How Sudden Censorship Can Increase Access to Information*

William Hobbs† and Margaret E. Roberts‡

October 23, 2016

Abstract

Conventional wisdom assumes that increased censorship will strictly decrease access to information. We delineate circumstances when increases in censorship will expand access to information. When governments suddenly impose censorship on previously uncensored information, citizens accustomed to acquiring this information will be incentivized to learn methods of censorship evasion. These tools provide continued access to the newly blocked information and also extend users' ability to access information that has long been censored. We illustrate this phenomenon using millions of individual-level actions of social media users in China before and after the block of Instagram. We show that the block inspired millions of Chinese users to acquire virtual private networks (VPNs) and join censored websites like Twitter and Facebook. Despite initially being apolitical, these new users began browsing blocked political pages on Wikipedia, following Chinese political activists on Twitter, and discussing highly politicized topics such as opposition protests in Hong Kong.

*We thank the following people for helpful comments and suggestions on this work: Alexei Abrahams, Eli Berman, Allen Carlson, Dean Eckles, James Fowler, Navid Hassanpour, Ruixue Jia, Holger Kern, Gary King, Shuhei Kurizaki, Jennifer Pan, Maria Petrova, Susan Shirk, Zachary Steinert-Threlkeld, Brian Tsay, Yiqing Xu, Jason Wu, the participants in the Social Media and Political Participation Global Conference and the Pacific Peace Science Conference. We specifically thank Zachary Steinert-Threlkeld for graciously supplying geo-located Twitter data from China and Pablo Barberá who helped us access geo-located Instagram data. We thank joint funding from the Policy Design and Evaluation Lab and Center for Global Transformation at UC San Diego and the Hellman Fellows Fund.

†Department of Political Science, University of California, San Diego, Social Sciences Building 301, 9500 Gilman Drive, #0521, La Jolla, CA 92093-0521, http://polisci2.ucsd.edu/whobbs/

‡Assistant Professor, Department of Political Science, University of California, San Diego, Social Sciences Building 301, 9500 Gilman Drive, #0521, La Jolla, CA 92093-0521, meroberts@ucsd.edu, MargaretRoberts.net

Introduction

Scholars have studied censorship by assuming it has an exogenous impact on a passive citizenry. Conventional wisdom posits that censorship should lower the probability that citizens access information (Morozov, 2012; Lessig, 1999). Not surprisingly, then, authoritarian governments appear to be tightening their grip on information environments, increasing their use of search filtering, content removal, and website blocking (Deibert et al., 2010; Kelly et al., 2012; Shirk, 2011), along with rapid online censorship during large-scale collective action events (King, Pan and Roberts, 2013, 2014),

Here, we show that information environments more realistically function like ecosystems and are therefore not always predictable. Because citizens quickly adapt to censorship, the imposition of Internet restrictions can sometimes have unexpected consequences. In certain circumstances, sudden censorship can even result in the opposite of the intended effect: an *increase* in access to off-limits information.

We show that censorship can increase access to information when a government extends censorship to previously uncensored media, especially when individuals have no ready alternatives to replace it. When favorite and difficult to replace media are suddenly blocked, those who would otherwise not take the time to evade censorship restrictions will have a stronger motive to learn how to continue to access the newly censored information. Because censorship evading technology (and, more broadly, social networks) that assist in evasion are rarely specific to particular information sources, acquisition of these tools and networks can give users unhindered access to *many* country-specific blocked sources. Although many users might learn censorship evasion only to maintain access to a preferred source, censorship evasion incidentally expands the set of information these users can easily access. We name this phenomenon a "gateway effect" as the motivation to access the newly censored information provides a gateway into access to information that has long been censored or blocked.

We illustrate this phenomenon by bringing to bear one of the first large, detailed, and individual-level datasets of the real-time censorship evasion behavior of Internet users under policies of increased censorship. Using geo-located social media data from websites that are blocked by the Great Firewall in China, we measure the number and types of people who use Virtual Private Networks (VPNs) to "jump" the Great Firewall to access information and networks blocked by the Chinese government. We show how increases in censorship through blocking of the popular social networking website Instagram in China disrupted the habits of millions of individuals accustomed to visiting that site and increased evasion of the Great Firewall. In doing so, censorship of Instagram created a sudden increase in the number of people from China accessing websites and networks such as Twitter and Facebook that the Chinese government has blocked for many years. We present evidence that previously apathetic Chinese social media users began to engage in topics related to ongoing protest events and with people in Chinese activist circles shortly after the Instagram block.

Our evidence suggests a mechanism that could explain previous work that has found a

positive correlation between censorship and the likelihood of government resistance (Hassanpour, 2014; Nabi, 2014). However, importantly, the "gateway effect" we explore does not rely on backlash against censorship which has commonly explained this phenomenon in the past. Our mechanism also does not require pre-existing or nascent political interest on the part of the censored individuals. As governments expand the types of information that are off limits, they can simply create more incentives for individuals to engage in networks and technologies that allow them to access the media they are accustomed to consuming, and these technologies in turn allow them to access to information they did not have access to before.

We want to be clear that our findings do not suggest that censorship is ineffective in general or that increases in censorship will *always* backfire against governments. Our evidence suggests that through the gateway effect, there are circumstances where censorship *can* increase access to information. As we explore in our discussion, our findings suggest that gateway effects are most likely when increases in censorship are sudden and blunt – when applied during emerging protests or crises, for example.[1] By motivating more people to acquire the ability to evade censorship, a sudden increase in censorship can erode its own effectiveness, can politicize previously apolitical citizens, and can accumulate collective action potential that it often seeks to suppress.

The paper proceeds as follows. First, we describe the previous literature on the effects of censorship and introduce the logic behind how censorship can increase access to information and politicize previously apolitical users. Second, we describe our research design and the case we use to test our theory – the sudden block of Instagram in China in September 2014. Using our data, we describe the types of users who evade censorship before the sudden block of Instagram. Then, we show how the Instagram block increased access to information for the majority of Instagram users. We end with implications of this research, including a description of the circumstances under which censorship will lead to the gateway effects that we describe.

Censorship and Access to Information

A large literature has contributed to our knowledge of how censorship influences the political views and behavior of its intended targets. Most of the literature finds that censorship generally decreases access to information among citizens. Scholars have shown that by reducing connectivity to information sources, censorship and repression can successfully restrict information and demobilize individuals from engaging in the behavior the government deems objectionable (Enikolopov, Petrova and Zhuravskaya, 2011; Pierskalla and Hollenbach, 2013; Edmond, 2013). Others have found that when media is restricted in authoritarian governments, citizens access what is readily available to them (Stockmann, 2012; Geddes and Zaller, 1989) and do not discuss alternative information with each other because they do not know what others believe (Kuran, 1997).

Government control of the media can also provide signals to citizens to follow the gov-

[1] Censorship might be more sudden and thus less effective when collective action events are decentralized and spontaneous (Steinert-Threlkeld, Forthcoming).

ernment line. Huang (2015) finds that propaganda in China can signal government strength and Stern and Hassid (2012) find that the ambiguous nature of off limits information can induce risk-averse journalists to self-censor instead of spreading information. Censorship may even have adverse effects on information outside of a country by reducing incentives for citizens of other countries to engage in social media without a wider audience (Zhang and Zhu, 2011). Perhaps it is no wonder, then, that governments tend to censor during time periods that are dangerous to them, in particular when they are threatened by large-scale collective action (King, Pan and Roberts, 2013, 2014).

Even the Internet, which many scholars thought would be a form of 'liberation technology' (Lynch, 2011; Diamond, 2010; Howard, 2010; Shirky, 2008), is thought be be highly affected by government censorship. 'Repression technology' on the Internet, in the form of search filtering, content filtering, or blocks from Firewalls are thought to effectively decrease access to information in authoritarian regimes (Kalathil and Boas, 2010; Morozov, 2012; MacKinnon, 2012; Rød and Weidmann, 2015). Even when autocrats facilitate access to information, scholars have found that they do so strategically, for information gathering purposes, indicating that many autocrats engage in sophisticated and calculated censorship (Lorentzen, 2014; Egorov, Guriev and Sonin, 2009).

