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

Xuebo Wang,[1] Hong Yang[2]

Abstract

This paper empirically examines how politically sensitive information spreads through social networks under strict media censorship in China. Using a unique dataset from Sina Weibo (China's version of Twitter), 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 is fighting 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 those who are not connected; (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 information 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

[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 individuals' beliefs and behaviors through the flow of information, 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). However, in authoritarian regimes, the free flow of information is often tightly controlled. Countries like China invest substantial resources in establishing and maintaining media censorship systems to suppress regime-threatening content and manipulate public opinion, thereby preserving 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 networks — 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 the effectiveness of censorship? Addressing these questions requires a comprehensive empirical investigation, a gap in the literature 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 globally, China generates millions of posts daily on vibrant domestic social networking platforms, providing rich data for research.[1] At the same time, the Chinese government operates one of the most sophisticated media censorship systems in the world, employing blocking, filtering, and censorship mechanisms to prevent citizens from accessing and sharing politically sensitive information. However, in the information age, such censorship cannot completely block alternative sources of information. Internet users residing abroad, as well as domestic residents using virtual private networks (VPNs), can bypass these restrictions to access uncensored content. These individuals, referred to as "information elites," are likely to share this content with their friends, potentially influencing their beliefs. In this sense, China's media censorship serves as a quasi-natural

[1] As of 2024, China remains the leading market in the world by the number of internet users (approximately 1.1 billion), followed by India and the United States. Source: https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users.

experiment: individuals connected to information elites are exposed to a radically different information environment than those without such connections. This allows us to empirically examine how politically sensitive information spreads through social networks and the pivotal role that information elites play 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 their friendship links. Within these networks, we identify key network nodes, referred to as "information elites," who can circumvent censorship to access uncensored content. Specifically, users with IP addresses originating from outside mainland China are designated as "information elites," indicating that they either reside abroad or use VPNs to bypass censorship. We then divide users into two groups: those connected to information elites (the treatment group) and those not connected (the control group). This division allows us to investigate 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 is fighting for justice; (3) Japan's discharge of nuclear wastewater into the ocean: Japan is extremely selfish and irresponsible. Given the Chinese government's extensive use of censorship and propaganda to shape public opinion, users' views on these issues can offer insight into their potential sources of information. Those who unconditionally support the government's propaganda are likely to rely primarily on official sources, 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 those not connected. However, the causal interpretation of these results relies on a strong assumption that connections to information elites are exogenous, which is unlikely to be the case. Specifically, social links are often endogenously formed, with individuals tending to connect with others who share similar characteristics — a phenomenon known as "homophily" in network theory (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 independent judgment, making them less susceptible to government propaganda. As such, the observed significant coefficients may reflect shared preferences or beliefs between information elites and their friends, rather than a causal effect of information elites on their friends.

To address this concern, we examine the shift in users' beliefs over time. While users' prior beliefs are often determined by their inherent characteristics, changes in beliefs are more likely to be driven by the acquisition of new information. For example, if a user initially agrees with government propaganda but later shifts to disagreeing with it, this change 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 in response to new information (Eil and Rao, 2011; Chen and Yang, 2019; Levy, 2021; Bursztyn et al., 2023).

Our findings show that users connected to information elites are more likely to shift their beliefs from agreeing with government propaganda to disagreeing with it over time. However, endogeneity remains a concern. Specifically, users connected to elites may be more open-minded to belief changes or better at seeking uncensored information, which could lead them to shift their beliefs over time. To further address this, we employ a matching approach to balance observable characteristics between users in the treatment and control groups. Our results remain robust, and the estimates increase in magnitude for the matched sample.

Next, we restrict the matched sample to users whose prior beliefs align with government propaganda. In this subsample, there are no systematic differences in observable characteristics between treatment and control users. Since all users in this group initially agree with government propaganda, they are less likely to differ significantly in unobservable characteristics. Within this subsample, the estimates become even larger, indicating that treatment users are more likely to shift toward disagreeing with government propaganda over time. These results suggest that information elites may have shared relevant, uncensored information with their friends, persuading them to deviate from government narratives.

