

Why Does China Allow Freer Social Media? Protests vs. Surveillance and Propaganda*

Bei Qin, David Strömberg, and Yanhui Wu[†]

Abstract

This paper documents basic facts regarding public debates about controversial political issues on Chinese social media. Our documentation is based on a dataset of 13.2 billion blog posts published on Sina Weibo—the most prominent Chinese microblogging platform—during the 2009–2013 period. Our primary finding is that a shockingly large number of posts on highly sensitive topics were published and circulated on social media. For instance, we find millions of posts discussing protests and an even larger number of posts with explicit corruption allegations. This content may spur and organize protests. However, it also makes social media effective tools for surveillance. We find that most protests can be predicted one day before their occurrence and that corruption charges of specific individuals can be predicted one year in advance. Finally, we estimate that our data contain 600,000 government-affiliated accounts which contribute 4% of all posts about political and economic issues on Sina Weibo. The share of government accounts is larger in areas with a higher level of internet censorship and where newspapers have a stronger pro-government bias. Overall, our findings suggest that the Chinese government regulates social media to balance threats to regime stability against the benefits of utilizing bottom-up information.

*This paper was previously circulated as "The Political Economy of Social Media in China." A shorter version will be published at the Journal of Economic Perspectives.

[†]Bei Qin is Assistant Professor at the School of Economics and Finance, Faculty of Business and Economics, University of Hong Kong, Hong Kong. David Strömberg is Professor at the Institute for International Economic Studies, Stockholm University, Stockholm, Sweden. Yanhui Wu is Assistant Professor of Finance and Business Economics, Marshall School of Business, University of Southern California, Los Angeles, California, USA. Their email addresses are beiqin@hku.hk, david.stromberg@iies.su.se, and yanhuiwu@marshall.usc.edu.

1 Introduction

Around midnight on March 29, 2014, some Chinese internet night owls noticed that the hazard factor of P-Xylene (PX) had been changed from “low” to “high” on Baidu Encyclopedia—the Chinese equivalent to Wikipedia. The next morning, hundreds of protestors assembled in Maoming—a city in southern China’s industrial heartland—where a large-scale PX plant was planned. At 8:38 am, a message with pictures of the protest was posted on Sina Weibo—the Chinese equivalent to Twitter. Tens of thousands of people joined the protests, demanding responses from local officials, burning a car, and throwing bottles until police dispersed the protestors with tear gas and batons. The next day, pictures of bloodied protestors circulated online. Thousands of posts debating the PX project and condemning the government’s action appeared on various social media platforms.

In the era of advanced information technology, social media can in some cases provide a huge information shock to a country like China, in which information and public communication has been limited by government control. How does such an information shock generated by social media affect the participation of Chinese citizens in political events? And how does the Chinese government respond to this information shock? A first step in addressing these questions is to document the information published and communicated on China’s social media.

In this paper, we document basic facts regarding public debates about controversial political issues on Chinese social media. Our documentation is based on a dataset of 13.2 billion blog posts published on Sina Weibo—the most prominent Chinese microblogging platform — during the 2009–2013 period. We also discuss the implications of our findings.

Our primary finding is that a shockingly large number of posts on highly sensitive topics were published and circulated on social media. For instance, we find millions of posts discussing protests such as the anti-PX event in 2014, and these posts are informative in predicting the occurrence of specific events. We find an even larger number of posts with explicit corruption allegations, and that these posts predict future corruption charges of specific individuals.

This type of social media content may increase the access of citizens to information and constrain the ability of authoritarian governments to act without oversight. In China, social media clearly have the potential to play such a role because of their immense popularity. Nearly one-half of the population has access to the internet, and two of every ten Chinese actively use Weibo. Every day, millions of blog posts are produced, exchanged, and commented upon. Many of these posts reach thousands or even millions of readers. It would be very costly for China’s government to monitor millions of users, especially when they are sometimes posted in large surges in the middle of the night.

However, social media also provide authoritarian governments with new opportunities for political control as noted by Egorov, Guriev, and Sonin (2009), Shirky (2011), Morozov (2012), and Lorentzen (2014). Social media messages are transmitted in electronic form through an infrastructure that is typically

controlled by the government. Recent advances in automated text analysis, machine learning techniques, and high-powered computing have substantially reduced the costs of identifying critical users and censoring messages (Edmond 2013). Governments can use these methods to track and analyze online activities, to gauge public opinion, and to contain threats before they spread.

In an autocracy, how effective can a surveillance system based on social media be? We evaluate this possibility by exploring a straightforward surveillance tool, and we find that social media can be very effective for protest surveillance. Most of the real-world protests and strikes that we study can be predicted one day in advance based on social media content. The method of detection is simple, and the cost is low.

Indeed, Chinese government agencies across the country have invested heavily in surveillance systems that exploit information on social media. The result is illustrated by another anti-PX event, which played out differently, a year earlier. Some citizens in Chengdu posted messages on Sina Weibo calling for demonstrations against the launch of a local PX project on Saturday, May 4, 2013. This call was picked up by the Chengdu government, which immediately took measures including making Saturday and Sunday of that week into working days and requiring students to be in school on those days. Consequently, the demonstration in Chengdu was limited to a few hundred participants, whereas a similar anti-PX protest in Kunming on the same day drew more than 10,000 participants. In this case, surveillance of social media content helped the government contain a protest.

Another important surveillance function of social media is to monitor local governments and officials. In China, many political and economic decisions are delegated to local governments. These decisions need to be monitored, but local news and internal reports are likely to be distorted because local politicians control the local press and administration. In contrast, national politicians regulate social media. In social media, relentless complaints about local officials are abundant. Posts exposing officials who wore Rolex watches, lived in mansions, or had inappropriate girlfriends have resulted in investigations and dismissals. Not surprisingly, we observe millions of posts with explicit corruption allegations in our data.

We find that social media posts related to corruption topics are effective for corruption surveillance. These posts help identify when and where corruption is more prevalent. Furthermore, we can predict which specific politicians will later be charged with corruption, up to one year before the first legal action. This result indicates that social media information can be a useful instrument to identify corruption, which probably holds in other countries as well.

It is not without risk to post information about protests and negative information about local leaders, and some users have been punished for doing so (Freedom House 2012; Reporters Without Borders 2013). However, as far as we know, there is no previous systematic research on the extent and effectiveness of this form of policing. We investigate the scale of these practices by tracking regular users who post on sensitive issues and seeing whether their accounts are subsequently closed.

Government can also use social media as a propaganda channel. In the above anti-PX examples, governments conducted vigorous propaganda campaigns via their accounts on social media. In one case, government accounts continuously blogged that “PX is no more carcinogenic than coffee.” The extent of this type of government postings is difficult to know because they emanate from a plethora of accounts across regions and government levels. In 2012, Sina Weibo reported that approximately 50,000 accounts were operated by government offices or individual officials, but there are no external estimates of this kind of government presence on social media.

We seek to identify government accounts from user names and text analysis of the posts in our data. Based on this approach, we estimate that there are 600,000 government-affiliated accounts, which contribute four percent of all posts about political and economic issues on Sina Weibo, using a measure that includes government organization, mass-organization, and media users. Even when limited to the most restrictive definition of government user, Sina Weibo’s reported number substantially underestimates the government presence on Sina Weibo.

Our findings challenge a popular view that an authoritarian regime would relentlessly censor or even ban social media. Instead, the interaction of an authoritarian government with social media seems more complex. From the government point of view, social media is not only (1) unattractive as a potential outlet for organized social protest but is also (2) useful as a method of monitoring local officials and (3) gauging public sentiments, as well as (4) a method for disseminating propaganda. From the point of view of citizens, any perceived benefits of social media need to be evaluated in a context of (5) possible pervasive policing, punishment, and (6) censorship of such media. As mentioned above, this complex interaction has been discussed by researchers. However, rigorous empirical study on this subject is scant. An exception is Enikolopov, Makarin, and Petrova (2016), who find that social media affected protests in Russia in 2011. The other exception is censoring, which has been studied extensively (for example, Bamman, O’Connor, and Smith 2012; Fu, Chan, and Chau 2013; King, Pan, and Roberts 2013, 2014; Zhu, Phipps, Pridgen, Crandall, and Wallach 2013). In contrast, we will not discuss censoring, but present evidence key to all the other five points listed above.

We begin with an overview of the development of social media in China, followed by a description of the data. We proceed to analyze protests, corruption, and the government presence on Sina Weibo. We finally discuss the implications of social media, given our results.

2 Background

By 2013, there were 618 million Chinese internet users, accounting for approximately 46 percent of the Chinese population. This rate is slightly higher than the global average of 39 percent (China Internet Network Information Center 2014; International Telecommunication Union 2013). Of these internet users,

281 million (45 percent) actively participated in microblogging, which refers to social media outlets that focus on short messages, individual images, or perhaps video links (as opposed to social media like Facebook that have the potential for longer-form or more detailed communication).

