Social Media and Protests in China

Bei Qin, David Strömberg, and Yanhui Wu*

March 2021

Abstract

This paper studies whether the explosive growth of social media in China affects the spread and incidence of protests. We combine a unique dataset of 13.2 billion microblog posts published during 2009-2013 with detailed information on thousands of protests and strikes during 2006-2017. We use retweets to measure the network of social media information flows across cities, and estimate the effects of this rapidly exapanding network. Despite the strict media control in China and a lack of explicit coordination, we find that the social media network has a sizeable and significant effect on the spread of both protests and strikes across cities. Tacit coordination and emotional reactions could drive these effects. While the spread of events over social media is fast and predominantly between similar events, we find that social media increases the scope of protests. These findings shed light on the recent debate regarding the political role of social media in autocracies.

*Qin: Hong Kong Baptist University, beiqin@hkbu.edu.hk; Stromberg: University of Stockholm, david.stromberg@ne.su.se; Wu: University of Hong Kong, yanhuiwu@hku.hk

1 Introduction

Social media are transforming political communication and interaction, and potentially more strongly so in non-democracies with little pre-existing media diversity. Because social media can be used both by and against authoritarian regimes, there is a heated debate on whether social media weakens or strengthens these (Shirky 2011; Morozov 2012). A focal question is whether social media facilitates the formation and dynamics of anti-regime collective action. Moreover, the recent rapid development of information technology and big data techniques has increased authoritarian governments' ability to exploit social media in their favor. In this paper, we study whether and how social media affects protests and strikes in China, the largest autocracy known for extensive censorship of social media and the extensive use of AI technology for information and social control.

Unraveling the effect of social media in China depends crucially on two factors. First, it depends on whether social media affects protest dynamics–whether social media makes events break geographical boundaries and generate far-reaching social movements. In China, protests of narrow scale and scope are allowed in part because they can inform the regime of unpopular policies and local corruption (e.g., Cai 2008, Lorenzten 2017). An obvious risk is that they will snowball into larger social movements that target national problems and challenge the regime.[1] This risk is likely to increase with information flow on social media, whose transmission is speedy and long-distanced.

Second, it depends on the mechanisms by which social media affect protests. Some content is highly valuable for protesters but not for regime surveillance and monitoring, such as logistic information to protesters on the ground and information about protest tactics. If this content is indispensable for widespread grass-roots collective action, then strategic censorship that removes it (e.g., King et al. 2013, 2014; Roberts 2018) will cripple social media as a means to affect protest dynamics without reducing its usefulness for surveillance and monitoring. Such a strategy will not be successful if the mere reporting from a protest help people to coordinate their protest activities (e.g., Barberà and Jackson, 2016; Little 2016) or make people so emotionally upset that they join the protest (e.g., Pasarelli and Tabellini 2017). In this case, the regime can only stop the spread by removing exactly the content that is useful for surveillance and monitoring, namely, reports of the cause of the protests, such as corruption, and expressions of anger and other public sentiment. In this case, a fundamental tradeoff between utilizing bottom-up information and maintaining political control arises.

To clearly understand the first factor, we focus our study on the effect of social media

[1] A cautionary example is Solidarity in Poland, which originated from limited demand for workers' rights and better economic conditions but soon developed into pervasive resistance that proved fatal to the regime. More recent examples involve social media. In 2013, small groups of environmentalists protested against the proposed destruction of Gezi Park in Istanbul, and exploded to attract millions of tweets on social media and eventually the participation of hundreds of thousands of people in Turkey. In 2013-14, after Ukrainian president, Viktor Yanukovych, failed to sign an agreement with the EU, opposition leaders posted calls for protests on Twitter and Facebook. Subsequently, protests quickly grew and spread widely, eventually culminating in the ouster of Yanukovych.

on the dynamics of protests and strikes, rather than on event incidence. To clarify, by the effect on incidence, we mean the impact of an exogenous shock on the probability that an event takes place at a particular location and time. By the effect on dynamics, we mean the impact on how much an event increases the probability of a subsequent event in another location within a certain time window. To estimate the causal effect on event dynamics, we use high-frequency information networks across cities based on large-scale textual data from Sina Weibo—the most prominent microblogging platform in China. We then investigate the mechanisms that drive the effect by analyzing the content of social media posts.

The first empirical question is whether information about protests and strikes circulates in Chinese social media at all, or whether this content is uniformly censored. In Qin et al. (2017), we examine the content of 13.2 billion posts published on Sina Weibo during 2009-2013 and find that there were approximately 4 million microblog posts mentioning protest or strikes. These posts predict real-world protests and strikes. Moreover, users who published these sensitive posts continue to post similar content. This suggests that the Chinese government's social-media-control strategy allows for sensitive content that is useful for surveillance, within a generally strict information environment.

To avoid protest waves, perhaps censoring instead blocks the visibility of this sensitive content in other regions and deletes posts that start to spread geographically. To investigate this, we take advantage of our unique data on forwards (retweets). The fact that a user forwards a message indicates that the user has read the message, so forwarding is a measure of information diffusion. We identify approximately 40 million forwards of the protest and strike posts. In a subset of 3 million forwards for which we have precise timing and location information, we find that information about protests and strikes spread fast, far, and wide. Approximately 30% of the forwards occur within one hour after the posting of the original messages, and 80% within one day. After one hour, the mean distance between the user who posts a message and the user who forwards it is more than 800 km. Evidently, Chinese social media are not only abundant in information about protests and strikes; the rapid diffusion of this information implies that social media may generate sweeping protest activities.

Estimating the effect of social media on the spread of real-world events is intricate. First, social media posts are likely to increase the observability of events to researchers. Second, cities with strong informational ties through social media typically also have strong ties through other information channels, such as phone calls and face-to-face meetings. Third, cities with strong social media ties tend to be similar in, for example, industry structure and hence exposed to similar shocks affecting the propensity to protest and strike. Therefore, it is challenging to separately identify the informational effect of social media on event spread. Related identification challenges have been extensively discussed in the studies of social networks, pioneered by Manski (1993). Existing 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).

We instead exploit the time-series variation in communication capacity arising from the rapid social media network expansion for identification. Specifically, we use forwards on topics other than protests and strikes to measure how information diffusion between cities increases with social media. We use two research designs. First, we use an estimator that allows for arbitrary time-invariant heterogeneity in the spread of events across city pairs. Second, we investigate how the spread of events has evolved across cities that eventually become closely connected through social media over time. Monte Carlo simulations justify the validity of our methods. We also show that, even if social media allow researchers to observe more events, our estimation of the effect of social media network expansion on event spread is not affected. The intuition is that increased observability per se does not affect the temporal clustering of events across cities.

The main findings of our empirical analysis are as follows. First, our two research designs consistently show that protests and strikes spread rapidly (within 2 days) through the social media network. We estimate that, because of information flows through social media, the occurrence of a protest within one prefecture in the last two days increases the probability of a protest taking place in any of the other prefectures by 26 percent. Relative to the mean event probability, the stimated effecst amount to an increase of 68 percent for protests and 32 percent for strikes. Both protests and strikes start spreading across cities that eventually become highly connected through social media in 2010-2011. For strikes, which are politically less sensitive, the spread continues at high levels throughout our entire sample period. In contrast, the spread of protests across cities connected by social media covaries with changes in political sensitivity and censoring strictness in China. The spreading effect is lower during 2012, the year of the 18th National Congress of the Chinese Communist Party (CCP) when Xi Jinping replaced Hu Jintao as General Secretary of CCP and falls significantly after 2014 when the censoring scheme became stricter as measured in terms of Press Freedom index by Freedom House.

Third, we find that social media increases the scope of protests and strikes. Not surprisingly, the spread of events is predominantly narrow in scope– predominantly, protests spread to other protests for the same cause and strikes spread to strikes in the same industry. However, we find that social media also induce significant, albeit smaller, spread of events across categories. Importantly, despite the smaller effect of each individual event, the aggregate effect of the across-category spread is larger than within-category spread, because the total number of events across all categories is much larger than within the same category.

Next, we investigate what mechanisms drive the spread of events through social media. A prominent mechanism in existing studies is that protesters use social media to organize events. For instance, Enikolopov et al. (2020) report that two-thirds of the Russian cities in their sample had social media communities created for the purpose of organizing protest

demonstrations.[2] However, our manual reading shows that posts that contain logistic information (e.g., specifying where and when to meet) is extremely rare, likely because of extensive censorship of this sort of content. Another potential mechanism is that social media carries information about protest tactics (e.g., what to demand and how to deal with governments), from which other protesters can learn about the benefits and costs of protests (e.g., Cai 2008; Little 2016; Chen and Suen 2016). If such a learning-tactics mechanism plays a significant role, we should observe a persistent effect on event spread through the social media network because effective tactics are unlikely to change in the short run and relevant information remains accessible for a long time. However, we find the estimated effects on either strikes or protests to be short lived, with the strong effect appearing within a week after the initial event and dying out after a month. Moreover, close reading of post content suggests that posts discussing protest tactics or government responses are scarce.

While we can rule out explicit organization and learning about tactics as important mechanisms, our empirical findings are consistent with other potential mechanisms, such as tacit coordination and emotional reaction. Tacit coordination arises when the mere visibility of events on social media induces people to protest simultaneously, and thus protesters have a higher chance of achieving their objectives and expect a lower risk of repression (Barberà and Jackson (2016); Edmond 2013; Little 2016). Protests can also spread through emotional reactions if angry people derive utility from protesting against social injustices (Pasarelli and Tabellini 2017). These two mechanisms are likely to generate short-lived effects. Furthermore, our manual reading shows that, among the posts that mention strikes and protests, a large share of them discuss the causes of events, such as corruption or wage arrears, and criticize governments for bad policies and misbehavior. Among the most-forwarded posts, many express anger and sympathy for protesters.

Our findings shed light on the censoring strategy in China. The Chinese way of censoring social media appears to be geared toward achieving the dual goals of utilizing bottom-up information and mitigating the risk of collective action movements. Information about social problems that cause protests and emotional reactions to them is allowed to be posted and to spread widely, making it possible to gauge whether sentiments are merely local or nationally prevalent. In contrast, content that is useful for protesters but not for the regime, such as calling for action and discussion of protest tactics, is deleted. Nevertheless, the content that is useful for surveillance and gauging public sentiment is sufficient to spread protests and strikes. In order to limit the spread of collective action, the government must silence discussion of causes and emotional reaction to social problems. The cost is losing the bottom-up information that is crucial for surveillance and monitoring.

[2] In the political science literature, Tucker et al. (2016) find that during the Turkish protests in Gezi Park in 2013 and the Maidan protests in Ukraine 2013-14, tweets provided logistic information instructing demonstrators to converge on particular locations, warning them when the police were approaching those locations, as well as advice on protest tactics, such as how to counteract the effects of tear gas. Similarly, Steinert-Threlkeld (2017) describe posts that provide logistic information during the Arab Spring.

This paper contributes to the emerging literature on how information and communication technology (ICT)–in particular social media–affects collective action and regime changes in non-democratic countries. A number of papers examine the political effects of the expansion of communication infrastructure, notably high-speed internet or mobile phone networks. For example, Manacorda and Tesei (2020) and Christensen and Garfias (2018) find that cell phone access affected protests in a panel of African countries. Guriev et al. (2021) find that 3G mobile networks reduce government approval. A couple of papers aim to identify the causal effects of social media on protests; see a recent discussion by Zhuravskaya et al (2020). Enikolopov et al. (2020) show that penetration of the dominant Russian online social network led to protests against election fraud in Russia. Ferguson and Molina (2020) find a positive effect of Facebook 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 stress the explicit-coordination role of social media, because they find ample social media content containing logistical communication among protesters on the ground and for tactical advice.

Our paper advances this literature in several important aspects. First, we study how social media affects political collective action in China, which is a considerably more strictly-controlled environment in which logistical communication and tactical advice is absent. Second, we shift focus from cross-sectional incidence to the dynamic interaction between geographically dispersed protests. This dynamic interaction is key to understanding whether local protests expand to widespread social movements that are consequential to 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. In this last regard, our paper is related to the recent research on how information affects participants' behavior in social movements (e.g., Cantoni et al. 2019; Bursztyn et al. 2020).

Our paper is also casts light on the theoretical literature on protests, government policy and regime change. Lohmann (1993) models how costly protests may help politicians avoid bad policies. Battaglini (2017) extends the analysis to the case where policy makers and protesters have partially divergent interests and argues that social media may reduce the informativeness of protests since better coordination means that they no longer act independently. Barberà and Jackson (2016) argue that social media helps coordinate revolt. Our findings suggest that endogenous censoring of social media, where the regime trades off the benefits and cost in the above models would be a useful extension of these models, related to the analysis in Lorentzen (2014).

Finally, our paper contributes to the literature on media control strategies in non-democratic

[3]For example, Gelbach et al (2016) note in their survey on the theory of nondemocratic politics: "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."

countries. Several studies take the revealed-preference approach and infer authoritarian governments' goals from their observed strategy of censoring social media (e.g., King et al. 2013, 2014, Qin et al. 2017). We extend this line of research by examining the effect of social media after censoring on real-world events, from which we infer that there is an unavoidable trade-off between the regime's dual goals. This has direct implications for the debate about whether social media strengthens authoritarian regimes.

The remainder of this paper proceeds as follows. Section 2 provides a brief description of the institutional background. Section 3 describes our data and provides descriptive statistics. Section 4 presents the empirical analysis and results. Section 5 concludes.

