan: Annotation Guidelines

For the issue of Japan’s discharge plan, we need to determine whether a user agrees with the Chinese government’s propaganda that “Japan is extremely selfish and irresponsible.” Below are the types of expressions in the posts that should be classified as *Agree*, *Disagree*, or *Unidentifiable*, respectively.

1. Posts to be labeled as *Agree*

If a user’s post explicitly or implicitly supports the view that “Japan is extremely selfish and irresponsible,” it should be classified as *Agree*. Specifically, the following types of expressions should be labeled as *Agree*:

- Explicit condemnation of Japan’s discharge actions: e.g., “Japan’s decision to release nuclear wastewater into the ocean is incredibly irresponsible! It will have long-term consequences for marine life and the environment,” or “This is a violation of global trust. The Pacific Ocean does not belong to Japan alone, and such reckless action puts the entire world at risk. This isn’t just an environmental issue; it’s a global one.”
- Belief that the discharge will cause severe health or environmental harm: e.g., “The discharge will pollute the entire Pacific Ocean, and it could spell the end for marine life,” or “Japan’s selfish behavior

will result in people being unable to eat seafood in the future due to the cancer risks associated with consuming it."

- Questioning or criticizing the IAEA's safety assessments: e.g., "The IAEA has been bought by Japan, so it cannot be trusted to report the truth," or "They claim the water is treated, but no one can truly guarantee its safety. The risk to the ecosystem is too high!"
- Supporting the Chinese government position: e.g., "The Chinese government is absolutely right to condemn Japan's reckless actions. Every country should stand united against this!" or "Japan thinks they can get away with poisoning the ocean, but the world is watching. China is speaking for everyone who values the environment and the well-being of humanity."

2. Posts to be labeled as *Disagree*

If a user's post explicitly or implicitly opposes the view that "Japan is extremely selfish and irresponsible," it should be classified as *Disagree*. Specifically, the following types of expressions should be labeled as *Disagree*:

- Belief that Japan's actions are scientifically safe: e.g., "The wastewater has been treated to remove most of the radioactive elements, and what remains is at levels considered harmless to both marine life and humans. This is a much safer solution than storing it indefinitely." or "I don't understand why Japan is being singled out for this. Other countries, including China, have discharged nuclear wastewater as well, and it has been scientifically proven to be safe. This is nothing more than fear-mongering and misinformation."
- Criticism of domestic media's exaggerated coverage: e.g., "Why is the Chinese government focusing so much on Japan's discharge when similar practices occur in other countries without such an outcry? It seems the issue is being politicized rather than addressed from a scientific standpoint," or "It's frustrating to see the media exaggerating the risks simply to score political points. The wastewater is treated and released under strict regulations. This is a scientific issue, not a political one."
- Trust in the scientific assessments of international authorities: e.g., "We must trust the evaluations of authoritative organizations and let the scientific data speak for itself," or "The IAEA has extensively reviewed Japan's plan and concluded that it is scientifically safe. There's no need for panic."

3. Posts to be labeled as *Unidentifiable*

If a user's post does not clearly indicate support for or opposition to the view that "Japan is extremely selfish and irresponsible," it should be classified as *Unidentifiable*. Specifically, the following types of expressions should be labeled as *Unidentifiable*:

- Neutral or ambiguous discussions: e.g., "I'm not sure about the specific impact of nuclear wastewater, but I hope the issue is resolved properly," or "Regardless of where I stand on Japan's

discharge, I believe this is a wake-up call to begin discussions on global standards for nuclear waste management."

- Emotional expressions without a clear stance: e.g., "The news of Japan's wastewater discharge makes me anxious, but the extent of its impact is still unclear."

4. Additional Notes

- Implied views: Users may not explicitly state their stance, but their view can be inferred from the context. For example, if a user criticizes domestic biased reporting, this may suggest a disagreement with the propaganda and, therefore, an opposition to the view that "Japan is extremely selfish and irresponsible." In such cases, annotators can infer the user's position based on contextual cues.
- Sarcasm: Users may express their opinions in a euphemistic or sarcastic manner. For example, comments like "It's amazing how quickly the government can turn a scientific issue into a global crisis. If Japan dumps a tiny bit of treated water into the ocean, it is labeled a 'disaster,' but when we do something similar, it's just 'normal operations.' Such interesting double standards..." should be interpreted as sarcastic criticism of the government's narrative. Annotators should carefully assess the context and wording to identify the user's true stance.

A.1.3 Model Training and Automatic Prediction

Model Training. We use Bidirectional Encoder Representations from Transformers (BERT), a model widely recognized for its strong performance in NLP tasks (Devlin et al., 2018), to automatically classify the universe of posts and comments within our dataset. BERT is a pre-trained deep learning model known for its exceptional performance across various NLP tasks, particularly in understanding the semantic meaning of text and capturing the contextual relationships between words within a sentence (Jawahar, Sagot, and Seddah, 2019; Rogers, Kovaleva, and Rumshisky, 2021).

Following established practices in the literature (Sun et al., 2019; Zhang et al., 2021), we fine-tune three distinct pre-trained BERT models, each tailed to one of the three issues, in order to adapt them to our specific text classification task. For each model, the input consists of annotated data related to a specific topic, and the output is one of three labels: *Agree*, *Disagree*, or *Unidentifiable*. Each model is fine-tuned to predict the stance of a given text based on its corresponding topic and the provided annotations.

Consistent with de Boer et al. (2005), we use the cross-entropy loss function to train the models, iteratively updating the weights based on the training data. For other hyperparameters, including learning rate, batch size, and number of epochs, we use the default values provided by BERT (Devlin et al., 2018).

Model Performance. Consistent with established practices in the literature (Reitermanová, 2010), we divide the annotated sample into three sets: training (80%), validation (10%), and test (10%). The training set is used to fine-tune the BERT model, the validation set is used to select best-performing model and monitor overfitting, and the test set was used to evaluate the final performance of the model. Specifically, we assess the models' performance on the test set using several commonly used metrics: accuracy, precision, recall, F1-score, and the Area Under the ROC Curve (AUC) (Goutte and Gaussier, 2005; Lobo, Jiménez-Valverde, and Real, 2008; Hand, 2009; Yacouby and Axman, 2020). Table A2 presents the accuracy, precision, recall, and F1-score for predicting the *Agree* and *Disagree* labels for the three issues,[1] while Figure A2 displays the AUC for each model.

The results indicate that our trained models perform well across all topics in the text classification tasks:

- **Accuracy**: Accuracy measures the proportion of correct predictions among all predictions. Our models achieved an average accuracy of over 0.75 on the test set. Notably, for the Russia-Ukraine War topic, accuracy reaches approximately 0.9.
- **Precision**: Precision measures the proportion of samples predicted as a given category that are actually correct. Our models achieve excellent precision across all topics, especially for the Russia-Ukraine War topic, where precision is approximately 0.85.
- **Recall**: Recall measures the proportion of relevant samples that are correctly identified by the model. Our models perform well in terms of recall across all topics, particularly for the Russia-Ukraine War topic, where recall is approximately 0.90. On average, the models' recall exceeds 0.75.
- **F1-**score: The F1-score is the harmonic mean of precision and recall, and it is especially useful for evaluating performance on imbalanced datasets. Our models perform well across all topics, with the highest F1-score achieved for the Russia-Ukraine War topic (0.87). Overall, the models' average F1-score across all topics is 0.75.
- **AUC**: AUC reflects a model's ability to distinguish between classes. The value of AUC ranges from 0 to 1, where a value closer to 1 indicates excellent model performance, and a value closer to 0.5 suggests performance no better than random guessing. Our models achieve high AUC scores across all topics, particularly for the Russia-Ukraine War topic, where the AUC exceeds 0.95. For the other two topics, AUC is close to 0.90.

