Social Media and Collective Action in China

Bei Qin, David Strömberg, and Yanhui Wu*

January 2024

Abstract

This paper studies how social media affects the dynamics of protests and strikes in China during 2009-2017. Based on 13.2 billion microblog posts, we use tweets and retweets to measure social media communication across cities and exploit its rapid expansion for identification. We find that, despite strict government censorship, Chinese social media has a sizeable effect on the geographical spread of protests and strikes. Furthermore, social media communication considerably expands the scope of protests by spreading events across different causes (e.g., from anti-corruption protests to environmental protests) and dramatically increases the probability of far-reaching protest waves with simultaneous events occurring in many cities. These effects arise even though Chinese social media barely circulates content that explicitly helps organize protests.

Keywords: Social Media, Protests, Strikes, China, Media Control

*Qin: Department of Accountancy, Economics and Finance, Hong Kong Baptist University, Hong Kong, beiqin@hkbu.edu.hk; Strömberg: Economics Department, University of Stockholm, SE-106 91 Stockholm, Sweden, david.stromberg@ne.su.se; Wu: HKU Business School, University of Hong Kong, Hong Kong, yanhuiwu@hku.hk. We are grateful to Davide Cantoni, Chun-Fang Chiang, Peter Hull, Ruixue Jia, Adam Szeidl, Jaya Wen, David Yang, Noam Yuchtman, and numerous seminar and conference participants for helpful comments. This study was approved by the Regional Ethics Review Board of Stockholm, number 2017/1288-31. The project has received funding from the European Research Council (ERC project number 742983) and the Swedish Research Council (diary number 2016-03220).

1 Introduction

Recent studies show that communication through social media is powerful in organizing grassroots protests (Tucker et al. 2016; Steinert-Threlkeld 2017; Acemoglu et al. 2018; Enikolopov et al. 2020; Fergusson and Molina 2020). However, it remains an open question whether this extends to a more strictly controlled media environment. Answering this question is important, as an increasing number of governments are tightening their control over social media to suppress protests (Howard 2011; Woolley and Howard 2019).

In this paper, we study whether and how social media affects protests and strikes in China, the world's largest nondemocratic country, which is known for its extensive media censorship. In China, protests of limited scale and scope are allowed (e.g., Cai 2008; Lorenzten 2017), but the regime is highly cautious of protests that may snowball into large social movements targeting national problems and challenging the ruling party.[1]

A key innovation of our study is the focus on the impact of social media on the dynamics of protests and strikes; that is, how information diffusion via social media helps spread events from one city to another, potentially leading them to evolve into substantial protest waves. Grasping how events spread is essential to understanding the emergence of nationwide protest waves, and, in particular, understanding the probability of the occurrence of very large waves of simultaneous events across many cities. To estimate the causal effect of social media on event dynamics, we use the rapid expansion of the information network across cities (prefectures) based on large-scale textual data from Sina Weibo (Weibo for short)—the leading microblogging platform in China.

The first empirical question is whether information about protests and strikes circulates on Weibo. Based on the same database used in Qin et al. (2017), we identify around 4 million microblog tweets mentioning protests or strikes during 2009-2013. These tweets appear in bursts around real-world protests and strikes. A large proportion of them discuss the causes of the events, such as corruption or wage arrears, and criticize the government for poor policies and misbehavior. Many of the most retweeted posts express anger towards the government and sympathy for the protesters. This type of information is largely absent from traditional Chinese media. In fact, before the advent of social media, no other sources inform large audiences of protests and strikes in China.

To measure how information diffuses on Weibo, we take advantage of our unique data on retweets (forwards). We use retweets on topics other than protests and strikes from one city to another within the past few months to measure information diffusion between cities.

[1] A cautionary example is Solidarity in Poland, which originated from limited demands for workers' rights and better economic conditions but quickly turned into pervasive resistance that proved fatal to the regime. More recent examples involve social media. In 2013, small groups of environmentalists protested the proposed destruction of Gezi Park in Istanbul; the protest exploded to attract millions of tweets on social media and eventually the participation of hundreds of thousands of people in Turkey. In 2013-2014, after the Ukrainian president Viktor Yanukovych failed to sign an agreement with the EU, opposition leaders posted calls for protests on Twitter and Facebook. The protests then grew rapidly and spread widely, eventually culminating in the ouster of Yanukovych.

The fact that a user retweets a message indicates that the user has seen it. Thus, retweets are an effective measure of information diffusion (e.g., Kwak et al. 2010). In a subset of 3 million initial retweets for which we have precise timing and location information, we find that approximately 28% of the retweets occur within one hour of the posting of the original messages, and 75% within one day. After one hour, geographical distance plays no role in explaining who retweets a tweet. In a country as vast and geographically dispersed as China, Weibo provides an unprecedented information network that connects distant individuals and instantly informs a substantial share of the Chinese population of sensitive events.

We examine how this rapidly expanding social media network affects the spread of protests and strikes in a panel of Chinese cities during 2006-2017. Estimating the causal effect of social media on the dynamics of real-world events is challenging. First, cities with strong social media ties typically have strong ties through other information channels, such as phone calls and face-to-face meetings. Second, cities with strong social media ties tend to share similar characteristics (e.g., industry structure) and hence are likely to be exposed to similar shocks that affect the propensity to protest and strike. These two issues correspond to what Manski (1993) calls contextual and correlated effects, which are endemic to network analysis.[2] In addition, a reflection or simultaneity problem appears if simultaneous events are observed in several cities, and these events affect each other at the same time.

We address these problems using the time-series dimension in our panel data. First, we measure the correlated and contextual effects in the pre-Weibo period (when the Weibo network did not exist) and partial them out when estimating the effect of the social media network in the post-Weibo period. Second, our panel structure helps mitigate the simultaneity problem. In a cross-sectional network setting, all events are simultaneous, which makes the reflection problem more acute. In a panel network setting, some events are pre-determined and cannot be affected by later events. We use high-frequency daily data with accurate information on which events come first.

The main findings of our empirical analysis are as follows. First, information diffusion on Weibo creates a strong ripple effect that rapidly spreads protests and strikes across cities. We estimate that, because of Weibo, the occurrence of a protest in one city increases the probability of a protest taking place within two days in any of the other cities by 17%. Relative to the mean event probability, the estimated effect amounts to a 34% increase for protests and a 21% increase for strikes. These results are robust to the use of student mobility across cities as an instrumental variable (IV) for social media connections and are immune to the consideration that social media may increase event observability.

Second, information on Weibo increases the scope of protests and strikes. Not surprisingly, we find that the spread of events is predominantly narrow in scope, in the sense that protests spread to other protests concerning the same cause and strikes spread to other strikes in

[2] Proposed solutions include exploiting the network structure to identify instruments (e.g., Bramoulle et al. 2009), using instruments that are uncorrelated with the error terms and the network (e.g., Acemoglu et al. 2015; Konig et al. 2017), matching (Aral et al. 2009), and explicit randomization (e.g., Bakshy et al. 2012).

the same industry. However, Weibo also induces a significant, albeit smaller, effect on event spread across categories. Although the spreading effect per event is smaller, the aggregate effect of such cross-category spread is greater than within-category spread because the number of events across all of the categories is much larger than within the same category. This finding suggests that Chinese social media helps connect protesters with different goals, which facilitates the formation of substantial protest waves.

Third, Weibo escalates the probability of very large protest waves. We measure the size of a protest wave by the number of cities experiencing protests within the same week. Immediately before the introduction of Weibo in 2009, the largest wave of protests affects cities comprising approximately 4% of China's population. This corresponds roughly to the population share of the two largest Chinese cities combined. We estimate that without Weibo, the share of the affected population would have increased to 6% by 2013, whereas with Weibo, it would rise to nearly 15%, only slightly below the share in reality. This large effect is driven by the strong information connections that social media creates between China's large and distant cities. In addition, by amplifying the feedback from one event to another, social media increases not only the mean but also the variance of the wave size and hence the likelihood of extreme event waves. We estimate that, in 2013, there is a 5% probability of event waves affecting cities containing 22% or more of China's population.

Finally, we find that purging social media of content that explicitly helps organize events does not mute its effect on collective action. By far the most direct explanation (and the one most emphasized in the literature) for social media's influence on protests is that it provides a platform for disseminating logistics information (e.g., where and when to meet) and organizing events. For instance, Enikolopov et al. (2020) report that two thirds of the Russian cities in their sample have social media communities created to organize protest demonstrations. Similar content is found in studies of protests during the Arab Spring and in Turkey and Ukraine (Tucker et al. 2016; Steinert-Threlkeld 2017). However, because of strict government censorship, this type of content is scarce on Chinese social media. Instead, most Weibo tweets about protests and strikes report what causes the event, coupled with emotional reactions.

This type of content can spread events for several reasons. For example, information about an event may indicate the opening of a "window of opportunity" in which protests and strikes are allowed without substantial political risks (Zhou 1993; Weiss 2013; Truex 2019). Such a window may also open if simultaneous protests across several cities increases the likelihood of obtaining concessions or reduces the risk of punishment (e.g., Granovetter 1978; Edmond 2013; Little 2016; Barberà and Jackson 2020). Another explanation is that protesters are motivated purely by emotions (e.g., Jasper 2008; Passarelli and Tabellini 2017), which rapidly spread across regions through social media (e.g., Kramer et al., 2014).

Overall, our findings demonstrate that although social media coverage of protests and strikes provides useful information for government surveillance (Qin et al. 2017), it also

spreads protests and strikes across cities and escalates isolated events into far-reaching waves. Strategic censorship that deletes content useful for organizing protests while retaining information that helps identify social problems and gauge public sentiment cannot prevent protests and strikes from spreading. This points to a fundamental trade-off in autocracies between eliciting information from the bottom up and containing anti-government collection action.

This paper contributes to the emerging literature on how information and communication technology affects collective action and regime changes in nondemocratic countries. A number of papers examine the political effects of the expansion of communication infrastructure, notably high-speed Internet or mobile phone networks (e.g. Christensen and Garfias 2018; Manacorda and Tesei 2020; Guriev et al. 2021). A few studies aim to identify the causal effects of social media on protests (Zhuravskaya et al. 2020). In particular, Enikolopov et al. (2020) show that penetration of the dominant Russian online social network leads to protests against election fraud in Russia. Fergusson and Molina (2020) find that Facebook has a positive effect on collective action on a global scale. Together with other research (e.g., Tucker et al. 2016; Steinert-Threlkeld 2017; Acemoglu et al. 2018), these studies highlight the role of social media in organizing events, as they find extensive social media content containing information on where and when to meet and how to act.

Our paper advances this literature in several ways. First, we study how social media affects political collective action in China, which is a considerably more strictly controlled environment than those previously studied. Second, we focus on the dynamic interaction between geographically dispersed protests. This dynamic interaction is key to understanding whether local protests turn into widespread social movements that have important consequences for an authoritarian regime, as emphasized in the theoretical literature on nondemocratic politics.[3] Finally, our examination of various mechanisms furthers the understanding of the role of social media in collective action (e.g., Cantoni et al. 2019; Bursztyn et al. 2020).

Our paper also relates to the theoretical literature on protests, government policy, and regime change (e.g., Lohmann 1993; Battaglini 2017), and in particular Barberà and Jackson (2020) who argue that social media helps coordinate uprisings. Our findings suggest that accounting for endogenous censorship of social media, where the regime trades off the benefits and costs of information control (e.g., Egorov et al. 2009; Lorentzen 2014), would be an valuable extension of these models.

Finally, our paper contributes to the literature on media control strategies in nondemocratic countries. Several studies take the revealed-preference approach and infer authoritarian governments' goals from their observed strategy of social media censorship (e.g., King et al. 2013, 2014; Qin et al. 2017). We extend this line of research by examining the effect of

[3] In their survey of the theory of nondemocratic politics, Gelbach et al (2016) note: "A number of models examine the dynamics of collective action in anti-regime protests and the means by which dictators can prevent their success. An important intuition in the study of protest and revolutions concerns cascades: the possibility that a protest today spurs more protests tomorrow by revealing information about the degree of popular support for the regime."

social media on real-world events after the imposition of censorship, from which we infer that there is an unavoidable trade-off between the regime's dual goals of collecting information and suppressing dissent.

The remainder of this paper proceeds as follows. Section 2 describes the institutional background. Section 3 describes the data and provides descriptive statistics. Section 4 explains the empirical specification. Section 5 presents the main results, and Section 6 discusses their implications. Section 7 discusses censorship and mechanisms. Section 8 concludes. Details about econometric formulation, the Monte Carlo simulations used to assess potential biases, and investigation of censorship are relegated to three online appendices.

2 Background

2.1 Social Media in China

Social media in China today is as vibrant and extensive as in any Western country. Our study focuses on the early period of rapid social media expansion in China during 2009-2013. Because the Chinese government blocked Twitter and Facebook and strictly controlled domestic microblogging services, the use of social media was limited until Sina Weibo appeared in August 2009. A hybrid of Twitter and Facebook, Weibo allows users to tweet and retweet short messages with embedded pictures or videos, send private messages, and write comments. Between 2009 and 2012, Weibo use increased exponentially. By 2010, Weibo had 50 million registered users, and this number doubled in 2011, reaching a peak of over 500 million at the end of 2012 (China Internet Network Information Center 2014). Weibo adoption was faster in some areas, notably, those with high pre-existing levels of mobile phone use. During our main sample period, Weibo was the dominant microblogging platform in China. Since then, it has lost ground to WeChat, a cellphone-based social networking service, but it remains an influential platform for public communication.

The popularity of social media in China is coupled with extreme government control. Early in our sample period, China was ranked by Freedom House as among the countries with exceptionally low internet freedom, surpassed only by Iran, Myanmar, and Cuba. After 2013, media control was further tightened, and since 2015, China has been ranked as possessing the lowest internet freedom in the world.[4] Although censorship is pervasive, it is far from complete even on very sensitive topics such as protests and strikes, likely because of the Chinese government's intention to gather information from social media (Qin et al. 2017). We discuss censorship in detail in Section 7.

[4] https://freedomhouse.org/report/freedom-net

2.2 Protests and Strikes

Although a lack of opportunities to protest is a hallmark of authoritarian regimes, there have been numerous instances of collective resistance in China in the last two decades. The Chinese government's response to protests is multifaceted. Protests are often met with violent repression, and their leaders are taken into custody (Lorentzen 2017). Even if some protest demands are accommodated, organizers and active participants risk losing their jobs and being arrested or placed under close watch. At the same time, concessions are frequent; protests viewed as sufficiently innocuous are even ignored (Su and He 2010; Lee and Zhang 2013). As numerous public statements by top Chinese Communist Party (CCP) leaders make clear, the Chinese central government requires local officials to handle collective action events strategically rather than simply suppressing all such events with police force (Steinhardt 2017). For instance, in many anti-corruption protests, high-ranking CCP officials were eventually sent to converse with protesters to re-establish the public's trust in the government.

In China, strikes are often triggered by firms' violations of labor laws, such as wage arrears and illegal work conditions. Due to weak law enforcement and inadequate government intervention, strikes are often the most effective way for workers to claim their rights and express their disapproval. In this sense, strikes are similar to protests, and it is not surprising that many strikes are associated with protest activities. Government reactions to strikes range from repressive acts (e.g., police arrests) to mediation to concessions. The Chinese government does not regard strikes as politically sensitive unless they escalate into violent events that lead to social unrest.

There are several reasons for the Chinese regime's relative tolerance of protests and strikes. First, China is large and diverse, and most political and policy decisions are decentralized to local governments. Protests are a costly and hence credible way for citizens to communicate their concerns, which may help the regime identify and correct policy oversights, gauge public sentiment, and monitor local officials (Lorentzen, 2017). Second, the absolute suppression of collective action may generate distrust and undermine the legitimacy of the regime. Finally, some collective actions such as strikes may improve welfare and even enhance productivity if they result in better working conditions (O'Brien and Li 2006; Cai 2010).

As long as protests remain within a restricted scope and location, they pose little threat to the regime. The regime is only threatened when local protests evolve into larger political action and social movements that shift the focus from local to national policies and leaders, thereby diminishing the legitimacy of the regime and trust in the CCP (Cai 2008).

3 Data

We assemble a unique dataset combining detailed information on thousands of collective action events from 2006 to 2017 together with posts published on Sina Weibo from 2009 to 2013. Data on protests and strikes in China are not available from any official sources, and

media coverage of such events in mainland China is rather limited. Hence, we collect data from sources outside mainland China. Below, we explain how we collect the data and provide a description of our dataset.

3.1 Protests

The data on protests are collected from the website of Radio Free Asia (RFA), a private nonprofit broadcasting corporation based in Washington DC. We obtain relevant information from the Chinese version of the website, which is widely used by Chinese news portals outside mainland China. The news reported on the RFA website comes from RFA's correspondents, from media outlets in mainland China, Hong Kong, and Taiwan, and from Western media outlets, such as the New York Times and BBC. One advantage of using RFA as the information source is that it hires correspondents on the ground to verify the authenticity of information. To the best of our knowledge, the RFA website is the most reliable and well-structured data source for protest events in mainland China.[5] We searched for keywords related to "protest" and "demonstration" (in Chinese) on the RFA website and downloaded the relevant news reports. Several research assistants were hired to verify the news source and purge duplicate information. Next, they extracted relevant information from each report and coded the date, location, cause, and size (number of participants) of each event.

The resulting dataset contains 1,172 protests between July 2006 and December 2013, which is the focal period of our main analysis, and 1,603 protests between 2014 and 2017. We plot these protests by size and cause in Figure 1. More than four out of every five protests in our sample concern the government (government policy or corruption, police and courts, and housing and land reforms) or livelihood issues (employment, environment, and health). Although many of the events are small and confined to certain localities, some of them are large-scale and widespread. For instance, an event widely reported by Western media is the 2011 Wukan protest, when thousands of villagers in a village in Guangdong protested against the corruption of local officials. The event led to direct confrontation between the villagers and local officials, violent conflicts between protesters and police, and demonstrations in multiple cities in support of the villagers.

These protests span 290 cities, accounting for 85% of all cities in China. Figure A1 plots the number of protests across cities. Around one-fifth of all locations experience more than 10 events during our sample period, with Beijing being an outlier (234 protests).

[5] One potential data source of protests is the GDELT project, which contains information on massive events collected from the world's news media. However, we find the data related to China extremely noisy. Another data source of protests is the Mass Incident Dataset constructed by the Chinese Academy of Social Sciences (e.g., Miao et al. 2021). But this dataset relies on media outlets in mainland China and covers only events with more than 100 participants. Recently, some researchers have applied machine-learning methods to Chinese social media data to collect collective action events (e.g., Zhang and Pan 2019). However, these data lack event specifics and are difficult to verify. Moreover, this method is likely to overestimate the number of events due to repeated counting of the same event that was discussed by users in different locations.

3.2 Strikes

We collect data on strikes mostly from the China Labor Bulletin (CLB), a non-governmental organization based in Hong Kong that supports the development of trade unions in China. The CLB has collected data on strikes in mainland China since 2007. From 2011 onward, this dataset is electronically available and contains detailed information on the timing, location, employers involved, industry, scale, worker action, and government responses for each event. For earlier strikes, we extract similar information from the annual reports published by CLB and supplement them with data from Boxun, a US-based political news website in Chinese.

The CLB draws on information from overseas Chinese media outlets, labor movement activists in China, and internet searches including social media. An important source for the 2013-2016 period is Wickedonna, a mass event tracking blog that searched Weibo and other Chinese social media platforms.

Our dataset contains 1,558 strikes during 2007-2013 and 7,967 strikes during 2014-2017. As shown in Figure 1, strikes occur in a wide range of economic sectors, with a concentration in the manufacturing and transportation industries until 2013 and then a shift away from transportation and towards the construction industry. The most common cause of strikes is demands for payment of wage arrears. Geographically, these strikes cover 324 cities, approximately 95% of all cities in China. The developed coastal areas are over-represented, notably Guangdong, Jiangsu and Shandong, but a significant number of strikes occur in some inland areas. See Figure A1 for the distribution of strikes across cities.

3.3 Social Media

Our social media data come from a database including 13.2 billion Weibo posts published from 2009 to 2013. The database was created by Weibook Corp, which conducted a massive data collection project of downloading blogposts from more than 200 million active users. 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 on a weekly basis. For each post, the data provide the content, posting time, and (self-reported) user location. According to our estimates, the Weibook dataset contains approximately 95% of the total posts published on Weibo before 2012 when we have an alternative measure of the total post volume (Qin et al. 2017).

From the Weibook database, we obtain two datasets. The first one contains the aggregate information on the number of posts per city and month, based on the entire 13.2 billion posts available. We use this aggregate measure, labelled as Weibo penetration, to capture how the popularity of Weibo changes over time and across cities.