A few empirical studies have provided evidence to the contrary, suggesting that sometimes disruption of the media seems to have the opposite effect, stimulating increased public engagement or providing opportunities for the opposition (Rasler, 1996; Hassanpour, 2014). Explanations for why censorship may backfire often concentrate on political and social evaluations of increased censorship, including signaling and psychological models of reactance. In popular culture, this type of backlash against censorship has commonly been known as the "Streisand effect,"[2] which occurs when information is attempted to be censored and, in the process, gains increased attention and increases in perceived value because the censorship efforts unintentionally publicize it to an audience who opposes censorship. For example, when Wikimedia, a Wikipedia-like website in France, was allegedly asked by France's intelligence agency to remove part of an article, the report of the attempted censorship launched an international campaign against it, and the page ended up becoming the most-viewed page on website.[3] In this case, censorship itself caused political backlash that attracted attention to the information the government had hoped to make less accessible.

In this vein, scholars have argued that increases in censorship or sudden media disruptions may at times anger a broader population that opposes censorship. Evidence of Streisand effects around the world have led scholars to categorically call censorship futile (Nabi, 2014) and are a potential explanation for some instances of censorship backfire (Jansen and Martin, 2003). In some cases, increased censorship might signal regime weakness (Roberts, 2014), poten-

[2]Named for Barbara Streisand, who in an attempt to have pictures of her home in Malibu, CA removed from the Internet instead brought more attention to them.

[3]`http://www.economist.com/blogs/economist-explains/2013/04/economist-explains-what-streisand-effect`

tially creating common knowledge that has been shown to coordinate collective action (Kuran, 1991; Lohmann, 1994; Chwe, 2001). In other cases, media disruption removes regime-solidifying distractions, decentralizing information sharing to face-to-face interactions that can accelerate collective action (Hassanpour, 2014; Kern and Hainmueller, 2009).[4]

Theory: Gateway effects in information access

For the most part, the literature has posited that censorship acts to restrict information access and that citizens are unlikely to have any recourse. In the cases that evidence of censorship backfire has existed, it has relied on emotional and political backlash where citizens access the information because they are aware of government efforts to suppress it. This often requires a well-organized opposition that can make the public aware of government censorship (Jansen and Martin, 2003; Hess and Martin, 2006). Here, we suggest that there may be more circumstances outside of backlash where censorship does not produce its intended effects. In particular, we posit that censorship has the capability to politicize and empower previously apathetic citizens without a well-organized effort.

We propose a 'gateway effect' mechanism to explain how sudden censorship can increase information access. This gateway effect occurs when, because of censorship, citizens seek out the *same* information that they previously had access to by acquiring tools for censorship evasion – networks or technologies that enable them to bypass government restrictions on information. These citizens then gradually become exposed to off-limits information, as this information is suddenly very easy for them to access.

Individuals might not intend to seek out political information – and they do not necessarily begin to seek out off-limits information in order to arm themselves against the government. The gateway effect is driven by habit: citizens are used to accessing the newly censored information and would like to continue to do so. Information that is indispensible or addictive will produce stronger gateway effects when censored than that which can be easily substituted.

This access will accumulate over time, unless the government continually adjusts its behavior to counter that cumulative effect. Sudden censorship will arm citizens with tools for bypassing government restrictions as they seek these tools out to access newly censored information. Once censorship evasion tools are acquired and learned, they can be used more easily in all situations. Censoring more information sources further increases the total benefits of evasion. Evasion technology will grant access to a broader array of off-limits media with each website block.

Consider a few examples of this gateway effect. Citizens might participate in a religious organization that is suddenly banned by the government, forcing the organization underground. To continue worshiping, the religious organization may then meet in time periods or places that are less likely to be detected by the government, and in doing so, may be exposed to other organizations that are using the same strategies to evade government censure. Alternatively,

[4]Streisand effect arguments are similar to that in some of the repression literature, where government violence can cause political backlash. See Davenport (2007) for an overview.

books by a particular author may suddenly become banned by the government. Loyal readers of the author may then seek out black market book stores in order to continue following the author, which would then become a gateway to a list of other off-limits titles.

Gateway effects, of course, are not specific to government censorship. For example, changes in alcohol consumption, drug use, and prostitution during the Prohibition Era in the United States could have displayed this dynamic. During the Prohibition Era, alcohol consumption was approximately 60% to 70% of pre-prohibition consumption (Miron and Zwiebel, 1991). In other words, many Americans kept drinking. Because underground alcohol distributors were also engaged in other black markets, prohibition could have given habitual drinkers ready access to many other illicit goods and services. More recently, a desired reduction in this shared provider gateway effect (perceived to be caused by shared markets and social networks for both cannabis and harder drugs) helped drive policies in the Netherlands that permitted the sale of cannabis in coffee shops (MacCoun, 2010).

The logic of this proposed mechanism contrasts with that of a "Streisand effect" because it does not require backlash against the censorship itself, whether that backlash is the product of anger or mere curiosity to see what was blocked. In a Streisand effect, access to the newly censored information should increase, as citizens become more interested in the information *because of* censorship. In contrast, in the gateway effect overall access to the newly censored information decreases since it has become more difficult to access – users do not rally around the censored information. Instead, access to *other* information that has long been banned increases incidentally and even without political motivation among those who seek out new technologies or networks to access the information they are in the habit of consuming.

The Great Firewall and Geo-location

In this paper, we study the gateway effect in detail in the context of China's Great Firewall. The Great Firewall of China blocks foreign websites from mainland Chinese IP addresses, preventing Chinese citizens from accessing websites that the government deems objectionable. Mainland Chinese users – importantly not including Hong Kong users - cannot access a whole host of foreign social media platforms from Twitter to Facebook to Google.[5]

In order to access these websites from China, a user must first log in to a computer outside of China using a Virtual Private Network (VPN) or a proxy. VPNs are not terribly difficult to acquire for a sophisticated user and are not illegal to use – citizens are not sanctioned for using them. However, the small costs in terms of time and money for Chinese users in combination with the availability of attractive alternative social media websites like Sina Weibo (the Chinese version of Twitter) and WeChat (the Chinese version of WhatsApp) mean that relatively few users regularly 'jump' (evade) the Wall.

[5] `Greatfire.org`, which monitors the constantly updating group of blocked sites from China, estimates that 131 of 937 Alexa top 1000 domains are blocked in China, as of May 16, 2016, including media sites such as the New York Times, the Wall Street Journal, Reuters and Bloomberg to social media sites like Blogspot, Twitter, Facebook, and Youtube.

Those who do use a VPN to evade censorship can obtain full access to blocked social media websites and can use the Internet without restrictions. Like their counterparts around the world, these users will sometimes use geo-location to 'tag' themselves at a particular location. When users tag themselves in mainland China on blocked websites like Instagram or Twitter to specific locations, we know that they are using evasion technology to access these social media websites. We can estimate the dynamics of evasion of the Great Firewall by examining how the number of Instagram and Twitter posts geo-tagged to China changes over time and comparing these rates to locations which are not subject to the Firewall, like Hong Kong. Of course, geo-located users are not a random sample of the entire population of users. To ensure that our findings are generalizable, we supplement the geo-located social media data with data from sources that do not rely on geo-location, such as application download statistics, Wikipedia page views, and accumulation of Twitter followers, which we describe in detail in the next section.

Data

To estimate the dynamics of censorship evasion in China, we first collect a sample of 419,279 geo-located Instagram posts from mainland China, between September 1, 2014 and October 30, 2014. To obtain a representative sample, we randomly sampled coordinate grids in mainland China proportional to their population and downloaded all geo-located Instagram posts for the time period with each sampled grid.[6] Instagram was not blocked by the Great Firewall until September 29, 2014, when it was suddenly blocked during the third day of large-scale protests in Hong Kong.[7] This allows for a convenient discontinuity to estimate what proportion of users persisted in accessing Instagram even after it was blocked. It also allows us to explore the types of users for which the block expanded access to information and the types of users who did not seek or succeed in evading censorship.

