Article

The International Journal of Press/Politics
1–23

DOI: 10.1177/19401612221117470
journals.sagepub.com/home/hij
SAGE

How Information Flows from the World to China

Yingdan Lu[1], Jack Schaefer[2], Kunwoo Park[3], Jungseock Joo[2], and Jennifer Pan[1]

Abstract

Government censorship—internet shutdowns, blockages, firewalls—impose significant barriers to the transnational flow of information despite the connective power of digital technologies. In this paper, we examine whether and how information flows across borders despite government censorship. We develop a semi-automated system that combines deep learning and human annotation to find co-occurring content across different social media platforms and languages. We use this system to detect co-occurring content between Twitter and Sina Weibo as Covid-19 spread globally, and we conduct in-depth investigations of co-occurring content to identify those that constitute an inflow of information from the global information ecosystem into China. We find that approximately one-fourth of content with relevance for China that gains widespread public attention on Twitter makes its way to Weibo. Unsurprisingly, Chinese state-controlled media and commercialized domestic media play a dominant role in facilitating these inflows of information. However, we find that Weibo users without traditional media or government affiliations are also an important mechanism for transmitting information into China. These results imply that while censorship combined with media control provide substantial leeway for the government to set the agenda, social media provides opportunities for non-institutional actors to influence the information environment. Methodologically, the system we develop offers a new approach for the quantitative analysis of cross-platform and cross-lingual communication.

[1]Department of Communication, Stanford University, Stanford, CA, USA
[2]Department of Communication, UCLA, Los Angeles, CA, USA
[3]School of AI Convergence, Soongsil University, Seoul, South Korea

Corresponding Author:
Jennifer Pan, Department of Communication, Stanford University, 450 Jane Stanford Way, Building 120, Room 110, Stanford, CA 94305-6104, USA.
Email: jp1@stanford.edu

Keywords
censorship, social media, Covid-19, deep learning, global communication

Introduction

While digital communication technologies have revolutionized the way information can flow across borders and national boundaries, information does not flow freely everywhere. Governments all over the world impose restrictions on access to digital information using strategies such as internet shutdowns, internet filtering, denial of service attacks, active distraction, as well as regulatory controls (Earl, Maher, and Pan 2022; Deibert 2008; Freyburg and Garbe 2018; Gohdes 2015; Howard, Agarwal, and Hussain 2011; Keremoğlu and Weidmann 2020; King, Pan, and Roberts 2017; Roberts 2018; Wagner 2018). In 2020 alone, 155 internet shutdowns occurred in twenty nine countries, and countries all over the world—from Russia to Iran to Cambodia to Uganda—have implemented country-wide firewalls to control information.[1]

Nowhere is the effort to control the transnational flow of digital information more extensive and sustained than in China. Although abrupt and visible forms of censorship often result in backlash and increased demand for censored information,[2] China's decades-long effort to stem the flow of information into the country appears to have resulted in low demand for uncensored information from beyond China's borders and low demand for (and usage of) censorship circumvention technology (Chen and Yang 2019; Pan and Roberts 2020). However, neither China, nor indeed any country in the world, is completely disconnected. Events and ideas originating outside of a country's borders that have implications for domestic politics will inevitably arise. To what extent does China's ruling Chinese Communist Party (CCP) control what events, actions, opinions, and ideas originating from outside of China's borders are presented to the domestic audience and how they are presented? The Chinese government has extensive control over domestic media and imposes stringent limits on foreign media. This may allow the state to set the agenda when it comes to new information that is not institutionally driven or managed (Lawrence 2000; Livingston and Bennett 2003). In other words, perhaps global information is always transmitted into China through a gatekeeping process that is under the full control of the state. However, a great deal of research has shown that technology can enable the flow of unmanaged information—information that is not institutionally driven—thereby forcing states to respond rather than set the agenda (Livingston 1997). Does technology still play this role in the face of China's extensive system of information control?

To answer this question, we develop a semi-automated system to find co-occurring content across platforms and languages in a three-step process—retrieval, ranking, and human verification—that combines deep learning with human annotation. We deploy this system to detect co-occurring content between the English-language Twittersphere and Chinese-language Sina Weibo as Covid-19 emerged in China and spread around

the world, a time when information was most likely to flow into China.[3] We then conduct in-depth investigations of co-occurring content to determine what content flows from the global information ecosystem into China, and how this inflow occurs. We define inflow as discussion among the Chinese public of events, actions, ideas, or opinions originating outside of China's borders.

We find that 24 percent to 28 percent of the most retweeted posts on Twitter pertaining to China and Covid-19 flowed into the Chinese information ecosystem. This means that only one-fourth of topics relevant to China that gained public attention on global social media made their way into public Chinese social media discourse during a time when Covid-19 was highly salient. Chinese state-controlled media and commercialized domestic media played the largest role in facilitating the inflow of information, including content antagonistic toward China. However, Weibo users without any media or government affiliation were also consistently transmitting global information to China, and in limited instances, foreign governments and media outlets directly transmitted information to China by posting on Weibo.

The paper proceeds in five sections. "The Power and Limits of Information Control" discusses the theoretical relevance of this research. The "Data" section describes our data sources. The "Method" section details our method, including the semi-automated system for detecting co-occurring content as well as our conceptualization and operationalization of information inflow. The "Results" section presents the results. The "Discussion" section concludes by discussing the implications the findings, as well as the applicability of the semi-automated system to cross-platform and cross-lingual research.

The Power and Limits of Information Control

The Chinese government has created one of the most restrictive media and online ecosystems in the world (Chen and Yang 2019; Xu and Albert 2014). The CCP controls domestic media through ownership and sanctions. All traditional media outlets in China must have total or majority state ownership and must be supervised by a government or CCP agency, which is responsible for licensing, appointment of key personnel, and monitoring editorial decisions. While media outlets, like other state-owned firms, are subject to commercial incentives—for example, the need to compete for advertising revenue and audience—and editorial decisions and content vary by the degree of state control (Qin, Strömberg, and Wu 2018; Stockmann 2013), all media outlets, as well as journalists, face the threat of economic and political penalties if they publish content the state deems objectionable. While the CCP does issue content guidelines, the state does not always make clear what is within or out of bounds, leading media outlets and journalists to engage in self-censorship (Stern and Hassid 2012). In addition to its controls on the media, the CCP controls China's domestic internet—including websites and social media platforms—through regulation, repression, and numerous censorship strategies including search filtering, keyword blocking, account deletion, and post-hoc censorship that are implemented by domestic Chinese social media firms (Earl, Maher, and Pan 2022; Fu, Chan, and Chau 2013; Jiang 2014; King,

Pan, and Roberts 2013, 2014; Miller 2018; Qiang 2019; Qin, Strömberg, and Wu 2017; Roberts 2018).

