ORIGINAL ARTICLE

Specificity, Conflict, and Focal Point: A Systematic Investigation into Social Media Censorship in China

Yun Tai[ID][1] & King-wa Fu[ID][2]

[1]Department of Mass Communication, Tamkang University, Taiwan
[2]Journalism and Media Studies Centre, The University of Hong Kong, Hong Kong

Internet censorship mechanisms in China are highly dynamic and yet to be fully accounted for by existing theories. This study interrogates postpublication censorship on Chinese social media by examining the differences between 2,280 pairs of censored WeChat articles and matched remaining articles. With the effects of account attributes and article topics excluded, we find that article specificity raises the odds of being censored. Also, an examination on a collection of international trade articles indicates that such articles with textual units disclosing conflicts, even pro-regime messages, are also removed by the censors. This mixed-method study introduces focal point as a theoretical angle to understand China's contextually contingent content regulation system and offers evidence based on large-scale, nonproprietary, and original social media data to investigate the evolving censorship mechanisms in China.

Public access to the Internet has been deemed by some scholars as a force facilitating the development of the public sphere, civil society, and democracy in nondemocratic countries such as China (Lei, 2017; Qian & Bandurski, 2011; Yang, 2009). However, this rather optimistic view stands in contrast to numerous studies showing that Internet freedom can be suppressed by state interventions through effectively employing and adapting measures of regulating online public speech. Indeed, scholars have suggested that Internet technologies largely contribute to resilience of autocratic regimes as states take on proactive tactics to direct or distract public discourse as well as demobilize online activism (Gunitsky, 2015; MacKinnon, 2008; Sullivan, 2014; Yang, 2018).

Corresponding author: King-wa Fu; e-mail: kwfu@hku.hk

 1

Residing in the heart of the discussion regarding civil society in China is its model of Internet censorship. It is widely known that the use of the Internet for public expression in China is vigorously governed. A sophisticated information control system, consisting of Internet infrastructure (e.g., the Great Firewall), central and local government officials, and technology firms, is constantly evolving (Freedom House, 2018b). Studies focusing on Internet censorship mechanisms in China have examined the targets of censorship and the ways in which Chinese netizens circumvent governmental control, as well as the regime's active actions to manipulate public opinion (Fu, Chan, & Chau, 2013; King, Pan, & Roberts, 2013, 2014, 2017). Though most studies agree that the main purpose of censorship is maintaining the legitimacy and stability of the regime, they diverge over exactly how the censorship system is operated to achieve this goal, that is, what are the major targets of censorship? The divergence between studies may have to do with the dynamics of the censorship mechanism, as well as the development of information and communication technology (ICT)—since it is relevant to how online public sphere is enabled and regulated.

On the one hand, scholars suggest that the Chinese government has become more adaptive and dynamic in restricting media content by deploying approaches of strategic censorship (Lorentzen, 2014), positing that content regulation is not simply a set of fixed rules but might be sometimes relaxed and then tightened in response to social tensions and public opinion (Chen & Xu, 2017). For example, online nationalism posts were not uniformly allowed or censored throughout the course of the 2012 Diaoyu/Senkaku Crisis (Cairns & Carlson, 2016), and even the views of pro-regime nationalists are sometimes censored by the authorities (Zhang, Liu, & Wen, 2018). Collective action posts are known to be the most likely censorable topic (King et al., 2013); however, study also reveals that a large amount of collective action posts usually did not go viral and were not censored at all. It is believed that the level of governmental tolerance regarding the press and public expression can wax and wane over time and across (virtual) spaces.

On the other hand, at least two changes in China regarding ICT development and cyberspace regulation are worth noting. First, the diversity of social networking sites in the market has been remarkably reduced. QQSchool (QQ校友录), RenRen (人人网), Sina Space (新浪空间), 51.com, and Kaixin001 (开心网) were once the most popular sites around 2009 (CNNIC, 2019), whereas the tech trinity—Baidu, Alibaba, and Tencent, collectively known as BAT—started dominating the digital market in China in just a few years. It was reported in 2015 that 80% of market capitalization came from the Big Three with a combined market value of more than $500 billion. Also, the dominance of a few players made it difficult for start-up challengers or international players to succeed (Sender, 2015)—with very few exceptions such as the popular social app TikTok (Douyin). Among the three, WeChat's owner, Tencent, which specializes in social media, is seen by some as the most "fearsome" one (The Economist, 2017).

The second change is, as noted by Lv and Luo (2018), Internet governance at the central level in China is institutionally fragmented, with the State Council and its Ministry of Industry and Information Technology (MIIT) in charge of Internet development whereas regulation of the Internet is overseen by the Central Cyberspace Affairs Commission (CCAC)[1] headed by President Xi Jinping, which may result in disagreements between the two organs as their primary goals (i.e., economic growth versus stability maintenance) are somewhat contradictory. However, the fragmentation may not be as prominent since the CCAC concentrated its power regarding Internet governance from other departments while holding on to censorship as its primary task (Lv & Luo, 2018). Both of the changes signal centralization in China, one in political power reflected by e-governance and another in the ICT business, if not in other commercial sectors.

Against this dynamic backdrop of evolving digital governance in China, as well as the inconsistent findings among previous studies, we revisit the role of information regulation in shaping the public sphere at a time when the development and regulation of ICT in China are both centralized and when Internet control is believed to have been further tightened since a new Cybersecurity Law commenced in 2017, online anonymity was ended (Denyer, 2017), and a setback of online activism was observed (Yang, 2018). In particular, we investigate what contributes in defining the boundaries of the online public sphere through the conducts of postpublication censorship (i.e., published information removal). The questions guide this study are as follows: What are the nuanced differences (e.g., textual units such as terms, phrases, and paragraphs) between the censored and remaining public contents on social media in China? How could we conceptually understand the role of such differences in the Chinese censorship mechanism?

Very few China censorship studies have been able to draw insights from large-scale and longitudinal empirical evidence due to the unavailability of systematic data collection. Among those limited studies conducted, most of them focus on short online textual content (e.g., the Twitter-like Chinese social media platform Sina Weibo), some are conducted by nontransparent sampling methodology, and a few draw conclusions from their findings on the basis of content contributed by rather limited numbers and types of users (e.g., online forums and emails between local government officials; Fu et al., 2013; King et al., 2013, 2014, 2017). We, therefore, develop a transparent, systematic, and longitudinal data gathering approach, *WeChatscope*, to automatically monitor online censorship, and obtain primary data concerning censored contents over a period of time.

In this article, we take an inductive approach with a mixed-method design to explore the likely mechanisms in online censorship on the basis of comparisons between censored and remaining social media data. We first introduce the WeChatscope system developed for monitoring and detecting Internet censorship in China, a system that can be adopted (with some modifications) for automation of data collection when data access via platform-provided application programming interface (API) is not available, for example, closed messaging system. We then

present our major finding showing that, aggregately, the censored WeChat articles differ from the remaining ones in having higher content specificity. An example based on a subcollection of articles is also presented to demonstrate how other textual units (e.g., phrases, sentences, and paragraphs signaling conflicts) are also possible factors of censorship decisions. As existing literature mostly understands Internet censorship decision based on the content's political implication, this study then suggests "killing the focal point" as an explanation for the empirical observations that certain textual characteristics of social media content are linked to a higher likelihood of being censored.