Second, we use application download statistics from the app tracker App Annie to directly measure how much the Instagram block spurred new downloads of censorship evasion technology. App Annie tracks the relative rank of application downloads on iPhones by country. We use this data source to measure the popularity of new downloads of VPN applications in China around the time period of the Instagram block.

Third, we study how the Instagram block influenced the popularity of websites that have been blocked in China for an extended period of time. We use App Annie to measure new application downloads that direct to blocked websites, such as Facebook and Twitter. In addition, we collect 2,321,861 geo-located Twitter posts from mainland China, beginning in March of 2014, using the Twitter garden hose, approximately 1% of all Tweets.[8] Twitter has been blocked by

[6]We sampled approximately 25% of the grid. A users' Instagram post will appear in our data if it was posted with geo-location from one of our sampled cells in mainland China during September or October 2014.

[7]To our knowledge, no other websites were suddenly blocked that day if they had not been blocked before.

[8]This was part of a broader collection of worldwide tweets. Using the Twitter API, all worldwide geo-located tweets were requested every hour starting in March 2014. Because the garden hose only allows for a total of 1% of worldwide (geo-located or not geo-located) tweets to be collected, if geo-located tweets were greater than 1% of all tweets that hour, the request returns a random sample of all geo-located tweets up to 1%. Scholars have estimated

the Great Firewall since 2009, therefore users whose posts are geo-located to mainland China are using evasion software to access the website. We use this data set to study new users who sign up for Twitter from China around the time period of the Instagram block. For comparison, we use 1,773,678 geo-located Twitter posts from Hong Kong over the same time period.[9] Hong Kong is not subject to the constraints of the Great Firewall, but is similar to China culturally and linguistically.

Fourth, we collect a sample of 238,918 geo-located social media posts from Sina Weibo, the Chinese version of Twitter, as an additional comparison set. We use these data in the subsequent section to compare the types of users who are savvy and interested enough to participate in Twitter by evasion to those who participate in the more accessible Sina Weibo.

Finally, we compare page views of blocked Chinese language Wikipedia pages before and after the Instagram block using Wikipedia page view data.[10] We also estimate how the Instagram block affected the accumulation of new followers of Chinese activists and blocked news websites on Twitter. These data do not rely on geo-location and directly measure whether the block increased access to off-limits information through page views and Twitter follows.

Research design

In this section, we describe the empirical setup of our argument that increased censorship can lead to expanded access to information by disrupting users' habitual behavior. Our empirics will demonstrate that Chinese users will begin to access blocked information (and social networks) because a website that they habitually access becomes suddenly censored, inspiring them to acquire evasion technology. We demonstrate this finding with four sequential tests, summarized in Table 1. First, we estimate the proportion of users who stay on Instagram after the block. Because the block completely prevented Chinese Internet users from accessing Instagram, we can be confident that anyone accessing Instagram from China is doing so through a VPN. Second, we test whether the block inspired the acquisition of censorship technology by examining download ranks for popular VPNs from Chinese iPhone users. Third, we test whether the block inspired the use of websites that have long been blocked by the Chinese government by measuring Facebook and Twitter application downloads and directly measuring activity of Chinese social media users on Twitter. Last, we estimate whether the block resulted in users engaging more in political information by estimating how the participation of new users affected the popularity of blocked media and activists on Twitter, measuring changes in access to off-limits information on Wikipedia, and testing whether the Instagram users began engaging in political conversation.

that geo-located tweets are approximately 2-3% of the whole Twitter sample (Leetaru et al., 2013). This means that our collection contains approximately one half to one third of all geo-located Tweets, and approximately 1% of Twitter users in China.

[9] These posts were collected in the same worldwide API call as the mainland China tweets.

[10] Page view data at `http://stats.grok.se/`

Question	Test
1. Do users evade censorship to access Instagram?	Number/proportion of Instagram users who *continue* to access site after block
2. Does total censorship evasion increase?	VPN application download ranks
3. Do users access other, already blocked information?	Facebook/Twitter application download ranks, active Twitter users, Twitter sign-ups, Wikipedia page views
4. Do users absorb and engage with blocked political information?	Mentions of Hong Kong protests on Twitter and number of followers of Twitter activists and media

Table 1: *Empirical tests.*

Who Typically Evades Censorship?

Before we analyze how user behavior is affected by the sudden increase in censorship, we use a comparison of geo-located users on Twitter and geo-located users on Sina Weibo, the Chinese version of Twitter, to describe the profile of the typical Chinese user who evades censorship before the Instagram block. Those who are already using a VPN to jump the Firewall at the time of the block would not be affected by the block as they already have access to off-limits information. We include this section in order to 1) describe the static efficacy of the Great Firewall and 2) describe the profiles of those who are already evading censorship in order to show what types of social networks new individuals would be introduced to.

Consistent with previous research, the data on Twitter users in China before the Instagram block reveals that very few individuals regularly evade the Great Firewall of China and the Firewall practically prevents many users from accessing information blocked by the Wall. Previous studies of censorship evasion in China have estimated that 3%-15% of Chinese Internet users engage in censorship evasion.[11] Consistent with this research, we find that very few users post from China on Twitter, particularly in comparison to similar locations where Twitter is not censored. In our sample of geo-located Twitter posts, on average 1,600 unique Twitter users geo-locate to mainland China every day. Since we believe our sample covers approximately 1% of all Twitter posts from China, we expect that there are around 160,000 Twitter users who tweet each day from mainland China, or about 0.026% of all Internet users in China.[12] In comparison, on average 900 unique Twitter users every day in our sample geo-located to Hong Kong.

[11] `https://cyber.law.harvard.edu/publications/2010/Circumvention_Tool_Usage`, `http://www.globalwebindex.net/blog/vpn-in-china`

[12] We estimate that there are 610,650,000 Internet users in China, approximately 45% of the Chinese population. If the late 2014 Twitter population was close to 2016 estimates (`https://techcrunch.com/2016/07/05/twitter-estimates-that-it-has-10-million-users-in-china/`), then this corresponds to 2% of Chinese Twitter users tweeting any given day.

This suggests that around 90,000 Twitter users tweet each day from Hong Kong, or about 1.7% of Internet users in Hong Kong.[13] This rough, back-of-the-envelope calculation suggests that Hong Kong has around 65 times more Twitter users per Internet user than China does, despite having similar language, culture, alternative social media sites like Sina Weibo and Wechat, and political reasons to join Twitter. The Hong Kong-China Twitter comparison suggests that the small costs of evasion of the Firewall are generally effective in keeping mainland users off of the blocked website, a finding that is largely consistent with previous research (Song, Faris and Kelly, 2015).

What types of people in China evade censorship to get on Twitter? Not surprisingly, those using Twitter from mainland China are much more likely to speak a language other than Chinese than those who are using Sina Weibo. When users sign up for a social media account, they can indicate the language they would like their social media account to appear in. Of the geo-located Twitter users before the Instagram block, only 39% of them indicated that their primary language was Chinese when they signed up for Twitter, and 48% indicated that their primary language was English. In comparison, 99% of Sina Weibo users indicate that their primary language is Chinese, and only .05% English. Many of the users of Twitter in China are either using Twitter as ex-pats or have extensive foreign language training and experience abroad. These users are likely to have more resources and reasons to evade censorship by jumping the Firewall. In the supporting information, we show that Chinese language Twitter users before the Instagram block are highly clustered in urban areas in comparison to Sina Weibo users.

