

How Information Flows from the World to China

Yingdan Lu[†] Jack Schaefer[‡] Kunwoo Park[§] Jungseock Joo[‡]
Jennifer Pan[†]

Abstract

A large body of research shows that 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-fifth of content with relevance for China that gain widespread public attention on Twitter appear on Weibo. Unsurprisingly, Chinese state-controlled media and commercialized domestic media play a dominant role in facilitating these inflows of information. However, we also find that Weibo users without media or government affiliation are consistently transmitting global information to China, and in limited instances, foreign governments and media outlets also transmit information directly to Weibo. These results imply that while China’s censorship apparatus severely limits the inflow of global information into China, state-controlled and domestic media outlets do not completely dominate information transmitted into the country.

Keywords: Censorship, social media, Covid-19, deep learning, global communication

*We thank participants of the 7th International Journal of Press/Politics Conference for comments and suggestions; Lingxiao Chen, Huazhi Qin, and Gou Wu for research assistance; the National Science Foundation (grant #1831481, #1831848) and the Stanford Graduate Fellowship for funding support.

[†]Department of Communication, Stanford University

[‡]Department of Communication, UCLA

[§]School of AI Convergence, Soongsil University

1 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 (Deibert 2008, Earl et al. 2022, Freyburg and Garbe 2018, Gohdes 2015, Howard et al. 2011, Keremoğlu and Weidmann 2020, King et al. 2017, Munger et al. 2019, Roberts 2018, Wagner 2018). In 2020 alone, 155 internet shutdowns occurred in 29 countries, and countries all over the world—from Russia to Iran to Cambodia to Uganda—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). However, China, and indeed no 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

¹See <https://www.accessnow.org/keepiton/>.

²Online censorship has been found to generate backlash in the form of greater social mobilization and demand for censored information (Hassanpour 2014, Hobbs and Roberts 2018, Pan and Siegel 2020).

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 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 Sina Weibo as Covid-19 emerged in China and spread around the world. We then conduct in-depth investigations of co-occurring content to determine which co-occurrences constitute an inflow of information from the global information ecosystem into China and how this inflow occurs.

We find that less than half of the most retweeted tweets pertaining to China and Covid-19 co-occur on Weibo in the time period of our study (January to April, 2020). Among co-occurring content, roughly half represent an inflow of information into the Chinese information ecosystem, which means that only around 20% of topics with relevance for China that gain public attention on global social media made its way into public Chinese social media discourse. Chinese state-controlled media and commercialized domestic media played a dominant role in facilitating the inflow of information; however, Weibo users without traditional 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. These results suggest that social media plays a role in enabling the flow of unmanaged information despite censorship and state control of media.

The paper proceeds in five sections. Section 2 discusses the theoretical relevance of this research. Section 3 describes our data sources. Section 4 details our method, including the semi-automated system for detecting co-occurring content as well as our conceptualization and operationalization of information inflow. Section 5 presents the results. Section 6 concludes by discussing the findings, along with their limitations and their implications beyond China

and Covid-19.

2 The Power and Limits of Information Control

The Chinese government has created one of the most restricted media and online ecosystem in the world (Anderson 2012, 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. These outlets are required to have a government or CCP supervising 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—e.g., competition for advertising revenue and audience—and editorial decisions and content vary by the degree of state control (Qin et al. 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 journalist to engage in self-censorship (Stern and Hassid 2012). In addition to controls on the media, the CCP controls China’s domestic internet—including websites, social media platforms—through regulations, 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 et al. 2022, Fu et al. 2013, Gunitsky 2015, Jiang 2014, King et al. 2013; 2014, Miller 2018, Qiang 2019, Qin et al. 2017, Roberts 2018, Zhu et al. 2013).

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 et al. 2006,

Deibert 2002, Ensafi et al. 2015, Griffiths 2021, Marczak et al. 2015a;b, Winter and Lindskog 2012, Xu et al. 2011). For example, social media platforms such as Facebook 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 banned by China’s Ministry of Industry and Information Technology and using circumvention technology without permission is considered a crime. Even before the VPN ban, however, only 3% 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 so that the central government is always 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, all events, even those originating outside of China’s borders, should be reported on only 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 is selective, not wholesale. Some foreign media outlets, foreign websites, and sources of information, including their Chinese versions as 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 who are not based in China and who have unlimited access to the global internet can create accounts on Chinese social media and share content. The network structures of social media facilitates fast-paced dissemination of information across national boundaries (Bakshy et al. 2012, Friggeri et al. 2014, Goel et al. 2016, Ugander et al. 2011). Although 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, there is an unanswered question of whether digital technology enables any inflow of information that is not controlled by the state.

The question we set out to answer in this paper is whether censorship and state-control of media allows authoritarian governments 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.

3 Data

To answer this question, our research design relies on data from both Twitter and Sina Weibo as measures of public discourse. As of June 2021, Twitter had 211 million daily active users,³ and Weibo had 246 million daily active users.⁴ Nearly all Weibo users are based in mainland China, where the company is based, while only 20% of Twitter users are based in the US, where Twitter is headquartered.

We focus on data from these two social media platforms for three reasons. First, both are

³See 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.

⁴See Weibo “Second Quarter 2021 Unaudited Financial Results” (<https://bit.ly/2ZMthH0>).

microblogging sites where all posts are primarily public. In other words, most posts are not limited to friends or followers as is the case on sites like Facebook and WeChat. Second, the public nature of Twitter and Weibo means that they often serve as sites of public discourse, and they are often the places where information is shared when a highly salient event with high levels of uncertainty emerges. For example, if you hear police sirens in your neighborhood or feel the house shaking, you are more likely to go to Twitter than another social media platform like Facebook or YouTube to try to seek timely, pertinent information. Last, viral Twitter content—by which we mean content that reaches a large audience⁵—often transcends local communities and physical borders. We focus on Twitter because the public nature of the platform is such that content is most likely to be seen by people outside of close friend networks or tight-knit communities.

As a result, for the purposes of this paper, 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 is representative of global public opinion; it is not. This also does not imply that Weibo is representative of Chinese public opinion; it is also not representative. Using data from the two platforms simply allows us to determine whether discussion of events, actions, ideas, and opinion originating outside of China that gained widespread traction in the English-speaking world made their into public discussions in China.

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. Focusing on information flow into China during the global spread of Covid-19 represents a

⁵We use the term virality to denote reach and audience engagement rather than the structure through which content reached certain audiences.

case where information is most likely to flow into the country. As a result, estimates of the information inflow will be closer to a ceiling than an average.

During the emergence of Covid-19, a number of researchers collected social media data pertaining to the pandemic and made these datasets publicly available. For this paper, we use these existing, public social media datasets.

Twitter Data: We utilized the existing scholarly repository of Twitter posts related to Covid-19 (Chen et al. 2020). This repository contains all tweet IDs of tweets containing keywords related to Covid-19, such as “Corona,” “N95,” “pandemic,” and “China.”⁶ This repository is based on real-time collection of tweet IDs of tweets containing keywords using Twitter’s Streaming API that began on January 28, 2020, as well as tweet IDs of keyword-containing tweets from the 7 previous days using Twitter’s historical search API (back to January 21, 2020). We focus on the 14 million tweets found in the repository during our study time period of January 21, 2020 to April 30, 2020. Using the tweet IDs of these 14 million posts, we retrieved the full content of associated tweets and tweet metadata. We then filtered the 14 million Twitter posts to the 1.8 million tweets tagged as English-language tweets and containing one or more China-related keywords.⁷

To obtain viral tweets, we stratified our sample of 1.8 million tweets by week and extracted the top 10 most retweeted tweets by week that are related to Covid-19, related to China, and talk about specific events or issues. Tweets related to China includes tweets about the Chinese regime and government, Chinese people, as well as Chinese society and culture. We excluded tweets that were general (e.g., “Covid-19 in China is terrible”) because such tweets could have many corresponding matches in Chinese social media but having corresponding match may or may not denote the flow of information. Finally, we ensured that the 10 most retweeted tweets for each week were not duplicates or near duplicates of other sampled tweets with high retweets. If there were tweets across weeks or within a week discussing the same issue, we

⁶Full list of current keywords found at <https://github.com/echen102/Covid-19-TweetIDs/blob/master/keywords.txt>; look at previous commits to see earlier lists of keywords.

⁷We used four China-related keywords: China, Chinese, Wuhan, CCP.

would keep the earliest occurring tweet and sample the tweet with the next highest number of retweets so that each week contained 10 unique tweets.

This means our final analysis sample consisted of 150 tweets: 10 tweets for each of the 15 weeks within our sample (for text of all 150 tweets, see Appendix).⁸ 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

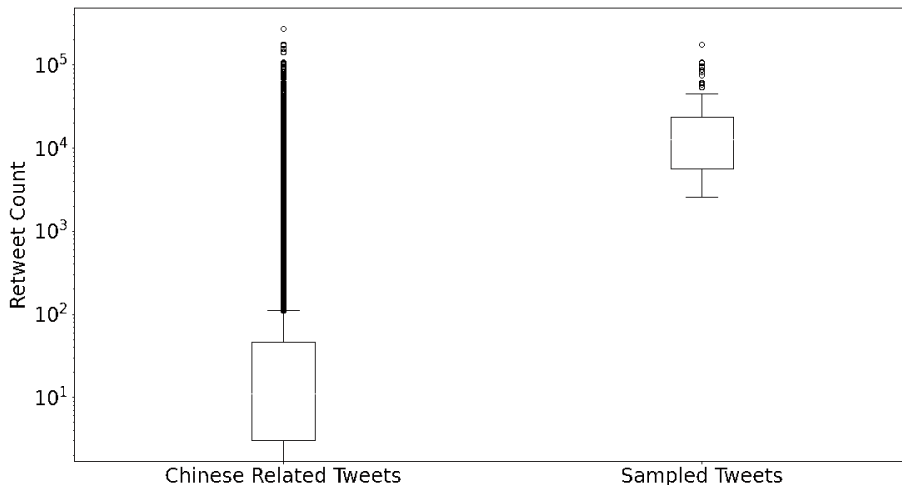


Figure 1: Comparing retweet count of 150 sampled tweets to China-related tweets

the sample is orders of magnitude higher than the mean retweet of all China-related tweets. However, the set of 1.8 million tweets does contain some outliers that have a higher retweet count than tweets in the sample. This is 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 20th most retweeted tweet in a week where Covid-19 was highly discussed on English-language Twitter might have more retweets than the most retweet tweet in another week where Covid-19 was not discussed as much.