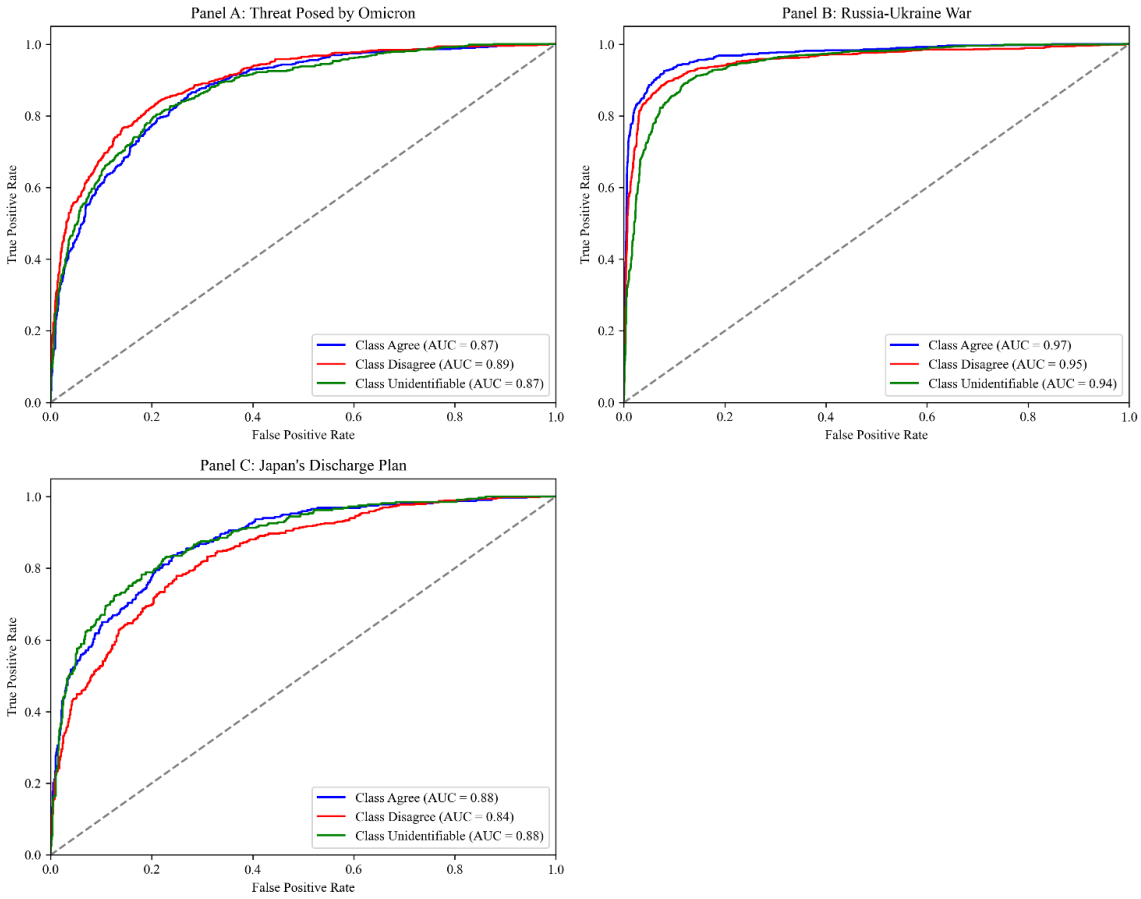
[1] Since text predicted as *Unidentifiable* is treated as noise, we did not report the model's performance on this label.

Table A2: Performance Metrics of NLP Model Across Three Sensitive Issues

Metrics & Model		Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
Agree	Accuracy	0.7357	0.9124	0.7097
	Precision	0.6698	0.8410	0.6838
	Recall	0.7357	0.9124	0.7097
	F1-score	0.7012	0.8753	0.6965
Disagree	Accuracy	0.7552	0.8457	0.7355
	Precision	0.7362	0.8946	0.7401
	Recall	0.7552	0.8457	0.7355
	F1-score	0.7456	0.8695	0.7378

Notes. This table presents the performance metrics (accuracy, precision, recall, and F1-score) of our trained NLP models to classify user posts into *Agree* and *Disagree* categories for each of the three issues.

Figure A2: ROC–AUC for Our Trained Models



Notes. This figure illustrates the ROC curves for the topic-specific NLP models, assessing their performance in multi-class classification tasks. Panels A–C correspond to the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. Each panel presents the model's discriminative ability for three classes: *Agree*, *Disagree*, and *Unidentifiable*, with Area Under the Curve (AUC) scores calculated for each class. Higher AUC scores for a specific class indicate a stronger ability of the model to distinguish posts belonging to that class from posts belonging to other classes, reflecting superior classification performance for that label.

Notably, the models' performance is particularly outstanding for the Russia-Ukraine War topic. A possible reason for this is that user comments on this topic tend to be more straightforward, making it easier to distinguish between pro-Russia, pro-Ukraine, or neutral stances. This clear differentiation of stances helps the BERT model classify texts on this topic more effectively.

Automatic Prediction. We apply the trained models to automatically classify all texts in the corpus. In total, nearly 3 million texts are predicted as *Agree* or *Disagree*. These texts were posted by over 800,000 users, forming our valid sample. The remaining texts were predicted as *Unidentifiable* because the explicit beliefs of the users who posted them could not be clearly identified. These texts were therefore treated as "noise."

A.2 Users' Information Preferences

We identify users' interest areas or information preferences based on the types of content they follow on the Sina Weibo platform. Specifically, we categorize the platform's content into five distinct types:

- **Entertainment and Culture:** Content focusing on entertainment, film, humor, and arts.
- **Lifestyle and Consumption:** Content related to lifestyle, consumption, and fashion.
- **News and Current Affairs:** Content providing news, financial, and military information.
- **Knowledge and Education:** Content focusing on education, science, or technical skills.
- **Public Services and Social Responsibility:** Content centered on public services, charity, and religion.

However, users' profile pages do not explicitly provide information about the types of content they are interested in. Instead, we inferred users' interests based on the types of bloggers they follow. During our study period, the Sina Weibo platform featured 36 types of bloggers, each specializing in producing content related to a specific category to attract audience. As detailed in Table A3, we map each blogger type to its corresponding content category.

Based on this, we create five indicator variables to capture users' information preferences: *Entertainment*, *Lifestyle*, *News*, *Knowledge*, and *Responsibility*. Each variable takes a value of 1 if a user follows at least one type of blogger associated with the corresponding content category; otherwise, it is set to 0. For example, if a user follows at least one charity blogger, public service blogger, or religious blogger, we infer that the user is interested in the Public Services and Social Responsibility category, and the variable *Responsibility* is coded as 1. Conversely, if a user does not follow any charity blogger, public service blogger, or religious blogger, we infer that the user is not interested in the Public Services and Social Responsibility category, and the variable *Responsibility* is coded as 0.

Table A3: Mapping Sina Weibo Blogger Types to Five Content Categories

Type of Content	Type of Bloggers
Entertainment and Culture	entertainment celebrity bloggers, comic bloggers, animation bloggers, music bloggers, movie bloggers, game bloggers, pet bloggers, variety show bloggers, TV drama bloggers, art bloggers
Lifestyle and Consumption	beauty bloggers, makeup bloggers, fashion bloggers, fitness bloggers, sports bloggers, digital bloggers, travel bloggers, parenting bloggers, photography bloggers, beauty bloggers, automotive bloggers, relationship and lifestyle bloggers
News and Current Affairs	current affairs bloggers, financial bloggers, military bloggers, influential bloggers, real estate bloggers, internet bloggers
Knowledge and Education	education bloggers, science popularization bloggers, writer bloggers, health and medical bloggers, legal bloggers
Public Services and Social Responsibility	charity bloggers, public service bloggers, religious bloggers

Notes. This table maps the 36 blogger types recognized by the Sina Weibo platform to the five content categories we establish. Each type of blogger specializes in sharing a specific kind of information to attract an audience.

A.3 Residing Countries of Overseas Information Elites and Their Potential Beliefs

As discussed in the main text, we infer the potential beliefs of overseas information elites based on whether their residing countries align with Chinese government narratives on the three issues. Specifically, information elites residing in countries where mainstream viewpoints conflict with Chinese government propaganda are more likely to be influenced by local media and oppose the Chinese official narratives. In contrast, those living in countries where mainstream views align with Chinese government propaganda are more likely to support it.

To determine these potential beliefs, we categorize foreign countries into two groups for each issue: those whose mainstream views are likely to oppose Chinese government propaganda and those whose mainstream views are more likely to align with it.

The Threat Posed by Omicron. For the Omicron topic, countries or regions are categorized based on their policy responses to COVID-19. We infer that countries or regions which adopted similarly stringent restrictions to China's during this period are more likely to align with Chinese government propaganda. Conversely, countries or regions that adopted relatively relaxed policies are more likely to oppose Chinese official narratives.

In practice, we classify countries or regions using dynamic data from the *Our World in Data* (OWID) platform, with the original data sourced from the Oxford COVID-19 Government Response Tracker (OxCGRT) at the Blavatnik School of Government, University of Oxford (Hale et al., 2021). The OxCGRT systematically collected information on pandemic response measures

adopted by governments worldwide over time. Based on this data, OWID classifies countries into four levels based on the stringency of their stay-at-home measures, with higher levels indicating stricter measures[2]:

- Level 1: No measures
- Level 2: Recommendations to stay home
- Level 3: Required to stay home with exceptions
- Level 4: Required to stay home with minimal exceptions

Our classification criterion is straightforward: During 2022, if a country was classified as Level 3 or Level 4 for more than 30 days, it is considered to have implemented strict control measures similar to China's and is thus more likely to align with the Chinese government propaganda. Otherwise, the country is considered to have implemented relatively relaxed policies and is more likely to conflict with the Chinese official narratives.