The second dataset contains posts that mention any of approximately 5,000 keywords related to various social and political topics.[6] This dataset comprises 202 million original

[6] Details on our selection of keywords and extraction of posts can be found in Qin et al. (2017).

posts and 133 million first retweets of them. We only have the direct retweets of the original posts, not the retweets of the initial retweets. For each original post and initial retweet, we obtain the text and information on the posting time, number of retweets, and author location. We use this dataset to construct a network of social media information flows across cities and measure its expansion over time. We also use posts referencing protests and strikes to examine how information about protests and strikes diffuses on social media.

Figure 2 depicts the total numbers of protests and strikes per month along with Weibo use per month. Clearly, there is a positive correlation between the incidence of protests/strikes and Weibo penetration over time. The number of strikes per month is approximately six in the 2007-2010 period and it increases sharply to over 52 in 2013. The pattern for protests is similar, with around three protests per month until 2010, followed by a rapid increase to around 54 in 2013. The green triangles indicate the number of Weibo posts per capita, which increases significantly after the start of 2010. This trend of increasing protests and strikes has not gone unnoticed. It is commented on in numerous news sources including the BBC, CNN, The New York Times, and The Washington Post.[7] Of course, the positive relationship between social media and events may not be causal. The increase in strikes could be driven by other factors, such as a slowdown in Chinese exports (Campante et al. 2019) and by increased observability of the events through social media.

3.4 Information Diffusion

Tweeting on protests and strikes As described in Qin et al. (2017), there is extensive coverage of protests and strikes on Weibo. From the dataset of original posts, we find 2.5 million posts containing keywords related to protests and 1.3 million posts containing keywords related to strikes. To characterize these posts, we inspect the content of a random sample of 1,000 posts in each category. Around 30% of the posts are indeed about protests and strikes.

These tweets appear in bursts in the cities where strikes and protests take place, starting from the day before the event, peaking at the day of the event, and remaining high the following several days. For example, the average number of posts in a city mentioning protests increases from 4.4 on an average day to 62.6 on days of protests in this city, while the number of strike tweets jumps from an average of 2.5 per day to 167.3 on strike days (Qin et al. 2017).

Retweets and information diffusion We find that the tweets mentioning protests and strikes indeed reach many other users. As mentioned before, retweets are an effective measure of information diffusion (e.g., Kwak et al. 2010). Our tweet data contains a variable measuring the total number of times each tweet is retweeted. According to this variable, the 3.8 million protest and strike posts generate 37 million retweets, amounting to an average of 10 retweets per original post. Conditional on being retweeted at least once, there are on

[7] "Can China keep its workers happy as strikes and protests rise?," Mukul Devichand, BBC, December 15, 2011; "China on strike," James Griffiths, CNN, March 30, 2016; "Strikes by Taxi Drivers Spread Across China," Andrew Jacobs, NYT, January 14, 2015; "Strikes and workers protests multiply in China, testing party authority," Simon Denyer, Washington Post, February 25, 2016.

average 50 retweets per strike post and 100 retweets per protest post.

To investigate how information in these tweets diffuses, we exploit the 3 million direct retweets of the protest and strike posts for which we have information on the time of the retweet and the user's location; see Figure A2. As shown in the left panel, for posts about protests and strikes, approximately 28% of the retweets occur within one hour of publication of the original posts, and 75% within one day. The right panel shows that retweets within the first hour are more likely to be geographically close. After that, distance plays no role: the average distance between the user who posts the original post and the user who retweets it is the same as the average distance between Weibo users. In general, information about protests and strikes on social media disperses rapidly and widely across China.

Documenting the rapid and wide spread of protest and strike information is important in its own right. Even if protests and strikes are only local, the accumulation of widespread information about them could be detrimental to the legitimacy of the regime. Nevertheless, posting and retweeting of such information is permitted.

Social media information network We use retweets across all topics, except protests and strikes, to construct a time-varying social media information network, through which information flows across cities. We use the number of posts from users in city i that are retweeted by users in city j as a proxy for the flow of information from city i to city j through the Weibo network. A post from city i being retweeted in city j implies that someone in city j has seen it and decided to retweet it. Of course, many others in city j may see the post without retweeting it. Therefore, our retweeting measure is a conservative measure of information spread.[8] In summary, the information network is built on the 133 million first retweets across all topics excluding the 3 million retweets of the protest or strike posts.

3.5 An Example

In Guangzhou, Guangdong, a male worker wearing a bomb vest staged a protest against wage arrears in a company in the afternoon of January 18, 2013. Later, he detonated the bomb, killing one person and severely injuring seven people including himself. Immediately after the tragedy, many Weibo posts, including some from direct witnesses, described and commented on the event. We identify 374 posts mentioning our protest-related keywords on that day from Guangzhou, 261 of them indeed talking about this event. These posts were first retweeted by Weibo users in nine cities closely connected to Guangzhou according to our retweet-based measure.[9] Most retweets express sympathy for the worker, condemn employers who default on wages, and decry the local government for disregarding citizens' rights and

[8] A less conservative measure is to use followers because many followers will not read each blog post. A widely cited study on Twitter (Kwak et al. 2010) probes whether the number of followers or the number of retweets is a better measure of influence and settles on retweets.

[9] Controlling for city and time fixed effects, the ranking percentile of the retweets between these cities and Guangzhou are 1.19% (Shanwei), 2.3% (Shenzhen), 4.36% (Wuhan), 5.03% (Shanghai), 5.3% (Hangzhou), 5.77% (Chengdu), 6.25% (Xi'an), 6.98% (Zhengzhou), and 8.62% (Qingdao).

condoning wage arrears.

Among the nine cities that first retweeted posts about the event, three experienced protests in the subsequent two days. In Shanghai, thousands of workers from a Sino-Japan joint venture protested the unfair new labor rules, detaining 18 senior managers, and the situation lasted until policemen stormed the factory two days later. In Shenzhen, hundreds of people took to the street, protesting the construction of a heavily polluting LCD factory in their neighborhood. In Shanwei, thousands of villagers protested in front of the city government, demanding the return of lands taken by the government without appropriate compensation. These protests occurred at the same time in different provinces. Although targeting different causes, they all involved people who had suffered enough and protested against perceived injustices. It is possible that the later protests were inspired by the earlier ones, but there is no direct evidence of explicit organization or coordination across events.

Our empirical task is to investigate whether such episodes are isolated random incidents or part of a large-scale systematic pattern. Specifically, do strikes and protests spread more after the expansion of Weibo use and across cities that are closely connected through Weibo?

4 Specification

We analyze whether information diffusion on Sina Weibo affects the spread of protests and strikes across Chinese cities using a panel of N cities at daily frequency, t. Because the specification is the same for both types of events, we use protests for illustration. Let y_{it} be a dummy variable indicating the occurrence of a protest. Suppose that the probability of a protest in city i on day t, $\Pr(y_{it})$, depends on the number of people who are informed about protests y_{jt-1} in another city j at time $t-1$. Let f_{ijt} be the number of people in city i who typically read tweets posted by users in city j. On days when there is no protest in city j, there is nothing to learn. Thus, the number of people in city i who learn about protests in city j can be represented by $f_{ijt}y_{jt-1}$. Under the assumption of linearly additive effects, the reading of protest tweets (and hence the potential spread of events to city i) is increasing in

$$\sum_{i \neq j} f_{ijt} y_{jt-1}.$$

This is a model of event spread through the social media network, as measured by f_{ijt}.

We use two measures of social media connections between city pairs: one time-varying and the other time-constant. The time-varying measure, f_{ijt}, is the logarithm of 1 plus the number of retweets by users in city i of posts from users in city j on all subjects except protests and strikes in the prior six months up to one week before day t.[10] We normalize

[10] Using cumulative retweets in a 4-month window produces qualitatively similar results. Although we have millions of retweets, some averaging over time is necessary because we estimate over 90,000 city-pair connections. A longer time window increases accuracy but reduces the time variation in connectedness across cities.

the f_{ijt} matrix so that the average sum of all elements in a row of a weighting matrix is one. Then, we define

$$s_{it-1} = \sum_{i \neq j} f_{ijt} y_{jt-1}$$

to capture the time-varying diffusion of information on protests through social media.

The time-invariant measure, f_{ij}, captures the average social media connectedness (f_{ijt}) between cities i and j in the post-Weibo period. Then, our measure of protest diffusion through social media is:

$$\overline{s}_{it-1} = \sum_{i \neq j} f_{ij} y_{jt-1}.$$

The f_{ij} matrix is normalized so that a marginal change in $\overline{s}_{it-1}$ is associated with a marginal change in s_{it-1} of the same size. Therefore, the estimated coefficients of the above two measures are comparable with each other.

As we later show, these two measures exploit different variations that lead to two complementary econometric models. The benefit of the time-varying measure is that it provides an accurate measure of the actual social media connections between cities at each point in time. Thus, it uses the exact timing of the expanding social media network to identify effects. In contrast, the time-invariant measure allows us to better spell out the identification assumptions and investigate the spread due to the pre-Weibo correlated and contextual effects between cities that eventually become connected in the post-Weibo period. An additional advantage is that we can examine the effect of social media after 2013, when Weibo data are not available but a large number of events are added to our data, under the reasonable assumption that the average number of retweets between cities during 2009-2013 is a good proxy for the average number of retweets during 2014-2017.[11]

4.1 Time-Varying Measure of Connections

When using the time-varying measure of social media connections, f_{ijt}, we estimate the model

$$y_{it} = \alpha y_{it-1} + \beta s_{it-1} + \gamma d_{it-1} + \theta_0 w_{it} + \theta' x_{it} + \delta_i + \delta_t + \varepsilon_{it}, \tag{1}$$

where y_{it} is a binary event dummy and s_{it-1} is defined as above with y_{jt-1} being the number of events in city j within two days before t. The variable d_{it-1} captures the spread to geographically close cites and is defined as

$$d_{it-1} = \sum_{i \neq j} d_{ij} y_{jt-1},$$

[11] Weibo passed its rapid expansion phase in 2012 and stabilized in 2013. We show that a regression of the number of retweets between cities during the past six months before the last day of 2013 on the same variable before the last day of 2011 yields an R^2 of 0.96.

where d_{ij} is the inverse of the geographic distance between cities i and j. We include Weibo penetration, w_{it}, because social media may directly affect the incidence of protests. Practically, w_{it} is the logarithm of 1 plus the number of Weibo posts, excluding protests- and strike-related posts, per capita based on the aforementioned 13.2 billion posts in the Weibook dataset. Note that w_{it} is measured by the total number of Weibo posts, most of which concern nonsensitive personal communication, and thus this number is only marginally affected by posting on protests and strikes or by censorship. We add another set of controls, x_{it}, including the total number of retweets by users in city i ($\sum_{i \neq j} f_{ijt}$), the logarithm of population, GDP, the shares of the industrial and tertiary sectors, and the number of cell-phone users.[12] We also include an auto-regressive term y_{it-1} as well as time and city fixed effects, δ_t and δ_i, respectively. We estimate this model for the period up to the end of 2013, during which the measure f_{ijt} is available. In a robustness specification, we allow for an arbitrary time-invariant heterogeneity c_{ij} in event spread across city pairs by adding the following set of interaction terms

$$\sum_{i \neq j} c_{ij} y_{jt-1}.$$

These variables control for any time-invariant spurious correlations between y_{it} and y_{jt-1} (e.g., due to correlated or contextual effects). However, adding a large set of $N(N-1)$ controls in a very demanding specification may remove some useful information.

4.2 Time-Constant Connections

When using our time-constant measure of connections, we divide the data into three periods: period 0 is the pre-Weibo years (2006-2009); period 1 is the first post-Weibo period (2010-2013) for which we have Weibo data, and period 2 is the second post-Weibo period (2014-2017) for which we do not have Weibo data. We estimate the following model

$$y_{it} = \alpha^p y_{it-1} + \beta^p \overline{s}_{it-1} + \gamma^p d_{it-1} + \theta_0 w_{it} + \theta' x_{it} + \delta_{ip} + \delta_t + \varepsilon_{it}, \tag{2}$$

where the variables are defined as above. What is new is that we allow coefficients on y_{it-1}, $\overline{s}_{it-1}$, d_{it-1}, and the intercept to differ across the three periods, as indicated by the superscript p. As shown in Appendix A.1, the above equation can be derived as a generalization of a fully saturated triple-difference estimator, in which the treatment is (lagged) event shocks in informationally connected cities in the post-Weibo period.

Now, we discuss the identification assumptions; see Appendix A.1 for more details. Let $y_{t-1} = (y_{1t-1}, y_{2t-1}, ..., y_{Nt-1})$ be the $N \times 1$ vector of lagged y's. We define F and D to be $N \times N$ matrices with elements f_{ij} and d_{ij}, respectively. We assume that

$$E\left[\varepsilon_{it} \mid y_{t-1}, D, F, \delta_{ip}, \delta_t, w_{it}, x_{it}\right] = 0. \tag{3}$$

[12] These data on city-level characteristics are obtained from the Chinese City Statistics Yearbooks.

Hence, the F matrix of social media connections is assumed to be conditionally exogenous, after conditioning on the other included variables, notably, the city-by-period fixed effects. This specification allows for the probability of a protest in city i to be higher in more connected cities in the pre-Weibo period (δ_{i0}) and differentially so in the post-Weibo period (δ_{i1}). By including these terms, we allow the network to be endogenous to the average protest probability in city i in each period. For example, cities that are more central in the network can be more prone to protests.

Note that spurious correlations may be caused by unobserved shocks that are correlated with the network and that arise from endogenous sorting or a common environment. This is the problem of *correlated effects*. The correlations may also be driven by the characteristics of connected cities, which is the problem of *contextual effects*. To address these issues, we allow protests in city i to be more likely after a protest occurs in an eventually connected city j with or without Weibo. Specifically, we measure the correlated and contextual effects in the pre-Weibo period by the estimate of β^0 and partial them out when estimating the network effects in the post-Weibo period.

The parameter of interest is $\beta = \beta^1 - \beta^0$, which captures the effect of the social media network on the spread of events. It is positive if the spread of protests increases more across the connected cities than across the nonconnected cities in the post-Weibo period. The identification of this parameter requires a parallel-trend assumption: the difference in the outcome on days with and without a protest in city j would have had the same trend across the pre-Weibo and post-Weibo periods in the connected and nonconnected cities, had it not been for the expansion of the social media network. Such an identifying assumption cannot be tested directly, but we can examine the pre-trends by separately estimating coefficients β^b on $\overline{s}_{it-1}$ at a biannual frequency (indexed by b) in the pre-Weibo period.

4.3 Econometric Issues

In addition to the aforementioned correlated and contextual effects, several econometric challenges remain. First, a *reflection* (or *simultaneity*) problem would appear if events in city i simultaneously affect events in city j, which in turn affect events in city i, and so on and so forth in an infinite loop. This problem is more severe in the typical cross-sectional network analysis, where all events are simultaneous. It is less severe in our panel data setting, where some events are pre-determined. In addition, we can measure protest shocks at a daily frequency and assign the temporal ordering of events. It is possible that we sometimes measure the event date with error and hence assign an incorrect temporal ordering. This measurement error would likely to lead to attenuation bias. By design, our specification does not capture the spread of events within the same day. Event spread within the same day is likely to be limited, because it takes time to organize a protest. To the extent that this happens, our estimates will capture the total "reduced-form" effect, including any potential spread within the same day.

Second, logistic and probit models are prone to bias with rare events data (King and Zeng 2001) and do not work well for panel data with a large set of fixed effects. Thus, we estimate a linear probability model, which is immune to these problems.[13] Third, we check whether the estimated process is stationary, as discussed in Appendix A.2.1. Fourth, consistency requires no serial autocorrelation of errors. We test for serial autocorrelation in the first-differenced residuals. Fifth, the error term in (1) may be correlated across both time and spatial units. We use two-way clustering of errors in the temporal and spatial dimensions to address this.[14]

Finally, our model includes location fixed effects and lagged dependent variables. In general, the estimates in this type of model are inconsistent, resulting in a so-called "Nickell bias" when T is fixed (e.g., Nickell 1981; Arrelano 2003; Baltagi 2008; Hsiao 2014). In our model, T is large, and the bias is likely to be small. Using Monte Carlo simulations, we show that the bias is indeed negligible in Appendix A.2.2.[15] Another potential issue for the specification allowing for time-invariant heterogeneity is that the autocorrelation test that we use may not be well suited; see Appendix A.2.2 for a detailed discussion.

5 Main Results

5.1 Time-Varying Measure of Connections

We estimate Equation (1) using the sample period until the end of 2013 (i.e., periods 0 and 1 in our three-period division). Table 1 reports the regression results: the first three columns for protests and the last three for strikes. Table A1 reports the summary statistics of the key variables.

The estimated coefficient β is positive and statistically significant across specifications. The magnitude is barely affected by the inclusion of controls. Columns (3) and (6) allow for arbitrary time-invariant heterogeneity in event spread across cities as described above. This increases the estimated coefficient by 0.04 for protests and decreases it by 0.013 for strikes, but neither of these changes is statistically significant.

The estimate in Column (2) implies that a protest in a given city in the last two days increases the expected number of cities with protest incidence by 0.17, whereas the corresponding number for strikes, based on the estimate in Column (5), is 0.12. Relative to the mean event probability, the estimates in Table 1 imply increases of 34% for protests and 21%

[13] The outcomes from probit models are reported in Appendix Table A7.

[14] A common approach in a cross-sectional network analysis model is to compute standard errors corrected for spatial correlation using the Conley (1999) adjustment, adapted to the network structure. Studies using panel data with time-invariant networks, such as König et al. (2017), typically use the Conley approach to correct for spatial correlation within each time unit. Our approach is more conservative than this because we allow for arbitrary spatial correlations within each unit of time.

[15] If the bias were large, one could try to address this issue by adapting the generalized method of moments (GMM) estimators proposed in Arrelano and Bond (1991) and Blundell and Bond (1998) to our setting with a dynamic spatial panel. However, instrumenting rare events in a setting like ours with lagged differences and levels is likely to perform poorly.

for strikes.[16]

The estimate of γ (the effect of geographical proximity) is positive and statistically significant for strikes, but not for protests. This suggests that strikes spread to nearby cities but protests do not. The magnitude of the spread effect through geographical connections is similar to that driven through social media. Note that the coefficients of both geographic spread and within-city spread are not identified in Columns (3) and (6), because the additional controls absorb all time-invariant spread across locations. The total number of retweets and Weibo penetration are both positively correlated with event incidence.

5.2 Time-Constant Measure of Connections

With the time-constant measure of social media connections, we extend the sample duration to 2017, including all three periods. Specifically, we estimate Equation (2) in a panel of cities with at least one protest or strike for the respective outcomes.

Table 2 presents the results. Columns (1) and (2) report the OLS estimates for protests, and Columns (4) and (5) report those for strikes. In the pre-Weibo period, neither protests nor strikes spread significantly across cities that eventually become closely connected through Weibo. In the period of rapid Weibo-expansion (period 1), both protests and strikes start to spread across cities with strong Weibo connections. The average effect on event spread through Weibo in period 1, $\beta^1 - \beta^0$, is approximately 0.14 for protests and 0.10 for strikes. That is, an event in a given city in the last two days increases the expected number of cities with protests and strikes by 14% and 10%, respectively. These estimated effects are comparable to, although slightly smaller than, the corresponding effects of β in Table 1. Relative to the mean event probabilities, the estimates in Table 2 imply an increase of 28% for protests and 18% for strikes.[17] In period 2, the estimated spread of protests falls to a marginally significant level, whereas for strikes, it remains roughly the same and highly significant, as shown by the F-tests towards the bottom of Table 2.

The estimated β-coefficients seem unrelated to the number of events in the post-Weibo period. Although there are six times as many strike events in period 2 than in period 1 (7,857 compared with 1,365), the estimated effect on strike spread in period 2 is even slightly smaller than that in period 1. Similarly, although there are 50% more protests in period 2 than in period 1 (1,517 compared with 1,046), the estimated spread in period 2 is 50% lower than in period 1. The reduced spread of protests over social media in period 2 is likely to be related to the increased strictness of media control in China, as we discuss later.

[16] As shown in Table A1, the mean of protest incidence is 0.002, and there are 247 cities other than the one where the first protest takes place. Thus, the effect is 0.17/(247*.002)=0.344. For strikes, the corresponding number is 0.12/(281*0.002)=0.214.

[17] When we discuss the coefficients concerning period 0 and period 1, the mean and number of cities are calculated based on the same sample as the one in Section 5.1 (using the time-varying measure). Thus, the effect for protests is 0.14/(247*0.002)=0.283. For strikes, the corresponding number is 0.10/(281*0.002)=0.178.

Pre-trends and dynamic effects We investigate the presence of pre-trends by separately estimating coefficients β on $\overline{s}_{it-1}$ at a biannual frequency in the pre-Weibo and post-Weibo periods. The results are depicted in Figure 3. Neither the estimates regarding protests nor those regarding strikes exhibit pre-trends.