The popularity of microblogs is a recent phenomenon. In 2006, Chinese people became aware of Twitter; the next year, major Chinese counterparts - Fanfou, Digu, and Jiwai - were launched. However, the number of microbloggers grew slowly at first. After the Urumqi riots in July 2009, the Chinese government not only blocked Twitter and Facebook but also shut down most domestic microblogging services. The microblog market was essentially vacant until Sina Weibo appeared in August 2009, and NetEase, Sohu and Tencent followed in 2010. The number of microblog users surged from 63 million at the end of 2010 to 195 million by mid-2011 (China Internet Network Information Center 2011).

Sina Weibo is a hybrid of Twitter and Facebook: up to 140 Chinese characters per tweet, embedded pictures or videos are allowed, and users can send private messages, comment, and re-post. With its easy access and use, Sina Weibo soon became the most popular microblogging platform in China. By 2010, it had 50 million registered users, and this number doubled in 2011, reaching a peak of over 500 million at the end of 2012. Since 2013, Sina Weibo has lost some ground to WeChat, a cellphone-based social networking service, but has remained an influential platform.¹

In recent years, microblogs have provided some of the most extensive and vivid discussions and debates in China. According to the Reports on Public Opinion 2010-2013, for example, in 2012, the two most popular Facebook-type social media platforms in China - Renren and Kaixin - covered the top 20 public events listed by the Public Opinion Monitoring Agency (which is run by the government newspaper People's Daily) in 20 million posts. However, Sina Weibo - the leading microblog site at the time - covered the same events in more than 230 million posts.

The Chinese central government has the power and tools to limit sensitive content on social media, if it wishes. The sensitive social media content that we find is unlikely to be the result of inability to clean up social media. Instead, the government must perceive benefits from leaving this content visible. Two primary tools are used to limit content.

The first tool is policing - to punish users who post sensitive content to induce self-censorship and to avoid content being posted. Policing is carried out by tens of thousands of information officers and internet monitors who are active at all levels of government (Chen and Ang 2011). Local politicians may use their own internet police to suppress negative information about the regions under their administration, even if blogging about this information is tolerated or encouraged by the central government. Users who post undesired content may receive warnings, have their accounts shut down, and even be imprisoned. Reporters Without Borders (2013) documented a total of 69 netizens in Chinese

¹Our data for this study ends in 2013. In that year, the number of Weibo users dropped by almost 28 million and the utilization ratio dropped by 9.2 percentage points, according to the China Internet Network Information Center (2014).

jails as of February 2013, although the number of unreported cases may be much bigger. There is no previous systematic research on the extent of policing, as far as we know. We will investigate the scale of these practices by tracking regular users who post on sensitive issues to see whether their accounts are subsequently closed.

Of course, personal punishments can occur only if a user is identified. The Chinese government initially allowed users on Sina Weibo to post anonymously. In March 2012, the media control authority required users to reveal their identities to social media providers. However, three years later, service providers had yet to implement this regulation in its entirety.

The other control tool is censoring. Censorship is regulated by the national Propaganda Department of the Chinese Communist Party, as well as by a number of national media control offices. However, in practice, censorship is implemented largely by private service providers who are registered in Beijing. The estimated extent of censorship of Sina Weibo ranges from 0.01 percent of posts by a sample of prioritized users, including dissidents, writers, scholars, journalists, and VIP users (Fu et al. 2013), to 13 percent of posts on selected sensitive topics (King et al. 2013). King et al. find that the Chinese government allows criticism of officials and bureaucrats but censors information about collective action more strictly (27 percent of posts censored). More generally, Bamman (2012) and Fu et al. (2013) find that internet censorship in China focuses on political and minority group issues. Zhu et al. (2013) find that the implementation of censorship is speedy: 30 percent of deletions occur within the first half hour and 90 percent within 24 hours. But unlike these studies, our paper examines the content that is available on microblogs rather than what is removed.

The government can affect debates and sentiments on social media by actively posting their own content. Chinese governments at all levels have opened microblog accounts in an effort to steer public opinion. In 2012, Sina Weibo reported that approximately 50,000 accounts were operated by government offices or individual officials. Governments at different levels also hire internet trolls, nicknamed "the 50-cent party" because some are paid at a piece-rate of 50 cents per post. Some commentators, at the behest of local politicians, may post fake positive reviews about the politicians or the regions under their administration.

3 Data

Our primary data, Sina Weibo posts, were collected by Weibook Corp. During the 2009-2013 period. This company executed a massive data collection strategy to download the posts of active users. First, they identified users as 200-300 million authentic active persons based on the individual's information and interaction with other users. Second, they categorized users into six tiers based on the number of followers. They downloaded the microblogs of the top-tier users at least daily, the second and third tiers every 2-3 days, and the lowest tier downloaded on a weekly basis. Thus, the data include at least some posts

that are later censored. For each post, they provided the content, posting time, and user information (including self-reported location).

In total, the data set that we study contains 13.2 billion posts published from 2009 to 2013. According to our estimation, the Weibook data contains approximately 95 percent of the total posts published on Sina Weibo.² As illustrated in Figure 1, the blue line indicates the number of posts per month included in the Weibook data, and the red line is our estimate of the total number of posts published on Sina Weibo.

From this Weibook database, we extract microblogs mentioning any of approximately 5,000 key words that are related to social and political topics. The key words fall into two groups. The first group refers to categories of issues, including political positions from the central to the village levels, names of top political leaders, social and economic issues (such as corruption, pollution, food and drug problems, disasters and accidents, and crimes), and collective action events (such as strikes, protests, petitions, and mass conflicts). Some words occur at a very high frequency. We collect a random sample that consists of only 10-percent of the posts mentioning these words. The second group of key words refers to specific events that we have recorded, including those in censorship directives issued by the Chinese media control authorities and a large number of massive events from 2009 to 2013. In total, our extracted data contain 202 million posts from 30.6 million different users.³

4 Conflicts, Protests and Strikes

We analyze 545 large collective action events that took place in Mainland China between 2009 and 2012. The list of events was derived from coverage by Radio Free Asia, a nonprofit radio station based in Washington, D.C. We classify these collective action events into four categories, ranked by sensitivity. The first category contains the most sensitive events, which involve direct confrontations between government and the public, including those involving riots and violence. The second category contains protests, including street demonstrations and mass protests, which are typically more expected and organized, less violent, and even often approved by the government. In several cases, protests evolved into riots, as in the Wansheng Event in Chongqing in 2012; we code such events as "conflicts." The third category contains strikes, including strikes in factories and schools and among taxi drivers. The last category includes

²Using the Sina Weibo public API, we downloaded all posts containing the neutral words "ya" or "hei" during four five-minute intervals each day and then divided by the average share of posts that contained these words and the average share of posts contained in these five-minute intervals in a day. We were not able to do this for later years because the public timeline API denied access.

³To analyze word frequencies in the Chinese text, we use the Stanford Word Segmenter to segment the words in each microblog post. We remove stopwords, punctuations, URLs, usernames and non-Chinese characters except meaningful English abbreviations) from the text. We exclude words with more than 30 characters and words occurring fewer than 5 times. We obtain 3.2 million distinct words and 6.0 billion tokens (word occurrences).

anti-Japan demonstrations.

We select key words that identify posts about each event type and extract all posts that mention these key words from the entire Weibook data set. The method for extracting keywords is described in the appendix.

4.1 Content and Users

We initially thought that coverage of these events on social media would be very limited. As just noted, it is well documented that Chinese internet users have been punished after posting about protests and other collective action events (for example, Freedom House 2012) and that these types of posts are censored (King, Pan, and Roberts 2013). But to our surprise, we find a large number of posts covering even the most sensitive collective action events based on our classification. In our data, we identify 382,000 posts in the “conflict” category and over 2.5 million posts in the “protest” category. As this finding attracts some doubt when we presented the paper, we use various approaches to examine it further.

As a starting point, we characterize the "hot topics" in posts about collective action. These topics are identified by words that are used more often in collective action posts than in the entire sample of post. More precisely, we compare the frequency of each word in a given category with the overall frequency of this word in our data set, as in Kleinberg (2006). Table 1 presents the hot topics in order of statistical significance. For example, in the conflict category, "suppression" has the most abnormally high use. Note that the topic ranking is not based on the absolute frequency of the words, but on the use of the word relative to its general usage. For example, "tear-gas bomb" is ranked above "government" because the latter word is more commonly used in general. Other topic words in this category include "police" "violence," "revolt," and "gunfire."

To characterize these data further, we investigate a random sample of 1,000 posts for each of the first three collective action categories in Table 1 and for anti-Japan demonstrations. We manually code whether and how the posts cover a particular type of event, with the results shown in Table 2. Out of our 1,000 post samples, the share of posts that actually cover the events ranges from 50.4 percent for the anti-Japan category to 31.2 percent for the strike category. The more sensitive events like conflict and protests receive more coverage in the form of general and retrospective comments. Here are a few examples to convey a sense of our coding.