Russia-Ukraine War. For the issue of Russia-Ukraine war, we source our data on global stances toward the war from Statista, a platform that provides statistics on the global economy and politics.[3]

Using this data, we categorize countries or regions based on whether they publicly condemned Russia or imposed sanctions during our study period. The rationale behind this classification is that residents in countries or regions that publicly condemned Russia or imposed sanctions on it are more likely to be exposed to media environments where mainstream public opinion strongly opposes Russia's actions. This exposure may lead them to perceive Russia as the aggressor. In contrast, countries did not publicly condemn Russia or impose sanctions may not have a strong stance on this issue. As a result, residents in these countries may not be exposed to media narratives that intensively criticize Russia, potentially leading to weaker opposition toward its actions.

Japan's Discharge Plan. Before launching the plan, the Japanese government collaborated with the International Atomic Energy Agency (IAEA) to ensure it met international safety standards. As nuclear wastewater discharge is a common practice in the world, this event did not attract broad international media attention, and Western mainstream media coverage has been relatively limited.

For this topic, we identify global stances based on international responses documented on Wikipedia.[4] Countries and regions are categorized according to whether they explicitly expressed support for Japan's discharge plan. As noted in Wikipedia, countries such as the United States and the

[2] Source: https://ourworldindata.org/covid-stay-home-restrictions (accessed September 10 2024).
[3] Source: https://www.statista.com/statistics/1293535/global-stance-in-russia-ukraine-war/#statisticContainer (accessed December 15, 2024).
[4] Source: https://zh.wikipedia.org/wiki/福島核廢水排放後續國際反應#政府反应 (accessed December 15, 2024).

United Kingdom have expressed understanding and support for Japan's decision, commending the country for adopting a scientific and responsible approach to handling the nuclear wastewater. In contrast, other countries like China have raised significant concerns about the plan.

We infer that residents of countries or regions supporting Japan's actions are likely to be exposed to media environments that report the facts of the event in an unbiased manner, making them more informed about the issue. As a result, they are also more likely to share information that contradicts Chinese official narratives with their friends in China. Conversely, countries that criticized Japan's actions tend to align with Chinese official narratives. Additionally, countries with a neutral stance may not focus on this issue. Therefore, residents in these countries may either agree with Chinese government propaganda or have limited knowledge about the matter, resulting in minimal influence on the beliefs of their friends in China.

Table A4 shows the categorization of countries and regions for each issue according to the classifications outlined above.

Table A4: Categorization of Countries and Regions based on Their Stances on the Three Issues

Topic	Countries and Regions
Threat Posed by Omicron	**Countries or regions with strict measures:** Albania, Barbados, Belize, Bhutan, Canada, Central African Republic, Congo, Democratic Republic of Congo, Eritrea, Fiji, Gabon, Greece, Guyana, Honduras, India, Jamaica, Kiribati, Kosovo, Laos, Lebanon, Lesotho, Macao, Madagascar, Malawi, Mozambique, Myanmar, Nigeria, Papua New Guinea, Peru, Philippines, Romania, Russia, Rwanda, Seychelles, Solomon Islands, South Sudan, Suriname, Togo, Tonga, Vanuatu, Vietnam, Yemen, Zimbabwe **Countries or regions with relax measures:** Other countries
Russia-Ukraine War	**Countries or regions that publicly condemned Russia or impose sanctions:** United States, United Kingdom, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Albania, Australia, Bahamas, Canada, Japan, Kosovo, Montenegro, New Zealand, North Macedonia, Norway, Singapore, South Korea, Switzerland, Taiwan (China) **Countries or regions that did not publicly condemned Russia or impose sanctions:** Other countries
Japan's Discharge Plan	**Countries or regions that publicly support Japan's actions:** United States, United Kingdom, Fiji, Palau, Papua New Guinea, New Zealand, Australia, France, Cook Islands, Argentina, Saudi Arabia, Jordan, Egypt, Indonesia, Netherlands, Morocco, Tunisia, Kuwait, Oman, Bahrain, Mongolia **Countries or regions that did not publicly support Japan's actions:** Other countries

Notes. This table lists the categorization of countries and regions based on their potential stances on the three issues.

In the main text, we highlight a notable example of the divergent stances of the USA and Canada on Japan's discharge plan. The USA explicitly endorsed Japan's plan, expressing understanding and support, whereas Canada took no clear stance, with its local media largely ignoring the issue. Table A6 reveals that users connected to information elites in the USA exhibit minimal differences in most observable characteristics compared to those connected to elites in Canada. Although some differences are statistically significant, their magnitudes remain modest.

With users connected to U.S.-based elites as the treatment group and those connected to Canada-based elites as the control group, we estimate Equations (1) and (2) to assess whether U.S.-based elites exert a greater influence on their friends' beliefs than their Canada-based counterparts. Column (1) of Table A7 shows that the coefficient for the belief variable is positive but statistically insignificant, suggesting that both groups hold similar views on Japan's discharge plan. However, Column (2) of Table A7 indicates that the former are significantly more likely to shift from agreement to disagreement with Chinese government propaganda over time than the latter. These findings suggest that U.S.-based elites may have significantly shaped their friends' beliefs by sharing relevant, uncensored information.

B. Coarsened Exact Matching

In the main text, we use the Coarsened Exact Matching (CEM) in two different identification strategies to strengthen the credibility of our causal claims. In Section 4.2, we use CEM to balance observable characteristics between users with high versus low exposure to lockdowns (through their friends' locations). By matching each high-exposure user to a similar low-exposure user, we create a natural experiment in which different levels of lockdown exposure are as-if randomly assigned conditional on the same covariates. In Section 4.3, CEM is used to balance the observable differences between treatment and control users. This section first briefly describes the CEM procedure and then details the specific implementation choices for each analysis.

B.1 Overview of the CEM Procedure

CEM, which aims to match each treated observation with one or more suitable control, is widely used in the literature (Iacus, King, and Porro, 2008; 2012). Unlike matching methods based on

estimating the propensity score, CEM is a non-parametric procedure, making it particularly suitable for our context, where most covariates are categorical variables. The key steps are:

1. Covariate Selection and Coarsening

The first step involves selecting the covariates to be matched and coarsening them into discrete categories. We use all control variables from column (6) of Table 2 in the main text as matching covariates. We coarsen continuous variables, such as the number of friends, account registration duration, and VIP level, into three categories based on percentiles. The remaining covariates, being categorical, do not require further coarsening.

2. Stratum Creation

In the second step, we create strata that cover the entire support of the joint distribution of the selected covariates. For example, suppose there are three covariates (X1, X2, X3) with coarsened categories of (2, 3, 5), this results in a total of 2 * 3 * 5 = 30 strata. In our case, this step generates 27,648 strata.

3. Matching Treatment and Control Users

In the third step, each user is allocated to a unique stratum, and treatment users are matched with control users within the same strata. If no control users are available in a given stratum, the treatment users in that strata remain unmatched.

The final matching performance is evaluated using the L1 statistic, which measures the overall imbalance in the sample. The L1 statistic ranges from 0 to 1, with a value of 0 indicating perfect matching and a higher value signifying greater imbalance between groups.

B.2 CEM Implementation in Section 4.2

In Section 4.2 of the main text, we use CEM to balance observable characteristics between users with high versus low friend exposure to lockdowns. To do this, we match each high-exposure user (whose first-stage instrument lies in the top quintile) to a similar low-exposure user (whose first-stage instrument lies in the remaining quintiles).

Each stratum is forced to contain the same number of high-exposure and low-exposure users. The k2k option in Stata's *cem* command accomplishes this by pruning observations from a CEM solution within each stratum until the solution contains the same number of high-exposure and low-exposure users within all strata. For strata with multiple eligible low-exposure users, one low-exposure user is randomly selected; high-exposure users without an eligible low-exposure user are dropped. This yields a matched sample of equal numbers of high-exposure and low-exposure users.

Before matching, the L1 statistic is 0.455 for the Omicron discussion sample, 0.552 for the Russian-Ukraine War sample, and 0.590 for the Japan's Discharge Plan sample. After matching, the L1 statistic falls to zero, confirming excellent global balance.