2 Background

2.1 Social Media in China

After an intense period of explosive growth, social media in China today is as vibrant and extensive as in any Western country. Since we will examine effects during this period of rapid growth, we now describe it shortly. With the Chinese government blocking Twitter and Facebook and strictly controlling domestic microblogging services, the use of social media was limited until Sina Weibo appeared in August 2009. Sina Weibo is a hybrid of Twitter and Facebook, allowing users to tweet and retweet short messages with embedded pictures or videos and send private messages and write comments. Between 2009 and 2012, Sina Weibo use increased exponentially. By 2010, Sina 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, where pre-existing levels of mobile phone use was higher. Over our main sample period, which ends in 2013, Sina Weibo was the dominant microblogging platform in China. Since then, it has lost ground to WeChat, a cellphone-based social networking service, but has remained an influential platform for public communication.

This extensive use of social media in China is coupled with extreme government control. To suppress the posting of unwanted content, tens of thousands of information officers and internet monitors police the internet to punish users posting this content and induce self-censoring (Chen and Ang 2011). To remove unwanted content, censorship on Chinese social media is regulated by the national Propaganda Department of the CCP and is implemented largely by private service providers who are registered in Beijing. The estimated extent of censorship of Sina Weibo ranges from 0.01 percent of posts published by a sample of prioritized users (Fu, Chan, and Chau 2013) to 13 percent of posts on collective action events such as protests and strikes (King et al. 2013). Even though censorship is pervasive, it is far from complete even on these topics.

The government also controls content by active posting on Sina Weibo. In Qin et al. (2017), we estimate that there are 600,000 government-affiliated accounts contributing 4% of

all posts regarding political and economic issues on Sina Weibo. The share of government users is higher in areas with more extensive censoring of social media posts and in areas where newspapers are more strictly controlled. Local governments also employ a large number of internet trolls that write posts involving cheerleading for China, the revolutionary history of the Communist Party, or other symbols of the regime (King et al., 2017).

2.2 Protests and Strikes

Although the 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 the protests is multi-faceted. Protests are often met with violent repression by the police or plainclothes thugs, and leaders are arrested (Lorentzen, 2017). Even if some protest demands are accommodated, organizers and active participants risk being arrested, losing their job and being under the close watch of the government. At the same time, concessions are frequent, as well as simply ignoring events viewed as sufficiently innocuous (Cai, 2010; O'Brien and Li, 2006; Su and He, 2010; Lee and Zhang, 2013). Central policies require local officials to handle collective action events strategically rather than simply suppressing all events and top CCP leaders have made public statements urging restraint in police handling of protests (Steinhardt 2017). In many anti-corruption protests, high-rank CCP officials were eventually sent to converse with the protesters so as to re-establish the public's trust in the government. In terms of task division, local governments are typically responsible for repression while central leaders deliver concessions (Cai, 2010).

Strikes in China are often aimed at violations of labor laws, such as wage arrears or illegal work conditions. Local government officials are responsible for enforcing these laws, but often lack resources and will to do so. Strikes and related protests are often the workers' only recourse and, in this sense, similar to protests. The reactions to strikes span from repressive acts, such as police arrests to concessions, such as mediation or worker compensation. Strikes are typically not regarded as sensitive by the Chinese government unless they escalate into violent events that threatens public security.

What explains the regime's relative tolerance of protests and strikes? The literature discusses a number of possible benefits to the regime from allowing local protests on narrow issues. China is large and diverse, and most political and economic decisions are decentralized to local governments. Protests are a costly and hence credible way for citizens to communicate their concerns and may help the central government to identify and correct policy oversights, gauge public sentiment, and monitor local officials (Lorentzen, 2017). Second, absolute suppression of collective action may generate distrust and undermine the legitimacy of the ruling party. Finally, some collective action events such as strikes may be welfare improving and even productivity enhancing if they result in fairer competition, redistribution and better working conditions (Cai, 2010; O'Brien and Li, 2006; Su and He, 2010; Lee and Zhang, 2013).

As long as the protests are contained within scope and geography, they pose no threat to the regime. The risk the regime faces when allowing local protests is that they exhibit a scale shift and evolve into better-organized political action and social movements that shifts the focus from local to national policies and leaders, undermining the legitimacy and trust in the CCP (Cai, 2008).

3 Data

We assemble a unique dataset combining detailed information about thousands of collective action events from 2006 to 2017 together with posts published on Sina Weibo over the period of 2009-2013. Data about protests and strikes in China are not available from any official sources and media coverage of these kinds of events in mainland China is rather limited. Hence, our collection of data must rely on sources outside of mainland China. Below, we explain how we collect these event data, and then provide a description of the data.

3.1 Protests

The data on protests are collected manually from the website of Radio Free Asia (RFA), which is a private non-profit international broadcasting corporation based in Washington DC. We focus on the news content published in Chinese. The news reported on the RFA website comes from news collected and written by RFA's special correspondents, as well as media outlets in mainland China, Hong Kong, Taiwan and Western media outlets such as the New York Times and the BBC. News produced by the RFA's special correspondents are widely used by Chinese news portals outside of mainland China. To the best of our knowledge, the RFA website is the most comprehensive and well-structured data source for protest events in mainland China. We searched key words related to "protest" and "demonstration" (in Chinese) on the RFA website and downloaded the relevant news reports. Several research assistants were hired to first verify that the news is indeed about events of interest and to purge duplicate information. Then, they extracted relevant information from each news report and coded the date, location, cause, and scale (number of participants) of each event.

There are in total 1153 protest events in our dataset between July 2006 and December 2013 when our main analysis is conducted, and an additional 1576 protests 2014-2017. While many of these instances are small-scale protests confined within certain localities, some of them are large scale, spreading across regions, and disruptive. For instance, an event that was widely reported by Western media is the Wukan Event in 2011 when thousands of farmers in a city in Guangdong province protested against corruption of local officials. The event led to direct confrontation between farmers and officials, violent conflicts between protesters and police, and demonstrations in multiple cities in support to the farmers.

As shown in Table 1, many protests concern governments (policy, police and court, and housing and land reforms) and livelihood issues (employment, environment, and health). In

terms of scale, more than 70% of the events involve hundreds or thousands of people. Geographically, these protest events span 224 prefectures, or around two-thirds of all prefectures in China. Many locations experience over ten events during our sample period, even though the distribution is right-skewed. Beijing is an outlier with 95 protests, and coastal areas in Shanghai and Guangdong and inland areas in Hebei, Shaanxi, Chongqing and Sichuan witnessed frequent occurrences of protests.[4]

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 and the enforcement of China's labor laws. The CLB has collected data on strikes in Mainland China since 2007. From 2011 and onward, this data contains detailed information on the timing, location, employers involved, industry, scale, worker action, and government responses for each event. For earlier strikes, we extracted this information from the annual reports published by CLB and supplemented this with data from Boxun, a widely-read Chinese website that is based in the US and specializes in political news.

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 period 2013 to 2016 was Wickedonna, a mass-event-tracing blog that manually searched Sina Weibo and other social media platforms.

In our dataset, there are 1558 strikes between January 2007 and December 2013, and an additional 7946 strikes 2014-2017. As shown in Table 2, strikes occur in a wide range of economic sectors, with concentration in manufacturing and transportation industries (including taxi). Over 60% of the strikes involve more than 100 people. The most common cause of strikes is to demand payment of wage arrears. Geographically, these strikes cover 265 prefectures, approximately 78% of all prefectures in China. Geographically, the developed coastal areas are over-represented, notably in Guangdong, Shanghai and Jiangsu, but a significant number of strikes occurred in some inland areas such as Chengdu and Chongqing.[5]

3.3 Social media

Our social media data was extracted from a database including 13.2 billion posts published on Sina Weibo from 2009 to 2013. The database was collected by Weibook Corp, which executed a massive data collection strategy to download blogpost from the 200-300 million active users they identified. 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

[4] Figure A1 in the Appendix shows the distribution of the number of protests per prefecture. The upper right panel of Figure A2 shows the total number of protests by prefecture.

[5] Figure A1 in the Appendix shows the distribution of the number of strikes per prefecture. The upper left panel of Figure A2 shows the total number of strikes by prefecture.

2–3 days, and the lowest tier downloaded on a weekly basis. The data include posts that are later censored. For each post, they provided the content, posting time, and user information (including self-reported location). According to our estimates, the Weibook dataset contains approximately 95% of the total posts published on Sina Weibo before 2012 when we have an alternative measure of the total post volume (Qin et al., 2017).

From this Weibook database, we obtain two datasets. One data set contains monthly Weibo penetration across prefectures. This consists of the aggregate number of blogposts per prefecture and month, based on all 13.2 billion posts in the Weibook data. The penetration data will be used to measure how the use of Sina Weibo was spreading across time and prefectures.[6]

The second data set contains individual microblog posts extracted from the 13.2 billion posts in the Weibook data. These posts are of two types: original posts and forwarding posts. The original posts are the 202 million blog posts that mention any of approximately 5,000 key words related to various social and political topics. The forwarding posts are the 133 million retweets of the original posts. We only have the direct forwards of the original post, not the forwards of these forwards. For each original and forwarding post, we obtain the textual content as well as the posting time, how many times it has been forwarded, and the location of its author.[7] This data will be used to measure how information about protests and strikes spread, and more generally to measure the expanding social media communication across Chinese cities.

Over time, there is a positive correlation between the incidence of protests/strikes and the Weibo penetration. Figure 1 shows the total number of protests and strikes per month along with the Weibo use per month. There were around six strikes per month in the period 2007-2010. The number of strikes rapidly rose to over 54 per month in 2013. The pattern for protests is similar, with around 3 protests per month until 2010, followed by a rapid increase to around 55 protests per month in 2013. The green dots show the number of Weibo posts per capita, which increased drastically after early 2010. This trend of increasing protests and strikes has not gone unnoticed. It has been commented on in numerous news sources including the BBC, CNN, the New York Times and the Washington Post.[8] Of course, this relationship between social media and events may not be causal. The increase in strikes could be driven by other factors, such a slowdown in Chinese exports (Campante et al., 2019) and by increased observability of the events because of social media.

[6] The geographical distribution of social media posts is shown in the bottom panel of Figure A2.

[7] Details about Weibook's data collection strategy and our selection of keywords and extraction of posts can be found in Qin et al. (2017).

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

3.4 Social media spread of protest and strike information

What content about protests and strikes is found on social media in China obviously depends on censoring. One possible strategy would be blanket censoring of all posts involving these events, another to allow the posting but not the spread of information across the country, a third to limit content to that useful for surveillance. We investigate the first two strategies here and the last in section , that discusses mechanisms behind event spread.

As described in Qin et al. (2017), there is extensive coverage of protests and strikes on Sina Weibo. From the original post dataset, we find 2.5 million posts mentioning keywords related to protest and 1.3 million posts mentioning keywords related strikes. To characterize these posts, we drew a random sample of 1000 posts in each category and read them manually. Around 30 percent of the posts are indeed about protests and strikes. As shown in Qin et al. (2017), these posts predict and identify the real-world protests and strikes. Hence, there is no blanket censoring of protest and strike posts.

The information in these posts spreads to many other users. The post data contain a variable measuring the total number of times each post was forwarded. Based on this variable, we observe that the 3.8 million protest and strike posts are forwarded in 37 million retweets, which implies that on average, there are around ten forwards per original post. This average masks a huge dispersion: some posts are forwarded millions of times and others none. Conditional on being forwarded at least once, there are on average 50 forwards per strike post and 100 forwards per protest post.

The information also spreads fast, far and wide. We investigate this using the data on individual retweets, exploiting the time and user location of the original post and the retweet. Since we only have the direct forwards of the original post, our dataset contains around 3 million of the 37 million forwards. For protests and strikes, around 30% of these forwards occur within 1 hour of the original post and 80% of the forwards within one day of the original post, see the left panel of Figure A5. Forwards within the first hour are more likely to be geographically close. After that, distance plays no role: the average distance between the user who posted the original post and the user who forwarded it is the same as the unconditional distance between users. In general, the forwarding data shows that information about protests and strikes on social media disperses rapidly and widely across China, see the right panel of Figure A5.

Documenting the fast and wide spread of protest and strike information is important in its own right. Even if all these events are just local, reports about them spread out geographically across China. The accumulation of such reports could be detrimental for the government's legitimacy. Still, it is allowed.

We will use forwards across all topics to measure how information flows across prefectures change as the use of Sina Weibo expands. A post being forwarded implies that someone has read the post and decided to forward it. Of course, many others may read the post without

forwarding it, so our forwarding measure is a conservative measure of information spread.[9] We will proxy the information flow from city i to city j by the number of posts by users in city i that are forwarded by users in city j. We use the 133 million forwards across all topics except the 3 million forwards of posts discussing protests or strikes.

We find that cities that are close in different ways are more strongly connected through the social media network. Geographical closeness is most important, but cities with similar population density and tertiary share are also more closely connected (see Figure A4). People tend to form links with others who are like them, a network property called homophily. For this reason, information from other sources is likely correlated with information through the social media network. In addition, errors are likely to be correlated across cities that are more connected through the social media network. For example, even absent social media, strikes are likely to erupt at similar points in time in areas which are tightly connected through the social media network, because these areas share a similar industry structure. Our empirical design aims to address this concern.

3.5 An example

We end this section with one concrete example. In Guangzhou, Guangdong, a male worker wearing a bomb west staged a protest against wage arrears in a company in the afternoon of January 18, 2013. Later, he detonated the bomb, causing the death of one person and injuring seven people seriously. Immediately after the event, many Weibo posts described and commented the event, including many posts from direct witnesses. We identify in total 374 Weibo posts mentioning our protest-related key words on that day from Guangzhou, Guangdong, and 261 of them are talking about this event. These posts were first forwarded by Weibo users in nine prefectures tightly connected to Guangzhou by our forwarding measure.[10] Most retweets express sympathy for the worker, condemn employers who default on wages, and decry the government who disregards citizens' rights and allow wage arrears.