More importantly, as we progressively reduce differences between the treatment and control groups by restricting the sample according to the above procedure, the coefficients increase in magnitude and become more statistically significant. This suggests that unobservable differences between the two groups may cause downward bias in our estimates. While concerns about unobservable differences may persist even in the strongest results from the final sample, further efforts to eliminate these differences would likely result in even larger coefficients. Thus, the estimates derived from the steps outlined above may represent a lower bound for the persuasive effect of information elites on their friends' belief shifts.

We further investigate the spillover effects of social networks on information dissemination. Specifically, the impact of social networks may extend beyond users directly connected to information elites, potentially spilling over to those indirectly connected as well. To identify such network

spillovers, we compare the beliefs of the friends of treatment users with a social distance of 2 from information elites to those of friends of control users with a social distance greater than 2. We find that, compared to the friends of control users, the friends of treatment users are significantly more likely to disagree with government propaganda, and they are also more likely to shift from agreeing with to disagreeing with the propaganda over time. These results suggest that treatment users may have further disseminated obtained uncensored information to their other friends, implying that politically sensitive information can propagate widely through social networks, significantly reshaping public opinion.

We provide supporting evidence for the information-sharing mechanism within social networks. First, we examine the heterogeneous effects of information elites with different beliefs. Our findings show that information elites who disagree with Chinese government propaganda have a much larger impact on their friends' beliefs compared to those who agree with it. Similarly, overseas information elites residing in countries where mainstream views contradict Chinese government propaganda exert a stronger persuasive effect than those in countries where mainstream views align with the government narrative. Finally, 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.

In summary, our findings demonstrate that social networks play a significant role in undermining the efficacy of media censorship by facilitating the spread of uncensored information among citizens. However, we do not claim that censorship is entirely ineffective in the information age. Indeed, censorship still plays a significant role in shaping public opinion, particularly in the short term. For instance, in our dataset, over 70% of users agree with Chinese government propaganda on the three selected issues. While social networks may not immediately dismantle censorship systems, their long-term effects could be much more profound. As our results suggest, social networks have the potential to gradually and deeply reshape public beliefs, leading to a lasting impact on society. As more citizens gain access to uncensored information and begin questioning government narratives, both censorship systems and authoritarian regimes may face growing challenges. A single spark can start a prairie fire: the spread of uncensored information could gradually undermine the grip of censorship and control over public thought.

This fragility of censorship in the information age is striking. While it may be possible to enforce stringent information and thought controls in nearly closed regimes such as North Korea, this becomes a formidable challenge in countries with extensive interactions with the outside world. In nations like

China, which heavily rely on international trade, the free flow of people, goods, and information across borders can significantly reshape domestic citizens' beliefs, despite the government's efforts to control the media. Consequently, it is becoming increasingly difficult to control the flow of information in a highly interconnected and open economy. While censorship systems are likely to eventually collapse, the exact moment 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 effectiveness 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 scarce. A notable exception is the influential study by Chen and Yang (2019), which finds that China's censorship system is robust, making citizens more supportive of the regime. While they observe the social transmission of uncensored information among college roommates, these effects appear to be of small magnitude. They acknowledge 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).[2]

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 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).[3] 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

[2] Source: http://access.opennet.net/wp-content/uploads/2011/12/accesscontested-china.pdf.
[3] 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.

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[4]—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 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.

[4] 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.

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, the Chinese government consistently exaggerated the threat posed by Omicron to justify its continued 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 firmly believe that Omicron remains highly fatal and unconditionally support government propaganda are likely loyal followers of the government, relying primarily on state-controlled media for information. In contrast, users who disagree with government narratives—viewing Omicron as mild and citing examples of countries successfully coexisting with the virus—are likely exposed to alternative information sources and are more aware of events happening 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.[5] 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 $Belief\ Shift_i$).[6] 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).[7] 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.

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

[6] 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.

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

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.

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.

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 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.