⁸To prepare for later text processing, we removed non-ASCII characters and hyperlinks to additional media.

Weibo Data: We use 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 retrospectively, in April 2020, from all posts made by a sample of more than 20 million active Weibo users. The 20 million Weibo users were identified from a pool of 250 million unique users as those with more than 50 followings, followers, or posts, as well as a post in the proceeding 30-day window. Posts from these 20 million active users were then filtered using Covid-19-related keywords.⁹ Although several Weibo Covid-19 datasets are available, we use 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 et al. 2021). We trained a deep learning classifier to filter out posts that contain Covid-19-related keywords but that do not focus on Covid-19 (e.g., a post containing the word “outbreak” may be 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 Appendix for details).¹⁰ The final dataset contains 6.7 million Covid-19-related Weibo posts made between January 16, 2020, and April 30, 2020.

Because the Weibo-COV dataset is collected retrospectively in April 2020, it is possible that some Weibo posts are not contained in the dataset because they were removed, i.e., censored, after they were posted and before the April 2020 collection date. To address this concern, we conduct a robustness check where we also search 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). Specifically, we search the Weiboscope Covid-19 dataset for tweets that were not found in the Weibo-COV dataset in the six weeks between January 21 and February 27, 2020 (see Appendix for details).

⁹See (Lu et al. 2021) for list of keywords and English translations.

¹⁰We 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.

4 Method

Answering our research question requires that we measure the transmission of information—events, actions, ideas, opinions—across national borders into China. In other words, are events, actions, ideas, and opinions originating outside of China making their way into Chinese public discourse on Weibo? This 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 and a task that is too large for human annotation alone. Third, because social media text is very short yet nuanced, any fully automatic method may 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 border 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 et al. 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. 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 whether co-occurring

content represents an inflow of information into China.

Step 1 Retrieval: The objective of the retrieval step is to reduce the number of target (Weibo) posts. For a tweet, we only consider Weibo posts made within a ± 5 day time period of the timestamp of a tweet.¹¹ This narrow time window increases our chances of detecting co-occurrence. If we do not narrow the time period, it increases the possibility that the same content could emerge in two settings independently. For example, if a content from a Twitter post from March is found several months later on Weibo, it is difficult to say whether information flowed between the two countries or whether social media users in two different countries independently arrived at a similar topic or issue. We set the window to five days because of the fast pace of the social media news cycle (Leskovec et al. 2009).

Next, we use word2vec (Mikolov et al. 2013) embedding to retrieve the most relevant 10,000 Weibo posts for each viral tweet. Word2vec is a word embedding method that represents words as vectors based on their semantic meanings. Words with similar meanings are located based on similarities between their word embedding vectors, e.g., the corresponding vector of “dog” will be closer to “cat” than “owl.” We can compute semantic similarity based on a vector similarity metric between words and average this across a sentence. Specifically, we first translate an English-language tweet to Chinese using Google Translate API (Google 2021).¹² We then create a vector for the translated social media post by averaging the word vectors of words in the post. We use a word2vec embedding trained on 20 million Weibo posts, with each word represented as a 128-dimension vector (Zhang and Pan 2019). For each source tweet, we then retrieve the top 10,000 Weibo posts with the highest cosine similarity measured over the embedding (see Appendix for the formalization of this algorithm). This retrieval step leaves a filtered subset of Weibo posts that occur within a reasonable time span of the sample tweet

¹¹To validate the time window, we also expanded the time window up to a ± 10 day period, and we found that the majority of tweets retrieved in a ± 10 day time window overlap with posts retrieved in ± 5 days, see Appendix for details.

¹²Although Google translation can introduce errors (Cornelison et al. 2021), we conduct pilot analysis and find that using tweets translated to English with Google translation generate similar levels of performance to tweets translated by bilingual speakers.

to determine co-occurrence.

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. Here, we utilize a multi-lingual version of the Universal Sentence Encoder (USE) (Cer et al. 2018) for ranking.¹³

USE is a contextualized embedding that has several advantages. It uses a convolutional neural network to capture the context of sentence, including varying word semantics across different contexts (Kim 2014). 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. It directly compares English tweets against Chinese Weibo posts without translation.¹⁴ We use USE for this ranking step rather than for the retrieval step because of its high computational cost. While it takes word2vec 133 minutes to match 150 tweets to an average of 705,759 Weibo posts, the estimated time of USE is approximately 135 hours.¹⁵

An important analytical decision is to select a proper K , which is the number of Weibo posts human annotators should verify. If K is too small, there could be Weibo posts that co-occur with a tweet but are missed because the relevant posts were not included in K and thus humans have no opportunity to verify them. If K is too large, human annotators would have to review more Weibo posts for every tweet, increasing the cost of covering larger numbers

¹³We 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 60 million scale web corpora including Reddit, Wikipedia, and Stanford Natural Language Inference (Bowman et al. 2015). They used 90% of the data for training, and the remaining 10% 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).

¹⁴Each 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.

¹⁵Speed is gained if GPUs are used for computing USE, but the computation is still 2.5 times slower than word2vec without GPUs.

of tweets for analysis. We select K based on our initial investigation of how K influences detection of co-occurring Weibo posts. Figure 2 shows how the total number of Weibo posts matched to tweets as a function of K . Figure 2 shows a rapid increase in the number of tweets

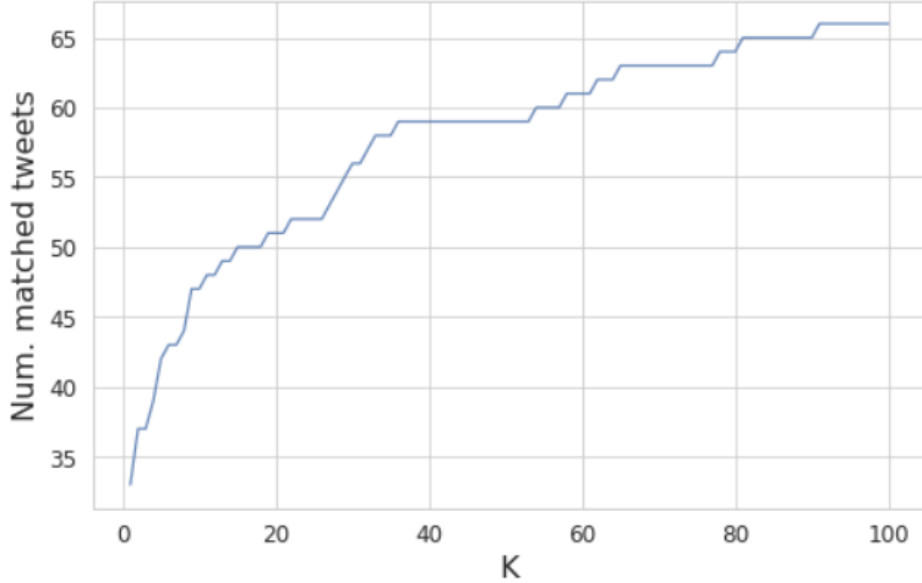


Figure 2: Number of matched tweets with varying K

with detected, co-occurring Weibo posts as K increases to 40. The rate of increase slows down afterwards, and is largely stable after $K = 60$. To err on the side of conservatism, and to maximize our chances of detecting co-occurring Weibo posts, we set $K = 100$.

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 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 a 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. For example, we would say that a tweet criticizing the Chinese government for locking down Wuhan matches a Weibo post praising the Chinese government for lockdown. We would also say that a Weibo post matches a tweet if the Weibo post criticizes a tweet, and vice versa. We would say there is a match if a Weibo post criticized a tweet recounting stories of anti-Chinese racism. 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.

Step 4 Inflow investigation: In this last step, we investigate each tweet containing Weibo matches to determine whether the information contained in the tweet represents information inflow into China. We do so by looking at references, images, and links contained within the tweet and Weibo post, as well as by searching for the same content on Chinese and English-language media sites, Baidu and Google search engines, as well as search functionality on Twitter and Weibo in Chinese and in English. We conduct this investigation for all co-occurring content to capture the context surrounding the tweet and Weibo post and to uncover the chain of events leading up to the spread of the topic on Twitter and Weibo.

We define inflow to mean that the event, action, idea, or opinion contained in the tweet must have originated outside of China. We exclude from this definition actions of the Chinese government as well as events or idea originating in China. However, suppose a new Chinese government policy generate discussion outside of China, if this discussion of a Chinese policy is then picked up on Chinese social media, it (the discussion of the policy, not original reporting about the policy) is considered an 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. In other words, a Weibo post can criticize the originating idea. However, to be considered inflow, the Weibo post must reference the specific event, action, idea, or opinion. It is not sufficient for a tweet and a Weibo post to be talking about the same general topic. This is a conservative approach to identifying information flow. We do not consider discussion of the same topic sufficient to indicate inflow or outflow unless specific details match. For example, if a tweet and Weibo post are both about electric vehicles, but do not make the same point (e.g., tweet says electric vehicle market is growing and 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 information inflow or outflow.

Because we treat Twitter and Weibo as proxies for public discussion, there can be an inflow of information if the viral tweet and Weibo post occur on the same day. This is because tweets and Weibo posts may all be responding to the same action occurring outside of China simultaneously. This represent inflow because our definition is not limited to the flow of social media content from Twitter to Weibo per se but rather to the flow of information into China. Similarly, there can be inflow of information even if the viral tweet occurs after the matched Weibo post. Again, this is because we are using viral tweets as a proxy for English-language discourse that captures global attention rather than analyzing whether a specific tweet transmitted into China. As a result, information can flow into China because it is about an event that originated outside of China even if social media discussion on Weibo preceded the viral discussion on Twitter by hours or days.

We are particularly interested in the mechanism of inflow because our research question pertains to whether the state dominates the inflow of information. We group mechanisms into four types. First, information can flow into Weibo because it is covered by Chinese state-controlled media or referenced by the Chinese government or government representatives and then picked up by Weibo users. State-controlled media and 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.