Among the nine prefectures who first forwarded the event posts, three incurred protests in the subsequent two days. In Shanghai, thousands of workers from a Sino-Japan joint venture protested for the unfair new labor rules, detaining 18 senior managerial persons, and the process lasted until over 400 policemen broke into the factory two days later. In Shenzhen, hundreds of citizens went to the street, protesting against a polluting LCD factory being constructed in their neighborhood. In Shanwei, thousands of villagers protested in front of the city government asking for land back, which was taken by the government without appropriate compensation. These protests spread across provinces, and although targeting

[9] A much less conservative measure is to use followers, since many followers will not read each blog post. One of the most cited Twitter studies (Kwak et al., 2010) asks whether the number of followers or the number of retweets is a better measure of influence and settles on retweets.

[10] Controlling for the location and time fixed effects, the ranking percentile of the forwarding posts between them and Guangzhou Guangdong are 1.19% (Shanwei Guangdong), 2.3% (Shenzhen Guangdong), 4.36% (Wuhan Hubei), 5.03% (Shanghai), 5.3% (Hangzhou Zhejiang), 5.77% (Chengdu Sichuan), 6.25% (Xi'an Shaanxi), 6.98% (Zhengzhou, Henan), 8.62% (Qingdao Shandong).

different causes, they all involve people who have had enough and protest against perceived injustices. Perhaps the later protests were inspired by the earlier ones, but there is no direct evidence of explicit organization or coordination across events.

We now investigate whether this episode was an isolated random incident or part of a large-scale systematic pattern. Do strikes and protests start to spread more after social media use expands, and exactly across those cities who are closely connected through social media?

4 Empirical analysis

We describe our baseline specification in Section 4.1 and corresponding results in Section 4.2. The following Section 4.3 analyzes how the event spread through social media has evolved over time. Section 4.4 studies whether social media increase the scope of event waves by spreading protests across causes and strikes across industries. Section 4.5 evaluates possible mechanism behind the spread. In Section 4.6, we collapse the data to the level at which we observe aggregate micro-blog penetration (prefecture-by-month) and describe how the incidence of protests and strikes is associated with this penetration.

4.1 Specification

We analyze whether information on Sina Weibo affects the spread of protests and strikes across Chinese prefectures (cities) using a panel of N cities at a daily frequency, t. Since the specification is the same for both types of events, we will in this section use protests as example. Let y_{it} be a dummy variable for protest occurrence. Suppose that the probability of a protest in city i at day t, $\Pr(y_{it})$ depends on the number of people there who are informed about protests y_{jt-1} in another city j at time $t-1$. Let f_{ijt} be the number of people who learn about the protest through social media and c_{ij} be the number informed from other sources, depending on time-invariant factors such as geographical distance. The model of protest spread from location j to location i is then

$$\Pr\left(y_{it}\right) = \left(\beta f_{ijt} + \gamma c_{ij}\right) y_{jt-1}. \tag{1}$$

As information flows, f_{ijt}, over the social media network between locations i and j increase, protests will spread more from location j to i. In addition, there is the time-invariant propensity for events to spread between location captured by c_{ij}.

Protests can spread simultaneously from multiple previous protest locations. The marginal effect of an additional protest may be constant. However, if the mechanism is through information, then it is likely that the marginal effect from additional events is falling. We will capture this by the function $h(\cdot)$, specified below.

Social media may also directly affect the incidence of protests (not just via spread from other locations). For example, this could happen because people learn about problems such

as corruption or wage arrears or because social media increases the visibility and hence effectiveness of protests. To capture this, we add Sina Weibo penetration, w_{it} to the above model. Furthermore, we add x_{it}, a set of controls. We also include an auto-regressive term and add time and prefecture fixed effects. We then have the following baseline specification:

$$\Pr(y_{it}) = \alpha h(y_{it-1}) + \beta h(\sum_{i \neq j} f_{ijt-1} y_{jt-1}) + \gamma h(\sum_{i \neq j} d_{ij} y_{jt-1}) + \beta_0 w_{it} + \theta' x_{it} + \delta_t + \delta_i. \quad (2)$$

Here, y_{it} is a binary event dummy, y_{jt-1} is the number of events in location j within two days prior to t, and, δ_t and δ_i are time and prefecture fixed effects.

The measures of information flows between cities, f_{ijt-1} and c_{ij} are key to our analysis of protest spread. Our variable measuring information flows over social media, f_{ijt-1}, is the log of one plus the number of forwards by users in i by posts from users in j on all subjects other than protests and strikes for the past 6 months up until one week before day t.[11] We exclude forwards of posts about protests and strikes and forwards less than one week before day t to avoid reverse causality that may arise, for example, if people who are planning protests and strikes are more likely to forward posts about protests and strikes. We also allow for distance related spread, by weighting past events by d_{ij}, the inverse geographic distance between cities i and j. Below, we will allow for arbitrary time-constant spread c_{ij} across locations. The matrices capturing information flows, F with the element f_{ijt-1} and D with the distance element d_{ij} are normalized to increase interpretability. F is arranged as T stacked $N \times N$ matrices F_t, and hence has $N \times T$ rows and N columns. We normalize these matrices so that the average sum of all elements in a row of a weighting matrix equals one.[12]

Weibo penetration, w_{it}, is measured as the log of one plus the number of Sina Weibo posts per capita. As mentioned, this includes the total number of blog posts per prefecture and month, based on all 13.2 billion blog posts in the Weibook dataset. The set of controls, x_{it}, includes the log of one plus the sum of all forwards by users in location i by posts from all users j in all other locations, population, GDP, tertiary share, industrial share, and the number of cell phone users and landline users.

We will use two functional forms $h(x) = x$ and $h(x) = \ln(5x + 1)$. In our data, the values of x are predominantly small. For these values, two functions are similar (see Figure A6) and hence yield similar estimated marginal effects. However, the log transformation implies smaller marginal effects at very high levels of x, which ensures that the Monte Carlo

[11] Using all cumulated forwards or a 4-month window produces qualitatively similar results. Even though we have millions of forwards, some averaging over time is necessary since we are estimating over 90,000 city-pair connections. A longer time-window increases precision but reduces time-variation in the connectedness across cities.

[12] A common form of normalization is row-normalization by which all coefficients in a row sum to one. However, in our case this normalization would imply that all locations would be equally affected (on average) by protests in other locations, which is clearly incorrect (e.g. Elhorst, 2001, argues against row-normalization of distance matrices for analogous reasons). Our weighting maintains the relative magnitude between all elements in the weighting matrix. In a constant $N \times N$ matrix, such as the distance matrix, the total sum of all elements is N and in a$(N \times T) \times N$ matrix, such as the forwarding matrix, this sum is $N \times T$.

simulations produce non-exploding paths. In contrast, the linear function produce exploding paths that are inconsistent with the bounded increases in event occurrence that we see in the real data. For this reason, we use the log transformation in the Monte Carlo simulations and present the results using the log model here, while the results from the linear model are presented in the appendix.

Several econometric issues require consideration. First, logistic and probit models are biased in rare events data (King and Zeng, 2001) and do not work well in panel data with a large set of fixed effects. Thus, we estimate a linear probability model that is immune to these problems. Second, our model includes location-fixed effects and lagged dependent variables. In general, the estimates in this type of model are inconsistent with T fixed (e.g. Hsiao 2014, Arrelano 2003, Baltagi 2008). In the current model, T is large and the bias is likely to be small. We will show that the bias is indeed small using Mont Carlo simulations.[13] Third, we need to check whether the estimated process is stationary.[14] Fourth, consistency requires no serial auto-correlation in the errors. Serial correlated in the error term implies that ε_{it} will be correlated with y_{it-1}. We will test for serial autocorrelation in the first-differenced residuals. Finally, errors may be correlated across both time and spatial units. We use two-way clustering of errors in the time and spatial dimensions.

4.2 Results

We estimate how protests and strikes spread across Chinese cities, using the specification in Equation (2) in our panel of prefectures with at least one protest or strike, for these respective outcomes. Table 3 shows the summary statistics for the protest and strike samples.

The results from our baseline regressions, specified in equation (2) are shown in Table 4. The first two columns show the results for protest and the last two for strikes. Our main interest is in the coefficient on the variable with lagged events, weighted by the amount of forwarding of posts from city j by users from city i. This is significant and positive across specifications. The coefficient estimate is not much affected by the inclusion of controls; see columns II and IV.

In terms of effect magnitudes, the estimates in column 2 imply that a protest in a given location the previous two days increased the expected number of locations with protests this day by 0.26, while the equivalent number for strikes, based on the estimate in column 4, is 0.17.[15] Relative to the mean event probability, this is an increase of 68 percent for

[13] If the bias was larger, one could address this issue by using the GMM estimators of Arrelano and Bond (1991) or Blundell and Bond (1998). However, instrumenting rare events, such as our protests and strikes, with lagged differences and levels is not likely to perform well.

[14] Sufficient analytical conditions for stationarity are that the errors are not auto-correlated and that $|\beta\omega_{F_t,\max}| + |\gamma\omega_{D,\max}| < 1$, where $\omega_{D,\max}$ and $\omega_{F_t,\max}$are the largest eigenvalue of the matrices D and F respectively (largest negative if $\gamma < 0$ or $\delta < 0$). We will evaluate this criterion and check whether the process is stationary in Monte Carlo simulations.

[15] For the case $h(x) = \ln(5x + 1)$, the marginal effect equals $\beta s f_{ijt-1}$. The scaling factor s equals 4.56 for strikes and 4.63 for protest. So, the marginal effect of equals $4.56{*}0.038 = 0.17$. for strikes and $4.63{*}.055 = .25$ for protsts. These effects are somewhat larger than the corresponding estimates from the linear model reported

protests and 32 percent for strikes.[16] These reported numbers are averages that conceal a considerable underlying heterogeneity, since the marginal spread-effect is proportional to forwarding, f_{ijt-1}, which is zero before the entry of Sina Weibo in 2009 and rapidly increasing thereafter. The marginal effects at high levels of forwarding in the later part of the sample period are much higher than the averages, and discussed in Section 4.3.

In addition, the incidence of observed protests and strikes are both increasing in Weibo penetration. The results suggest that strikes, but not protests, spread to nearby locations as the distance-weighted effect of lagged strikes is positive. Both protests and strikes are autoregressive processes, so that having an event during the last two days significantly increases the predicted probability of an event occurring in the same location. However, this only affects one location and is much less important for the overall dynamics of protest and strike waves than the spread across locations.

To assess the Nickell bias in our coefficient estimates, we run a set of Monte Carlo simulations. The bias is very small, the difference between the true and the mean estimated β is in the third value digit for both protests and strikes, see Appendix A.4.

Observability Social media was used as a source to identify protests and strikes. This is obviously likely to bias our estimate of β_0, the effect of social media penetration on events. However, the estimate of β, capturing the effect of social media on the spread of events across prefectures, is less likely to be biased by observability. Suppose that events do not spread across locations connected by social media, so that the probability of an event in location i is independent of whether there was an event in a connected location j in the previous two days. Suppose further that we observe only a share of the events in location i (and similarly in location j). Reporting on social media may make more events observable, but it will do so whether there was an event in location j the past two days or not. Hence, social media will increase the number of observed events, but not the probability that one observed event occurs just two days after an another relative to other days. In this case, a regression of event incidence in location i on incidence in a location j within two days will yield a zero coefficient. A bias could arise, for example, if the social-media induced observability increases and falls with two-day intervals but not by a general increase in observability.

The real situation is a bit more complicated than this simple example, since events may spread across cities i and j for other reasons such as geographic proximity. We use Monte Carlo simulations to verify that we can consistently test our null hypothesis of no effects of social media on event spread, even if event observability is increasing in Weibo penetration. In particular, we simulated event data, y_{it}, using the estimated parameters from Equation 2, except that we set $\beta = 0$ so that there is no spread of events through the social media network. We assume that the probability of observing a simulated event, p_{it}, increases linearly

in columns II and IV of Table A1, which are .17 for protests and .12 for strikes. See Appendix A.1 for details.

[16] The mean of the protest incidence is .0015 and there are 246 other prefectures than the one where the first strike took place, .25/(246*.0015)=.68. For strikes, the corresponding number is .17/(281*.0019)=0.32.

in Weibo penetration, w_{it}, adjusting the size of the effect so that Weibo penetration increases observability by 30%. This is higher than what would be implied if the estimated coefficient on Weibo penetration, $\widehat{\beta}_0$, only reflected increased observability. We then draw a set of observed events $\widetilde{y}_{it}$ with probability p_{it} from the simulated events y_{it}. We finally estimate the model in Equation 2 on the observed simulated events, $\widetilde{y}_{it}$. Figure 2 shows that no spurious correlation in spread is generated because social media increase observability. Hence, we conclude that we can consistently test the hypothesis $\beta = 0$, even if social media affects observability.

Homophily and correlated shocks Another identification concern is that it is difficult to identify the role of social media in spreading events if information from other communication channels or shocks are correlated with the social media network. We showed in the previous section that the social media network exhibits homophily: cities that are similar in different ways are more connected through the social media network. Consequently, strikes and protests may spread more between cities that are closely connected in the social media network even though social media play no role, as these cities may have strong information flows through other channels as well as correlated shocks. Such shocks would make ε_{it} correlated with the lagged outcome in other cities y_{jt-1}. A large literature, pioneered by Manski (1993), has discussed this issue and related identification problems. Solutions include exploiting the network structure to identify instruments (e.g. Bramoulle et al. 2009), of identifying instruments that are uncorrelated with the error terms and the network (e.g. Acemoglu et al., 2015, and König et al., 2017), matching (Aral et al., 2009) and explicit randomization (e.g. Bakshy et al., 2012).

We instead use time-series variation in the rapidly expanding network to identify effects, by adapting standard difference-in-difference and event-study methods to estimate network interactions. We will allow strikes and protest to spread through channels that are correlated with the network and allow that errors are correlated with the network, but assume that these correlations are constant during our short sample period. We assume that, although the average spread of strikes between cities may be related to the average network strength, the exact timing of the network expansion is unrelated to other changes in the spread of events across cities.