- "I saw hundreds of policemen armed with weapons. Fire was everywhere, after some gas containers were bombed." [Conflict, ongoing]
- "A big crowd is gathering in front of the government building holding 'No Forced Demolition of House' signs." [Protest, ongoing]
- "The money from selling lands all went into the pockets of officials. They are nothing but gangsters. We have no choice but to rebel." [Protest, general]

- "Seriously? Taxi-drivers strike again!" [Strike, ongoing.]
- "Low wages, cheap labor. We make tons of Made-in-China, but receive little in return. Migrant workers, strike!" [Strike, general]
- "We will march towards the Japanese Embassy today. Gathering at the People's Square at 10 am. Anyone wanna join?" [Anti-Japan, forthcoming]

This Weibo content predicts real-world collective action events. Panel A of Table 3 reports the average number of posts for each event type published by users in the prefecture where an event took place on the day of the event and on the day before. Suppose, for example, that a strike took place in a given prefecture and day. We then count the number of posts that contain any of our key words related to strikes by users from this prefecture on the same day as the strike and on the day before the strike. We do this for all strikes and report the averages in the table. The average number of posts is much higher on the day of and the day before a collective action event than on other days. To make sure that the posts the day before really do predict these events, and do not arise for spurious reasons such as miscoding of the posting dates, the final column of Table 3 examines coal mine accidents. These should clearly not be predicted by microblogging posts. We obtain data on the locations and days of 253 coalmine accidents during the 2010–2012 period from the State Administration of Coal Mine Safety. We search for word strings related to coalmine accidents in our dataset. While coal mine accidents are covered much more on the day of the accident, they are not discussed more frequently on the day before the accident than on other days. This finding is related to Acemoglu, Hassan, and Tahoun (2014) who find that the number of protesters in Tahrir Square on any given day was predicted by the number of tweets with Tahrir hashtags. Of course, media control in Egypt in 2011 was much less strict than in China.

We also carry out a more detailed analysis, which examines how well microblog discussion, compared to newspaper coverage, predicts when and where these collective action events would occur. We use news reports from 62 general-interest newspapers that covered at least one of these events during the 2010–2012 period. Panels B and C of Table 3 present the results from regressing an indicator for an event occurring on the number of Weibo posts from users in a prefecture that mention the event key words on the event day (panel B) or on the previous day (panel C). While microblogs are highly significant in predicting where and when collective action events take place, newspaper coverage of this event type is uninformative.

In sum, we find literally millions of posts on Sina Weibo that discuss sensitive collective action events. Many of these posts are posted before or concurrent with the events and predict the events. The fact that people begin discussing events before they happen indicates that Sina Weibo may be used to organize or at least to coordinate collective action events.

To investigate whether the users who post this type of sensitive content are identified and perhaps punished, we examine the subsequent posts of the users who blogged about collective action events. 16 percent of these posts are the

last post published by a user in the data that contain any of the 5,000 keywords. In the “conflict” and “protest” categories, the corresponding rates are 17 and 23 percent, respectively. The share of users who exit from our data within five or ten more posts is slightly higher in the full data (38 and 49 percent) than in the conflict and protest categories (33–34 and 41–42 percent). In short, we find that users who posted on these topics continued to post to a similar extent as other users, indicating that their accounts were not more likely to be closed nor were their posts muted.

Another way to investigate whether users are concerned about censorship is to see whether posts on sensitive topics tend to come from user accounts with relatively few posts – which could be a sign that users create separate Sina Weibo accounts for controversial messages, perhaps even with hidden IP addresses. However, the average number of posts from users who blog on sensitive topics is not significantly lower than that of a randomly drawn comparison sample of users (drawn using the number of posts by each user as sampling weights). The bottom line is that, although there are documented cases of people being punished after posting sensitive content, this does not seem to happen on a large scale. Broadly speaking, people do not seem afraid to post on these topics.

4.2 Surveillance

We examine how effective social media information is for government surveillance of collective action events. As noted earlier, government agencies across China have invested heavily in software to track and analyze online activities, to gauge public opinion, and to contain threats before they spread (as reported in Epstein 2013). Presumably, these government agencies desire an early warning system for collective action events. We imagine that such a system might work in two steps. The first step is automated: just have software be alert for days when mentions of certain events spike on social media; second, hire actual human beings to read the posts published on those days.

We explore this automation-plus-manpower method. We study 316 prefectures for all days from January 1, 2010, to December 31, 2012. We study anti-Japan events and strikes because these events are unlikely to be censored and hence we have the same Weibo post information as the government. For the first step, we have the software alert us all days when a user from a city writes a post mentioning any of our event keywords. One day in advance, this alarm would identify 42 of the 43 anti-Japan events and 115 of the 130 strikes. This would require investigating 109,726 and 104,241 prefecture-days, respectively (out of 346,020).

It is very easy to improve this crude method, to find more events and at the same time reduce the number of observations searched. Machine learning tools have been developed for exactly this type of problem (see e.g. Sasaki et al., 2010). We explore something much simpler. We predict the probability of the event occurring in a given day and location, based on the number of Weibo posts by users in this location mentioning the key-words for this event on this day, the day before, and the total number of Weibo posts.

In setting up this kind of early warning system, local governments face a standard trade-off between type 1 and type 2 errors. They can investigate all cases where the predicted probability of an event is positive, however small. This will find most events (i.e. it has a high true positive rate) but this approach will also pick up a lot of false positives. Alternatively, they can restrict attention to cases where this probability is very high, leading to a lower false positive rate but also a lower true positive rate.

Figure 2 shows this trade-off for the anti-Japan demonstrations and strikes. It plots the true positive rate against the false positive rate, for different event probability thresholds. For example, to find all anti-Japan events (true positive rate of one), the local governments would have to search 40 percent of the observations with no event (false positive rate of .4) using information available one day in advance and around 15 percent with concurrent information. In absolute numbers, one would have to search 55,000 prefecture-days to identify 100 of the 130 strikes one day in advance. An uninformative warning system would produce a curve along the 45-degree line. The graph shows that a simple statistic based on keywords is very informative in identifying days when events are likely.

For the second step, we manually read the strike-related social media posts in the 100 prefecture-days with the highest predicted probabilities of having a strike, which took us about 2 hours in total. (Remember, these are very short messages!) Thus, our estimated time-cost of analyzing the 55,000 prefecture-days necessary to discern 100 strikes one day in advance is 1,100 person-hours. This would be the aggregate time cost for all prefectures spread over three years. From that perspective, the cost is very small. The bottom line is that collective action events that are large enough to pose potential threats to the regime are likely to be easily detected using social media data, and they can be detected one day in advance.

Several other interesting points emerge from the above analysis. One is that, when manually reading all of the strike-related social media posts for the top 100 days, we detected 23 strikes during these prefecture-days that were also in our original dataset. However we also identified 14 additional strikes. Our procedure thus shows how social media can be used as a data-collection device in countries where data on relevant social outcomes are scarce but data from social media are abundant.

5 Monitoring Local Politicians

Can social media provide information relevant to holding local politicians accountable to higher-level politicians? We will first describe the content on Sina Weibo related to corruption. We then analyze 200 corruption cases involving high-ranking Chinese government or Communist Party of China leaders, drawing on examples of corruption from the Central Disciplinary Committee of the Communist Party and the Ministry of Supervision, as well as news reports published by the Chinese government's mouthpiece news agency, Xinhua News. We

find that Weibo posts predicts a number of corruption cases one year in advance.

To examine coverage of corruption on social media, we combine two types of microblog posts: those mentioning politicians or political positions and those mentioning corrupt behavior. For the first category, we retrieve posts that mention any major political positions at the central, provincial, prefectural, county, and village levels. We obtain over 11 million total posts in this category. Column 1 of Table 4 shows the number of posts covering each position or top leader. The table is sorted by the number of posts per office, shown in Column 2 – (for instance, there are 31 offices for provincial-level positions). Xi Jinping, the current president of China and the general secretary of the Communist Party of China, is the most discussed leader, with over 1.3 million posts mentioning his name, followed by Wen Jiabao, the former prime minister of China. In general, officials at higher levels are more extensively discussed, and executive positions are covered more than are party secretaries.

Regarding the second category of posts mentioning corrupt behavior, we search for words that are widely used to describe corrupt behavior, wrongdoing, and punishment of officials. The hot topic words in this category are "embezzlement," "corrupt," "government money," "take bribe," and "give bribe" (as shown earlier in Table 1, Column 4). We identify over 5.3 million posts in this category.

5.1 Content and Users

To characterize posts about corruption, we manually inspect 1,000 randomly selected posts. Most of these posts make general comments on corruption. Of the 419 posts that discuss specific corruption cases, 293 were written after the government had taken action. However, 126 posts discuss instances of corruption before government action. These 126 posts can be divided into two types. One type targets specific government officials, illustrated in the following two examples.