The estimated spread of events through social media increases rapidly during 2010-2011, when the use of Weibo took off (recall Figure 2). For protests, the estimated spread effect drops significantly in 2012, bounces back to its highest level in 2013, and then trends downward after 2014. This pattern coincides with changes in political sensitivity and the strictness of censorship in China. The political situation and media control was relatively stable over the period of 2006-2012. In late 2012, protests became politically sensitive when Xi Jinping replaced Hu Jintao as the new government and party leader. From 2014, media control became increasingly strict, as reflected in China's fall in the Media Freedom Index published by Freedom House (see Figure A3). For strikes, the estimated effect increases sharply until 2013 and then remains at roughly the same level. A likely reason is that the Chinese central government does not view the spread of strikes as regime-threatening (e.g., Kuruvilla and Zhang 2016; China Labor Bulletin 2018).

5.3 Robustness

We perform a number of robustness checks. The first two are specific to the model using the time-constant measure of Weibo connections. One involves instrumenting the time-constant connections by student flows across cities. The other shows that the lower frequency of events in the pre-Weibo (vs. post-Weibo) period does not mechanically produce non-significant estimates during that period. The remaining robustness checks apply to both models using the time-varying and the time-constant measures of Weibo connections. In particular, we evaluate the impact of event observability on our estimation, the sensitivity of functional forms, and the robustness of estimates when additional controls are added.

5.3.1 Student-Mobility Instrumental Variable

A potential concern with regard to the constant measure of the Weibo network is that the F matrix is not exogenous, even when we extensively control for fixed effects and other variables. One possible issue is that the strength of Weibo connections may be significantly influenced by tweets about protests and strikes, which in turn are affected by the events. However, this seems unlikely, given that our information network is constructed based on the overall number of tweets on topics excluding protests and strikes, and that the tweets about protests and strikes comprise a very small share of all tweets.

To further alleviate the endogeneity concern, we conduct an IV analysis, exploiting an arguably exogenous variation in social media communication across cities caused by student mobility across cities before Weibo entry. College students constitute a large proportion of the early adopters of Weibo, as they could use it to maintain communication with their

hometown friends and relatives.[18] Hence, the flows of college students from their home cities to their college cities may drive information flows across cities on Weibo.

We collect data on student flows from their home (origin) to college (destination) cities in 2005 and 2009, based on student registration information at the city level.[19] Chinese students typically spend four years in college, and many stay in their college cities after graduation. Others go back to their home cities but maintain contact via Weibo with their friends who remain in their college cities. Thus, student mobility data in 2005 may predict social media communication four years later when Weibo launched in 2009.

We find that the student flows across origin-destination city pairs correlate strongly with Weibo connections between them. Let $student_{ij}$ be the log of 1 plus the mean number of students in home city i who start college in city j in 2005 and 2009. We regress social media connections, f_{ij}, on student flows, $student_{ij}$, controlling for city fixed effects, province-pair fixed effects, and destination-city by origin-province fixed effects.[20] The estimated coefficient on $student_{ij}$ is 0.083 with a standard error of 0.013 (see Appendix Table A2).

Next, we construct the variable

$$z_{it-1} = \sum_{i \neq j} student_{ij} y_{jt-1}.$$

This variable is used to instrument $\overline{s}_{it-1}$ in Equation 2. Let Z be the matrix of student flows across city pairs. The identifying assumption in the IV-estimation can be expressed by replacing the matrix of social media connections, F, with Z in equation (3). Hence, the Z matrix is assumed to be conditionally exogenous, after conditioning on the other included variables, notably, the city-by-period fixed effects. This specification permits the probability of a protest in city i to be higher in cities with larger student flows, or any other node-specific network-related statistics, in the pre-Weibo period and differentially so in the post-Weibo period. Similar to the discussion of equation (3), we also allow for differential spread across city pairs with different intensity of student flows. This flexibility is important because city pairs with large student flows probably also have strong information flows through communication channels other than Weibo. In both the OLS and IV specifications, the contextual and correlated effects are measured in the pre-Weibo period and partialed out in the estimation of the Weibo effect. Consequently, we are able to identify the effect of social media through the

[18] The launch of Weibo occurred immediately after the Chinese government banned Twitter and Facebook and shut down domestic service providers. The users of these terminated forms of social media were mostly "technology geeks" and college students, who later comprised the majority of early Weibo users. In a dataset containing Weibo posts referencing "vaccine" with detailed user information, we find that, in 2019 and 2010, the age of nearly 70% of users is below 25.

[19] The student registration information is from an administrative dataset covering students who took the Chinese national college extrance exam in 2005 and 2009. The data are maintained by CIEFR at Peking University and have been used in several studies (e.g., Graff Zivin et al. 2020). We thank Ruixue Jia and Hongbin Li for sharing the data.

[20] We control for destination-city by origin-province fixed effects because some universities set quotas for origin provinces.

changes in the F-matrix instrumented by the Z-matrix before and after Weibo was launched.

The coefficients in the first-stage regression are significant in all three periods and are stable across periods, as shown in Appendix Table A3. The IV estimates are reported in Columns (3) and (6) of Table 2. The estimated social media effects on event spread are generally similar to those obtained from the OLS regressions. A caveat is that the IV estimate is smaller in period 1 for strikes and in period 2 for protests than its OLS counterparts. This implies that, in the IV estimation, the difference $\beta^1 - \beta^0$ is no longer significantly different from zero for strikes, while it is significantly different from zero for protests. The dynamic effects, estimated using the student-mobility IV, are presented in Figure A6. The estimates are similar to their OLS counterparts in Figure 3, although the standard errors increase modestly, and the marginal effect appears slightly lower for protests during 2010-2012.

5.3.2 Mechanical Zero Effect in the Pre-Weibo Period

As noted above, the estimated effect of Weibo on event spread is not closely related to the number of events in different post-Weibo periods. Concerns may remain that the smaller number of events in the pre-Weibo period mechanically reduces the estimated effects.

We investigate this concern by simulating events with a process that matches the observed event frequencies within each period, while keeping the underlying propagation of event waves across cities equally strong across periods. The details are provided in Appendix A.2.3. In brief, in each of the 100 simulated data sets, we estimate β^0 and β^1 using the specification in Equation (2). We find that there is no mechanical bias in the estimated β^0 coefficient, and we have power to test the hypothesis $\beta^1 > \beta^0$ (see Appendix Table A4). Given the described data-generating process, we conclude that at the observed event frequency there is no mechanical zero effect. Of course, one could imagine some other distinct features in the pre-Weibo data-generating process that would depress all the coefficients in Equation (2). However, the geographical spread of strikes, captured by the estimated γ-coefficients in Table 2, is equally large in the pre-Weibo period and during 2010-2013, and then falls during 2014-2017. Hence, the coefficients in the data-generating process do not seem to be uniformly depressed in the pre-Weibo period.

5.3.3 Observability

Social media can be used by our data sources to observe and record protest and strike events. Thus, growing Weibo use in a city may increase the probability that an event in that city is observed, causing a correlation between Weibo penetration, w_{it}, and the number of reported events, y_{it}, even in the absence of causal effects. Similarly, an increase in observability may increase the frequencies of reported events y_{it} and y_{jt-1} (and hence s_{it-1}), but it is less clear whether observability will induce a spurious correlation between y_{it} and s_{it-1}.

We examine this issue using Monte Carlo simulations. Specifically, we investigate whether we can consistently test the null hypothesis that $\beta = 0$ in Equation 1, allowing for a positive

correlation between event observability and Weibo penetration. To this end, we first simulate the event data, $\widetilde{y}_{it}$, under the null hypothesis that $\beta = 0$. Next, we assume that the probability of a simulated event ($\widetilde{y}_{it}$) being reported is strongly and linearly increasing in w_{it}. Then, we draw a set of observed events y_{it} with probability p_{it} from the simulated events $\widetilde{y}_{it}$.[21] Finally, we estimate the model in Equation 1 on the observed simulated events, y_{it}.

As shown in Figure A7, observability induced by social media does not generate any spurious event-spread effect. Hence, we conclude that we can consistently test the hypothesis that $\beta = 0$, even if event observability increases in the use of social media.

5.3.4 Functional Form

We assume that s_{it-1} and $\overline{s}_{it-1}$ enter Equations 1 and 2 linearly. Appendix Tables A5 and A6 report the results from estimating a logarithmic model, in which s_{it-1} is replaced by the transformation $ln(5s_{it-1}+1)$, and so is $\overline{s}_{it-1}$. As a further robustness check, Table A7 shows the results from a probit regression of Equation 1, where the date fixed effects are replaced by a quadratic time trend to avoid the problem of incidental parameters. The estimates are statistically significant across all specifications, and the implied marginal effects on event probabilities are similar to those estimated from the linear model.

5.3.5 Additional Controls

We create additional control variables to capture the changing spread of events related to other city-pair characteristics in the following form:

$$c_p \sum_{j \neq i} \omega_{ij} y_{jt-1},$$

where c_p is a period indicator dummy variable and $\omega_{ij} = \omega_i \omega_j$ is constructed for each of our control variables ω_{ii} in 2008. For example, when ω_{ij} refers to population, it is the product of the log population in city i in 2008 and the log population in city j in 2008. We interact these variables with period-fixed effects and include them in the specifications underlying Table 2. The results are reported in Appendix Table A8. The estimates are not significantly affected by these additional controls. The only exception is the β^1 coefficient in the estimation regarding strikes, which falls, after we include the number of cell-phone users, rendering $\beta^1 - \beta^0$ non-significant. The fall in the estimated β^1 could occur because the spread of strikes between cities with high-cell phone coverage increases exactly at the time of Weibo expansion but for reasons unrelated to social media. However, we suspect that it

[21] To assess the size of this increase in observability, we use the estimated coefficient $\widehat{\beta}_0$ on Weibo posts, w_{it}, in Table 1. This coefficient is likely to capture both increased observability and a presumably positive causal effect of w_{it} on y_{it}. Hence, $\widehat{\beta}_0$ is an upper bound on the response in observability due to w_{it}. The linear increase in observability due to Weibo penetration in our simulations is set so that it exceeds this upper bound. More precisely, p_{it} increases by 30% as w_{it} increases from zero to its maximum value.

is caused by over-controlling one channel through which social media operates because cell phones are the main device used by Chinese people to access Weibo.

To summarize, the Weibo effect on strike spread in period 1 is sensitive to the control for cell-phone use but non-significant when we instrument Weibo connections with student mobility across cities. Hence, our conclusion regarding this estimate is more guarded.

6 Implications: Speed, Scope and Size

In China, protests confined to a small number of cities and with a narrow scope have limited implications for national policies and regime changes. By establishing strong information links between distant cities, social media has the potential to help break the boundary of event spread. Therefore, it is natural to investigate whether social media enables individual protests to evolve into widespread movements that affect a large number of people in many cities and from diverse social backgrounds. To this end, we examine how social media affects the speed and duration of event spread as well as the spread of protests across various causes and strikes across multiple industries. Next, we estimate the effect of social media on the magnitude of protest waves, paying particular attention to the likelihood of very large event waves manifested by simultaneous events across many cities that influence a substantial share of China's population.

6.1 Speed and Duration

So far, we have analyzed the short-term (within two days) response to an event shock mediated via information flows on social media. To examine how the effect persists over time, we estimate the same model with the time window extended to various durations. Figure 4 depicts the estimated spread effect of events occurring within 1-2 days, 3-7 days, 8-30 days, 31-90 days, and 91-180 days. For both protests and strikes, the effects are greatest in the 2-day window, then drop drastically, and become small and non-significant after 3-7 days. Thus, social media rapidly spreads events but the effect dies out quickly.

The rapid but short-lived response to social media communication is consistent with the notion of "the window of opportunities" stressed in the studies of protests in China by sociologists (Zhou 1993) and political scientists (Weiss 2013; Truex 2019). The basic idea is that there are certain narrow windows during which protests are allowed without substantial political risks. Tweets about protests in other cities may signal that such a window has opened, which encourages users in other cities to grasp the opportunity to protest before it closes. The short-lived response may also arise if protesters are motivated by spontaneous emotions upon seeing tweets about protests and strikes (e.g., Jasper 2008; Passarelli and Tabellini 2017). In our dataset, the abnormal bursts of tweets about protests and strikes last for approximately two days, similar to the duration of the significant social media effects.

6.2 Scope

As previously discussed, events that only spread within the same cause and industry are bounded in size and influence and are unlikely to be perceived as regime-threatening. It seems likely that strikes predominantly spread within industries and that protests predominantly spread within causes. For example, the waves of school teacher strikes in 2014-2015 spread within the education sector (Chang and Hess 2018). The protests against corruption that originated from Wukan and spread among farmers were for the same cause (recall Section 3.1). However, the example presented in Section 3.5 suggests that strikes and protests may also spread across industries and causes.

To study whether social media increases the scope of events, we separately estimate the spread effect induced by social media within and across the protest causes and strike industries that are listed in Figure 1. Table 3 shows that the estimated spread effect through social media is six to eight times higher within the same category (protest cause and strike industry) than its cross-category counterpart. This result confirms that the spread of events through social media predominantly occurs within categories.

Nevertheless, the cross-category spread induced by social media is statistically significant. Although the effect of an individual event is smaller across categories than within categories, the total effect of spread across categories is larger because there are many more events across all categories. The means of the Weibo-weighted events within and across categories are reported in the last two rows of Table 3. The aggregate effect of protests spreading across protest causes is 60% larger than that within causes. Similarly, the aggregate effect across strike industries is 20% larger than that within industries. Thus, social media helps break down the boundaries of protest causes and strike industries, leading to a greater overall impact on event waves than through within-category spread.

6.3 Size and Probability of Large Event Waves

China's big cities lie far apart, and each contains only a small fraction of the country's population. Therefore, protests in one or a few cities are easy to contain and barely have any national implications. Our setting provides a unique opportunity to study the impact of social media on event waves, measured by the number of essentially simultaneous events across multiple cities. We are interested in estimating the mean size of event waves as well as the likelihood of very large event waves.

This inquiry requires additional investigation of the dynamic process of event waves. A key question is whether the number of events will continue to grow at a constant rate even when the waves increase in size, as implied by the model in Equation (1) which is linear in s_{it-1}. We examine this linearity assumption using a nonparametric least squares regression (Cattaneo et al., 2022). Figure 5 plots the nonparametric conditional mean function (the dots), together with a linear, a logarithmic, and a 5th-order polynomial approximation, using the specification in Equation 1.

The conditional mean function is approximately linear for most of the support. This can be seen more clearly in the lower panel, which zooms in on observations with s_{it-1} less than 0.4 (the vertical line marks the 99th percentile in the distribution of s_{it-1}). As all three approximation functions are roughly linear in this range, the estimated average marginal effects they produce are similar to that from the linear model.

However, as an event wave grows, the marginal spread effect of an additional event (the slope of the curve) falls. For sufficiently high values of s_{it-1}, there is no spread of events through social media. This curbs the size of protest and strike waves. There are several possible reasons for this collapse of large event waves. First, censorship and repressive actions tend to increase when a protest wave intensifies. Second, there exists a natural limit to the number of cities susceptible to a particular protest wave. For example, a protest against improper conversions of agricultural land to commercial land cannot spread further when most of the susceptible cities have already protested.

We examine the impact of social media on the size of protest waves by simulating events under two scenarios: with and without social media. In the scenario with social media, we allow protests to spread through Weibo using the estimated 5th-order polynomial approximation of the conditional mean function. In the scenario without social media, this term is dropped from the data-generating process. Other aspects of the data-generating process remain the same in both scenarios.[22] We simulate the data 1,000 times, obtaining 1,000 possible histories of protest events in our city panel up to 2013. For each simulation and year, we compute two measures of the largest event wave: the maximum number of cities experiencing protests within the same week and the corresponding maximum share of the population in the cities experiencing protests.

Figure 6 shows the yearly average of these two variables in the scenarios with and without social media, as well as in the real-world data. The mean of the simulated data with social media matches the real-world data fairly well. In the pre-Weibo period, on average, the largest protest waves affect cities comprising only 4% of the population, which approximately equals the sum of the population of the two largest cities in China. This is consistent with the fact that, before the advent of social media, no source informed large audiences of protests and strikes in China. Hence, limited information flows restrict event spreads.

In the post-Weibo period, we estimate that the size of event waves would increase even without social media spread, but the increase would have been substantially smaller. In 2013, without social media, the largest event waves affect only 11 cities comprising 6% of the population. With social media, the largest event waves affect 28 cities comprising around 15% of the national population. Apparently, social media connect these large cities, enabling protests to spread rapidly across them.

Furthermore, the probability of very large event waves increases dramatically with social media. Consider 2013 again. With social media, there is a 5% probability that an event

[22] For details about the data-generating process, see Appendix A.2.4.

wave affects 40 cities or more, containing 22% or more of China's population. Without social media, the corresponding numbers are 15 cities containing 8% of the population. The reason for this large difference is that the dynamic feedback induced by social media increases not only the mean but also the variance of the protest outcomes, thereby substantially increasing the likelihood of very large protest waves.[23]

7 Censorship and Mechanisms

We have shown that, even in a strictly controlled information environment, social media has a considerable effect on the spread of protests and strikes. This appears puzzling because content that is typically deemed crucial for social media to affect protests is likely to be censored in China. To better understand what potential mechanisms may drive the estimated social media effects, we need a clear understanding of how censorship is performed in China and what content is visible to social media users. In this section, after a brief description of censorship in China, we provide a measure of local censoring intensity and show that it does not significantly bias our estimated effects on event spread. Then, we analyze whether we can use the tweets in our downloaded data to infer what content is visible to users, given that some tweets are censored before we download them. Finally, we discuss what mechanisms may be at play given the available content.

7.1 Background

Censorship of Weibo tweets is conducted in two ways (King et al., 2014). First, tweets that contain sensitive keywords are automatically held in quarantine for review by censors and are then released or deleted. Second, tweets without sensitive keywords are immediately published but later censored. Researchers have studied the latter type of censorship by recording whether posted tweets are later deleted. In the daily operation of Weibo, the vast majority of posts are on nonsensitive topics, and only a tiny share of regular posts are censored. For example, Fu et al. (2013) estimate that only 0.01% of the posts published by a large sample of VIP Weibo users are censored. When tweets involve sensitive topics such as collective action, the share of censored posts can be as high as 13% (King et al. 2013).

[23] To see this, consider the simple first-order serial autoregressive process

$$y_t = \rho y_{t-1} + \varepsilon_t.$$

Let $m = \frac{1}{1-\rho}$. Then, the mean and variance of the process are $\overline{y} = m\overline{\varepsilon}$ and $var\,(y) = m^2 var\,(\overline{\varepsilon})$, respectively. Therefore, an increase in the autocorrelation coefficient, ρ, increases both the mean and variance of y_t. The increase in variance implies an increase in the probability of very high realizations of y_t relative to its mean. Our setting is more complex than this simple example, but the basic mechanism remains the same.

7.2 Local Censoring Intensity

In 2015, we went back and recorded which posts in our dataset were still online. We choose a subset of 156,553 posts on sensitive topics from our dataset, including 10,000 in each of the protest and strike categories. The deletion rate is 27%. We do not know for certain which posts were censored, and which were deleted by users themselves. However, as Figure A8 shows, our regional deletion rates correlate strongly with other regional measures of media control in China, such as the measure of social media censorship in Bamman et al. (2012) and the measure of pro-government media bias in Qin et al. (2018).

In Appendix A.3.1, we examine whether our measure of social media connections is endogenous to censorship, by regressing f_{ij} on the share of deleted posts in city i and city j. We find no evidence of such endogeneity. This is hardly surprising, because our construction of f_{ij} is based on tweets across all subjects excluding protests and strikes, and only a negligible fraction of these are censored. Hence, we find no evidence that censorship biases our estimated effect on event spread through f_{ij}.

Moreover, we investigate whether the spread effects are smaller in areas with higher censorship intensity. We include the interaction between s_{it-1} and the share of deleted posts in city i in the regression Equation (1). We find that the Weibo effect on event spread does not vary significantly across regions with different levels of censorship. A possible reason is that censorship in a city does not affect the reading of incoming tweets from other cities. The effect of censoring outgoing tweets is difficult to identify, as discussed in A.3.1.

7.3 Censoring, Downloads, and User Exposure

In section 7.4, we will evaluate possible mechanisms that may explain our large and systematic effects of social media on event spread using a simple necessary condition: for a certain mechanism to play a significant role, the content supporting this mechanism must be highly visible. For example, if information on where and when to meet on social media explain our findings of large and systematic effects, then tweets with this logistics content must circulate widely. Conversely, if this type of content is read by very few users, we can rule out the associated mechanism as the main explanation of our empirical findings. The assessment of content visibility is more complicated in the presence of censorship than in a free media environment because some content may have circulated widely before being censored. A particular issue is that our dataset may miss some posts that were first published and then censored if our downloading was slower than censoring.