We explore three different designs. Our first solution was to estimate a model of dyads of locations with time-constant dyad-fixed coefficients $\beta_{i,j}$. However, Monte Carlo simulations showed that this method produced strongly (Nickell) biased estimates and hence this approach was abandoned.

Time-invariant heterogeneity in event spread across city pairs Our second solution is to use a model that allows for arbitrary time-invariant heterogeneity in event spread across location pairs. We implement this by including a fixed effect for $h(y_{jt-1})$ for each $j \neq i$, in the equation for each $\Pr(y_i)$ Two potential issues arise with this specification. First, Monte Carlo

simulations show that a (Nickel) bias is induced by including these fixed effects, although this bias is small; see Appendix A.4. Second, the auto-correlation test we use is not well suited for this case; see Appendix A.4.

The results from this specification are shown in Table 5. The pure auto-regressive and the distance-weighted effects are not identified in this specification. The estimates of β remain highly significant. The coefficient magnitude is somewhat smaller for protests. For strikes the coefficient is reduced from .037 to .024, perhaps because correlated shocks or communication channels is a bigger concern.

4.3 Effects over time

Our third solution aims to produce an event-study type analysis of the spread across cities connected through social media. We now estimate a model as in equation (2), but with the actual forwarding between city pairs at time t, f_{ijt-1}, replaced by the average, $\overline{f}_{ij}$, and the constant parameters β and γ replaced by the time-varying paramters β_p and γ_p, respectively, and where p indexes time periods.[17] Our baseline specification in equation (2) assumes that the spread of events across social media is increasing communication over social media, measured by forwarding across cities, f_{ijt}. By fixing $\overline{f}_{ij}$ at the average forwarding in location i of posts written in location j, we can instead let the data tell us when protests and strikes start spreading across cities that eventually develop strong social media ties. This allows us to more directly measure how the effects of social media on event spread change over time, as well as to check if events were spreading across these cities even before social media was introduced.

This specification also allows us to investigate social media effects after 2013, under the assumption that the average forwarding across locations in 2013 is a good proxy for average forwarding 2014-2017. This is likely the case. By 2013, the social media network was past its rapid expansion phase and had stabilized. We can evaluate the stability of the network connections after 2011 by investigating how much our measure of social media connectedness change in the last two years of our sample. A regression of the number of forwards between locations for the past six months before the last day of 2013 and the same variable the last day of 2011 has an R2 of 0.96. A significant number of events are added when extending the end of the sample period to 2017. Between 2013 and 2017 there are 1576 protests and 7946 strikes in our data.

We do this for two different time frequencies. We first estimate separate coefficients on event spread across cities for each biannual period. We then estimate a more demanding model that includes separate location-fixed effects and coefficients on our control variables for each period. To increase power, we split the sample into just three periods: before 2010, 2010-2013 and 2014-2017. The first period is before the entry of Sina Weibo. The second

[17] To make the coeffcients comparable to those in Tables 4 and 5, we normalized so that the sum of all elements $\overline{f}_{ij}$ in the sample 2010-2013 equals the sum of all elements of f_{ijt} for this period.

period is that for which we have data from Sina Weibo and the last is the period after this.

The resulting estimate of β_p contains both the causal effect of the expanding social media network, as well as any bias created by spurious correlations from spread through pre-existing communication channels and shocks correlated with the network. Before August 2009, Sina Weibo did not exist and hence for this period β_p only contains the bias. This means that we can estimate the causal effect as the increase in β_p after the pre-period. The key identifying assumption is that the correlation between on one hand the average network $\overline{f}_{ij}$ and on the other hand communication channels c_{ij} and shocks, is constant over time.

Biannual periods Figure 3 plots the estimated marginal effects for the biannual periods.[18] For both protests and strikes the spread increases rapidly across cities that eventually become highly connected through social media in 2010-2011 when actual social media use takes off (see Figure 1). The spread of protests across connected cities drops significantly in 2012, bounces back to its highest level around 2013 and then trends downward after 2014. This coincides with changes in political sensitivity and censorship tightness. In 2012, protests were highly politically sensitive because it of the 18th national congress of the CPC when Xi Jinping replaced Hu Jinato as General Secretary of the Communist Party. In addition, social media regulation was tightened by the real name reform this year, causing a temporary drop in social media use. During our early sample period 2006-2013, the media freedom index of Freedom House was at a level close to the average between 1990 and 2016. However, after 2013, the index has fallen to an historical low, comparable to the levels in the early 1990s. In contrast, the spreading of strikes across social media connected cities continues at a constant and high level after 2011. A likely reason is that the spread of strikes is not regarded as politically sensitive.

The magnitude of the marginal effect during the later period of our sample is very high, in particular for strikes. The fact that the number of events are still falling despite the strong feedback through social media in this period indicates that some other factor is limiting event occurrence, such as stronger suppression or reduced incentives to protest and strike.

Note finally that the large increase in the number of reported events 2013-2016, presumably caused by increased observability through the Wickedonna blog that searched Sina Weibo for events, does not seem to affect the estimated marginal effects much. This is additional evidence that the estimate is robust to increased observability via social media.

[18] In the logarithmic case,

$$h\left(x\right)=\ln\left(5x+1\right)$$

and the reported marginal effect is

$$\beta_b \widetilde{E}(\frac{5\sum_{j\neq i}\overline{f}_{ij}}{5\sum_{j\neq i}\overline{f}_{ij}y_{kt-1}+1}),$$

where $\widetilde{E}()$ is the sample mean.

The marginal effects for the linear model with $h(x)=x$ is simply the estimated β_b-coefficients. These are very similar to those in Figure 3; see Figure A9.

Three periods We now report estimates of a more demanding model that includes separate location-fixed effects and coefficients on controls for each period. The estimated marginal effects are shown in Table 6. Before Sina Weibo was introduced protests were spreading across cities that eventually became closely connected through social media while this was not true for strikes. In the period 2010-2013, this spread increased significantly for both protests and strikes. If one subtracts the pre-period β-coefficient, the average spread through social media is for the second period is around 0.6 for both protests and strikes. In the last period, 2014-2017, the spread of protests falls to insignificant levels while the spread of strikes remains at a high level.

The spread across social media is not just picking up a general time trend in event spread. The distance-related spread of strikes does not move in tandem with social-media induced spread, see Table 6. The spread of strikes to geographically close areas remained roughly the same in the pre-Weibo period as in the period 2010-2013 and then dropped to an insignificant level in the period 2014-2017.

4.4 Do social media break industry and cause bounds?

We now investigate whether social media increase the scope of event waves by spreading protests across causes and strikes across industries. Movements that only spread within the same narrow industry and cause are bounded in size and are unlikely to be perceived as a threat by the regime. It seems a priori likely that strikes predominantly spread within industries and protests within causes. For example, the wave of school teacher strikes in 2014 and 2015 documented by Chang and Hess (2018) spread within the same industry, and the spreading protests among farmers against corruption in Wukan were for the same cause. On the other hand, strikes and protests may also spread across industries and causes.

We study this using regressions that estimate separately the spread of events through social media within and across the protest categories listed in Table 1 (government policy and corruption, housing and land, etc.) and the strike industries listed in Table 2. Table 7 shows that the estimated spread through social media is 6-8 times higher within strike industry and protest cause than across, also the latter is also significant.

It is important to note that while the effect of an individual event is smaller across than within categories, the total effect of spread across categories is larger. The reason is that there are many more events across all categories. The mean of the social media weighted events within and across categories is reported in the two last rows of the table. The total average effect of protests spreading across protests causes is 60% higher than within. Similarly, for the total effect across strike industries is 20% larger than the total effect within industries. Note also that there is no distance related spread of events across strike industries or protest causes. We conclude that social media are breaking the bounds of strikes industries and protest causes and that the aggregate impact is larger than effects through within-category spread.

4.5 Mechanisms

We evaluate possible mechanisms by investigating their specific implications and by looking for content that is necessary to support them.

Effect duration Alternative mechanisms differ in their implications for effect duration. Learning about protest tactics or government responses from protests posts is likely to have long-lasting effects, since what is a useful tactic is not likely to change abruptly and since people who read the posts are not likely to forget the information quickly (and even if they do, the posts remain available online). A short-term response is more likely if the posts arouse emotional responses or if people see the opportunity to combine efforts across location and protest for a common cause.

So far, we have analyzed short-run responses, within two days, since we saw that posts about protests and strikes spread over long distances within a short time span. However, we now expand the effects window and study how effects persist over time. Figure 4 shows the estimated spread of events occurring within 1-2 days, 3-7 days, 8-30 days 31-90 days and 91-180 days. The effect is largest just after the strike and then turn insignificant. The lack of long-run effects is more supportive of mechanisms that go through emotional responses or coordination across cities than learning effects.

Content A prominent mechanism in existing studies is that protesters use social media to organize events and provide logistical information, such as a call for action or at least a mention of where and when a protest or strike is to take place. Enikolopov et al. (2020) report that the vast majority of protest participants in Russia learned about upcoming protest events from the social media and that social media was also widely used for organizing protest activities. Similar content is found in other studies of protests, for example, during the Arab Spring, and in Turkey and Ukraine (Tucker et al. 2016, Steinert-Threlkeld, 2017). The situation in China is very different. In our sample, only 15 in 1000 protest posts involve a call for action and only two explicitly state a time and location of action. These 15 posts are rarely forwarded, with only one forward per post on average. A likely reason is that explicit calls for protest actions are censored. In our random sample of 1000 posts about strikes, none call for action. Hence, explicit organization and coordination is unlikely to be an important channel for social media effects on collective action in China.

Social media may also make protests spread because it carries *information about effective protest tactics* (e.g. Little 2016; Chen and Suen 2016). However, relatively few posts mention tactics (5 protest posts and 13 strike posts in 1000). For example, some posts discuss whether violence is necessary or helpful for a solution while others describe particular tactics. Hence, tactics is reported, but in very few cases. One could imagine that the few posts that introduce new tactics are still very influential. This is not what we see in this data. The posts discussing tactics only receive two forwards per post, on average. It seems unlikely that this is a main

channel for the spreading or protests and strikes across events. Similarly, only a few posts discuss the *outcome of the events*, for example, whether the protest was successful or not.

Most of the posts instead report other facts about the protest, coupled with emotional reactions such as anger and sympathy for the protesters. Many posts mention the cause of the event, such as misbehaviors of officials or persistent wage arrears (164 in the protest category and 223 in the strike category). These posts are among the most forwarded. Information on social media about the problems causing the protests and strikes is important since it is very difficult for outsiders to get this information in any other way.

It is also worth noting that many posts also question social institutions (137 of the protest posts and 42 of the strike posts). This includes complaints about the legal system and absence of freedom of speech. Because of their general nature, these types of posts may induce protests to spread more broadly, for example, across causes and industries.

Perhaps a few posts are very influential. We looked specifically at the 100 most forwarded posts. After removing the repeated posts and the posts that are not relevant to protests, we are left with 91 posts. Of these, 55 talk about ongoing events, while the others comment on past events, government policy, and social problems. Of the posts about ongoing event, almost all discuss the cause of the protest and a majority discuss persecution of protesters posts and government repression. Many of these posts express anger. Another common type of emotional content are reactions by outsiders to the events, with posts stating that the protesters were unfairly treated and expressing sympathy and moral support.

To sum up, because the effects are short-term, substantial social media effects on protests and strikes are unlikely to be driven by learning about protest tactics or government responses. These mechanisms are also unlikely because the content to support them is very infrequent in the blog posts. In addition, lack of social media content allows us to rule out explicit coordination and organization of protests as mechanisms.

A more likely mechanisms is that *information about an ongoing protest and its cause* on social media sparks a short-term response. The short-term response could arise as a consequence of an emotional reaction that motivates people to protest (as in Pasarelli and Tabellini, 2017). It could also arise because of *tacit coordination* as simultaneous protests may reduce the risk of punishment and increase the pressure to deal with the problem (as in e.g Barberà and Jackson, 2016, Edmond, 2013, Little, 2016).

4.6 Event incidence and social media penetration

Finally, we show how protest and strike incidence are associated with social media penetration at the prefecture-month level. We already know from the results presented in Tables 4 and 5 that social media penetration is associated with event incidence. However, this alternative specification at the level at which we observe penetration is simpler and we can more easily discuss magnitudes and heterogenous effects on incidence. Specifically, we will study effect by event size, by whether the prefecture is inland or coastal, and by whether a strike results

in concessions or repression.

To this end, we now estimate a linear probability model of the form

$$y_{im} = \alpha_i + \alpha_m + \beta_0 w_{im} + \beta' x_{im} + \varepsilon_{im},$$

where i indicates prefecture, m indicates month, y_{im} is an indicator variable for an event taking place in prefecture i at month m, w_{im} is Sina Weibo penetration. We use the same set of controls x_{im} as above, with one exception. We drop the total number of forwards by users in location i because it is strongly correlated with Weibo penetration and we wish to have only one measure of the intensity of social media use.

Table 8 shows the results. The first three columns use an indicator variable for a strike as the dependent variable, whereas the last three columns use an indicator variable for protest as dependent variable. The protest and strike incidences are both positively associated with the number of Weibo posts. Adding controls does not much affect the estimates. The magnitudes of the estimated coefficients are large. In 2012, the variable Weibo posts has an average of 0.3. The estimated coefficients imply that an increase in Weibo posts by 0.3 is associated with an increase in the total number of protests by 13 per month (.3*.186*224) and in the total number of strikes by 15 per month (.3*.188*265).

Figure 5 shows the dynamic response in the form of results from a regression with quarterly leads and lags of Weibo penetration, w_{im}. There is no pre-trend and the effects seem to appear with a lag of 6 months for strikes and immediately for protests.