WeChat and WeChatscope

As of December 2019, WeChat is reported to have more than 1.1 billion monthly active users (Tencent, 2020) and is now the leading social media platform in China. As a mega all-in-one application, WeChat integrates essential online functions, private messaging, group chat, online gaming, e-payment, and blogging; it also offers "public accounts" (or "official accounts") to which users can subscribe. Functioning in a similar way to Facebook Pages, WeChat public accounts are publicly accessible and are utilized by companies, organizations, media, celebrities, and individuals to broadcast media content and engage with their target users; users receive push notices when the public accounts they follow post new messages, articles, pictures, or videos. In 2017, the number of monthly active WeChat public accounts reached 3.5 million with 797,000,000 active followers (Tencent, 2017). Given the importance of WeChat public accounts in the Chinese media system, systematic data collection is much needed for current and future scholarship—however, even public content on WeChat is not conveniently accessible via a web browser and without using the application,[2] which adds another layer of difficulty to gather WeChat public account data on a large scale, not to mention that WeChat provides no developer API to the public. We thus develop the WeChatscope system, which deploys an "app crawling" technique to collect the data of WeChat public accounts utilizing WeChat's desktop application.[3]

The WeChatscope system works in three major parts (Figure 1). First, we create a number of individual WeChat accounts, each of which subscribes to a number of public accounts in our sample. We then use the WeChat desktop application to monitor the sampled public accounts each day from 7:00 a.m. to 12:30 a.m. of the next day to obtain the unique uniform resource locators (URLs) of the new articles they published. Second, the obtained URLs of the published articles are stored in a buffer, where a set of background-running web crawlers then get the URLs one-by-one and visit the articles in headless browsers and store the metadata (article title, publication date, public account name) and media contents (texts and images contained in the articles) in a relational database. Third, the system continues to monitor these published articles for another 48 hours by revisiting them each hour between 7:00 a.m. and 2:00 a.m. of the next day to check if the articles have been
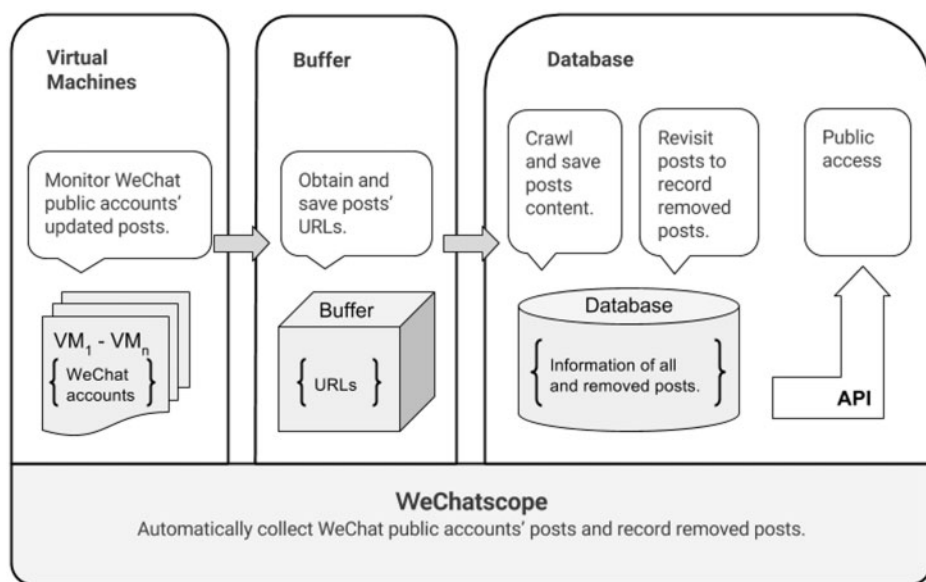
Figure 1 The system diagram of WeChatscope.

removed—if removed, record the returned removal system message displayed on the web page. Finally, WeChatscope allows public access to the open data through our self-developed Application Programming Interfaces (APIs),[4] by which the public can obtain information about removed articles such as title, public account name, publication date, removal date, removal message, as well as full text of the re-moved articles captured by the system.

Data and method

In this present study, we use a sample collected by WeChatscope, which obtains 818,393 WeChat public articles published between 1 March and 31 October 2018[5] by 2,560 public accounts.[6] In this sample, 8,736 articles were removed after publica-tion for six reasons[7]: removed by content owner (5,945), content violation (2,345), account blocked (272), copyright issue (148), account migrated (25), and reported by multiple users (1). We define the articles removed due to "content violation" as "censored articles" as the removal was forced by WeChat platform administrator, whose decisions are directed by the governmental censorship authorities, with no clear reason offered to the public (e.g., which regulation was violated).

We deploy a case–control matching strategy to pair each censored article with one remaining article from the same public account that is topically most similar to the censored one. This strategy is deployed to adjust for the possible confounding influence of the attributes of a public account (e.g., account type and popularity) and the variations of topics published by a public account, as we seek to examine

nuanced factors, that is, textual terms, to account for the differences between censored and remaining articles. We do this case–control matching by constructing a topic model from a corpus containing the censored articles and the remaining articles from the same accounts published within 1 week before and after the target censored article was published. Topic models, such as Latent Dirichlet Allocation (LDA), are useful tools for exploring large collections of documents by detecting latent themes (i.e., clusters of words) without additional knowledge other than texts in documents. We adopt the correlated topic model (CTM) because it offers a more realistic model that relaxes the assumption of independence between topics[8] and also fits the data better (Blei & Lafferty, 2007). Similar to LDA and many other topic models, CTM treats each document as a combination of topics and each topic is a mixture of words.

By performing residual analysis (Taddy, 2012) as well as considering semantic coherence and exclusivity, a model with 160 topics[9] is found to be the most parsimoniously fitted model, and a document-topic matrix with topic distributions of each article in the corpus is obtained; in other words, each article is vectorized into 160 dimensions. With this document-topic matrix, we select the most similar remaining "candidates" for each of the censored article based on cosine similarity score. Finally, we arrive at a sample of 4,560 articles[10] with 2,280 pairs of censored and remaining articles from 751 public accounts.[11] Our final sample thus contains the censored articles we observed and their counterparts selected, so that each member of the pair is similar to each other in two senses: they are published by the same WeChat public account and they are of similar composition concerning latent topics; however, only one of them was censored.

Since one important goal of the present study is to find unobserved nuances that vary the decision regarding online censorship, we first adopt the random forest (RF) method to mimic the classification of censoring an article or not. Specifically, we do so to discover textual terms that matter more than others in predicting if an article is censored or not. RF is thus selected for its advantage of handling large numbers of input variables with each of them offering a small amount of information (Breiman, 2001); in our case, we have lots of words in the collected articles for which we would like to learn about their fraction of influence on censorship. RF is an ensemble method built on the Classification and Regression Tree (CART) methodology (Breiman, Friedman, Stone, & Olshen, 1984). CART produces decision trees that make predictions (i.e., predict classes for classification trees) based on binary splits: at each split, one condition (e.g., a value of a variable) is considered to divide the cases in the data into two groups that are internally more homogeneous (or purer) than the current group. RF, by definition, is "a classifier consisting of a collection of tree-structured classifiers where . . . each tree casts a unit vote for most popular class . . .." (Breiman, 2001, p. 6). It extends CART by bagging[12] and randomly selecting a subset of input variables at each split; for classification, it predicts classes using "out-of-bag" (OOB)[13] estimation based on the most common decision in the forest (i.e., the majority vote). RF is preferred in this study and many others for at least

two advantages: it is nonparametric with no distributional assumptions and it does not overfit because of the law of large numbers.

