COVID-19 Increased Censorship Circumvention And Access To Sensitive Topics in China

Keng-Chi Chang1, William R. Hobbs2, Margaret E. Roberts1,1, and Zachary C. Steinert-Threlkeld3

1Department of Political Science, University of California San Diego, La Jolla, CA 92093; 2Department of Human Development and Department of Government, Cornell University, Ithaca, NY 14850; 3Department of Public Policy, Luskin School of Public Affairs, University of California Los Angeles, Los Angeles, CA 90095

This manuscript was compiled on July 1, 2021

Crisis motivates people to track news closely, and this increased engagement can expose individuals to politically sensitive information unrelated to the initial crisis. We use the case of the COVID-19 outbreak in China to examine how crisis affects information seeking in countries that normally exert significant control over access to media. The crisis spurred censorship circumvention and access to international news and political content on websites blocked in China. Once individuals circumvented censorship, they not only received more information about the crisis itself but also accessed unrelated information that the regime has long censored. Using comparisons to democratic and other authoritarian countries also affected by early outbreaks, the findings suggest that information seeking spurred by crisis can undermine censorship and provide abrupt access to previously hidden content.

COVID-19 | CHINA | CENSORSHIP | TWITTER

Significance Statement

We study the impact of crisis on information seeking in authoritarian regimes. Using digital trace data from China during the COVID-19 crisis, we show that crisis motivates citizens to seek out crisis-related information, which subsequently exposes them to unrelated and potentially regime-damaging information. This gateway to both current and historically sensitive content is not found for individuals in countries without extensive online censorship. While information seeking increases during crisis under all forms of governance, the added gateway to previously unknown and sensitive content is disproportionate in authoritarian contexts.

K.C., W.H., M.R., and Z.S-T. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no competing interest.

1To whom correspondence should be addressed. E-mail: meroberts@ucsd.edu
has large effects on the opinions of the general public and the trustworthiness of their government. While the impact on the autocrat may not be immediate and could be outweighed by a successful, rapid government response to the crisis, gradual accumulation of censored information over time could have negative impacts on authoritarian resilience.

### The COVID-19 Crisis in China

On December 31, 2019, officials in Wuhan, China confirmed that a pneumonia-like illness had infected dozens of people. By January 7, 2020, Chinese health officials had identified the disease – a new type of coronavirus called novel coronavirus, later renamed COVID-19. By January 10, the first death from COVID-19 was reported in China, and soon the first case of COVID-19 was reported outside of China, in Thailand. As of December 2020, COVID-19 has infected over 91,000 people in China with over 4,500 deaths, and at least 73.5 million people worldwide with over 1.6 million deaths.†

While initial reports of COVID-19 were delayed by officials in Wuhan (39), Chinese officials took quick steps to contain the virus after it was officially identified and the first deaths reported. On January 23, 2020, the entire city was placed under quarantine – the government disallowed travel to and from the city and placed residents of the city on lockdown (40). The next day, similar restrictions were placed on 9 other cities in Hubei province (41). While Hubei province and Wuhan were most affected by the outbreak, cities all over China were subject to similar lockdowns. By mid-February, about half of China – 780 million people – were living under some sort of travel restrictions (42). Between January 10 and February 29, 2020, 2,169 people in Wuhan died of the virus (43).

### The Effect of Crisis on Information Seeking and Censorship Circumvention

We use digital trace data to understand the effect of the COVID-19 crisis on information seeking. Table 1 summarizes the empirical tests conducted in this paper.§ First, we show that the crisis increased the popularity of virtual private network (VPN) applications, which are necessary to jump the Great Firewall, downloaded on iPhones in China. We also show that the crisis expanded the number of Twitter users in China, which has been blocked by the Great Firewall since 2009. The crisis further increased the number of page views of Chinese language Wikipedia, which has been blocked by the Great Firewall since 2015. We also show that the areas more affected by the crisis – such as Wuhan and Hubei Province – were more likely to see increases in circumvention.

Next, we show that the increase in circumvention caused by the crisis not only expanded access to information about the crisis, but also expanded access to information that the Chinese government censors. On Twitter, blocked Chinese language news organizations and exiled dissidents disproportionately increased their followings from mainland China users.

---

†Stockmann (2012) (23) provides evidence that consumers of newspapers in China are unlikely to go out of their way to seek out alternative information sources. Chen and Yang (2019) (33) provided censorship circumvention software to college students in China, but found that students chose not to evade the Firewall unless they were incentivized monetarily. Roberts (2018) (25) provides survey evidence that very few people choose to circumvent the Great Firewall because they are unaware that the Firewall exists or find evading it difficult and bothersome.


Replication materials will be posted on Dataverse.
On Wikipedia, sensitive pages such as those pertaining to Chinese officials, sensitive historical events, and dissidents showed large increases in page views due to the crisis. Last, the fourth subsection shows that these dynamics do not occur on Wikipedia in countries with similar crisis but where Wikipedia is uncensored.

### Table 1. Empirical Tests

<table>
<thead>
<tr>
<th>Question</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Do individuals circumvent censorship more during crisis?</td>
<td>VPN ranking; increased use of blocked services; new Twitter users.</td>
</tr>
<tr>
<td>2. Do individuals access crisis information?</td>
<td>Wikipedia traffic about current leaders; new mainland China followers for certain account types.</td>
</tr>
<tr>
<td>3. Do individuals access non-crisis sensitive information?</td>
<td>Wikipedia traffic to blocked pages; new mainland China followers for activists and foreign political figures.</td>
</tr>
<tr>
<td>4. Do these same dynamics occur in democracies and less censored environments?</td>
<td>Wikipedia page views in German, Italian, Persian, and Russian.</td>
</tr>
</tbody>
</table>

**Crisis Increased Censorship Circumvention.** We show that censorship circumvention increased in China as a result of the crisis using data from application analytics firm AppAnnie, which tracks the ranking of iPhone applications in China. While most VPN applications are blocked from the iPhone App Store, we identified one still available on it. Around the time of the Hubei lockdown, its rank popularity increased significantly and maintained that ranking (top panel of Figure 1).5

Concurrent with the increase in popularity of the VPN application is a sudden increase in popularity of Facebook, Twitter, and Wikipedia applications, as Figure 1 shows. These increases indicate that those jumping the Firewall as a result of the crisis were engaging in part with long blocked websites in China – Twitter and Facebook have been blocked since 2009 and Chinese language Wikipedia since 2015.

This finding is consistent with data we collected directly from Twitter and Wikipedia. The top panel of Figure 2 shows the number of geolocating users in China posting in Chinese in the time period of interest. Immediately following the lockdown, Chinese language accounts geolocating to China increased 1.4 fold, and post-lockdown, 10% more accounts were active from China than before. The bottom panel of Figure 2 shows that the crisis also coincided with increases of new users, indicating that increases are due to new users and not dormant ones reactivating.6 We provide a rough, back-of-the-envelope calculation for the absolute size of these effects. If there were 3.2 million Twitter users in China (44) prior to the COVID-19 pandemic and the 10% increase in usage applies generally to Twitter users (i.e. not just those geotagging), then

---

5To protect the application and its users, we are not disclosing its name or the exact ranking, though results are available for review upon request.

6Note that increase in popularity is not comparable across applications because popularity is measured in terms of ranks. More highly ranked applications (like Facebook and Twitter) may need many more downloads to achieve a more popular ranking.

SI Appendix S2 provides more detail, and Figure A1 shows trends per province.