Second, information can flow into Weibo because it is covered by Chinese commercialized media or transmitted by media professional from commercialized media outlets. Although commercialized media are also state-owned, they have a greater degree of independence because of advertising, subscriber, and other commercial incentives, and prior research has found difference in their patterns of coverage compared with state-owned media outlets that function as government mouthpieces (Lu and Pan 2022, Qin et al. 2018, Stockmann 2013). We identify commercialized Chinese media outlets as those that are not directly controlled by a Chinese Communist Party (CCP) propaganda department, other CCP organization, or a government bureau. For example, *Dailies* (日报), controlled by local propaganda departments, are not considered commercialized media. The *Global Times*, owned by the party-run *People's Daily Publishing* (人民日报社) is not considered commercial media, and neither is *Beijing Review* (北京周报) because it is a subsidiary of China's Foreign Languages Publishing Administration (中国外文出版发行事业局), a government agency. Based on this definition, commercial media outlets include self-media and non-government affiliated online news platforms, such as Netease News. Chinese commercialized media facilitates inflow when it report on or respond to events, actions, ideas, and opinions from outside China by posting on Weibo or through television, newspaper, websites, or other media channels.

Third, information can flow into Weibo because a Weibo user, not affiliated with any traditional media outlet or government, picks up information and posts it to Weibo. Users can do this if they see foreign media reporting, foreign government announcement, tweets, and other social media content from outside of China. Finally, information can flow into Weibo because a foreign entity (e.g., non-Chinese media outlet with Weibo account, foreign embassy with Weibo account) posts information directly on Weibo.

In this project, we are interested in measuring the extent to which there is inflow of information into China, but more importantly, we are focused on understanding how infor-

mation flow occurs, and whether this process is dominated by domestic media outlets, both state-controlled and commercialized. Although commercialized media outlets do exhibit some editorial independence, they received government directive and face sanctions and punishment if they deviate. Research has 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 (Stockmann and Gallagher 2011). As a result, in the paper, state-controlled media and commercialized media outlets are assumed to serve the interests of the state and play a gatekeeping role for the CCP.

5 Results

Among the 150 viral tweets, our system identified 66 tweets with at least one matching Weibo post made within five days of the viral tweet (see Table 1). We conduct robustness checks using pre-censored Weibo data to check whether more tweets might have Weibo matches and we find that our dataset is not substantively affected by online censorship (see Appendix for details). When we conduct in-depth investigation of all 66 matching tweets, we find

Table 1: Matched Viral Tweets

	Number	% viral tweets
All viral tweets	150	
Tweets with matched Weibo posts	66	44%
Information inflow to China	31	21%
Information outflow from China	18	12%
No clear indication of flow	17	11%

that 31 tweets (21% of all viral tweets) represent inflow of information to China. In other words, these tweets contain information about events, actions, ideas, or opinions originating outside of China that made its way into China. Another 18 tweets (12%) represent outflow of information from China. These tweets contains content about events, actions, ideas, or

opinions originating from China. For example, one viral tweet contained a voice recording of a Chinese nurse working in Wuhan from the early days of the outbreak, and another contained aerial drone video taken of Wuhan during the initial 2020 lockdown. These tweets and Weibo posts are identified as co-occurring and represent an outflow of information because the actions originated within China. The remaining 17 matches do not provide any clear indication for information flow. Rather, it appears that people within and outside of China were talking about the same issues around the same time. For example, in late January, a viral tweet describes how Covid-19 is being called “Chinese” just as ebola was touted as “African.” In the same week, a Weibo post criticizes how Covid-19 is being called a “Chinese” virus and states that no one would dare call HIV/AIDS, which originated from Africa, an African virus. There is similarity between these two posts in that they both criticize the racialization of Covid-19. However, the tweet focuses on ebola while the Weibo posts focuses on HIV/AIDS, and importantly, the tweet expresses the opinion that ebola was similarly racialized while the Weibo post expresses a different opinion, that China is being singled out and other diseases, such as HIV/AIDS, were not racialized.

We find 10 instances where the Chinese government or state-controlled media facilitated inflow of information, seven instances where commercialized Chinese media facilitated inflow of information, 11 instances where Weibo users without media or government affiliation facilitated in-flow of information, and three instances where foreign entities facilitated inflow of information by posting directly on Weibo. The numbers we provide here are imprecise and unlikely to be representative of overall proportions. The relative share of inflow mechanisms may differ over time, or if we were working with a broader sample of tweets in the same time period. We provide these numbers only to show that all four mechanisms are at play in facilitating the inflow of information to China.

Government/state-media facilitated inflow: Figure 3 illustrates the timeline for one instance of an inflow of information that is facilitated by Chinese state media. On February 3, 2020, the *Wall Street Journal* (WSJ) published an opinion piece titled “China Is the



Figure 3: Timeline of Wall Street Journal controversy

Real Sick Man of Asia.” Approximately two weeks later, on February 19, Chinese state-controlled media outlets, including CGTN and *People's Daily* reported that China has expelled three WSJ reporters from China. Less than four hours later, US Secretary of State Mike Pompeo tweeted that the “United States condemns the move by China” to expel the WSJ reporters and “Mature, responsible countries understand that a free press reports facts and expresses opinions. China should not restrict #freespeech.” This condemnation of China is one of the 150 viral tweets in our sample, and is picked up by various news outlets, including WSJ. Four days later on February 23, *Beijing Review* (北京周报), a state-controlled outlet, transmits information about Pompeo’s criticism of China’s expulsion of WSJ journalists to a domestic Chinese audience by posting a video on Weibo criticizing the WSJ and responding to Pompeo’s comments about free speech. In this example, although some actions were initiated by China (e.g., expulsion of journalists), state-controlled media facilitated the inflow of information concerning US condemnation of China’s action toward WSJ. Other examples of state-controlled media facilitating inflows of information include *The Global Times* reporting in February that India was donating personal protection equipment to China and Xinhua reporting in April 2020 that Japan would invest to relocate Japanese businesses out of China.

In addition to state-controlled media outlets, Chinese government agencies also facilitate the inflow of information to China. In the data we examine, these inflows often occur because Chinese embassies abroad are responding to criticisms of the Chinese government. For example, in late March, Spanish media reported that Chinese testing kits were defective. The Chinese Embassy in Spain refuted this allegation on Twitter. Chinese media outlets reported on the dispute in China, and this reporting was picked up on Weibo. 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, which was posted on the website of China's Ministry of Foreign Affairs and accessible within China.¹⁶ Shortly thereafter, Chinese media outlets began reporting on the issue.¹⁷

Commercialized media facilitated inflow: The viral tweet of the WHO announcement of Covid-19 as a global emergency contained the largest number of matching Weibo posts. Close to midnight on January 30, 2020 in China, Netease News, a commercialized internet news platform in China, reported that WHO convened a meeting to declare Covid-19 as a global emergency. Weibo users immediately began to share this information. In the early hours of January 31 in China, or mid-day in North America, news of the WHO announcement gains 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 because of Chinese commercialized media reporting, and hence meets our definition of inflow. While it may seem preposterous that this WHO announcement would not have spread globally and flowed into China, we should keep in mind that in the most extreme cases of information controls—perhaps in North Korea—we do not know for sure whether the public has knowledge of actions that are known to the rest of the world, such as the WHO declaration.

Figure 4 illustrates the timeline for another instance of information inflow facilitated by Chinese commercialized media. On April 22, 2020, US Secretary of State Pompeo criticizes

¹⁶See <https://www.fmprc.gov.cn/ce/cede/chn/sgyw/t1770853.htm> (accessed March 22, 2022).

¹⁷See <https://bit.ly/3queU14> (accessed March 22, 2022).

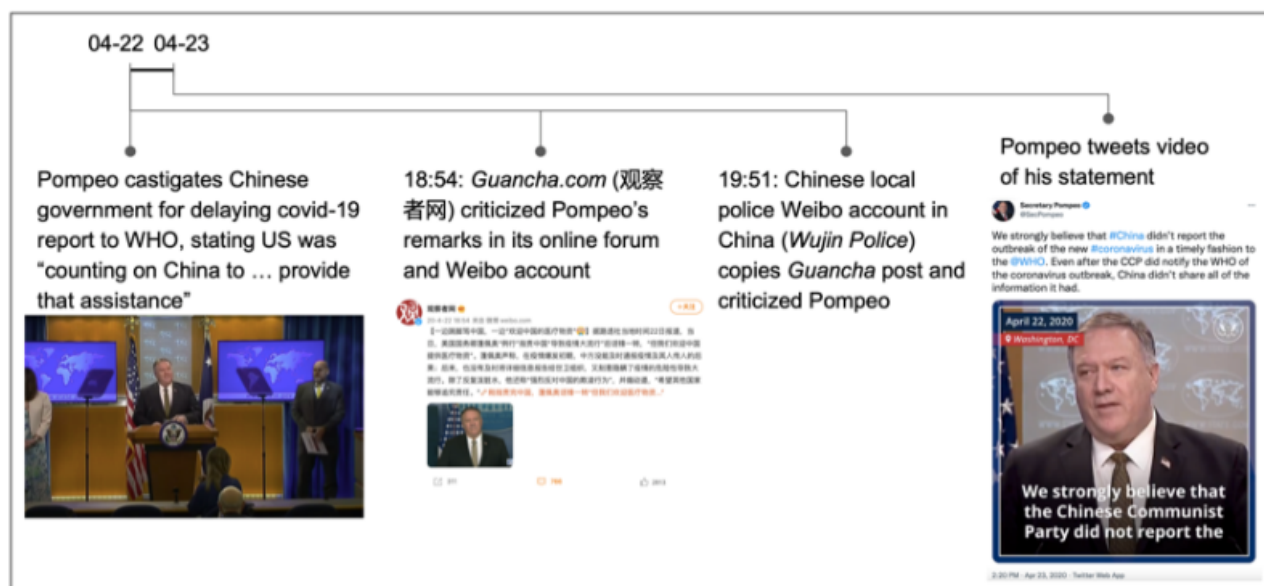


Figure 4: Timeline of Pompeo critique

the Chinese government for delaying its Covid-19 report to the WHO in a press conference. Pompeo states that the US was “counting on China to...provide that assistance.” 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 is copied by other Weibo accounts, including government accounts such as the Weibo account of a lower-level police department. The next day, on April 23, Pompeo tweets a video of his statement from the press conference, which goes gains attention on Twitter. 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 US was building mortuaries. Chinese internet news platforms, Netease News and Sohu News, picked up this reporting, and Weibo users began discussing the topic thereafter. A few days later, a tweet, which gains traction on Twitter, picks up the same story, stating: “China were building coro hospitals England are building mortuary.” This case represents information inflow because it is about an event that 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 traditional media outlets or government facilitates the inflow of information to China. In some instances, content is flowing directly from Twitter to Weibo. For example, on February 28, 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 found no evidence that any traditional Chinese media outlet reported on this. This example suggests that Weibo users living outside of China play a role in information transmission. Other inflows facilitated by Weibo users are sourced from foreign media. For example, a Weibo user posts 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 New segment or saw the tweet of the clip, which was widely viewed. Regardless, this news clip made its way into China because a non-VIP Weibo user posted about it. There was no response from the Chinese government or any reporting by Chinese media outlets.