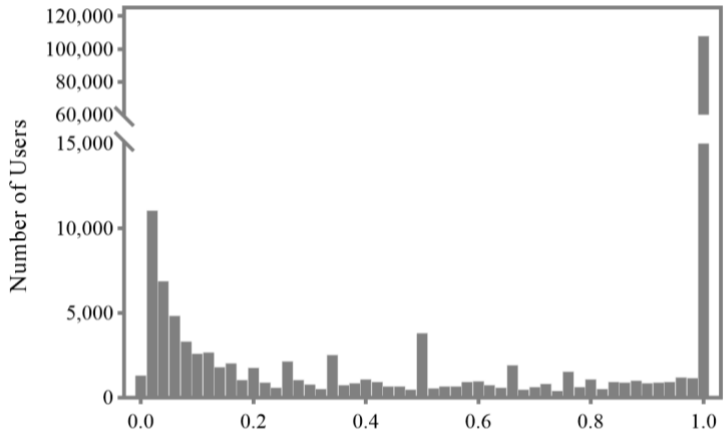
Figure 1 illustrates the distribution of information elites based on the probability that their IP addresses are located outside mainland China. An IP address with a probability of 1 indicates an overseas elite, while lower probabilities correspond to VPN elites. Figure A3 in the Appendix shows the global 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

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

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.

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.

Strength of Relationship with Information Elites. We collect all public comments (over 400 million in total) under the original posts from our sample users to map the social interaction networks.[9] 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, and (2) whether information elites are among the user's top 10 closest friends.

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

[9] In the interaction network, an undirected edge is formed between two users if one replies to or comments on a post made by the other user.

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.

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: Proximity to Info. Elites						
Colocation	0.009	0.096	0.004	0.065	0.003	0.053
Non-Colocation	0.059	0.236	0.034	0.181	0.024	0.154
Ever Interacted	0.028	0.166	0.013	0.114	0.006	0.076
Never Interacted	0.098	0.297	0.068	0.251	0.051	0.219
Top 10 Closest Friends	0.025	0.155	0.012	0.107	0.005	0.070
Non-Top 10 Closest Friends	0.102	0.303	0.069	0.253	0.051	0.221
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	

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 presents our measures of proximity between users and information elites: whether the user is friends with VPN elites and located in the same province (labeled *Colocation* or *Non-Colocation*); whether the user is friends with information elites and has ever interacted with them (labeled *Ever Interacted* or *Never Interacted*); and whether information elites are among the user's top 10 closest friends (labeled *Top 10 Closest Friends* or *Non-Top 10 Closest Friends*).

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.

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

This section presents the empirical results. We begin by outlining the empirical strategy, followed by presenting the baseline results, and then address potential concerns that may bias our estimates.

4.1 Empirical Strategy

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).

The control variables include users' demographic characteristics (e.g., gender, education), network features (e.g., number of friends, followers, and followees), account characteristics (e.g., VIP status, account registration age, iPhone user status), and users' information preferences. Additionally, we control for provincial fixed effects.

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 concern, we examine the shift in users' beliefs over time.

We first employ a matching approach to balance observable characteristics between the treatment and control groups. To further mitigate potential biases from unobservable user characteristics, we restrict the matched sample to users whose prior belief was in agreement with government propaganda. By focusing on this subsample, we investigate whether users connected to information elites are more likely to change their beliefs over time. Specifically, we estimate the following equation:

$$Belief\ Shift_i = \alpha + \delta Connection_i + \gamma X_i + \varepsilon_i, \quad (2)$$

Equation (2) 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.

4.2 Baseline Results

We estimate Equation (1) 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.

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.

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).

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.

Given that our results remain robust across different sets of control variables, we will henceforth report only the results that include all the control variables, as presented in Column (6) of Table 2.

Connection Intensity. We define connection intensity as the proportion of information elites among a user's friends and use it as the treatment variable to re-estimate Equation (1), with the results presented in the first row of Table 3. 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. In the Appendix, we further 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), and re-estimate Equation (1), with the results presented in Table A5. The results remain consistent, showing that users with more information elite friends are more likely to disagree with government propaganda.

Type of Connection. To examine the heterogeneous effects across different types of connections, we categorize treatment users into two subgroups: those connected to overseas elites and those connected to VPN elites, with users having no elite friends serving as the reference group.[10] We then

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

estimate Equation (1) and present the results in the last two rows of Table 3. The findings indicate that users in both subgroups are significantly more likely to disagree with government propaganda.