Although RF is capable of handling a large number of predictors (often hundreds or thousands), the numbers of potential predictors (i.e., words) in our sampled WeChat articles are a lot more than thousands—this is a common challenge faced in conducting text analysis with a large number of documents, more so when the documents are lengthy articles like what we have in our sample. We thus pre-process the data, first by segmenting Chinese words in the text and removing punctuation, numbers, stopwords, and words contain only one character. Second, we tokenize each document in the corpus with bigram tokenizer and create a document-term matrix with term frequency-inverse document frequency (tf-idf) weighting. Third, we calculate chi-squared scores for each term to obtain a ranking reflecting the discriminatory ability for the two distinct sets of articles (i.e., censored and remaining articles). We then select those terms with higher chi-squared values as input features in predicting the status of the articles: censored and remaining.[14]

We also supplement the computational textual analysis described above by conducting a close reading (close textual analysis; Allen, 2019) on a subset of articles as an example to demonstrate how other textual units jointly contribute in determining the decision of censorship that may not be captured by simply examining segmented words. Close reading is a method which "investigates the relationship between the internal workings of discourse in order to discover what makes a particular text function persuasively" (Allen, 2019, p. 137). We thus apply this method on the subset of articles, as a collection of text, to look for what the text bears persuades the censors that it has to be removed from the Internet. In particular, the major goal of this analysis is finding the repetitions in the text—the common elements and themes that the censored articles in this subset share.[15]

Results

Topical variety of challenging articles

The final sample of 4,560 WeChat public articles that are either censored or semantically similar to the censored ones represents "challenging" articles carrying the potential of being undesired content on social media in China—they can be about a variety of topics (see Table Si) as identified by the CTM based on these articles. The topic labels are assigned based on associated words and articles of each topic.[16] One major theme among the identified topics is the economy, with the most common topic being international trade with a focus on trade among China, the United States, Japan, and North Korea, which constitutes 6.8% of all articles in the sample. In fact, among the top 20 topics, almost half are relevant to the economy with a 27.5% combined proportion: trade, technology companies, economics and finance, the housing market, "guo jin min tui" (国进民退),[17] Cui Yongyuan (崔永元),[18] multilevel marketing, and the social security tax, as well as poverty.

We first tested if some topics play more important role than others in determining if an article is censored. To do so, we construct a RF model with the proportion of topics for each article as the input variables to predict if an article is censored or not. This initial RF model[19] performs only slightly better than a simple coin toss as it can only correctly predict 52.68% of article status in terms of censored or not censored. This result is not surprising as we select the remaining articles to pair with the censored ones based on the distribution of topics across all the articles to pick the most similar remaining article for each censored one. This low successful prediction rate also confirms in a way that our selection strategy works in reducing the topical differences between censored and remaining articles as the identified topics do not serve as good predictors in differentiating the status of the articles. Although the topical difference between censored and remaining articles within each individual pair is purposely reduced as much as possible, aggregately, topical distribution differences still contribute in predicting article status as the overall error rate of prediction (47.32%) is slightly lower than random chance. Among the sixty topics, topic 32 (fiction and stories I) and topic 49 (food and health) are the two predictors that matter more than all the others in providing predictive accuracy.[20] Both of them are more like "housekeeping" issues when it gets to the censoring decision. Many of the censored ones that are highly composed of one of the two topics were either with heavy erotic content or content that may be deemed as "rumors"[21] (谣言), for example, health suggestions without scientific evidence.

Specificity and censorship

We tested the second model with higher granularity: using textual terms as predictors for classifying censored versus remaining WeChat articles. As mentioned above, we ranked the terms contained in the articles based on their chi-squared values and selected those with higher values as input variables for this RF model.[22] This RF model is constructed with 2,029 terms to predict the status of the 4,560 articles[23]—its OOB estimate of error rate is reduced to 33.25%, compared with the 47.32% error rate of the initial model.

We further measure the importance and perilousness of the terms by, respectively, calculating an RF importance score[24] for each term that indicates how important a term is in classifying if the articles are censored (the higher the more important), and its censored-to-remaining ratio (CR ratio)[25] is shown as Equation 1:

$$CR \ ratio \ = \ \frac{(T_c \ + \ .5)}{(T_r \ + \ .5)} \tag{1}$$

T_c = Total occurrence of a term in censored articles;
T_r = Total occurrence of a term in remaining articles.

For instance, the term "government" appeared 2,377 times in the group of censored articles and 1,747 times in the group of remaining articles, then its CR ratio is 1.36—that is, the number of the term "government" appeared in censored articles is about 1.36 times the number in remaining articles. Terms with its CR ratio > 1 are
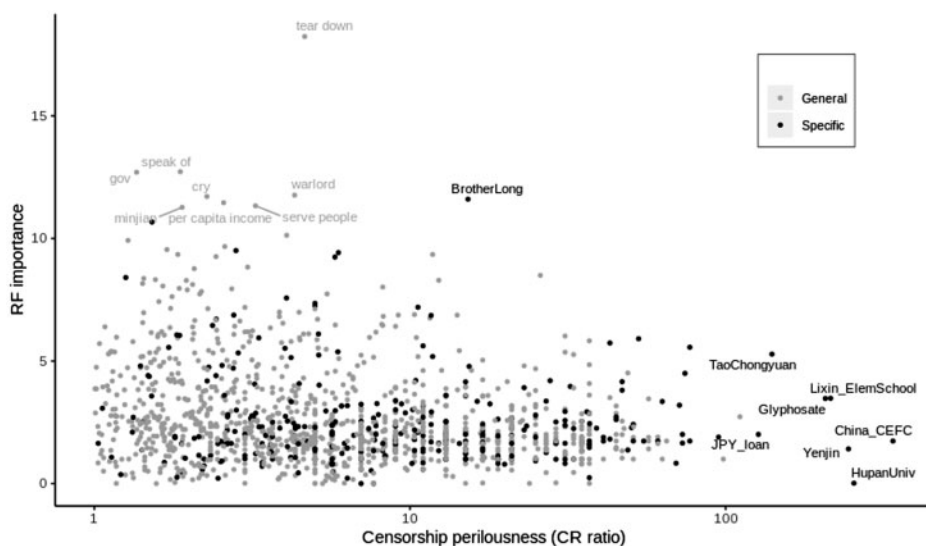
Figure 2 Important perilous terms. Censorship perilousness is plotted with log.

terms used more often in the censored articles than in the remaining articles; in other words, terms with their CR ratio higher than 1 (i.e., perilous terms) can be seen as terms with higher possibility of triggering the censorship system.

Figure 2 plots the distribution of the 1,528 perilous terms that are important in recognizing censored articles (i.e., terms with their RF importance score larger than 0 and their CR ratio higher than 1), with the two colors representing if it is a specific term, that is, whether it is a unique identifier of an object.[26] Although more than 80% of the important perilous terms are general terms and those with the highest importance are also mostly general terms such as government (政府), speak of (谈到), and tear down (拆除), the terms with the highest perilousness are, instead, mostly specific terms such as CEFC China (中国华信), Hupan University (湖畔大学), and TaoChongyuan (陶崇园).

Figure 3 shows the various distribution[27] of the key specific terms (i.e., terms that are important, perilous and specific) for censored and remaining articles—though both of the groups are mostly concentrated at 0 (i.e., no key-specific terms), the number of remaining articles (1,280) is higher than censored articles (775) at 0, whereas the numbers are generally higher for censored articles once the number of key specific terms goes over 3. We further examine the relationship between article specificity and the chance of being censored by evaluating the correlations between the number of key specific terms (n) in each article and the odds ratio of being censored (OR_n) shown in Equation 2.
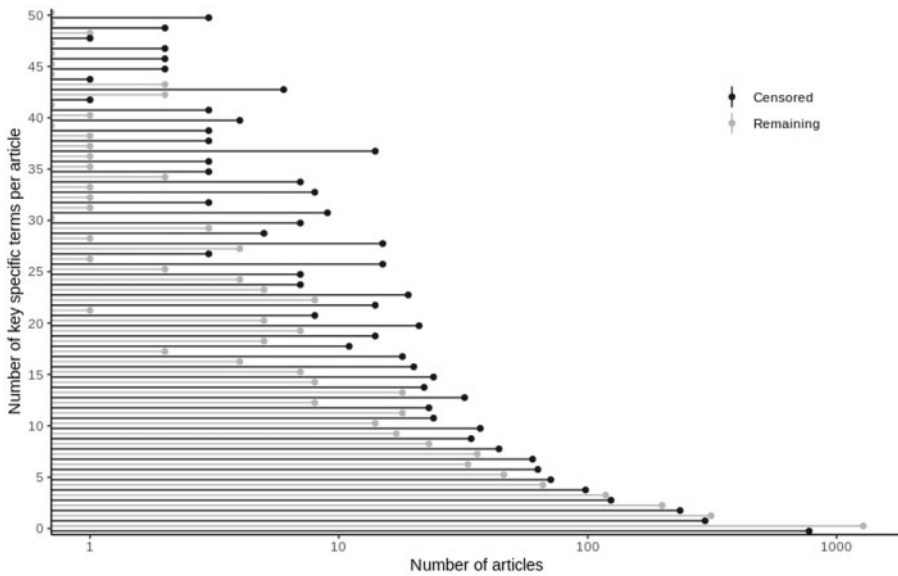
Figure 3 Distribution of key specific terms in censored and remaining articles. $N = 4,497$. As 98.6% of articles contain 50 or less key specific terms, in this figure, we exclude those contain more than 50 key specific terms; number of articles is plotted with log; the data points with 0 article are represented without segments: for example, there is no remaining article contains 50 key specific terms.