In line with the previous example, we observe several instances where Weibo users are picking up content from US-based media outlets, including *Buzzfeed*, *New York Times*, and PBS. In addition, overseas Chinese media outlets are another sources of information for Weibo users. Figure 5 illustrates the timeline for an instance of information inflow where overseas Chinese media plays a role. On April 3, 2022, the editor-in-chief of an Indian media platform *Chanakya Forum* posted a news clip from the Urdu television channel, NDTV, 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 publishes a story on the topic. Later that day, *New Tang Dynasty Television*, a US-based media outlet founded by Falun Gong adherents and highly critical of the CCP, posted a story about “underwear masks,” citing both the *Chanakya Forum* editor tweet and *Business Standard* article. Two hours later, a non-VIP Weibo user copies text from the *New Tang Dynasty Television* article citing the *Business*



Figure 5: Timeline of the mask donation controversy

Standard article and asks whether it is fake news. In this instance, a Weibo user without media or government affiliations facilitates the transmission of information into China, but this story also involves overseas Chinese media, media professional with a twitter professional, and media outlets in Pakistan and India.

Foreign entity facilitated inflow to Weibo: The final mechanism is one in which foreign entities post information directly on Weibo. Although many foreign media outlets are blocked in China and are not present on Chinese social media, some foreign entities do have a social media presence in China. In our viral tweets, we find three instances where this mechanism of transmission is at play. First, in February, news that Russia was banning travel from China makes 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, posts to its Weibo account that Australia's calls for investigations into the origins of Covid-19 have sparked intense criticism in China. The same entities posting on Weibo also shared the same content on Twitter, where

this information went viral.

6 Discussion

These results show that the inflow of global information into China is severely limited. A small share of information pertaining to China that capture global public discussion, as measured by viral tweets, make their way to Weibo. However, of the information that is transmitted into the country, state-controlled and domestic media outlets are not the only institutions responsible for transmission. Weibo users without traditional media or government affiliation are also responsible for bringing information from international media sources as well as global social media platforms into the country. This suggests that despite stringent and extensive controls over its information environment, the Chinese government—through government mouthpieces, state-controlled media, and commercialized media—is into the sole gatekeeper of event-driven news and other unmanaged content originating outside of its borders. Social media plays a direct role in the transmission of information inflow

There are several limitations to this study. One is that we only focus on one Chinese social media platform. It may be the case that 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 where relatively larger share of global attention was focused on the country. A final limitation is that we focus on a relatively small number of tweets. Taken together, these limitations caution us against interpreting the quantitative metrics of this paper, e.g., share of co-occurring viral tweets, quantity of inflow, as precise or representative. Rather, the quantitative metrics should be interpreted as indications of general magnitude, and quantitative methods employed allow us to investigate in detail how inflows occur.

This paper contributes methodologically to cross-lingual, cross-platform studies of digital communication. The analytical framework we design for this study can be used to investigate

textual co-occurrence across social media platforms and spanning diverse media contexts. For example, how does content co-occur between Twitter and Reddit, or between Facebook and Vkontakte? How does content co-occur between speeches of American politicians and media reporting about US politics taking place outside of the United States? This could be relevant to many areas of communication, from the spread of misinformation to social mobilization, which have typically focused on single platforms rather than the dynamics of communication throughout the digital ecosystem.

These results have implications for our understanding of the consequences and limitations of government censorship in the digital age. These results confirm existing findings that internet blocks do significantly limit information flows across borders. They extend prior insights to show that the effect of such barriers in preventing the cross-border flow of information persists even when demand for information is high. However, these results also suggest that information barriers, even substantial ones, are not impenetrable when connections to the world remain. Whether transmission occurs because of social media users living outside of China, Chinese social media users who circumvent censorship, or foreign entities that have presence on Chinese social media, social media does allow for the transmission of information that is not institutionally managed by the Chinese state.

The empirical analyses of this paper are focused on the flow of digital information into China. This means that on the one hand, the effects of censorship on preventing cross-border flows of information are high because they are implemented by a state with high capacity. On the other hand, however, the role of social media users in facilitating cross-platform and censorship-circumventing flows of information likely extends beyond China to other regions where government-imposed barriers may not be as stringent or extensive as those in China. In other contexts where governments are limiting transnational flows of information, ordinary users may play an even greater role in overcoming censorship and attempts by the government to control the information environment.

References

- Anderson, D. (2012). Splinternet behind the great firewall of china: Once china opened its door to the world, it could not close it again. *Queue* 10(11), 40–49.
- Bakshy, E., I. Rosenn, C. Marlow, and L. Adamic (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pp. 519–528.
- Boorstin, D. J. (1992). *The image: A guide to pseudo-events in America*. Vintage.
- Bowman, S., G. Angeli, C. Potts, and C. 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. 632–642.
- Cer, D., Y. Yang, S.-y. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar, et al. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.
- Chen, E., K. Lerman, and E. 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, K., A. Chen, J. Zhang, J. Meng, and C. 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, Y. and D. Y. Yang (2019). The impact of media censorship: 1984 or brave new world? *American Economic Review* 109(6), 2294–2332.
- Chidambaram, M., Y. Yang, D. Cer, S. Yuan, Y.-H. Sung, B. Strope, and R. Kurzweil (2018). Learning cross-lingual sentence representations via a multi-task dual-encoder model. *arXiv preprint arXiv:1810.12836*.

- Clayton, R., S. J. Murdoch, and R. N. Watson (2006). Ignoring the great firewall of china. In *International Workshop on Privacy Enhancing Technologies*, pp. 20–35. Springer.
- Cornelison, B. R., S. Al-Mohaish, Y. Sun, and C. 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), 2053–2058.
- Deibert, R. J. (2002). Dark guests and great firewalls: The internet and chinese security policy. *Journal of Social Issues* 58(1), 143–159.
- Deibert, R. J. (2008). The geopolitics of internet control: Censorship, sovereignty, and cyberspace. In *Routledge handbook of Internet politics*, pp. 339–352. Routledge.
- Earl, J., T. V. Maher, and J. Pan (2022). The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances* 8(10), eabl8198.
- Ensaifi, R., P. Winter, A. Mueen, and J. R. Crandall (2015). Analyzing the great firewall of china over space and time. *Proc. Priv. Enhancing Technol.* 2015(1), 61–76.
- Freyburg, T. and L. Garbe (2018). Blocking the bottleneck: Internet shutdowns and ownership at election times in sub-saharan africa. *International Journal of Communication* 12, 3896–3916.
- Friggeri, A., L. Adamic, D. Eckles, and J. Cheng (2014). Rumor cascades. In *proceedings of the international AAAI conference on web and social media*, Volume 8, pp. 101–110.
- Fu, K.-w., C.-h. Chan, and M. 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, K.-w. and Y. Zhu (2020). Did the world overlook the media’ s early warning of covid-19? *Journal of Risk Research* 23(7-8), 1047–1051.

- Goel, S., A. Anderson, J. Hofman, and D. J. Watts (2016). The structural virality of online diffusion. *Management Science* 62(1), 180–196.
- Gohdes, A. R. (2015). Pulling the plug: Network disruptions and violence in civil conflict. *Journal of Peace Research* 52(3), 352–367.
- Golan, G. (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), 323–333.
- 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, J. (2021). *The great firewall of China: How to build and control an alternative version of the internet*. Bloomsbury Publishing.
- Gunitsky, S. (2015). Corrupting the cyber-commons: Social media as a tool of autocratic stability. *Perspectives on Politics* 13(1), 42–54.
- Hassanpour, N. (2014). Media disruption and revolutionary unrest: Evidence from mubarak’s quasi-experiment. *Political Communication* 31(1), 1–24.
- Himmelboim, I., T.-K. Chang, and S. 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, W. R. and M. E. Roberts (2018). How sudden censorship can increase access to information. *American Political Science Review* 112(3), 621–636.
- Howard, P. N., S. D. Agarwal, and M. M. Hussain (2011). The dictators’ digital dilemma: When do states disconnect their digital networks?

- Hu, Y., H. Huang, A. Chen, and X.-L. Mao (2020). Weibo-cov: A large-scale covid-19 social media dataset from weibo. *arXiv preprint arXiv:2005.09174*.
- Jiang, M. (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), 212–233.
- Jiang, M. and K.-W. Fu (2018). Chinese social media and big data: big data, big brother, big profit? *Policy & Internet* 10(4), 372–392.
- Keremoglu, E. and N. B. Weidmann (2020). How dictators control the internet: a review essay. *Comparative Political Studies* 53(10-11), 1690–1703.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751.
- King, G., J. Pan, and M. E. Roberts (2013). How censorship in china allows government criticism but silences collective expression. *American political science Review* 107(2), 326–343.
- King, G., J. Pan, and M. E. Roberts (2014). Reverse-engineering censorship in china: Randomized experimentation and participant observation. *Science* 345(6199), 1251722.
- King, G., J. Pan, and M. 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, R. G. (2000). *The politics of force: Media and the construction of police brutality*. Univ of California Press.
- Lee, J.-A. and C.-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.