With these complications, the key question is whether we can draw conclusions about the visibility of specific content based on our downloaded tweets. The answer is yes for tweets that are never censored or are automatically quarantined before further action, because, in this case, the downloaded posts are precisely those that are visible to users. For posts that are first published and later censored, the answer depends on the share of tweets read by

users before they are censored, as well as the share of tweets that are downloaded before censorship, as illustrated below.

Consider two types of content referencing protests, labelled A and B. Type-A content reveals logistics information (e.g., where and when to meet). Type-B content just talks about the occurrence of a protest or expresses sentiment. Because Type-A content imposes a greater political risk to the government than Type-B content, the Chinese government censors Type-A content rapidly but not Type-B content. Given that data downloading is slower than user reading of tweets, users may read Type-A tweets that are censored before we download them. This is not the case for Type-B tweets, because we eventually download all of them. Hence, the ratio of users who have read type-A tweets to those who have read Type-B tweets will be larger than the corresponding shares of downloaded tweets. Consequently, the ratio in the downloaded tweets may be a biased measure of the ratio of the share of tweets read by users (user exposure for short).

The magnitude of this bias depends on the speed of (i) censoring, (ii) user reading, and (iii) tweet downloading. Below, we briefly describe how we measure it, leaving details to Appendix A.3. Zhu et al. (2010) measure the speed of censoring by the minutes that pass after a tweet is posted. Censorship of this type is fast: 30% are deleted within half an hour and 90% within one day. The speed of reading tweets is not directly observable but can be reasonably approximated by the speed of retweeting. In our dataset, user retweeting is almost as fast as censoring: 28% of the retweets occur within one hour after the original posting and 75% within one day. Our downloading is considerably slower. Tweets are downloaded at a constant frequency (typically daily), independent of the timing of tweet posting. The empirical distributions of censoring and retweet speed are plotted in Figure A2.

We estimate the share of tweets that are read and downloaded before censorship by sampling independent draws from the three empirical distributions described above. On average, approximately one third of the users who would have retweeted (read) a post in the absence of censorship have done so by the time of censoring. For users from whom we download tweets at a daily frequency, the simulated probability of downloading a tweet that is later censored is 0.23.

Based on these numbers, we conclude that using downloaded posts as a proxy for user exposure underestimates user exposure to censored content by around 30%, relative to uncensored content. This estimate is for content that does not contain sensitive keywords. In reality, many posts susceptible to censorship contain sensitive words and are automatically quarantined. For these posts, there is no underestimation because we see the same posts as users do. Hence, the average underestimation across all posts is likely to be considerably smaller than 30%.

Another discrepancy between user exposure to content and the number of downloaded tweets arises because the latter does not account for the number of readers of each post. Even though the number of posts are similar across content types, some content may be read by

many more users. As retweets are increasing in the number of users reading tweets, we also compare the number of retweets of different types of content below.

7.4 Mechanisms and Supporting Content

Now, we describe the availability of different types of Weibo content and its implications for determining which mechanisms may explain our findings.

7.4.1 Scarce Content

Where and when to meet The most direct way in which social media may affect protests and strikes is by providing logistics information on where and when to meet. Theoretically, logistics information is typically modelled as reducing the cost of protesting (e.g., Little 2016; Enikolopov at al. 2020). Empirically, several studies show that protesters use social media to communicate logistics information and organize events, such as calling for action or at least mentioning where and when a protest is to take place, as documented in Russia (Enikolopov et al. 2020), in Turkey and Ukraine (Tucker et al. 2016), and during the Arab Spring (Steinert-Threlkeld 2017). The existence of logistics information on social media is even proposed as a necessary criterion for social media to affect protest participation (Tucker et al. 2016).

However, in China, content related to protest logistics is scarce on social media. Based on a random sample of 1,000 posts about protests, only 15 posts, retweeted a total of 15 times, involve a call for action, and only two explicitly state the time and location of action. In a random sample of 1,000 posts about strikes, none calls for action. Because of censorship, it could be that the visibility of posts with logistics content is more than proportional to the corresponding share in our downloaded posts. But the difference is likely to be small. Recall from Section 7.3 that most users do not read posts momentarily, and posts that exist for a short period and later censored are unlikely to be widely visible. Further, the few tweets with logistics content that we find do not seem to have been widely read. On average, they only have one retweet per tweet.

Regardless of whether it is caused by censorship or self-censorship, communication of logistics information is unlikely to be an important reason for the widespread social media effects on collective action that we find in China. This is consistent with our interviews with staff of the China Labor Bulletin in 2018, who stated that strike participants typically share logistics information via private text messaging services and use Weibo to gain public attention.

Other protest tactics Social media may spread protests and strikes because it carries information about effective protest/strike tactics that lowers the cost or increases the benefits of protesting (e.g., Chen and Suen 2016; Little 2016). Examples of such information are how to counteract the effects of tear gas (Tucker et al., 2016) and whether it is effective

to complement strikes with other actions, such as demonstrations and petitions to local administrators (Chang and Hess 2018).

However, in our random samples of 1,000 posts, only five protest posts and 13 strike posts mention tactics, such as whether violent or peaceful protests are more effective and what to demand from and how to cope with the government. Consistently, only a few posts discuss the outcome of an event, such as whether the protest is met with concessions. The posts that indeed talk about tactics receive, on average, only two retweets. Given the scarcity and limited circulation of posts mentioning tactics, the tactic-learning mechanism seems unlikely to be a significant driver of the spread of protests and strikes in China.

7.4.2 Abundant Content

Causes and emotional reactions Most posts about protests and strikes in our random samples report facts about the events other than logistics and tactics, coupled with emotional reactions, such as the protesters' anger and sympathy. Many posts —164 about protests and 223 about strikes—discuss what causes the event and the underlying social problems such as corruption and persistent wage arrears. These posts are among the most retweeted—3,389 times for the protest posts and 2,602 for the strike posts.

To investigate which posts circulate widely, we inspect a sample of the 100 most retweeted posts about protests and strikes. After removing those that are repeated or irrelevant, the sample contains 91 posts. Of them, 55 talk about ongoing events, almost all mentioning the cause of the event. The remaining 36 posts comment on past events, government policies, or social problems. The majority of these popular posts convey certain emotional elements. Many posts express protesters' anger after a description of government repression. Another common type of emotional content is outsiders' reaction to events, with posts stating that the protesters are unfairly treated and expressing sympathy and moral support.

A significant number of posts—137 about protests and 42 about strikes—question existing social institutions. The content of these posts ranges from criticism of the legal system to complaints about the lack of free speech. Because of their general nature, these posts have the potential to spread events across causes and industries.

Potential mechanisms The type of abundant content described above can spread protests and strikes for several reasons. For example, people who learn from social media that others are protesting may conclude that a "window of opportunity" has opened, in which protests and strikes are allowed without substantial political risks (Zhou 1993; Weiss 2013; Truex 2019). Such a window may also open if there are complementarities among simultaneous protests across cities, which increase the chance of obtaining concessions or reduce the risk of punishment (e.g., Granovetter 1978; Edmond 2013; Little 2016; Barberà and Jackson 2020). It seems that merely observing a protest for a certain cause in another city will increase the expected benefits and lower the expected cost of protesting. People simply need a public

signal of a focal period to implicitly coordinate their actions.

Protests may also spread if protesters are motivated by emotions such as anger (Jasper, 2008; Passarelli and Tabellini 2017), and social media is indeed an important channel to spread emotions (Kramer et al. 2014). The escalation of protests due to strategic complementarities or emotional mobilization is a general phenomenon, not restricted to China. In other countries, these mechanisms are also likely to play a significant role in spreading protests even if social media does not circulate content that explicitly helps organize protests. They are particularly important in authoritarian regimes where the opportunities to explicitly organize anti-government protests are limited.

7.5 Discussion

As mentioned above, there is ample evidence that communication of logistics or tactics information via social media facilitates protests in a relatively free media environment. In China, however, this type of content is largely absent, likely because the Chinese government censors posts that have the potential to spur collective action (King et al. 2013, 2014).

This provides an opportunity to study whether removing logistics information mutes the effect of social media on protests and strikes. We find that, despite the scarcity of information on where and when to meet and how to organize events, Chinese social media still have a notable impact on the spread of protests and strikes. This finding suggests the relevance of the other potential mechanisms, such as those discussed above. It also points to a trade-off between eliciting information from social media and curbing the spread of protests. The Chinese government cannot prevent events from spreading without removing exactly the information it needs for surveilling local policies and monitoring corruption.

One natural question is why the Chinese government allows the circulation of tweets describing protests and strikes but restricts the exposure to information on logistics and tactics. One likely explanation is that the government allows the former content because this information is valuable for surveillance but removes the latter because its visibility would lower participants' cost of protesting without providing the government with new information. Suppose that the regime allows protests of limited scale and scope, which can inform the government about unpopular policies and local corruption. Protests are a costly and thus informative device to convey public grievances. Social media provides a cheap way for the government to gather information about the causes and the severity of social problems. Hence, protests and social media combined produce credible and detailed information. Suppose further that the government wishes to address problems with severity above some threshold value, and that the number of protesters increases in the severity of the problem and decreases in the cost of protesting. Social media content about logistics and tactics lowers the cost of protesting and increases the size of protests at a given level of problem severity, but it does not provide additional information value to the government.

8 Conclusion

This paper studies how social media affects the dynamics of protests and strikes in China. The media environment in China is significantly less free than in most other countries. At the same time, the Chinese government is known for actively collecting information from social media for surveillance. China's model of strategic control of social media may be used by other regimes to maintain stability. Thus, it is important to systematically examine the effect of social media on the spread of protests and the probability of very large waves of simultaneous events with significant consequences at a national scale.

Despite China's strict control of anti-government collective action and extensive media censorship, we document thousands of protest and strike events and millions of tweets referencing them on social media. We use retweets from one city of tweets from another city within the last few months to measure information diffusion between cities through the social media network. Then, we estimate the effect of the changing social media connections on the spread of protests and strikes in a panel of Chinese cities during 2006-2017. We exploit the time-series dimension of our data to address major issues in network econometrics and formulate a difference-in-differences type of estimator, aided by instrumental variables, to establish causality.

We find that social media rapidly spreads protests and strikes across cities. Moreover, social media significantly increases the scope of events such that protests spread to other protests with different causes and strikes spread to strikes in other industries. These effects dramatically increase the likelihood of very large protest waves, which simultaneously affect dispersed cities comprising a significant share of the Chinese population.

Remarkably, the strong effects of social media are found amidst China's substantial efforts to censor content used to organize unauthorized collective action. Given the scarcity of posts specifying where and when to meet and how to organize protests, the explicit-organization role of social media in China is less likely to be as important as in a freer media environment. The content circulating widely on Chinese social media supports other mechanisms, which enable coordination through signaling the opening of a "window of opportunity" or mobilization through people's emotional responses. These mechanisms are not specific to the Chinese context; they play a prominent role in the theory of protests (e.g. Granovetter 1978; Jasper 2008). Hence, our finding that purging social media of logistics or tactics content does not mute its effect on spreading protests and strikes is also relevant to other settings.

Our findings suggest that authoritarian regimes face an unavoidable censorship trade-off. To limit the spread of collective action, China must shut down public discussion about causes of social problems and silence people's emotional reactions and thus bear the cost of losing bottom-up information that is valuable for surveillance and monitoring.

References

[1] Acemoglu, Daron, Camilo Garcia-Jimeno, and James A. Robinson. 2015. "State capacity and economic development: A network approach." American Economic Review 105(8): 2364-2409.

[2] Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun. 2018. "The Power of the Street: Evidence from Egypt's Arab Spring." The Review of Financial Studies 31 (1): 1-42

[3] Aral, Sinan, Lev Muchnik, and Arun Sundararajan. 2009. "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks." Proceedings of the National Academy of Sciences 106.51: 21544-21549.

[4] Arellano, Manuel. Panel data econometrics. 2003. Oxford university press.

[5] Arellano, Manuel, and Stephen Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." The review of economic studies 58.2: 277-297.

[6] Bamman, D., O'Connor, B., & Smith, N. 2012. Censorship and deletion practices in Chinese social media. First Monday.

[7] Bakshy, Eytan, tamar Rosenn, Cameron Marlow, and Lada Adamicet. 2012. "The role of social networks in information diffusion." Proceedings of the 21st international conference on World Wide Web. ACM, p. 16-20.

[8] Baltagi, Badi. Econometric analysis of panel data. John Wiley & Sons, 2008.

[9] Barberà, Salvador, and Matthew O. Jackson. 2020. "A model of protests, revolution, and information." Quarterly Journal of Political Science, 15 (3), 297-335.

[10] Battaglini, Marco. "Public protests and policy making." 2017. The Quarterly Journal of Economics 132.1: 485-549.

[11] Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." Journal of econometrics 87.1: 115-143.

[12] Born, Benjamin and Jorg Breitung. 2016. "Testing for Serial Correlation in Fixed-Effects Panel Data Models," Econometric Reviews, 35:7, 1290-1316.

[13] Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. "Identification of peer effects through social networks." Journal of econometrics 150.1 (2009): 41-55.

[14] Brauer, A., and I. C. Gentry. 1974. "Bounds for the greatest characteristic root of an irreducible nonnegative matrix." Linear Algebra and its Applications 8.2: 105-107.

[15] Bursztyn, L., Cantoni, D., Yang, D. Y., Yuchtman, N., & Zhang, Y. J. 2020. Persistent political engagement: Social interactions and the dynamics of protest movements. American Economic Review: Insights.

[16] Cai, Yongshun. 2008. "Power structure and regime resilience: contentious politics in China." British Journal of Political Science: 411-432.

[17] Cai, Yongshun. 2010. Collective resistance in China: Why popular protests succeed or fail. Stanford University Press.

[18] Campante, Filipe R., Davin Chor, and Bingjing Li. 2019. The Political Economy Con-

sequences of China's Export Slowdown. NBER w25925.

[19] Cantoni, D., Yang, D. Y., Yuchtman, N., & Zhang, Y. J. (2019). Protests as strategic games: experimental evidence from Hong Kong's antiauthoritarian movement. The Quarterly Journal of Economics, 134(2), 1021-1077.

[20] Cattaneo, M. D., Crump, R. K., Farrell, M. H., & Feng, Y. (2022). On binscatter. arXiv:1902.09608v3.

[21] Chang, Shengping, and Steve Hess. 2018. "The Diffusion of Contention in Contemporary China: An investigation of the 2014–15 wave of teacher strikes." Modern Asian Studies 52.4: 1172-1193.

[22] Chen, Heng, and Wing Suen. 2016. "Falling Dominoes: A Theory of Rare Events and Crisis Contagion." American Economic Journal: Microeconomics, 8(1): 228-255.

[23] China Internet Network Information Center. 2014. "The 34th Statistical Report on Internet Development in China."

[24] China Labor Bulletin. 2018. "The Workers' Movement in China: 2015-2017."

[25] Christensen, Darin, and Francisco Garfias. 2018. "Can you hear me now? How communication technology affects protest and repression." Quarterly journal of political science 13.1: 89.

[26] Conley, T. G. (1999). GMM estimation with cross sectional dependence. Journal of econometrics, 92(1), 1-45.

[27] Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." Review of Economic Studies 80(4): 1422–1458.

[28] 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(4): 645–68.

[29] Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." American Economic Review, 111(7): 3253-85

[30] Fergusson, Leopoldo and Carlos Molina, 2020. "Facebook Causes Protests," Documentos de Trabajo LACEA 018004, The Latin American and Caribbean Economic Association.

[31] 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." IEEE Internet Computing 17(3): 42–50.

[32] Gehlbach, Scott, Konstantin Sonin, and Milan W. Svolik. "Formal models of nondemocratic politics." Annual Review of Political Science 19 (2016): 565-584.

[33] Graff Zivin, Joshua, Yingquan Song, Qu Tang, and Peng Zhang. 2020. "Temperature and High-Stakes Cognitive Performance: Evidence from the National College Entrance Examination in China", Journal of Environmental Economics and Management, 104.

[34] Granovetter, M. 1978. "Threshold models of collective behavior." American journal of sociology, 83(6), 1420-1443.

[35] Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya. "3g internet and confidence

in government." The Quarterly Journal of Economics (2020).

[36] Howard, Philip. 2011. The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam. Oxford University Press.

[37] Hsiao, Cheng. 2014. Analysis of panel data. Cambridge university press.

[38] Jasper, James M. 2008. The art of moral protest: Culture, biography, and creativity in social movements. University of Chicago Press

[39] 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): 1–18

[40] King, Gary, Jennifer Pan, and Margaret E Roberts. 2014. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation." Science 345(6199): 1–10.

[41] King, Gary, and Langche Zeng. "Logistic regression in rare events data." Political analysis 9.2 (2001): 137-163.

[42] König, Michael D., Dominic Rohner, Mathias Thoenig, and Fabrizio Zilibotti. 2017. "Networks in conflict: Theory and evidence from the great war of Africa." Econometrica 85.4: 1093-1132.

[43] Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788-8790.

[44] Kuruvilla, Sarosh and Hao Zhang. 2016. "Labor Unrest and Incipient Collective Bargaining in China," Management and Organization Review, Vol 12(1): 159-187.

[45] Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. 2010. "What is Twitter, a social network or a news media?." Proceedings of the 19th international conference on World wide web. AcM, p. 591-600.

[46] Lee, Ching Kwan, and Yonghong Zhang. 2013. "The power of instability: Unraveling the microfoundations of bargained authoritarianism in China." American Journal of Sociology 118.6: 1475-1508.

[47] Little, Andrew T. 2016. "Communication technology and protest." The Journal of Politics 78.1: 152-166.

[48] Lohmann, Susanne. 1993. "A signaling model of informative and manipulative political action." American Political Science Review: 319-333.

[49] Lorentzen, Peter. 2014. "China's Strategic Censorship." American Journal of Political Science 58(2): 402–414.

[50] Lorentzen, Peter. "Designing contentious politics in post-1989 China." Modern China 43.5 (2017): 459-493.

[51] Manacorda, Marco, and Andrea Tesei. "Liberation technology: Mobile phones and political mobilization in Africa." Econometrica 88.2 (2020): 533-567.

[52] Manski, C., 1993. Identification of endogenous social effects: The reflection problem.

Review of Economic Studies 60 (3), 531-542.

[53] Miao, M. J. Ponticelli, and Y. Shao. 2021. "Eclipses and the Memory of Revolutions: Evidence from China," working paper, Northwestern University.

[54] Nickell, Stephen. 1981. Biases in Dynamic Models with Fixed Effects. Econometrica, Vol. 49, No. 6, pp. 1417-1426.

[55] O'brien, Kevin J., and Lianjiang Li. 2006. Rightful resistance in rural China. Cambridge University Press.

[56] Passarelli, Francesco, and Guido Tabellini. 2017. "Emotions and political unrest." Journal of Political Economy 125.3: 903-946.

[57] Qin, Bei, David Strömberg, and Yanhui Wu. 2017. "Why does China allow freer social media? Protests versus surveillance and propaganda." Journal of Economic Perspectives 31.1: 117-40.

[58] Qin, Bei, David Strömberg, and Yanhui Wu. 2018. "Media bias in China." American Economic Review 108.9: 2442-76.

[59] Steinert-Threlkeld, Z. C. (2017). Spontaneous collective action: Peripheral mobilization during the Arab Spring. American Political Science Review, 111(2), 379-403.

[60] Steinhardt, H. C. 2017. "Discursive Accommodation: Popular Protest and Strategic Elite Communication in China." European Political Science Review, 9(4):539-560.

[61] Su, Yang, and Xin He. "Street as courtroom: state accommodation of labor protest in South China." Law & Society Review 44.1 (2010): 157-184.

[62] Truex, Rory. 2019. Focal points, Dissident Calendars, and Preemptive Repression, Journal of Conflict Resolution, Vol. 63(4) 1032-1052.

[63] Tucker, J. A., Nagler, J., MacDuffee, M., Metzger, P. B., Penfold-Brown, D., & Bonneau, R. (2016). Big data, social media, and protest. Computational social science, 199.

[64] Weiss, Jessica Chen. 2013. "Authoritarian Signaling, Mass Audiences, and Nationalist Protest in China." International Organization 67:1-35.

[65] Woolley, Samuel and Philip Howard. 2019. Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media. Oxford University Press.

[66] Zhang, Han and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media" Sociological Methodology 49(1): 1-57.

[67] Zhou, Xueguang, 1993. Unorganized Interests and Collective Action in Communist China, American Sociological Review, Vol. 58, No. 1 (Feb., 1993), pp. 54-73

[68] Zhu, T., Phipps, D., Pridgen, A., Crandall, J. R., & Wallach, D. S. (2013, August). The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions. In USENIX Security Symposium (pp. 227-240).

[69] Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. "Political effects of the internet and social media." Annual Review of Economics 12 (2020): 415-438.