Table 3: The Effects of Connections to Information Elites on User Beliefs Regarding Politically Sensitive Issues, by Connection Intensity and Type

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)	
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	403966	393187	299382	295702	121136	120379
R-Square	0.012	0.011	0.011	0.011	0.007	0.008
Mean Y	0.230	0.229	0.194	0.193	0.026	0.026
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table examines the effects of connections to information elites on user beliefs regarding politically sensitive issues, disaggregated by connection intensity and type. Specifically, the first row assesses the impact of connection intensity on users' beliefs, presenting the estimated coefficients (β) from Equation (1), with the dependent variable being connection intensity. The second and third rows explore the heterogeneous effects across different types of connections, presenting the estimated coefficients (β) from Equation (1), where the control group consists of users unconnected to information elites, and the treatment groups consist of users connected to overseas elites and VPN elites, respectively. 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.

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.

4.3 Matching Strategy

We begin by employing a matching strategy to balance the observable characteristics between treatment and control users. Specifically, we use the Coarsened Exact Matching (CEM) approach, which is widely used in the literature (Iacus, King, and Porro, 2008; 2012), to match each user connected to information elites with an unconnected user, based on a set of covariates. 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. This procedure ensures that the resulting sample shows no systematic differences in observable characteristics between the two groups. For further details on the matching procedure, please refer to Appendix B.

We then re-estimate the same specification as in Table 2 using the matched sample, with the results presented in Column (7) of Table 2. The findings show that the coefficients from the matched sample are consistent with the baseline results.

4.4 Connection to Information Elites and Belief Updating

To more precisely identify the impact of information elites on their friends' beliefs, we examine the effect of connections to information elites on users' belief shifts.

We begin by estimating Equation (2) for the full sample across the three issues, with results presented in Columns (1), (4), and (7) of Table 4. Next, we repeat the same regressions on the matched sample and report the results in Columns (2), (5), and (8) of Table 4. Finally, we further restrict the matched sample to users who initially agree with government propaganda and run the same regressions on this subsample, with the results shown in Columns (3), (6), and (9) of Table 4. In this restricted subsample, both treatment and control users are similar in terms of observable characteristics and share the same prior beliefs, making them the most comparable to each other.

We find that as we gradually eliminate 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 two groups may cause a downward bias in our estimates. While there may still be concerns that unobservable differences persist, even in the strongest results presented in Columns (3), (6), and (9) of Table 4, further efforts to eliminate these differences would likely result in even larger coefficients. Thus, the estimates derived from the above steps may represent a lower bound for 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 4 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.

Table 4: 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 (2). 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.

Finally, Columns (7)–(9) of Table 4 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.

Notably, the posts on Japan's discharge plan were predominantly concentrated in the one- to two-week period around the start of Japan's discharge, when the Chinese government strongly condemned Japan's actions, and both domestic propaganda and public attention were at their peak (see Figure A1 in the Appendix). During this brief window, only 0.9% of users in our dataset experienced belief shifts. As a result, when we split the sample to analyze heterogeneity in Section 5, the number of users who experienced a belief shift in each subsample becomes even smaller, making it challenging to obtain statistically significant estimates.

In summary, these results demonstrate 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 significant role in reshaping their friends' beliefs by providing updated, uncensored information that challenges official narratives.

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.

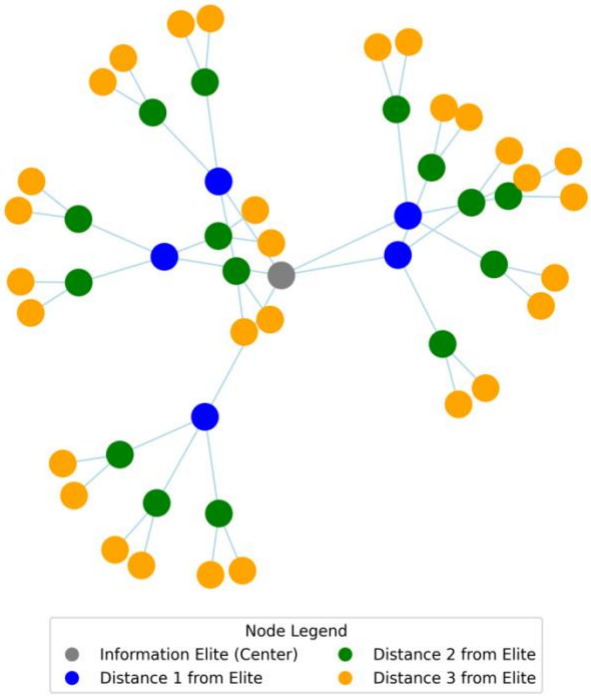
However, this concern is unlikely to undermine our findings. As mentioned earlier, the majority of Weibo users' online friends are also their real-world friends, making it unlikely that their friendship links are formed solely through online interactions. Moreover, the results from the previous section 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.