The Chinese government does not have direct control over foreign media outlets or international social media platforms. These sources report on and disseminate news, events, and ideas that deviate from the agenda and narratives the CCP wants to promote. To prevent domestic audiences from accessing such information, the CCP uses numerous technical strategies to selectively block online information from entering the country's borders (Clayton, Murdoch, and Watson 2006; Ensafi et al. 2015; Griffiths 2021; Marczak et al. 2015; Winter and Lindskog 2012; Xu, Mao, and Halderman 2011). For example, social media platforms such as Twitter and YouTube and media sites such as *The New York Times* are not accessible in China. Users must employ a virtual private network (VPN) or other censorship circumvention technology to access this content, but VPNs are increasingly restricted by China's Ministry of Industry and Information Technology and using circumvention technology without permission has in recent years become a crime. Even before the VPN ban, however, only 3 percent of Chinese internet users regularly purchased tools to circumvent censorship (Roberts et al. 2010). This combination of technical infrastructure and regulation has created a unique Chinese online information environment where China-specific digital apps, which are subject to greater government controls, have flourished (Jiang and Fu 2018; Lee and Liu 2012; Pan 2017).

With high levels of control over domestic media and blocks on the global internet, the Chinese government may be able to set the agenda even when faced with event-driven news of foreign origin. Event-driven news is coverage of activities and actions that are not managed by officials in an institutional setting (Lawrence 2000; Livingston and Bennett 2003). Event-driven news contrasts with "pseudo-events" or institutionally driven and managed news that is planned, organized, or created by an institutional actor, often the government (Boorstin 1992). In settings with competitive, free media, event-driven news is thought to have the potential to prevent officials from setting the agenda, forcing them instead to respond to the news agenda (Livingston 1997). In a setting with stringent state controls on media, such as China, one of the main goals of information control is for the central government to always be positioned to set the agenda, even when faced with event-driven news and events and ideas originating outside of its borders. Indeed, if domestic media outlets are under state control, they will report on events in the manner desired by the state. If domestic audiences are blocked from accessing foreign media and foreign information sources, events originating outside of China's borders will also be reported on by domestic media, and again, in the manner desired by the state.

However, China's efforts to block information from entering the country's borders are selective, not wholesale. Some foreign media outlets, foreign websites, and sources of information, including their Chinese versions, are not blocked, and users in China can access these sources. A very small minority of users use circumvention technology to access global information, and they may also transmit information into the country. Finally, people with uncensored access to the global internet who are not based in China can create accounts on Chinese social media and share content. The network

structures of social media facilitate the fast-paced dissemination of information across national boundaries (Bakshy et al. 2012; Goel et al. 2016). Thus, while the scale and technological sophistication of the CCP censorship program has created distance between the country's information ecosystem and the rest of the world,digital technology may enable inflows of information outside of state control.

The question we set out to answer in this paper is whether censorship and state control of media allow the Chinese government to act as the sole gatekeeper of information about events, actions, ideas, and opinions that occur outside of the regime's control, or whether, despite stringent information control measures, social media allows for the transmission of unmanaged information.

Data

To answer this question, our research design uses Twitter and Sina Weibo data as measures of public discourse. As of June 2021, Twitter had 211 million daily active users,[4] and Weibo had 246 million daily active users.[5] Nearly all Weibo users are based in mainland China, where the company is based, while only 20 percent of Twitter users are based in the United States, where Twitter is headquartered. On both platforms, content is primarily public and serves as a source of timely information. Viral content on the platforms—by which we mean content that reaches a large audience[6] —often transcends local communities and physical borders.

We take viral, English-language tweets as a measure of visible, global English-language public discourse, and we take Weibo posts as an indicator of the presence of public discourse in China. This does not mean that Twitter or Weibo is representative of public opinion; they are not. Using data from the two platforms simply allows us to identify discussions that gained widespread traction in the English-speaking world and discussions that appeared in a public space in China, which enables us to determine whether discussions originating outside of China appear in the country.

Our data and research are centered on the time period when Covid-19 emerged. The emergence of Covid-19 was a highly salient event that brought global attention to China. Because of this global attention, China's actions, as well as the global actions directed at China, were of deep interest to Chinese audiences. Furthermore, the emergence of Covid-19 represents a situation where greater flows of information would be beneficial for individual decision-making, policymaking, and public health in China and countries around the world. The salience of Covid-19 and China during this time may push estimates of information inflow closer to a ceiling.

Twitter Data:

We utilized an existing scholarly repository of Twitter posts related to Covid-19 (Chen, Lerman, and Ferrara 2020). This repository contains all tweet IDs of tweets containing keywords related to Covid-19, such as "Corona," "N95," "pandemic," and "China."[7] This repository is based on real-time collection of tweet IDs of all tweets containing keywords using Twitter's streaming API.[8] Using the tweet IDs in the repository, we retrieved fourteen

million tweets and their metadata in the repository from January 21, 2020 to April 30, 2020. We then filtered the fourteen million Twitter posts to the 1.8 million tweets tagged as English-language tweets and containing one or more China-related keywords.[9]

We then stratified the sample of 1.8 million tweets by week and extracted the top ten most retweeted tweets by week related to Covid-19, China, and focused specific events, actions, issues, or opinions regarding specific events, actions, or issues. Tweets related to China include tweets about the Chinese regime and government, the Chinese people, as well as Chinese society and culture (for details on coding rules see Supplemental Information (SI) file).

Our final sample for analysis consisted of 150 tweets: ten tweets for each of the fifteen weeks within our sample (for text of all 150 tweets, see SI).[10] Figure 1 shows the distribution of retweets for the entire sample of 1.8 million tweets pertaining to Covid-19 and China and for our 150 sampled tweets.

As Figure 1 shows, the mean retweet of tweets in the sample is orders of magnitude higher than the mean retweet of all China-related tweets. The set of 1.8 million tweets does contain outliers with a higher retweet count than tweets in the sample because some highly retweeted tweets did not satisfy our sampling criteria and because the volume of discussions of Covid-19 varied by week. For example, the twentieth most retweeted tweet in a week where Covid-19 was highly discussed on English-language Twitter might have more retweets than the most retweeted tweet in another week where Covid-19 was not discussed as much.

Weibo Data:

We used the publicly available Weibo-COV dataset (Hu et al. 2020) as the basis on which to search for viral Twitter content. The Weibo-COV dataset was collected

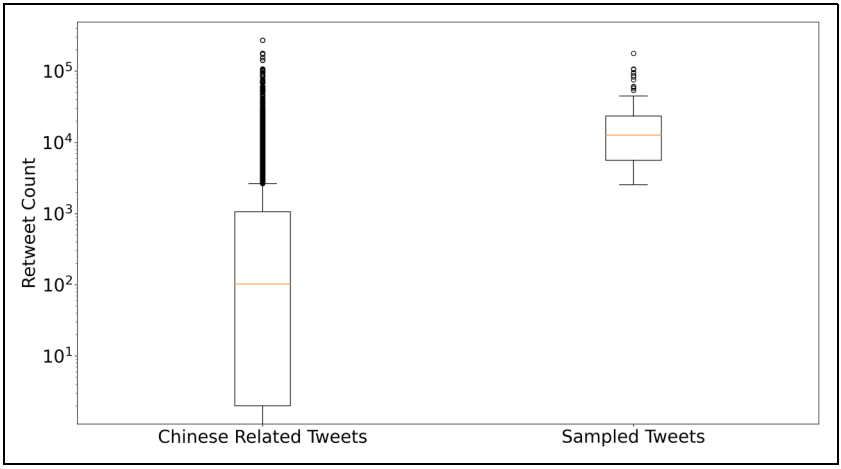

Figure 1. Comparing retweet count of 150 sampled tweets to China-related tweets

retrospectively, in April 2020, from all posts made by a sample of more than twenty million active Weibo users. The twenty million Weibo users were identified from a pool of 250 million unique users as those with more than fifty followings, followers, or posts, as well as a post in the preceding 30-day window. Posts from these twenty million active users were then filtered using Covid-19-related keywords.[11] Although several Weibo Covid-19 datasets were available, we used the Weibo-COV dataset because it contains a more diverse sample of Weibo users and has been used to study public sentiment during Covid-19 in China (Chen et al. 2020; Lu, Pan, and Xu 2021). We trained a deep learning classifier to filter out posts that contained Covid-19-related keywords, but that did not focus on Covid-19 (e.g., a post containing the word "outbreak" talking about previous disease outbreaks). Specifically, we fine-tuned a pretrained Chinese BERT with the Whole Word Masking model (Chinese BERT-wwm-ext) on a sample of 5,484 human annotations, achieving an accuracy of 0.97 and an F1 score of 0.99 (see SI for details).[12] The final dataset contained 6.7 million Covid-19-related Weibo posts made between January 16, 2020, and April 30, 2020.