---

Fig. 1. Download Rank of iPhone Application in China: Facebook, Twitter, and Wikipedia. Data from AppAnnie.
320,000 new users joined Twitter because of the crisis, including users who do not post or post publicly. We assess this estimate in SI Appendix S4 using the estimated fraction of posts in Chinese that are geotagged (1.95%) and the total number of unique Twitter users in our sample (47,389 users posting in Chinese and in China).

Data from Wikipedia on the number of views of Wikipedia pages by language matches the App Annie and Twitter patterns.** We measure the total number of views for Chinese language Wikipedia by day from before the coronavirus crisis to the time of writing. Figure 3 reveals large and sustained increases in views of Chinese language Wikipedia, beginning at the Wuhan lockdown and continuing above pre-COVID levels through May 2020. Views of all Wikipedia pages in Chinese increased by around 10% during lockdown and by around 15% after the first month of lockdown. This increase persisted long after the crisis subsided. In absolute terms, the total number of page views increases from around 12.8 million views per day in December 2019, to 13.9 million during the lockdown period (January 24 through March 13), and up to 14.7 million views per day from mid-February through the end of April.

**This page view data is publicly available: https://dumps.wikimedia.org/other/pagecounts-az/migrated/
media and officials, who have become increasingly vocal on the banned platform (45). New users of Wikipedia might only seek out information about the virus and not about politics. If the crisis produced a gateway effect, we should see increases in consumption of sensitive political information unrelated to the crisis.

**Types of Twitter accounts mainland Twitter users started to follow as a result of the crisis.**

We use data from Twitter to examine what types of accounts received the largest increases in followers from China due to the crisis. For this purpose, we identify 5,000 accounts that are commonly followed by Chinese Twitter users. The Materials and Methods and SI Appendix S2 detail how we identified these accounts.

We assigned each of the 5,000 popular accounts into one of six categories: 1) international sources of political information, including international news agencies; 2) Chinese citizen journalists or political commentators, which include non-state media discussions of politics within China; 3) activists, or accounts disseminating information about politics in the U.S., Taiwan, or Hong Kong; 4) accounts disseminating pornography; 5) state media and political figures; and 6) entertainment or commercial influencers. Categories 1, 2, and 3 are accounts that might distribute information sensitive to the Chinese government, such as international media blocked by the Great Firewall (e.g., *New York Times Chinese* and *Wall Street Journal Chinese*); Chinese citizen journalists and political commentators such as exiled political cartoonist Badiucao and currently detained blogger Yang Hengjun; and political activists such as free speech advocate Wen Yunchao and Wu’er Kaixi, former student leader of the 1989 Tiananmen Square Protests. Accounts in Category 4 are pornography, which we consider sensitive because it is generally censored by the Chinese government, but not politically sensitive like Categories 1-3. Accounts in Category 5 include accounts linked to the Chinese government, including the government’s news mouthpieces *Xinhua* and *People’s Daily*, as well as the Twitter accounts of Chinese embassies in Pakistan and Japan. Category 6 is also not sensitive, as these accounts mostly do not tweet about politics, but instead are entertainment or commercial accounts or accounts of non-political individuals.

We want to understand how the coronavirus crisis affected trends in follower counts of each of the six categories, and in particular, compare how the crisis affected the followings of categories 1-3 to those in categories 5 and 6. We therefore downloaded the profile information of all accounts that began following these popular accounts after November 1, 2019. We then use the location field to identify which of the 38,050,454 followers are from mainland China or Hong Kong (see SI Appendix S2 for more details).

Because Twitter returns follower lists in reverse chronological order, we can infer when an account started following another account (46). For the accounts in the six categories, we compare the increase in followers from mainland China to the increase in followers from Hong Kong accounts relative to their December 2019 baselines; we chose Hong Kong because it is part of the PRC but is not affected by the Firewall. The ultimate quantity of interest is the ratio of these two increases. If the ratio is greater than one, then the increase in following relationships is more pronounced among mainland Twitter users as compared to those from Hong Kong.

Figure 5 shows this ratio by category-day. Relative to Hong Kong, the crisis in mainland China inspired disproportionate increases in the number of followers of international news agencies, Chinese citizen journalists, and activists (some of whom might otherwise, without exposure on Twitter, be obscure within China, especially ones who have been banned from public discourse for a long time) – all information considered sensitive that has long been censored. In comparison, there is only a small increase in mainland followers of Chinese state media and political figures during the lockdown period and a slight decrease for non-political bloggers and entertainers. Figure 6 reports the regression estimate for the relative ratio of number of new followers (akin to a difference-in-differences design with Hong Kong as control group and December 2019 as pre-treatment period). The result is the same.

We then demonstrate that the result does not depend on the choice of comparison group and that the relative increase starts no earlier than Wuhan lockdown. Figure A6 in SI Appendix S2 conducts a placebo test by running weekly regressions, showing that the relative increase in followers in China starts precisely during the week of lockdown. Figures A7, A8, and A9 show that the same pattern holds with alternative comparison groups such as overseas Chinese in Taiwan and the United States.

SI Appendix S4 provides effect size estimates. There, we roughly estimate that around 320,000 new users came from China. Further, based on December 2019 follower growth rates, 53,860 excess accounts follow citizen journalists and political bloggers; 52,144 for international news agencies. By the end of the lockdown, citizen journalists and political bloggers benefit from 3.63 times the number of followers they otherwise would have had; activists, 2.97. Importantly, 88-90% of
Fig. 5. Increases in Twitter Followers from China vs Hong Kong By Category

New Followers Compared to Baseline, China / Hong Kong

- Activists or US / Taiwan / Hong Kong Politics
- Citizen Journalists / Political Bloggers
- Pornography Accounts
- State Media or Chinese Officials
- Non-Political Bloggers or Entertainment Accounts

Note: Gain in followers from mainland China compared to Hong Kong across six types of popular accounts, relative to December 2019 trends. Ratios here approximate the incidence rate ratios estimated in the models for Figure 6. A value greater than 1 means more followers than expected from mainland China than from Hong Kong. Accounts creating sensitive, censored information receive more followers than expected once the Wuhan lockdown starts. Accounts that are not sensitive or censored, such as state media or entertainment, do not see greater than expected increases.

Fig. 6. Increases in Twitter Followers China vs Hong Kong By Category (Regression Estimate)

Relative Size of New Followers, China / Hong Kong

- International News Agencies
- Citizen Journalists / Political Bloggers
- Activists or US / Taiwan / Hong Kong Politics
- Pornography Accounts
- State Media or Chinese Officials
- Non-Political Bloggers or Entertainment Accounts

Note: Incidence rate ratios shown above are from negative binomial regressions of number of new followers on the interaction between indicator variables for ‘in lockdown period’ and ’in mainland China’, with December 2019 as control period and Hong Kong as control group.

the followers from China follow accounts in these categories one year later, and these rates are higher than for accounts which start following in the weeks after the end of the Hubei lockdown. In addition, Figure A10 in SI Appendix S5 shows that new users from China persist in tweeting at the same rates as those from Hong Kong and Taiwan.

Types of Chinese language Wikipedia pages that received the most attention.

To better understand patterns of political views in the Wikipedia data, we leverage existing lists (see Materials and Methods for additional details) to categorize the Chinese language Wikipedia views into three different categories: 1) Wikipedia pages that were selectively blocked by the Great Firewall prior to Wikipedia’s move to https (after which all of Wikipedia was blocked), 2) pages that describe high level Chinese officials, and 3) historical leaders of China since Mao Zedong. Whereas we would expect that a crisis in any country should inspire more information seeking about current leaders in Category 2, only if crisis created a gateway to historically sensitive information would we expect proportional increases in information seeking about historical leaders in Category 3 or information about sensitive events that were selectively blocked by the Great Firewall on Wikipedia prior to 2015 in Category 1.