B.3 CEM Implementation in Section 4.3

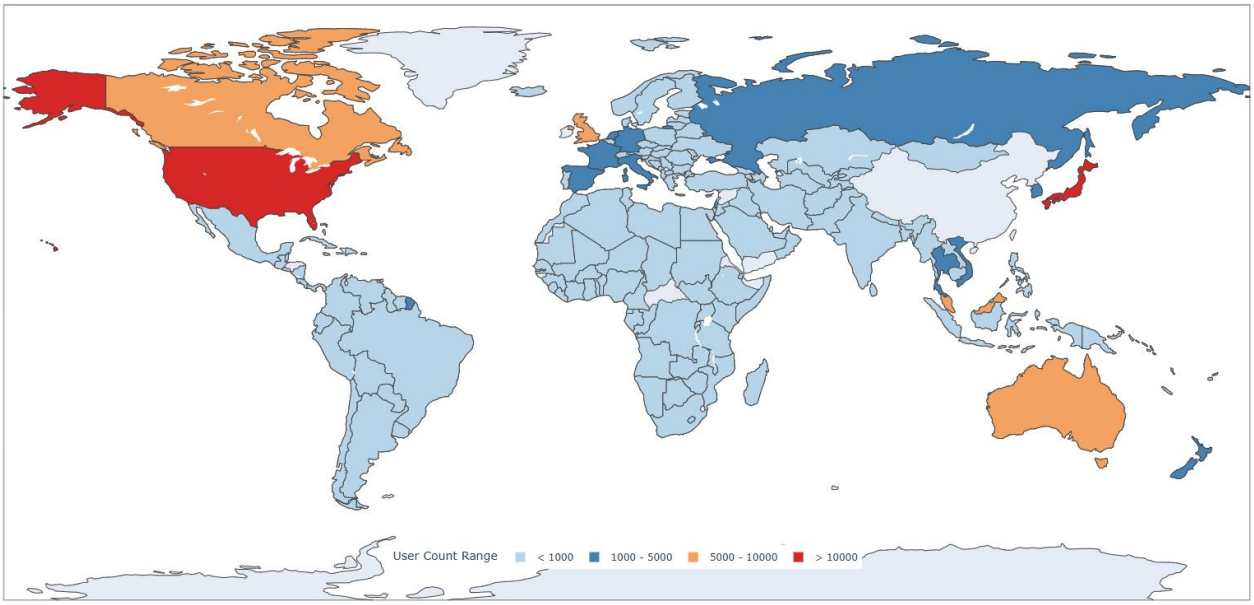
In Section 4.3 of the main text, we use CEM to balance observable characteristics between treated users (users connected to information elites) and control users (users not connected to information elites).

We allow strata to include different numbers of treated and control users. Since the treatment group represents just 10% of the sample, enforcing k-to-k matching would cap the matched sample at 20% of the original observations, dramatically reducing statistical power. Allowing many-to-one matching preserves balance while maximizing sample size and efficiency.

Before matching, significant differences exist between treatment and control users across the selected covariates, as reflected by L1 statistics greater than 0.6. Specifically, the L1 statistic is 0.613 for the Omicron discussion sample, 0.635 for the Russian-Ukraine War sample, and 0.660 for the Japan's Discharge Plan sample. However, after matching, the differences are almost completely eliminated, with the L1 statistic equal to zero in all cases.

C. Additional Figures and Tables

Figure A3: Global Distribution of Overseas Information Elites



Notes. This map visualizes the global distribution of overseas information elites. Countries are color-coded based on the concentration of these elites: red indicates over 5,000 elites, yellow denotes 2,000–5,000, darker blue represents 1,000–2,000, and lighter blue signifies fewer than 1,000.

Figure A4: Distribution of VPN Elites across Provinces in China

Notes. This map illustrates the spatial distribution of VPN information elites across provinces in China, with darker shades representing higher concentrations.

Table A5: IV Estimates of the Effect of Connections to Information Elites on Shifts in User Beliefs Regarding Politically Sensitive Issues

Dependent Variable:	*Belief Shift*		
	Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
	(1)	(2)	(3)
Connected to Elites	-0.002	0.020*	-0.003
	(0.007)	(0.012)	(0.009)
Observations	403966	299382	121136
R-Square	0.005	0.003	0.003
Mean Y	0.046	0.055	0.009
Baseline Controls	Yes	Yes	Yes

Notes. This table reports IV estimates of the effect of connections to information elites on shifts in user beliefs regarding politically sensitive issues. Specifically, we estimated Equation (4) but replace the dependent variable with *Belief Shift*, with estimated coefficients (β^{IV}) presented in the table. Columns (1)–(3) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Observable Characteristic Differences Between Users Connected to U.S.- and Canada-Based Information Elites Regarding Japan's Discharge Plan

	Connected to U.S.-based elites		Connected to Canada-based elites			
Variables	Obs.	Mean	Obs.	Mean	Diff in Mean	*p*-value
Male	667	0.439	226	0.416	0.023	0.541
Reported School	667	0.247	226	0.226	0.022	0.511
Number of Friends	667	47.123	226	49.398	-2.275	0.475
High Followers	667	0.841	226	0.889	-0.048	0.076
High Following	667	0.541	226	0.496	0.046	0.235
Account Age (years)	667	9.004	226	9.212	-0.208	0.458
VIP Level	667	2.426	226	2.469	-0.043	0.829
Has iPhone	667	0.364	226	0.341	0.024	0.523
Entertainment	667	0.738	226	0.721	0.016	0.630
Lifestyle	667	0.465	226	0.389	0.075	0.049
News	667	0.510	226	0.425	0.085	0.027
Knowledge	667	0.574	226	0.558	0.017	0.662
Responsibility	667	0.079	226	0.049	0.031	0.121

Notes. This table presents the observable differences between users connected to information elites in the United States and those connected to information elites in Canada, within discussions about Japanese nuclear wastewater discharge plan. The observables include all control variables as specified in Column (6) of Table 2.

Table A7: Beliefs and Belief Shifts Regarding Japan's Discharge Plan: Connections to U.S.- vs. Canada-based Information Elites

Dependent Variable:	*Belief*	*Belief Shift*
	(1)	(2)
U.S. Elite Connection	0.008	0.024***
	(0.016)	(0.007)
Observations	959	893
R-Square	0.047	0.044
Mean Y	0.044	0.018
Prior: Agree Gov	No	Yes
Baseline Controls	Yes	Yes

Notes. This table examines the differences in beliefs and belief shifts about Japan's discharge plan between users connected to information elites in the United States and those connected to information elites in Canada. The treatment group consists of users connected to information elites in the United States (Connection to U.S.-based Elites =1), while the control group consists of those connected to information elites in Canada (Connection to U.S.-based Elites =0). Column (1) presents the estimated coefficients (β) from Equation (1) in the main text. Column (2) evaluates the impact of connections to U.S.-based elites on user belief shifts, presenting the estimated coefficients (δ) from Equation (5), with the sample restricted to users whose prior beliefs align with government propaganda. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Robustness Checks through Flexible Definitions of User Beliefs

Dependent Variable:	*Belief: 75%*					
	Threat Posed by Omicron		Russia-Ukraine War		Japan's Discharge Plan	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A.						
Connected to Elites	0.027***		0.030***		0.015***	
	(0.002)		(0.003)		(0.003)	
Connection Intensity		0.093***		0.102***		0.035***
		(0.010)		(0.011)		(0.011)
Observations	403966	403966	299382	299382	121136	121136
R-Square	0.012	0.012	0.014	0.014	0.009	0.008
Mean Y	0.239	0.239	0.215	0. 215	0.029	0.029
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable:	*Belief: 90%*					
Panel B.						
Connected to Elites	0.025***		0.024***		0.013***	
	(0.002)		(0.003)		(0.002)	
Connection Intensity		0.088***		0.081***		0.031***
		(0.010)		(0.011)		(0.010)
Observations	403966	403966	299382	299382	121136	121136
R-Square	0.012	0.012	0.012	0.012	0.008	0.007
Mean Y	0.231	0.231	0.197	0.197	0.026	0.026
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table assesses the robustness of our results by flexibly defining user beliefs. In Panel A, a user is classified as disagreeing with government narratives if over 75% of their posts are predicted as *Disagree*, while in Panel B, the threshold is set at 90%. We then estimate Equation (1), with Columns (1)–(2), (3)–(4), and (5)–(6) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

References

Ainslie, Joshua, Santiago Ontanon, Chris Alberti, Vaclav Cvicek, Zachary Fisher, Philip Pham, Anirudh Ravula, Sumit Sanghai, Qifan Wang, and Li Yang, "ETC: Encoding Long and Structured Inputs in Transformers," arXiv Working Paper, 2020. https://doi.org/10.48550/arXiv.2004.08483.

de Boer, Pieter-Tjerk, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein, "A Tutorial on the Cross-Entropy Method," *Annals of Operations Research*, 134 (2005), 19–67.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv Working Paper, 2018. https://doi.org/10.48550/arXiv.1810.04805.

Egger, Roman, and Enes Gokce, “Natural Language Processing (NLP): An Introduction,” in *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications*, Roman Egger, ed. (Cham: Springer International Publishing, 2022), 307–334.