Tables and Figures

Table 1. Effects of social media on event spread: time-varying measure of connections

VARIABLES	(1) Protest	(2) Protest	(3) Protest	(4) Strike	(5) Strike	(6) Strike
Social-media spread (β)	0.170***	0.169***	0.209***	0.122***	0.120***	0.107***
	(0.044)	(0.043)	(0.060)	(0.026)	(0.025)	(0.038)
Geo-distance spread (γ)	-0.031	-0.030		0.143**	0.137**	
	(0.037)	(0.036)		(0.058)	(0.055)	
Number events 1-2 days prior(α)	0.015**	0.015**	0.000	0.030***	0.030***	0.000
	(0.007)	(0.007)	(0.000)	(0.007)	(0.006)	(0.000)
Total number retweets	0.002***	0.002***	0.003***	0.002**	0.002	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Weibo posts	0.006***	0.006***	0.004**	0.010***	0.009***	0.007**
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)
Controls	No	Yes	Yes	No	Yes	Yes
Observations	670,996	670,996	670,996	713,702	713,702	713,702
R-squared	0.017	0.017	0.220	0.027	0.027	0.239
QPtest	0.05	0.13		0.14	0.34	

Notes: Results are from a linear regression of an event dummy. The unit of observation is city by date. The regression includes city and date fixed effects. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Columns (3) and (6) allow for arbitrary time-invariant heterogeneity in the spread across city pairs. The QP statistic reports the p-value of the test for serial correlation in fixed-effects models (Born and Breitung, 2016). Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table 2. Effects of social media on event spread: time-constant measure of connections

VARIABLES		(1) Protest OLS	(2) Protest OLS	(3) Protest IV	(4) Strike OLS	(5) Strike OLS	(6) Strike IV
Social-media spread, s_{it-1}	β^0	0.029*	0.027	0.032	0.026	0.031	0.059
		(0.015)	(0.017)	(0.028)	(0.029)	(0.032)	(0.037)
	β^1	0.150***	0.169***	0.162***	0.127***	0.134***	0.098***
		(0.041)	(0.043)	(0.056)	(0.028)	(0.030)	(0.034)
	β^2	0.078***	0.084***	0.051**	0.116***	0.119***	0.132***
		(0.020)	(0.021)	(0.023)	(0.015)	(0.016)	(0.021)
Geo-distance spread, d_{it-1}	γ^0	-0.004	-0.011	-0.013	0.104	0.105	0.093
		(0.016)	(0.016)	(0.018)	(0.063)	(0.065)	(0.074)
	γ^1	-0.026	-0.021	-0.016	0.107*	0.106*	0.131*
		(0.036)	(0.035)	(0.040)	(0.056)	(0.058)	(0.074)
	γ^2	0.022	0.021	0.035	0.040	0.048	0.041
		(0.023)	(0.024)	(0.023)	(0.032)	(0.031)	(0.032)
Controls		No	Yes	Yes	No	Yes	Yes
Observations		1,140,224	1,028,085	999,472	1,119,858	999,472	1,092,477
R-squared		0.018	0.019	0.001	0.050	0.001	0.003
P-value: $\beta^1 = \beta^0$		0.006	0.002	0.035	0.008	0.011	0.307
P-value: $\beta^2 = \beta^0$		0.067	0.046	0.661	0.001	0.004	0.041
Kleibergen-Paap F				34.26			38.49
		Period 0	Period 1	Period2			
Total No. Of Protests		126	1046	1517			
Total No. Of Strikes		193	1365	7857			

Notes: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period, and date fixed effects. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Columns (3) and (6) report results from the IV (student-mobility) regressions. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table 3. Event spread, within and across categories

	(1) *Protest*	(2) *Strike*
Within		
Number events 1-2 days prior, retweet weighted	0.0639***	0.0607***
	(0.0132)	(0.0140)
Number events 1-2 days prior, distance weighted	-0.0094	0.0943***
	(0.0106)	(0.0266)
Across		
Number events 1-2 days prior, retweet weighted	0.0086***	0.0064***
	(0.0029)	(0.0020)
Number events 1-2 days prior, distance weighted	-0.0017	0.0056
	(0.0026)	(0.0040)
Observations	8,722,948	7,137,020
R-squared	0.0077	0.0162
Category	cause	industry
Mean (within-category)	0.0013	0.0020
Mean (across-category)	0.0152	0.0179

Notes: Results are from a linear regression of an event dummy. The unit of observation is city by date. The regression includes city and date fixed effects. Both regressions include controls such as population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1. Events by size and cause

Figure 2. Monthly number of events and number of Weibo posts per capita over time

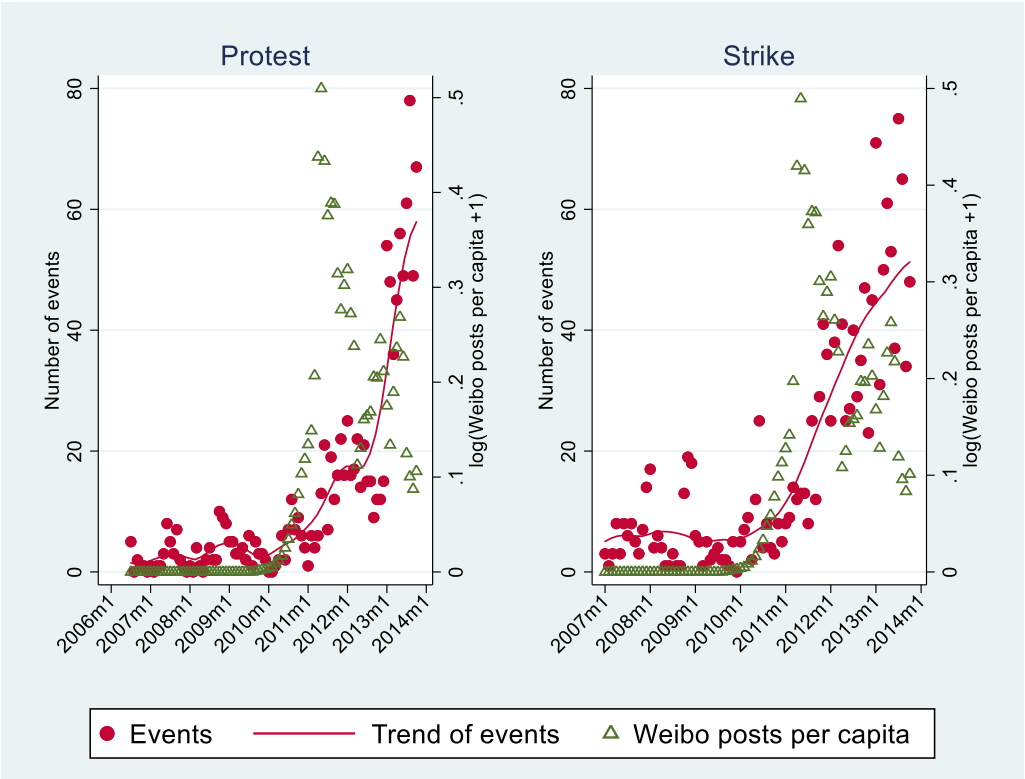
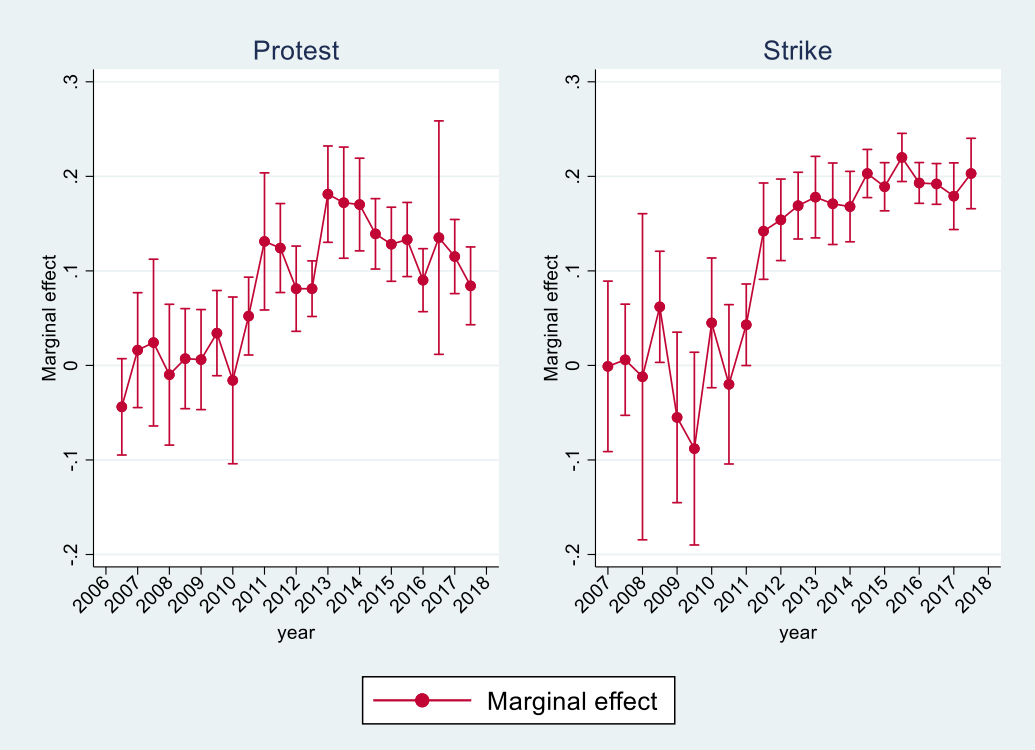
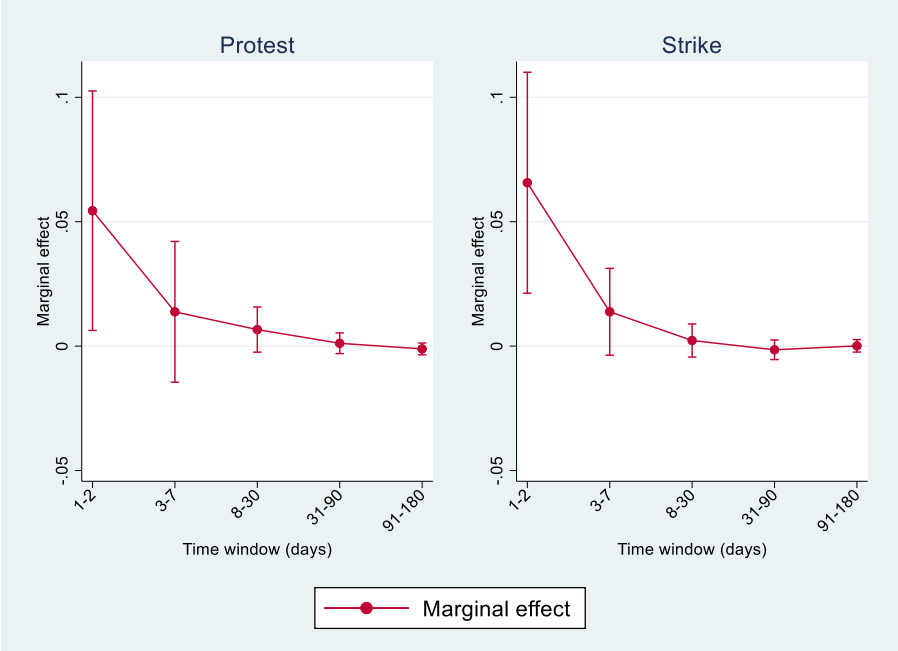


Figure 3. Dynamics of marginal effects using constant measure of social media connections



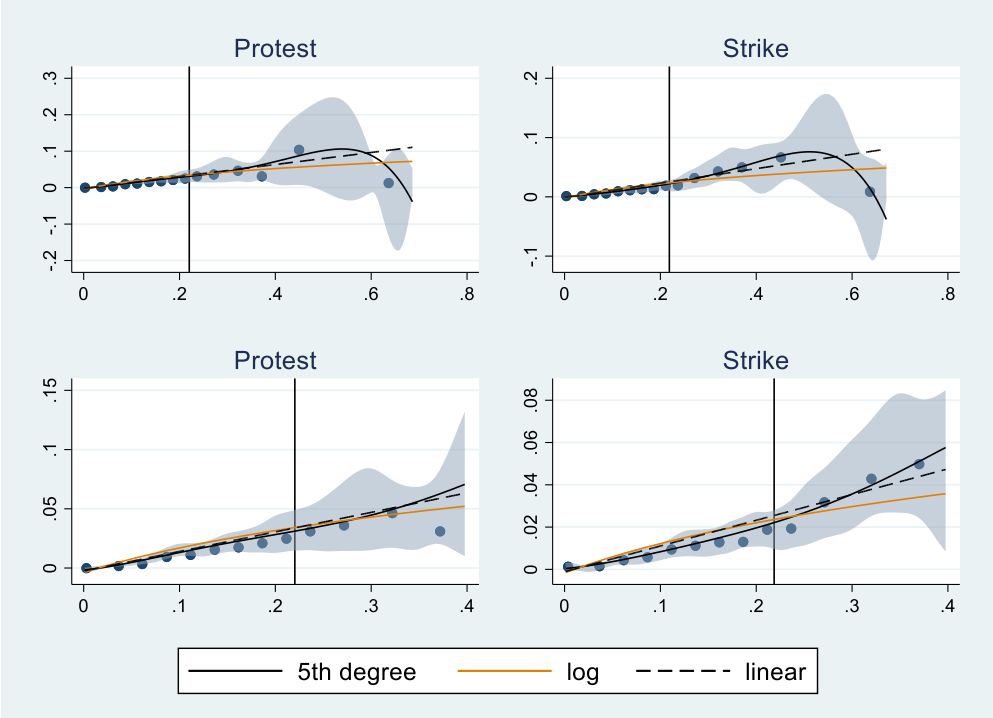
Notes: The graphs plot the estimated biannual coefficients on the time-constant measure of social media connections, relative to the mean in the pre-Weibo period. The bars indicate the 95% confidence interval.

Figure 4. Effect of social media on event spread: speed and duration



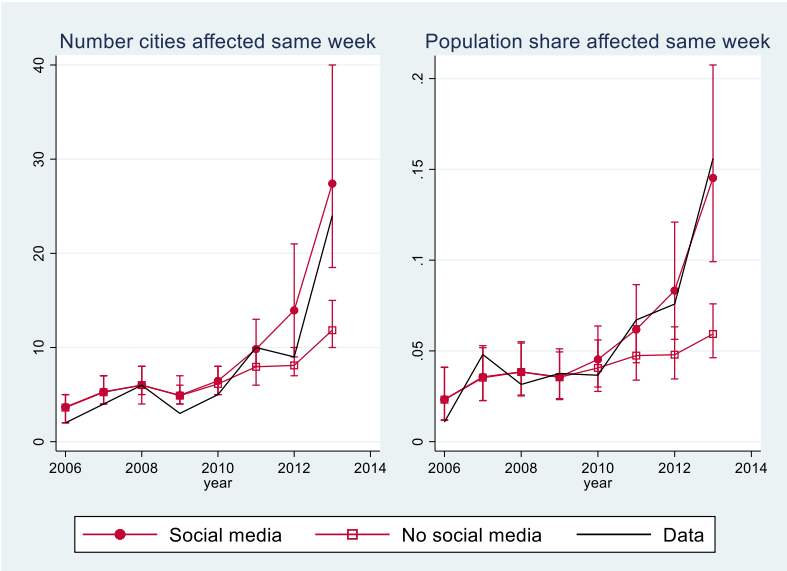
Notes: The graphs plot the estimates of the Weibo spread effects specified in the baseline model with various time windows. The bars indicate the 95% confidence interval.

Figure 5. Functional forms



Notes: The x-axis indicates s_{it-1}. The y-axis plots the nonparametric conditional mean function (the dots), together with a linear, a logarithmic, and a 5th-order polynomial approximation, using the specification in Equation (1). The vertical line marks the 99th percentile of the s_{it-1}-distribution. The bottom panels zoom in on the lower segment which contains most of our data.

Figure 6. Simulated event waves with and without social media



Notes: The graphs plot the number of cities (left) and share of population (right) that the largest event waves reach in the same week in simulations of the model with and without social media. The points denote the yearly average, and the bars denote the corresponding 5-95 percentile range.

A Online Appendix to Social Media and Collective Action in China

A.1 Specification

To examine the problem of correlated and contextual effects, we start with the formulation similar to Moffitt (2001). Suppose that we have $g = 1, ...G$ groups and that there are only two cities ($i = 1, 2$) per group. Let y_{ig} be the outcome variable of interest for city i in group g, x_{ig} be an observed characteristic of i, which in our case is the lagged outcome, and v_{ig} be an unobservable characteristic. Assume that the model that we wish to estimate is

$$\begin{aligned} y_{1g} &= \alpha_0 + \alpha x_{1g} + \beta x_{2g} + v_{1g}, \\ y_{2g} &= \alpha_0 + \alpha x_{2g} + \beta x_{1g} + v_{2g}. \end{aligned}$$

In particular, we are interested in β, which captures how lagged events spread from city $1g$ to connected city $2g$, and vice versa.

A concern is that v_{1g} and v_{2g} are correlated, for example, because of correlated and contextual effects. The correlated effects are typically parametrized by an unobserved group-specific unobserved shock, δ_g, which is part of the error and for which

$$E(\delta_g x_{ig}) \neq 0.$$

The contextual effects are typically parameterized by including a term γx_{2g} in the error term v_{1g}. In our case, contextual effects could arise from events spreading through communication channels that are correlated with the social media network, such as cell-phone use. Thus, the error term, v_{1g}, can be decomposed as:

$$v_{1g} = \gamma x_{2g} + \delta_g + \varepsilon_{1g}.$$

By the same token, v_{2g} is similarly defined.

In this model, $\gamma \neq 0$ implies contextual effects and $\delta_g \neq 0$ implies correlated effects. Clearly, β is not identified in the presence of these effects.

A.1.1 Setup

We estimate our model using a panel of N Chinese cities at daily frequency, t. Let y_{it} be a dummy variable indicating the occurrence of a protest. Suppose that the probability of a protest in city i on day t, $\Pr(y_{it})$ depends on the number of people who are informed about protests y_{jt-1} in another city j at time $t-1$. Let f_{ijt} be the number of people in city i who read posts from users in city j at time t and assume that the number of people who learn about the protest is $f_{ijt}y_{jt-1}$. We wish to model how protests spread from city j to city i as

a function of $f_{ijt}y_{jt-1}$.

We illustrate the identification assumptions with the following simple example. There are two periods (pre-Weibo and post-Weibo) and two groups of cities (N, C) that may be affected by a protest in another city j on a previous date. Cities in group N are never connected via social media to city j, whereas cities in group C are connected via social media to city j in the post-Weibo period only. Social media connections increase the probability that a protest spreads from j to other cities. In both periods, there are some days with protests in location j, and some days without such protests. Let y_{ji-1} be an indicator variable for a protest in city j at $t-1$, f_i be an indicator variable for city i belonging to group C and P_t be an indicator variable for the post-Weibo period. Then, we have

$$y_{it} = \delta + \alpha y_{it-1} + \beta^p f_i y_{jt-1} + v_{it}, \tag{4}$$

where β^p measures the effect of events spreading through the social media network with $\beta^0 = 0$ in the pre-Weibo period ($p = 0$) and $\beta^1 = \beta$ in the post-Weibo period ($p = 1$). Further, let

$$v_{it} = \gamma f_i y_{jt-1} + f_i \delta_g + \varepsilon_{it}, \tag{5}$$

where $\gamma f_i y_{jt-1}$ captures the contextual effects and $f_i \delta_g$ captures the correlated effects.

A.1.2 Triple-Differences Estimator

Consider the following triple-differences equation in the linear probability model, with outcome variable y_{it} being a protest in city i at time t,

$$\begin{aligned} y_{it} \quad = \quad & \beta_0 + \beta_1 f_i + \beta_2 y_{jt-1} + \beta_3 P_t \\ & + \beta_4 f_i y_{jt-1} + \beta_5 f_i P_t + \beta_6 P_t y_{jt-1} \\ & + \beta P_t f_i y_{jt-1} + \varepsilon_{it}. \end{aligned} \tag{6}$$

To simplify the exposition, we drop from the equation the term αy_{it-1}, which captures within-city spread. The conditional mean function $E[y_{it}|f_i, y_{jt-1}, P_t]$ can take on eight values, and the model is saturated because it has eight parameters. The first row of Equation (6) contains the three main effects and the constant, the second row contains the three two-way interactions, and the third row contains the triple interaction. This model allows for the probability of a protest in city i to be larger in connected cities (β_1) than in nonconnected cities, for example, because of correlated effects (absorbing $f_i \delta_g$) in the post-Weibo period (β_3), and additionally higher for connected cities in the post-Weibo period (β_5). Protests are also allowed to spread across all cities (β_2), and differentially so in the post-Weibo period (β_6). In particular, note that protests in city i may be more likely to occur immediately after a protest in a connected city j in the pre-Weibo period (β_4), such as because of contextual effects (absorbing $\gamma f_i y_{jt-1}$). The coefficient β captures the difference in the spread of protests

between connected and nonconnected cities before and after Weibo entry.