Because the Weibo-COV dataset was collected retrospectively in April 2020, it is possible that some Weibo posts are not contained in the dataset because they were removed, that is, censored, after they were posted and before the April 2020 collection date. To address this concern, we conducted a robustness check where we also searched for viral Twitter content in the Weiboscope Covid-19 dataset—a real-time, pre-censored collection of Weibo posts made between December 1, 2019 and February 27, 2020 (Fu and Zhu 2020). We used the Weiboscope Covid-19 dataset to search for tweets that were not found in the Weibo-COV dataset in the six weeks between January 21, 2020 and February 27, 2020 (see SI for details).

Method

Measuring the transmission of information—events, actions, ideas, opinions—across national borders into China is a challenging task for several reasons. First, the task is a multilingual, cross-platform matching problem where source and target texts are written in different languages (English and Chinese), in different formats (e.g., Twitter has a maximum character limit of 140 characters, Weibo does not), and by different authors. Second, a very large number of candidate pairs must be evaluated for matching. Assuming there are n tweets and m Weibo posts, $n \times m$ pairs need to be compared and evaluated. Given the volume of social media posts (i.e., n and m are very large), this is an intensive computational task that is also too large for human annotation alone. Third, because the social media text is very short yet nuanced, fully automatic methods can have difficulty achieving high levels of performance in determining whether two social media posts talk about the same thing. Finally, co-occurrence does not necessarily represent inflow of information and information may transmit across borders directly on social media or through other communication channels. This means that an investigation that looks beyond the specific social media content is required to determine the direction of information flow. To date, most studies of

information flows across borders have used manual coding to identify and count common topics or country names between selected articles in a small sample set (Golan 2006; Himelboim, Chang, and McCreery 2010; Wu 2000).

We overcome these challenges by developing a deep-learning-based semi-automated method to identify the co-occurrence of content from viral English-language tweets and Chinese social media posts in three steps. This method utilizes deep-learning-based natural language processing and information retrieval methodologies to retrieve and rank Weibo content, and then employs human verification and annotation to make the final decision on co-occurring content. To measure the inflow of information, we then conduct in-depth investigations of co-occurring content.

Step 1 Retrieval:

The objective of the retrieval step is to reduce the number of target (Weibo) posts. For a tweet, we consider Weibo posts made within a $+-5$ day time period of the timestamp of a tweet, which we validated to increase the chances of detecting co-occurrence.[13] We examine Weibo posts made prior to the tweet because we are interested in whether conversations that capture global attention make their way into public discourse in China. Viral tweets are the proxy for global attention and Weibo posts are the proxy for public discourse in China, which means there may be events or actions originating outside of China that are picked up on Weibo (and therefore represent inflow of information) before they go viral on Twitter. Next, we use word2vec (Mikolov et al. 2013) embedding trained on twenty million Weibo posts (Zhang and Pan 2019) to retrieve the most relevant 10,000 Weibo posts for each viral tweet. We first translate an English-language tweet to Chinese using Google Translate API (Google 2021).[14] We then create a vector for the translated social media post by averaging the word vectors of words in the post. For each source tweet, we retrieve the top 10,000 Weibo posts with the highest cosine similarity measured over the embedding (see SI for the formalization of this algorithm).

Step 2 Ranking:

The objective of the ranking step is to identify, for every tweet, the most similar K Weibo posts among the 10,000 candidates obtained from Step 1. It would still be impractical for human annotators to verify 10,000 Weibo posts for every tweet, and thus this ranking step reduces the time and cost of human labor. For ranking, we use a multi-lingual version of the Universal Sentence Encoder (USE) (Cer et al. 2018).[15] USE uses a convolutional neural network to capture the context of a sentence, including varying word semantics across different contexts (Kim 2014).[16] Unlike word2vec, used in Step 1, USE can directly compare English tweets against Chinese Weibo posts without translation.[17] We use USE for this ranking step rather than for the retrieval step because of its higher computational cost.[18] We set $K = 100$ most similar Weibo posts for each tweet (for details on selection of K, see SI).

Step 3 Human matching:

The third step of the approach employs bilingual Chinese and English speakers to evaluate the top 100 Weibo posts produced by the ranking step and to decide whether each post corresponds to its matched tweet. These bilingual annotators are provided with the Weibo post in Chinese, the Chinese translation of the tweet, and the original English tweet with all relevant links (e.g., to images, videos). For each tweet-Weibo post pairing, two research assistants read the original English-language tweet and reviewed 100 potential Weibo pairs in Chinese (a total of 15,000 tweet-Weibo post pairs). They annotate the Weibo post as matching the tweet in all of the following cases: (1) the Weibo post covers the same issue as the tweet and the sentiment of the Weibo post is identical with that of the tweet; (2) the Weibo post covers the same issue as the tweet but expresses a different opinion, attitude, or perspective on the issue; or (3) the Weibo post covers the same issue as the tweet and contains multiple viewpoints and/or differing opinions on the issue. This means that the Weibo post matches the tweet if it is talking about the same event or issue (e.g., overcrowding at a hospital, lockdown in Wuhan, donation of medical supplies by a foreign government), even if the opinions expressed in the tweet and Weibo post are different. When research assistants do not agree on whether a Weibo post matches a tweet, an additional annotator reviews the pair, and the final decision is based on majority rule.

Inflow investigation:

Taking the co-occurring posts generated by the system, we investigate each tweet-Weibo match to determine whether it captures the flow of information into China. We examine the text, metadata, and associated images and links of the tweet and Weibo post. We search for the content contained in the tweet and Weibo post on Chinese and English-language media sites, Baidu and Google search engines, as well as search functionality on Twitter and Weibo in Chinese and English. This investigation captures the context surrounding the tweet and Weibo post and the chain of events leading up to the discussions on Twitter and Weibo.

Inflow occurs if the events, actions, ideas, or opinions discussed in the tweet originated outside of China. This excludes actions of the Chinese government as well as events and ideas originating in China. Suppose a new Chinese government policy is set and Weibo users discuss it; this is not considered inflow. However, if the same policy results in an opinion outside of China that is then picked up on Weibo, it is considered information inflow. This is because what is transmitting is the opinion, which originated outside of China. Inflow occurs if the Weibo post is echoing the event, action, idea, or opinion in question or if the Weibo post is responding to it, regardless of agreement. This means information inflow can be a critique of the originating idea as long as the specific event, action, idea, or opinion is referenced. It is not sufficient for a tweet and a Weibo post to be talking about the same general topic. We take this conservative approach to increase the certainty in identifying information inflow. For example, if a tweet and Weibo post are both about electric vehicles, but do not

make the same point (e.g., the tweet says the electric vehicle market is growing, but the Weibo post talks about the supply chain issues hampering the electric vehicle market) or do not reference each other, we do not consider this to be an information inflow.