Goutte, Cyril, and Eric Gaussier, “A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation,” in *Advances in Information Retrieval*, David E. Losada and Juan M. Fernández-Luna, eds. (Berlin, Heidelberg: Springer, 2005), 345–359.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow, “A global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker),” *Nature Human Behaviour*, 5 (2021), 529–538.

Hand, David J., “Measuring Classifier Performance: A Coherent Alternative to the Area Under the ROC Curve,” *Machine Learning*, 77 (2009), 103–123.

Iacus, Stefano Maria, Gary King, and Giuseppe Porro, “Matching for Causal Inference without Balance Checking,” available at SSRN (2008): http://dx.doi.org/10.2139/ssrn.1152391.

——, “Causal Inference without Balance Checking: Coarsened Exact Matching,” *Political Analysis*, 20 (2012), 1–24.

Jawahar, Ganesh, Benoît Sagot, and Djamé Seddah, “What Does BERT Learn about the Structure of Language?,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Anna Korhonen, David Traum, and Lluís Màrquez, eds. (Florence, Italy: Association for Computational Linguistics, 2019), 3651–3657.

Li, Yuan, Benjamin Rubinstein, and Trevor Cohn, “Exploiting Worker Correlation for Label Aggregation in Crowdsourcing,” in *Proceedings of the 36th International Conference on Machine Learning*, Kamalika Chaudhuri and Ruslan Salakhutdinov, eds. (Long Beach, California: PMLR, 2019), 3886–3895.

Lobo, Jorge M., Alberto Jiménez-Valverde, and Raimundo Real, “AUC: A Misleading Measure of the Performance of Predictive Distribution Models,” *Global Ecology and Biogeography*, 17 (2008), 145–151.

Reitermanová, Z., “Data Splitting,” in *WDS’10 Proceedings of Contributed Papers*, J. Safrankova and J. Pavlu, eds. (Prague: Matfyzpress, 2010), 31–36.

Rogers, Anna, Olga Kovaleva, and Anna Rumshisky, “A Primer in BERTology: What We Know about How BERT Works,” *Transactions of the Association for Computational Linguistics*, 8 (2021), 842–866.

Tabassum, Ayisha, and Dr Rajendra R. Patil, “A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing,” *International Research Journal of Engineering and Technology (IRJET)*, 7(2020), 4864-4867.

Yacouby, Reda, and Dustin Axman, “Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models,” in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, Steffen Eger, Yang Gao, Maxime Peyrard, Wei Zhao, and Eduard Hovy, eds. (Online: Association for Computational Linguistics, 2020), 79–91.

Yu, Lili, Daniel Simig, Colin Flaherty, Armen Aghajanyan, Luke Zettlemoyer, and Mike Lewis, "MEGABYTE: Predicting Million-Byte Sequences with Multiscale Transformers," *Advances in Neural Information Processing Systems*, 36 (2023), 78808–78823.

Supplementary Information for

The Power of Social Networks: Information Elites and the Spread of Politically Sensitive Information under Media Censorship

Xuebo Wang,[1] Hong Yang[2]

December 17, 2025

[1] School of Economics, Shanghai University of Finance and Economics. E-mail: wang.xuebo@mail.shufe.edu.cn.
[2] School of Economics, Shanghai University of Finance and Economics. E-mail: yanghong@stu.sufe.edu.cn.

S1 Impact of Connections to Elites on User Belief—Connection Intensity and Type

In the main text, users are categorized into the treatment group if they have at least one information elite friend—either an overseas elite or a domestic VPN elite—while those without such connections are placed in the control group. This classification addresses the extensive margin of elite exposure; however, it does not consider the variation in the intensity of exposure among users in the treatment group. In this section, we further investigate the intensity of exposure to information elites, along with the heterogeneity based on elite type.

We construct two measures of treatment intensity. The first measure, Connection Intensity, is defined as the proportion of a user's friends who are information elites. By replacing the binary treatment indicator with this continuous measure, we re-estimate Equation (1) from the main text. The results are reported in the first row of Table S1. Column (1) shows that, for the Omicron issue, the coefficient is 0.088, indicating that a 10% increase in the connection intensity corresponds to a 0.88 percentage point increase in the likelihood of perceiving Omicron as mild. Columns (3) and (5) present similar results for the other issues.

As a second approach, we categorize users into four groups based on the number of information elite friends they have: 0, 1, 2, and 2+ (with users having no elite friends, labeled 0, serving as the reference group). We then re-estimate Equation (1), with the results presented in last three rows. The results remain consistent, showing that users with more information elite friends are more likely to disagree with government propaganda.

We next explore heterogeneity by type of information elite. Treatment users are divided into two mutually exclusive subgroups: (i) those connected to overseas elites and (ii) those connected to VPN elites.[1] Users with no elite friends remain the reference group. We estimate separate treatment effects for these two subgroups by replacing the single treatment indicator in Equation (1) with two indicators. Results are presented in Table S2. The findings indicate that users in both subgroups are significantly more likely to disagree with government propaganda.

[1] Here, we exclude users who are connected to both types of information elites.

Table S1: The Effects of Connections to Information Elites on User Beliefs Regarding Politically Sensitive Issues, by Connection Intensity

Dependent Variable:	*Belief*					
	Threat Posed by Omicron		Russia-Ukraine War		Japan's Discharge Plan	
	(1)	(2)	(3)	(4)	(5)	(6)
Connection Intensity	0.088***		0.074***		0.031***	
	(0.010)		(0.011)		(0.010)	
Num. of Elite Friends: 1		0.019***		0.016***		0.012***
		(0.003)		(0.003)		(0.003)
Num. of Elite Friends: 2		0.029***		0.022***		0.016***
		(0.005)		(0.007)		(0.006)
Num. of Elite Friends: 2+		0.052***		0.065***		0.014*
		(0.005)		(0.009)		(0.008)
Observations	403966	403966	299382	299382	121136	121136
R-Square	0.012	0.012	0.011	0.012	0.007	0.008
Mean Y	0.230	0.230	0.194	0.194	0.026	0.026
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table examines whether the effect of connections to information elites increases with treatment intensity. Specifically, the first row replaces the binary treatment indicator in Equation (1) with *Connection Intensity*, defined as the proportion of a user's friends who are information elites, and reports the corresponding estimated coefficients (β). The last three rows replace the binary treatment indicator with three mutually exclusive indicator variables for having exactly 1, exactly 2, or 2+ information-elite friends. Users with zero information-elite friends serve as the omitted reference category. The table reports the estimated coefficients on these three indicators. Columns (1)–(2), (3)–(4), and (5)–(6) report the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table S2: The Effects of Connections to Information Elites on User Beliefs Regarding Politically Sensitive Issues, by Connection Type

Dependent Variable:	*Belief*		
	Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
	(1)	(2)	(3)
Connected to Overseas Elites	0.018***	0.024***	0.011***
	(0.003)	(0.004)	(0.003)
Connected to VPN Elites	0.025***	0.008*	0.016***
	(0.004)	(0.005)	(0.004)
Observations	393187	295702	120379
R-Square	0.011	0.011	0.008
Mean Y	0.229	0.193	0.026
Baseline Controls	Yes	Yes	Yes

Notes. This table examines heterogeneity in treatment effects by the type of information elite. We replace the binary treatment indicator in Equation (1) with two mutually exclusive indicator variables for connected to overseas elites and connected to VPN elites. Users with zero information-elite friends serve as the omitted reference category. The table reports the estimated coefficients on these two indicators. Columns (1)–(3) report the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

S2 COVID-19 Lockdown Increased Censorship Circumvention

In the main text, we exploit the intensity of China's 2022 zero-COVID lockdowns as an exogenous shock to the demand for uncensored information, which allows us to address the endogenous selection of users into social networks that include information elites. This identification strategy rests on the well-documented surge in censorship-circumvention behavior during periods of strict lockdown (Chang et al., 2022). For example, during the Shanghai lockdown in April 2022, Twitter visits originating from the city surged by approximately 41 % as residents sought to bypassed internet controls.[2] This supplementary section provides additional institutional background by outlining three plausible mechanisms that explain the observed surge.

First, prolonged lockdowns created a significant disparity between official narratives and citizens' lived experiences, severely undermining the credibility of state-controlled information and compelling citizens to seek external sources for verification. In 2022, the dominant Omicron variant was characterized by rapid transmission but predominantly mild symptoms. While most countries had

[2] See https://www.wired.com/story/shanghai-lockdown-china-censorship/.

shifted toward living with the virus, China adhered rigidly to its zero-COVID strategy, exemplified by Shanghai's citywide lockdown, which lasted over two months in the first half of the year. Such extended and disruptive measures may lead a significant portion of the population to question the rationality and sustainability of these policies, incentivizing them to bypass censorship to access international reporting and assess the trustworthiness of their government's claims.