- Leskovec, J., L. Backstrom, and J. Kleinberg (2009). Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 497–506.
- Livingston, S. (1997). Clarifying the cnn effect: An examination of media effects according to type of military intervention.
- Livingston, S. and W. L. Bennett (2003). Gatekeeping, indexing, and live-event news: Is technology altering the construction of news? *Political Communication*, 20(4), 363–380.
- Lu, Y. and J. Pan (2022). The pervasive presence of chinese government content on douyin trending videos. *Computational Communication Research*.
- Lu, Y., J. Pan, and Y. Xu (2021). Public sentiment on chinese social media during the emergence of covid19. *Journal of Quantitative Description: Digital Media* 1.
- Marczak, B., N. Weaver, J. Dalek, R. Ensafi, D. Fifield, S. McKune, A. Rey, J. Scott-Railton, R. Deibert, and V. Paxson (2015a). An analysis of china’ s “great cannon” . In *5th {USENIX} Workshop on Free and Open Communications on the Internet ({FOCI} 15)*.
- Marczak, B., N. Weaver, J. Dalek, R. Ensafi, D. Fifield, S. McKune, A. Rey, J. Scott-Railton, R. Deibert, and V. Paxson (2015b). China’ s great cannon. *Citizen Lab* 10.
- Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Miller, B. (2018). The limits of commercialized censorship in china. *Working Paper*.
- Munger, K., R. Bonneau, J. Nagler, and J. A. Tucker (2019). Elites tweet to get feet off the streets: Measuring regime social media strategies during protest. *Political Science Research and Methods* 7(4), 815–834.
- Pan, J. (2017). How market dynamics of domestic and foreign social media firms shape strategies of internet censorship. *Problems of Post-Communism* 64(3-4), 167–188.

- Pan, J. and A. A. Siegel (2020). How saudi crackdowns fail to silence online dissent. *American Political Science Review* 114(1), 109–125.
- Qiang, X. (2019). The road to digital unfreedom: President xi’s surveillance state. *Journal of Democracy* 30(1), 53–67.
- Qin, B., D. Strömberg, and Y. Wu (2017). Why does china allow freer social media? protests versus surveillance and propaganda. *Journal of Economic Perspectives* 31(1), 117–40.
- Qin, B., D. Strömberg, and Y. Wu (2018). Media bias in china. *American Economic Review* 108(9), 2442–76.
- Roberts, H., E. Zuckerman, J. York, R. Faris, and J. Palfrey (2010). Circumvention tool usage report. *Cambridge, MA: Berkman Klein Center for Internet & Society*.
- Roberts, M. E. (2018). *Censored: Distraction and Diversion Inside China’s Great Firewall*. Princeton University Press.
- Stern, R. E. and J. Hassid (2012). Amplifying silence: uncertainty and control parables in contemporary china. *Comparative Political Studies* 45(10), 1230–1254.
- Stockmann, D. (2013). *Media commercialization and authoritarian rule in China*. Cambridge University Press.
- Stockmann, D. and M. E. Gallagher (2011). Remote control: How the media sustain authoritarian rule in china. *Comparative Political Studies* 44(4), 436–467.
- Ugander, J., B. Karrer, L. Backstrom, and C. Marlow (2011). The anatomy of the facebook social graph. *arXiv preprint arXiv:1111.4503*.
- Wagner, B. (2018). Understanding internet shutdowns: A case study from pakistan. *International Journal of Communication* 12(1), 3917–3938.

- Winter, P. and S. Lindskog (2012). *How the great firewall of china is blocking tor*. USENIX-The Advanced Computing Systems Association.
- Wu, H. D. (2000). Systemic determinants of international news coverage: A comparison of 38 countries. *Journal of communication* 50(2), 110–130.
- Xu, B. and E. Albert (2014). Media censorship in china. *Council on Foreign Relations* 25, 243.
- Xu, X., Z. M. Mao, and J. A. Halderman (2011). Internet censorship in china: Where does the filtering occur? In *International Conference on Passive and Active Network Measurement*, pp. 133–142. Springer.
- Yang, Y., D. Cer, A. Ahmad, M. Guo, J. Law, N. Constant, G. H. Abrego, S. Yuan, C. Tar, Y.-H. Sung, et al. (2019). Multilingual universal sentence encoder for semantic retrieval. *arXiv preprint arXiv:1907.04307*.
- Zhang, H. and J. 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.
- Zhu, T., D. Phipps, A. Pridgen, J. R. Crandall, and D. S. Wallach (2013). The velocity of censorship:{High-Fidelity} detection of microblog post deletions. In *22nd USENIX Security Symposium (USENIX Security 13)*, pp. 227–240.

Supplemental Appendix

Measuring Global Information Flow into Chinese Social Media

Appendix

A. Viral Tweets

Index	Tweet content	Date	Retweets	Matched
1	westerners are like wow the chinese govt is doing a bad job with coronavirus while their kids are coming back from school with measles because their government cant/wont be more convincing than some hippie mum with a wordpress site	1/24/20	22857	No
2	Those Wuhans people who gets infected with this disease literally said that even if they die, they want everyone to die together. They even pull doctors face mask & spit on them just so that the doctors would get infected too. Hope us Malaysian will be safe	1/24/20	39695	Yes
3	Wuhan China Corona Virus Update Disturbing Leaked Video shows People literally falling over as they succumb to Virus. 10 cities 33 million people have now been quarantined. 1 case found outside quarantine area This Morning. Retweet The Truth #china #Wuhan #Coronavirus	1/24/20	11672	No
4	this is allegedly a voice note from a nurse in Wuhan, China at the epicenter of where the coronavirus began. heartbreaking if its real.	1/25/20	19707	Yes
5	White people are really making anti-Chinese jokes because of this outbreak as if their ancestors didnt spread catastrophic diseases everywhere because they didnt know how to bathe	1/25/20	57928	No
6	I live in #Wuhan. Today is the third day of the lockdown of the city. Here are some things I see or hear from other citizens these days. #Coronavirus The situation is far more severe than you might think, I think support from international organizations is in grave need.@WHO	1/25/20	20346	No
7	Liang Wudong, just become the first doctor to die from the new Coronavirus in #China A reminder that while everyone else is trying to avoid it, frontline medical staff are risking their lives to cure it + save others. #CoronavirusOutbreak	1/25/20	14175	No
8	Translation of the Wuhan Doctor video that's going viral.. #CoronavirusOutbreak #coronavirus #coronavirusPH	1/25/20	15413	No
9	From my experience as a cabin crew, Wuhan passengers are the worst passenger i have ever encountered. They have no manners, stubborn, uncivilised and dirty. They even push me at the aisle and run my feet over with their luggage wheel	1/25/20	12582	No
10	A man from China is risking his life to spill the truth of the current situation of Wuhan, China in the middle of #CoronavirusOutbreak This man is seeking help from the world. Go watch before China take this video down	1/26/20	87394	No
11	If #coronavirus isnt such a big deal, then why are Italian authorities scanning every single passenger landing in the country? And why are we still allowing Flights originating from China to land in the US ? Something aint right.	1/28/20	36213	No
12	coronavirus has been racialised as a "chinese" illness and for this reason chinese people, regardless of their proximity to wuhan are being treated like carriers of the virus. similar to how ebola was subtly touted as an "african" illness.	1/29/20	75668	Yes
13	A man in China left 500 face masks at a police station - then jogged off before officers could thank him	1/29/20	22236	Yes
14	Coronavirus is declared global emergency by World Health Organization as outbreak continues to spread outside China	1/30/20	12429	Yes
15	RETWEET IF YOU WANT A CHINA-WIDE TRAVEL BAN NOW.	1/30/20	34507	Yes
16	Is it racist to suggest a ban on flights? NO. Is it racist to criticize Wuhan residents for traveling overseas? NO. Is it racist to look at every Chinese person and see disease? YES. Is it racist to refer to Chinese people as chinks, chekwa, etc.? YES	1/31/20	9566	No
17	the flu: *results in 500,000 hospitalizations and 30,000 deaths a year in the usa* americans: *don't even get the fucking flu shot consistently* coronavirus: *resulted in 0 deaths in any country outside of china* americans: *HYSTERICALLY USING IT AS AN EXCUSE TO BE ANTI-CHINESE*	1/31/20	31354	No
18	Instead of blaming them, lets pray for China especially those who risk their lives in helping those who are greatly affected by this crisis. Hate the virus, but don't hate the Chinese people. Remember they are still human and they're struggling too. #PrayforChina	1/31/20	30628	Yes