If observability drives these correlations, then we would expect to see larger correlations where observability prior to social media was lowest. The RFA has better information in certain coastal provinces (Guangdong, Fujian, Zhejiang, Jiangsu) than inland. Table 8, columns III and VI, include an interaction term between our Weibo penetration variable and a dummy variable for areas other than these four coastal provinces. The effects on strike incidence is not significantly different in coastal and inland areas. However, the effect on protests is significantly larger in inland areas, consistent with the RFA having fewer sources there.

It is also likely that prior observability was worse for small events. The lower left panel of Figure 5 shows the estimated coefficients by the number of event participants (less than 100, in the hundreds, thousands or tens of thousands). Estimated effects are larger for smaller and medium-sized protests. To conclude, our evidence suggests that observability explains part of the association between social media penetration and protest in inland areas and small strikes.

5 Conclusion and discussion

In response to new information and communication technologies, authoritarian countries, most notably China, have exerted substantial efforts to strategically control online informa-

tion flows. We find evidence consistent with the view that China's strategy aims at achieving the dual goals of utilizing bottom-up information and mitigating the risk of grass-roots anti-regime protests. Chinese social media are void of content that supports the main mechanisms underpinning the effect of social media on protests that are stressed in the previous studies. Very few posts contain explicit information about how to organize protests and information about protest logistics and tactics, likely due to the Chinese government's censorship of social media. On the other hand, Chinese social media massively diffuses the type of information about protests and strikes that is potentially useful for surveillance and gauging public sentiment. In particular, a large number of posts discuss the causes of the events, criticize governments and policies, and express moral anger and sympathy for protesters.

Despite the absence of information that is explicitly useful for organizing protests, we find that social media posting and forwarding had a sizeable effect on the spread of both protests and strikes across cities. The conclusion is that the diffusion of content that the regime can use to maintain regime stability is sufficient to forcefully spread protests and strikes. One reason, consistent with our empirical findings, is that people can use social media information for tacit coordination, for example, simultaneous protests to increase the chance of obtaining their objective and lowering the risk of punishment. Another possible reason is that social media information triggers widespread emotional response, leading to spontaneous protests. The lesson is that, to limit the spread of collective action, an authoritarian government has to shut down discussion of causes and emotional reactions to problems and thus bear the cost of losing precious bottom-up information valuable for identifying unpopular policies, monitoring local officials, and tracking social unrest and anti-regime sentiment.

Another important finding of our study is that social media increases the scope of protests. Although the spread of events induced by social media is fast and predominantly local, spread across event categories is significant and of greater magnitude in aggregation. Even though the effect of social media on event spread is short-lived, the effect caused by a sequence of events may lead to waves of social movements.

A natural question arises following our study: with further advancement of information and communication technology, can an authoritarian regime adopt an alternative media control strategy to collect truthful information but avoid the risk of event spread? For example, the government may limit the spread of information across regions. The success of this type of strategy depends on the value of knowing how people in other regions respond to an event, which reveals information about national attitudes towards the discussed problems and the regime. This question is highly relevant given that some authoritarian governments such as China have invested heavily in surveillance and AI technology to improve social monitoring. With the capability of closely monitoring people's action, an authoritarian regime may allow for a free social media to collect information without worrying about widespread protests– people dare not take to the street. The feasibility of such a strategy hinges on the complementary between online information and offline action. Without street protests, communication

online may be merely cheap talks and spread rumors and noise. We leave investigation into this issue to future research.

Our study also has important implications for the effect of information technology on political accountability. Information is indispensable for holding political leaders accountable to the public in democracies and to higher-level leaders in autocracies. In a large autocracy like China where traditional media are operated by sub-national governments, local officials have an informational advantage over citizens and the central government. This creates severe agency problems within the Chinese political system. However, social media substantially reduces the information asymmetry among the central government, local officials, and citizens. In particular, citizens can easily make their information public while the central government has the technological capability to collect and aggregate this information. Therefore, when the central government's goal is aligned with the citizens' interest, social media may help solve agency problems and hold local officials accountable. One caveat is that the low cost of posting complaints and allegation online may reduce the informativeness of social media. Consequently, the informational value of social media relies critically on real events that are costlier (e.g., protests and strikes) and thus more informative. Investigating how this nuanced complementarity between online allegation and offline protests affects local accountability will be an interesting extension of our current research.

References

[1] Acemoglu, Daron, Camilo Garcia-Jimeno, and James A. Robinson. "State capacity and economic development: A network approach." American Economic Review 105, no. 8 (2015): 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, no. 1(January 2018): 1-42

[3] Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." The Quarterly Journal of Economics, 130(4): 1885-1939, https://doi.org/10.1093/qje/qjv030

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

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

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

[7] Bakshy, Eytan, et al. "The role of social networks in information diffusion." Proceedings of the 21st international conference on World Wide Web. ACM, 2012.

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

[9] Barberà, Salvador, and Matthew O. Jackson. "A model of protests, revolution, and information." Matthew O., A Model of Protests, Revolution, and Information (March 1, 2018) (2018).

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

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

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

[13] Brauer, Alfred, and Ivey C. Gentry. "Bounds for the greatest characteristic root of an irreducible nonnegative matrix." Linear Algebra and its Applications 8.2 (1974): 105-107.

[14] 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.

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

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

[17] Campante, Filipe R., Davin Chor, and Bingjing Li. The Political Economy Consequences of China's Export Slowdown. No. w25925. National Bureau of Economic Research, 2019.

[18] 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.

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

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

[21] Chen, Xiaoyan and Peng Hwa Ang. 2011. "The Internet Police in China: Regulation, Scope and Myths." In Online Society in China: Creating, Celebrating, and Instrumentalising the Online Carnival, edited by David Kurt Herold and Peter Marolt, 40–52. New York: Routledge.

[22] China Internet Network Information Center. 2014. "The 34th Statistical Report on Internet Development in China." January 2014, Beijing.

[23] China Labor Bulletin. 2015. Search for the Union: the Workers' Movement in China 2011-2013, Research Report by China Labor Bulletin.

[24] China Labor Bulletin. 2018. The Workers' Movement in China: 2015-2017, Research Report by China Labor Bulletin.

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

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

[27] 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.

[28] Elhorst, J. Paul. "Dynamic models in space and time." Geographical Analysis 33.2 (2001): 119-140.

[29] Elhorst, J. Paul. "Spatial panel data models." Spatial econometrics. Springer, Berlin, Heidelberg, 2014. 37-93.

[30] Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. "Social media and protest participation: Evidence from Russia." Econometrica 88.4 (2020): 1479-1514.

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

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

[33] 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.

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

[35] Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya. "3g internet and confidence in government." The Quarterly Journal of Economics (2020);, qjaa040, https://doi.org/10.1093/qje/qjaa040.

[36] Hsiao, Cheng. Analysis of panel data. No. 54. Cambridge university press, 2014.

[37] 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

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

[39] King, Gary, Jennifer Pan, and Margaret E. Roberts. "How the Chinese government fabricates social media posts for strategic distraction, not engaged argument." American political science review 111.3 (2017): 484-501.

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

[41] König, Michael D., Dominic Rohner, Mathias Thoenig, and Fabrizio Zilibotti. "Networks in conflict: Theory and evidence from the great war of africa." Econometrica 85, no. 4 (2017): 1093-1132.

[42] Kwak, Haewoon, et al. "What is Twitter, a social network or a news media?." Proceedings of the 19th international conference on World wide web. AcM, 2010.

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

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

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

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

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

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

[49] Manski, C., 1993. Identification of endogenous social effects: The reflection problem. Review of Economic Studies 60 (3), 531 542.

[50] Morozov, Evgeny. 2012. The Net Delusion: The Dark Side of Internet Freedom. Reprint edition. Public Affairs.

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

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

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

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

[55] Roberts, Margaret E. Censored: distraction and diversion inside China's Great Firewall. Princeton University Press, 2018.

[56] Shirky, Clay. 2011. "The Political Power of Social Media: Technology, the Public Sphere, and Political Change." Foreign Affairs, January/February.

[57] Steinhardt, H. Christoph. 2017. "Discursive Accommodation: Popular Protest and Strategic Elite Communication in China." European Political Science Review, 9(4):539-560 https://doi.org/10.1017/S1755773916000102

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

[59] 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.

[60] Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." The Quarterly Journal of Economics, 129(4): 1947-1994 https://doi.org/10.1093/qje/qju020.

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

A Appendix

A.1 Marginal effects

To interpret β in equation (2), note that the average marginal effect of y_{jt-1} across locations $j \neq i$ on $\Pr(y_{it})$ is

$$\beta h'(x_{it}) \frac{1}{N} \sum_{j \neq i} f_{ijt-1}.$$

The sum of these effects across all locations i is

$$\beta \sum_i h'(x_{it}) \frac{1}{N} \sum_{j \neq i} f_{ijt-1}$$

For the linear function $h'(x_{it}) = 1$. Hence, the above expression equals β time the average row sum of f_{ijt-1},

$$\beta \frac{1}{N} \sum_i \sum_{j \neq i} f_{ijt-1}.$$

The above expression, averaged across time, equals

$$\beta \frac{1}{N} \frac{1}{T} \sum_t \sum_i \sum_{j \neq i} f_{ijt-1} = \widetilde{E}(\sum_{j \neq i} f_{ijt-1}) = \beta,$$

where $\widetilde{E}$ is the sample mean. The last step follows since the average row sum of f_{ijt-1} is normalized to one.

For the logarithmic function,

$$\beta \sum_i h'(x_{it}) \frac{1}{N} \sum_{j \neq i} f_{ijt-1} = \beta \frac{1}{N} \sum_i \frac{5 \sum_{j \neq i} f_{ikt-1}}{5 \sum_{j \neq i} f_{ijt-1} y_{kt-1} + 1}.$$

The above expression, averaged across time, equals

$$\beta \widetilde{E}(\frac{5 \sum_{j \neq i} f_{ikt-1}}{5 \sum_{j \neq i} f_{ijt-1} y_{kt-1} + 1}) = s\beta.$$

Evaluated at sample means equals

$$s = \frac{5}{5 \overline{\sum_{i \neq j} f_{ijt-1} y_{jt-1}} + 1},$$

since the average row sum of f_{ijt-1} is normalized to one. For strikes, at the sample mean value of

$$\overline{\sum_{i \neq j} f_{ijt-1} y_{jt-1}} = 0.0193,$$

so that $s_t = 4.56$ So the estimate $\widehat{\beta} = 0.037$ in column IV should be multiplied by 4.56 to be

comparable to the estimate of the linear model in column IV of Table A1: $4.56 * 0.038 = 0.17$. For protests, the scaling factor is $s_t = 4.63$ and the marginal effect is $4.63 * .055 = .25$.

These effects are somewhat larger than the corresponding estimates from the linear model reported in columns II and IV of Table A1, which are .17 for protests and .12 for strikes.

A.2 Stationarity

In our setting, stationarity of the protest and strike processes is not simply an econometric issue. Whether these processes are stable or exploding is likely to be a core concern of the top political leadership. Stationarity in a dynamic spatial panel data model depends on the parameters of the model as well as on the spatial weights matrix that determine the amount of feedback in the process. For the location's own autoregressive term and the distance weighted term, this feedback is constant over time (since α, γ and the distance matrix D are constant). However, increased use of social media increases the feedback because each individual row in the forwarding matrix, F, does not sum up to one. This implies that marginal effect of a change in y_{t-1} on the probability of a protest or strike differs across localities and time. In the linear model, the average effect at a particular date t equals $\beta\overline{f}_{t-1}$, where $\overline{f}_{t-1}$ is the average row sum across locations that day. The maximum such row sum equals 10.5, an order of magnitude larger than the average row sum. This implies that sufficient conditions for stationarity are not fulfilled for the linear model.[19]

However, the actual data do not exhibit the explosive path suggested by the estimated linear model. This could be because the government observed that the process was exploding and stepped in and struck down on protests and strikes, but we do not have direct evidence of this. A more likely reason is that the model is mis-specified in that the marginal effects of protests and strikes are assumed linear in the number of cross-sectional events, in other words, in how wide spread protests are. The linearity assumption may be incorrect for several reasons. If the mechanism is through information about protest methods, then it is likely that the marginal value of information from an additional event is falling. If the mechanism is through information about the number of protests about an issue, then the incentive to organize an additional protest so as to increase government awareness is likely to be decreasing. It may also be the case that there is a limited number of areas where people are upset enough about a particular issue to potentially protest or strike if they see that other people protesting for this issue. Hence there is a cap to the total number of protests and the process will eventually be concave.

We model this by using the concave function $h(x) = \ln(5x + 1)$, for terms x involv-

[19] Formally, sufficient conditions for stationarity are $|\beta\omega_{F_t,\max}| + |\gamma\omega_{D,\max}| < 1$, where $\omega_{D,\max}$ is the largest real characteristic root of the matrix D (largest negative if $\gamma < 0$) and $\omega_{F_t,\max}$ correspondingly for the F_t matrix (from Elhorst, 2014). The greatest characteristic roots z of a irreducible non-negative matrix A with maximum row-sum $R(A)$ satisfy $|z| < R(A)$. (Brauer and Gentry, 1970). Hence, the second criterion is fulfilled if $|\gamma R(D)| + |R(\beta F_t)| < 1$. It is clear that this condition is not fulfilled in our case with $h(x) = x$. The estimated process is explosive.

ing lagged events in equation 2, for example, we insert $\ln(5\sum_{i\neq j} f_{ijt-1}y_{jt-1}+1)$ instead of $\sum_{i\neq j} f_{ijt-1}y_{jt-1}$ (in this case$x = \sum_{i\neq j} f_{ijt-1}y_{jt-1}$). This function was chosen to be sufficiently concave to make the process stable and is used in the simulations. Figure A3 shows that the linear model and the concave model have similar slopes in the region where the data density is high. Therefore, the estimated effects are very similar. However, some simulations reach high levels of $\sum_{i\neq j} f_{ijt-1}y_{jt-1}$. In these cases, the concave model is stable while the linear model explodes. For this reason, we use the concave model in the simulations. We prefer to report the linear model results in the main tables since the coefficients are simpler to interpret and report the results from the concave model here.