$$OR_n = \frac{\frac{(N_{cn}+.5)}{N_n}}{\frac{(N_{rn}+.5)}{N_n}} = \frac{(N_{cn}+ .5)}{(N_{rn}+ .5)} \tag{2}$$

N_{cn} = Number of censored articles contain n key specific terms;
N_{rn} = Number of remaining articles contain n key specific terms;
N_n = Total number of articles contain n key specific terms.

OR_n is computed as the ratio of the number of censored articles to remaining articles for each key-specific-term frequency n, defined as the count of key specific terms that appeared in each article. For instance, there are 775 censored articles containing zero key specific term and 1,280 remaining articles containing 0 key specific term, then $OR_0 = (775 + 0.5)/(1,280 + 0.5) = 0.606$; that is, at the frequency of zero key specific term, the number of censored articles is about 0.6 times the number of remaining articles. To see if the odds ratio of being censored increases when the frequency of key specific term increases, Equation 3 shows the odds ratios given cumulative frequency of key specific terms ($ORCUM_n$):
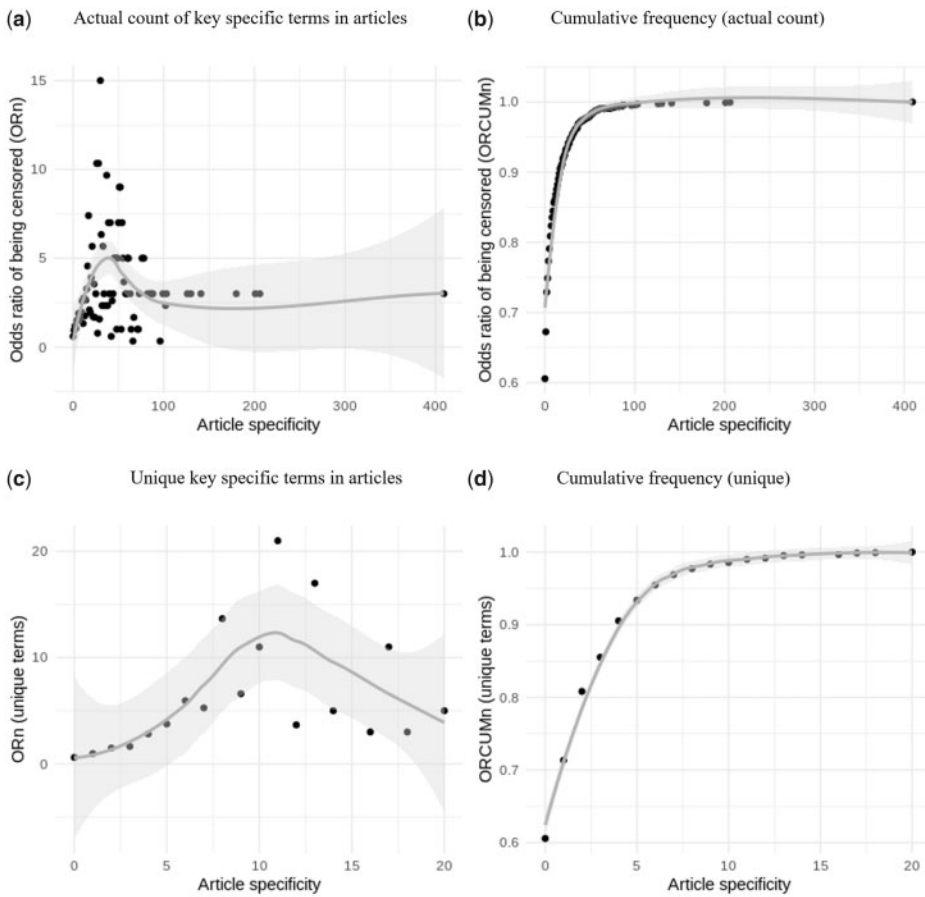
Figure 4 Article specificity and the odds ratio of being censored: (**a**) actual count of key specific terms in articles, (**b**) cumulative frequency (actual count), (**c**) unique key specific terms in articles, and (**d**) cumulative frequency (unique).

$$ORCUM_n = \frac{\frac{(CUM_{cn}+.5)}{CUM_n}}{\frac{(CUM_{rn}+.5)}{CUM_n}} = \frac{(CUM_{cn}+ .5)}{(CUM_{rn}+ .5)} \tag{3}$$

CUM_{cn} = Number of censored articles contain less than or equal to n key specific terms;

CUM_{rn} = Number of remaining articles contain less than or equal to n key specific terms;

CUM_n = Total Number of articles contain less than or equal to n key specific terms.

Figure 4a shows that the odds ratio of being censored is low (0.606) when the articles contain no key specific terms, whereas it peaks at 15 when the articles contain 30 of such terms. Figure 4b demonstrates that as the number of key specific terms contained in articles increases, the odds ratio of being censored increases as

Table 1 The Impact of Key Specific Terms on Censoring

Variable	(1)	(2)
	Odds ratio of being censored	
Number of key specific terms per article	1.064***	
Number of unique key specific terms per article		1.445***

***$p < .01$.
$N = 4,560$ articles.

well. Since repeated specific terms in one single article may not contribute further, for example, finding a sensitive term that appears in an article twice or 10 times may not locate much of a difference in a censor's decision, we further test the relationship between specificity on the basis of unique terms and the odds ratio of being censored. In Figure 4c and d, the specificity is calculated with the number of unique key specific terms in each article instead of total occurrence (i.e., such terms are only counted once no matter how many more times they appear in an article). Figure 4c shows that the odds ratio peaks at 21 when the articles contain 11 unique key specific terms, and Figure 4d demonstrates again a growing trend of the censoring odds ratio as article specificity increases. Finally, we examine the effect of the number of key specific terms in an article on censoring with logistic regression models. Table 1 shows that on average, the odds of being censored are 1.06 (95% CI: 1.054, 1.074) higher when adding one key specific term in a WeChat article. Moreover, the odds are 1.45 (95% CI: 1.385, 1.508) higher if one unique key specific term is added.

Conflict and censorship: An example of international trade articles

One of the possible explanations for the approximately 33% misclassification rate for Model 2 is the diversity of this collection of articles as a good number of topics are identified by the CTM mentioned above and a very large quantity of distinct terms are used in the articles (i.e., 2,783,450 unique terms excluding those with only one character). The third model thus focuses on the articles that are mainly about the most common topic, that is, trade,[28] as well as the terms[29] contained in such articles. This final model with 1,236 terms for classifying a total of 270 trade articles —132 censored articles and 138 remaining ones—results in a substantially reduced error rate of 21.11%,[30] that is, almost 80% of the articles' status are correctly classified (Table Sii). Among the 1,236 features, 730 are estimated to have an importance score larger than 0. For articles that are mainly about international business trading, 484 important terms (i.e., words with RF variable importance larger than 0) are perilous terms (i.e., terms with CR ratio larger than 1) and another 239 terms are not perilous (CR ratio smaller than 1).[31] We find that a number of important perilous terms are named entities directly relevant to business (e.g., names of companies and

products), whereas this is not the case for the group of important nonperilous words. It appears the content of censored articles on trade as a whole is more specific than that of those remaining.