Index	Tweet content	Date	Retweets	Matched
19	#Coronavirus On a train in Italy. A teenage Chinese boy boards the train. A woman comments loudly: There you go, we are all going to be infected. He replies in perfect Italian with a Roman inflection: Maam, in my whole life Ive seen China only on google maps. Applauses.	2/1/20	31973	No
20	Amdist the Corona Virus, please don't forget about the now 3 Million Muslims suffering in concentration camps in China. They are subject to accounts of torture, forced conversions, mass rape, forced sterilization, Organ harvesting...	2/2/20	8876	No
21	Turkey has sent medical aid to China and told the world Islam does not respond to evil with evil. #Coronavirus	2/3/20	20505	No
22	A Chinese nurse in a coronavirus-hit hospital in Henan province gives her sobbing daughter an air hug A tree is not the forest. The Mainland Chinese citizens need our sympathy and support. They are not the problem. Please be very mindful of the things we say. : Xinhua News	2/5/20	6424	No
23	rumor spread on whats*pp of a employee having coronavirus and this bakery thats known to have lines out the door is getting very little business. chinatown here in SF has decreased by 50% in foot traffic since the virus made headlines. no one in SF has been reported diagnosed.	2/5/20	8904	No
24	Proving he doesn't hold grudges, @realDonaldTrump just announced he's sending a delegation immediately to China consisting of Mitt Romney, Nancy Pelosi, Adam Schiff, and Jerry Nadler.	2/7/20	9028	No
25	The Chinese public have staged what amounts to an online revolt last night after the death of Dr. Li Wenliang. Government officials, biz figures & ordinary folks, have posted numerous messages expressing their grief and anger. It's sth I haven't seen in years	2/7/20	7380	Yes
26	Drone shots of Wuhan. Its literally a Ghost Town. #Wuhan #CoronaVirus	2/7/20	13862	Yes
27	We have coordinated with U.S. organizations to transport more humanitarian relief to people in Wuhan. Personal protective equipment and other medical supplies donated by these organizations can help save lives in #China and help protect people from the #coronavirus.	2/7/20	7608	Yes
28	the number of Chinese people in Britain with confirmed cases of coronavirus is still fewer than the number of Chinese people doing comedy so if you see a Chinese person on the tube and assume they are a comedian that would be statistically less racist	2/9/20	7943	No
29	Chen Qiushi, a citizen journalist who had been doing critical reporting from #Wuhan, the epicenter of the #coronavirus outbreak, went missing on Thursday evening, just as hundreds of thousands of people in #China began demanding freedom of speech online.	2/9/20	7626	No
30	Xiaozhan donated much needed medical supplies and ventilators directly to hospital in Wuhan He does this quietly without saying anything, if the hospital staff didnt post this we would never even know.	2/9/20	5456	No
31	Coronavirus: China is facing shortage of masks & gloves To help China, India has lifted ban on export of personal protection equipments. India has allowed the immediate export of at least two large consignments of protective clothing including masks to China. This is our India	2/10/20	7607	Yes
32	Sick: Border Patrol officers intercepted a passenger from China carrying 200 dead birds in his suitcase. He was trying to enter the country through Virginia with what could have been infected carcasses What was he planning? RT if we should ban all travel to and from China!	2/11/20	18596	No
33	These videos of China spraying disinfectant around #WuHan arent giving me the impression that they have the #coronavirus under control.	2/11/20	4548	Yes
34	Pakistan stands with the people & govt of China in their difficult & trying time and it will always stand by them. We will be extending every material & moral support to China just as China has always stood by us during all our times of trial and tribulation.	2/12/20	5192	Yes
35	A Chinese girl killed two other customers, including a 12-years girl in a fight for a bottle of antiseptic little in a supermarket in China over #coronavirus pressure. Now she is under arrest.	2/12/20	4316	No
36	Many subway stations in China have implemented an AI fever screening system developed by Megvii, an #AI company founded by Tsinghua alumni. The system supports temperature screening from a distance of 3 meters, with a 0.3 degree Celsius margin of error.	2/12/20	4340	Yes
37	It's disgraceful that the Chinese community have faced increased racism and over 50% business losses since the coronavirus outbreak. I've come to Chinatown here in Manchester, a city that has been twinned with Wuhan for 33 years. We stand in solidarity with the Chinese community.	2/13/20	4457	No
38	Gross sial. 5 star hotels too! why China why	2/13/20	7646	No
39	the 7days bulit propaganda hospital in #wuhan shows its true color now. after all its Made in Chian by Chinese Gov #WuhanCoronavirus #coronavirus #CoVid2019	2/15/20	4097	No
40	BREAKING: People in Wuhan, China have been banned from leaving their homes	2/15/20	4196	Yes
41	EXCLUSIVE: China is tracking every movement of its Uighur minority to find reasons to detain them, a leaked document DW obtained from a whistleblower shows. People have been arrested for growing beards or having too many children: #KarakaxList	2/17/20	10928	No
42	Chinese fans of #BTS member #jhope donate funds originally meant to celebrate the stars birthday to charity amid #coronavirus epidemic #JHOPEday @BTS.twt	2/17/20	3422	No
43	Japan's boy band #ARASHI # announced on #Weibo their concert scheduled to open in Beijing in April has to be canceled due to #COVID19. They sang a song in Chinese in the video to show their support to Chinas fight against #coronavirus @arashi5official	2/17/20	2862	Yes
44	BREAKING: Many leading Chinese scientists are starting to speak out and say #COVID-19 (#coronavirus) originated at a government bioweapons research facility in #Wuhan, rather than the widely-held belief that it emerged from the city's #Huanan seafood markets. #WuhanCoronavirus	2/18/20	3173	No
45	This is shocking stuff. Bloomberg News reportedly tries to silence & then ruin female journalist for speaking out about corrupt #China Communist Party. Will @MikeBloomberg be asked about this at Dem Debate?	2/18/20	2577	No
46	Due to coronavirus, Russia has banned all Chinese citizens from entering the country for tourism, work and for private purposes. Full ban.	2/18/20	3792	Yes
47	Facts: Mike Bloomberg said Chinas dictator, Xi Jinping, is not a dictator Joe Biden says that China is not a threat Bernie Sanders said China has done more to address extreme poverty than any country in history Why are Democrats so friendly to Americas greatest enemy?	2/19/20	5506	No
48	The United States condemns the move by China to expel three @WSJ journalists. Mature, responsible countries understand that free press reports facts and expresses opinions. China should not restrict #freepress.	2/19/20	2540	Yes
49	Well, look at that. The Chinese Communist Partys propaganda rag finally admits what Ive said for a month: coronavirus didnt start in Wuhan food market. So where did it originate? Time for answers from CCP.	2/22/20	6590	Yes
50	Wow: Bernie Sanders just shamed the United States by saying that we have more people locked up than Communist China What hes not telling anyone isnt counting the at least 700,000 Uyghur Muslims locked away in Concentration Camps I guess that doesnt fit his narrative	2/23/20	6213	Yes
51	BREAKING: Three Chinese nationals were apprehended trying to cross our Southern border illegally Each had flu-like symptoms. Border patrol quickly quarantined them and assessed any threat of Coronavirus Our weak border is a health risk Close the borderBuild. The. Wall. RT!	2/26/20	28465	No
52	B.I has donated 100,000 masks to help those who are having trouble obtaining masks due to the Coronavirus The masks will be distributed in Korea, China, Thailand, Japan, Indonesia, Vietnam	2/26/20	15748	Yes
53	Ive seen Democrats be tougher on Trump for his coronavirus response than theyve been on China, whose communist government censored and hid the outbreak, leading to its unnecessary spread around the world.	2/27/20	6939	No

Index	Tweet content	Create date	#Retweets	Matched
54	My friend Jimmy Lai, brave voice for free speech & basic rights in #HongKong, was arrested just hours ago by #Beijing puppet government. Transparent attempt to silence pro-democracy voices. While #coronavirus ravages China, repression is Beijings priority	2/28/20	5895	No
55	Airborne Nitrogen Dioxide Plummets Over China #NASA	2/28/20	10724	Yes
56	This is beyond shocking China shut down the lab that published the worlds first genome sequence of the #coronavirus last month, barring its scientists from finding ways to contain the outbreak Their only crime? Publishing the sequence before authorities	2/28/20	7535	No
57	Whats a reason the #coronavirus spread so quickly in China? Governmental incompetence plus the government downplaying the risk. Unfortunately, the @realDonaldTrump Administration has also shown incompetence and is actively trying to downplay the risk of coronavirus.	2/28/20	4881	No
58	Corona virus is like Pasta. The Chinese invented it, but the Italians spread it all over the world.	2/28/20	26049	Yes
59	Tzuyu donated 50 million won to Korea and 300k RMB to China to help prevent spreading of coronavirus. Shes donated quietly and used her own expenses to support both countries at the same time. She has the best heart of all and deserves the best. @JYPETWICE	2/29/20	5166	Yes
60	This blog is the opinion of one solitary American with a unique experience fighting for his childs life in a communist healthcare system (China). Boy, is it worth a read.	3/1/20	8531	No
61	Our antibiotics and penicillin are Made in China This is a national security issue The President is now considering activating the Defense Production Act to rectify the situation I say he should do it	3/2/20	5944	Yes
62	To fight the coronavirus China is undertaking its biggest experiment yet in rule by data: an app that automates quarantines. Alibaba software, using unexplained data, now decides whether millions can leave their home. It also shares data with the police. As it damages both the lungs and immune systems, and can cause irreversible lung damage even if the patient survives. @thespybrief	3/2/20	3397	Yes
63	Countries with largest numbers of #coronavirus diagnoses 1. China 2. South Korea 3. Italy 4. Iran 5. Japan 6. France 7. Germany 8. Spain 9. Singapore 10. USA y'all	3/3/20	5029	No
64	need to stop using photos of RANDOM ASS chinatown aunties and uncles for your #coronavirus headlines especially when it's about a SPECIFIC person becoming infected. THIS UNCLE WAS JUST MINDING HIS OWN DAMN BUSINESS BEING COURTEOUS WEARING A MASK. @thehill do the fck better.	3/3/20	4017	No
65	Chinese doctors say autopsies of #coronavirus victims suggest that #COVID19 is like a combination of #SARS and #AIDS	3/5/20	19673	No
66	China has socialized medicine.	3/6/20	3786	Yes
67	US was late to develop test 4 #Coronavirus. Germany had a test a week after Chinese scientists released the genome info. World Health Org had one shortly after & gave it to 60 countries. US said no to it. In #NYC we have 2700+ ppl quarantine. Most may be infection free.	3/7/20	3677	No
68	THREAD 1/ Of all the articles I read this week on #COVID19, this @nytimes really caught my eye. It was an interview w/ Dr. Bruce Aylward, the lead on the @WHO team that visited China to examine their response. It surprised me for multiple reasons.	3/8/20	4393	Yes
69	Has anyone wondered why the #coronavirus started in China?	3/8/20	5400	No
70	THREAD: Fri Jan 31, 2020, a few weeks before #Coronavirus has officially spread to other countries (which led to the bad stock market week Feb 24-Feb 28), the Trump Admin announced travel restrictions on China. Here is some of the reporting it generated. Take Politico of 2/4/20.	3/8/20	4091	No
71	yt people blaming chinese people for bringing the corona virus to the U.S. is the most hypocritical thing ive ever seen..did you colonizers forget when yall brought the flu, smallpox, and measles here? killing 90% of my indigenous ancestors because i sure didnt	3/9/20	24110	Yes
72	This is going to sound kinda mad, but this week, pls consider making your weekly takeout a Chinese takeaway. My family's businesses have all been impacted hugely by coronavirus panic	3/9/20	60146	No
73	reminder that there are actual businesses owned across america by chinese people that are hurting immensely because of the racism and xenophobia which come with pandemics, and specifically sinophobia following covid-19	3/12/20	41589	No
74	This is the medical team from Wuhan responsible for recovering 50K+ patients in just a few months. They're now risking their lives again to help out Italy. Rag on China all you want for the mishandlings early on but you can't deny that these people are true selfless heroes.	3/13/20	18684	Yes
75	SCOOP: Cuban doctors have discovered an antiviral for #COVID19 and are collaborating with China, Italy, and Venezuela. This is what happens when you invest in science and public health rather than spend \$800,000,000,000 a year blowing up children.	3/13/20	61117	No
76	Chinese medical workers who have been fighting the #coronavirus day and night in Wuhan celebrated the closing of the last temporary hospital in Wuhan. The reported #covid19 cases went from a surge in February of 15,000 in one day to only 15 this week.	3/13/20	82890	Yes
77	A photo of Chinese medical workers and law enforcement officers riding horses to remote places in the Altai Mountains for #coronavirus prevention efforts.	3/13/20	26304	Yes
78	Interesting that a Chinese person would help Americans when most of yall are fucking assholes	3/13/20	107181	No
79	If China loses this legal battle against Gilead to produce the Coronavirus Vaccine they should just disregard it and distribute it worldwide for free anyways, it's the moral choice, fuck the law. The world will see that Gilead are the real villains.	3/14/20	25579	No
80	Never forget who is responsible for the coronavirus pandemic: China. China: -Lied to the world about outbreak, they knew about it in NOVEMBER -Refused outside help to contain it -Failed to lock down borders to prevent it from escaping -Silenced whistleblowers who sounded alarm	3/15/20	35124	No
81	Cuba: We have an antiviral with demonstrably high success rates in treating patients with COVID-19 China: Our studies show this Cuban drug has incredibly high success rates US: If only WE could find a treatment, someone in the PRIVATE SECTOR needs to find a PATENTABLE treatment	3/16/20	60586	No
82	if you guys keep calling it the chinese virus then we get to call it the white mass shooting	3/16/20	105763	No
83	china went from having 15,000 new cases of coronavirus every day to just 15 a day within a month of wants to sit around being quarantined but the racist orange haystack in office and call it china virus while spring breakers are currently snorting eachs spit down in florida	3/17/20	22455	No