Twitter and Weibo users are also quite different in terms of the content they post. The differences in content suggest that Chinese Twitter users are more technologically savvy and more interested in politics than the average Chinese user on Weibo. In Table 2, we use mutual information to calculate the words most associated with Twitter posts and most associated with Weibo posts. Mutual information measures of the amount of information a word contains about whether the document that the word is in is part of a class, in this case Twitter and Weibo (Manning, Raghavan and Schütze, 2008). We only consider Twitter and Weibo users where the account is registered in Chinese. Even so, Twitters users are much more likely to inject English into their Tweets than those on Sina Weibo – 15 out of the 35 words most associated with the Twitter posts are English rather than Chinese. In addition, Twitter users are much more likely to talk about technology – iPhone and Apple appear in the list of words associated with Twitter. Last, Twitter users are much more likely to talk about politics – political words like international, citizens, government, China, country, freedom, and the politically restive province Xinjiang (not shown, 35th most distinctive) are included in the words associated with Twitter, whereas there are no explicitly political words in the list associated with Weibo.

Consistent with previous research, this comparison suggests that the existence of the Firewall has created two different social media communities in China – one for ordinary individuals and one for individuals who find it worth their while to spend time and money to participate in

[13] We estimate that there are 5,254,000 Internet users in Hong Kong or about 74% of the population.

	Words Associated with Twitter	Words Associated with Weibo
1	im: i'm	自己: myself
2	发布: released	知道: know
3	香港: Hong Kong	今天: today
4	刚刚: just	播放: play
5	照片: photos	生活: life
6	北京: Beijing	努力: work hard
7	good: good	别人: others
8	the: the	看看: watch
9	night: night	mv: mv
10	中心: center	时候: time
11	和平: peace	什么: what
12	apple: Apple	永远: forever
13	day: day	通过: through
14	my: my	快乐: happy
15	国际: international	开心: happy
16	will: will	坚持: persistence
17	公民: citizens	喜欢: like
18	政府: government	没有: none
19	中国: China	事情: situation
20	happy: happy	我们: we
21	人民: the people	明星: stars
22	see: see	生日: birthday
23	自由: freedom	觉得: think
24	国家: country	现在: now
25	like: like	心里: in my heart

Table 2: *Words most associated with Twitter and Weibo users, mutual information.* This figure shows that Chinese language Twitter users were more likely to use political words than Weibo users.

websites that are censored. While the typical citizen participates in domestic social media and might find participating in blocked social media sites inconvenient, those in China who jump the Firewall to participate on Twitter live in urban areas, are more technologically savvy, seem to have international language experience and are relatively likely to discuss politics.

1. Direct effects of the Instagram block

What happens to the size and composition of Instagram users in China when censorship increases? On September 26, 2014, Hong Kong protesters began sit-in protests in response to reforms being considered by the Hong Kong government to the electoral system. On September 29, 2014, the Chinese government suddenly disrupted social media by blocking Instagram, a popular social media website used for posting photos.

Instagram was widely used in China before it was suddenly blocked. In 2011, when Instagram was first taking hold in China, one of the founders of the site indicated that Instagram had around 100,000 downloads each week from China.[14] In the month before Instagram was blocked in China, it was in the top 100 Apple application downloads for users in China and in

[14] `https://www.techinasia.com/instagram-china-weibo`

the top ten for photo and video sharing.[15]

The Instagram block had a large, negative impact on the number of people using Instagram from China. Our data indicate that a little less than half of the users who were previously geo-locating to China using Instagram discontinued their use of Instagram. As shown in Figure 1, before the block, our sample contains on average 6,368 unique users geo-locating to China every day. After the block, the number of unique users within our sample geo-locating with Instagram decreases to a mean of 3,376 (note that the vast majority of Instagram users do not geotag posts and most Instagram users do not post photos on a given day, so we observe only a small fraction of activity). If, similar to Twitter, only 2 to 3% of Instagram users geo-tag their posts and we collected 1 out of 4 geo-tagged posts,[16] then there were approximately 1 million unique users posting to Instagram from China every day before the block, and half a million unique users posting to Instagram from China after the block.

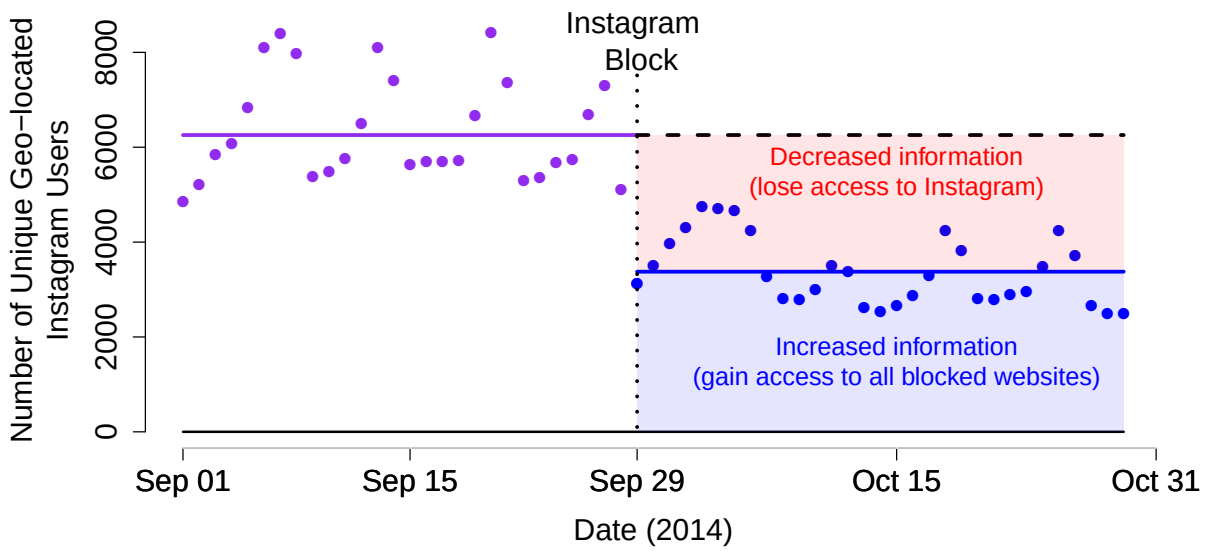

Figure 1: *The Instagram block's effect on the number of unique Instagram users geo-locating from mainland China.* The Instagram block led to an approximately 50% decrease in active, geo-located Chinese Instagram users. The x-axis in this figure is the date and the y-axis is the number of unique users on Instagram who added a geo-tagged post to the site that day (a small fraction of all users). We add horizontal lines for the before (purple) and after (blue) Instagram block means of the number of daily, unique, geo-tagging users, as well as the number of these users we would have expected had there been no block (black dotted line). The blue shaded area highlights that 50% of active Chinese Instagram users were accessing an uncensored version of the Internet after Instagram was blocked, while the red shaded area highlights that 50% of Chinese Instagram users were no longer active on Instagram after it was blocked.

[15]Data from application analytics firm App Annie.

[16]We randomly sampled and searched grid coordinates that covered the residences of approximately one fourth the population of China.

The evidence provided here shows that the Instagram block did not create a Streisand effect as censorship of Instagram did not create *more* interest in or attract new users to Instagram. Yet, given that experts have long estimated that very few people in China use or have ever used Virtual Private Networks, the persistence of Instagram users should be surprising. A full 53% of Instagram users continued using Instagram, despite the block. While there is no direct data available on the total number of Instagram users in China, survey data suggests that approximately 5% of Internet users use Facebook.[17] Facebook – which is blocked in China – was downloaded with less frequency than Instagram before the Instagram block according to AppAnnie. To give a very rough estimate of the extent of Instagram use, if 5% of Internet users in China used Instagram before the block, there would be 30,532,500 Instagram users in China. Roughly, this would suggest that around 16.1 million people – three times the number of all Internet users in Hong Kong – were motivated enough to evade the Firewall in the days following the block, in order to access a social media site that they were accustomed to using.[18]

2. Effects of the Instagram block on VPN acquisition

It could be that the users who persisted in using Instagram were already using VPNs and therefore the block did not increase their access to information. However, the applications analytics website, App Annie, suggests that the Instagram block caused a large, sudden increase in new VPN application downloads, suggesting that the Instagram block created millions of *new* VPN users. On September 28, 2014, VPNs did not make it on the top 10 downloaded productivity applications for iPhones. But on September 29, 2014, four of the top ten downloaded free productivity applications in China were VPNs: VPN Express, GreenVPN, VPNArtifact, VPN in Touch. To take one stark example, on September 28, 2014, VPN Express was the 1,229th most downloaded iPhone app in China. By the next day, September 29, 2014, it was the 6th most downloaded iPhone app in China. The right panel of Figure 2 shows the time series rank of two of these VPN applications during the time period surrounding the Instagram block.