Table A1 and A3 correspond to Tables 4 and 5, but uses the linear transformation $h(x) = x$. To interpret the coefficient magnitudes, the estimates in column 2 implies that a protest in a given location the previous two days increased the expected number of locations with protests this day by 0.163.[20] The equivalent number of strikes is 0.093. Relative to the mean event probability, this is an increase of 48 percent for protests and 28 percent for strikes.[21]

A.3 Probit regression

Table A2 shows the results from a probit regression of Equation 2, where the time fixed effects have been replaced by a quadratic time trend to avoid the incidental parameters problem. The estimates of β are significant across all specifications. However, the implied marginal effects on event probabilities are smaller.

A.4 Monte Carlo simulations

A.4.1 The baseline model

We run a set of Monte Carlo simulations to assess the Nickel bias in the coefficient estimates of our baseline model. We first estimate the parameters $\alpha, \beta, \gamma, \delta_t$ and δ_i from a regression specified as in Equation 2 but for simplicity without Weibo penetration and controls. We then generate data using the estimated parameters, adjusted so that $\delta_t + \delta_i \geq 0$ and estimate the model on this data. We repeat this procedure 100 times.

Figure A6 in the Appendix plots the distribution of t-statistics of coefficients α, β and γ in 2 against the standard normal density. The bias is very small, the difference between the true and the mean estimated β is in the third value digit for both protests and strikes.

[20] To assess the magnitude of the estimated effects, consider first the case $h(x) = x$. The marginal effect of a strike at $y_{j,t-1}$ on the strike probability $\Pr(y_{it})$, through the forwards-weighted term, equals βf_{ijt-1}. Our weighting matrices, D, and F, are normalized so that the average row-sum equals one. Hence, β measures the average increase in strike probability if there was a strike in all locations on the previous day. Since the average column-sum of F_t is one, β also measures the expected total increase in strike probability across all locations at date t due to one strike at a random locality the previous day.

[21] The mean of the protest incidence is .0015 and there are 224 other prefectures than the one where the first strike took place, .16/(224*.0015)=0.48. For strikes, the corresponding number is .09/(218*.0015)=.28.

A.4.2 Time-invariant heterogeneity in spread across locations

We also ran a set of Monte Carlo simulations to assess the Nickel bias in the coefficient estimates of model including fixed effects for arbitrary time-constant spread across locations. As in the case above, the baseline model was used for the data generating process. We then estimated the interaction-fixed-effects model using the simulated data.

Figure A7 shows the distribution of β-estimates from Monte Carlo simulations. The graphs to the left show results from estimations without interaction-fixed effects (Equation 2). The graphs to the right are based on regressions that include interaction-fixed effects, corresponding to the results in Table 5. These graphs show a clear, albeit small, bias. The blue line shows the coefficients from the estimations without interaction-fixed effects (Equation 1) while the red line shows the mean coefficients from the simulated data.

We also tested for auto-correlated errors on simulated data (where there is no auto-correlation in the errors by construction). The test for auto-correlation in the baseline case slightly over-rejects auto-correlation. However, the model with interaction-fixed effects clearly rejects auto-correlation. The likely reason is that the many interaction-fixed effects are imprecisely estimated. Partialling out slightly incorrect auto-correlated terms generates auto-correlated errors.

The data generating process has no auto-correlated errors. The test for auto-correlation in the baseline model verifies this, although it slightly over-rejects no auto-correlation. The model with interaction-fixed effects can control for the same pattern as the baseline model (since it is more general), but the many interaction-fixed effects are imprecisely estimated. Partialling out slightly incorrect auto-correlated terms generates auto-correlated errors. Hence, it is not surprising that the auto-correlation test for the model with interaction-fixed effects rejects no auto-correlation. Presumably, the auto-correlation would disappear as the sample size grows to infinity, because the coefficients are consistently estimated and would converge to the DGP. The lack of auto-correlation in the errors in Table 4 shows that there is no other significant spread of protest other than what is captured by the baseline model. Although the specification used in Table 5 will generate auto-correlated errors in small samples, this will not severely bias the estimated coefficients as shown in Figure A7.

Table 1: Protests by cause 2006-2013

	Size					
Cause	<100	100s	1,000s	10,000s	Unknown	Total
Government policy or corruption	58	105	93	6	51	313
Housing and land	23	112	79	5	62	281
Labor	7	52	45	1	15	120
Police and court	12	44	77	7	19	159
Environment and health	3	36	51	5	26	121
Firm and finance		18	12	3	3	36
Ethnic and religion	9	21	8	3	9	50
School and education	5	17	12		5	39
Social conflicts		5	9	2	1	17
International relationship			1		1	2
Transportation					1	1
Human right	2	2	2		5	11
Unknown				1	2	3
Total	119	412	389	33	200	1,153

Table 2. Strikes by industry 2007-2013

	Size					
Industry	<100	100s	1,000s	>=10,000s	Unknown	Total
Manufacturing	99	400	216	5	72	792
Taxi	26	58	14	2	19	119
Transportation	138	83	13	0	23	257
Services	63	60	4	0	11	138
Education	36	32	12	0	6	86
Construction	50	29	12	2	3	96
Retail	15	13	2	0	5	35
Mining	1	8	3	0	3	15
Other	3	8	4	0	2	17
Unknown	0	0	3	0	0	3
Total	431	691	283	9	144	1,558

Table 3. Summary statistics

	count	mean	sd	min	max
Protests					
event	668,260	.0015458	.0392864	0	1
y_{it-1}	668,260	.0054728	.0991698	0	2.397895
$h(\sum d_{ij} y_{jt-1})$	668,260	.0155264	.0336983	0	.5459623
$h(\sum f_{ijt} y_{jt-1})$	668,260	.0666955	.1485338	0	1.497624
Strikes					
event	713,702	.0020415	.0451365	0	1
y_{it-1}	713,702	.0071719	.1138407	0	2.397895
$h(\sum d_{ij} y_{jt-1})$	713,702	.0199501	.0383343	0	.7732889
$h(\sum f_{ijt} y_{jt-1})$	713,702	.0808735	.1582861	0	1.476421

Table 4. Event spread across locations

	I	II	III	IV
	Protest	*Protest*	*Strike*	*Strike*
y_{it-1}	0.009**	0.009**	0.017***	0.017***
	(0.004)	(0.004)	(0.004)	(0.004)
$h(\sum d_{ij} y_{jt-1})$	-0.008	-0.007	0.031**	0.030**
	(0.008)	(0.008)	(0.013)	(0.012)
$h(\sum f_{ijt} y_{jt-1})$	0.055***	0.055***	0.039***	0.037***
	(0.013)	(0.012)	(0.008)	(0.007)
$h(\sum f_{ijt})$	-0.000	-0.000	-0.001	-0.001
	(0.000)	(0.000)	(0.001)	(0.001)
Weibo posts	0.006***	0.006***	0.011***	0.010***
	(0.002)	(0.002)	(0.004)	(0.003)
Controls	No	Yes	No	Yes
Observations	668,260	668,260	713,702	713,702
R-squared	0.017	0.017	0.027	0.027
QPtest	0.07	0.15	0.27	0.32

Results from a linear regression on an event dummy variable. The unit of observation is prefecture by day. The function h(x)=ln(5x+1). The regression includes prefecture and day fixed effects. Controls are log of one plus the sum of all forwards by users in location i by posts from all users j in all other locations h($\sum f_{ijt}$), population, GDP, tertiary share, industrial share, and the number of cell phone users . The QP statistic reports the p-value of the test for serial correlation in fixed-effects model of Born and Breitung (2016). Standard errors are two-way clustered by date and location: *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Eventspread across locations

	I	II	III	IV
	Protest	Protest	Strike	Strike
$h(\sum f_{ijt} y_{jt-1})$	0.049***	0.047***	0.025***	0.024***
	(0.013)	(0.013)	(0.008)	(0.008)
$h(\sum f_{ijt})$	0.000	0.000	-0.000	-0.001
	(0.000)	(0.000)	(0.001)	(0.001)
Weibo posts	0.006**	0.005**	0.009***	0.009***
	(0.002)	(0.002)	(0.003)	(0.003)
Observations	668,260	668,260	713,702	713,702
R-squared	0.223	0.223	0.239	0.239
Controls	No	Yes	No	Yes

Results from a linear regression on an event dummy variable. The unit of observation is prefecture by day. The function h(x)=ln(5x+1). The regression includes prefecture and day fixed effects. Controls are log of one plus the sum of all forwards by users in location i by posts from all users j in all other locations, population, GDP, tertiary share, industrial share, and the number of cell phone users . Standard errors are two-way clustered by date and location: *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Three-period analysis

VARIABLES		I Protest	II Protest	III Strike	IV Strike
	Period	0.004	0.003	0.119*	0.116
$h(\sum d_{ij}y_{jt-1})$:	0	0.004	0.003	0.119*	0.116
		(0.017)	(0.018)	(0.070)	(0.071)
	1	-0.032	-0.029	0.113*	0.106*
		(0.040)	(0.040)	(0.061)	(0.058)
	2	0.024	0.022	0.025	0.032
		(0.026)	(0.026)	(0.036)	(0.035)
$h(\sum \bar{f}_{ij}y_{jt-1})$:	0	0.260***	0.270***	0.025	0.020
		(0.085)	(0.092)	(0.214)	(0.235)
	1	0.796***	0.895***	0.622***	0.632***
		(0.200)	(0.206)	(0.135)	(0.133)
	2	0.379***	0.399***	0.752***	0.777***
		(0.111)	(0.118)	(0.096)	(0.104)
Observations		1,131,840	1,023,533	1,224,880	1,119,493
R-squared		0.019	0.020	0.048	0.049
Pr(β period: 1=0)		0.009	0.003	0.028	0.027
Pr(β period: 2=0)		0.379	0.384	0.001	0.002

Results from a linear regression on an event dummy variable. The coefficients are scaled to measure marginal effects. The unit of observation is prefecture by day. The function h(x)=ln(5x+1). The regression includes prefecture-by-period and day fixed effects. Controls are interacted with period-fixed effects: population, GDP, tertiary share, industrial share, and the number of cell phone users (the sum of weights, $\sum \bar{f}_{ij,}$, is constant and perfectly collinear with the location fixed effect). Standard errors are two-way clustered by date and location: *** p<0.01, ** p<0.05, * p<0.1.

Table 7. Event spread, within and across categories

	I	II
	Protest	*Strike*
Within		
Number events 1-2 days prior, cumulative forwards weighted	0.0642***	0.0606***
	(0.0132)	(0.0140)
Number events 1-2 days prior, distance weighted	-0.0089	0.0943***
	(0.0107)	(0.0266)
Across		
Number events 1-2 days prior, cumulative forwards weighted	0.0086***	0.0064***
	(0.0029)	(0.0020)
Number events 1-2 days prior, distance weighted categories	-0.0017	0.0056
	(0.0026)	(0.0040)
Observations	8,687,380	7,137,020
R-squared	0.0078	0.0162
Controls	Yes	Yes
Category	cause	industry
Mean within	0.0013	0.0020
Mean across	0.0152	0.0179

Results from a linear regression on an event dummy variable. The unit of observation is prefecture by day. The regression includes prefecture and day fixed effects. Controls are h($\sum f_{ijt}$), population, GDP, tertiary share, industrial share, and the number of cell phone users . Standard errors are two-way clustered by date and location.

Table 8: Effects on incidence: dependent variable event dummy

	I	II	III	IV	V	VI
	Strike			Protest		
Weibo posts	0.203***	0.188***	0.179***	0.199***	0.186***	0.136***
	(0.045)	(0.045)	(0.045)	(0.043)	(0.042)	(0.030)
Weibo posts, inland			0.038			0.236***
			(0.053)			(0.047)
Observations	23,475	23,463	23,463	21,993	21,981	21,981
R-squared	0.185	0.187	0.187	0.146	0.148	0.154
Controls	No	Yes	Yes	No	Yes	Yes

Results from a linear regression on an event dummy variable. Unit of observation :prefecture by month. The regression includes prefecture and month fixed effects. Controls are population, GDP, tertiary share, industrial share, and the number of cell phone users . Standard errors two-way clustered by prefecture and month in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Figure 1. Number of events and Weibo posts per capita per month