Index	Tweet content	Date	Retweets	Matched
84	@realDonaldTrump You see this? Two children and two adults were stabbed at Sams in Midland, Texas, bc ppl associate Asians with the coronavirus. Calling it the Chinese virus encourages fear and racism. I will hold you personally responsible for every hate crime here on after.	3/17/20	27219	No
85	Imagine blaming china when americans are out here doing stupid shit like this	3/17/20	177575	No
86	Citizens of Wuhan can finally hear birds chirping after years, Venices water canals are clear and full of fish, and you can even see the Tatra mountains from Krakw because the smog has lifted. This isnt an apocalypse. Its an awakening.	3/18/20	53932	Yes
87	China followed protocol, quarantined, and has been deep cleaning the city since the epidemic started. Europeans and americans on the other hand have been partying and going to the beach with hundreds of other people. Shut the fuck up.	3/19/20	94720	No
88	"Yes, Im Asian. And yes, I have coronavirus. But I did not get it from China. I got it in America. In New York City. And despite what certain political leaders want to call it, I dont consider the place where it's from as important as the people who are sick and dying."	3/19/20	40855	Yes
89	China were building coro hospitals England are building mortuary	3/19/20	22063	Yes
90	The H1N1 swine flu pandemic that infected up to 1.4 billion people and killed up to 575,000 originated in factory farmed pigs in the United States. Yet nobody ever calls it the American Pig Flu. Lets stop calling the coronavirus the Chinese Virus. Were all in this together.	3/22/20	96013	Yes
91	I'm sitting home in Wuhan and eating free vegetables and fish govt sent me and you tell me it hurts my human rights???	3/23/20	36858	No
92	china are lifting their isolation on the 8th of april, it was enforced on the 23rd january, and people really believe were gonna be out in 3 weeks	3/25/20	31976	Yes
93	BREAKING: Another migrant caravan is heading to the US border While we suffer under the strain of the China Virus, we can't afford to let a single unchecked, unvetted, possibly unhealthy illegal alien enter the country RT if the southern border should be COMPLETELY CLOSED	3/26/20	19344	No
94	China is a fucking liar too. The refuse to accept new Covid19 suspected patients just to keep zero new patient record clean	3/26/20	18283	No
95	Knowledge is power! We launched an online platform for doctors and nurses around the world to exchange ideas, lessons and know-how to fight the virus. We welcome all hospitals to join Chinese hospitals on this open platform One world, one fight!	3/26/20	18164	Yes
96	Spain bought 640,000 testing kits from China. They dont work	3/26/20	14940	Yes
97	Just two weeks ago the head of Emergency at Wuhan Central hospital went public, saying authorities had stopped her and her colleagues from warning the world. She has now disappeared, her whereabouts unknown. #60Mins	3/29/20	22637	No
98	#Wuhan residents estimate, based on calculations of cremations and urns now being returned to families, that between 42k-46k (!) died in city + surrounding areas in the 2.5 months of lockdown. Far more than official figure of 2535 deaths. #COVID19	3/29/20	21183	Yes
99	Chinese scientists have developed a new weapon to combat the #coronavirus. They say they have found a nanomaterial that can absorb and deactivate the virus with 96.5-99.9% efficiency.	3/29/20	14628	No
100	NETHERLANDS: Country's government recalls 600,000 masks from hospitals for being defective after being bought from China	3/29/20	13475	Yes
101	Chinese medics have just arrived in London to help us fight Covid-19. The media wont tell you for some reason.	3/30/20	54050	Yes
102	An Asian family in Texas was attacked due to the racist narrative that Chinese people are the ones spreading the Coronavirus.The perpetrator stabbed a 2 y.o and 6 y.o. LITERAL CHILDREN!! So excuse me if I dont find your racist Coronavirus jokes funny.	4/1/20	17024	Yes
103	BREAKING: China has concealed the extent of the coronavirus outbreak in its country, under-reporting both total cases and deaths, the U.S. intelligence community concluded in a classified report	4/1/20	28376	Yes
104	BREAKING: US Intelligence has officially confirmed what weve known all along China intentionally concealed the extent of the Coronavirus and under-reported the number of cases & deaths China lied, people died, and now its confirmed The CCP must pay RT!	4/1/20	15093	Yes
105	CHINA: City of Shenzhen becomes the first in the country to ban the eating of cats and dogs	4/2/20	16408	Yes
106	Official Chinese statistics put Wuhans coronavirus death toll at 2,563. Evidence coming out of the city suggests it was actually more than 40,000. Our latest:	4/3/20	14146	No
107	USA just discovered the man who manufactured and sold the #coronavirus to #China. Dr #CharlesLieber, head of the chemistry and biology department at #HarvardUniversity, USA. He was just arrested today according to American department sources.	4/4/20	13436	No
108	China promised to send top quality N-95 masks to Pakistan. When the consignment landed, Pakistanis found that China had sent masks made of underwear. Pakistani anchor says China ne Choona laga diya. #ChineseVirusCorona	4/4/20	12776	Yes
109	The U.S. had a program - PREDICT - that identified lethal viruses abroad to prevent their spread to the U.S.. It had found 1,200 viruses (and 160 coronaviruses) in 10 years. It was actively working in China. Two months before the Wuhan outbreak, Trump shut the program down.	4/5/20	44807	No
110	Report: China says it will ship France a billion face masks – but only if they buy their 5G equipment from Huawei:	4/5/20	12743	Yes
111	Who are you working for? China?	4/6/20	13214	Yes
112	Today, China has flown in 300 ventilators, 33 million masks and 1 million gloves for the NHS. The media really needs to start reporting positive stories of international solidarity.	4/6/20	9976	Yes
113	Somewhere in #Wuhan, people moving body secretly as "zero new cases" has become a political task, and communities dare not break the "zero new cases" record , so they arrange to have the body moved at midnight. #CCPVirus #COVID2019 #Coronavirus #CoronavirusPandemic	4/7/20	31159	No
114	Japan just announced it would make the investment to relocate Japanese businesses OUT of China This is a smart, bold move by Japan RT if the United States should immediately do the same thing bring business back home!	4/9/20	17776	Yes
115	Great! First, I would like to speak with Dr. AI Fen. She worked at Wuhan Central Hospital and tried to sound the alarm on the virus. Could you un-disappear her so we could speak?	4/10/20	10487	Yes
116	Medical clothing sent from China has just arrived in France.	4/10/20	23576	No
117	China is a dictatorship that, for decades, enforced a one child per family policy under penalty of forced sterilization. But they can't close down the farmer's market from hell? #CoronaVirus #WetMarkets	4/11/20	41626	No