Figure 7 shows the increase in page views for each of these categories on Chinese Wikipedia relative to the rest of Chinese language Wikipedia. We find that the lockdown not only increased views of current leaders (purple), but also views of historical leaders (yellow) and views of pages selectively blocked by the Great Firewall (red). Tables A2 and A3 in SI Appendix show specific pages disproportionately affected by the increase in views of Wikipedia. While pages related to coronavirus experienced a jump in popularity, other unrelated sensitive pages including the “June 4 Incident,” “Ai Weiwei,” and “New Tang Dynasty Television” (a television broadcaster affiliated with Falun Gong) also experienced an increase in page views.

For more detail on this analysis as well as the Wikipedia pages that received the largest absolute and relative increases in traffic, see SI Appendix S6.

Comparison with Other Countries Affected by the Crisis.

Since information seeking during crisis is common (1), we investigate Wikipedia data from other countries affected by the crisis. We show that the gateway effect of crisis on historically sensitive information is unique to the currently censored web-pages in China. For comparison, we focus on Iran, another authoritarian country affected by COVID-19 that previously censored Wikipedia (but does not any longer), and Russia, an

‡‡ Using data from https://www.greatfire.org/.
¶¶ These lists are based on offices in the CIA World Facebook. We use this list for ease of comparisons with other countries and remove the Ambassador to the United States from each list. China’s list is available here (and there are links to leaders of other countries on the same page): https://www.cia.gov/library/publications/resources/world-leaders-1/CH.html, excluding Hong Kong and Macau.

The June 2020 increase in China is due to the anniversary Tiananmen Square protests. Our claim is not that only the COVID-19 crisis causes increases in views of sensitive content. That the same behavior is observed around another crisis event supports this paper’s argument.
To make the comparison, we use lists of current leaders from these countries (based on office lists in the CIA World Factbook, see Materials and Methods), and create lists of historical leaders using de facto country leaders since World War II (see Table A4 in the appendix for a list of these titles and offices). All of these countries were affected by the crisis in late February or early March and Italy imposed relatively stringent lockdowns. Therefore, we expect increases in information seeking for current leaders, as citizens begin to pay more attention to current politics as the crisis hits. However, none of these countries block Wikipedia. Information seeking about the current crisis therefore should not act as a gateway to information about historical events or controversies, as these pages are always available to the public.

Table 2 shows these results. While overall Wikipedia views and page views of current leaders increase in three out of four comparison countries, only in China do historical leaders increase disproportionately and consistently throughout the whole time period. That is, we see an overall effect on information-seeking throughout the world, including for historical leaders; in China, we see larger increases for historical leaders compared to Wikipedia page views in general. The small increases in historical political leader page views in German and Italian did not correspond with the start of the COVID-19 crisis or their respective lockdowns (Figure 7).

Further, we do not see increased attention to pages previously blocked in Iran (47) during the crisis – Wikipedia pages that can now be accessed without restriction in Iran.

In SI Appendix S6.1, we replicate these results for much larger sets of 1) historical leaders and 2) ‘politically sensitive’ pages (pages related to the pre-https blocked pages in Iran and China, and political opposition pages in Russia). We expand these sets of pages using Wikipedia2vec (48), and find that very broad information-seeking about historical leaders and politically sensitive topics occurred only in China.

Table 2. During the lockdown period, Wikipedia views in Chinese increased relative to overall views for politically sensitive Wikipedia pages and political leader pages, as well as for historical political leaders.

<table>
<thead>
<tr>
<th>Change: Overall</th>
<th>Blocked, pre-https relative to overall:</th>
<th>Leaders</th>
<th>Historical Leaders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>1.09</td>
<td>1.19</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td>(1.05 - 1.12)</td>
<td>(1.09 - 1.22)</td>
<td>(1.67 - 2.07)</td>
</tr>
<tr>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Russian</td>
<td>1.37</td>
<td>1.46</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>(1.37 - 1.46)</td>
<td>(0.79 - 0.89)</td>
<td>(0.80 - 1.05)</td>
</tr>
<tr>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.20</td>
</tr>
<tr>
<td>German</td>
<td>1.23</td>
<td>1.73</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>(1.18 - 1.28)</td>
<td>(1.48 - 2.02)</td>
<td>(0.82 - 0.99)</td>
</tr>
<tr>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.03</td>
</tr>
<tr>
<td>Italian</td>
<td>1.40</td>
<td>1.47</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>(1.40 - 1.53)</td>
<td>(1.40 - 1.53)</td>
<td>(1.27 - 2.07)</td>
</tr>
<tr>
<td></td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Note: Incidence rate ratios shown above are from a negative binomial regression estimating the daily number of views within a category in the lockdown period compared to December 2019 relative to the number of views across the rest of Wikipedia compared to December 2019 (using the same difference-in-difference specification as the Twitter follower analysis). Observations are the total views per category by day. 95% confidence intervals are shown in parentheses, and p-values are shown in the third row for each language. See the SI for over-time ratios by day for all comparison languages (Figure A11), and for the dates of the lockdowns used (Table A4). German and Italian pages of historical leaders (shown in orange in the figures above) saw several large and short-lived spikes in views not clearly related to those countries’ lockdowns. Figures A12, A13, A14, and A15 in the SI replicate these results for much larger sets of Wikipedia pages, including Russian language pages related to opposition leaders and movements (which did not see broad increases in views).

**Discussion**

Crisis in highly censored environments creates a gateway to sensitive, censored information unrelated to the crisis. Like in democracies, consumers of information in autocracies seek out information and depend on the media during crisis.
ever, in highly censored environments, increased information seeking also incentivizes censorship circumvention. This new ability to evade censorship allows users to discover a wider variety of information than they may have initially sought.

While evaluations of responses to an ongoing crisis and comparisons to other governments’ responses to the same crisis may have benefited government officials in China in this particular circumstance (12), beyond these evaluations, increased access to historical and long-censored information, as documented here, has the potential to dampen positive changes in trust or compound negative changes in trust, and may also contribute to easier access to uncensored information about a government in the future. Natural disasters, including epidemics, tend to alter trust in government officials. When a policy response is perceived as efficacious, support for the level of government perceived to have directed the response increases (13, 49). On the other hand, neglectful responses can induce subsequent protest participation (10). In China, the average effect of natural disasters from 2007–2011 was to decrease political trust, and internet users have decreased baseline levels of political trust (13, 14).

While the results do not link the COVID-19 crisis gateway effect to the political fortunes of the Chinese government, they do suggest that crisis has the ability to undermine censorship in highly censored environments. While in normal times censorship can be highly effective, it may create unintended side effects of facilitating access to sensitive information during crisis.

Materials and Methods

Application download rank data. Download rank data for Facebook, Twitter, Wikipedia, and the VPN app come from application analytics firm AppAnnie (https://www.appannie.com), which tracks the popularity of iPhone application downloads in China. While most VPN applications are blocked from the iPhone Apple Store (and there are other means of obtaining VPNs), we identified one still available on it. VPN download rank shown in the text is for that VPN application. This data contains the ranking of an application for Wikipedia, its rank within the Reference App category, rather than the number of downloads. To protect the VPN application and its users, we do not disclose its name or the exact ranking, though results are available for review upon request.