We then examine if and how specific terms and other textual units are used in censored articles that are mainly about trade. From the 270 trade articles, a total of 102 articles (96 censored and 6 remaining articles) were selected for a close textual analysis. The selected articles contain the terms that matter the most in differentiating whether an article was censored and characterizing the censored articles (i.e., terms with high RF importance and high CR ratio[32]). The analysis indicates that WeChat public article authors not only address international affairs but also domestic ones when writing about international trade. The representative censored articles are summarized in Table Siii. As expected, a large number of trade articles are entirely or at least partially concerned with the trade dispute between China and the United States; the relations between China and Japan, North Korea, and a few countries in Africa as well as Europe are also covered. The major themes of the international affairs covered in such articles include the impact of foreign relations (in particular, China–U.S. relations), the China–U.S., trade imbalance, and the ambition and aggression of China in the world.

The impact of relations between China and the United States can be illustrated with at least four elements indicating: (a) for China, the relationship with the United States plays a pivotal role in the past and the future of its growth and development, (b) the trade war between the two countries is further escalated, (c) the trade war is not coming to an end any time soon, and (d) the war is beyond trade and may be the beginning of a "new cold war" or even leading to a hot war. The following example excerpts demonstrate how the impact of the China–U.S. relations or trade dispute is revealed in such articles as they mention:

The importance of such relationships (e.g., assistance from the United States and settlement between the two countries) in history and the future economic development of China,

> The US offered China military assistance with a higher level than what it offered to the allies, which ... laid the foundation of the great strategy of China's "reform and opening-up" as well as the rapid growth and development of China in the past 40 years.

> In a month, we will face two (different) outcomes: one is that China and the US settle and work together on China's market reform, the other is the breakdown of negotiations, which will be like watching a disaster film. Perhaps this coming month is the crucial time of China's fate.

> The trade war escalation,

> Trade war escalated! 200 billion isn't enough? Tariffs on an additional US $276 billion in goods...

China might have misjudged the intention of the Trump government twice. . . the repeated misjudgments escalated the China-US trade fight again and again.

Its long-lasting nature (e.g., predicting a long-term trade war), and

Its broader influence (e.g., other disputes between the two countries beyond trade):

Like Jack Ma just said, the trade war is going to last for 20 years . . . the two countries will fall into a comprehensive fight against each other once the trade war goes over 3 years, and the rest of the 17 years won't be as easy as the trade war period - financial war, economic blockade, and cold war may all happen one after another.

The trade war started by the US poses a threat to China that goes beyond trade itself. Other fields such as bilateral investment, intellectual property, and strategic industry, especially high-tech industry, are also involved. It also extends to the South China Sea disputes and the Taiwan issue . . .

. . . though it (the trade war) is represented as an economic war, the possibility of military war can't be ruled out if it continues with escalation.

Besides the importance and impact of the U.S.–China trade disputes, a few censored articles either attribute such a fight to the trade imbalance due to the deficit of the United States with China, or suggest such a business relationship is not fair for the United States by stating, for example:

. . . the China-US trade relation is extremely imbalanced; the investment of the US in China is highly regulated . . . China and the US urgently need to eliminate the trade dispute through opening the market in China to the US as well as creating a market that is fair and without discrimination for American investors.

The third theme indicates that China is an aggressive player in the world and may be seen as a common enemy of many other countries by some. The articles contain elements of this feature mentioning: (a) terms and sentences representing that China is "great" and doing better than others, in particular, the United States (e.g., "China has won," "Made in China 2025" plan that calls for becoming a "manufacturing superpower," China's strength and confidence in fighting with others), (b) China is ungrateful to countries such as the United States and Japan that have been contributing to China's economic progress—article titles such as "How generous has the US been in the past 100 years?" and "Japan to terminate Official Development Assistance to China: should we say 'thank you'?" call into question China's attitude regarding such foreign aids, (c) the formation of unions of countries that are either against China, for instance,

. . . the real goal of Trump is forcing them (the EU & Japan) to join the containment of China.

> North Korea-US Summit: the two countries "reconciled," what does China do next?

> The US, EU and Japan have started to form a new common market... we can't stop the global economy trend and may become an outsider again.

or even line up with China,

> Members of the Shanghai Cooperation Organization can be seen as China's best friends. India is a new member... just hope it won't align with the US and Japan.

> ... Italy has stepped on the road of cooperating with China. This may bring economic opportunities but may also put this country under China's influence. Italy's acceptance of China may also damage the agreement between European countries regarding their China policy.

and (d)) China's engagement with African nations may be seen by some as economic colonialism, for instance:

> As the world worries that China is going to economically colonize Africa, it is however questioned in China: is China a "international doofus"? Facing a severe economic downturn, small businesses in need of help, as well as the budget of medical services, social insurance and education falling short for a long time—why share China's $600 billion foreign exchange reserves with friends but not with family?

The above excerpt also reveals that other than international relations, some of the censored representative articles extend the agenda to domestic affairs while the major topic remains as international trade. An author directly points out in a censored article on the trade war that what really matters for China is its internal problem:

> Every problem is an internal problem. The history of China shows that problems are mostly from inside, and oftentimes external factors are only the final push. Take the current economy of China as an example: (the government) prints more money to maintain economic growth, but excess printing of currency brings a series of problems - excess capacity, huge debt, high leverage, an asset bubble ...

The domestic affairs addressed are mainly featured with economic elements such as the structure and policies of domestic economy—other economic elements include unemployment due to capital flight, possible short supply of manufacturing materials from the United States (e.g., cotton and high-performance chips), and bad debts of banks. Some domestic affairs addressed are extended from economic to political issues or even national defense such as governmental corruption, distrust in government officials, political regime (democratic vs. nondemocratic) and armed forces. For example, an article titled "The US and North Korea reach agreement, China faces internal and external troubles" mentions:

> The internal problem (of China) is reflected by the aggravation in corruption…and the lowering of public trust in governmental officials… Corrupted officials of the army like Xu Caihou have damaged the public trust in their combat power… The most urgent problem for China is its 'domestic contradiction.'

> In addition, a few social elements regarding agenda setting and freedom of speech are also part of these censored articles. For instance, the author of a censored article titled "Is the US really attempting to curb China's rise?" expresses an alternative view to an official newspaper of the CCP:

> On August 9th, People's Daily, the major official media of China, published an editorial commentary asserting that the US is definitely going to thwarts the country…because it threatens the status of the US. Since China is now the world's second largest economy, the US will curb China's rise no matter what China does… I think this argument is not supported by a firm theory… how China rises determines whether the US is to curb China's rise.

The author continues with a brief summary of the "theory" (Thucydides's Trap) based on which the argument made by the People's Daily is formed. He then states that the competition between the two countries may be "fair and friendly" if they are "both democratic countries" with "open" speech and "the government is 'freely' elected by the people." The author also makes it clear that the leaders of China warned that the two countries should not fall into Thucydides's Trap and questions why the People's Daily still made this argument. This example shows a viewpoint different from a mainstream media outlet—even not questioning the government itself—may not survive the censorship system, nor does implying the lack of expression freedom in China, not to say explicitly stating that China hinders the information flow, as well as monitors and controls the life of its citizens, as shown in other censored articles on foreign relations and trade.