Consistent with our hypothesis that habits were driving censorship evasion, those who continued to use Instagram by downloading a VPN were those who were most accustomed to using Instagram. We use the numbers of likes and comments for users on Instagram as a proxy for the amount of interaction each user experiences on Instagram. Figure 7 in the supporting information shows the difference between the number of likes and comments before the Instagram block between the users who stayed on Instagram in comparison to those who stopped using Instagram. Those who continued to use Instagram had substantially more interactions with other users before the block, including more comments and likes.

[17] `http://www.slideshare.net/wearesocialsg/digital-in-2016`

[18] Note that, typically, a small fraction of social media users post content on a daily basis and the 'lurkers' show up in our daily activity data.

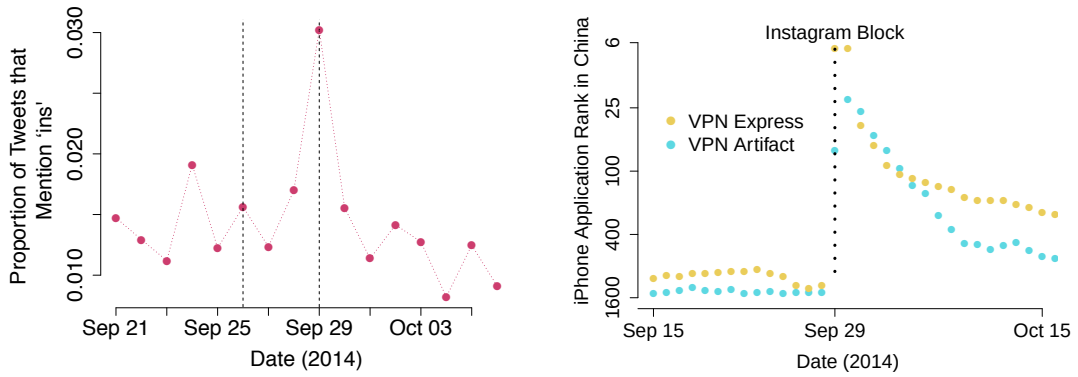

Figure 2: *Left: Proportion of Tweets from China mentioning 'ins' by day. Right: The Instagram block's effect on the rank of VPN applications on iPhones from mainland China, from AppAnnie.com.* In the left panel of this figure we show that 3% of tweets in China mentioned Instagram on the day of the Instagram block. The right panel of this figures shows that the download ranks of VPN Express and VPN Artifact increased from ranks lower than 1,000 to the top ten most popular applications in China on the day of the Instagram block.

3. Expanded Access to Blocked Websites

Users who downloaded a VPN to continue to access Instagram did not just remain on Instagram, but also appear to have begun interacting with a number of websites that have long been off-limits in China. Websites already blocked in China saw a surge in use from China. The left panel of Figure 3 shows the number of application downloads of Twitter and Facebook – both blocked from mainland – in China. Both applications peak on the day of the Instagram block, indicating that the new users evading censorship expanded their presence on social media to websites that had long been blocked and that they had not previously interacted with.

In the sample of Twitter data we collected geo-located to China, we see unprecedented numbers of new users joining Twitter. The right panel of Figure 3 shows the effect of the Instagram block on Twitter use. The increase in Twitter use was driven largely by new account creations. Figure 8 in the Appendix shows new account creations by day. On the day of the Instagram block, the rate of new user account creation jumped more than 600%.

Of course, it could be that the new users of Twitter joined not because of the Instagram block but instead to follow developments in the Hong Kong protests. However, our data do not support this interpretation of events. First, we show in Figure 3 that the timing of the user increase corresponds with the Instagram block and not the protests. The Hong Kong protests began on September 26, and the first two days of the protest did not seem to generate enough interest among mainland users to significantly increase the number of users on Twitter. The increase in new users also corresponds to a spike in conversation about Instagram (left panel of Figure 2). A full 3% of tweets within our sample use the word "ins" on September 29,[19] three

[19]We use "ins" because it will pick up both those using the word "instagram" and also it's nickname "ins".

times the amount that it is typically used within the sample.

Second, if new Twitter use were inspired by the Hong Kong protests, we would expect a similar increase in the area most affected by the protest – Hong Kong. The right panel of Figure 3 shows that the jump in Twitter use only affects mainland China. This suggests that the increase in Twitter use is not due to events within Hong Kong, but indeed due to the Instagram block, which only affected mainland China.

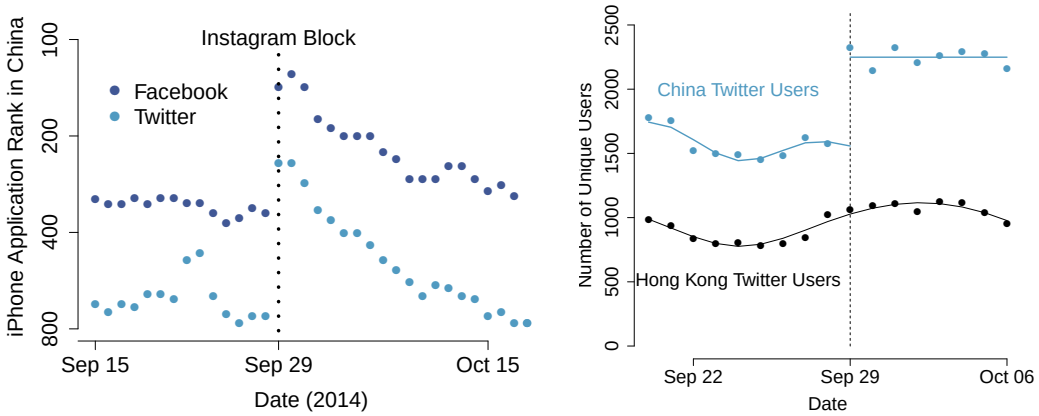

Figure 3: *Left: The Instagram block's effect on the rank of Facebook and Twitter on iPhones from mainland China, from AppAnnie.com. Right: Comparison of tweets per day from Mainland China and Hong Kong before and after the Instagram block.* The left panel of this figure shows the change in download ranks for Facebook and Twitter before and after Instagram was blocked. The right panel of this figure shows that the Chinese Twitter users in our sample increased 30% the same day that we observe a spike in Instagram mentions and several days after the beginning of the Hong Kong protests. This increase only occurred in China and not in Hong Kong. The lines in this panel were fit using a smoothing spline.

To further understand the sudden increase in Twitter users on September 29, we take a closer look at the new, geo-locating users within our sample who joined Twitter on September 29th and compare them to existing Twitter users from China. Overall, these new users look very different from the typical Twitter user, indicating that the Instagram block encouraged less political, more typical Chinese users to jump the Firewall. First, the new users are much less international than the typical Chinese Twitter user. 80% of the new Twitter users indicated that their preferred language was Chinese, in comparison to only 39% of regular Twitter users. Only 18% of new users indicated that they preferred English, compared to 49% of regular Twitter users. New users are also much less likely to reside in China's most international cities. Whereas approximately 30% of regular Twitter users geo-locate to Beijing and Shanghai, only 15% of the new users were geo-located to these two major cities.