Twitter data. For the Twitter analyses, we collected 1,448,850 tweets (101,553 accounts) from mainland China from December 1, 2019 until June 30, 2020. These tweets were identified using Twitter’s POST statuses/filtered endpoint. Our analyses are limited to the 367,875 that were posted in Chinese (47,389 accounts that posted in Chinese, 43,114 that had names or descriptions in Chinese).

The Twitter follower analysis examines accounts that Twitter users from China commonly follow. To find those accounts, we randomly sampled 5,000 users geolocated to China. For each of these users, we gathered the entire list of whom they follow, their Twitter "friends." From these 1,818,159 friends, we extracted the 5,000 most common accounts. We also selected only accounts that were Chinese language accounts or had Chinese characters in their name or description field to ensure that we were studying relevant accounts: those disseminating information easily accessible to most Chinese users. SI Appendix S2 provides more detail.

We downloaded the profile information of all accounts that began following these popular accounts after November 1, 2019. Because Twitter returns follower lists in reverse chronological order, we can infer when an account started following another account (46). We then use the location field to identify which of these 38,050,454 followers are from mainland China or Hong Kong (see SI Appendix S2 for more details). We downloaded all new followers of non-pornography accounts and all new followers of a random selection of 200 pornography accounts (the majority of the accounts were pornography). This sampling allows us to estimate the impact of the coronavirus on pornography while decreasing our requests to the Twitter API.

Mobility data. Human mobility data is publicly available from Baidu Qianxi (https://qianxi.baidu.com/2020/), which tracks real-time movement of mobile devices and is used in studies of human mobility and COVID-19 containment measures (50). Our robustness checks use data across China during the Lunar New Year period in both 2020 and 2019. We scraped the daily within-city movement index (an indexed measure of commute population relative to the population of the city), as well as daily moving out index (an indexed measure based on the volume of population moving out of the province relative to the total volume of migrating population on that day across all provinces in China). See Section S3 for more details.

Wikipedia data. Data on the number of Wikipedia page views is publicly available here: https://dumps.wikimedia.org/other/pagecounts-ez/ (March 2020). To better understand patterns of political views in the Wikipedia data, we use existing lists to categorize the Chinese language Wikipedia views into three different categories: 1) Wikipedia pages that were selectively blocked by the Great Firewall (https://www.greatfire.org/) maintains a list of websites censored by the Great Firewall prior to Wikipedia’s move to https, after which all of Wikipedia was blocked, 2) pages about high level Chinese officials (using offices listed in the CIA World Factbook https://www.cia.gov/library/publications/resources/world-leaders-1/CH.html, excluding Hong Kong and Macau as well as the Ambassador to the United States), and 3) historical ‘paramount’ leaders of China since Mao Zedong.

In comparing multiple languages and countries, we use the same offices listed in the CIA World Factbook to create lists of current leaders from Iran, Russia, Italy, and Germany (for office holders as of February 2020), and create daily lists of historical leaders using de facto country leaders since World War II. See Table A4 in the appendix for a list of these titles and offices, as well as the lockdown start and end dates used for the language by language Wikipedia page view models displayed in Table 2. The list of pages of Wikipedia pages blocked in Iran was published by (47).

In SI Appendix S6.1, we replicate the Wikipedia page view page results for much larger sets of 1) historical leaders and 2) ‘politically sensitive’ pages (pages related to the pre-hits blocked pages in Iran and China, and political opposition pages in Russia). We expand these sets of pages using Wikipedia2vec (48).

Models. Incidence rate ratios for the follower analyses and the Wikipedia page view analyses are from negative binomial regressions. In the follower analysis, this models the number of new followers per day, with a separate model for each account category. Independent variables are ‘in lockdown period’ and ‘in mainland China’, and the effect of interest is the interaction between these indicator variables (i.e. a difference-in-difference), with December 2019 as control period and Hong Kong as control group. The Wikipedia page view analyses use the same specification, reporting the coefficient for ‘in lockdown period’ and ‘in page set’ (current leader, historical leader, previously blocked) relative to December 2019 and relative to page views for Wikipedia. Observations are pre-censored for a page view event per day. Figures displaying (log scale) ratios of followers/Wikipedia page views approximate coefficients from these negative binomial regressions. Negative binomial regressions were estimated using the MASS library in R.

Increases in geolocated Twitter activity (unique users) by day and by province were modeled using a five-term polynomial regression (by day) for time trends after the Hubei lockdown and a mean without any time trend prior to lockdown (see Figure A1 for a province by province visualization of this model). The points in Figure 2 are predicted values by province for the first day of lockdown and day 30 of lockdown.


512 | Chang et al.
42. J Griffiths, A Woodyatt, 80 million people in china are living under travel restrictions due to the coronavirus outbreak. CNN (2020).
43. P Belluck, Coronavirus death rate in wuhan is lower than previously thought, study finds. New York Times (2020).
44. P Mozur, Twitter Users in China Face Detention and Threats in New Beijing Crackdown (2019).
45. L Zhou, Chinese officials have finally discovered twitter. what could possibly go wrong? South China Morning Post (2020).
47. N Nazari, C Anderson, Citation Filtered: Iran’s Censorship of Wikipedia. Iran Media Program (2013).
52. E Kinetz, Army of fake fans boosts China’s messaging on Twitter (2021).
Supplementary Materials for “Crisis is a Gateway to Censored Information”

S1. Twitter Activity by Province

Figure A1 shows the number of unique, geolocating users who are tweeting in Chinese by province. The x-axis is the number of months before (negative) or after (positive) the initial coronavirus lockdown in Hubei province. The blue line is a pre-lockdown average for x less than 0 and a five term polynomial regression for x greater than or equal to 0 (where 0 is the first day of Hubei’s lockdown). The points in Figure 2 are the values of the blue line by province for x equals 1/30 (first day of lockdown) and x equals 1 (day 30 of lockdown).

Fig. A1. Increases in Geolocated Twitter Activity by Province (modeled)
S2. Twitter Data

From a global sample of tweets with GPS coordinates, we found the 1,448,850 tweets from China from December 1, 2019 through June 30, 2020, 367,875 of which are in Chinese. This corpus contains 101,553 unique users, 43,114 of whom had names or descriptions in Chinese. These dates were chosen to encompass a baseline period and the height of COVID-19 in China. This corpus is used for Figures 2 and 4, evaluating the impact of the lockdown on tweeting behavior.

For the follower analysis (Figures 5 and 6), we sample 5,000 of these 43,114 accounts. For these 5,000 random users in China, we download who they follow, their “friends” in Twitter parlance. From these friends, we identify the 5,000 most commonly followed accounts that are either a Chinese language account or have Chinese characters in their name or description field. Of these 5,000 most common friends, the vast majority were pornography accounts. We therefore hand-categorized the accounts into pornography or not pornography. We keep the 354 non-porn accounts and sample 200 from the remaining 4,646 porn accounts.

We then download the followers of these 554 accounts. We identify 38,050,454 total followers. For each, we identify the location of the users. Because very few of these followers have geolocated information, we rely on the language of their Twitter status and their self-reported location to distinguish between mainland and overseas followers. We only include users whose status language is Chinese in order to study only Chinese language followers of these accounts. Followers are classified as Mainland Chinese if the location field contains the name of a Chinese city, town, or province. Followers are classified as from Hong Kong if the location field contains the name of a district in Hong Kong. Followers are classified as Taiwanese if the location field contains the name of a Taiwanese city, county, or district. Followers are classified as US if if the location field contains the name of states or state abbreviations (in capital letters).
S3. Mobility and Twitter Usage

To better understand the relationship between lockdown and Twitter usage, we scrape the publicly-available human mobility data from Baidu Qianxi (https://qianxi.baidu.com/2020/), which tracks real-time migration (including moves in & out of provinces and within city movements) across China during the Lunar New Year period in both 2020 and 2019.