By closely examining the representative censored articles, we first find that specific terms (e.g., names of individuals, organizations, countries, and policies, as well as monetary values) are indeed frequently used in such articles (see terms in bold in the above excerpts, for example). Second, we find that although various themes and elements are identified in the representative censored articles, these textual units have one thing in common: they either directly point out the possibility of external or internal conflict that China is or will be encountering, or reveal, imply, or suggest an intensification of such conflict. The China–U.S. trade dispute, as one of the most discussed issues in 2018, is itself a conflict between the two countries; when it is described as intensified in terms of, for example, time (i.e., a long-term conflict), field (i.e., conflicts beyond trade) and magnification (e.g., importance, imbalance, and escalation), the articles may not be allowed to remain on the social media platform. Indeed, not all the WeChat articles regarding the trade dispute were censored: for instance, an article indicating the improvement of the relation between China and Japan may create positive conditions for the easing of tensions between China and

the United States was not censored. In this case, China's external conflict is mentioned but in lighter terms with a possibility of being mitigated. External conflicts concerning China are also revealed when indicating or implying that China is not necessarily a "friend" of other countries, such that China is ungrateful for international aid, is a target of containment actions, or is taking advantage of others (e.g., African countries).

The impression of intensified conflict can also be conveyed to readers, of course, in criticisms that either indicates domestic economic problems or social and political features revealing weak institutions of conflict management (Rodrik, 1999) such as high corruption and low democracy represented in the lack of political rights (e.g., "free" election) and civil liberties (e.g., freedom of speech) (Freedom House, 2018a). However, diverging from a popular belief that "positive energy" such as supporting or praising the government or its policies (Yang, 2018) is welcomed by the government, we find that such pro-regime articles were also censored when they signal conflicts. Several articles illustrating China's strength and confidence by addressing the influence of policies such as "Made in China 2015" and the "One Belt, One Road" initiative or describing China as a strong "fighter" in trade disputes with the United States were censored. Although these articles do deliver "positive energy," they in the meantime disclose external conflicts confronting China.

Focal point and online censorship: Making sense of the findings

Why, then, the articles with higher specificity and/or signaling conflicts are the targets of censors? Thomas Schelling's concept of focal point could be a way of understanding their role in the censorship mechanism. and other scholars' (Mehta, Starmer, & Sugden, 1994) experiments involving pure coordination games show that people use focal points with salience in coordinating their behavior when prior communication is absent. For instance, the number "1" tends to be the most common choice when people are asked to pick a positive number that is most clearly unique; or, in another example, the information booth at the Grand Central Terminal tends to be selected as the spot when a group of students was asked to choose where to meet with a stranger in New York City . The commonly picked number "1" and the Grand Central as a popularly selected meeting point represent a "focal point for each person's expectation of what the other expects him to expect to be expected to do" (p. 57). In a game, some labels (e.g., words, pictures, other symbols) of strategies are more salient or prominent than others, often because they are associated with common knowledge, experience, or culture of the players (Mehta et al., 1994). Strategies with salient labels are more likely to be chosen by players and a focal point is likely to be formed.

In a study on legislative politics (Ringe, 2005), focal points are defined as "ideas and images that communicate information about the consequences of a given legislative proposal" (Ringe, 2005, p. 733) with regard to legislators' ideology. The provision of focal points is considered as a mediation through which ideology can be

linked to specific policy proposals, and thus influences policy preferences and outcomes. For instance, in the case of the European Union (EU) cross-border takeover bill debate, the deployment of three different dominant focal points (i.e., "single market," "workers' rights," and the notion of creating a "level playing-field" across the EU) at the three reading stages contributed in shifting the voting outcomes of this bill. The focal points linked to ideological preferences and constituency concerns serve as "decision-making shortcuts" (Ringe, 2005, p. 738) that simplify and make legislative proposals more tangible. The concept of focal point is also employed in understanding the pre-emptive repression of collective actions and the crackdowns on political dissidents. Scholars (Carter & Carter, 2019; Truex, 2019) suggest that high profile events (e.g., prodemocracy movement anniversaries, high-level regime meetings, and international sporting events) serve as natural focal events of coordination for activists, as this "dissident calendar" is known in advance, with high salience, and able to reduce coordinating problems for citizens when attempting to form collective actions; these focal events are thus also targets of state suppression.

Similarly, Internet censorship can also be thought as a multiparty game setting where the censors intend to regulate the contents, whereas some social media users aspire to circumvent the censorship system so that they can communicate with each other to, for instance, facilitate public discussions and form opinions regarding social issues, or even mobilize social actions. In this content regulation mechanism, censors preferentially restrict the sharing of common knowledge by, for instance, removing "focal points" addressed in media contents created by social media users. The role of "focal points" in such multiparty game setting can also be understood with the communication theory of Elaboration Likelihood Model (Petty & Cacioppo, 1986), which posits a "dual process" of persuasive communication: the route of systematic (rational) thinking or/and the route of heuristic thinking via cognitive shortcuts. Since sensitive topics or keywords are not allowed to be directly published in China, the censor and social media users both tend to read media text beyond literal meaning and are then inclined to explore "hidden meaning" in (social) media content or actively read "between the lines." In other words, they rely more on heuristic thinking than on systematic thinking. For some social media users in China, whereas the actual idea they intend to convey cannot be directly presented, as a censorship circumvention, they would purposively write in a manner of referring to "focal points" in order to provide possible cues for readers to form inferences associating with the original idea through cognitive shortcuts. For the censor, such cues contained in texts or other social media content could be seen as heuristics that are more likely to persuade social media readers to think toward undesirable directions, for example, believing in information related to the issues that may lead to social instability.

Drawing from the empirical findings presented above, it appears that textual units signaling conflicts and specific terms providing details and concrete cases would potentially contribute to the construction of perceived crises when they are

widely shared and becoming common knowledge, that is, heuristics. Conflicts are in friction with stability (i.e., such textual units delineate an "image" in which the Chinese society is not as stable as the regime claims) and specific details could provide evidence for the public to verify if, for example, a controversial issue regarding the government is a rumor, a false accusation, or an inconvenient truth. If a controversial issue is considered as an inconvenient truth because of detailed information provided by individuals or organizations other than the government, a trust crisis may be successfully constructed. For instance, an article describing an escalated trade war between the United States and China by providing detailed lists of goods hit with higher tariff effectively informs the public that there is an intensified tension between the two countries, and it also reveals who and what are going to be affected negatively by the trade war. In other words, such messages could contribute in persuading the public that a crisis is or will be happening; the censorship system thus reacts to the potential of a successful persuasion by removing such messages in order to prevent the construction of a crisis, that is, prevent people from believing that they are or will be in trouble due to a crisis.

Therefore, times of crisis can be considered as one type of potential focal points—the salience resides in the (often negative) common memories and experiences of previous crises that are likely to be evoked. Such negative memories and experiences associated with previous crises could be seen as "invisible" labels shared and thus are easier to be recognized by citizens. Among other choices, times of crisis may be more likely to be selected by social media users (e.g., WeChat public account holders or ordinary netizens) to form dissents, demand changes, or even mobilize collective actions by, for instance, addressing their grievances, as they would expect other netizens with such shared negative experiences to do the same. The online messages containing elements that could potentially join the portrayal of a crisis are thus likely to be removed as a precaution against the formation of perceived crisis, as it may signal problems of the party-state governance. In other words, such elements can serve as the heuristics that enable netizens to reassemble the information regarding the crisis (the focal point) and associate it with the original idea (e.g., criticizing the government or mobilizing the people). In short, the censor does not only suppress the original ideas, the focal points, but also the "materials" that could be linked to the focal points.

Conclusion and limitations

With a selection of more than 2,200 pairs of WeChat public articles that are similar in content but different in their censorship status, we find that, first, topics carrying the potential of experiencing postpublication censorship range widely—they are not limited to those that have long been deemed as "problematic" (e.g., pornography and rumors) or "sensitive" ones such as political issues researched extensively in previous studies; instead, more than a quarter of the content of this set of articles is relevant to the economy, a topic that was previously perceived as safe to talk about, even under the censorship system in China. Although the quantity of economic

content on social media may be partially attributable to China's economic slowdown (Harada, 2019) and partially to the China–U.S. trade dispute during the data collection period of the present study, it is worth noting that "sensitive" topics in China as well as its censorship system are both contextually contingent.