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, since the friends of treatment users are at a social distance of 2 from information elites, while the friends of control users are farther than 2 steps away, we use the latter group as the reference. We then examine whether users in the former 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 5: 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)
friends of treatment users	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 (δ) from Equation (2). Specifically, the treatment group consists of the friends of users connected to information elites, while the control group comprises the friends of users unconnected to information elites. The independent variable, *friends of treatment users*, is a dummy variable that equals one if the user is a friend of treatment users, and 0 if the user is a friend of control users. 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.

We first assess whether the friends of treatment users are more likely to disagree with government propaganda, compared to the friends of control users. To do this, we create a dummy variable (labeled *friends of treatment users*), which takes the value of 1 if the user is a friend of treatment users, and 0 if the user is a friend of control users. We use this variable as the treatment in Equation (1), with results reported in Columns (1)–(3) of Table 5. The coefficients for all three issues are all positive and statistically significant. These results suggest that, compared to the friends of control users, friends of treatment users 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 the friends of treatment users are more likely to experience belief shifts over time, compared to the friends of control users. Following a similar approach, we re-estimate Equation (2) and present the results in Table 6. 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 6 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 friends of control users, friends of treatment users are significantly more likely to change their beliefs, shifting from perceiving Omicron as fatal to viewing it as mild. In our strongest specification in Column (3), the estimate indicates that friends of 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.

Table 6: 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)
friends of treatment users	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 (2). Specifically, the treatment group consists of the friends of users connected to information elites, while the control group comprises the friends of users unconnected to information elites. The independent variable, *friends of treatment users*, is a dummy variable that equals one if the user is a friend of treatment users, and 0 if the user is a friend of control users. 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.

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 A6 in the Appendix, our main results remain robust across these different thresholds for defining user beliefs.

5. Mechanism Analysis

In the previous section, we demonstrate that information elites have a significant persuasive effect on their friends' beliefs, leading them to deviate from government propaganda. This section further explores the underlying influencing mechanisms. One plausible mechanism is information-sharing within social networks, where users who initially with government propaganda become fully informed and shift their beliefs after receiving relevant, uncensored information shared by their information elite friends.

To test this hypothesis, we examine heterogeneous treatment effects across various characteristics of users, information elites, and user-elite dyads. This approach helps us better understand the conditions under which the influence of information elites is most pronounced and how information flows within social networks contribute to belief shifts.

5.1 Heterogenous Treatment Effects by Information Elites' Beliefs

We first examine how the treatment effects vary based on the beliefs of information elites. Given that information elites can bypass censorship and access uncensored information that contradicts government narratives, they are likely to disagree with government propaganda. However, not all information elites necessarily hold this view. In reality, some Chinese citizens with access to uncensored internet may still support government propaganda on certain issues, either due to a lack of interest in politics, or because of unawareness or distrust of foreign news outlets (Chen and Yang, 2019).

Understanding this variation is important because the persuasive power of information elites may depend on whether they are exposed to and adopt uncensored views themselves. We hypothesize that information elites with different beliefs exert different impacts on their friends' beliefs. Specifically, we expect that information elites who continue to support government propaganda are more likely to share information that reinforces those narratives, which are already widely known to the public, resulting in minimal influence on their friends' beliefs. In contrast, information elites who disagree with government propaganda are more likely to share uncensored information with their friends, thus leading their friends to deviate from government narratives. This hypothesis aligns with the concept of selective sharing, which suggests that individuals tend to share information that reinforces their preexisting beliefs, while dismissing or ignoring information that contradicts them (Bakshy, Messing, and Adamic, 2015; Mocanu et al., 2015; Shin and Thorson, 2017; Bowen, Dmitriev, and Galperti, 2023).