- “XXX, the Party secretary of XXX village, misused the money transferred from the central government for low-income villagers to pay his family members and relatives.”
- “XXX, the chief officer of XXX county, embezzled public money by awarding all major government project contracts to his brother’s company. Even worse, he hired gangsters to stab people who reported his corruption to the upper-level government.”

The other type conveys resentment of and anger toward certain corrupt officials. In most cases, these posts talk about positions and government divisions without specifying the names of the officials. Several examples are documented as follows.

- “The black market for government positions in XXX prefecture is rampant. The price is getting higher and higher, the top officials in this prefecture

are becoming richer and richer, and corruption will be more and more severe because the buyers need to make sufficient money to cover their costs.”

- “Without support from the prefecture party secretary and the vice governor, how dare these prefecture officials sell government positions? Crack down on the tigers!”
- “Billions of money went into the pockets of local officials and their business partners! President Xi, Premier Li, and Secretary Wang in the Central Discipline Inspection Department, do you read our microblogs? Can you hear our voice? Please eradicate these corrupt officials! Right now!”

Column 3 of Table 4 shows the estimated percentage of posts mentioning a leader’s position that discuss specific corruption cases. Specifically, we predict the probability that a post about a position discusses specific cases of corruption based on the frequencies of words used. For example, the last row shows that over four percent of all posts that mention village or county party secretaries also mention specific corruption cases. To obtain a broader measure of people’s sentiments towards their leaders, we subtract the number of negative words from the number of positive words in all posts mentioning these leaders (using the National Taiwan University Sentiment Dictionary).

Column 4 of Table 4 shows that County and village party secretaries have the most negative sentiment and represent the largest share of corruption posts. One interpretation of this finding is that these two types of officials are usually viewed as the most powerful low-ranked politicians who have opportunities to be corrupt. An alternative view is that they are the most vulnerable officials in anti-corruption campaigns because they are at the bottom of the Chinese government hierarchy.

Estimating from our random sample of 1,000 posts, our dataset contains approximately 668,000 posts that discuss specific instances of corruption before government action. This provides a wealth of information for upper-level governments seeking to hold lower-level politicians accountable. Clearly, posts of this type are not censored by the central government.

We find some posts explicitly criticize top national leaders, although these posts do not contain explicit corruption allegations. Such posts claim, for example, that democracy and social stability decreased under Hu Jintao’s reign, that the campaign against Bo Xilai was initiated by Xi Jinping as part of a political fight, and that Wen Jiabao shifted capital to Wenzhou to help the children of some top leaders.

We find no evidence that users who post this type of sensitive content are systematically identified and punished. Users who post about corruption continue to post to the same extent as other users. It also seems that people are not afraid of posting concrete corruption allegations implicating powerful local politicians, as these posts are not generated from special accounts with few posts. One possible reason is that, even if local governments can identify bold users, it is risky for them to take action against these users because this may

trigger an even bigger response in social media, which may further jeopardize the implicated politicians' career.

5.2 Prediction and Surveillance

To investigate whether social media posts predict future corruption charges, we study a sample of 200 corruption charges: 15 at the national level, 39 at the provincial level, 114 at the prefecture level, and 32 at the county level. For comparison, we construct a matched control sample of 480 politicians who were not charged with corruption. The control politicians hold similar political positions to and are located in areas geographically near to the charged politicians.

We count the number of posts mentioning the name of each of these 680 politicians and the number of posts that mention both the politician and any word in our corruption category. We calculate the number of posts 2–7 months (as well as 12–23 months) before a corruption charge. Table 5a shows that corrupt and noncorrupt officials are mentioned in roughly the same number of posts 2–7 months before a corruption charge: 49 and 44.4 posts, respectively. However, corrupt officials appear much more frequently in posts that mention our corruption words (3.9 compared to .4). A similar pattern is found in posts published 12–23 months before a charge. Given the substantial difference in the number of corruption posts, it is not surprising that in more formal regression analysis, these posts are highly predictive of corruption charges.

Table 5b presents the results of a regression of the corruption-charge indicator variable on the number of posts mentioning an official's name and corruption. The unit of observation is the official (200 charged and 480 not charged). The dependent variable is a dummy variable indicating whether the official was charged with corruption. The regression also includes the number of posts mentioning only the official's name. This variable is never significant. Columns 2, 4 and 5 include dummy variables for case indicators, which are assigned the same value for an official charged with corruption and his or her matched officials. Standard errors in parenthesis, clustered by case id (charged leader and matched control leaders). The table shows that corruption charges are strongly predicted by the number of posts mentioning corruption 2–7 and 12–23 months before the first government action.

However, a significant number of corrupt officials fly under the social media radar. In particular, 133 officials later found to be corrupt were never mentioned in a corruption post two months or more before the first government action against them. From the perspective of the Chinese central government, which aims to crack down on corruption, a simple rule would be to investigate all officials with at least one corruption post. In our sample, this rule would lead to the investigation of 192 officials, 67 of whom were later charged with corruption.

Social media posts predict which top politicians will be charged with corruption one year before the first legal action. The reason may be that these individuals are indeed more corrupt. Another reason may be that the central government lifts censorship or event that government plants corruption posts about leaders who have lost political support and who will later be charged

with corruption. To investigate the planting of stories, we examined a well-reported scandal involving Bo Xilai, a high-ranking official. We find that there was blanket censoring of posts mentioning Bo Xilai between the start of investigation on March 15, 2012, and the ultimate action undertaken by the Communist Party on September 28, 2012. We find no evidence that censorship focused on posts that were supportive of Bo Xilai or that there was a trend in corruption stories prior to his downfall.

In summary, a massive volume of Sina Weibo posts discuss corruption. These posts help identify the political positions, regions, time, and individuals involved in instances of corruption. The lack of censorship shows that for the central Chinese government, improved monitoring of lower-level officials outweighs the negative publicity of corruption coverage. The results also suggest that local politicians are at least not fully effective in imposing self-censorship on users or otherwise distorting the information.

6 Propaganda

Propaganda posted on social media is largely generated by government-affiliated users: government departments; mass organizations, such as schools and hospitals and industrial associations that are part of the public sector; and state-owned media (note that, per regulation, all general-interest media that are allowed to publish political content are owned and supervised by the government.) We will study these types of high-powered users. We do not study internet trolls hired by the government to distract the public debate (King, Pan, and Roberts 2016). We use two approaches to identify government posts. On a small scale, we manually code posts published by randomly selected users; on a large scale, we use machine learning techniques to discern the language patterns used by well-known government users and thus predict which accounts are affiliated with the Chinese government. We then investigate the goals of these government-affiliated users.

6.1 Volume

In 2012, Sina Weibo reported that approximately 50,000 accounts on Sina Weibo were operated by government offices or individual officials. Our estimation shows that even when limited to the most restrictive definition of government user (excluding mass-organization and media users), this reported number substantially underestimates the government presence on Sina Weibo.

We manually code a sample of 1,000 Sina Weibo users randomly selected from our entire database of 30 million users. A user is classified as a government user if the posts explicitly reveal the user's identity or are mostly related to the activities of a government function; mass-organization users are analogously coded. An account is classified as a media account if the posts reveal that the user is a media outlet or a division. Table 6 shows the result. In our random sample of 1,000 users, 0.5 percent are government users, implying that there are

approximately 150,000 (with a standard deviation of 67,000) government users in the entire data set. State-owned media and mass-organization users contribute an even larger share. In total, these three types of government-affiliated users comprise 2 percent – or 600,000 – users.

Thus, we estimate that the government-affiliated accounts contribute 3.6 percent of all posts in our database (with bootstrapped standard errors of 1.6 percent); see the right panel in Table 6. This percentage is greater than the 2 percent of government-affiliated users because these users publish more posts than others do. Note that these estimates are restricted to the sample of posts that mention words related to political and economic issues. Because we do not include users who write on other topics, the total number of government-affiliated accounts on Sina Weibo is likely to be higher than our estimates. However, the share of government posts may be substantially lower on topics outside politics and economics.

6.2 Identifying Government Affiliation by Language

We also use a linguistics-based approach to predict the probability that a user is affiliated with the government. We restrict our attention to the 5.6 million users who publish more than five posts in our dataset. These users contribute more than three-quarters of all posts. First we identify 1,042 official, government-affiliated, and 538 newspaper accounts by manually inspecting the blogs of thousands of users with user names typically associated with these functions. This is only a small subset of all official accounts, but we can use it to determine what words are characteristic of government accounts and then use this to estimate the total number of government accounts.