Figure A2 plots the average within city movement index in both 2020 (real black line) and 2019 during the same period in the Chinese Lunar New Year (dotted line). Specifically, since the New Year’s day is on February 5 in 2019 and January 25 in 2020, we shifted the dates in 2019 backwards for 12 days to match the dates in 2020. Red vertical line indicates the day of Wuhan lockdown. One can see that almost all provinces experienced a huge decrease in human mobility after January 23 in 2020, compared to the same period in 2019. In 2019, we only see significant decreases in mobility in Beijing, Shanghai, and Tianjin.

Within City Movement Index (Black: 2020, Dotted: 2019, Dates Adjust for Lunar New Year)

We also validate that the increase in geolocated Twitter users is correlated with the decrease in human mobility. The left panel of Figure A3 plots this correlation. Let $M_{i,t}$ denote the mobility index for province $i$ on date $t$. The x-axis plots the decrease in within city movement index from January 22 (the day before the Wuhan lockdown) to February 22, $M_{i,Jan22} - M_{i, Feb22}$. The y-axis plots the increase in geolocated Twitter users for 30 days after the Wuhan lockdown compared to the average number of geolocated Twitter users in a province before the Wuhan lockdown. This shows that the more reduction in human mobility, the more increase in geolocated Twitter users, comparing to the levels before the lockdown. Hubei province experience the most reduction in mobility and most increase in the number of geolocated Twitter users.

In Figure A1 we see increases in geolocated Twitter users in most provinces except Beijing and Shanghai. One explanation for this is that Twitter users in Beijing and Shanghai left the cities during the outbreak. Mobility data supports this explanation. The right panel of Figure A3 plots the relationship between moving out of a province on January 23, the day of Wuhan lockdown, and the increase of number of geolocated Twitter users on the same day, compared to the average number of geolocated Twitter users in a province before the lockdown. One can see that the more people moving out, the less jump in Twitter user on the day.
of lockdown. Beijing, Shanghai, and Guangdong all experienced large outflows of individuals on the day of Wuhan lockdown.

Fig. A3. Reduction in within city movement and increase in geolocated Twitter users during the month of Wuhan lockdown (left); degree of moving out and increase in geolocated Twitter users on the day of Wuhan lockdown (right).

Note: The left panel plots the correlation between decreased mobility and increased geolocated Twitter users during the first 30 days of Wuhan lockdown. The x-axis plots the decrease in within city movement index from January 22 (the day before Wuhan lockdown) to Feb 22. The y-axis plots the increase in geolocated Twitter users for 30 days after the Wuhan lockdown compared to the average number of geolocated Twitter users in a province before the Wuhan lockdown. The right panel plots the relationship between moving out of province and the increase of geolocated Twitter users on January 23, the day of Wuhan lockdown. Estimates and day of lockdown are drawn from a five term polynomial regression on the number of unique geolocated Twitter users per day after the lockdown. These province-by-province polynomials are displayed over the raw data in Figure A1.

Since the period of Wuhan lockdown overlaps with the Chinese Lunar New Year, increased Twitter usage could partly be due to general boredom during the New Year. To explore New Year versus pandemic effects, we normalize both mobility and number of Twitter users in 2020 by those in the same period in 2019. To do so, we first adjust the dates in 2019 backwards for 6 days to match the dates of 2020 Lunar New Year. Then, we create normalized mobility and Twitter usage. Specifically, denote $M_{i,y,t}$ the mobility index and $T_{i,y,t}$ the Twitter usage for province $i$ in year $y$ on date $t$. The normalized mobility index would be

$$M_{i,2020,t}/M_{i,2019,t}$$

and the normalized Twitter usage would be

$$T_{i,2020,t}/T_{i,2019,t}.$$  

We then plot the weekly change in mobility and Twitter usage after Wuhan lockdown, comparing to the period before Wuhan lockdown. Figure A4 shows the plots. In mathematical notations, for the first week of Wuhan lockdown, we plot

$$M_{i,2020,\text{Week 1}}/M_{i,2019,\text{Week 1}}$$

$$M_{i,2020,\text{Week 0}}/M_{i,2019,\text{Week 0}}$$

on the x-axis and

$$T_{i,2020,\text{Week 1}}/T_{i,2019,\text{Week 1}}$$

$$T_{i,2020,\text{Week 0}}/T_{i,2019,\text{Week 0}}$$ on the y-axis. This is shown in the top left panel in Figure A4. The other panels shows the corresponding ratios for the 2nd, 3rd, and 4th week, respectively (all relative to the week before lockdown).

Figure A4 shows that we still find effects of reduced mobility on increased Twitter usage, after adjusting for the decrease in movement driven purely from New Year, in the early periods of lockdown (at least for the first week of lockdown, the correlation for the second week is not statistically significant). In the 3rd and 4th weeks, we find a general increase in Twitter usage in most provinces, regardless of the relative decrease in mobility in these weeks. In other words, the mobility-induced effect specific to Wuhan lockdown fades out in around 2 weeks, and there’s a general increase in Twitter usage across China that is not related to reduced human mobility. This pattern suggests that the increase in Twitter usage is not driven only by people’s staying at home because, if that is the case, we expect to see a continued relationship between relative reduction in mobility and increase in Twitter usage, as other Provinces started to announce stay-at-home orders. This pattern is also not driven only by New Year because we should not expect to see an overall increase in Twitter usage after normalizing with the same New Year period in 2019.
**Geo-located Twitter Users by Day**

2020 versus 2019 aligned by lunar calendar

Lunar New Year (day of)

Note also that the first week of Twitter use in 2020 was not much higher than in 2019 because 2019 saw a very large number of posts on Chinese Lunar New Year. This increase was presumably related to New Year related posts, and these celebratory posts did not increase to the same extent during the start of the COVID-19 pandemic.
S4. Effect Size

S4.1. New Twitter Users. This section provides rough estimates of absolute increases in Twitter use in China, and sections below expand it to consider increased Twitter followings and increased Wikipedia use. Note that these are estimates for increased usage on only these sites, which require that new users from Mainland China (where these sites are blocked) 1) create an account to view Twitter content and 2) use cookies (to be recorded in the Wikipedia unique device data). Other sites that do not require accounts could have seen larger increases, and the Wikipedia unique device counts are underestimates.

The top panel of Figure 2 shows a 10% long-term increase in the number of geolocating users from China. In 2019, as reported in (44), Professor Daniela Stockman of the Hertie School of Governance surveyed 1,627 internet users in China and found .4% of them use Twitter; the article reports that number as 3,200,000. Roughly, if the same 10% increase applies to all users from China and the long-term increase reflects a new pool of users (the number of unique geotagging users in our sample in May 2020 was around 10% higher than in December 2019), then 320,000 new users joined Twitter because of the crisis.

We can assess this estimate by considering 1) the fraction of (posting) users who geotag and 2) the number of unique geotagging users in our data. For 1), using a sample of 100 hours of non-geolocated tweets from 2019.01.01-2020.12.31, we found 37,957 in Chinese. Assigning location to these tweets using the same code that was used to assign location to followers of the most commonly followed accounts, we then found that 1.79% of tweets and 1.95% of users from China geotag. For 2), we find that 47,389 unique Twitter users geotagged (in Chinese and in China) in our sample (note, however, that our 1% sample captures approximately 56% of tweets that are geotagged). Dividing this number by 0.0195 gives us 2.4 million Twitter users, suggesting that somewhere around 70% (\(\frac{47,389}{2.4 \times 10^6}\)) of geotagging Twitter users in China publicly geotagged posts and were in our sample.