In addition to basic user characteristics, we also coded the first 10 people each of the new users followed. We find that almost 80% of the users the new users followed fall into the

category of Arts & Entertainment or Sports. One new Twitter user best summarizes the new users' motivation to join Twitter, suggesting that the motivation for joining Twitter was not political, but rather incidental:

> "Today Chinese not on INS, want to download something to use, then try the twitter, the results can, can pay attention to Justin Bieber proud" [sic]

4. Politicization

The Instagram block precipitated a large increase in followers for Chinese Twitter users, enhancing the network of the blocked website among mainland citizens. Not only did new users follow celebrities, they also began following political accounts including accounts of news sites such as New York Times Chinese and Apple Daily (a Hong Kong news source highly critical of the mainland Chinese government). We show in the right panel of Figure 4 the total additional followers to these accounts, by day, compared to their average in the 30 days preceding the block. Overall, by November 1st, there were 33,750 more followers of Chinese Twitter users than we would have expected based on data from the month before the block.

Even though the new Twitter users were not originally interested in politics, preliminary evidence suggests that they started becoming politically interested a few days after joining Twitter. We code the first ten tweets of each of the new users starting September 29th into categories of whether or not these Tweets are political mentions of Hong Kong. Only 1% of the first ten tweets of new users mention politics in Hong Kong, in comparison to the first ten tweets of old users on Twitter on September 29th, where 3% of them mention politics in Hong Kong.

The left panel of Figure 5 displays the proportion of Tweets that mention politics in Hong Kong for new users and old users. Although new users show relatively little interest in Hong Kong their first day on the site, by the second day, they begin to look like the other Twitter users. As one new Twitter user, whose Twitter activity mostly revolves around a Thai pop star, puts it: "Something have happened in HK, but people on mainland do not know because of the government of mainland, pray for you."

These effects extended beyond geo-located social media and application downloads. We were able to measure off-limits information access directly by using page views of Chinese language Wikipedia pages (zh.wikipedia.org) that were blocked in China before and after the Instagram block.[20] The pages we study are political, as only *specific*, political pages were blocked at the time in mainland China, including pages such as "human rights" (人权) and Chinese dissident Liu Xiaobo (刘晓波).

In Figure 5, we show that there was a large spike in views of Chinese language Wikipedia pages blocked in China on the days after the Instagram block. There were approximately 160,000 more page views of blocked Wikipedia pages on September 29th than in the week preceding the

[20]Page view data at `http://stats.grok.se/`. Data on which Wikipedia pages were blocked on September 29,2014 was taken from `greatfire.org`

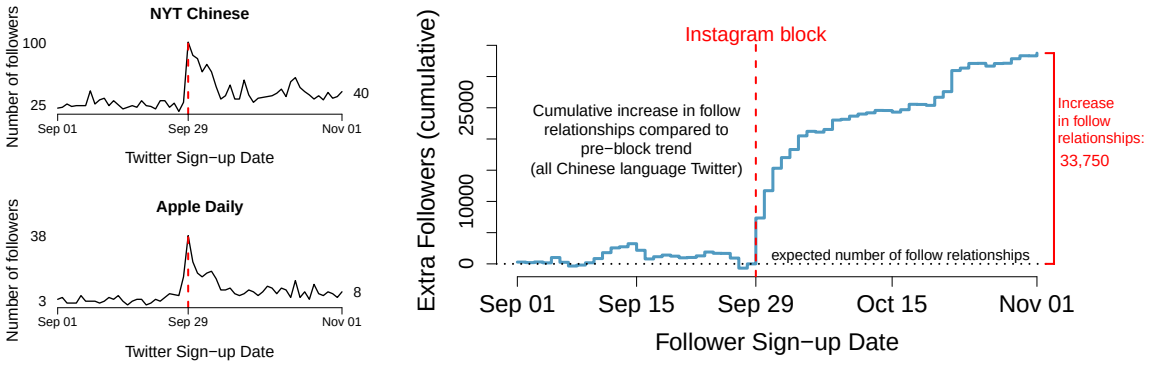

Figure 4: *Left: Daily new followers to New York Times Chinese and Apple Daily Twitter accounts (based on new user sign-up dates). Right: Cumulative increase in followers, compared to pre-block trend, of any Chinese language user (based on new user sign-up dates) compared to expected increase in followers.* The left panel of this figure shows the sign-up dates of followers of the New York Times Chinese and Apple Daily Twitter accounts. Many followers of these accounts signed up for Twitter immediately following the Instagram block. This increase in sign-ups – users who eventually followed NYT Chinese and Apple Daily – continues long after the Instagram block. The right panel of this figure shows that all Chinese language Twitter users accumulated approximately 33,750 more followers from new Twitter sign-ups than what we would expect based on pre-block trends. This cumulative increase was calculated using a cumulative sum of the number of new followers minus the number of expected followers, where the expected followers was the mean daily number of new followers prior to the Instagram block.

Instagram block. Unlike other analyses, we were not able to distinguish worldwide page views of Chinese language Wikipedia pages from page views coming from Mainland China, however, the patterns of information access very closely reflected those in our geo-tagged results. We find an effect specifically among Wikipedia page views were limited to presumably political pages deemed unfavorable by the Chinese government.

Implications

The Instagram block resulted in the unintended consequence of giving a large number of 'normal' Chinese citizens incentives to evade censorship and therefore access to websites and information that many had likely never previously encountered. What situations do we expect that the gateway effect will be relevant? And what are the implications of this expanded access to information? In this section, we describe the situations in which we expect that gateway effect to be strongest and the fallout of this increased access in terms of the government's reaction and user engagement in political conversations.

The increased activity in political discussion on the part of new users could shed light on the mechanisms that underlie existing empirical evidence that media disruption can increase the potential for collective action. However, we do not expect that the gateway effect will be important in all instances of increased censorship. We hypothesize that the increase in access

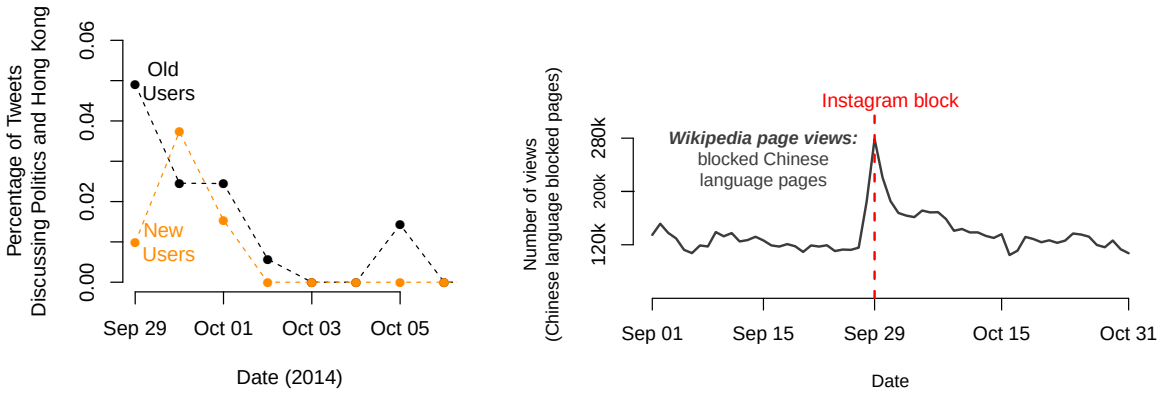

Figure 5: *Left: Tweets that mention politics in Hong Kong, comparison of new users and old users. Right: Page views for Chinese language Wikipedia pages blocked in China.* The left panel of this figure shows that users who signed up for Twitter after the Instagram block began mentioning protest events in Hong Kong about a day after their arrival on the site. The right panel of this figure shows page views of Chinese language Wikipedia pages that were blocked in China before and after Instagram was blocked. The increase in Wikipedia page views was more short-lived than the increase in Twitter follow relationships. Users newly accessing off-limits information appeared to check the relatively static, blocked Wikipedia pages in a one-off burst.

to information will occur primarily in two situations 1) when censorship is sudden enough to disrupt citizen habits and 2) when the newly censored information is difficult to substitute with uncensored alternatives.

Sudden censorship is more likely to cause the habit disruption that creates gateway effects. When users have full access to a website one day which is completely blocked the next, they may be in the middle of conversations or projects that create short-term incentives for them to seek out evasion tools. The gateway effect that sudden censorship creates may explain why the Chinese government has opted for more gradual blocks of websites in the past. In 2010, following a conflict with Google, the Chinese government throttled Google for years, choosing not to outright block the website until 2014 (Milward, June 30, 2011). In retrospect, this may have been an effort to wean users off of Google without creating immediate incentives to jump the Firewall.