To this end, we first adopt a widely used classification algorithm (Support Vector Machine (SVM)) to identify this type of user from a 1 percent sample (28,440) of randomly drawn users based on the frequencies of certain words in their posts.⁴ Based on performance in other classification tasks, SVMs have been identified as one of the most efficient classification methods (Dumais, Platt, Heckerman, and Sahami 1998; Joachims 1998; Sebastiani 2002). In the SVM classification, a large number of words are important. However, just to give a sense of the classification, the words with the highest weight are “Communist Youth League”, “Municipal Party Committee” and “Convention”. To assess how well the SVM performs we use crossvalidation where we iteratively estimate the model leaving out 1 government account and 17 nongovernment accounts and then classify the left-out observations. This classifier has a precision of .81 and a recall of .41. A more familiar statistic is perhaps the t-statistic of a probit regression of a variable indicating a government account on the SVM-output parameter used for classification. This t-statistic is 56, meaning that language

⁴The word frequencies in each post are computed after the pre-processing described at the earlier section in note 4. As inputs to the SVM, we use term-frequency inverse document frequencies. We use the software SVM-light Joachims (1999). Using a new random sample of 500 users, we estimate a probit model of the probability of being a government account conditional on the SVM parameter. See the online appendix for details.

is highly predictive of government accounts.

Since government accounts were over-sampled in the above estimation sample, we cannot use it to estimate the share of government accounts. We instead draw a new random sample of 500 users. In this sample, we estimate a probit model of the probability of being a government account conditional on the SVM parameter. This process is known as Platt scaling and is a

common way to map the SVM parameter estimates to probabilities (Platt, 1999). We combine the SVM-parameters with the probit estimates to predict the probability that each account is a government account. We use the estimated result from this test sample to predict the probability that each of the 5.6 million users is government-affiliated. We compute the average of this probability in total, by province, and by prefecture. This provides us with a measure of the share of government-affiliated users across geographic regions.

At the national level, we estimate that 3.1 percent of the 5.6 million users are government affiliated (with a standard error of .8 percent). This is higher than the 2 percent in the overall sample, but the two estimates are nonetheless roughly consistent, because government users contribute more posts and are thus more strongly represented in the sample of users with more than five posts. The estimated share of posts published by government-affiliated users in this sample is 3.9 percent (with a standard deviation of 1.0 percent).

6.3 Goals of Government Users

Government users of social media may provide neutral information or propaganda. There are several patterns that can help us to distinguish between the two possibilities. For example, in areas where the government perceives that the need for influence is high, we should observe more of both censorship and propaganda and a strong positive correlation them. We should also observe a positive correlation between posts from government users and pro-government bias in traditional media, which are subject to greater government control than social media. Conversely, these correlations should be absent if government users mainly provide neutral information.

Earlier research suggests some other theses about propaganda, as well. Propaganda may be more effective on audiences that share the message sender's view, while the effect of propaganda may be negative when the audience holds opposing views. For example, Adena, Enikolopov, Santarosa, and Zhuravskaya (2014) find that Nazi radio in the 1930s was most effective in places where anti-Semitism was historically high and had a negative effect on support for Nazi policies in places with historically low levels of anti-Semitism. Similarly, in a laboratory experiment, DellaVigna, Enikolopov, Mironova, Petrova, and Zuravskaya (2014) find that Serbian radio exposure caused anti-Serbian sentiment among Croats. If the Chinese regime believes in this argument, then we would expect to find more government-affiliated accounts in Communist Party strongholds.

Finally, propaganda is likely to reduce consumers' valuation of social media. To the extent that service providers can affect the amount of propaganda, we

should see fewer official accounts in areas where the advertisement market is valuable and where competition for consumers is high. Although we lack direct measures of these factors, they are likely to be positively related to local incomes or GDP per capita.

We test these hypotheses using our own measure of government users on Sina Weibo across provinces of China against a measure of censorship and against a measure of bias in Chinese newspapers. The left panel of Figure 3 plots the estimated share of government users against the measure of media bias in the daily newspapers strictly controlled by the Communist Party (from Qin, Ströberg, and Wu 2016). The latter measure is based on nine content categories, including leader mentions, citations of the official Communist Party news agency, and coverage of stories that criticize the regime. The right panel of Figure 3 plots estimated share of government users against a measure of censorship developed by Bamman, O’Connor, and Smith (2012): the share of deleted posts. Guangdong has the lowest share of government users (2.5 percent), whereas Ningxia and Gansu have the highest share (6 percent). The graph looks virtually the same if we use the share of posts published by government users instead of the share of government users.

The two figures show that the estimated share of government users is strongly correlated with both the share of deleted posts and newspaper bias (the correlation coefficient is 0.7 in both cases). This positive correlation indicates that censorship, newspaper bias, and official accounts on Sina Weibo are used for the same propaganda purpose. Note that, in Figure 2, Tibet has more deleted posts than expected. Perhaps this is an indication that propaganda is not viewed as particularly effective in Tibet because of weaker underlying support for the Chinese central government.

The two figures show that the estimated share of government users is strongly correlated with both the share of deleted posts and newspaper bias (the correlation coefficient is 0.7 in both cases). This positive correlation is consistent with the hypothesis that censorship, newspaper bias, and official accounts on Sina Weibo are used for the same propaganda purpose. Note that, in Figure 3, Tibet has fewer government users than expected. Perhaps this is an indication that propaganda is not viewed as particularly effective in Tibet because of weaker underlying support for the Chinese central government.

Other correlations are also consistent with the belief that government users of social media are engaging in propaganda. Table 7 offers an illustrative cross-section regression. The dependent variable is the share of government users of Sina Weibo across prefectures of China. We use GDP as a measure of economic development. We include a variable, “CPC stronghold,” indicating areas where the Communist Party is comparatively more appreciated due to historical reasons (Qin, Ströberg, and Wu 2016). Conversely, some areas have a history of Western influence, notably, the areas that were part of a treaty port controlled by Western powers during the 1840–1910 period (Jia 2014). Other variables in the regression include the distance to Beijing, latitude, longitude, and population.

Our estimated share of government users across prefectures is significantly

lower in areas with high levels of GDP and is higher in Communist Party of China strongholds. The latter result is consistent with the view that propaganda is more effective in areas where the audience shares the ideology of the sender. The estimated share of government users also appears higher in areas that are closer to Beijing and in areas that are more populous. To sum up, these patterns are consistent with propaganda being the main goal of this government social media content.

7 Conclusions

We use a large dataset of blog posts from the most prominent Chinese microblogging platform—Sina Weibo—over the 2009–2013 period to document a number of basic facts. We now discuss how to make sense of these facts and what they may imply for outcomes that we are ultimately interested in, such as corruption, regime stability, local and central accountability, and the central–local balance of power.

Given the immense efforts to police and censor social media, it may seem puzzling that we find so much sensitive material available on social media. What are the private returns to posting this material? Why does the regime not punish users on a large scale, and why does the regime not censor all the sensitive material? We suggest the following explanation.

The central government limits its censoring because only a small fraction of potentially sensitive material is likely to pose a meaningful threat to the regime. Although diverse and even dissenting public opinion may displease the regime, a complete cleanup of sensitive content can impair the regime’s ability to learn from bottom-up information and to address social problems before they become threatening. Therefore, there exists a subtle trade-off in information control faced by an authoritarian regime.

Social media users have an incentive to speak out about local problems as they expect the central government to address these problems. In the corruption posts, we find frequent explicit appeals for central-government action. The posts about protests and strikes may help organize the events, but, importantly, they also make these events visible to the central government. This may force local leaders to deal with the problems that cause the protests and strikes. This line of reasoning may also explain why protests and strikes are rising in China⁶ even though we find it is simple and cheap to use social media to identify these events one day in advance. Local officials may not want to suppress them through police and violence because this will trigger an even larger response on social media.

The visibility-enhancement effect of social media may also explain why we find no evidence of large-scale punishment of dissent users. One might think that this finding just reflects the consistency between censorship and policing—the posts that we observe were not sensitive enough to be censored, and people are thus not punished for posting them. However, the censors and the internet police are not the same people. The central government implements censorship while local governments can rely only on policing. So the question is why

local politicians do not punish users whose posts may jeopardize their career, for instance, those who accuse them of corruption. One reason is that local politicians are not able to identify these users. Another reason is that social media can make local politicians' punishment of dissenting users visible to the national leaders, which exposes local politicians to great political risks.

Given that the central government uses information from social media to monitor local official, it is not surprising that local officials actively run their microblogs on Sina Weibo. Local politicians can signal their loyalty to the central government by engineering propaganda that promotes the party line while also deflecting corruption charges and demonstrating their ability to identify and solve local problems. However, it is costly to produce high-quality propaganda to dominate a discussion among millions of users. We estimate that government-affiliated accounts contribute around 4 percent of all posts about politics and economics on Sina Weibo. While this is substantially larger than Sina's official numbers, it is not overwhelming. In this light, it is not surprising that we find that the local leaders are unable to effectively dilute negative information about, for example, corruption and collective action.

What are the implications for the outcomes that we ultimately care about? Our findings suggest that social media in China primarily affects those outcomes in which the central regime and general users share a common interest. For example, the regime and social media users both benefit from fighting local corruption and other abuse of power by local leaders. In this aspect, Chinese social media seem to play a positive role in public affairs at the local level, improving the public's access to information, engagement in public debate, and their ability to coordinate mass actions and respond to local problems.