To test this hypothesis, we first need to identify the beliefs held by information elites on the three issues under study. As outlined in Section 3.1, we use our trained NLP model to analyze users' posts to identify their beliefs. However, this approach presents challenges, as many information elites may not engage in discussions on these issues, making it difficult to determine their relevant beliefs. In our dataset, only a small proportion of information elites participate in discussions on the issues under study, resulting in identifiable beliefs.

Based on the beliefs identified for these informational elites, we categorize the sample into four groups: (1) *Group TD:* Treatment users connected to information elites who disagree with government propaganda, (2) *Group TA:* Treatment users connected to information elites who agree with government propaganda, (3) *Group TU:* Treatment users connected to information elites with unidentifiable beliefs, and (4) *Group C:* Control users who are not connected to any information elites, serving as the reference group.

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

$$Belief_i = \alpha + \sum_{g \neq C} \beta^g Group\ g_i + \gamma X_i + \varepsilon_i, \quad (3)$$

This equation is similar to Equation (1), with the reference group consisting of users who are not connected to information elites (*Group C*). The coefficient β^g captures the deviation in beliefs for each treatment group (*Groups TD, TA, and TU*) relative to the reference group (*Group C*).

Table 7: The Heterogenous Effects of Connections to Information Elites on User Beliefs, Classified by Elite Beliefs

Dependent Variable:	*Belief*		
	Threat Posed by Omicron	Russia-Ukraine War	Japan's Discharge Plan
	(1)	(2)	(3)
Panel A.			
Group TU	0.025***	0.020***	0.014***
	(0.002)	(0.003)	(0.002)
Group TD	0.068***	0.136***	0.116
	(0.010)	(0.020)	(0.129)
Group TA	0.007	-0.020	-0.002
	(0.007)	(0.013)	(0.010)
Observations	403781	299350	121136
R-Square	0.012	0.012	0.008
Mean Y	0.230	0.194	0.026
p-value (*Group TD==Group TA*)	0.000	0.000	0.363
Baseline Controls	Yes	Yes	Yes
Dependent Variable:	*Belief Shift*		
Panel B.			
Group TU	0.009***	0.006**	
	(0.002)	(0.002)	
Group TD	0.039***	0.024	
	(0.008)	(0.017)	
Group TA	0.028***	0.009	
	(0.006)	(0.012)	
Observations	284198	217184	
R-Square	0.009	0.006	
Mean Y	0.066	0.075	
p-value (*Group TD==Group TA*)	0.227	0.480	
Prior: Agree Gov	Yes	Yes	
Baseline Controls	Yes	Yes	

Notes. This table examines heterogenous effects based on information elites' beliefs. Panel A presents the estimated coefficients (β^g) from Equation (3), while Panel B presents the estimated coefficients (δ^g) from Equation (4), with the sample restricted to users whose prior beliefs align with government propaganda. Group TD consists of users connected to information elites who disagree with government propaganda, Group TA consists of users connected to information elites who agree with government propaganda, and Group TU consists of users connected to information elites with unidentifiable beliefs. 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 TD and Group TA. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Panel A of Table 7 presents the results for the three issues. In Column (1), which focuses on the Omicron issue, the coefficient for *Group TD* is 0.068, indicating that these users are 0.068 more likely to disagree with government propaganda compared to the reference group. This represents a 30% increase relative to the sample mean (0.068/0.23). For *Group TU*, the corresponding coefficient is 0.025, representing an 11% increase relative to the sample mean. In contrast, the coefficient for *Group TA* is very small and statistically insignificant, suggesting that information elites who agree with government propaganda have a negligible effect on their friends' beliefs.

Similar patterns emerge for the other two issues, as shown in Columns (2) and (3).