Second, the ruthless efficiency of media censorship in suppressing extreme individual tragedies paradoxically fueled intense impulses to express outrage on foreign websites. During the lockdowns, numerous distressing incidents—such as pets being forcibly culled by pandemic workers, patients with chronic illnesses being denied medication, pregnant women experiencing miscarriages due to lack of timely care, and reported suicides—were censored on domestic social media platforms. Nevertheless, videos, photos, and firsthand accounts of these events circulated widely on overseas platforms like Twitter and YouTube via VPNs.[3] This suggests that the combination of stringent lockdown measures and pervasive censorship pushed citizens to circumvent the Great Firewall to voice their anger on foreign websites.

Third, citizens may also have circumvented censorship to access entertainment content due to lack of mobility and boredom during quarantine and lockdown. Prior research has demonstrated that entertainment-driven censorship circumvention often serves as a gateway to accessing censored politically sensitive content (Hobbs and Roberts, 2018). This spillover effect is particularly pronounced in authoritarian regimes such as China, where a large amount of political information has long been censored. Once circumvention tools are adopted—even initially for apolitical reasons—users gain unrestricted access to a broad spectrum of previously inaccessible information.

S3 Measuring China's 2022 Lockdown Intensity across Regions

In this section, we first explain how to measure China's 2022 lockdown intensity across provinces by examining reductions in human mobility relative to normal times. We then show that this intensity is primarily influenced by local pandemic control pressures but uncorrelated with pre-existing regional characteristics, supporting the exogeneity of this variation.

S3.1 Measuring Provincial Lockdown Intensity Using Human Mobility Data

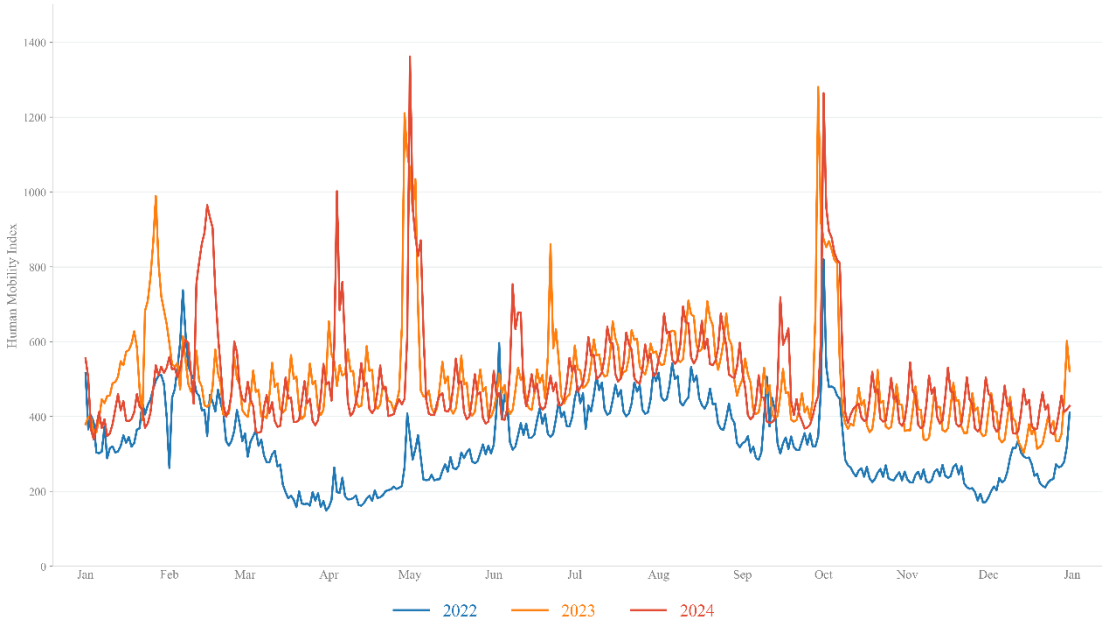
Human mobility data are publicly available from Baidu Migration Big Data (https://qianxi.baidu.com/#/), which tracks real-time migration (including daily national human mobility index and daily provincial inflows/outflows index) using location data from Baidu Maps, a

[3] See, for example, https://edition.cnn.com/2022/04/08/china/shanghai-corgi-death-china-covid-intl-hnk/index.html.

leading Chinese mapping service. These data have been widely used in studies of COVID-19 containment measures, where reductions in mobility serve as a proxy for lockdown intensity (Kraemer et al., 2020; Chang et al., 2022).

Figure S1 shows the daily national human mobility index in China for 2022 (blue curve), 2023 (orange curve), and 2024 (orange-red curve). This figure offers several key insights: First, human mobility in 2022 was significantly lower throughout nearly the entire year compared to 2023 and 2024, following the complete lifting of all COVID-19 restrictions at the end of 2022. This indicates that China's strict zero-COVID policy in 2022 substantially suppressed human mobility.

Figure S1: Daily National Human Mobility Index (2022-2024)

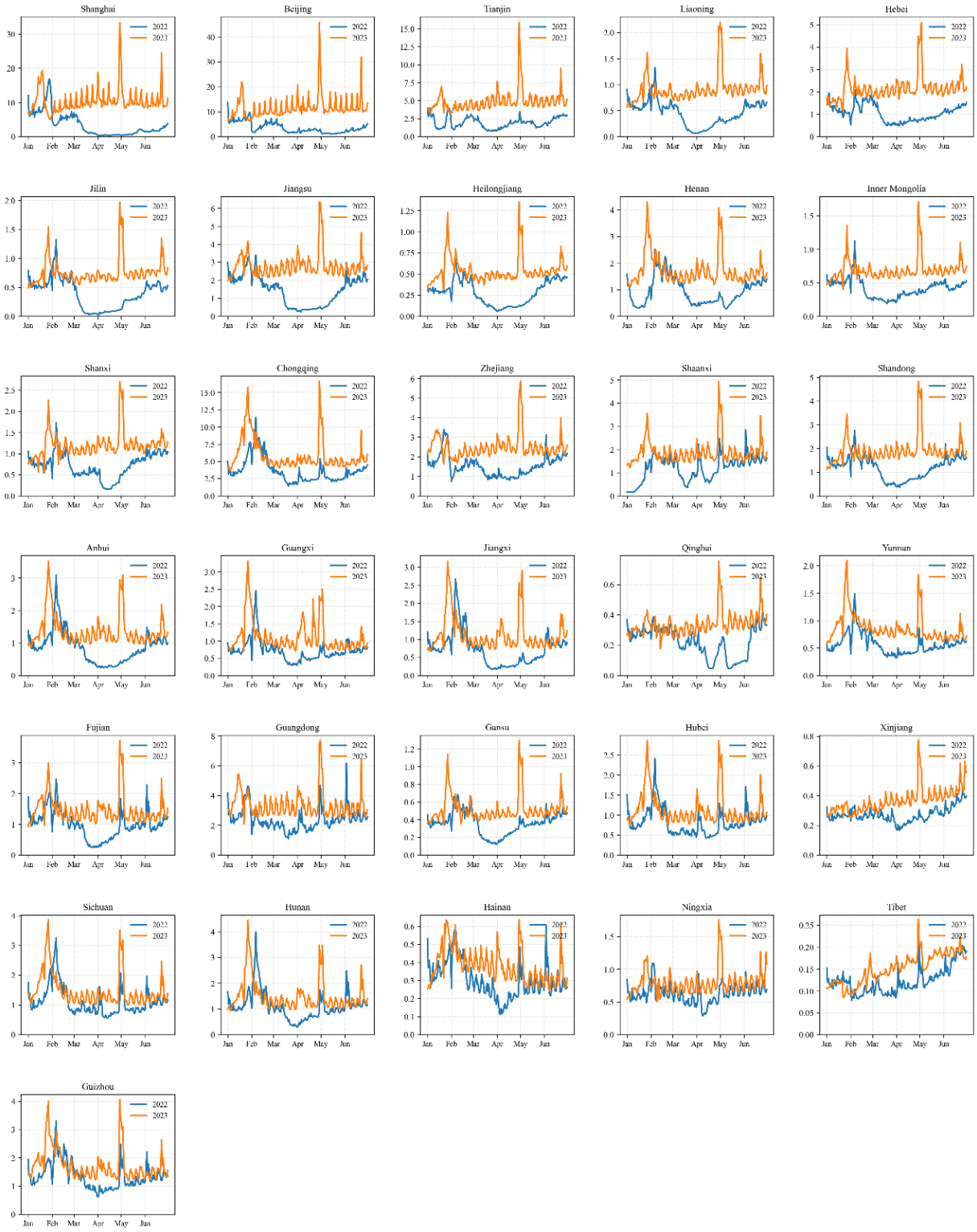


Notes. This figure plots the daily national human mobility index in China for 2022 (blue line), 2023 (orange line), and 2024 (orange–red line). The index is constructed by Baidu Migration Big Data and reflects nationwide travel intensity (higher values indicate greater population flows).