Conversely, outcomes in which the central regime and users have opposing interests are less likely to be affected. For instance, a very limited number of posts discuss top Chinese national leaders in a negative manner. Similarly, social media coverage of large-scale conflicts is muted, either by censorship or by self-censorship. Consequently, one might expect social media to increase the incidence of small and medium-sized protests, while decreasing the probability of large regime-threatening protests. Thus, it is far from clear that social media is capable of restraining the Chinese central government.

Our findings finally suggest that social media in China is likely to increase the power of the central government at the expense of local governments, which will be more closely monitored and have less discretion. Eventually, it may also lead to increased centralization of power as the central government's incentive to decentralize decisions is diminished by its improved ability to acquire local information.

Given the advancing IT technology and the changing Chinese political landscape, it is natural to ask the question of how relevant the results of this paper are outside our sample period that ends in 2013. Since 2012, Sina Weibo has lost ground to other services, most notably WeChat and Qzone. Overall, user behavior and government censoring on these services are similar to that on Sina Weibo, but some differences should be noted. According to Ng (2015), a smaller share of posts are censored on WeChat than on Weibo. WeChat provides a

somewhat less-effective tool to organize large-scale collective action or to expose corrupt officials, since only subscribers to an account can see a post. Thus, it is not surprising that, while WeChat is the most popular platform for instant messaging, Sina Weibo and Qzone are the most popular platforms for public information sharing. A survey-based report issued by CNNIC shows that in 2015, the number of social media users who listed Sina Weibo as their first choice to follow current news events and social issues was twice as many as those who listed Qzone.⁵

Since 2012, Chinese governments have escalated their efforts to control social media. As part of this endeavor, since March 2012, users have been required to reveal their true identities to social media providers. Because our sample period spans until the end of 2013, we can provide a glimpse of the effect of tightening control and increased competition from WeChat. Our data show a fall in the number of total Weibo posts from 2012 to 2013 by around 30 percent. However, the number of posts about sensitive topics (conflict, protests, strikes, and corruption) changed only marginally. This result indicates that the traffic that moved to WeChat and other services was concentrated on nonpublic topics such as personal life and private chatting. It also suggests that the stricter regime still found it in their interest to not fully censor posts about the sensitive topics that we study.

References

- [1] Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun. 2014. "The Power of the Street: Evidence from Egypt's Arab Spring." NBER Working Paper 20665.
- [2] Adena, Maja, Ruben Enikolopov, Veronica Santarosa, and Katia Zhuravskaya. 2014. "Radio and the Rise of Nazis in Pre-War Germany", forthcoming in *Quarterly Journal of Economics* 130(4): 1885–1939.
- [3] Bamman, David, Brendan O'Connor, and Noah Smith. 2012. "Censorship and Deletion Practices in Chinese Social Media", *First Monday* 17(3).
- [4] China Internet Network Information Center. 2011. "The 28nd Statistical Report on Internet Development in China" July 2011, Beijing.
- [5] China Internet Network Information Center. 2013. "The 32nd Statistical Report on Internet Development in China" January 2013, Beijing.
- [6] China Internet Network Information Center. 2014. "The 34th Statistical Report on Internet Development in China" January 2014, Beijing.
- [7] Chen, Xiaoyan and Peng Hwa Ang. 2011. "Internet Police in China: Regulation, Scope and Myths". In *Online Society in China: Creating, Celebrating, and Instrumentalising the Online Carnival*, ed. David Herold and Peter Marolt, 40–52. New York: Routledge.

⁵The Chinese version of the report can be downloaded from <http://www.cnnic.cn/>.

- [8] DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova and Ekaterina Zhuravskaya. 2014. "Cross-border media and nationalism: Evidence from Serbian radio in Croatia." *American Economic Journal: Applied Economics* 6(3): 103-32.
- [9] Dumais, S., Platt, J., Heckerman, D., & Sahami, M.. 1998. "Inductive learning algorithms and representations for text categorization". Proceedings of the 7th international conference on information and knowledge management, 48-155. ACM Digital Library.
- [10] Edmond, Chris. "Information manipulation, coordination, and regime change." *The Review of Economic Studies* (2013): rdt020.
- [11] Egorov, Georgy, Sergei Guriev, and Konstantin Sonin. 2009. "Why resource-poor dictators allow freer media: A theory and evidence from panel data." *American Political Science Review* 103.04: 645-668.
- [12] Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2016. "Social Media and Protest Participation: Evidence from Russia." Available at SSRN 2696236.
- [13] Epstein, Gady. 2013. "China's Internet: A Giant Cage," *The Economist*. April 6. <http://www.economist.com/news/special-report/21574628-internet-was-expected-help-democratise-china-instead-it-has-enabled>
- [14] Freedom House. 2015. "2015 Freedom of the Press Data" <https://freedomhouse.org/report/freedom-world/freedom-world-2015#.WFxACX3CB1A>
- [15] Fu, King-wa, Chung-hong Chan, and Marie Chau. 2013. "Assessing censorship on microblogs in China: Discriminatory keyword analysis and the real-name registration policy." *Internet Computing, IEEE* 17(3): 42-50.
- [16] International Telecommunication Union. 2013. "The World in 2013: ICT Facts and Figures," Geneva. <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf>
- [17] Jia, Ruixue. 2014. "The Legacies of Forced Freedom: China's Treaty Ports", *Review of Economics and Statistics*, Vol.96(4): 596-608.
- [18] Joachims, Thorsten. 1998. "Text categorization with Support Vector Machines: learning with many relevant features". 10th European Conference on Machine Learning, volume 1398 of Lecture Notes in Computer Science, 137-142. Berlin: Springer Verlag.
- [19] Joachims, Thorsten. 1999. "Making large-Scale SVM Learning Practical". *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press.

- [20] King, Gary, Jennifer Pan, and Margaret E Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression", *American Political Science Review*, 107(2(May)): 1-18
- [21] King, Gary, Jennifer Pan, and Margaret E Roberts. 2014. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation." *Science* 345 (6199): 1-10.
- [22] Kleinberg, Jon. 2006. "Complex Networks and Decentralized Search Algorithms.", *Proceedings of the International Congress of Mathematicians (ICM)*.
- [23] Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58.2 : 402-414.
- [24] Morozov, Evgeny. 2012. "The Net Delusion: The Dark Side of Internet Freedom." *Public Affairs*, Reprint edition (February 28, 2012).
- [25] Ng, Jason Q. 2015. "Politics, Rumors, and Ambiguity: Tracking Censorship on WeChat's Public Accounts Platform." University of Toronto, mimeo.
- [26] Platt, John C. 1999. "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods." *Advances in large margin classifiers* 10(3): 61-74.
- [27] Public Opinion Monitoring Agency. Various years. *Reports on the Online Public Opinion (2010 – 2013)*. Published by People's Daily.
- [28] Qin, Bei, David Stromberg, and Yanhui Wu. 2016. "Media Bias in China," working paper.
- [29] Reporters Without Borders. 2013. "2013 World Press Freedom Index: Dashed Hopes after Spring." <https://rsf.org/en/news/2013-world-press-freedom-index-dashed-hopes-after-spring>.
- [30] Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. 2010. "Earthquake shakes Twitter users: real-time event detection by social sensors." *Proceedings of the 19th international conference on World wide web*. ACM.
- [31] Sebastiani, Fabrizio. 2002. "Machine learning in automated text categorization", *ACM Computing Surveys*, 34(1), 1-47.
- [32] Shirky, Clay. 2011. "The Political Power of Social Media: Technology, the Public Sphere, and Political Change", *Foreign Affairs*, January/February.
- [33] Zhu, Tao, David Phipps, Adam Pridgen, Jedidiah R. Crandall, and Dan S. Wallach. 2013. "The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions." *arXiv preprint*, arXiv:1303.0597.