Under the standard OLS assumption, $E[\varepsilon_{it}|f_i, y_{jt-1}, P_t] = 0$. For compactness, we use the notation

$$E[Y|f_i = 1, y_{jt-1} = 1, P_t = 1] = Y_{C,S,Post},$$

where C denotes a connected city ($f_i = 1$), S denotes lagged protests in city j ($y_{jt-1} = 1$) and $Post$ denotes $P_t = 1$. Similarly, let N denotes nonconnected city ($f_i = 0$), W denote a day without a protest in city j, $y_{jt-1} = 0$, and Pre denote that $P_t = 0$. It is straightforward to show that $\beta = \delta_C - \delta_N$, where

$$\delta_C = (Y_{C,S,Post} - Y_{C,W,Post}) - (Y_{C,S,Pre} - Y_{C,W,Pre})$$

is the difference in city C between the pre-Weibo and post-Weibo periods in the difference in protest incidence when there is a protest or not in city j, and

$$\delta_N = (Y_{N,S,Post} - Y_{N,W,Post}) - (Y_{N,S,Pre} - Y_{N,W,Pre})$$

is the equivalent difference in city N.

By re-arranging terms, we can also write the estimator as $\beta = \delta_1 - \delta_0$, where

$$\delta_1 = (Y_{C,S,Post} - Y_{C,W,Post}) - (Y_{N,S,Post} - Y_{N,W,Post})$$

is the difference between connected and nonconnected cities in the difference between days with and without protests in the post-Weibo period, and

$$\delta_0 = (Y_{C,S,Pre} - Y_{C,W,Pre}) - (Y_{N,S,Pre} - Y_{N,W,Pre})$$

is the equivalent difference in differences in the pre-Weibo period.

A.1.3 Parallel-Trends Assumption

Below, we articulate the identification conditions for the triple-difference estimator we have derived under the potential outcomes framework.

Let $E[Y_1|C, S, Post]$ denote the expected outcome in state $(C, S, Post)$ in the case that a city is connected and $E[Y_0|C, S, Post]$ denote the expected outcome in the counter factual case in which the city is not connected. The estimate β_{TT} captures the causal effect of being connected to a city with a protest in the post-Weibo period for the connected cities, formulated as follows:

$$\beta_{TT} = E[Y_1|C, S, Post] - E[Y_0|C, S, Post].$$

Our estimator is

$$\begin{aligned}\beta &= \delta_C - \delta_N \\ &= \left(E\left[Y_1|C,S,Post\right] - E\left[Y_0|C,W,Post\right]\right) - \left(E\left[Y_0|C,S,Pre\right] - E\left[Y_0|C,W,Pre\right]\right) \\ &\quad - \left(E\left[Y_0|N,S,Post\right] - E\left[Y_0|N,W,Post\right]\right) - \left(E\left[Y_0|N,S,Pre\right] - E\left[Y_0|N,W,Pre\right]\right).\end{aligned}$$

The condition for $\beta = \beta_{TT}$ is that absence treatment, $\delta_C = \delta_N$. Since the only term that contains treatment in δ_C is the first term, $E\left[Y_1|C,S,Post\right]$, the parallel-trend condition is

$$\begin{aligned}&\left(E\left[Y_0|C,S,Post\right] - E\left[Y_0|C,W,Post\right]\right) - \left(E\left[Y_0|C,S,Pre\right] - E\left[Y_0|C,W,Pre\right]\right) \quad (7) \\ = \;&\left(E\left[Y_0|N,S,Post\right] - E\left[Y_0|N,W,Post\right]\right) - \left(E\left[Y_0|N,S,Pre\right] - E\left[Y_0|N,W,Pre\right]\right).\end{aligned}$$

A.1.4 Contextual and Correlated Effects

Our estimator differences out the contextual and correlated effects in Equations 4 and 5. The correlated effects, $f_i\delta_g$, are differenced out because both δ_C and δ_N are constructed solely from the within-group differences.

The contextual effects are partialed out because they are contained in both δ_1 and δ_0, and $\delta_1 - \delta_0 = (\beta + \gamma) - \gamma = \beta$. The parallel-trends assumption in Equation 7 implies that the difference-in-differences captured by δ_0 would have been the same in the post-Weibo period, in the counter factual case that the social media network did not exist,

$$\left(E\left[Y_0|C,S,Post\right] - E\left[Y_0|C,W,Post\right]\right) - \left(E\left[Y_0|N,S,Post\right] - E\left[Y_0|N,W,Post\right]\right) = \gamma.$$

We cannot test the assumptions directly because we do not observe $E\left[Y_0|C,S,Post\right]$. However, we can estimate γ separately for each half-year, b, using

$$\widehat{\gamma}_b = \left(\overline{Y}_{C,S,Pre} - \overline{Y}_{C,N,Pre}\right)_b - \left(\overline{Y}_{N,S,Pre} - \overline{Y}_{N,N,Pre}\right)_b$$

to see whether this exhibits a trend in the pre-Weibo period.

A.1.5 Multiple cities

In the real-world data that we use, protests occur in all cities, and social media connections in the post-Weibo period are of different strengths. Let f_{ij} be the strength of city i's connection to city j. Now, we generalize the above model by linearly adding the contributions described in Equation 6 from each city j,

$$\begin{aligned}y_{it} &= \beta_0 + \beta_1\overline{f}_i + \beta_2\left(y_{t-1} - y_{it-1}\right) + \beta_3 P_t \\ &\quad + \beta_4 s_{it-1} + \beta_5\overline{f}_i P_t + \beta_6 P_t\left(y_{t-1} - y_{it-1}\right) \\ &\quad + \beta P_t s_{it-1} + \varepsilon_{it}.\end{aligned}$$

where $\overline{f}_i = \sum_j \overline{f}_{ij}$, $y_{t-1} = \sum_j y_{jt-1}$, and $s_{it-1} = \sum_{j \neq i} f_{ij} y_{jt-1}$. To allow for more flexibility, we will estimate an equation of the form

$$y_{it} = \alpha^p y_{it-1} + \beta^p s_{it-1} + \delta_i^p + \delta_t + \varepsilon_{it}, \tag{8}$$

where superscripts indicate the pre-Weibo and post-Weibo periods. The city–by-period fixed effects absorb the following terms:

$$\begin{aligned} \delta_i^0 &= \beta_0 + \beta_1 \overline{f}_i \\ \delta_i^1 &= \beta_0 + \beta_1 \overline{f}_i + \beta_3 P_t + \beta_5 \overline{f}_i P_t. \end{aligned}$$

The date-fixed effects and the lagged outcome in city i, y_{it-1}, interacted with the period-fixed effects p_t, absorb the following terms:

$$\beta_2 y_{t-1} + \beta_3 P_t + \beta_6 P_t y_{t-1} + (-\beta_2 - \beta_6 P_{t2})\, y_{it-1} = \delta_t + \alpha^0 y_{it-1} + \alpha^1 y_{it-1},$$

where $\delta_t = \beta_2 y_{t-1} + \beta_3 P_t + \beta_6 P_t y_{t-1}$, $\alpha^0 = -\beta_2$, and $\alpha^1 = -\beta_2 - \beta_6 P_{t2}$. To simplify the exposition, we omit the term involving the within-city spread, αy_{it-1}, from Equation (4) when we constructed Equation (6). To include this spread, α should be added to α^0 and α^1. Finally, our specification includes s_{it-1}, interacted with period-fixed effects p_t. This absorbs the terms

$$\beta_4 s_{it-1} + \beta P_t s_{it-1} = \beta^0 s_{it-1} + \beta^1 s_{it-1},$$

where $\beta^0 = \beta_4$, and $\beta^1 = \beta_4 + \beta$. After adding controls, Equation (8) corresponds to Equation (2) in the main manuscript.

A.2 Monte Carlo Simulations

A.2.1 Stationarity

In our setting, the stationarity of the process of protests or strikes is not simply an econometric issue. Whether these processes are stable or exploding is likely to be a core concern of an authoritarian regime (the Chinese central government in our context). Consider the following simple first-order serial autoregressive process: $y_t = \rho y_{t-1} + \varepsilon_t$.

This process is stationary if $|\rho| < 1$. If $\rho > 1$, then, on average, each y_t is larger than the past y_{t-1}, and in expectation, y_t grows exponentially over time. We can investigate whether a process is on such an exploding path by simulating outcomes generated from drawing random shocks from the distribution of ε_t, and iteratively computing sequences of y_t. Our setting is more complex than this simple example because of the network structure, multiple time lags, and discrete outcomes, but the principle remains the same.

In a dynamic spatial panel data model, stationarity depends on the parameters of the

model as well as on the spatial weight matrix, which determines the amount of autocorrelation, or feedback, in the process. For the location's own autoregressive term and the distance weighted term, this feedback is constant over time because α, γ, and the distance matrix D are constant. However, more intensive use of social media will increase the feedback, and each individual row in the forwarding matrix, F, does not sum up to one. This implies that the marginal effect of a change in y_{t-1} on the probability of an event differs across location and time. In the linear model, the average effect on a particular date t equals $\beta\overline{f}_{t-1}$, where $\overline{f}_{t-1}$ is the average row sum across locations on that date. The maximum of such a row sum is 10.5, an order of magnitude larger than the average row sum across all dates, t, which is normalized to one. This implies that the sufficient conditions for stationarity are not fulfilled for the linear model. In other words, stationarity requires a concave function for high values of s_{it-1}, as we demonstrate below.

In the baseline estimation, we assume that s_{it-1} enters linearly in Equation (1). Now, we investigate this assumption using a nonparametric least squares regression (Cattaneo et al. 2022). Specifically, we estimate the nonparametric conditional mean function $h(\cdot)$ using the following specification:

$$y_{it} = \alpha y_{it-1} + h\left(s_{it-1}\right) + \gamma d_{it-1} + \theta_0 w_{it} + \theta' x_{it} + \delta_i + \delta_t + \varepsilon_{it}. \tag{9}$$

Figure 5 in the paper plots the nonparametric conditional mean function (shown by the dots), estimated using the specification in Equation 1, together with three parametric approximations: a linear approximation, a logarithmic ($log(5s_{it-1}+1)$) approximation, and a 5-th order polynomial approximation. The conditional mean function is approximately linear for most of the support, and all three approximation functions yield very similar results for the estimated average marginal effects.[24]

However, as shown in Figure 5, when an event wave grows above a certain extent, the marginal spread effect of an additional event (the slope of the curve) falls. Thus, for sufficiently high values of s_{it-1}, there is no spread of events through social media. This causes the magnitude of y_t to fall rapidly in the right tail. Because of this feature, the process is never on an exploding path when we use either the 5th-order polynomial or the log approximations of the conditional mean functions for the data-generating process. Given that the log-model only has one parameter, and is sufficiently concave to avoid exploding paths, we use this functional form in most of the simulations.

[24] This is evident from results using the logarithmic function in the Appendix Tables A5 and A6, in comparison with our baseline results (Tables 1 and 2). The logarithmic model uses the conditional mean function $\ln(5s_{it\ \ 1}+1)$. The estimates of β are statistically significant across all specifications and the implied marginal effects on event probabilities are similar to those estimated using the linear model.

A.2.2 Nickell Bias

The Nickell bias arises in dynamic panels with fixed effects. In our estimation, we presume that the Nickell bias is small because the T in our panel model is very large. Nevertheless, we run a set of Monte Carlo simulations to assess the Nickell bias in the estimated coefficients of our baseline model. We first estimate the parameters $\alpha, \beta, \gamma, \delta_t$ and δ_i from a regression specified as in Equation 1, using the logarithmic function described above and without Weibo penetration and controls (for the sake of simplicity). Next, we generate data using the same model with the estimated parameters, adjusted such that $\delta_t + \delta_i \geq 0$. Then, we estimate the model parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ on the simulated data. We repeat this procedure 100 times. Figure A4 plots the distribution of t-statistics of coefficients α, β and γ in Equation (1) against the standard normal density. The bias is very small, as evident from the negligible difference between the true and the mean estimated β for both protests and strikes.

In Section 5, we report the estimates of a model including fixed effects for arbitrary time-constant spread across locations. We also run a set of Monte Carlo simulations to assess the Nickel bias in this model. Specifically, we use the baseline model for the data-generating process and then estimate the interaction-fixed-effects model using the simulated data. Figure A5 depicts the distribution of β-estimates from the Monte Carlo simulations. The graphs to the left show the results from estimations without interaction-fixed effects (Equation 1), corresponding to specifications in Columns (1) and (4) of Table 1. The graphs to the right are based on the regressions with interaction-fixed effects, corresponding to the specifications used in Columns (3) and (6). The blue line shows the true coefficient used in the data-generating process. The red line represents the mean coefficients from the estimated on the simulated data. These graphs show a bias of 0.011 for both protests and strikes.

We also test for the possibility of auto-correlated errors on the simulated data in which the autocorrelation in errors is absent by construction. The test for autocorrelation in the baseline model verifies this, although it slightly over-rejects the non-autocorrelation hypothesis. The model with interaction-fixed effects can control for the same pattern as the baseline model, but many interaction-fixed effects are imprecisely estimated. Removing slightly incorrect auto-correlated terms generates auto-correlated errors. Hence, it is not surprising that the autocorrelation test for the model with interaction-fixed effects incorrectly rejects no autocorrelation in our data which is simulated with no autocorrelation.

Presumably, autocorrelation would disappear as the sample size grows to infinity, because the coefficients are consistently estimated and would converge to the data generation process. The lack of autocorrelation in the errors in Table 1 shows that there is no significant spread of protests other than what is captured in the baseline model. Although the specifications used in Columns (3) and (6) of Table 1 generate auto-correlated errors in small samples, they will not severely bias the estimated coefficients, as shown in Figure A5.

A.2.3 Mechanical Zero Effect

The estimated social media effect on event spread is not closely related to the number of events in each period. For example, there are 50% more protests in period 2 than period 1 (1516 compared with 1,037), but the estimated spread in period 2 is smaller than in period 1. Similarly, while there are six times as many strikes in period 2 than in period 1 (7,857 compared with 1,365), the estimated spread effect is not significantly different. The lack of, or even a negative, relationship between the number of events and the size of the estimated effects suggests that there is no mechanical relationship between these two variables.

Concerns may remain that the smaller number of events in the pre-Weibo period would mechanically reduce the estimated effects. To further explore this, we simulate event data using a process that matches the observed event frequencies within each period, with the propagation of protest waves across cities being equally strong in the pre-Weibo and post-Weibo periods. Then, we estimate the effects on the simulated data and test whether the coefficients in the pre-Weibo period are significantly lower than in the post-Weibo period.

To implement this research design, we first run a set of Monte Carlo simulations to generate event data for the period from 2010 to 2013. To simulate data with lower event frequency for the pre-Weibo period (2006-2010), we use the post-Weibo data-generating process, and then sample sequences starting from a random date and continuing until the same number of events as in the real data are reached. We retain simulated events $\widetilde{y}_{it}$ in the sampled sequences but drop all other events (i.e., y_{it} is set to zero).

Then, we re-estimate the model of Table 2 based on these simulated data. Table A4 shows the results. Despite the much smaller number of events in the pre-Weibo period, the average estimated $\beta-$coefficient is only slightly lower in the pre-Weibo period than in the post-Weibo period (0.10 compared with 0.12). When we test whether the coefficient in the pre-Weibo period is lower than in the post-Weibo period at the 5% significance level, the hypothesis is rejected in 8% of the simulations.

A.2.4 Magnitude of Event Waves

In this section, we estimate the effect of social media on the size of an event wave, measured by the number of essentially simultaneous events across multiple cities. We are not only interested in the mean number of events but also in the likelihood of very large event waves. This requires that we specify some additional details about the dynamic process of event waves.

In addition to the effects through social media, measured by s_{it-1}, the propagation of protests may depend on the size of event waves in the real world, measured by the number of essentially simultaneous events, $y_{t-1} = \sum y_{it-1}$. On the same date, y_{t-1} is constant across all cities and hence is absorbed by the date fixed effect, δ_t, in Equation (1). The specification including date-fixed effects is preferred when the purpose is to identify the marginal effect of s_{it-1}, because it flexibly controls for all time-constant heterogeneity, including that through

y_{t-1}. However, this specification does not explicitly model the dynamic effects through y_{t-1}. Therefore, when the objective is to explore the size of event waves, we need to bring the effect of y_{t-1} to the scene. We do so by adding y_{t-1} to Equation (1) and replacing the date-fixed effects with fixed effects for running months. More specifically, we use the parameters $(\widehat{\alpha}, \widehat{\gamma})$ and the 5th-order polynomial $\widehat{h}_5(s_{it-1})$, estimated as described above, which are obtained from the specification in Equation 1 that includes date fixed effects (because this model most convincingly identifies these parameters). Then, we estimate the parameters δ_i, ρ, δ_m using the following specification:

$$y_{it} = \widehat{\alpha} y_{it-1} + \widehat{h}_5(s_{it-1}) + \widehat{\gamma} d_{it-1} + \delta_i + \rho y_{t-1} + \delta_m + \varepsilon_{it}, \tag{10}$$

where the date fixed effects, δ_t, are replaced with date-constant variables: the total number of events in the past two days, y_{t-1}, and the month-fixed effects, δ_m. The estimate of $\widehat{\rho} \approx -0.002$ implies that the probability of an event falls by 0.02 for every 10 additional events that occurred in the past two days.

Then, we use the above model to simulate event data under two scenarios, with and without social media. In both scenarios, we use the estimated probabilities that events erupt independently of other events, $\widehat{\delta}_i + \widehat{\delta}_m$, and we keep the within-city and geographical and aggregate spread, $(\widehat{\alpha}, \widehat{\gamma}, \widehat{\rho})$ at the estimated levels. In one scenario, we allow events to spread through social media using the estimated nonparametric conditional mean function, $\widehat{h}_5(s_{it-1})$. In the other scenario, we do not allow for event spread over social media by omitting the function $\widehat{h}_5(s_{it-1})$ from the data generating process. We simulate the data 1,000 times, obtaining 1,000 possible histories of protest events in our city panel up to 2013. For each simulation and year, we compute the maximum number of cities with protests within the same week and the corresponding maximum share of population in the cities affected by protests. The results are shown in Figure 6 and discussed in Section 6.3 of the paper.

A.3 Censorship

Censorship and aggregate retweets In 2015, we checked which posts in a subsample of posts in our dataset remained online. Based on this deletion rate at the regional level, we construct a measure of local censoring intensity. Given the literature on censorship in China, it seems unlikely that censorship significantly affects the number of retweets on nonsensitive topics. We verify this by regressing our time-constant measure of social media connections, f_{ij}, on the average share of deleted posts in city i and city j. All of the variables are standardized, and standard errors are clustered by city i and city j. Table A9 reports the results. It is clear that f_{ij} is not significantly related to the share of deleted posts in either the city of the tweeting user (j) or the city of the retweeting user (i). Column (2) adds city-level controls and Weibo penetration in 2012. The coefficients of interest are non-significant, and the implied magnitudes are small. Column (3) interacts the share of deleted posts with

Weibo penetration, because the effect on total retweets should be increasing in Weibo use. The coefficient of this interaction term is small and statistically insignificant.

Heterogeneity analysis We further investigate whether the spread of protests and strikes is correlated with our measure of local censorship intensity. First, censorship may affect outgoing messages from the city of event occurrence. Suppose that the event y_{jt-1} in city j at time $t-1$ causes a burst of tweets, a share of which, c_j, are censored. To capture this type of effect, we construct the variable $\sum_{i \neq j} f_{ij} c_j y_{jt-1}$ where c_j is the measure of the average share of deleted posts in city j. However, the variables $\sum_{i \neq j} f_{ij} c_j y_{jt-1}$ and $\sum_{i \neq j} f_{ij} y_{jt-1}$ are highly collinear, making it impossible to separately identify the effects from these two variables. Such multicollinearity arises because many events occur in cities with deletion rates close to the mean deletion rates, and thus the variance in the weights c_j is small, making the two variables highly correlated.

Second, we consider the situation in which censorship may prevent social media users in affected cities (to which the events could spread) from reading or retweeting incoming event-related tweets from other cities. We capture this by using the variable $c_i \sum_{i \neq j} f_{ij} y_{jt-1}$ where c_i is the share of deleted posts in city i. As shown in Table A10, neither the spread of protests nor the spread of strikes is significantly different in cities with different levels of censorship.

A.3.1 Censorship, Content Exposure, and Conclusions about Mechanisms

Because of censorship, the visibility to Weibo users of tweets with a certain content may differ from the availability of these tweets in our dataset.