Second, the temporal patterns of human mobility in 2023 and 2024 are highly similar, suggesting that, in the absence of pandemic-related restrictions, mobility exhibits strong year-on-year consistency in corresponding periods. This finding supports the validity of using relative reductions in human mobility compared to the same period in subsequent years as a proxy for the intensity of pandemic control measures.

Finally, the reduction in human mobility in 2022 was particularly pronounced in the first half of the year, driven by large-scale lockdowns, including the two-month city-wide lockdown in Shanghai and stringent measures in provinces such as Beijing, Tianjin, and Liaoning. Provincial-level mobility

Figure S2: Daily Human Mobility Index Across Chinese Provinces (First Half of 2022 vs. First Half of 2023)



Notes. This figure plots daily human mobility index for each Chinese province during January – June of 2022 (blue line) and the same period in 2023 (orange line). The y-axis scales differ across provinces to better illustrate relative changes within each province. The index is constructed by Baidu Migration Big Data and reflects provincial travel intensity (higher values indicate greater population flows).

data confirm much sharper mobility declines in these regions during the first half of 2022 compared to the same period in 2023 (Figure S2). Although population mobility declined across all provinces in the first half of 2022 compared to the same period in 2023, the extent of the decline varied significantly across provinces. This reflects the differing levels of lockdown measures implemented in each region.

We therefore measure provincial lockdown intensity as the average reduction in human mobility in the first half of 2022 relative to the first half of 2023. Specifically, the lockdown intensity in a province p is given by:

$$Lockdown\ intensity_p = \frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}}\frac{M^p_{2023,d} - M^p_{2022,d}}{M^p_{2023,d}}, \tag{X1}$$

where $\mathcal{D}$ denotes the set of all dates from January 1 to June 30. $M^p_{year,d}$ is the daily mobility index in province p on date d in the respective year. The year 2023 serves as the counterfactual baseline, as China had fully abandoned the zero-COVID policy and shifted to complete reopening. $Lockdown\ intensity_p$ thus captures the average reduction in mobility relative to this baseline, with higher values reflecting stricter lockdown measures.

S3.2 Evidence on the Exogeneity of Lockdown Intensity Variation

In the main text, we argue that provincial lockdown intensity is driven primarily by local pandemic control pressures rather than regional attributes. To validate this claim, we regress provincial lockdown intensity on a vector of pre-determined (2021) socioeconomic characteristics and report the results in Column (1) of Table S3. The regressors include (log) GDP, (log) GDP per capita, (log) fiscal expenditure per capita, share of primary industry in GDP, share of secondary industry in GDP, average years of schooling (separately for urban and rural populations), and number of hospital beds per 10,000 persons. All coefficients are statistically insignificant, indicating no systematic correlation between lockdown intensity and pre-pandemic provincial characteristics.

In Column (2), we add the (log) cumulative number of confirmed Omicron cases per 10,000 people in the first half of 2022. As expected, its coefficient is positive and statistically significant at 1% level, while the other coefficients remain insignificant. This further supports the conclusion that provincial lockdown intensity is primarily driven by local pandemic control pressures. Finally, Column (3) regresses case pressure itself on the same set of 2021 provincial characteristics. Once again, no coefficient is statistically significant, confirming that the geographic distribution and severity of the 2022 Omicron wave were largely unpredictable based on observable pre-pandemic traits.

Taken together, these results suggest that the variation in lockdown intensity was primarily driven by the exogenous arrival and spread of Omicron rather than by regional attributes. This supports our

identification assumption that the lockdowns acted as an exogenous shock on residents' demand for uncensored information.

Table S3: Correlation between Provincial Lockdown Intensity and Socioeconomic Characteristics

Dependent Variable:	*Lockdown intensity*		ln (Num. of confirmed cases per 10k persons)
	(1)	(2)	(3)
ln (Num. of confirmed cases per 10k persons)		0.049***	
		(0.013)	
ln (GDP)	0.035	0.038	0.001
	(0.053)	(0.053)	(0.261)
ln (GDP per capita)	-0.116	-0.109	-0.029
	(0.129)	(0.123)	(0.551)
ln (Fiscal expenditure per capita)	0.155	0.134	0.260
	(0.117)	(0.111)	(0.683)
Share of primary industry	-0.849	-0.590	-6.220
	(0.735)	(0.722)	(4.391)
Share of secondary industry	-0.142	-0.024	-3.701
	(0.272)	(0.257)	(2.706)
Average years of education in urban areas	0.072	0.068	-0.218
	(0.055)	(0.052)	(0.544)
Average years of education in rural areas	0.061	0.049	0.179
	(0.047)	(0.042)	(0.277)
Num. of hospital beds per 10k persons	0.001	-0.000	0.010
	(0.003)	(0.003)	(0.008)
Observations	31	31	31
R-Square	0.585	0.642	0.309
Mean Y	0.363	0.363	0.351

Notes. This table reports cross-sectional associations between provincial lockdown intensity, local Omicron epidemic severity, and a set of predetermined socioeconomic characteristics in 2021. The dependent variable in columns (1) and (2), Lockdown Intensity, is measured as the average reduction in human mobility during the first half of 2022 relative to the same period in 2023 (higher values indicate stricter lockdowns). Column (1) regresses lockdown intensity on the following predetermined provincial characteristics (listed in the order they appear in the table): (log) GDP, (log) GDP per capita, (log) fiscal expenditure per capita, share of primary industry in GDP, share of secondary industry in GDP, average years of schooling for urban residents, average years of schooling for rural residents, and number of hospital beds per 10,000 persons. Column (2) adds the (log) cumulative confirmed Omicron cases per 10,000 persons in the first half of 2022 as an additional regressor. Column (3) regresses the log cumulative confirmed Omicron cases per 10,000 persons in the first half of 2022 on the same set of 2021 provincial characteristics. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

S4 Heterogenous Treatment Effects by Geographic Proximity and Relationship Strength

Geographic proximity and relational closeness between information elites and their friends may play a critical role in the dissemination of uncensored information through social networks. This section explores how these factors affect the treatment effects based on users' proximity to information elites.

We begin by examining the heterogeneous effects of users' geographic proximity to domestic VPN information elites.[4] Intuitively, if these elites are geographically closer to their friends, they can meet more frequently and share uncensored information in person. Private conversations, in particular, provide a relatively safe space for exchanging uncensored content under strict media censorship. Therefore, sharing uncensored information would be more convenient and secure for friends who live closer, compared to those who live farther apart.

To assess geographic proximity, we identify the primary residence of domestic users based on the IP address that appears most frequently in their posts. Since the IP address data is only available at the provincial level, we determine whether a user resides in the same province as their VPN-elite friends. We acknowledge that this measure is somewhat imprecise and has limitations.

Based on this, we categorize the sample into three subgroups: (1) *Group TC*: Treatment users connected to VPN elites living in the same province (*Colocation=1*), (2) *Group TN:* Treatment users connected to VPN elites in different provinces (*Non-Colocation=1*), and (3) *Group C:* Control users not connected to any information elites, serving as the reference group.

We hypothesize that users in *Group TC* are more likely to be influenced by their elite friends and deviate from government propaganda, compared to those in *Group TN*.

We estimate Equations (6) and (7) from the main text for each issue and report the results in the first two rows of Table S4 and Table S5, respectively. In Table S4, the coefficients for *Colocation* are generally larger than those for *Non-Colocation*, suggesting that users residing in the same province as their VPN-elite friends are more likely to be influenced. However, the differences are not statistically significant. The corresponding results in Table S5 follow a similar pattern. While these findings suggest that geographic proximity might increase the likelihood of influence, the evidence is relatively weak due to data limitations.

Next, we explore the heterogeneous treatment effects based on the strength of users' relationship with their information elite friends. We hypothesize that users who are socially closer to, or more

[4] Since overseas information elites are based abroad, we do not consider the geographic proximity between them and their friends in China.

frequently interact with, their information elite friends are more likely to be influenced by these friends and deviate from government propaganda. To test this hypothesis, we collect all public comments (over 400 million in total) under the original posts from our sample users to map the social interaction networks. Using this data, we measure the strength of friendship links based on the frequency of interactions. To assess the strength of the relationship between treatment users and their information elite friends, we construct two indicators: (1) whether there has been any social interaction between a treatment user and information elites (labeled *Ever Interacted* or *Never Interacted*), and (2) whether information elites are among the user's top 10 closest friends (labeled *Top 10 Closest Friends* or *Non-Top 10 Closest Friends*).