Though this number is small in the context of China’s 1.4 billion inhabitants, it is nonetheless important for three reasons. First, the effects in this paper are a minimum effect size for Twitter since accounts do not have to use geolocation or provide an accurate self-reported location in their profile. Second, the effects documented herein focus only on one banned platform (Twitter) and website (Wikipedia), and there is no reason to think the same behavior did not occur on other banned platforms like Facebook, Telegram, and Instagram as well as banned websites such as Reddit or The New York Times. Third, the Chinese government behaves as though these relatively small numbers threaten it. Since 2018, it has become increasingly repressive in response to comments its citizens make on platforms unavailable in China. Recently, several individuals from China have been arrested for comments made on platforms such as WhatsApp (owned by Facebook, unavailable inside the Great Firewall) and Twitter (51). The government has also started large influence campaigns on social media platforms that are unavailable domestically, including Twitter (52). If the behaviors documented in this paper were immaterial, then we believe the government would not put such a priority on attempting to control speech on these platforms.

S4.2. Followers. In addition to causing new people to join Twitter, the crisis caused more people to follow accounts posting sensitive content. Here, we estimate the number of surplus followers from China and show that they persist after the crisis, perhaps at greater rates than users who follow after.

Figure A5 shows the absolute number of excess followers (top) and its ratio (bottom). The absolute number is the total number of new followers minus the total number of predicted new followers based on the December daily average growth rate per category; the bottom panel divides the new follower count by the predicted number of new followers. Several interesting patterns emerge. First, the crisis clearly causes all account types to gain followers; some, such as pornography and international news agencies, may even have served as early warning indicators since they receive excess followers before the Wuhan lockdown. Second, the categories with the most excess followers, citizen journalists/political bloggers and international news agencies, are exactly those people who would seek out in a crisis. By the end of March, 53,860 more accounts follow citizen journalists/political bloggers than would have happened without the crisis; for international news agencies, 52,144. Third, normalizing for the expected number of new followers reinforces that attention was paid to sensitive categories. Extra, early attention is paid to the citizen journalists and activist categories (which received almost 4 times as many new followers during the lockdown as we would expect based on December’s following rate), while international news agencies’ importance decreases to third place. Normalizing emphasizes the increased attention activists receive since they have relatively fewer followers than the other categories. Fourth, Chinese accounts increase their following of state media or Chinese officials once Hubei’s lockdown lifts, though from a low base.

Importantly, these excess followers persist a year after the lockdown. To make this claim, we crawled the follower list of the same popular accounts starting on May 31, 2021, more than one year after the first crawl, and assigned location using the same procedure as before. Comparing the 2021 follower lists to 2020 shows which followers stopped following the popular accounts. We then calculate the percentage of the 2020 followers that persist in 2021 by account type, follower location, and date. Table A1 shows these results.

Accounts from China that start following the popular accounts during the lockdown period persist at the same to slightly higher rates than those that start following before or after then. 87.31% of accounts from China that start following international news agencies before the lockdown persist versus 89.09% that start following after the lockdown. The difference is especially
stark for citizen journalists/political bloggers. Finally, since older followers should have a lower persistence rate since more time has passed, it is striking that accounts that start following during lockdown have higher persistence rates than newer accounts, those that start following during the seventeen days after the lockdown ends. The increased exposure to sensitive content persists after the crisis passes at rates equal to or greater than for non-crisis periods.

S4.3. Number of unique devices accessing Wikipedia with cookies enabled. Wikipedia tracks the number of unique devices that have accessed its site each day and month using a ‘privacy-sensitive access cookie’ (https://dumps.wikimedia.
Table A1. Persistence of Followers by Account Type and Period Following Starts

<table>
<thead>
<tr>
<th>Account Type</th>
<th>Pre-Lockdown</th>
<th>Lockdown</th>
<th>Post-Lockdown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>China</td>
<td>Hong Kong</td>
<td>Taiwan</td>
</tr>
<tr>
<td>International News Agencies</td>
<td>87.31</td>
<td>87.71</td>
<td>85.51</td>
</tr>
<tr>
<td>Citizen Journalists / Political Bloggers</td>
<td>72.84</td>
<td>79.58</td>
<td>78.73</td>
</tr>
<tr>
<td>Activists or US / Taiwan / Hong Kong Politics</td>
<td>78.27</td>
<td>76.74</td>
<td>76.45</td>
</tr>
<tr>
<td>Pornography Accounts</td>
<td>85.56</td>
<td>84.28</td>
<td>83.00</td>
</tr>
<tr>
<td>State Media or Chinese Officials</td>
<td>82.72</td>
<td>81.38</td>
<td>84.62</td>
</tr>
<tr>
<td>Non-Political Bloggers or Entertainment Accounts</td>
<td>73.75</td>
<td>72.94</td>
<td>65.66</td>
</tr>
</tbody>
</table>

Note: Each cell is the percent of followers from April 2020 that still follow the six account types (row) in May 2021, by follower location and period the follower started following the account. The lockdown period is January 23, 2020 - March 13, 2020. Post-lockdown refers to March 14-April 1.

By design, this number does not count devices not accepting cookies through private browsing (as we might expect from users accessing Wikipedia from within Mainland China) and so underestimates access (see https://diff.wikimedia.org/2016/03/30/unique-devices-dataset/). However, this estimate still provides some perspective on the number of individuals who might be accessing the Chinese language version of Wikipedia over time. For the Chinese language version of Wikipedia, 40.8 million devices accessed the site during December 2019 and 42.8 million per month during January, February, and March 2020, an increase of approximately 2 million devices. 3.3 million devices accessed the Chinese language Wikipedia per day in December 2019 and 3.54 million, an increase of approximately 300 thousand devices. These differences are somewhat smaller when comparing to the last half of 2019 (during ongoing protests in Hong Kong) – an increase of 1 million unique devices monthly during lockdown compared to July through December 2019, and an increase of 200 thousand devices daily.
S5. Robustness Checks

In this section, we assess whether the result is driven by (1) a misspecified treatment period, (2) the choice of comparison group, or (3) an increase of followers due to only a few accounts.

Figure A6 plots the estimates based on regressions for each week before and after the lockdown. We do not see pre-treatment increases in number of followers in China, and the increase starts precisely on the week of lockdown.

Figures A7 and A8 verify that the results in Figure 5 are not due to choosing Hong Kong for the denominator. Figure A7 uses accounts from Taiwan for the denominator, and Figure A8 uses accounts in the United States. These accounts are from any user using Chinese and their self-reported location is in Taiwan or the United States. Figure A9 reports the regression estimate for the relative ratio of number of new followers (akin to a Difference-in-differences design with December 2019 as control period and Hong Kong/Taiwan/China as control group). The result is not driven by Hong Kong-specific trend of news cycles.