We are interested in the mechanism of inflow because our research question pertains to whether the state dominates the inflow of information. We have grouped inflow mechanisms into four types. The first mechanism is through Chinese state-controlled media or the Chinese government. State-controlled media and the government can report on or respond to events, actions, ideas, and opinions from outside China by posting on Weibo or through other media channels such as television, newspaper, or websites. The second mechanism is through commercialized Chinese media. Although commercialized media are also state-owned, they exhibit different patterns of coverage compared with state-controlled media outlets because they are also responding to commercial incentives (Lu and Pan 2022; Qin, Strömberg, and Wu 2018; Stockmann 2013). We define commercialized Chinese media outlets as those that are not directly controlled by a CCP propaganda department, other CCP organization, or a government bureau. This means *Dailies* (日报) controlled by local propaganda departments are categorized as state-controlled media. The *Global Times*, owned by the party-run People's Daily Publishing (人民日报社), and *Beijing Review* (北京周报), a subsidiary of China's Foreign Languages Publishing Administration (中国外文出版发行事业局), a government agency, are both considered state-controlled media. Commercial media outlets include self-media (自媒体) and non-government-affiliated online news platforms such as Netease News. Inflow occurs when media outlets, media professionals, the Chinese government, or government representatives report on or respond to events, actions, ideas, and opinions from outside China by posting on Weibo or through television, newspapers, websites, or other channels. We separate state-controlled and commercialized media because this distinction has been identified as important in prior research (Qin, Strömberg, and Wu 2018; Stockmann 2013). However, research has also shown that commercialized media outlets in China can be particularly effective in influencing the public on behalf of the state because they are perceived as more independent even though they fulfill the propaganda goals of the CCP (Stockmann and Gallagher 2011). As a result, we assume that state-controlled and commercialized media outlets play a gatekeeping role for the CCP.

The third mechanism is through Weibo users not affiliated with any media outlet or government. Users can pick up information from foreign media reporting, foreign government announcements, or global social media and post it to Weibo. The last mechanism of inflow is the direct dissemination of information in China by a foreign entity. For example, some non-Chinese media outlets (e.g., Russia's RT) and foreign embassies in China (e.g., the Russian Embassy in China) have active accounts on Chinese social media and can post information directly on Chinese platforms.

Results

Among the 150 viral tweets, our system identified sixty-six tweets with at least one matching Weibo post based on the Weibo-COV dataset (see Table 1). The robustness check

Table 1. Matched Viral Tweets.

	Number	% viral tweets
All viral tweets	150	
Tweets with matched Weibo posts	68	45%
Information inflow to China	32	21%
Information outflow from China	19	13%
No clear indication of flow	17	11%

using pre-censored Weiboscope data yielded two additional co-occurrences, taking the total to sixty eight (see SI for details).

When we conducted an in-depth investigation of all sixty-eight matching tweets, we find thirty-two tweets that represent inflow of information to China, and another nineteen tweets that represent the outflow of information from China. Information outflow consists of content about events, actions, ideas, or opinions originating from China that gained attention on Twitter. For example, one viral tweet contained a voice recording of a Chinese nurse working in Wuhan from the early days of the Covid-19 outbreak; another contained an aerial drone video taken of Wuhan during the initial 2020 lockdown. The remaining seventeen co-occurring posts are instances where people within and outside of China were talking about the same issues around the same time. For example, in late January, a viral tweet described how Covid-19 was being called "Chinese" just as Ebola was touted as "African." In the same week, a Weibo post criticized how Covid-19 was being called a "Chinese" virus and states that no one would dare calling HIV/AIDS, which originated from Africa, an African virus. These two posts are similar in that they both criticize the racialization of Covid-19, but there is no clear indication of the direction of information flow. In sum, this means that around 24 percent to 28 percent of conversations that captured global attention originating outside of China made their way into China.[19]

How did information make its way into China? We find ten instances where the Chinese government or state-controlled media facilitated inflow of information, seven instances where commercialized Chinese media facilitated inflow of information, twelve instances where Weibo users without media or government affiliation facilitated inflow of information, and three instances where foreign entities facilitated inflow of information by posting directly on Weibo. These numbers are unlikely to be representative of overall proportions. These numbers may differ in other periods of time and for a broader sample of tweets. We provide this numerical breakdown to show that all four mechanisms are at play in facilitating the inflow of information to China.

Government/state-media facilitated inflow:

Figure 2 illustrates the timeline for one instance of information inflow facilitated by Chinese state-controlled media.

On February 3, 2020, *The Wall Street Journal* (WSJ) published an opinion piece titled "China Is the Real Sick Man of Asia." Approximately two weeks later, on February 19th, Chinese state-controlled media outlets, including CGTN and *People's Daily*, reported that China had expelled three WSJ reporters. Less than four hours later, U.S. Secretary of State Mike Pompeo tweeted that the "United States condemns the move by China" to expel the WSJ reporters and that "Mature, responsible countries understand that a free press reports facts and expresses opinions. China should not restrict #freespeech." Pompeo's condemnation of China is one of the 150 viral tweets in our sample, and it was picked up by various news outlets, including WSJ. Four days later, on February 23rd, *Beijing Review*, a state-controlled outlet, transmitted the U.S. government's criticism of China's expulsion of WSJ journalists to the domestic Chinese audience by posting a video on Weibo criticizing WSJ, responding to Pompeo's comments about free press, and criticizing the U.S. government for controlling the media. In this case, although some actions were initiated by China (e.g., expulsion of journalists), state-controlled media facilitated the inflow of information concerning Pompeo's comments about China. Other examples of state-controlled media facilitating inflows of information include *Global Times* reporting in February that India was donating personal protection equipment to China and Xinhua reporting in April 2020 that Japan would fund the relocation of Japanese businesses out of China.

Chinese government agencies also facilitate the inflow of information. These inflows often occur when Chinese embassies respond to criticism. In one case, in the late March, Spanish media reported that Chinese testing kits were defective. The Chinese Embassy in Spain refuted this allegation on Twitter. Weibo users discussed the dispute, and Chinese media outlets also reported on it. Similarly, in mid-April, a German media outlet's criticism of China's coronavirus response gained widespread attention on Twitter and prompted a response by the Chinese Embassy in Germany on Twitter and on China's Ministry of Foreign Affairs website.[20] Shortly thereafter, Chinese media outlets began reporting on the issue.[21]

Figure 2. Timeline of Wall Street Journal controversy.

Commercialized media facilitated inflow:

The World Health Organization (WHO) announcement of Covid-19 as a global emergency gained widespread attention on both Twitter and Weibo. Close to midnight on January 30, 2020 in China, Netease News reported that WHO had convened a meeting to declare Covid-19 a global emergency. Weibo users immediately began to share this information. In the early hours of January 31st in China, or mid-day in North America, news of the WHO announcement gained widespread attention on Twitter. In this case, information was not transmitted from Twitter to Weibo, but an action originating from outside of China flowed into China after Chinese commercialized media reported on it, which meets our definition of inflow.