Although these identified topics do not differ much in the chance of an article being actually removed by the system, as the findings suggest, one thing that does raise the odds of being censored is article specificity. In general, articles contain a higher number of key specific terms are more likely to be censored than those with fewer of such words, with the account attributes and topics of the articles controlled. Furthermore, in examining the representative foreign trade articles, we find that many censored articles share at least one major characteristic: they contain textual units signaling conflicts or tensions, both internal and external, that China has to deal with. Finally, drawing on the above findings, it appears that social media content with elements (e.g., specific terms and textual units signaling conflicts) contributing in constructing focal points (e.g., perceived crises) may be more likely to be suppressed by the censors.

With the development of technology coupling with the centralization of ICT in China, the censorship system further evolves and becomes even more comprehensive and sophisticated. Just as setting the agenda for traditional media reporting by, for instance, censoring through guiding directives (Tai, 2014) and threatening or punishing critical or "offending" investigative journalists and organizations (Tong & Sparks, 2009), the boundaries of the online public sphere are still defined and restricted by the regime, but with the strength of commercial social media platforms that own the means of facilitating the online public sphere. In particular, such boundaries are drawn by brutally warning media organizations and opinion leaders, that is, removing some of their published articles for content violation without clearly indicating any source of law. Such boundary drawing is especially done when key social media users create or spread online messages containing elements that could contribute to recognizing, disclosing, or constructing a potential focal point.

The present study offers a rigorous examination of censorship on a dominant social media platform in China by monitoring hundreds of WeChat public accounts' daily publishing activities longitudinally with a research design customized for the project objective and without reliance on any nontransparent methodology. Moreover, this study suggests a theoretical concept, focal point, for future study to investigate censorship, or "content moderation" (Roberts, 2019), which has become a global phenomenon. However, this study is still limited in terms of the scope of content being considered, as semiotic units other than textual ones (e.g., images, videos, and other symbols) are excluded in the analysis. As the WeChatscope system can also capture images contained in WeChat public articles, it would be a starting point for future studies to examine if and how the patterns of censoring are different when considering multimodal semiotic units. Also, although the present study offers new evidence of Internet censorship of the previously understudied but extremely influential social media platform WeChat, more studies on self-censorship with

solid empirical evidence are needed. The "removed by content owners" WeChat public articles captured by WeChatscope can also serve as possible proxies of self-censorship, which warrants further investigation.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Please note: Oxford University Press is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Acknowledgments

Notes

1. The CCAC is headed by President Xi Jinping and is China's policy making unit to setup Internet regulation and management. The Cyberspace Administration of China (CAC) is the working agency reporting to CCAC and is responsible for implementation of the Internet control measures. The CAC's director is one of the Vice Ministers of the Central Propaganda Department of the Chinese Communist Party. The CAC oversees the regulation of Internet service and content providers in China, for example, holding meetings with and issuing regulation directives to Internet service providers (Qiang, 2019; Tai, 2014), who are the actual entities to execute censorship decision and information control. In other words, the actual execution power of Internet censorship is directly exercised by the Internet service providers, who are directed by the Chinese authorities.
2. Some use the third party Weixin search engine, for example, Sogou, to gather WeChat public account data. However, the data access is restricted in many ways, such as anti-crawling mechanisms and time-limited weblink, which virtually makes systematic data collection impossible. Also, the sampling pool of WeChat public accounts and completeness of the contents are nontransparent and thus questionable.
3. The study was approved by the Human Research Ethics Committee for Non-Clinical Faculties, The University of Hong Kong (Reference number: EAl708007).
4. WeChatscope API: http://wechatscope.jmsc.hku.hk
5. This sample is the best available data set collected with WeChatscope. We believe it is representative as the 8-month time period allows us to monitor articles published on usual days as well as days with regular events that are commonly seen to be associated with tighter Internet control, such as the National People's Congress in March, the Tiananmen Square anniversary in June and the National Day in October.
6. We build our sample of WeChat public accounts on the basis of a fixed set of inclusive criteria: accounts whose general targets include publishing articles related to social and political news or commentary are selected. We perform a search on and beyond the WeChat platform to locate relevant public accounts using keywords of social issues, such as labor, LGBT, human rights, or environmental issues, so forth. We also sample "influential" public accounts by including: (a) public accounts for departments and official institutions of central and local government, as well as the Chinese Communist Party; (b) high ranked accounts, that is, accounts publishing highly read articles; and (c)

accounts with article links posted on a major discussion forum (i.e., Tianya Club) or indexed by Baidu search engine.

7. WeChat offers official explanation if an article is removed from the WeChat public account platform. For instance, the removal reason "content violation" is determined when the official explanation is indicated as "The content is unavailable for violation of related law and regulations."

8. The CTM considers the correlation between topics within documents; for instance, an article about the economy of China is likely also relevant to international relation issues but may not be about sexual harassment scandals.

9. The CTM is estimated using partially collapsed variational EM algorithm (Roberts, Stewart, & Airoldi, 2016) in R stm package (Roberts, Stewart, & Tingley, 2014). For model selection, we run the algorithm with topics from 50 to 200 with an interval of 10 and calculate the residuals, semantic coherence, and exclusivity for each model.

10. Cases are dropped if: (a) the censored article contains no text content (e.g., it contains only images); (b) no remaining articles from the same public account published during the 1-month period.

11. Among the 751 public accounts, about 55.9% of them are operated by individuals, 34.5% are enterprise accounts, 6.5% are media accounts, 1.86% are institution accounts, and 0.1% are organization accounts (8 accounts do not have such information available to the public). Comparing to a report published by Newrank (2018) with a much larger sample (0.54 million public accounts), our sample contains higher proportions of individual and media accounts (they report 41.7% and 1.7%, respectively), but smaller shares of enterprise, institution, and organization accounts (they report 46.9%, 7.3%, 2.4%, respectively). This is a natural difference as: (a) we focus more on the public accounts that tend to post more about social and political news and issues, therefore many enterprise accounts posting more about commercial or entertainment content are excluded, and (b) the accounts in our sample are those with at least one article censored in the 8-month period of our data collection, therefore less institution and organization accounts are included since many of them are state-affiliated accounts.

12. Bootstrap aggregation: resample the data × times to grow × trees.

13. Predictions are made with cases other than those used to construct the tree.

14. The above data processing and analysis are performed using R and the following packages: jiebaR (Qin & Wu, 2018) for word segment, quanteda (Benoit et al., 2018) for all other data processing, and randomForest (Liaw & Wiener, 2002) for Random Forest modeling.

15. The reading was mainly done by the first author. The detailed notes regarding the elements and themes each article contains were taken. The organized elements, themes, and conclusion derived from such reading was then shared and discussed with the second author and a research assistant, who is familiar with the media environment in China.

16. Topics are labeled with referencing to their associated words that are either with highest probabilities under each topic or are "unique" to the topic in question (e.g., appear less frequently in other topics), as well as to their highly associated documents (i.e., articles that are composed with high proportion of the topic).

17. Roughly translated as "state advances, private sector retreats" which indicates the negative influence of state-owned firms' success (Hsieh & Song, 2015).

18. Chinese television host and producer who uncovered contract and tax scandals of the film industry in China.

19. The estimation of this RF model is based on 500 trees and 7 variables tried at each split.

20. Permuting each of the three predictors increases the misclassification rate of the OOB sample (with all predictors untouched) by 6.11, 5.67, and 5.40%, respectively.

21. In China, the word "rumor" is more commonly used, by the authorities and the public, than words such as "misinformation," "disinformation," or "fake news"—although in many cases they use "rumor" when referring to groundless allegations rather than gossips. We go with "rumor" here as words such as "misinformation" is usually translated in other ways.