One might also curious about whether new users stayed on Twitter at different rates. Figure A10 plots the daily unique active users since their sign up dates in 2020. We don’t find that users from one location stayed on Twitter longer than others.
Fig. A6. Increases in Twitter Followers from mainland China versus Hong Kong by Week

Relative Size of New Followers by Week, China / Hong Kong

Note: Incidence rate ratios shown above are from Negative Binomial regressions of number of daily new followers on the interaction between dummy for each week and China, with December 2019 as control period and Hong Kong as control group.
Fig. A7. Increases in Twitter Followers from China versus Taiwan

New Followers Compared to Baseline, China / Taiwan

Note: Gain in followers from mainland China compared to Taiwan across six types of popular accounts, relative to December 2019 average. A value greater than 1 means more followers than expected from mainland China than from Taiwan. Accounts creating sensitive, censored information receive more followers than expected once the Wuhan lockdown starts. Fewer Taiwanese users follow Chinese state media or government officials than Hong Kong users do.
Fig. A8. Increases in Twitter Followers from China versus US

New Followers Compared to Baseline, China / US

Note: Gain in followers from mainland China compared to US across six types of popular accounts, relative to December 2019 average. A value greater than 1 means more followers than expected from mainland China than from the US. Accounts creating sensitive, censored information receive more followers than expected once the Wuhan lockdown starts. Fewer US users follow Chinese state media or government officials than Hong Kong users do.
Fig. A9. Increases in Twitter Followers from China versus Others (Regression Estimate)

Relative Size of New Followers, China / Control Group

Note: Incidence rate ratios shown above are from negative binomial regressions of number of new followers on the interaction between indicator variables for ‘in lockdown period’ and ‘in mainland China’, with December 2019 as the control period.

Fig. A10. New Users Stay on Twitter at the Same Rates across Locations

Decay of Daily Unique User Activity

Note: This figure plots the daily unique active users since their sign up date using the user panel across locations. A user is considered active between their sign up date and the last day they tweet (before July 2020). We find that users stay on Twitter at the same rate across locations.

Page view data analyzed in this paper is publicly available and hosted here: https://dumps.wikimedia.org/other/pagecounts-ez/merged/. In replication materials, we will additionally provide processed and aggregated versions of the page view data so that this paper’s findings can be more quickly replicated than would be possible with the above page view files.

Below, we show the top Wikipedia pages by relative and absolute increases in page views within each of the categories we analyzed in the main text, as well as pages about the coronavirus and COVID-19 (pages considered: coronavirus, COVID-19, ventilator, flu, pneumonia, fever). The largest relative increases among these pages and for current leaders were related to coronavirus – the COVID-19 pandemic Wikipedia page and the head of China’s National Health Commission. Top increases for pages that were blocked prior to the introduction of https on Wikipedia (after which China blocked all pages) were for an activist who criticized China’s pandemic response.

<table>
<thead>
<tr>
<th>Overall</th>
<th>Blocked</th>
<th>Current Leaders</th>
<th>Historical Leaders</th>
</tr>
</thead>
<tbody>
<tr>
<td>马晓伟_官员 (36.67)</td>
<td>许志永 (16.78)</td>
<td>马晓伟_官员 (36.67)</td>
<td>胡锦涛 (8.1)</td>
</tr>
<tr>
<td>Xu Zhiyong</td>
<td>Ma Xiaowei</td>
<td>Hu Jintao</td>
<td></td>
</tr>
<tr>
<td>许志永 (16.78)</td>
<td>2 月 17 日 (9.01)</td>
<td>孙春兰 (9.38)</td>
<td>邓小平 (1.75)</td>
</tr>
<tr>
<td>Xu Zhiyong</td>
<td>February 17</td>
<td>Sun Chunlan</td>
<td>Deng Xiaoping</td>
</tr>
<tr>
<td>孙春兰 (9.38)</td>
<td>西藏人民起义日 (7.04)</td>
<td>李克强 (2.52)</td>
<td>江泽民 (1.65)</td>
</tr>
<tr>
<td>Sun Chunlan</td>
<td>Tibetan Uprising Day</td>
<td>Li Keqiang</td>
<td>Jiang Zemin</td>
</tr>
<tr>
<td>2 月 17 日 (9.01)</td>
<td>台湾 (5.21)</td>
<td>王岐山 (2.50)</td>
<td>华国锋 (1.44)</td>
</tr>
<tr>
<td>February 17</td>
<td>Taiwan</td>
<td>Wan Qishan</td>
<td>Hua Guofeng</td>
</tr>
<tr>
<td>西藏人民起义日 (7.04)</td>
<td>圆周率 (4.15)</td>
<td>鸟捷 (2.45)</td>
<td>毛泽东 (1.15)</td>
</tr>
<tr>
<td>Tibetan Uprising Day</td>
<td>Pi Day</td>
<td>Xiao Jie</td>
<td>Mao Zedong</td>
</tr>
<tr>
<td>肺炎 (5.38)</td>
<td>艾未未 (3.9)</td>
<td>韩正 (2.14)</td>
<td></td>
</tr>
<tr>
<td>Pneumonia</td>
<td>Ai Weiwei</td>
<td>Han Zheng</td>
<td></td>
</tr>
<tr>
<td>台湾 (5.21)</td>
<td>李长春 (3.71)</td>
<td>胡锦涛 (1.99)</td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>Li Changchun</td>
<td>Hu Chunhua</td>
<td></td>
</tr>
<tr>
<td>流行性感冒 (5.04)</td>
<td>新唐人电视台 (3.51)</td>
<td>许家 (1.88)</td>
<td></td>
</tr>
<tr>
<td>Influenza</td>
<td>New Tang Dynasty Television</td>
<td>Miao Wei</td>
<td></td>
</tr>
<tr>
<td>圆周率 (4.15)</td>
<td>唐柏桥 (3.34)</td>
<td>习近平 (1.80)</td>
<td></td>
</tr>
<tr>
<td>Pi Day</td>
<td>Tang Baqiao</td>
<td>Xi Jinping</td>
<td></td>
</tr>
<tr>
<td>艾未未 (3.93)</td>
<td>长春国难战 (2.31)</td>
<td>杨晓波 (1.73)</td>
<td></td>
</tr>
<tr>
<td>Ai Weiwei</td>
<td>Siege of Changchun</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A2. Top relative increases for Wikipedia pages January 24 through March 13 compared to December 2019.

Note: Labels are limited to: blocked, leader, historical leader, COVID/coronavirus. All other pages are aggregated as “rest of Wikipedia”.

In Figure A11, we show the trajectories for categories matching those analyzed for China – current leaders (using offices listed in the CIA World Factbook), historical leaders, and, in Iran, pre-https blocked Wikipedia pages (47).

Russia, Germany, and Italy (none of which block Wikipedia) saw increases in current leader views without accompanying increases in historical leader views. Germany and Italy did see spikes views of in historical leader pages in the weeks leading up to the relaxation of lockdowns in early May, but saw no change during the initial crisis.

German and Russian political pages also saw an increase in political leader page views prior to their own lockdown, and approximately at the same time as the announcement of widespread lockdown in Italy (see Figure A11).
Table A3. Top absolute daily increases for Wikipedia pages January 24 through March 13 compared to December 2019.

Note: Studying average daily increases standardizes the different lengths of time before versus after the Wuhan lockdown. Labels are limited to: blocked, leader, historical leader, COVID/coronavirus. All other pages are aggregated as “rest of Wikipedia”.