Figure 3 illustrates the timeline for another instance of information inflow facilitated by Chinese commercialized media.

On April 22, 2020, U.S. Secretary of State Pompeo criticized the Chinese government in a press conference for delaying its Covid-19 report to WHO. Later that day, Guancha.com (观察者网), a commercialized Chinese media outlet known for its nationalistic discourse, criticized Pompeo for his remarks on its online forum and Weibo account. An hour later, Guancha's Weibo post was copied by other Weibo accounts, including Chinese government accounts. The next day, on April 23rd, Pompeo posted a video clip of the press conference to Twitter that became highly retweeted. This example is one in which inflow occurs even though Weibo posts appear before the viral tweet because Chinese social media picks up the action occurring outside China's border before the Twittersphere. Another similar case occurred in mid-March 2020 when the British tabloid *The Sun* reported that the United States was building mortuaries to hold the bodies of Covid-19 victims, and China-based Netease News and Sohu News picked up the report. Weibo users began discussing the topic thereafter. A few days later, a tweet picked up the same story, stating: "China were building coro hospitals England are building mortuary." Information inflow occurs

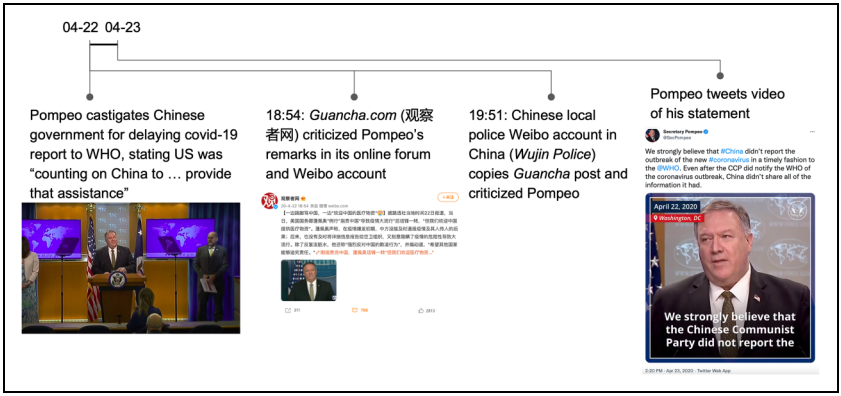

Figure 3. Timeline of Pompeo critique.

here because the action under discussion originated outside of China. The fact that the Weibo posts preceded the viral discussion on Twitter by three days does not alter the directionality of information flow.

Weibo user facilitated inflow:

We find a number of instances where Weibo users who are not affiliated with any media outlets or the government facilitated information inflow. In some instances, content flowed directly from Twitter to Weibo. In one case on February 28th, someone tweeted "Corona virus is like Pasta. The Chinese invented it, but the Italians spread it all over the world," which subsequently went viral. A few hours later, a Weibo user geolocated in Brazil posted the tweet to Weibo. We find no evidence that any traditional or online Chinese media outlets reported on this. This case suggests that Weibo users living outside of China play a role in information transmission. In other instances, Weibo users are sourcing information from foreign media outlets. For example, a Weibo user posted about a Fox News video of Trump asking an ethnically Chinese reporter, "Who are you working for? China?" We do not know whether the Weibo user saw the Fox News segment or saw the tweet of the clip, which was widely viewed. In this example, we also find no evidence of a response from the Chinese government or any reporting by Chinese media.

We observe instances where Weibo users are picking up content from U.S.-based media outlets, including BuzzFeed, *The New York Times*, and PBS. In addition, we also see that overseas Chinese media outlets are a source of information for Weibo users (see Figure 4).

On April 3, 2022, the editor-in-chief of an Indian media platform called Chanakya Forum posted a news clip from the Urdu television channel NBTV about low-quality masks allegedly made from underwear sent by China to Pakistan. Two days later, the *Business Standard*, an English-language Indian daily newspaper, published a story on the topic. Later that day, New Tang Dynasty Television, a U.S.-based media outlet founded by Falun Gong adherents that is highly critical of the CCP and blocked in China, reported on "underwear masks," citing both the Chanakya Forum editor tweet and the *Business Standard* article. Two hours later, a Weibo user copied text from the New Tang Dynasty Television article citing the *Business Standard* article and asked if it was fake news. In this instance, the Weibo user did not have a VIP (verified) account, any visible media or government affiliations facilitates, or any notable account characteristics (e.g., small number of followers, no information on geolocation). This ordinary Weibo user transmitted information into China, but this story also involves overseas Chinese media, a media professional with a Twitter key opinion leader, and media outlets in Pakistan and India.

Foreign entity facilitated inflow to Weibo:

The final mechanism is one in which foreign entities posted information directly on Weibo. Although many U.S.- and European-based media outlets are blocked and

Figure 4. Timeline of the mask donation controversy.

not present on Chinese social media, some foreign entities do have a presence. We find three instances where this mechanism of transmission was at play. First, in February, news that Russia was banning travel from China made its way to China because the Russian Embassy in China posted this information to its Weibo and WeChat accounts. Second, in April, *MIT Technology Review* posted news to its Weibo account that cybersecurity firms have identified hacks related to fraud around Covid-19. Third, in late April, SBS, an Australian TV channel, posted on its Weibo account that Australia's calls for investigations into the origins of Covid-19 had sparked intense criticism in China. The same entities posting on Weibo also shared the same content on Twitter, where this information went viral.

What types of information make their way into China? First, we examine the accuracy of information. We searched all 150 viral tweets on fact-checking websites[22] to identify tweets that contain verified misinformation. We find that 4 percent of the 150 viral tweets contain misinformation, and 6 percent of tweets that represent information flowing into China contain misinformation. This suggests that it is not the case that a disproportionate amount of inaccurate information makes its way into China.

Second, we examine whether tweets that are antagonistic toward the Chinese regime or Chinese people are more or less likely to flow into the country. We manually annotated tweets as antagonistic if they describe actions taken against the Chinese government or Chinese people or if they fault, blame, or denigrate the Chinese government, its policies, its leaders, the Chinese Communist Party, or Chinese people as a group.[23] Among the 150 viral tweets, 55 (37 percent) are antagonistic toward China, and among the thirty two tweets that represent information inflow, 21 (66 percent) are antagonistic. This suggests that a relatively larger share of content antagonistic toward China flows into China.[24] Strikingly, although ordinary Weibo users transmit some antagonistic information into China, government and state media accounts are the main channel facilitating the inflow of antagonistic information, transmitting nearly half (ten out of twenty one) of antagonistic content to China.

Discussion

These findings show that the inflow of global information into China is limited. During a global pandemic highly salient in China and around the world, only a small proportion of information pertaining to Covid-19 and China that sparked global public discussion made its way to Weibo. Because of the high interest in Covid-19 and China as well as high demand for information during this period, we believe our estimate represents a ceiling on the inflow of information that can be captured from Twitter and Weibo rather than an average.

Of the information that is transmitted in the country, state-controlled and domestic media outlets are not the only institutions responsible for transmission. Weibo users without media or government affiliation are also responsible for bringing information from international media sources as well as global social media platforms into China. This suggests that despite stringent and extensive controls over its information environment, the Chinese government—through government agencies, state-controlled media, and commercialized media—is not the sole gatekeeper of event-driven news. Social media platforms serve as a source of information (Twitter) and transmission channel (Weibo).