Consider two types of content referencing protests, labelled A and B. Type-A content reveals logistics information (e.g., where and when to meet), which lowers the cost of protesting. Type-B content just talks about the occurrence of a protest or expresses sentiment, which supports other mechanisms. The Chinese government censors Type-A content but not Type-B content. Given that data downloading is slower than user reading of tweets, users may read Type-A tweets that are censored before we download them. Hence, the ratio in the downloaded tweets may be a biased measure of the ratio of reader exposure (the share of tweets read by users). Below, we provide a simple formulation to assess this potential bias.

Suppose that there are N Type-A tweets. Among these tweets, let α be the share of those that contain sensitive keywords, which are therefore automatically held in quarantine. A share s of these pass the censorship test and are later published; these published tweets are downloaded by us and also visible to users. The remaining tweets, which amount for a share of $1-\alpha$, do not contain sensitive keywords, and are directly published but later reviewed and perhaps censored. We download a share δ of these tweets. Suppose that u_A users would read these tweets if they were never censored, and that on average a share r of these users read the tweets before they are censored. Thus, we define a measure of user exposure to Type-A

content, e_A, to be:

$$e_A = N u_A \left(\alpha s + (1-\alpha) r\right).$$

The number of tweets that we download is

$$d_A = N \left(\alpha s + (1-\alpha) \delta\right).$$

Hence, the ratio of user exposure to downloaded tweets is

$$\frac{e_A}{d_A} = u_A \frac{\alpha s + (1-\alpha) r}{\alpha s + (1-\alpha) \delta}.$$

The equivalent ratio for Type-B content, which is not censored, is u_B, because, in this case, $\alpha = 0$ and $r = \delta = 1$. The relative ratios of user exposure to the number of downloaded tweets for Type-A content and Type-B content depend on the difference between r and δ as well as the discrepancy between u_A and u_B.

Empirical estimates To gauge the discrepancy between reader exposure and data downloading (r and δ), we need to know the share of interested users who read the tweets before they are censored, and the share of tweets that are downloaded before censoring. This requires estimates of the speed of (i) censoring, (ii) user reading, and (iii) our downloading of tweets. Let $\widehat{F}_c(t)$ be the empirical distribution of censoring time as measured by Zhu et al. (2010), who provide an accurate measure of the speed of censorship in minutes after a tweet is posted. Let $\widehat{F}_r(t)$ be the empirical distribution of retweet time in our data, which serves as a proxy for the speed of reading tweets. Finally, we assume that the downloading time, $\widehat{F}_d(t)$, is uniformly distributed between 0 and 24 hours. This assumption is reasonable because most of our data are downloaded by lining up users over a 24-hour time window and the posts of each of these users are downloaded at a daily frequency.

The empirical distributions of censoring and retweet speed are plotted in Figure A2. This figure includes two distributions of retweet speed for (i) all retweets and (ii) retweets of posts referencing protest and strike within two days after a real world event. In the first 10 minutes after a tweet is posted, retweeting (especially for the protest or strike events) is faster than censoring. After this time, the speed of censoring is considerably higher than that of retweeting. Protest and strike tweets are retweeted more quickly than the average tweet. This could be because users are more active in responding to politically sensitive content or because censorship disables the retweeting of sensitive posts after a certain time. For the latter reason, rather than using protest and strikes tweets, we use the retweet distribution based on all tweets, for which the share of censored posts is very small, to calculate the probabilities below.

We simulate data to generate samples of 1,000 tweets. For each simulation, we draw independent realizations of the censoring time (c) from the empirical distribution $\widehat{F}_c(t)$, the

download time (d) from $\widehat{F}_d(t)$, and 10 realizations of retweet time (r) from $\widehat{F}_r(t)$.[25] The simulated probability of downloading a tweet that is later censored is 0.23, which implies that approximately 77% of the posts that are eventually censored are censored before they are downloaded. Our simulation show that around one third of the users who would read a tweet have done so by the time of censorship.

To gauge the ratio of r to δ, suppose that there are 100 Type-A tweets susceptible to censorship and 100 uncensored Type-B content. Each tweet would be read by 10 users if censorship did not occur. In this case the Type-B tweets are read 1000 times and the Type-A tweets are read 330 times. We download 100 of the Type-B tweets and 23 of the type A tweets. Hence, the ratio of downloaded Type-A tweets relative to Type-B tweets is 23/100, and the corresponding ratio of user exposure to tweets is 330/1000.

According to our previous formulation, if no posts in category A contain sensitive keywords and are automatically held for review ($\alpha = 0$), then we would underestimate the prevalence of censored Type-A tweets relative to uncensored Type-B tweets by around 30% (1-23/33). In practice, we expect that a significant share of Type-A tweets contain logistic information which typically includes sensitive keywords, and thus the bias is smaller than when $\alpha = 0$.

The other issue is that the number of interested users may differ across content types ($u_A \neq u_B$). To capture this difference, we compare the numbers of retweets of Type-A and Type-B content, because the average number of retweets per tweet tends to increase in the number of users who read the tweet.

References

[1] Cattaneo, M. D., Crump, R. K., Farrell, M. H., & Feng, Y. (2022). On binscatter. arXiv:1902.09608v3.

[2] Moffitt, Robert A. "Policy interventions, low-level equilibria, and social interactions." Social dynamics 4.45-82 (2001): 6-17.

[3] Zhu, T., Phipps, D., Pridgen, A., Crandall, J. R., & Wallach, D. S. (2013). The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions. In USENIX Security Symposium (pp. 227-240).

[25] Since the protest and strike tweets are retweeted ten times on average.

Table A1. Summary statistics of the main variables

	Protest				Strike			
	mean	sd	min	max	mean	sd	min	max
Event dummy	0.002	0.039	0	1	0.002	0.045	0	1
#events 1-2 days prior (retweet weighted)	0.016	0.04	0	0.69	0.02	0.043	0	0.675
#events 1-2 days prior	0.003	0.056	0	2	0.004	0.065	0	2
#events 1-2 days prior (distance weighted)	0.003	0.007	0	0.145	0.004	0.008	0	0.233
Total # retweets by users in focal city in the last 182 days	1.026	1.019	0	3.495	1.021	0.948	0	3.303
log (#posts per capita +1)	0.06	0.185	0	2.646	0.059	0.18	0	2.646
log(Population)	5.965	0.635	3.767	8.119	5.874	0.684	2.871	8.119
log(GDP)	15.147	1.204	11.989	19.179	15.093	1.184	12.031	19.179
log(Agriculture share of GDP)	1.484	1.068	-3.219	4.078	1.455	1.076	-3.219	4.071
log(Industrial share of GDP)	3.898	0.264	2.123	4.504	3.903	0.277	2.161	4.504
log(Tertiary share of GDP)	3.703	0.266	2.153	4.454	3.693	0.276	2.153	4.454
log(#Cellphone users, 10,000)	5.422	0.846	2.509	8.124	5.362	0.851	2.786	8.124
#cities	248				282			
#observations	670996				713702			

Table A2. Between-city mobility of college students and social media connections

	(1)
VARIABLES	f_{ij}
$Students_{ij}^{0509}$	0.083***
	(0.013)
Inverse geographical distance	39.722***
	(4.223)
Observations	86,308
R-squared	0.938
Fixed effects	Prov pair + city-prov pair

Notes: The dependent variable is log(1+# of retweets from city i of tweets from city j). $Students_{ij}^{0509}$ is log(1+mean # students from city i who move to city j in 2005 or in 2009). The regression includes origin and destination city fixed effects, as well as fixed effects for province pairs and destination city-province pairs. Standard errors are clustered by origin city and destination city. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Student-mobility IV: First stage

VARIABLES	Protest $\bar{s}_{it-1}$	Strike $\bar{s}_{it-1}$
$Students_{ij}^{0509}$ *x period 0*	1.196***	1.249***
	(0.081)	(0.078)
$Students_{ij}^{0509}$ *x period 1*	1.414***	1.422***
	(0.078)	(0.073)
$Students_{ij}^{0509}$ *x period 2*	1.246***	1.494***
	(0.075)	(0.081)
Observations	999,472	1,092,477
R-squared	0.980	0.985

Notes: The dependent variable is $\bar{s}_{it-1}$, the lagged events in other cities, weighted by social media connections. The unit of observation is city by date. The regression includes lagged events, y_{it-1}, and distance-weighted lagged events, d_{it-1}, interacted with period fixed effects, city-by-period, and date fixed effects. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table A4. Testing for mechanical zero effects at low event frequency

VARIABLES	mean
β^0	0.106
β^1	0.121
sd(β^0)	0.038
sd(β^1)	0.021
Tests of $\beta^1 > \beta^0$ at 5% significance	0.089

Notes: Data was generated so that the simulated event frequencies equal observed event frequencies for the pre-Weibo and post-Weibo periods, while the DGP-beta is the same for both periods. The model of Table 2 is then estimated on the simulated data.

Table A5. Time-varying measure of connections (log model)

VARIABLES	(1) Protest	(2) Protest	(3) Protest	(4) Strike	(5) Strike	(6) Strike
Social-media spread (β)	0.055***	0.054***	0.047***	0.039***	0.037***	0.024***
	(0.013)	(0.012)	(0.013)	(0.008)	(0.007)	(0.008)
Geo-distance spread (γ)	-0.008	-0.007		0.031**	0.030**	
	(0.008)	(0.008)		(0.013)	(0.012)	
Number events 1-2 days	0.009**	0.009**	0.000	0.017***	0.017***	0.000
	(0.004)	(0.004)	(0.000)	(0.004)	(0.004)	(0.000)
Total number retweets	0.000	-0.000	0.000	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Weibo posts	0.006***	0.006***	0.005**	0.011***	0.010***	0.009***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)
Observations	670,996	670,996	670,996	713,702	713,702	713,702
R-squared	0.016	0.016	0.219	0.027	0.027	0.239
Controls	No	Yes	Yes	No	Yes	Yes
QPtest	0.07	0.15		0.27	0.32	

Notes: Results are from a linear regression of an event dummy. The unit of observation is city by date. The estimated model is

$$y_{it} = \alpha h(y_{it-1}) + \beta h(s_{it-1}) + \gamma h(d_{it-1}) + \beta_0 w_{it} + \theta' x_{it} \quad + \alpha_{ip} + \delta_t + \varepsilon_{it}$$

where $h(x) = \ln(5x + 1)$. Controls are population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell phone users. Columns (3) and (6) allow for arbitrary time-invariant heterogeneity in the spread across city pairs. The QP statistic reports the p-value of the test for serial correlation in the fixed-effects model. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table A6. Time-constant measure of connections (log model)

VARIABLES		(1) Protest	(2) Protest	(3) Strike	(4) Strike
Social-media spread,	β^0	0.036**	0.033*	0.023	0.028
		(0.018)	(0.020)	(0.039)	(0.043)
	β^1	0.214***	0.246***	0.178***	0.190***
		(0.055)	(0.058)	(0.042)	(0.046)
	β^2	0.100***	0.109***	0.237***	0.249***
		(0.027)	(0.028)	(0.032)	(0.035)
Geo-distance spread,	γ^0	-0.003	-0.010	0.113*	0.114*
		(0.018)	(0.018)	(0.066)	(0.067)
	γ^1	-0.028	-0.026	0.116**	0.113*
		(0.038)	(0.038)	(0.059)	(0.061)
	γ^2	0.026	0.024	0.030	0.031
		(0.026)	(0.027)	(0.035)	(0.034)
Observations		1,140,224	1,028,085	1,224,880	1,119,858
R-squared		0.018	0.019	0.048	0.049
Controls		No	Yes	No	Yes
P-value: $\beta^1 = \beta^0$		0.002	0.001	0.006	0.008
P-value: $\beta^2 = \beta^0$		0.048	0.032	0.000	0.000

Notes: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The estimated model is

$$y_{it} = \alpha^p h(y_{it-1}) + \beta^p h(\bar{s}_{it-1}) + \gamma^p h(d_{it-1}) + \beta_0 w_{it} + \theta' x_{it} \quad + \alpha_{ip} + \delta_t + \varepsilon_{it}$$

where $h(x) = \ln(5x + 1)$. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table A7. Event spread across cities: probit model

VARIABLES	(1) Protest	(2) Protest	(3) Strike	(4) Strike
Social-media spread (β)	0.618**	0.568**	0.558***	0.572***
	(0.271)	(0.254)	(0.179)	(0.181)
Geo-distance spread (γ)	1.727	1.927	2.129***	2.103***
	(1.433)	(1.392)	(0.643)	(0.653)
Number events 1-2 days prior	0.178**	0.168**	0.317***	0.319***
	(0.079)	(0.077)	(0.049)	(0.048)
Total number retweets	-0.032	-0.042	0.194***	0.203***
	(0.046)	(0.044)	(0.036)	(0.036)
Weibo posts	0.238***	0.232***	-0.009	-0.008
	(0.058)	(0.060)	(0.034)	(0.038)
Observations	564,378	564,378	581,509	581,509
Controls	No	Yes	No	Yes

Notes: Results are from a probit regression on an event dummy. The unit of observation is city by date. The regression includes city-fixed effects and a quadratic function of the time trend in date. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table A8a. Three-period analysis of protests, additional controls

VARIABLES		(1) Protest	(2) Protest	(3) Protest	(4) Protest	(5) Protest	(6) Protest	(7) Protest
Social-media spread, s_{it-1}	β^0	0.043**	0.029*	0.033*	0.033**	0.031*	0.028	0.044**
		(0.021)	(0.018)	(0.019)	(0.016)	(0.017)	(0.018)	(0.020)
	β^1	0.204***	0.168***	0.204***	0.179***	0.175**	0.173***	0.189***
		(0.061)	(0.045)	(0.056)	(0.046)	(0.045)	(0.045)	(0.065)
	β^2	0.068**	0.086***	0.090***	0.083***	0.085**	0.086***	0.082***
		(0.028)	(0.022)	(0.024)	(0.021)	(0.021)	(0.022)	(0.026)
Geo-distance spread, d_{it-1}	γ^0	-0.014	-0.012	-0.013	-0.015	-0.011	-0.012	-0.014
		(0.018)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
	γ^1	-0.034	-0.024	-0.033	-0.030	-0.019	-0.026	-0.027
		(0.037)	(0.035)	(0.037)	(0.037)	(0.034)	(0.036)	(0.038)
	γ^2	0.024	0.020	0.019	0.020	0.022	0.019	0.021
		(0.025)	(0.024)	(0.024)	(0.024)	(0.023)	(0.024)	(0.024)
Observations		1,022,162	1,022,162	1,022,162	1,022,162	1,022,1	1,022,162	1,022,162
R-squared		0.020	0.020	0.020	0.020	0.020	0.020	0.020
Interacted controls		All	Populatio n	GDP	Agricultu re share	Industri al share	Tertiary Share	Cell-phone
P-value: $\beta^1 = \beta^0$		0.016	0.004	0.004	0.003	0.003	0.003	0.034
P-value: $\beta^2 = \beta^0$		0.544	0.059	0.089	0.096	0.074	0.059	0.302

Notes: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period and date fixed effects. Default controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell phone users. Interacted controls are constructed as $\sum_j w_i w_j y_{jt-1}$, where the weights w_i are population, GDP, etc., for city i. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table A8b. Three-period analysis of strikes, additional controls

VARIABLES		(1) Strike	(2) Strike	(3) Strike	(4) Strike	(5) Strike	(6) Strike	(7) Strike
Social-media spread, s_{it-1}	β^0	-0.025	0.034	0.027	0.025	0.031	0.028	0.023
		(0.038)	(0.039)	(0.036)	(0.032)	(0.033)	(0.033)	(0.035)
	β^1	0.023	0.151***	0.129***	0.126***	0.135***	0.133***	0.092***
		(0.038)	(0.035)	(0.030)	(0.026)	(0.029)	(0.029)	(0.031)
	β^2	0.074***	0.109***	0.133***	0.125***	0.124***	0.125***	0.087***
		(0.026)	(0.017)	(0.022)	(0.017)	(0.017)	(0.017)	(0.021)
Geo-distance spread, d_{it-1}	γ^0	0.117	0.100	0.102	0.104	0.102	0.102	0.103
		(0.071)	(0.064)	(0.066)	(0.066)	(0.065)	(0.065)	(0.066)
	γ^1	0.112**	0.093*	0.100*	0.102*	0.100*	0.098*	0.106*
		(0.052)	(0.051)	(0.054)	(0.056)	(0.054)	(0.054)	(0.055)
	γ^2	0.050	0.043	0.048	0.047	0.054*	0.047	0.048
		(0.032)	(0.032)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
Observations		1,113,920	1,113,920	1,113,920	1,113,920	1,113,920	1,113,920	1,113,920
R-squared		0.050	0.050	0.050	0.050	0.050	0.050	0.050
Interacted controls		All	Population	GDP	Agriculture share	Industrial share	Tertiary Share	Cell-phone
P-value: $\beta^1 = \beta^0$		0.263	0.025	0.022	0.005	0.010	0.011	0.113
P-value: $\beta^2 = \beta^0$		0.020	0.055	0.002	0.001	0.004	0.002	0.091

Notes: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period and date fixed effects. Default controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Interacted controls are constructed as $\sum_j w_i w_j y_{jt-1}$, where the weights w_i are population, GDP, etc., for city i. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table A8c. Three-period analysis of protests, additional controls (log model)

VARIABLES		(1) Protest	(2) Protest	(3) Protest	(4) Protest	(5) Protest	(6) Protest	(7) Protest
Social-media spread, s_{it-1}	β^0	0.045*	0.036*	0.036*	0.035*	0.036*	0.036*	0.037*
		(0.024)	(0.020)	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)
	β^1	0.268***	0.244***	0.244***	0.249***	0.245***	0.244***	0.245***
		(0.069)	(0.057)	(0.057)	(0.060)	(0.057)	(0.057)	(0.057)
	β^2	0.109***	0.107***	0.107***	0.102***	0.107***	0.107***	0.108***
		(0.030)	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)	(0.028)
Geo-distance spread, d_{it-1}	γ^0	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011
		(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
	γ^1	-0.034	-0.029	-0.029	-0.031	-0.029	-0.029	-0.030
		(0.039)	(0.037)	(0.037)	(0.038)	(0.038)	(0.037)	(0.037)
	γ^2	0.023	0.023	0.023	0.025	0.023	0.023	0.023
		(0.026)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Observations		1,022,162	1,022,162	1,022,162	1,022,162	1,022,162	1,022,162	1,022,162
R-squared		0.019	0.019	0.019	0.019	0.019	0.019	0.019
Interacted controls		All	Population	GDP	Agriculture share	Industrial share	Tertiary Share	Cell-phone
P-value: $\beta^1 = \beta^0$		0.003	0.001	0.001	0.001	0.001	0.001	0.001
P-value: $\beta^2 = \beta^0$		0.157	0.051	0.051	0.059	0.052	0.051	0.055

Notes: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period and date fixed effects. Default controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Interacted controls are constructed as $\sum_j w_i w_j y_{jt-1}$, where the weights w_i are population, GDP, etc., for city i. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table A8d. Three-period analysis of strikes, additional controls (log model)

VARIABLES		(1) Strike	(2) Strike	(3) Strike	(4) Strike	(5) Strike	(6) Strike	(7) Strike
Social-media spread, s_{it-1}	β^0	-0.017	0.019	0.018	0.026	0.020	0.019	0.015
		(0.058)	(0.045)	(0.044)	(0.039)	(0.044)	(0.044)	(0.045)
	β^1	0.167***	0.183***	0.182***	0.179***	0.183***	0.181***	0.181***
		(0.041)	(0.042)	(0.042)	(0.041)	(0.042)	(0.042)	(0.042)
	β^2	0.258***	0.248***	0.248***	0.247***	0.249***	0.249***	0.249***
		(0.036)	(0.035)	(0.034)	(0.034)	(0.035)	(0.035)	(0.035)
Geo-distance spread, d_{it-1}	γ^0	0.120	0.112*	0.112*	0.110	0.111	0.112*	0.113*
		(0.074)	(0.068)	(0.068)	(0.068)	(0.067)	(0.068)	(0.068)
	γ^1	0.110*	0.106*	0.106*	0.107*	0.105*	0.106*	0.106*
		(0.058)	(0.057)	(0.057)	(0.058)	(0.057)	(0.057)	(0.057)
	γ^2	0.030	0.031	0.031	0.032	0.031	0.031	0.031
		(0.035)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Observations		1,113,920	1,113,920	1,113,920	1,113,920	1,113,920	1,113,920	1,113,920
R-squared		0.049	0.049	0.049	0.049	0.049	0.049	0.049
Interacted controls		All	Population	GDP	Agriculture share	Industrial share	Tertiary Share	Cell-phone
P-value: $\beta^1 = \beta^0$		0.007	0.006	0.005	0.002	0.005	0.006	0.006
P-value: $\beta^2 = \beta^0$		0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period and date fixed effects. Default controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Interacted controls are constructed as $\sum_j w_i w_j y_{jt-1}$, where the weights w_i are population, GDP, etc., for city i. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Table A9. Correlation between retweeting and censoring intensity