We further examine the heterogenous effects on user belief shifts across these treatment groups by estimating the following equation:

$$Belief\ Shift_i = \alpha + \sum_{g \neq C} \delta^g Group\ g_i + \gamma X_i + \varepsilon_i, \quad (4)$$

This equation is similar to Equation (3), with the key difference being that the dependent variable is the belief shift of users.

The results are presented in Panel B of Table 7. While most of the findings follow a similar pattern to those in Panel A, some estimates are not statistically significant. This can largely be attributed to data limitations. As noted earlier, only a small number of information elites publicly discussed the three selected issues and had identifiable beliefs, leading to a limited number of users connected to them. In particular, for the issue of Japan's Discharge plan, only about 200 users were connected to information elites with identifiable beliefs, and fewer than 10 of these users experienced belief shifts. This limitation makes it difficult to detect heterogeneous effects on these users' belief shifts, which is why the corresponding results are not reported in Table 7. Taken together, the findings in this section should be interpreted as suggestive at best.

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

While we face challenges in identifying the beliefs of most information elites on the three issues due to data limitations, we can infer the potential beliefs of overseas elites based on their residing countries or regions. 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 view contradicts Chinese government propaganda are more likely to disagree with these narratives, thus sharing relevant, uncensored information with their friends in China and influencing their beliefs. In contrast, overseas information elites residing in countries where the prevailing viewpoint aligns with Chinese government propaganda are expected to have minimal influence on the beliefs of their friends in China.

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 it.

Russia-Ukraine war: Countries that condemned Russia or imposed sanctions are more likely to oppose Chinese narratives, whereas those with ambiguous or supportive stances towards Russia tend 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 oppose Chinese narratives, while countries that criticized Japan's actions tend to align with Chinese official narratives.

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 then estimate Equation (3) for each issue and present the results in Panel A of Table 8. Columns (1)–(3) of Panel A 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.

We further examine the heterogeneous effects on user belief shifts across treatment groups. Specifically, we restrict the sample to users who initially agree with government propaganda and estimate Equation (4) for each issue. The results presented in Panel B of Table 8 show a similar pattern to those in Panel A: overseas information elites residing in countries where mainstream views

contradict Chinese government propaganda exert strong persuasion effects, while elites from countries with aligned views do not significantly influence belief shifts.

Table 8: 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)
Panel A.			
Group TC	0.019***	0.035***	0.020***
	(0.003)	(0.005)	(0.006)
Group TA	0.001	-0.017**	0.006*
	(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	0.054
Baseline Controls	Yes	Yes	Yes
Dependent Variable:	*Belief Shift*		
Panel B.			
Group TC	0.014***	0.009***	0.007*
	(0.002)	(0.004)	(0.004)
Group TA	0.007	0.001	-0.002
	(0.007)	(0.006)	(0.002)
Observations	265441	209216	113160
R-Square	0.009	0.006	0.004
Mean Y	0.066	0.075	0.010
p-value (*Group TC==Group TA*)	0.300	0.179	0.032
Prior: Agree Gov	Yes	Yes	Yes
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. Group TC consists of users connected to overseas information elites residing in countries where the prevailing viewpoint contradicts Chinese government propaganda, while Group TA consists of users connected to overseas information elites residing in countries where the mainstream view aligns with Chinese government propaganda. Panels A and B present the estimated coefficients (β^g) and (δ^g) from Equations (3) and (4), respectively, 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.

Taken together, the results support our hypothesis that information elites who are more likely to be exposed to politically sensitive information tend to share relevant, uncensored content with their friends in China, thereby reshaping their beliefs.

5.3 Heterogeneous 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.[11] 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 (3) and (4) for each issue and report the results in the first two rows of Tables 9 and 10, respectively. In Table 9, 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 10 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.

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

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 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 create two indicators to measure the relationship strength or interaction intensity between users and their information elite friends: (1) whether a user has ever interacted with their information elite friends (labeled *Ever Interacted* or *Never Interacted*), and (2) whether information elites are among the user's top 10 closest or most frequently interacted 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 (3) and (4) for each issue and report the results in the last four rows of Tables 9 and 10. 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 9). 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 10).

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 9: 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 (3). 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 10: 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 (4). 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 9, 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.

6. Conclusions

In the information 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.

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.