The results show that content antagonistic toward the Chinese government is more likely to flow into China than other types of content. The inflow of antagonistic content could be interpreted as threatening to the Chinese regime if ordinary Weibo users bring in criticism of the Chinese government or information that is not reported in state-owned media. However, two factors suggest that inflows of antagonistic content may reinforce rather than threaten government control. First, antagonistic content is not likely to be censored. Of the twenty-one viral tweets that express antagonism, twenty are found in the post-censorship Weibo-COV dataset, and only one in the pre-censorship Weiboscope dataset. Second, the Chinese government and state media accounts are responsible for transmitting the largest share of antagonistic content into the country. An important narrative the Chinese government has reiterated in its domestic propaganda is that the "West" cannot tolerate the rise of China and interferes with Chinese politics to contain China and dampen its prospects (Mattingly and Yao 2022). Facilitating the inflow of content that is antagonistic toward China reaffirms this view, suggesting that this content may be regime-reinforcing rather than threatening.

These results should be interpreted with several limitations in mind. One is that we focus on only one Chinese social media platform. Perhaps more viral tweets make their way into China through other social media platforms such as WeChat or Douyin. Another limitation is that we focus on the period when Covid-19 first emerged, when China was in the global limelight and when a relatively larger share of global attention was focused on the country. A third limitation is that we focus on a relatively small number of tweets. Taken together, these limitations suggest that the quantitative metrics presented in this paper should be interpreted as indicative of the general magnitude rather than precise or representative.

The deep learning, human annotation system created for this paper contributes methodologically to cross-lingual, cross-platform studies of digital communication. The three-step analytical framework can be used to investigate textual co-occurrence

across social media platforms and spanning diverse media contexts—for example, between Twitter and Reddit, between Facebook and Vkontakte, between politician statements and media reporting. There is growing interest in cross-platform research. Methods to facilitate cross-platform analysis have focused on similarities in usernames and user metadata, which may capture only a subset of activity across platforms (Han Veiga and Eickhoff 2016; Malhotra et al. 2012; Goga et al. 2013; Zafarani and Liu 2009). This content-focused method offers an alternative approach for cross-platform research in social science, one that can capture the spread of ideas and information relevant to many areas such as the spread of misinformation and social mobilization.

These results have implications for our understanding of the consequences and limitations of government censorship in the digital age. While these results confirm existing findings that internet blocks do significantly limit information flows across borders, they suggest that even substantial barriers are not impenetrable when connections to the world exist. Whether transmission occurs because social media users posting in China live outside of China, Chinese social media users circumvent censorship, or foreign entities post on Chinese social media, digital media does allow for the transmission of information that is not institutionally managed by the Chinese state.

The empirical analyses of this paper focus on the Chinese government, with an information control system that is second to none. As other governments around the world impose greater limits on digital information and greater control over traditional media, our results suggest that we may observe social media users playing an increasing role in censorship-circumvention and in facilitating cross-platform and cross-national flows of information.

Acknowledgments

We thank Cherian George, Aliaksandr Herasimenka, and participants of the 7th International Journal of Press/Politics Conference for comments and suggestions; Lingxiao Chen, Huazhi Qin, and Gou Wu for research assistance.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Foundation SBE/SMA No. 1831481, No. 1831848 "RIDIR: Collaborative Research: Integrated Communication Database and Computational Tools." Y. Lu was supported by the Stanford Graduate Fellowship. K. Park was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R1F1A1062691).

ORCID iDs

Yingdan Lu https://orcid.org/0000-0002-9955-6070
Jennifer Pan https://orcid.org/0000-0002-4818-0122

Supplemental Material

Supplemental material for this article is available online.

Notes

1. See https://www.accessnow.org/keepiton/.
2. Online censorship has been found to generate a backlash in the form of greater social mobilization and demand for censored information (Hobbs and Roberts 2018; Pan and Siegel 2020).
3. Replication data for this paper can be found at: https://doi.org/10.7910/DVN/7C7FEI.
4. See Twitter "Q3 2021 Letter to Shareholders" (https://bit.ly/3wavxny); 211 million is based on Twitter's definition of monetizable daily active users, who are logged in or authenticated users Twitter can show ads to.
5. See Weibo "Second Quarter 2021 Unaudited Financial Results" (https://bit.ly/2ZMthH0).
6. We use the term "virality" to denote reach and audience engagement rather than the structure through which content reached certain audiences.
7. For full keyword list, see https://github.com/echen102/Covid-19-TweetIDs/blob/master/keywords.txt (prior commits show earlier lists of keywords).
8. Real-time collection began on January 28, 2020; also collected were tweet IDs of keyword-containing tweets from the seven previous days using Twitter's historical search API (back to January 21, 2022).
9. We used four China-related keywords: China, Chinese, Wuhan, CCP.
10. To prepare for later text processing, we removed non-ASCII characters and hyperlinks to additional media.
11. See Lu, Pan, and Xu (2021) for a list of keywords and English translations.
12. We also exclude auto-generated content. For example, after making an online donation to support Covid-19 relief efforts on Weibo, users can post an auto-generated post sharing donation information.
13. We tested time windows up to a $+-10$ day period, and we found that the marginal gain of going past a $+-5$ day window is low; see SI for details.
14. Although Google Translate can introduce errors (Cornelison et al. 2021), we conducted pilot analysis and found that using tweets translated to English with Google Translate generated similar levels of performance as tweets translated by bilingual speakers.
15. We employ a multi-lingual version of USE available in TensorFlow Hub 2.0 with Apache 2.0 License (Google 2019). The model was trained by Google on sixty million scale web corpora including Reddit, Wikipedia, and Stanford Natural Language Inference (Bowman et al. 2015). They used 90 percent of the data for training and the remaining 10 percent for validation. Since this previous research shows that the pre-trained USE can generalize well for unseen text in various configurations (Chidambaram et al. 2018; Yang et al. 2019), we adopt the pre-trained model without further training on our corpus (see TF Hub website (Google 2019) for more details).
16. For example, the Chinese word 传播 can be translated as "propagate" in the phrase 传播疾病 (propagate disease) but would be understood as "communication" in 信息传播 (information communication), and USE can capture this context-based semantic distinction.

17. Each sentence is returned as a fixed-dimensional vector. USE splits a given sentence into units, called "subwords," which are more granular than words. For example, the word "concentrate" can be split into con (together) + centr (center) + ate (make), and their semantic representation can be obtained separately and aggregated for the word. After subword tokenization, USE propagates raw features of subwords by considering neighboring subwords' features, thus capturing the context of the sentence.
18. While it takes word2vec 133 min to match 150 tweets to an average of 705,759 Weibo posts, the estimated time of USE is approximately 135 h. Speed is gained if GPUs are used for computing USE, but the computation is still 2.5 times slower than word2vec without GPUs.
19. If we subtract outflow from the 150 viral tweets, inflow as a share of tweets is 24 percent; if we subtract outflow and tweets with no clear indication of flow from the 150 viral tweets, inflow as a share of tweets is 28 percent.
20. See https://www.fmprc.gov.cn/ce/cede/chn/sgyw/t1770853.htm (accessed March 22, 2022).
21. See https://bit.ly/3queUl4 (accessed March 22, 2022).
22. FactCheck.org, PolitiFact, and Snopes.
23. Content that does not fall into this category includes descriptions of the situation on the ground, calls to stop anti-Asian racism, and topics that reference Covid-19 and China but are about politics or politicians outside of China.
24. Content is mainly antagonistic toward the CCP, the Chinese government, and its policies, not of Chinese people. Among the fifty six antagonistic tweets, seven (12 percent) are antagonistic toward Chinese people, and among the twenty one antagonistic tweets representing topics that flowed into China, three (14 percent) show antagonism against the Chinese people.