Note that users could communicate with their information elite friends through various channels other than Weibo, such as face-to-face conversations, phone calls, WeChat, or other private messaging apps. Due to data limitations, we use the frequency of interaction on Weibo as a proxy for overall social interaction intensity. While this measure is not perfect, it may not pose a significant issue. As the literature suggests, different communication channels are complements (Barwick et al., 2023). Therefore, users who frequently interact with their elite friends on Weibo are also more likely to engage with them via other channels. This suggests that the frequency of interaction on Weibo can effectively serve as an indicator of overall social interaction intensity.

Based on these measures, we divide the sample into three subgroups: (1) *Group TS:* Treatment users with a strong relationship with their information elite friends (i.e., *Ever Interacted=1* or *Top 10 Closest Friends=1)*, (2) *Group TW:* Treatment users with a weak relationship with their elite friends (i.e., *Never Interacted=1* or *Non-Top 10 Closest Friends=1*), and (3) *Group C:* Control users not connected to any information elites, serving as the reference group.

We then estimate Equations (6) and (7) for each issue and report the results in the last four rows of Tables S4 and Table S5. These results generally align with our expectations. For example, users who have interacted with information elites are more likely to disagree with government propaganda, compared to those who have never interacted (Columns (3)–(4) of Table S4). Moreover, even among users who initially agree with government propaganda, those who have interacted with their elite friends are more likely to change their beliefs over time (Columns (3)–(4) of Table S5).

In summary, this section provides evidence that social interactions between information elites and their friends play a critical role in the dissemination of politically sensitive information through social networks.

Table S4: Heterogenous Effects of Connections to Information Elites on User Beliefs, by Geographic Proximity and Relational Closeness to Elites

Dependent Variable:	*Belief*								
	Threat Posed by Omicron			Russia-Ukraine War			Japan's Discharge Plan		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Colocation	0.035***			0.009			0.021		
	(0.009)			(0.014)			(0.013)		
Non-Colocation	0.022***			0.008			0.015***		
	(0.004)			(0.005)			(0.004)		
Ever Interacted		0.039***			0.040***			0.024***	
		(0.004)			(0.007)			(0.008)	
Never Interacted		0.021***			0.018***			0.012***	
		(0.002)			(0.003)			(0.002)	
Top 10 Closest Friends			0.042***			0.042***			0.023***
			(0.005)			(0.007)			(0.009)
Non-Top 10 Closest Friends			0.021***			0.018***			0.012***
			(0.002)			(0.003)			(0.002)
Observations	369781	403966	403966	283005	299382	299382	116843	121136	121136
R-Square	0.011	0.012	0.012	0.010	0.012	0.012	0.007	0.008	0.008
Mean Y	0.228	0.230	0.230	0.192	0.194	0.194	0.026	0.026	0.026
p-value (*Group 1==Group 2*)	0.204	0.000	0.000	0.966	0.004	0.002	0.667	0.137	0.218
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table examines the heterogeneous effects of connections to information elites on user beliefs, specifically presenting the estimated coefficients (β^g) from Equation (6) from the main text. The first two rows report the heterogeneous effects of users' geographic proximity to domestic VPN elites, while the last four rows present heterogeneous effects based on users' relationship strength with their information elite friends. Columns (1)–(3), (4)–(6), and (7)–(9) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. The p-values test for the statistically significant differences between the coefficients for the group that is closer to information elites (Group 1) and the group that is farther away from information elites (Group 2). Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table S5: Heterogenous Effects of Connections to Information Elites on User Belief Shifts, by Geographic Proximity and Relational Closeness to Elites

Dependent Variable:	*Belief Shift*								
	Threat Posed by Omicron			Russia-Ukraine War			Japan's Discharge Plan		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Colocation	0.002			0.009			0.008		
	(0.006)			(0.011)			(0.008)		
Non-Colocation	0.004			0.003			-0.001		
	(0.002)			(0.004)			(0.002)		
Ever Interacted		0.015***			0.009*			0.003	
		(0.003)			(0.005)			(0.004)	
Never Interacted		0.011***			0.006**			0.002	
		(0.002)			(0.003)			(0.001)	
Top 10 Closest Friends			0.017***			0.009			0.004
			(0.003)			(0.006)			(0.005)
Non-Top 10 Closest Friends			0.010***			0.006**			0.002
			(0.002)			(0.003)			(0.001)
Observations	260915	284311	284311	206050	217202	217202	112221	116280	116280
R-Square	0.009	0.009	0.009	0.006	0.006	0.006	0.004	0.004	0.004
Mean Y	0.066	0.066	0.066	0.075	0.075	0.075	0.010	0.010	0.010
p-value (*Group 1==Group 2*)	0.764	0.178	0.039	0.647	0.541	0.535	0.254	0.798	0.575
Prior: Agree Gov	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table examines the heterogeneous effects of connections to information elites on user belief shifts, specifically presenting the estimated coefficients (δ^g) from Equation (7) from the main text. The first two rows report the heterogeneous effects of users' geographic proximity to domestic VPN elites, while the last four rows present heterogeneous effects based on users' relationship strength with their information elite friends. The sample are restricted to users whose prior beliefs align with government propaganda. Columns (1)–(3), (4)–(6), and (7)–(9) display the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. All specifications include control variables as in Column (6) of Table 2. Similar to Table S4, the p-values test for the statistically significant differences between the coefficients for Group 1 and Group 2. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

S5 Measurement Error in Social Networks

We measure the friendship network in May 2023 for the Omicron and Russia–Ukraine war issues, and in September 2023 for the Japan nuclear wastewater discharge issue. Because the Omicron and Russia–Ukraine events concluded in 2022, the network snapshot is taken after the periods of interest for these issues, raising the concern that the network may have changed in the interim. To assess network stability, we randomly draw 5 percent of users in our sample and compare their friendship networks in July and December 2025. More than 95 percent of friends remain the same between the two dates, indicating that online social networks are highly stable, especially over short horizons. Measurement error in network links is therefore unlikely to pose a serious threat to our identification strategy.

Table S6: Correlation between Individual-Level Instrument and User Characteristics

Independent Variable:	*Exposure*		
	Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
	(1)	(2)	(3)
User Characteristics			
Male	-0.315***	-0.245***	-0.316***
	(0.004)	(0.004)	(0.007)
Reported School	0.143***	0.157***	0.145***
	(0.003)	(0.003)	(0.005)
Number of Friends	44.705***	38.601***	55.088***
	(0.162)	(0.198)	(0.445)
High Followers	0.437***	0.510***	0.531***
	(0.003)	(0.004)	(0.007)
High Following	0.030***	0.062***	0.058***
	(0.004)	(0.004)	(0.007)
Account Age (years)	1.369***	1.666***	1.384***
	(0.024)	(0.030)	(0.053)
VIP Level	1.703***	1.551***	1.766***
	(0.016)	(0.019)	(0.030)
Has iPhone	0.191***	0.121***	0.144***
	(0.003)	(0.004)	(0.006)
Entertainment	0.244***	0.163***	0.166***
	(0.003)	(0.004)	(0.006)
Lifestyle	-0.038***	-0.074***	-0.149***
	(0.004)	(0.004)	(0.007)
News	-0.155***	-0.105***	-0.180***
	(0.004)	(0.004)	(0.007)
Knowledge	-0.065***	-0.037***	-0.140***
	(0.003)	(0.004)	(0.007)
Responsibility	-0.026***	-0.022***	-0.063***
	(0.002)	(0.002)	(0.004)
Observations	403966	299382	121136

Notes. This table reports correlations between the individual-level instrument (measured as users' social network exposure to lockdowns) and observable user characteristics in the full sample. Each entry is the coefficient from a univariate regression of the row variable on the instrument. Columns (1)–(3) report the results for the issues of Omicron, the Russia-Ukraine War, and Japan's Discharge Plan, respectively. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

References

Chang, Keng-Chi, William R. Hobbs, Margaret E. Roberts, and Zachary C. Steinert-Threlkeld, "COVID-19 Increased Censorship Circumvention and Access to Sensitive Topics in China," *Proceedings of the National Academy of Sciences*, 119 (2022). https://doi.org/10.1073/pnas.2102818119.

Hobbs, William R., and Margaret E. Roberts, "How Sudden Censorship Can Increase Access to Information," *American Political Science Review*, 112 (2018), 621–636.

Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., ... & Scarpino, S. V., "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, 368(2020), 493-497.

Barwick, Panle Jia, Yanyan Liu, Eleonora Patacchini, and Qi Wu, "Information, Mobile Communication, and Referral Effects," *American Economic Review*, 113 (2023), 1170–1207.