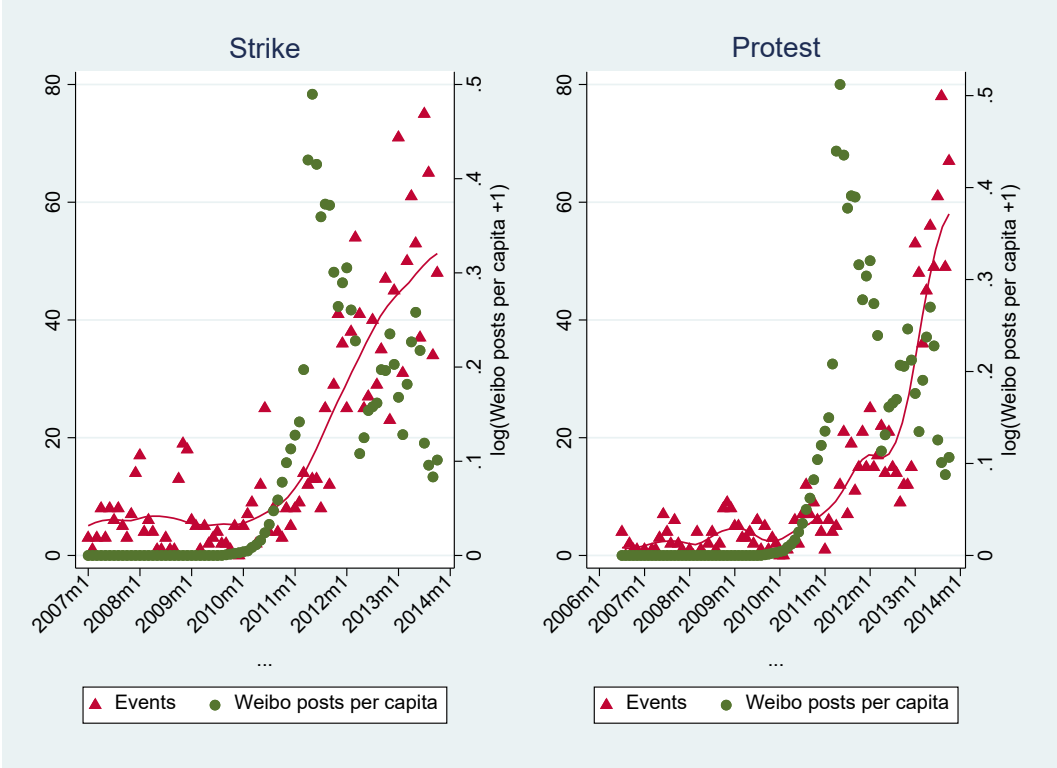


Figure 2. Monte Carlo simulations with observability driven by Weibo and no network spread effect



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

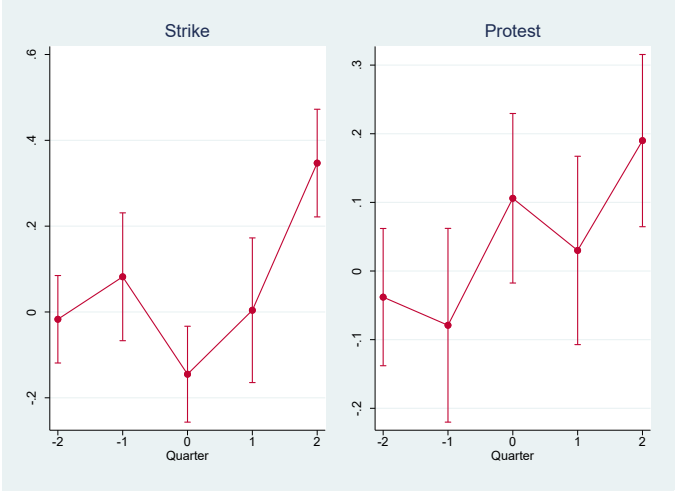
Figure 3. Estimated marginal effects using constant forwarding matrix

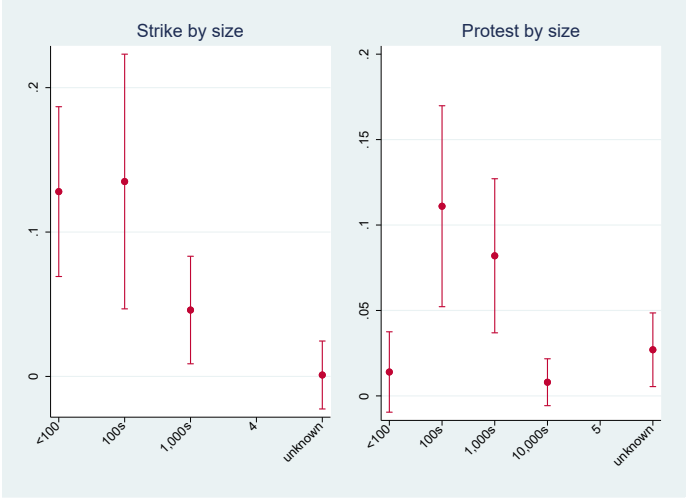


Figure 4. Effect duration



Figure 5. Dynamic and heterogenous effects on incidence (beta0)





The upper panel shows the dynamic effects. The lower panel shows heterogenous effects by size.

Appendix

Table A1. Event spread across locations

	I	II	III	IV
	Protest	*Protest*	*Strike*	*Strike*
y_{it-1}	0.015**	0.015**	0.030***	0.030***
	(0.007)	(0.007)	(0.007)	(0.006)
$\sum d_{ij} y_{jt-1}$	-0.031	-0.030	0.143**	0.137**
	(0.037)	(0.037)	(0.058)	(0.055)
$\sum f_{ijt} y_{jt-1}$	0.170***	0.169***	0.122***	0.120***
	(0.044)	(0.043)	(0.026)	(0.025)
$\sum f_{ijt}$	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)
Controls	No	Yes	No	Yes
Observations	668,260	668,260	713,702	713,702
R-squared	0.017	0.017	0.027	0.027
QPtest	0.05	0.13	0.14	0.34

Results from a linear regression on an event dummy variable. The unit of observation is prefecture by day. The regression includes prefecture and day fixed effects. Controls are log of one plus the sum of all forwards by users in location i by posts from all users j in all other locations, population, GDP, tertiary share, industrial share, and the number of cell phone users . Standard errors are two-way clustered by date and location. The QP statistic reports the p-value of the test for serial correlation in fixed-effects model of Born and Breitung (2016).

Table A2. Event spread across locations – probit model

VARIABLES	I Strike	II Strike	III Protest	IV Protest
y_{it-1}	0.190***	0.191***	0.103**	0.097**
	(0.027)	(0.027)	(0.048)	(0.047)
$h(\sum d_{ij} y_{jt-1})$	0.548***	0.541***	0.393	0.444
	(0.151)	(0.151)	(0.330)	(0.322)
$h(\sum f_{ijt} y_{jt-1})$	0.195***	0.199***	0.195**	0.181**
	(0.058)	(0.059)	(0.084)	(0.079)
$\sum f_{ijt}$	0.049*	0.051*	-0.094**	-0.095**
	(0.026)	(0.026)	(0.037)	(0.037)
Weibo posts	0.068**	0.070**	0.244***	0.236***
	(0.029)	(0.032)	(0.054)	(0.057)
Observations	581,509	581,509	558,906	558,906
Controls	No	Yes	No	Yes

Results from probit regressions on event dummy variable. All regressions contain prefecture fixed effects and quadratic time trends. Unit of observation: prefecture by day. Standard errors clustered by prefecture in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Event spread across locations, allowing time-constant heterogenous spread

	I	II	III	IV
	Protest	*Protest*	*Strike*	*Strike*
$\sum f_{ijt} y_{jt-1}$	0.217***	0.216***	0.109***	0.107***
	(0.061)	(0.061)	(0.039)	(0.038)
$\sum f_{ijt}$	0.003***	0.003***	0.003***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Weibo posts	0.004**	0.004**	0.008**	0.007**
	(0.002)	(0.002)	(0.004)	(0.003)
Observations	668,260	668,260	713,702	713,702
R-squared	0.223	0.223	0.239	0.239
Controls	No	Yes	No	Yes

Results from a linear regression on an event dummy variable. The unit of observation is prefecture by day. The regression includes prefecture and day fixed effects. Controls are log of one plus the sum of all forwards by users in location i by posts from all users j in all other locations, population, GDP, tertiary share, industrial share, and the number of cell phone users . Standard errors are two-way clustered by date and location. The QP statistic reports the p-value of the test for serial correlation in fixed-effects model of Born and Breitung (2016).

Figure A1. Distribution of collective action events across prefectures

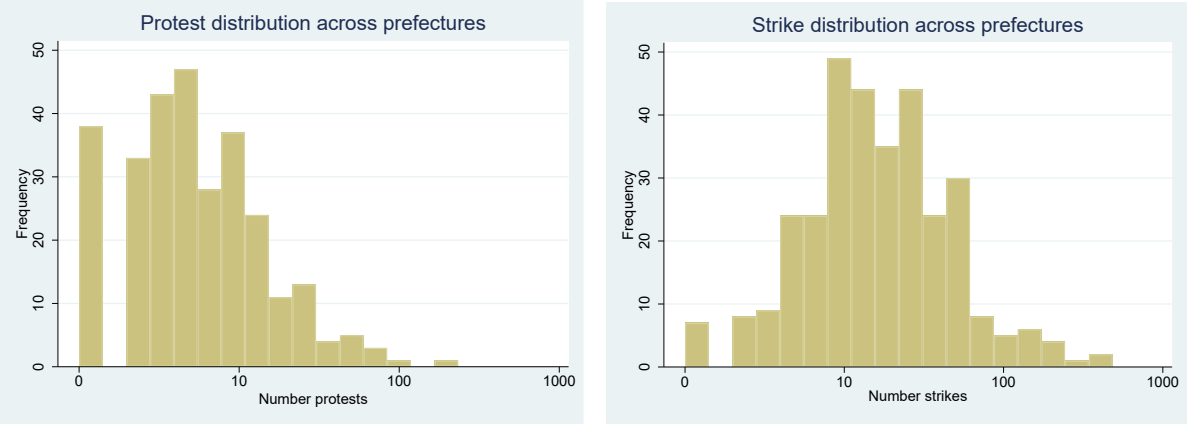


Figure A2. Strike count by prefecture 2007-2013 and protest count by prefecture 2006-2013