References

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada Adamic. 2012. "The Role of Social Networks in Information Diffusion." In *Proceedings of the 21st International Conference on World Wide Web*. pp. 519–28.

Boorstin, Daniel Joseph. 1992. *The Image: A Guide to Pseudo-Events in America*. New York: Vintage.

Bowman, Samuel, Gabor Angeli, Christopher Potts, and Christopher D Manning. 2015. "A Large Annotated Corpus for Learning Natural Language Inference." In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. pp. 632–42.

Cer, Daniel, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Céspedes, Steve Yuan, Chris Tar et al. 2018. "Universal Sentence Encoder." *arXiv preprint arXiv:1803.11175*.

Chen, Kaiping, Anfan Chen, Jingwen Zhang, Jingbo Meng, and Cuihua Shen. 2020. "Conspiracy and Debunking Narratives About COVID-19 Origination on Chinese Social Media: How it Started and Who is to Blame." *arXiv preprint arXiv:2011.08409*.

Chen, Emily, Kristina Lerman, and Emilio Ferrara. 2020. "Tracking Social Media Discourse About the Covid-19 Pandemic: Development of a Public Coronavirus Twitter Data Set." *JMIR Public Health and Surveillance* 6 (2): e19273.

Chen, Yuyu, and David Y Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review* 109 (6): 2294–332.

Chidambaram, Muthuraman, Yinfei Yang, Daniel Cer, Steve Yuan, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. 2018. "Learning Cross-Lingual Sentence Representations Via a Multi-task Dual-Encoder Model." *arXiv preprint arXiv:1810.12836*.

Clayton, Richard, Steven J Murdoch, and Robert N. M. Watson. 2006. "Ignoring the Great Firewall of China." In *International Workshop on Privacy Enhancing Technologies*. Springer, pp. 20–35.

Cornelison, Bernadette R, Sulaiman Al-Mohaish, Yizhou Sun, and Christopher J Edwards. 2021. "Accuracy of Google Translate in Translating the Directions and Counseling Points for Top-Selling Drugs From English to Arabic, Chinese, and Spanish." *American Journal of Health-System Pharmacy* 78 (22): 2053–58.

Deibert, Ronald J. 2008. "The Geopolitics of Internet Control: Censorship, Sovereignty, and Cyberspace." In *Routledge handbook of Internet politics*. Routledge, pp. 339–52.

Earl, Jennifer, Thomas V Maher, and Jennifer Pan. 2022. "The Digital Repression of Social Movements, Protest, and Activism: A Synthetic Review." *Science Advances* 8 (10): eabl8198.

Ensafi, Roya, Philipp Winter, Abdullah Mueen, and Jedidiah R Crandall. 2015. "Analyzing the Great Firewall of China Over Space and Time." *Proc. Priv. Enhancing Technol* 2015 (1): 61–76.

Freyburg, Tina, and Lisa Garbe. 2018. "Blocking the Bottleneck: Internet Shutdowns and Ownership at Election Times in Sub-Saharan Africa." *International Journal of Communication* 12, 3896–916.

Fu, King-wa, Chung-hong Chan, and Michael Chau. 2013. "Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and the Real-Name Registration Policy." *IEEE Internet Computing* 17 (3): 42–50.

Fu, King-wa, and Yuner Zhu. 2020. "Did the World Overlook the Media's Early Warning of COVID-19?" *Journal of Risk Research* 23 (7-8): 1047–51.

Goel, Sharad, Ashton Anderson, Jake Hofman, and Duncan J Watts. 2016. "The Structural Virality of Online Diffusion." *Management Science* 62 (1): 180–96.

Goga, Oana, Howard Lei, Sree Hari Krishnan Parthasarathi, Gerald Friedland, Robin Sommer, and Renata Teixeira. 2013. "Exploiting Innocuous Activity for Correlating Users Across Sites." *Proceedings of the 22nd International Conference on World Wide Web, WWW '13*, pp. 447–58.

Gohdes, Anita R. 2015. "Pulling the Plug: Network Disruptions and Violence in Civil Conflict." *Journal of Peace Research* 52 (3): 352–67.

Golan, Guy. 2006. "Inter-Media Agenda Setting and Global News Coverage: Assessing the Influence of the New York Times on Three Network Television Evening News Programs." *Journalism Studies* 7 (2): 323–33.

Google. 2019. "TensorFlow Hub." https://tfhub.dev/google/universal-sentence-encoder-multilingual/3. [Online; accessed 25-Aug-2021].

Google. 2021. "Cloud Translation — Google Cloud." https://cloud.google.com/translate/. [Online; accessed 25-Aug-2021].

Griffiths, James. 2021. *The Great Firewall of China: How to Build and Control an Alternative Version of the Internet*. London, UK: Zed Books.

Han Veiga, Maria, and Carsten Eickhoff. 2016. "A Cross-Platform Collection of Social Network Profile." *CoRR*.

Himelboim, Itai, Tsan-Kuo Chang, and Stephen McCreery. 2010. "International Network of Foreign News Coverage: Old Global Hierarchies in a New Online World." *Journalism & Mass Communication Quarterly* 87 (2): 297–314.

Hobbs, William R, and Margaret E Roberts. 2018. "How Sudden Censorship can Increase Access to Information." *American Political Science Review* 112 (3): 621–36.

Howard, Philip N, Sheetal D Agarwal, and Muzammil M Hussain. 2011. "The Dictators' Digital Dilemma: When do States Disconnect their Digital Networks?".

Hu, Yong, Heyan Huang, Anfan Chen, and Xian-Ling Mao. 2020. "Weibo-COV: A Large-scale COVID-19 Social Media Dataset from Weibo." *arXiv preprint arXiv:2005.09174*.

Jiang, Min. 2014. "The Business and Politics of Search Engines: A Comparative Study of Baidu and Google's Search Results of Internet Events in China." *New Media & Society* 16 (2): 212–33.

Jiang, Min, and King-Wa Fu. 2018. "Chinese Social Media and Big Data: Big Data, Big Brother, Big Profit?" *Policy & Internet* 10 (4): 372–92.

Keremoğlu, Eda, and Nils B Weidmann. 2020. "How Dictators Control the Internet: A Review Essay." *Comparative Political Studies* 53 (10-11): 1690–703.

Kim, Yoon. 2014. "Convolutional Neural Networks for Sentence Classification." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pp. 1746–51.

King, Gary, Jennifer Pan, and Margaret E Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107 (2): 326–43.

King, Gary, Jennifer Pan, and Margaret E Roberts. 2014. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation." *Science* 345 (6199): 1251722.

King, Gary, Jennifer Pan, and Margaret E Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument." *American Political Science Review* 111 (3): 484–501.