Index	Tweet content	Date	Retweets	Matched
118	Sad that I have to make a distinction between the Chinese people and the Chinese Communist Party (CCP). The people of China have suffered greatly because of their government. Please be kind to your Asian-American friends. It is not easy for them these days. Lets be our best!	4/11/20	11506	No
119	Bill Maher on Chinese coronavirus: "So when someone says, 'what if people hear Chinese virus and blame China?' The answer is we should blame China. We can't afford the luxury anymore of non-judgment towards a country w/habits that kill millions of people"	4/11/20	10423	No
120	Blacks: don't call it the Chinese virus that's racist China:	4/12/20	28564	No
121	Dear @realDonaldTrump: Your myopic focus on foreign nationals in China is why you failed to stop the virus. You let in hundreds of thousands from Europe, and US travelers from China. And then you failed to test domestically until it was too late. An epic presidential disaster.	4/13/20	9625	No
122	This is a a very obvious point and has been obvious from the beginning but as the global picture fills in over time, there is *absolutely* no fucking way China only had 3300 fatalities from this virus.	4/14/20	12626	No
123	President Trump says U.S. is pulling funding to WHO because the group praised China's "transparency." Below is President Trump's tweet from January 24 praising China's transparency.	4/14/20	12838	Yes
124	So Bill Gates is unhappy that USA halted our funding for the WHO, after they parroted CCP propaganda & aided Chinas Wuhan cover-up. Well, quick math: USA funds \$400m/yr to WHO & Gates is worth \$103bn... he could personally fund the US share for 257 years!	4/15/20	11377	Yes
125	China says the US should "fulfill its obligations" to the World Health Organization and restore funding I say China should fulfill ITS obligations and pay damages to the 22.5 million Americans who lost their job because of the virus China unleashed on the world RT if you agree!	4/16/20	9625	No
126	Bill Gates is seeing his utopian globalization project go up in smoke. We know China lied, Americans died, and a booming economy was fried. Time for real border enforcement and essential products MADE IN USA.	4/16/20	18215	No
127	Chinas embassy in Berlin wrote me an open letter because they werent too happy with our Corona coverage. I responded. Full version:	4/18/20	19135	Yes
128	President Trump has announced that he will end Obama era funding to the Wuhan lab accused of spreading the China virus. Why were we funding it in the first place?	4/18/20	8174	Yes
129	Germanys largest newspaper comes out swinging against China. This is a must watch for US journalists who seem intent on doing Chinas bidding. #BildTV @jreichelt	4/19/20	13238	Yes
130	The most economically, scientifically, and technologically advanced country on the planet is ranked 38th in tests per person. Donald Trump had to work hard to screw up such a massive advantage. His trust in China lost months, as did his belief COVID-19 would magically go away.	4/19/20	8102	No
131	Insanity: \$76,000 in US tax dollars went to a Chinese virology lab in Wuhan last year The money was specifically earmarked to study Coronavirus infections in bats This is sick. RT if not another DIME of US taxpayer money should go to funding the CCPs research	4/20/20	5460	Yes
132	So the Gates Foundation, WHO and Wuhan Institute of Virology have all been hacked and people are logging in via SSH and downloading everything, and I mean everything.	4/21/20	5096	Yes
133	In the past five weeks, the China Virus has caused: Deaths1.5/10,000 Americans New cases2/1,000 Americans Unemployment8/100 Americans 27 million jobs have been lost due to the China Virus This has gone on long enough. It's time to re-open America. RT if you agree!	4/23/20	3658	Yes
134	NEW: In deleted tweets, new HHS spokesman Michael Caputo said the coronavirus was because "millions of Chinese suck the blood out of rabid bats as an appetizer and eat the ass out of anteaters." He said Democrats had a "goal" of "massive deaths."	4/23/20	9152	Yes
135	China unleashed this deadly pandemic on the world and thats where Joe Biden should direct his anger. Only problem is Biden said a rising China is a positive development, his family made millions there and he opposed President Trumps China travel ban which saved American lives	4/23/20	4120	No
136	We strongly believe that #China didnt report the outbreak of the new #coronavirus in a timely fashion to the @WHO. Even after the CCP did notify the WHO of the coronavirus outbreak, China didnt share all of the information it had.	4/23/20	3977	Yes
137	Trump owes the Bank of China millions for helping him acquire one of his most valuable properties. The loan comes due in the middle of the next presidential term. Remember how much he echoed Chinese government propaganda about the virus being contained?	4/24/20	3943	Yes
138	BREAKING: US officials are investigating hundreds of cases of coronavirus brought in by Chinese tourists in January	4/24/20	3919	No
139	ATTENTION: A picture is emerging across 10 major-media investigative reports of the COVID-19 outbreak in the U.S. as a bribery scandal in which Trump looked the other way on intel about the virus in November in order to placate Chinese lenders and get political dirt on Joe Biden.	4/24/20	18410	No
140	Why did India buy Chinese rapid testing kits for #COVID19 when other countries have complained about its accuracy? In the words of a TV showman, the nation wants to know!	4/26/20	3868	No
141	#BREAKING: China's ambassador warns of potentially severe economic consequences if Australia doesn't back down from launching a coronavirus inquiry. Bring. It. On. You don't scare us.	4/27/20	3713	Yes
142	A new study from Wuhan finds the coronavirus can linger in the air in crowded spaces, as scientists struggle to determine exactly how it spreads	4/28/20	9761	Yes
143	BREAKING: The Trump Administration has suspended all US funding to the Wuhan Institute of Virology	4/28/20	9175	No
144	Meanwhile in China, they are treating Africans worse and blaming them because of the virus	4/28/20	13196	No
145	The same people who say Chinese Virus is hateful & racist are eerily silent about Bill De Blasios insanely anti-Semitic message to the Jewish Community last night	4/29/20	5147	No
146	Dr. Fauci backed controversial Wuhan lab with millions of U.S. dollars for risky coronavirus research	4/29/20	6849	Yes
147	@realDonaldTrump says he has seen intelligence that gives him a high degree of confidence that the coronavirus originated from the Wuhan Institute of Virology - though he would not go into detail I cant say	4/30/20	3592	No
148	Our mission at the @StateDept is to protect Americans from threats around the world. As we seek information about the origins of #COVID19, we call on #China to be transparent and cooperative. When countries engage in disinformation it creates risk. We need reliable partners.	4/30/20	2757	No
149	Well, THERE'S a shocker. Spy Agencies Told White House Theres No Evidence to Prove Virus Began in Wuhan Lab, Says Report via @thedailybeast	4/30/20	2844	No
150	China: We discovered a new virus. America: So what? China: It's Dangerous America: It's only a Flu China: Wear a Mask America: Don't wear a Mask ...	4/30/20	23897	No

B. Classifier for Weibo Data

To train a classifier that could identify covid-19-related sentiment, we first sampled 5,484 posts from the Weibo-COV dataset. We define a post related to COVID-19 if the post contains one topic about the SARS-CoV-2 virus, the resulting COVID-19 epidemic, as well as the emergence and impact of the outbreak. Three Mandarin-speaking research assistants annotated each post and we derived the final annotation result for each post by the majority vote. With this training data, we fine-tuned the pretrained Chinese BERT with the Whole Word Masking model (Chinese BERT-wwm-ext), which adapts whole word masking on Chinese text as an upgrade of BERT and more data sources for training to get better performance (Cui et al. 2020). After tuning for different hyper-parameters (e.g. learning rate, dropout rate), we selected hyper-parameters that maximizes F1 score. Our fine-tuned model reached 0.97 in accuracy and an F1 score of 0.99 in classifying whether a Weibo post is related to covid-19. Among 8.3 million original posts from Weibo-COV dataset, we identify 6.7 million covid-19-related posts.

C. Retrieval

C.1 Similarity calculation

Formally, let $T_i = [t_1, \dots, t_m]$ be a translated source tweet that comprises a sequence of m Chinese word tokens and $W = W_1, \dots, W_N$ be the pool of Weibo posts. Each $W_j = [w_1, \dots, w_n]$ is a Chinese word sequence. To retrieve the top-10,000 candidate Weibo posts for a given tweet T_i , we measure similarity of T_i and W_j by Algorithm 1 for every post in W and obtain the top-10,000 candidate posts according to the scores.

Algorithm 1: Word2vec-based similarity for a tweet and a Weibo post

Input: Translated source tweet $T = [t_1, \dots, t_m]$, Weibo post $W = [w_1, \dots, w_n]$

Output: Similarity score

```

tweet_vectors=<>;
weibo_vectors=<>;
for  $k = 1$  to  $m$  do
     $\mathbf{v}_{t_k} = \text{word2vec}(t_k)$ 
    tweet_vectors.append( $\mathbf{v}_{t_k}$ );
end
for  $k = 1$  to  $n$  do
     $\mathbf{v}_{w_k} = \text{word2vec}(w_k)$ 
    weibo_vectors.append( $\mathbf{v}_{w_k}$ );
end
 $\mathbf{v}_T = \text{mean\_vector}(\text{tweet\_vectors})$ ;
 $\mathbf{v}_W = \text{mean\_vector}(\text{weibo\_vectors})$ ;
 $\text{score} = \text{cosine\_similarity}(\mathbf{v}_T, \mathbf{v}_W)$ ;
return  $\text{score}$ ;

```

C.2 Time window validation

We used different window sizes (5 to 10) to retrieve data with the same search engine pipeline. We found that when the window size increases from 5 to 10, the majority of posts still overlap with the posts retrieved by the ± 5 day time window.

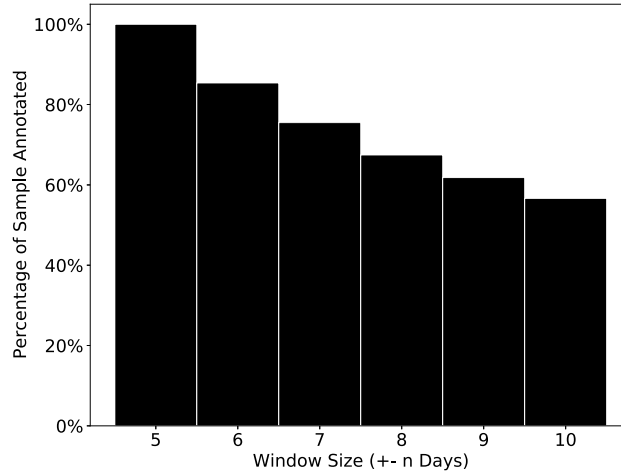


Figure A1: Overlapped posts with different window sizes

D. Robustness Check Using Pre-Censored Weibo Data

It may be the case that more viral tweets appeared on Weibo, but we do not find them because the corresponding Weibo posts were censored after they were posted. To assess the impact of this type of post-hoc censorship, we replicated all analyses, from classifier training to inflow investigation, on the Weiboscope Covid-19 dataset, a near real-time, pre-censored dataset collected by researchers in Hong Kong (Fu and Zhu 2020). The Weiboscope dataset ends on Feb 27, 2020, so we use our system to search for the 32 tweets in the Weiboscope dataset that were not found to have any Weibo matches in the WeiboCOV dataset.

We find two additional viral tweets with matching Weibo posts. One matched tweet represents information outflow from China. Two Weibo posts reported on the death of Liang Wudong, the first doctor to die in China from Covid-19, and this content found its way to Twitter and gained widespread attention. The other matched tweet represents information inflow. A Weibo public intellectual (verified as a celebrity account) copied reporting by Reuters on the relative lack of Covid-19 controls at French airports compared to other European airports where scanners were used to take the temperature of passengers arriving from China. These results suggest that post-hoc censorship does not alter the pattern of information inflow we find using the WeiboCOV dataset.

References

- Cui, Y., W. Che, T. Liu, B. Qin, S. Wang, and G. Hu (2020, November). Revisiting pre-trained models for Chinese natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, Online, pp. 657–668. Association for Computational Linguistics.
- Fu, K.-w. and Y. Zhu (2020). Did the world overlook the media’ s early warning of covid-19? *Journal of Risk Research* 23(7-8), 1047–1051.