Table A4. Lockdown dates

Note: This table lists the time periods we use to estimate the effects of crisis lockdowns on Wikipedia page views, along with the offices considered for the historical leaders analysis. Each country’s lockdown involved various levels of lockdown for different parts of the countries, and so there is no single time period for us to analyze. Figure A11 displays Wikipedia page views with solid, vertical gray lines for the periods listed above.
Fig. A11. Views of Blocked, Current Leader, and Historical Leader Wikipedia Pages in Other Countries
S6.1. Analysis of an expanded set of historical political pages and ‘politically sensitive’ pages using Wikipedia2vec. We replicated our analyses of historical Wikipedia pages and “politically sensitive” (pages specifically blocked in China and Iran prior to the introduction of https) Wikipedia pages by expanding the original set of pages to a much larger set of related pages. We expanded these lists of pages using Wikipedia2vec (48). This analysis assesses 1) whether the increase in views of Chinese historical leaders (and the lack of increase for other languages) was a relatively narrow effect or much broader one than what we see for that small set of pages and 2) whether a broader set of ‘politically sensitive’ pages are able to uncover increases in page views in Iran and Russia. Because, unlike China and Iran, Russia did not provide a list of politically sensitive pages (by blocking specific pages on Wikipedia), we assess Russian views of political opposition pages related to a) Alexei Navalny (arguably the most prominent opposition leader in Russia) and b) a list of opposition-related pages which we mine to discover increases in views – after this, we then looked to closely related pages to assess whether single page increases represented broader trends or were isolated and potentially random occurrences.

Wikipedia2vec finds similar pages (along with other entities and words) on Wikipedia by analyzing the network of page links, the co-occurrence of words, and the occurrences of specific words on pages. This analysis is accomplished using the same approach as in word2vec (53). At a high level, this approach involves placing words and entities into a shared n-dimensional space such that words and entities are placed closed together if they frequently share contexts (e.g. page links or co-occurring words). Shared contexts must occur beyond what would be expected from the frequency of a word or page, which is accomplished through ‘negative sampling’ – predicting the co-occurrence of words and entities against frequency weighted sampling of negative cases. Once in the n-dimensional space, we can find the most similar entities (pages) for any given entity (or the mean of a set of entities’ projections) using cosine similarity – and we can incorporate dissimilar entities in this calculation by flipping the sign those entities’ locations when calculation the mean of a set of entities. Wikipedia2vec can be run with hyperparameters that affect the size of the n-dimensional space and the exact weighting scheme used in negative sampling. Our estimations for each language used the same default settings as the wikipedia2vec pre-trained embeddings provided at https://wikipedia2vec.github.io/wikipedia2vec/pretrained/ with the number of dimensions set to 100.

For each set of pages (historical leaders, blocked pages, current leaders, Russian opposition pages), we found the top 100, 250, 500, and 1,000 pages that were most similar according to Wikipedia2vec. For historical leaders, we expanded to historical leader related pages not related to the current leader – using the current leader as a dissimilar case – and we expanded to current leader related pages not related historical leaders. With these sets, we re-estimated the changes in views during the first 30 days of lockdown. This excludes the late lockdown spikes in historical leader views visible for German and Italian (visible in Figure 7 in the main text and in Figure A11 above). Note that the German increases in views of historical leaders (in those figures and in the results below) began well prior to the German lockdown (in February).

For Alexei Navalny specifically, we also manually collected a list of Wikipedia pages closely related to his opposition activities, and re-estimated changes in lockdown for each of these pages. The list of Russian opposition-related pages checked for increases is shown in Table A5.

For the pages previously blocked in Iran, (47) provided labels for the category of each blocked page: academic, artistic/cultural, drugs or alcohol, human rights, media and journalism, other, political, profane non-sexual, religious, and sex and sexuality. We also replicated our analyses for the Persian language set subset to page categories human rights, media and journalism, and political.

The findings from these analyses are displayed in Figures A12, A13, and A14, and we also show findings for current leaders in Figure A15. In each cluster of estimates in the top panels, the first is the estimate for the seed pages (and is colored yellow for historical pages, red for blocked ‘politically sensitive’ pages, purple for current leaders). These exclude estimates for seeds which we mined for increases (i.e. selected them only because we saw increases during lockdowns – after a Bonferroni multiple testing correction). Given many tests when looking for increases, these pages have estimates that could very likely reflect random variation in page views, even though we are relatively certain that the increases were not zero, given the multiple testing correction.

Across the results, we see 1) that the increase in historical leader page views in Chinese also applies to a much larger set of pages and page views (bottom panel) and 2) we do not see comparable increases in historical leader pages or politically sensitive page views in other languages, despite increased interest in current leaders across almost all languages analyzed.

In the manual Alexei Navalny analysis, we see that views for his page specifically did rise and that this rise was comparable to what we see for historical leaders in Chinese. However, unlike the broad increase in views in China, we did not see similar increases for any other Navalny-related pages – and only one of the 9 considered showed a statistically significant increase without a multiple-testing correction (falling just short of significance at a 0.05 level after a Bonferroni correction for 9 tests).
Relative increase in historical leader views: first 30 days of lockdown

Leader set order:
- Seeds
- 100 most similar pages
- Not related to current leader
- 500
- 1,000

Increase in views of historical leader pages (compared to overall increase in views)

Relative increase in historical leader views:
- First 30 days of lockdown

Number of Historical Leader Page Views by Page Set

German increases in views of historical leaders began in February (see Figure A11 above)
Relative increase in "sensitive" political page views:
first 30 days of lockdown

Number of Sensitive Political Page Views by Page Set

FIG. A13. Changes in views of 'politically sensitive' Wikipedia pages (expanded set of pages).
Fig. A14. Changes in views of Alexei Navalny related Wikipedia pages. The Alexei Navalny-related pages in this figure are listed in alphabetical order.
Table A5. List of opposition-related pages in Russian that were checked for significant increases during lockdown.

Note: Russia did not block specific Wikipedia pages prior to Wikipedia’s introduction of https. Because of this, we do not have a government-provided list of politically sensitive or objectionable content. As an alternative, we mine a manual list of government opposition-related pages, and then check whether for those increases were narrow and perhaps random (i.e. only occurred for those specific pages) or represented broad increases similar to those seen for historical and previously blocked pages in China. This table lists those Wikipedia pages (translated) that were checked for significant associations during the Russian lockdown period when compared to December 2019. Pages with statistically significant increases (p < 0.05) after a Bonferroni multiple testing were used as seeds when expanding with Wikipedia2vec. These “biggest increase seeds” are in bold above.

---

<table>
<thead>
<tr>
<th>Year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-2013 Russian protests</td>
<td>He Is Not Dimon to You</td>
</tr>
<tr>
<td>2014 anti-war protests in Russia</td>
<td>Human rights in Russia</td>
</tr>
<tr>
<td>2017-2018 Russian protests</td>
<td>List of journalists killed in Russia</td>
</tr>
<tr>
<td>2018 Russian pension protests</td>
<td>Media freedom in Russia</td>
</tr>
<tr>
<td>2019 Moscow protests</td>
<td>Mikhail Khodorkovsky</td>
</tr>
<tr>
<td>Alexander Litvinenko</td>
<td>Open Russia</td>
</tr>
<tr>
<td>Anna Politkovskaya</td>
<td>Opposition to Vladimir Putin in Russia</td>
</tr>
<tr>
<td>Anti-Corruption Foundation</td>
<td>Party of crooks and thieves</td>
</tr>
<tr>
<td>Assassination of Anna Politkovskaya</td>
<td>Pussy Riot</td>
</tr>
<tr>
<td>Assassination of Boris Nemtsov</td>
<td>Russia of the Future</td>
</tr>
<tr>
<td>Boris Berezovsky (businessman)</td>
<td>Russian Opposition Coordination Council</td>
</tr>
<tr>
<td>Boris Nemtsov</td>
<td>Sergei Magnitsky</td>
</tr>
<tr>
<td>Corruption in Russia</td>
<td>Sergei Yushenkov</td>
</tr>
</tbody>
</table>

---

Fig. A15. Changes in views of current leader Wikipedia pages (expanded set of pages).