22. We experiment with a range of numbers of variables to select one with the best prediction rate.

23. The estimation of this RF model is based on 500 trees and 45 variables tried at each split. Among the 2,029 input terms, 1,830 terms are estimated with an importance score larger than 0.

24. It is calculated from OOB data: after constructed, for each tree, we record the error rate with the OOB data run down the tree, then do the same with the values of a term randomly permuted—this is repeated for each of the term. The raw importance score of each term is simply the differences of the recorded error rate and the error rate obtained after permuting the values of the corresponding term. Each score is then normalized by averaging the differences over all trees and divided by the standard deviation of the differences obtained from all trees. We thus only include the terms with a score larger than 0 in the further analysis as other terms do not increase the error rate of classification after their values are permuted and can be seen as carrying no significant information in classifying the articles.

25. The .5 constant is added to ensure the ratio is meaningful when the denominator is 0, that is, when the total occurrence of a term is 0 in the remaining articles.

26. A term is defined as a specific term if it is a "unique identifier" of entities, times, or quantities (Chinchor & Robinson, 1997), a general term otherwise.

27. The variance of the two distributions is detected as a Kolmogorov–Smirnov (K–S) test yields a small p-value ($<.001$).

28. An article is defined as a "trade article" if more than 50% of the article is constructed by the topic labeled as "trade."

29. The final set of terms is again filtered based on their chi-squared values as well as RF variable importance scores.

30. The estimation of this RF model is based on 500 trees and 35 variables tried at each split.

31. The CR ratio of the rest of the seven terms is equal to 1.

32. A total of 38 terms are selected as their CR ratio and RF importance scores are equal to or larger than the third quartile.

References

Allen, M. (2019). Close reading. In *The SAGE Encyclopedia of Communication Research Methods*. Retrieved from https://methods.sagepub.com/reference/the-sage-encyclopedia-of-communication-research-methods

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Mueller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774–778.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of Science. *The Annals of Applied Statistics*, *1*(1), 17–35.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Oxfordshire, England: Taylor & Francis.

Cairns, C., & Carlson, A. (2016). Real-world islands in a social media sea: Nationalism and censorship on Weibo during the 2012 Diaoyu/Senkaku crisis. *The China Quarterly*, *225*, 23–49.

Carter, E. B., & Carter, B. L. (2019). Protests and focal moments in autocracies: Evidence from China, *35*, Retrieved from http://www.brettlogancarter.org/May%202016/AnniversaryV30.pdf

Chen, J., & Xu, Y. (2017). Why do authoritarian regimes allow citizens to voice opinions publicly? *The Journal of Politics*, *79*(3), 792–803.

China Internet Network Information Center (CNNIC). (2019). The 43th Internet Development Report. Retrieved from http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201902/P020190318523029756345.pdf

Chinchor, N., & Robinson, P. (1997). MUC-7 named entity task definition. Proceedings of the 7th Conference on Message Understanding, *29*, 1–21.

Denyer, S. (2017, September 28). The walls are closing in: China finds new ways to tighten Internet controls. *The Washington Post*.

Freedom House. (2018a). *Freedom in the World 2018 Methodology*. Retrieved from Freedom House https://freedomhouse.org/report/methodology-freedom-world-2018

Freedom House. (2018b, September 23). China Media Bulletin: China's growing cyber power, entertainment crackdown, South Africa censorship (No. 129). Retrieved from https://freedomhouse.org/china-media/china-media-bulletin-chinas-growing-cyber-power-entertainment-crackdown-south-africa-censorship-no-129

Fu, K., Chan, C., & Chau, M. (2013). Assessing censorship on microblogs in China: Discriminatory keyword analysis and the real-name registration policy. *IEEE Internet Computing*, *17*(3), 42–50.

Gunitsky, S. (2015). Corrupting the cyber-commons: Social media as a tool of autocratic stability. *Perspectives on Politics*, *13*(1), 42–54.

Harada, I. (2019). China's GDP growth slows to 28-year low in 2018. *Nikkei Asian Review*. Retrieved from https://asia.nikkei.com/Economy/China-s-GDP-growth-slows-to-28-year-low-in-2018

Hsieh, C., & Song, Z. M. (2015). Grasp the large, Let go of the small: The transformation of the state sector in China. *Brookings Papers on Economic Activity*, *Spring 2015*, 295–346.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343.

King, G., Pan, J., & Roberts, M. E. (2014). Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science*, *345*(6199), 1251722.

King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, *111*(3), 484–501.

Lei, Y.-W. (2017). *The contentious public sphere: Law, media, and authoritarian rule in China* (Vol. *2*). Princeton, NJ: Princeton University Press.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lorentzen, P. (2014). China's strategic censorship. *American Journal of Political Science*, *58*(2), 402–414.

Lv, A., & Luo, T. (2018). Asymmetrical power between internet giants and users in China. *International Journal of Communication*, *12*, 19.

MacKinnon, R. (2008). Flatter world and thicker walls? Blogs, censorship and civic discourse in China. *Public Choice*, *134*(1–2), 31–46.

Mehta, J., Starmer, C., & Sugden, R. (1994). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, *84*(3), 658–673.

Newrank. (2018). 公众号6年, 多少已停更？｜新榜数洞. Retrieved from https://mp.weixin.qq.com/s/dMFdMId8Bf9qtQnBRvDbww.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). New York: Academic Press.

Qian, G., & Bandurski, D. (2011). China's emerging public sphere: The impact of media commercialization, professionalism, and the Internet in an era of transition. In *Changing media, changing China* (pp. 38–76). New York, NY: Oxford University Press.

Qiang, X. (2019). The road to digital unfreedom: President Xi's surveillance state. *Journal of Democracy*, *1*, 53–67.

Qin, W., & Wu, Y. (2018). jiebaR: Chinese text segmentation. (Version R package version 0.9.99.). Retrieved from https://CRAN.R-project.org/package=jiebaR

Ringe, N. (2005). Policy preference formation in legislative politics: Structures, actors, and focal points. *American Journal of Political Science*, *49*(4), 731–745.

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2014). *STM: R package for structural topic models. Technical report*. Cambridge, MA: Harvard University.

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. London, England: Yale University Press.

Rodrik, D. (1999). Where did all the growth go? External shocks, social conflict, and growth collapses. *Journal of Economic Growth*, *4*(4), 385–412.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Sender, H. (2015, January 13). China's tech winners set to consolidate. *The Financial Times*.

Sullivan, J. (2014). China's Weibo: Is faster different? *New Media & Society*, *16*(1), 24–37.

Taddy, M. (2012). On estimation and selection for topic models. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics* (AISTATS 2012), 1184–1193.

Tai, Q. (2014). China's media censorship: A dynamic and diversified regime. *Journal of East Asian Studies*, *14*(2), 185–209.

Tencent. (2017). *The 2017 third quarter results*. Retrieved from https://www.tencent.com/en-us/articles/8003451510737482.pdf

Tencent. (2020). *Announces 2019 fourth quarter and annual results [press release]*. Retrieved from https://www.tencent.com/en-us/investors/financial-news.html

The Economist. (2017). *Three kingdoms, two empires; China's internet giants*, *423*(9037), 57.

Tong, J., & Sparks, C. (2009). Investigative journalism in China today. *Journalism Studies*, *10*(3), 337–352.

Truex, R. (2019). Focal Points, Dissident Calendars, and Preemptive Repression. *Journal of Conflict Resolution*, *63*(4), 1032–1052.

Yang, G. (2009). *The power of the Internet in China: Citizen activism online*. New York, NY: Columbia University Press.

Yang, G. (2018). Demobilizing the emotions of online activism in China: A civilizing process. *International Journal of Communication*, *12*, 21.

Zhang, H., & Pan, J. (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.

Zhang, Y., Liu, J., & Wen, J.-R. (2018). Nationalism on Weibo: Towards a multifaceted understanding of chinese nationalism. *The China Quarterly*, *235*, 758–783.