Lawrence, Regina G. 2000. *The Politics of Force: Media and the Construction of Police Brutality*. Berkeley: University of California Press.

Lee, Jyh-An, and Ching-U Liu. 2012. "Forbidden City Enclosed by the Great Firewall: The Law and Power of Internet Filtering in China." *Minn. JL Sci. & Tech* 13, 125.

Livingston, Steven. 1997. "Clarifying the CNN Effect: An Examination of Media Effects According to Type of Military Intervention.".

Livingston, Steven, and W Lance Bennett. 2003. "Gatekeeping, Indexing, and Live-event News: Is Technology Altering the Construction of News?" *Political Communication* 20 (4): 363–80.

Lu, Yingdan, and Jennifer Pan. 2022. "The Pervasive Presence of Chinese Government Content on Douyin Trending Videos." *Computational Communication Research* 4 (1): 68–97.

Lu, Yingdan, Jennifer Pan, and Yiqing Xu. 2021. "Public Sentiment on Chinese Social Media During the Emergence of COVID-19." *Journal of Quantitative Description: Digital Media* 1: 1–47.

Malhotra, Anshu, Luam Totti, Wagner Meira Jr., Ponnurangam Kumaraguru, and Virgilio Almeida. 2012. "Studying User Footprints in Different Online Social Networks." *Proceedings of the 2012 International Conference on Advances in Social Networks Analy- sis and Mining (ASONAM 2012)*.

Marczak, Bill, Nicholas Weaver, Jakub Dalek, Roya Ensafi, David Fifield, Sarah McKune, Arn Rey, John Scott-Railton, Ron Deibert, and Vern Paxson. 2015. "An Analysis of China's "Great Cannon"." In *5th {USENIX} Workshop on Free and Open Communications on the Internet ({FOCI} 15)*.

Mattingly, Daniel C., and Elaine Yao. 2022. "How Soft Propaganda Persuades." *Comparative Political Studies* 0 (0): 00104140211047403.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv preprint arXiv:1301.3781*.

Miller, Blake. 2018. "The Limits of Commercialized Censorship in China." *Working Paper*.

Pan, Jennifer. 2017. "How Market Dynamics of Domestic and Foreign Social Media Firms Shape Strategies of Internet Censorship." *Problems of Post-Communism* 64 (3-4): 167–88.

Pan, Jennifer, and Margaret E Roberts. 2020. "Censorship's Effect on Incidental Exposure to Information: Evidence From Wikipedia." *SAGE Open* 10 (1): 2158244019894068.

Pan, Jennifer, and Alexandra A Siegel. 2020. "How Saudi Crackdowns Fail to Silence Online Dissent." *American Political Science Review* 114 (1): 109–25.

Qiang, Xiao. 2019. "The Road to Digital Unfreedom: President Xi's Surveillance State." *Journal of Democracy* 30 (1): 53–67.

Qin, Bei, David Strömberg, and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests Versus Surveillance and Propaganda." *Journal of Economic Perspectives* 31 (1): 117–40.

Qin, Bei, David Strömberg, and Yanhui Wu. 2018. "Media Bias in China." *American Economic Review* 108 (9): 2442–76.

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton: Princeton University Press.

Roberts, Hal, Ethan Zuckerman, Jillian York, Robert Faris, and John Palfrey. 2010. "Circumvention Tool Usage Report." *Cambridge, MA: Berkman Klein Center for Internet & Society*.

Stern, Rachel E, and Jonathan Hassid. 2012. "Amplifying Silence: Uncertainty and Control Parables in Contemporary China." *Comparative Political Studies* 45 (10): 1230–54.

Stockmann, Daniela. 2013. *Media Commercialization and Authoritarian Rule in China*. Cambridge, UK: Cambridge University Press.

Stockmann, Daniela, and Mary E Gallagher. 2011. "Remote Control: How the Media Sustain Authoritarian Rule in China." *Comparative Political Studies* 44 (4): 436–67.

Wagner, Benjamin. 2018. "Understanding Internet Shutdowns: A Case Study From Pakistan." *International Journal of Communication* 12 (1): 3917–38.

Winter, Philipp, and Stefan Lindskog. 2012. *How the Great Firewall of China is Blocking Tor*. USENIX-The Advanced Computing Systems Association.

Wu, H Denis. 2000. "Systemic Determinants of International News Coverage: A Comparison of 38 Countries." *Journal of Communication* 50 (2): 110–30.

Xu, Beina, and Eleanor Albert. 2014. "Media Censorship in China." *Council on Foreign Relations* 25, 243.

Xu, Xueyang, Z Morley Mao, and J Alex Halderman. 2011. "Internet Censorship in China: Where Does the Filtering Occur?" In *International Conference on Passive and Active Network Measurement*. Springer, pp. 133–42.

Yang, Yinfei, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, Steve Yuan, Chris Tar, Yun-Hsuan Sung et al. 2019. "Multilingual Universal Sentence Encoder for Semantic Retrieval." *arXiv preprint arXiv:1907.04307*.

Zafarani, Reza, and Huan Liu. 2009. ""Connecting Corresponding Identities Across Communities." *International AAAI Conference on Web and Social Media*.

Zhang, Han, and Jennifer Pan. 2019. "Casm: A Deep-Learning Approach for Identifying Collective Action Events With Text and Image Data From Social Media." *Sociological Methodology* 49 (1): 1–57.

Author Biographies

Yingdan Lu is a PhD candidate in the Department of Communication, Stanford University. Her research focuses on digital media, political communication, and online information manipulation. She employs computational methods on massive multimodal datasets to understand social media propaganda, media governance, misinformation and disinformation, and the presence of digital inequality around the world and on diverse platforms. Her research has been published in peer-reviewed journals, including *Political Communication*, *Human-Computer Interaction*, and the *Journal of Quantitative Description: Digital Media*.

Jack Schaefer is a research assistant in UCLA's Computational Media Lab and a Research Data Analyst in UCLA's Holocaust and Genocide Studies Digital Research Lab. His work utilizes computational tools and methods within social science and humanities domains. His research interests include: Data Science, Data Visualization, Machine Learning, and Computational Social Science.

Kunwoo Park is an assistant professor in School of AI Convergence, Soongsil University. He studies data science and machine learning methods, aiming to address social problems. Recent interests focus on detecting misinformation and social bias in online media platforms using natural language processing and multi-modal methods. His research has been published in premier computer science/interdisciplinary conferences (ACL, WWW, ICWSM, CSCW, AAAI) and social science/interdisciplinary journals (*Journal of Advertising*).

Jungseock Joo is an assistant professor of Communication at University of California, Los Angeles where he directs the Computational Media Lab. His research focuses on understanding multimodal human communication with computer vision and machine learning. He has received more than $2M in extramural funding from the National Science Foundation, Hellman Foundation, Samsung, and other awards. His research has appeared in prestigious venues in computer science and social sciences including the *Proceedings of the National Academy of Sciences*, IEEE Conference on Computer Vision and Pattern Recognition, *American Political Science Review*, and *The Journal of Politics*.

Jennifer Pan is an associate professor of Communication at Stanford University. Her research resides at the intersection of political communication and authoritarian politics, and she has published in peer-reviewed journals such as *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, *Political Communication*, and *Science*. Pan received her Ph.D. from Harvard University's Department of Government, and her B.A. from Princeton University.