VARIABLES	(1) f_{ij}	(2) f_{ij}	(3) f_{ij}
Share deleted i	0.026	0.032	
	(0.030)	(0.035)	
Share deleted j	-0.000	-0.002	
	(0.016)	(0.018)	
Share deleted i x Weibo penetration i			0.209
			(0.210)
Share deleted j x Weibo penetration j			-0.116
			(0.100)
Observations	99,048	73,015	73,015
R-squared	0.750	0.803	0.803
City controls	No	Yes	Yes

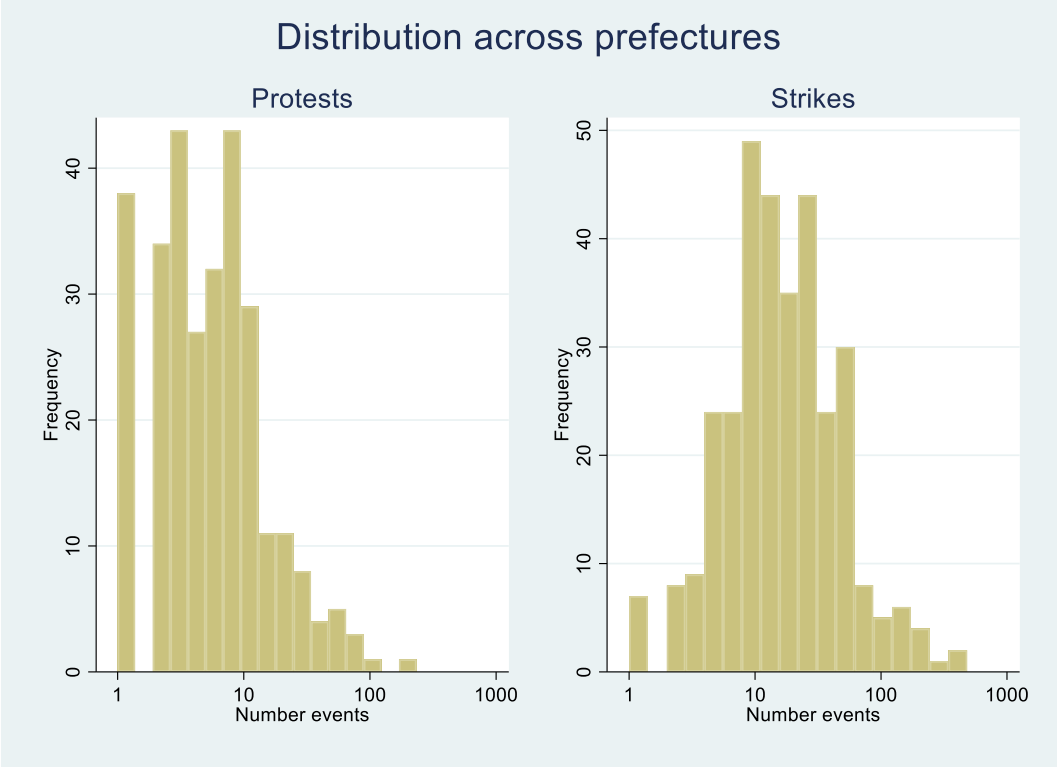
Notes: Results are from a linear regression of f_{ij} on the share of deleted Weibo posts in a city. All variables are standardized. The unit of observation is city i by city j. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors, clustered by city i and city j, in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A10. Time-varying measure of connections interacted with censoring intensity

VARIABLES	(1) Protest	(2) Protest	(3) Strike	(4) Strike
Social-media spread (β)	0.163***	0.162***	0.120***	0.116***
	(0.049)	(0.049)	(0.025)	(0.024)
Social-media spread (β) x share deleted posts	0.024	0.026	0.009	0.015
	(0.044)	(0.044)	(0.026)	(0.029)
Geo-distance spread (γ)	-0.031	-0.031	0.142**	0.137**
	(0.036)	(0.036)	(0.058)	(0.054)
Number events 1-2 days prior	0.015**	0.015**	0.030***	0.030***
	(0.007)	(0.007)	(0.007)	(0.006)
Total number retweets	0.002***	0.002***	0.002**	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Weibo posts	0.006***	0.006***	0.010***	0.009***
	(0.002)	(0.002)	(0.004)	(0.003)
Observations	670,996	670,996	713,702	713,702
R-squared	0.017	0.017	0.027	0.027
Controls	No	Yes	No	Yes

Notes: Results are from a linear regression of an event dummy. The unit of observation is city by date. The regression includes city-by-period and date fixed effects. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors are two-way clustered by date and city. *** p<0.01, ** p<0.05, * p<0.1.

Figure A1. Distribution of events across cities



Notes: The x-axis indicates the accumulated number of events within each city, and the y-axis indicates the frequency of cities. The sample period for protests is from 2006 to 2017 and that for strikes is from 2007 to 2017.

Figure A2. Speed of retweeting and censorship of posts about protests and strikes

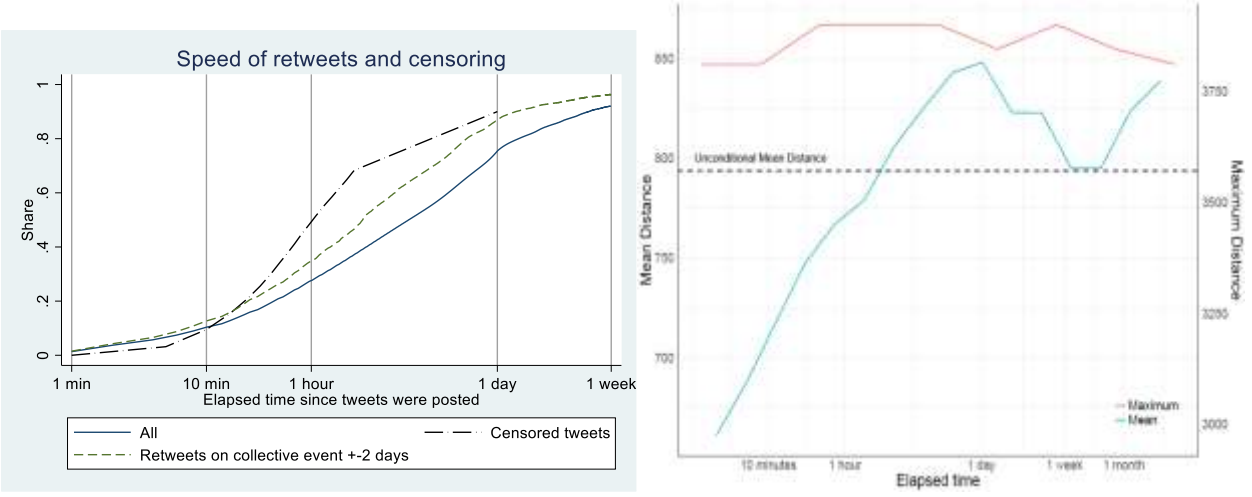
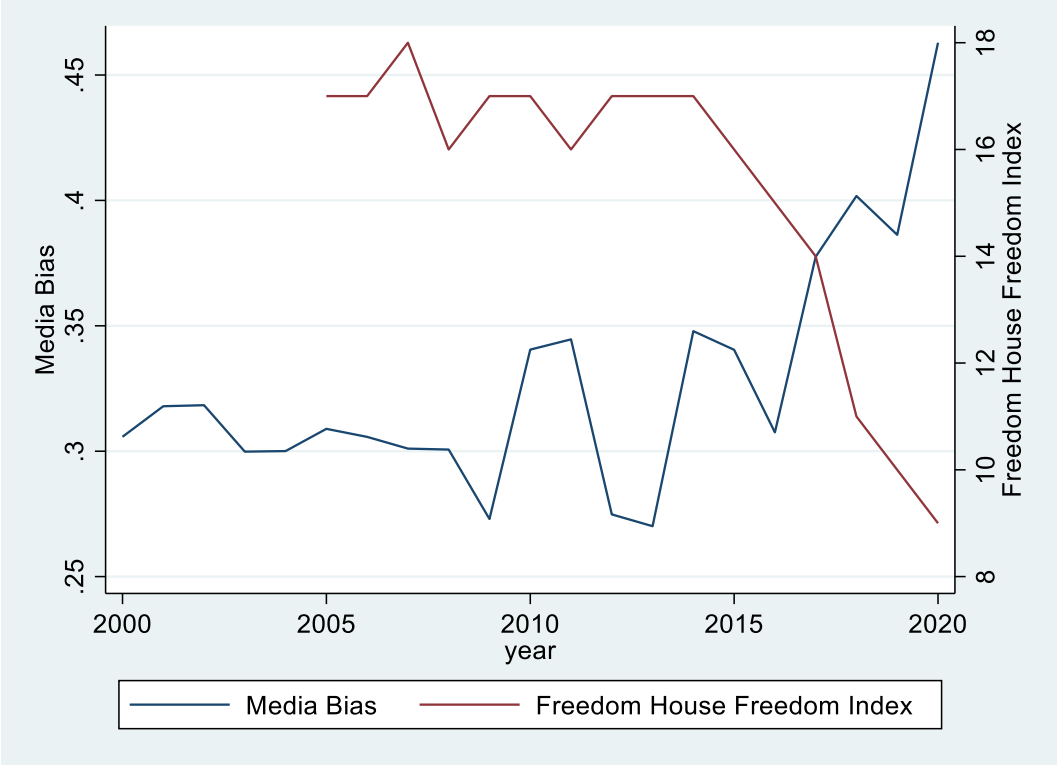
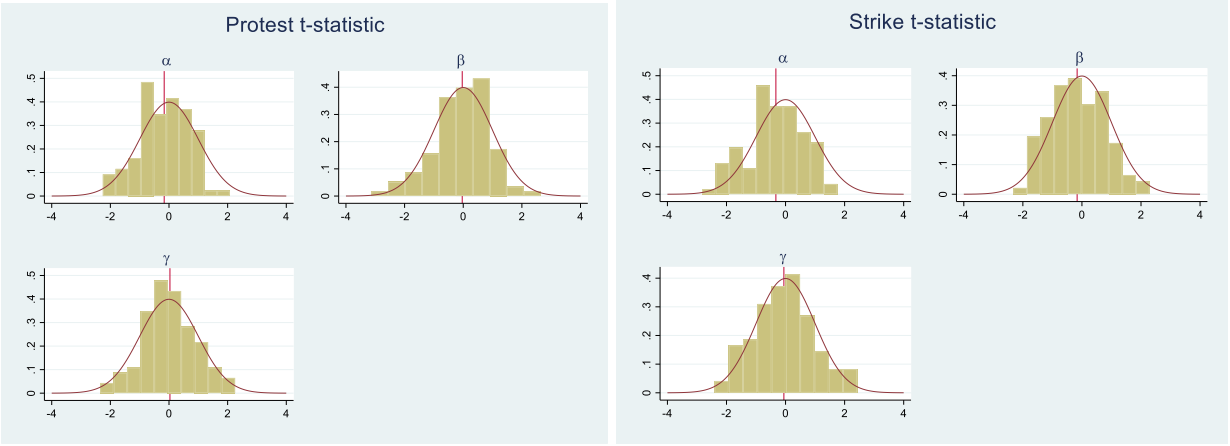


Figure A3. Media control in China during 2000-2020.



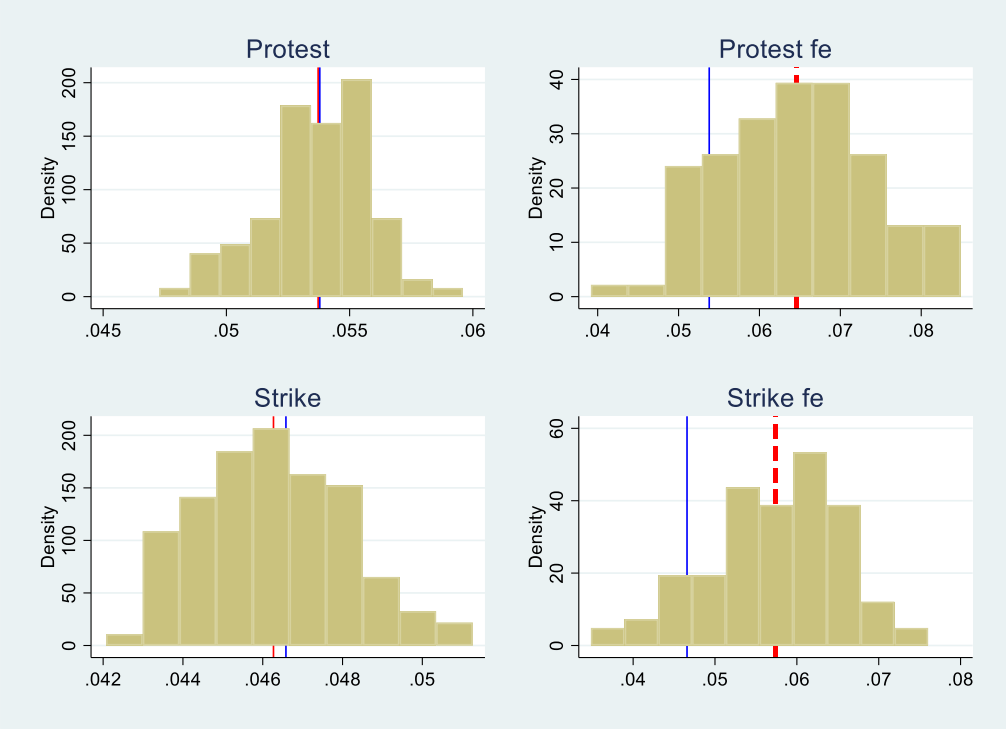
Notes: "Media Bias" is a measure of pro-government bias of Chinese newspapers based on the method used in Qin et al. (2018). The Freedom House Index is the aggregate score of freedom of expression and media freedom constructed by the Freedom House.

Figure A4. Monte Carlo simulations DGP, distribution of t-statistics



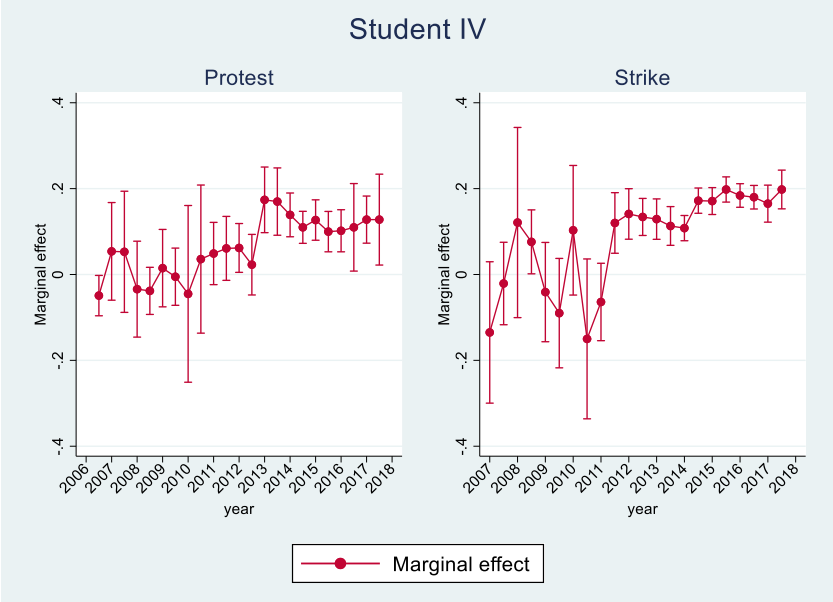
Notes: The graphs plot the distribution of t-statistics of coefficients of interest against the standard normal density. The red vertical line denotes the mean of the standardized t statistics. α , β, γ are coefficients of the three variables in equation (1): lagged event dummy in city i (y_{it-1}), the time-varying diffusion of information on protests through social media (s_{it-1}), the spread to geographically close cities (d_{it-1}).

Figure A5. Monte Carlo simulations of the fixed-effect model



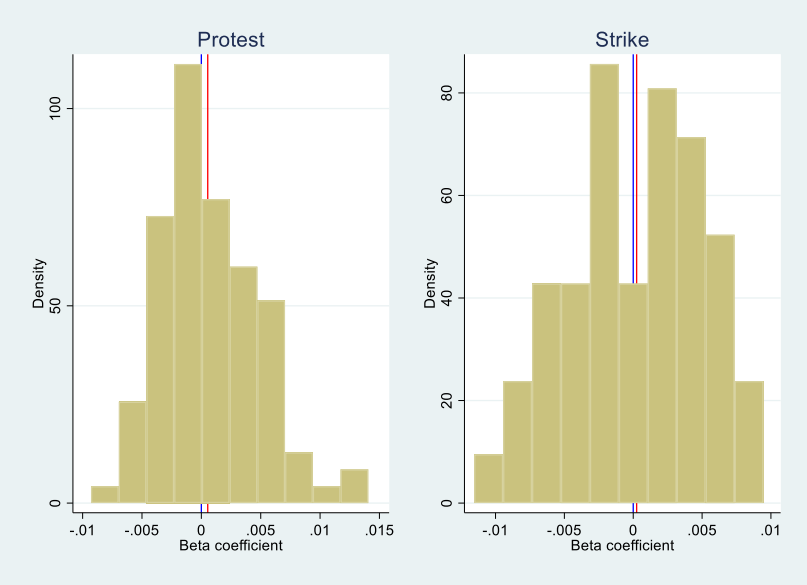
Notes: The graphs plot the distribution of β estimates. The blue line indicates the coefficient of the DGP. The red line indicates the mean estimated coefficient using the simulated data.

Figure A6. Student-IV dynamic effects



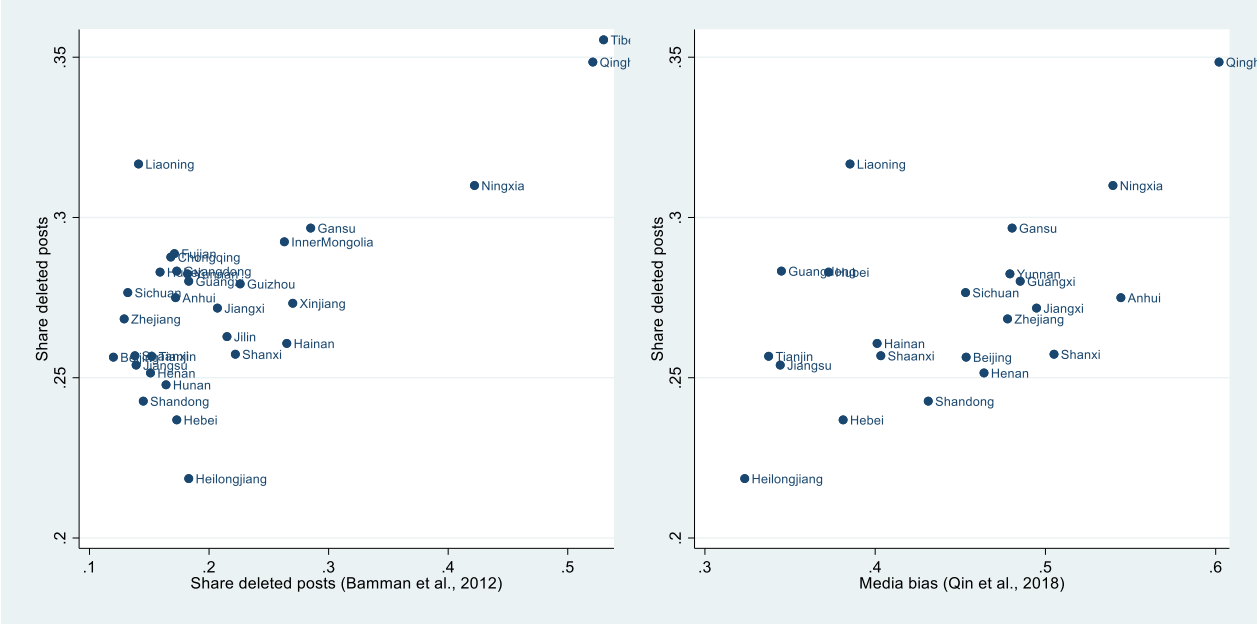
Notes: The graphs plot the estimated biannual coefficients on the time-constant measure $\bar{s}_{it-1}$ (Instrumented by the student-mobility variable z_{it-1}) relative to the pre-period mean. The bars indicate the 95% confidence interval.

Figure A7. Monte Carlo simulations with observability driven by Weibo and no network spread effect



Notes: The blue line is at the beta-coefficient of the DGP. The red line is at the mean estimated coefficient using the simulated data.

Figure A8. Local censoring intensity across provinces



Notes: The left graph plots the correlation between our measure of censoring intensity (the share of deleted posts) and the measure of the share of deleted posts by Bamman et al. (2012). The right graph plots a similar correlation with the x-variable replaced with a measure of pro-government media bias based on Qin et al. (2018).