Unlike the Google block, which was precipitated by a conflict between the tech giant and the Chinese government, censorship in reaction to collective action events may necessarily be sudden. In the case of the Instagram block, the Chinese government was purportedly worried about images of police using tear gas on protesters in Hong Kong circulating to mainland citizens on Instagram.[21] If these images were the most immediate, dangerous piece of information that

[21] `http://www.cnn.com/2014/09/29/world/asia/china-censorship-hong-kong/`

could accelerate the protest, a sudden block that decreases use of Instagram could be strategic for a government even if it increases access to evasion technology and other blocked website for millions of people in the longer term.

Second, gateway effects are more likely when the newly censored information has fewer substitutes. It is increasingly clear that one of the primary information control strategies of the Chinese government is to create distractions to off-limits information (King, Pan and Roberts, 2016). The government has actively encouraged the development of Chinese social media websites to provide substitutes for those that are blocked by the Firewall. Instagram, one of the most popular social networking websites in the world, not only has no clear substitute in China, but has been shown to be more addicting than websites (Roberts, Yaya and Manolis, 2014). Social networks formed on Instagram may be difficult for consumers to replace. The results described in this paper therefore may not hold for other websites.

The Government Reaction: VPN Crackdowns

The longer term impact of the gateway effect will be conditional of the government's strategic reaction to increased citizen evasion. Not long after the Instagram block, the Chinese government began increasing censorship in a different way – by raising the Great Firewall. Instead of blocking more websites, the government instead increased the difficulty of successfully jumping the Firewall. Though the crackdown on VPNs does not appear to have happened in one day, reports of difficulties of accessing VPNs begin in November 2014, culminating in January 2015 when Beijing is said to have 'upgraded' the Firewall. During this time period, VPN providers began reporting disruptions in their services and users reported being increasingly frustrated with VPN access.[22]

While we cannot tell for certain whether this crackdown in the months following the Instagram block was in reaction to the increased numbers of people evading censorship, the Chinese government's upgrade of the Firewall after the Instagram block was conceivably an effort to reduce a ballooning censorship-evading population. This implies that autocrats will have incentives to follow expansions in censorship like the Instagram block with crackdowns on censorship evasion. The VPN crackdown following the block is consistent other empirical evidence that media disruptions are accompanied by larger crackdowns, but suggests an alternative mechanism (Gohdes, 2015).

If governments frequently make evasion more difficult following an expansion of censorship, this likely churns networks of those engaged on blocked social media sites creating unknown implications for collective action. Some individuals who already access blocked information may drop out once evasion becomes more costly. This removes individuals for whom evasion is difficult or who have limited motivation to pay the costs of evasion. At the same time, as governments block more websites, it will create incentives for new individuals to jump the firewall

[22] http://www.theguardian.com/world/2014/nov/18/china-blocks-access-websites-before-hosting-world-internet-conference, http://www.reuters.com/article/us-china-internet-vpn-idUSKBN0KW0WS20150123

if they are accustomed to using the censored sites. This introduces newly censured individuals to existing, politically interested social networks and removes relatively disinterested and/or resource-constrained individuals.

The effects of this churn are currently unclear. On the one hand, the replacement of peripheral members in a social network could disrupt collective action potential by reducing continuity in a network (Barberá et al., 2015). New users might only gradually build new connections to replace old ones in the networks. On the other hand, new members could alter social dynamics in a network by introducing new ideas and behaviors to homogeneous and stale social groups, similar to dynamics observed in successful creative networks (Uzzi and Spiro, 2005). By disrupting old relationships and introducing unfamiliar ones to replace them, churn potentially increases the likelihood of emergent collective behaviors.

Conclusion

Despite conventional wisdom that increases in censorship strictly decreases access to information, we find that expansions in censorship to previously uncensored websites can incentivize citizens to invest in censorship evasion technology. Upon learning how to evade censorship, these individuals in turn have more, not less, access to information and begin engaging in conversations, social media sites, and networks that have long been off-limits to them. Censorship, unexpectedly, increases access to information for the individuals for whom it provides a reason to search out off-limits information. Our findings suggest a potential mechanism through which users actually have access to more information despite increased censorship. If censorship is applied to largely apolitical social media sites like Instagram, it can pull apolitical users into censorship evasion networks and engage them in political discussion and coordination.

Importantly, these patterns of behavior were not forms of psychological reactance – such as a "Streisand effect", where individuals explicitly seek out information that is being hidden. Instead, users gained access to more information simply by maintaining their current social media consumption (and there were fewer users of the newly blocked media itself).

In blocking Instagram, the Chinese government reduced the number of users on Instagram, but increased access to websites that are typically more political such as Twitter, Wikipedia, and Facebook. While this evidence seems to suggest that the sudden censorship of Instagram was a misguided move by the censors, we cannot say for sure. If Instagram were suddenly politicized by protests and police crackdowns in Hong Kong, then increasing the incentives for evasion of the Firewall may have been worth decreasing Instagram use from the perspective of Beijing. In its censorship efforts during a quickly unfolding crisis, Beijing was left with two bad options. We hope that future research can unravel the long-term political influence of these censorship efforts.

References

Barberá, Pablo, Ning Wang, Richard Bonneau, John T Jost, Jonathan Nagler, Joshua Tucker and Sandra González-Bailón. 2015. "The Critical Periphery in the Growth of Social Protests." *PLOS ONE* 10(11):e0143611.

Becker, Gary S. 1974. Crime and punishment: An economic approach. In *Essays in the Economics of Crime and Punishment*. NBER pp. 1–54.

Brown, William W and Morgan O Reynolds. 1973. "Crime and "punishment": Risk implications." *Journal of Economic Theory* 6(5):508–514.

Chwe, Michael Suk-Young. 2001. *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press.

Davenport, Christian. 2007. "State repression and political order." *Annu. Rev. Polit. Sci.* 10:1–23.

Deibert, Ronald, John Palfrey, Rafal Rohozinski, Jonathan Zittrain and Miklos Haraszti. 2010. *Access controlled: The shaping of power, rights, and rule in cyberspace*. Mit Press.

Diamond, Larry. 2010. "Liberation technology." *Journal of Democracy* 21(3):69–83.

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

Egorov, Georgy, Sergei Guriev and Konstantin Sonin. 2009. "Why resource-poor dictators allow freer media: a theory and evidence from panel data." *American Political Science Review* 103(04):645–668.

Eide, Erling, Paul H Rubin and Joanna Mehlop Shepherd. 2006. *Economics of crime*. Now Publishers Inc.

Enikolopov, Ruben, Maria Petrova and Ekaterina Zhuravskaya. 2011. "Media and political persuasion: Evidence from Russia." *The American Economic Review* 101(7):3253–3285.

Geddes, Barbara and John Zaller. 1989. "Sources of Popular Support for Authoritarian Regimes." *American Journal of Political Science* 33(2):319–347.

Gohdes, Anita R. 2015. "Pulling the plug Network disruptions and violence in civil conflict." *Journal of Peace Research* 52(3):352–367.

Hassanpour, Navid. 2014. "Media Disruption and Revolutionary Unrest: Evidence From Mubarak's Quasi-Experiment." *Political Communication* 31(1):1–24.

Hess, David and Brian Martin. 2006. "Repression, backfire, and the theory of transformative events." *Mobilization: an International Quarterly* 11(2):249–267.

Howard, Philip N. 2010. *The digital origins of dictatorship and democracy: Information technology and political Islam*. Oxford University Press.

Huang, Haifeng. 2015. "Propaganda as Signaling." *Comparative Politics* 47(4):419–444.

Jansen, Sue Curry and Brian Martin. 2003. "Making censorship backfire." *Counterpoise* 7(3):5.