Figure 1:

Estimated number of Sina Weibo post by Weibook and API

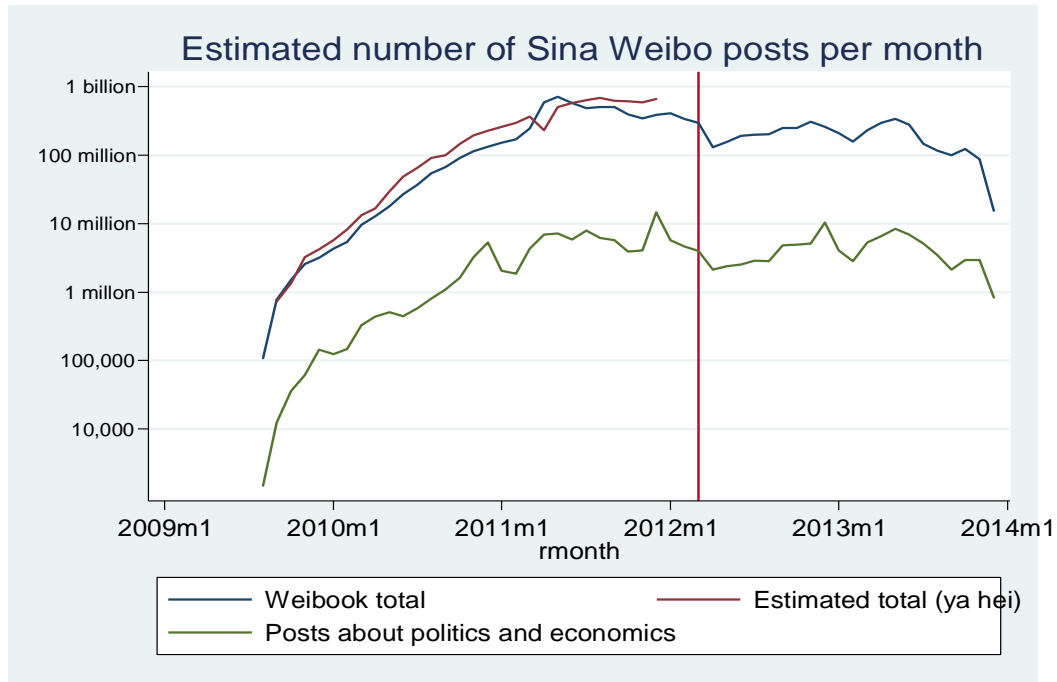


Figure 2:

Event prediction and detection

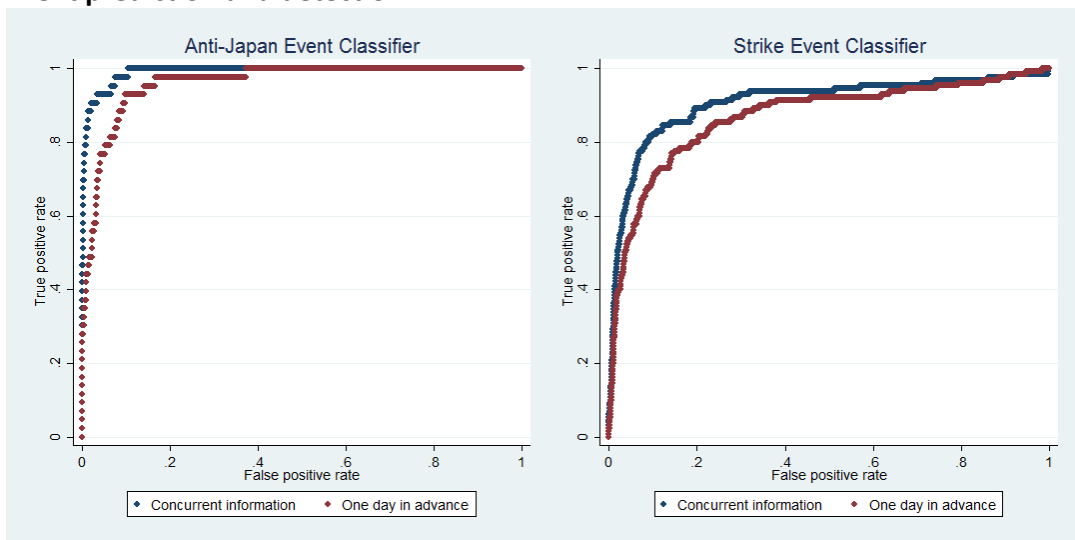
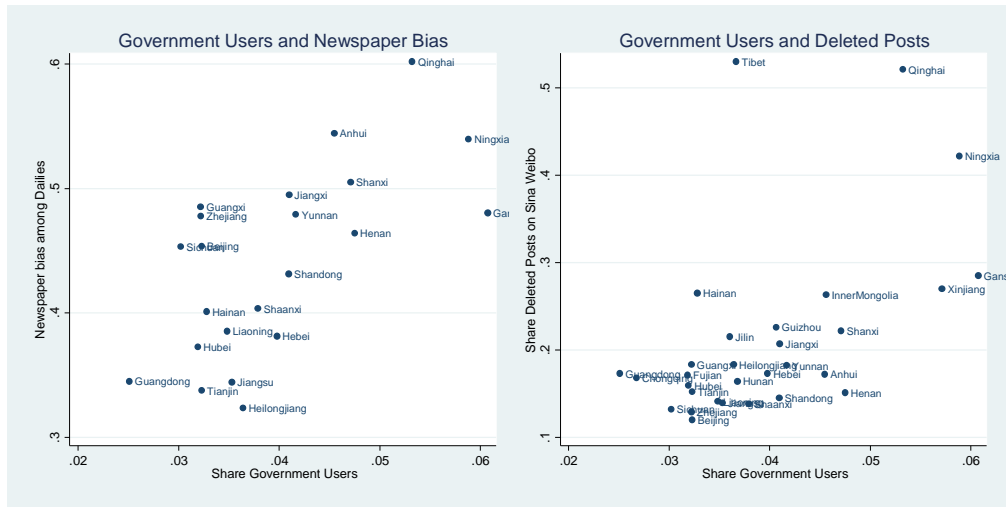


Figure 3

The Share of Government Users on Sina Weibo across Provinces versus Newspaper Bias and Censorship



Note: Each dot represents one province in China. The left panel plots estimated share of government users against the measure of media bias in the daily newspapers strictly controlled by the Communist Party (from Qin et al. 2016). The right panel plots estimated share of government users against the measure of censorship developed by Bamman et al. (2012), the share of deleted posts.

Table 1: Hot Topics by Category

<i>Conflict (sensitivity: very high)</i>			<i>Protest (sensitivity: high)</i>			<i>Strike (sensitivity: medium)</i>			<i>Corruption</i>		
# of posts: 382,232			# of posts: 2,526,325			# of posts: 1,348,964			# of posts: 5,326,897		
Frequency	Word	Translation	Frequency	Word	Translation	Frequency	Word	Translation	Frequency	Word	Translation
322,797	镇压	suppression	647,711	示威	demonstration	1,361,854	罢工	strike	1,455,878	贪污	embezzlement
32,117	冲突	conflict	534,784	静坐	sit-in	69,068	罢课	student strike	1,658,687	腐败	corrupt
19,124	警民	police and people	430,112	自焚	self-immolation	101,887	工人	workers	681,055	公款	money
17,460	催泪弹	tear-gas bomb	260,574	讨薪	ask for compensation	98,822	电脑	computer	674,503	受贿	take bribe
31,161	矛盾	contradictory	346,836	游行	parade	65,557	出租车	taxi	556,609	贿赂	give bribe
40,286	警察	police	164,367	请愿	petition	164,549	泪	tears	975,187	官员	officials
14,271	官民	officials and people	113,936	示威者	demonstrators	46,219	工会	trade union	393,125	廉政	government
31,935	暴力	violence	109,339	堵路	stops up the road	91,051	抓狂	driven nuts	639,293	利益	benefit
130,036	被	by	166,600	抗议	protest	55,687	司机	drivers	1,002,491	政府	government
74,391	政府	government	101,845	集会	assembly	48,845	集体	collective	245,606	挪用	diverting
12,002	宽恕	forgiveness	118,262	农民工	migrant workers	52,066	员工	staff	512,006	集团	group
12,764	武力	military force	103,975	思	thinking	157,937	今天	today	201,891	吃喝	food and drink
18,951	军队	army	80,481	静静	static	24,477	的士	taxi	153,731	职权	authority
29,566	民众	populace	60,237	闲谈	chat	22,559	法国人	French	572,569	钱	money
14,701	叙利亚	Syria	58,318	人非	shortcomings of people	51,479	上班	going to work	247,942	贪官	corrupt officials
20,170	抗议	protest	72,753	民工	laborers	16,290	罢市	merchant strike	156,363	滥用	abusiveness
60,068	人民	people	63,719	白宫	White House	40,827	抗议	protest	291,309	原	former
21,521	村民	villagers	130,198	坐	sitting	86,612	手机	cellphone	288,287	干部	cadres
10,264	起义	revolt	60,957	己	oneself	17,679	罢	strike	123,827	行贿	bribery
10,150	开枪	gunfire	37904	玩火自焚	being made to pay for one's evil doings	41586	工资	wages	126,820	情妇	mistress

Note: This table presents the hot-topic words, ordered by how abnormally high the frequency of a word is in the posts that cover a particular category (e.g. conflict) compared to that in the entire data set.

Table 2

Collective Action Posts

	<i>Total posts containing key words</i>	<i>Out of a random 1,000 post sample:</i>				
		<i>Posts actually about defined topic</i>	<i>Forthcoming event</i>	<i>On-going event</i>	<i>Past event</i>	<i>General comments</i>
Conflict	382,232	398	1	11	156	230
Protest	2,526,325	317	2	19	172	124
Strike	1,348,964	312	5	178	39	90
Anti-Japan	2,506,944	504	9	188	42	265

Note: For each category of hot topic, out of the total instances of the words used, we investigate a random sample of 1,000 posts. We manually coded whether and how the posts cover a particular type of event.