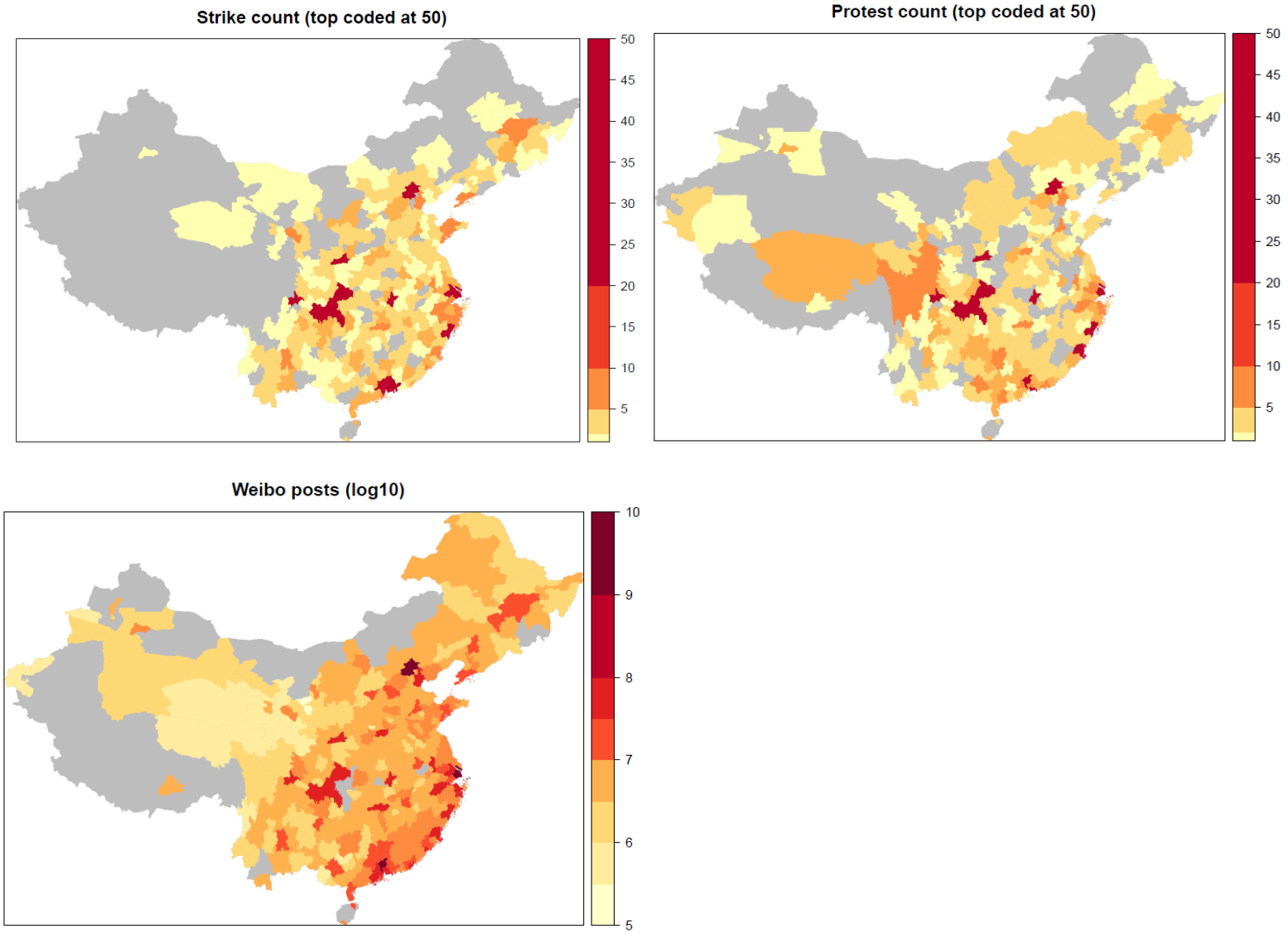


Figure A3. Forward connections from prefectures with many strikes

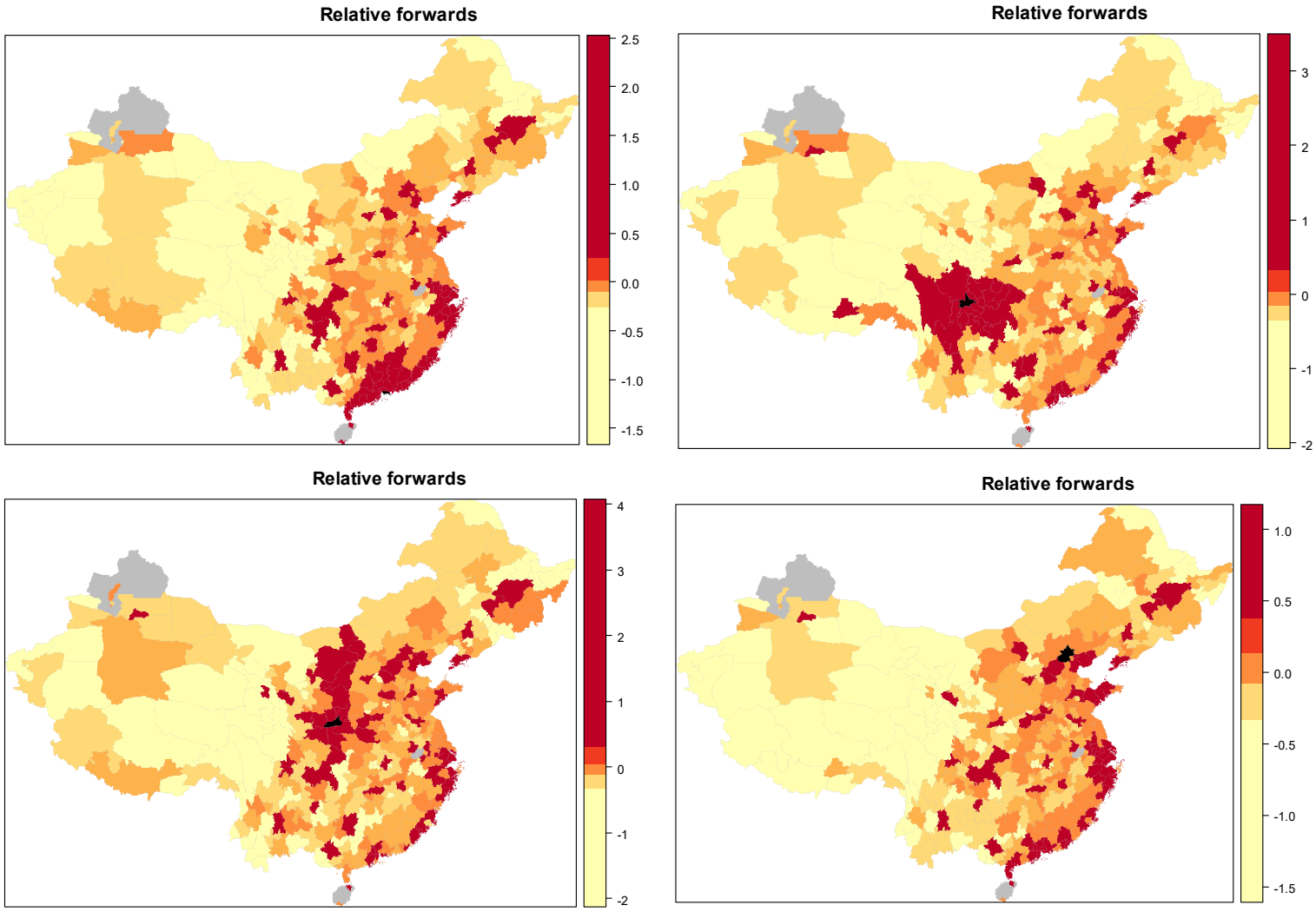


Figure A4: Predictors of relative forwarding between cities

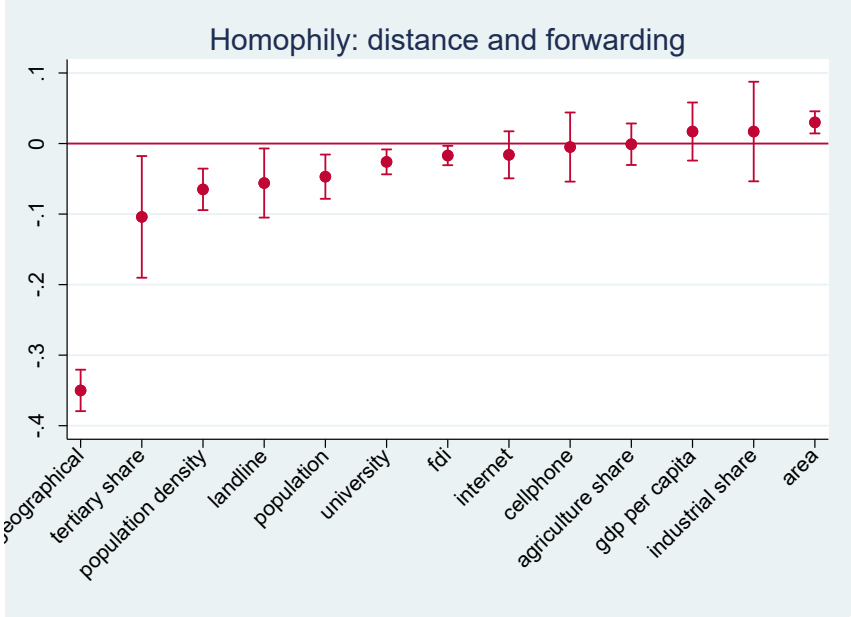


Figure A5. Retweet quantity and distance over time

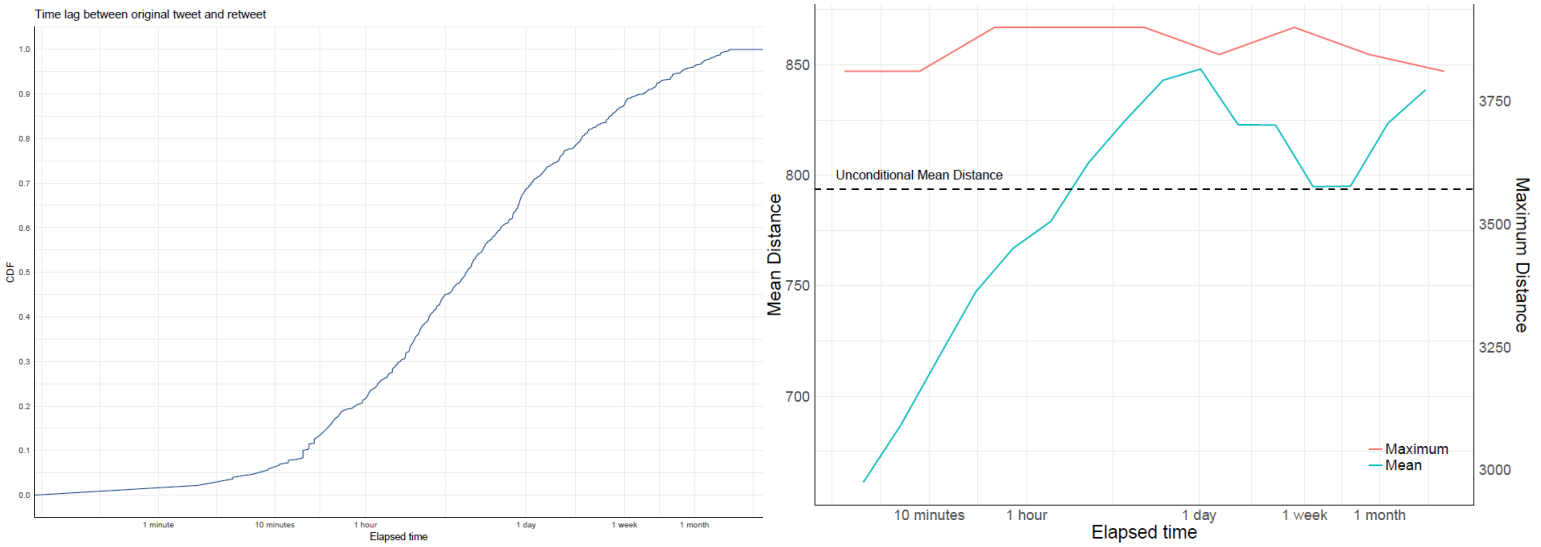
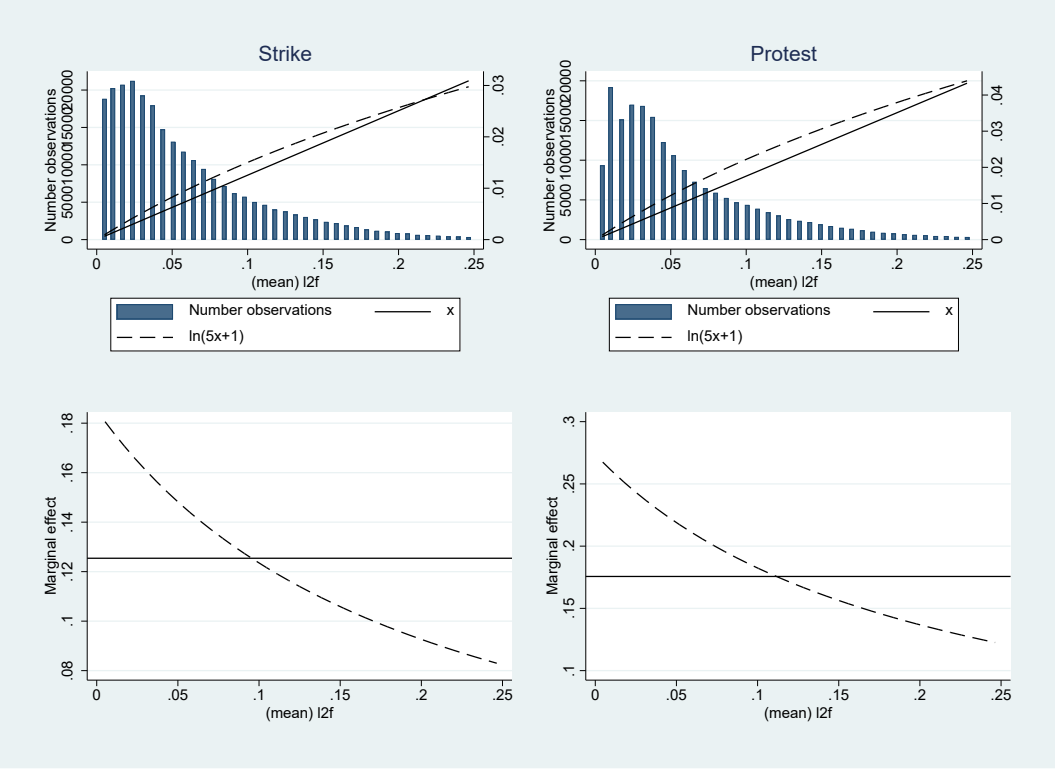
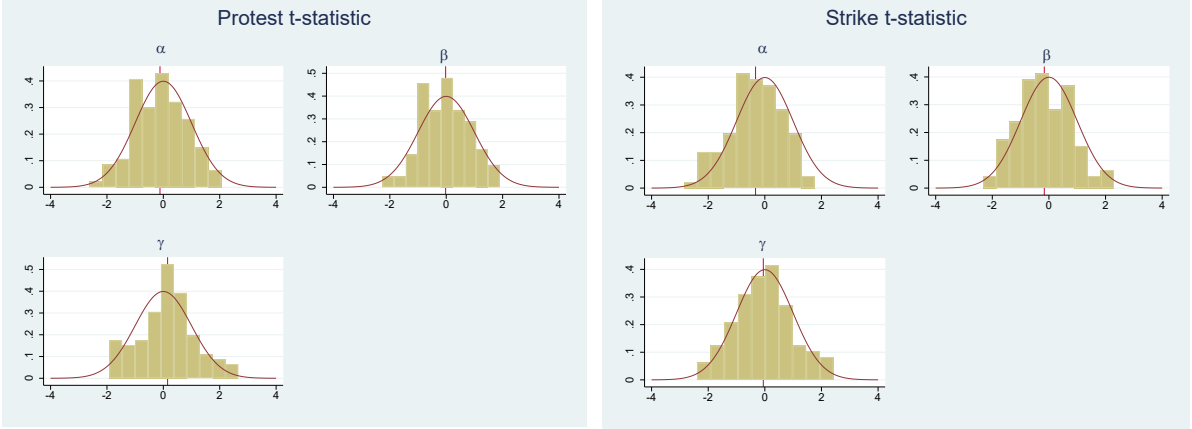


Figure A6. Functional-form comparison in sample range



The x-axis contains the sum of the forwards-weighted lagged events. The upper panel shows a histogram of the number of observations in each bin, together with beta*h(x) for the linear and the logarithim h(x)-functions. The lower pane shows the implied marginal effects from the two models. The estimated marginal effect using the logarithmic function (the dashed line) is higher for low values and lower for high values of x.

Figure A7. Monte Carlo Simulations: distribution of t-statistic of estimated parameter = true parameter



To assess the Nickell bias in our coefficient estimates, we run a set of Monte Carlo simulations. The bias is very small, the difference between the true and the mean estimated β is in the third value digit for both protests and strikes, see Appendix Figure A4.

Figure A8. Monte Carlo Simulations: distribution of coefficients around DGP parameter values.

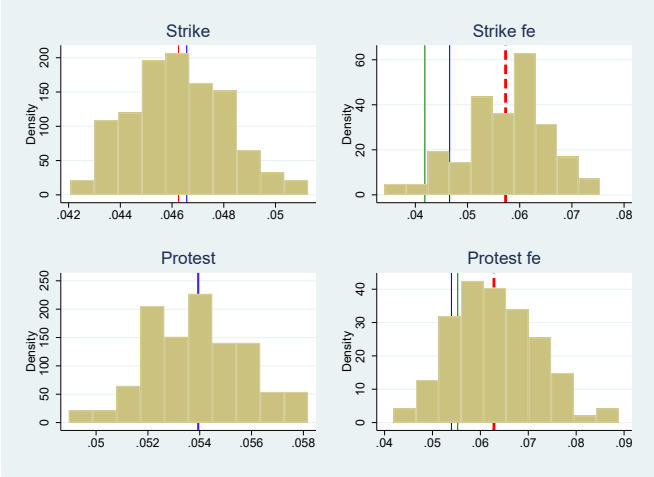


Figure A9. Estimated marginal effects using constant forwarding matrix and linear model, h(x)=x