Table 3: Event Prediction and Detection
(# per prefecture and day)

VARIABLES	Conflict	Protest	Strike	Anti-Japan	Coalmine Accident
<i>Panel A</i>					
# Weibo posts: day of event	6.1	62.6	167.3	2036.6	3.0
# Weibo posts day before event	3.4	54.3	48.1	924.6	0.7
# Weibo posts: no event	0.7	4.4	2.5	4.5	1.2
<i>Panel B</i>					
Regression coefficient					
# Weibo posts	0.647*** (0.196)	1.013*** (0.166)	1.777*** (0.310)	1.105*** (0.209)	1.213*** (0.286)
# newspaper articles	0.002* (0.001)	0.002* (0.001)	0.001 (0.002)	-0.000 (0.001)	
Observations	346,336	346,336	346,336	346,336	346,336
R-squared	0.002	0.006	0.007	0.005	0.004
<i>Panel C</i>					
Regression coefficient					
# Weibo posts day before event	0.381*** (0.137)	0.639*** (0.143)	0.802*** (0.200)	0.614*** (0.133)	-0.143* (0.082)
# newspaper articles day before event	-0.000 (0.001)	0.001 (0.001)	0.000 (0.002)	0.000 (0.000)	
Observations	346,336	346,336	346,336	346,336	346,336
R-squared	0.001	0.006	0.005	0.003	0.004

Note: Panel A: For each event type, this table reports the average number of posts published by users in the prefecture where an event took place on the day of the event (first row) and on the day before it (second row). The third row shows the analogous average number of posts on days when no such event took place. Panels B and C: The unit of observation is prefecture and day. The dependent variable is a dummy for the occurrence of an event. The key independent variables are the log of 1 + the number of Sina Weibo posts mentioning words related to the event and the log of 1 + the number of newspaper articles mentioning event words. Coefficients and standard errors are multiplied by 1,000 to make the table more readable. Controls include prefecture and year fixed effects. Standard errors, clustered by prefecture, in parenthesis.

Table 4
Coverage of Politicians

	(1)	(2)	(3)	(4)
			<i>% Posts discussing specific corruption cases</i>	<i>Measure of sentiment</i>
<i>Position</i>	<i># Posts</i>	<i># Posts per position</i>		
Xi Jinping	1,374,780	1,374,780	0.23	0.88
Wen Jiabao	1,318,345	1,318,345	0.15	0.51
Li Keqiang	401,451	401,451	0.14	0.81
Hu Jintao	326,621	326,621	0.10	1.16
Provincial governor	728,386	23,496	1.88	-0.19
Provincial Party secretary	403,074	13,002	1.91	0.52
City mayor	3,431,471	10,305	1.39	0.17
City party secretary	718,856	2,159	2.81	0.28
County governor	719,634	251	1.21	-0.70
County Party secretary	324,522	113	4.40	-0.88
Village chief	1,053,346	25	0.65	-0.51
Village Party secretary	144,742	3	4.26	-1.40

Note: Column 1 shows the number of posts covering each position or top leader. The table is sorted by column 2—the number of posts per office. Column 3 shows the estimated percentage of posts mentioning a leader’s position that discuss specific corruption cases. Column 4 presents a broader measure of people’s sentiments towards the leader or type of leader. See text for details.

Table 5a

Mean Number of Posts, by Corruption Charge

	<i>2–7 month lag</i>		<i>12–23 month lag</i>	
	<i>Name</i>	<i>Corruption</i>	<i>Name</i>	<i>Corruption</i>
Corrupt official	49.0	3.9	148.3	4.7
Noncorrupt official	44.4	0.4	121.1	1.8

Note: To investigate whether social media posts predict future corruption charges, we study a sample of 200 corruption charges. For comparison, we construct a matched control sample of 480 politicians who were not charged with corruption. We count the number of posts mentioning the name of each of these 680 politicians and the number of posts that mention both the politician and any word in our corruption category. We calculate the number of posts 2–7 months (as well as 12–23 months) before a corruption charge.

Table 5b

Dependent variable: corruption case dummy

VARIABLES	I	II	III	IV	V
# posts mentioning name and (2-7 months before first action)	0.0042*** (0.0010)	0.0065*** (0.0015)			0.0038*** (0.0009)
# posts mentioning name and (12-23 months before first action)			0.0035** (0.0014)	0.0050** (0.0024)	0.0029 (0.0019)
Observations	680	680	680	680	680
R-squared	0.014	0.053	0.009	0.044	0.052
Fixed Effects	No	Case Id	No	Case Id	Case Id

Note: Unit of observation is official. The regression also includes the number of posts mentioning the official’s name. This variable is always insignificant. Standard errors in parenthesis, clustered by case id (charged leader and matched control leaders).

Table 6

Government Presence on Sina Weibo

<u>Type</u>	<i>Users</i>			<i>Posts</i>	
	<i>Percent</i>	<i>Estimated #</i>	<i>Standard deviation</i>	<i>Percent</i>	<i>Standard deviation</i>
Government	.5	149,746	66,801	.2	.1
Media	.5	149,746	66,801	2.3	1.6
Mass organization	1.0	299,491	94,233	1.1	0.5
Government-affiliated	2.0	598,982	132,590	3.6	1.6
Others	98.0	29,350,118	132,590		

Note: Based on 1,000 Sina Weibo users randomly selected from our entire database of 30 million users. A user is classified as a government user if the posts explicitly reveal the user's identity or are mostly related to the activities of a government function; mass-organization users are analogously coded. An account is classified as a media account if the posts reveal that the user is a media outlet or a division.

"Government-affiliated" is the sum of "government," "media," and "mass organization."

Table 7

Dependent Variable: Share of Government Users

	I
GDP	-0.849*** (0.103)
CPC stronghold	0.533** (0.236)
Treaty port	-0.079 (0.166)
Distance to Beijing	-0.464*** (0.165)
Population	0.366*** (0.129)
Latitude	0.052*** (0.016)
Longitude	-0.037*** (0.014)
Observations	259
R-squared	0.358

Note: The unit of observation is the prefecture. The result is obtained by cross-sectional ordinary least-squares regression. GDP and Population values are from 2010, which is the first year Sina Weibo was in use. Robust standard errors are in parentheses. "CPC" is "Communist Party of China." "Treaty port" is a proxy for Western influence.

*** p<0.01, ** p<0.05, * p<0.1

Appendix: keywords

This sections lists the search strings that we used to identify posts in each of the topics conflict, protest, strike, corruption and political positions/politicians.

Table A1: Conflict, protests and strikes

Conflict	Protest	Strike
被袭击	堵路	罢弛
被袭击 and (政府 or 官员 or 干部)	非法集会	罢工
威胁政府	集会 and (群众 or 公众 or 大规模)	罢课
催泪弹 and (群众 or 政府 or 警察)	静坐	罢驶
官民 and (矛盾 or 冲突 or 暴力 or 对抗)	请愿	罢市
集体冲突	示威	罢运
警民 and (矛盾 or 冲突 or 暴力 or 对抗)	讨薪	
军民 and (矛盾 or 冲突 or 暴力 or 对抗)	学潮	
镇压	工潮	
	游行	
	学生 and 闹事	
	封堵 and (政府 or 群众 or 工人 or 公路)	
	自焚	
	千人下跪	
	not 反日	
	not 抗日	
	not 反日	

Table A2: Corruption

Corruption
腐败 and (政府 or 部门 or 官员 or 干部 or 官员)
腐败分子
公款
贿赂
廉政
买官
卖官
挪用
社保 and (贪污 or 腐败 or 挪用)
受贿
索贿
贪污
行政腐败
徇私
滥用职权
利益集团
侵占 and (政府 or 官员 or 部门 or 干部)
情妇 and (政府 or 官员 or 部门 or 干部)
失职 and (政府 or 官员 or 部门 or 干部)
私分 and (政府 or 官员 or 部门 or 干部)
私生 and (政府 or 官员 or 部门 or 干部)
伪造 and (政府 or 官员 or 部门 or 干部)
舞弊 and (政府 or 官员 or 部门 or 干部)
虚报 and (政府 or 官员 or 部门 or 干部)
虚开 and (政府 or 官员 or 部门 or 干部)
诈骗犯 and (政府 or 部门 or 官员 or 干部)
诈骗罪 and (政府 or 部门 or 官员 or 干部)

Table A3: Politicians

Political position/person	Keywords
Xi Jinping	习近平
Xi Jinping	习大大
Xi Jinping	习总
Li Keqiang	李克强
Hu Jintao	胡锦涛
Wen Jiabao	温家宝
Wen Jiabao	温总理
Provincial governor	省长 or 区主席 or 省主席 or 区副主席 or 省副主席
Provincial Party secretary	(书记 and (省委 or 自治区)) or 省委书记 or 省副书记
City mayor	市长 or 州主席 or 州专员 or 地区专员
City party secretary	(书记 and (市委 or 地委 or 自治州)) or 市委书记 or 市副书记
County governor	县长
County Party secretary	书记 and 县委
Village chief	村长
Village Party secretary	村支书