Kalathil, Shanthi and Taylor C Boas. 2010. *Open networks, closed regimes: The impact of the Internet on authoritarian rule*. Carnegie Endowment for International Peace.

Kelly, Sanja, Sarah G Cook, Mai Truong and Freedom House. 2012. *Freedom on the Net 2012: A Global Assessment of Internet and Digital Media*. Freedom House.

Kern, Holger Lutz and Jens Hainmueller. 2009. "Opium for the masses: How foreign media can stabilize authoritarian regimes." *Political Analysis* 17(4):377–399.

King, Gary, Jennifer Pan and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107:1–18. http://j.mp/LdVXqN.

King, Gary, Jennifer Pan and Margaret E Roberts. 2014. "Reverse-engineering censorship in China: Randomized experimentation and participant observation." *Science* 345(6199):1251722–1251722.

King, Gary, Jennifer Pan and Margaret E Roberts. 2016. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument.". http://j.mp/1Txxiz1.

Kuran, Timur. 1991. "Now out of never: The element of surprise in the East European revolution of 1989." *World politics* 44(01):7–48.

Kuran, Timur. 1997. *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

Leetaru, Kalev, Shaowen Wang, Guofeng Cao, Anand Padmanabhan and Eric Shook. 2013. "Mapping the global Twitter heartbeat: The geography of Twitter." *First Monday* 18(5).

Lessig, Lawrence. 1999. *Code: And other laws of cyberspace*. Basic Books.

Lohmann, Susanne. 1994. "The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989–91." *World politics* 47(01):42–101.

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

Lynch, Marc. 2011. "After Egypt: The limits and promise of online challenges to the authoritarian Arab state." *Perspectives on politics* 9(02):301–310.

MacCoun, Robert J. 2010. "What Can We Learn from the Dutch Cannabis Coffeeshop Experience?" *RAND working paper* pp. 1–32. https://www.rand.org/content/dam/rand/pubs/working_papers/2010/RAND_WR768.pdf.

MacKinnon, Rebecca. 2012. *Consent of the Networked: The Worldwide Struggle For Internet Freedom*. New York: Basic Books.

Manning, Christopher D., Prabhakar Raghavan and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. NY: Cambridge University Press.

Milward, Stephen. June 30, 2011. "Google+ Not Actually Blocked in China, Just Being Slowly Throttled." *TechAsia*. `http://www.techinasia.com/google-plus-china/`.

Miron, Jeffrey A and Jeffrey Zwiebel. 1991. "Alcohol Consumption During Prohibition." *National Bureau of Economic Research* .

Morozov, Evgeny. 2012. *The net delusion: The dark side of Internet freedom*. PublicAffairs.

Nabi, Zubair. 2014. "Resistence censorship is futile." *First Monday* 19(11).

Pierskalla, Jan H and Florian M Hollenbach. 2013. "Technology and collective action: The effect of cell phone coverage on political violence in Africa." *American Political Science Review* 107(02):207–224.

Rasler, Karen. 1996. "Concessions, repression, and political protest in the Iranian revolution." *American Sociological Review* pp. 132–152.

Roberts, James, Luc Yaya and Chris Manolis. 2014. "The invisible addiction: Cell-phone activities and addiction among male and female college students." *Journal of behavioral addictions* 3(4):254–265.

Roberts, Margaret Earling. 2014. Fear, Friction, and Flooding: Methods of Online Information Control PhD thesis Harvard University. https://dash.harvard.edu/handle/1/12274299.

Rød, Espen Geelmuyden and Nils B Weidmann. 2015. "Empowering activists or autocrats? The Internet in authoritarian regimes." *Journal of Peace Research* 52(3):338–351.

Shirk, Susan L. 2011. *Changing Media, Changing China*. New York: Oxford University Press.

Shirky, Clay. 2008. *Here comes everybody: The power of organizing without organizations*. Penguin.

Song, Sonya Yan, Robert Faris and John Kelly. 2015. "Beyond the Wall: Mapping Twitter in China." *Berkman Center Research Publication* (2015-14).

Steinert-Threlkeld, Zachary C. Forthcoming. "Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring." *American Political Science Review* .

Stern, Rachel E and Jonathan Hassid. 2012. "Amplifying silence uncertainty and control parables in contemporary China." *Comparative Political Studies* 45(10):1230–1254.

Stockmann, Daniela. 2012. *Media commercialization and authoritarian rule in China*. Cambridge University Press.

Uzzi, Brian and Jarrett Spiro. 2005. "Collaboration and Creativity: The Small World Problem." *American Journal of Sociology* 111(2):447–504.

Zhang, Xiaoquan and Feng Zhu. 2011. "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia." *The American economic review* 101(4):1601–1615.

Supplemental Information

Theory: a simple model and a few predictions

To make our point especially clear, we borrow from the literature on the economics of crime to develop a very simple, general model of how censorship influences the number and types of individuals willing to engage in censorship evasion. The model is general enough to apply not just to censorship via the Great Firewall, but also to other forms of censorship, including banned books, off-limits religious organizations, or banned political speech. In each of these cases, like crime, individuals can participate in the off-limits behavior if they are willing to incur government-imposed physical or financial costs of doing so.

We assume that an individual derives particular benefits from evading censorship and also incurs costs. These costs include the cost of punishment (multiplied by the probability of punishment) and the costs of the actual mechanics of evasion of government restriction. The benefits may range from economic benefits (employees at multinational firms may need to evade censorship to perform their job), to less concrete or immaterial benefits such as risk-loving, thrill, or the ability to support a particular political cause for speaking out. As in the crime literature, we model the benefits and costs of an individual to participate in the banned activity with a simple equation (Brown and Reynolds, 1973; Becker, 1974; Eide, Rubin and Shepherd, 2006):

$$E(U) = p * U(W_i + I_i - P_i) + (1 - p) * U(W_i + I_i) - C_i$$

where U is an individual's utility function, p is the probability of being caught and punished, W_i is the utility from participating in in-bounds behavior, I_i is the additional utility of participating in banned behavior, P_i is the magnitude of punishment and C_i is the cost to the individual each time they participate in the restricted behavior. In the case of evading the Great Firewall using a VPN, citizens who 'jump' the Firewall are not typically punished, so p and P_i should be very low if not zero. However, citizens do incur time and financial costs. To evade censorship, citizens must find and sometimes purchase VPN software. They also must deal with Internet slowdowns associated with using VPNs. These types of costs would be incorporated in C_i because they are incurred whether or not the individual is punished.

Both costs (C_i, P_i) and benefits (W_i, I_i) vary by individual. For some individuals, say with

more education and more income, the barriers to evading censorship might be more trivial than for individuals who are less savvy or who have fewer resources. In cases where censorship evasion is punished, some individuals may be protected from punishment that comes from participating in banned activities because they are politically protected. Benefits may also vary by individual's occupation and their commitment to a cause. Benefits will vary heterogeneously depending on what the banned behavior is: a very religious person may be derive more benefits from participating in off-limits religious activities, for example, but an academic could be more affected by a book ban.

We complicate this simple model of evasion by adding fixed learning costs for those who have not engaged in banned behavior before. In order to evade government restrictions on behavior, individuals must learn how to do so. To evade the Firewall, individuals must buy and learn how to use censorship technology; to buy banned books, the person must know a black-market book seller. Once a person has engaged in banned behavior once, the cost of doing so again is lower. We therefore add a term F_i for the fixed cost to the economic cost and benefit equation, which only appears for individuals who have never engaged in the banned behavior:

$$E(U) = p * U(W_i + I_i - P_i) + (1 - p) * U(W_i + I_i) - C_i - F_i$$

Individuals will participate in off limits behavior when:

$$U(W_i) < p * U(W_i + I_i - P_i) + (1 - p) * U(W_i + I_i) - C_i - F_i$$

If people participate in banned activities when their expected utility of doing so is greater than the utility derived simply from participating in in-bounds behavior W_i, then this simple model has very straight forward implications for behavior of individuals under increased censorship. Except when there are extreme costs of censorship evasion, censorship will typically not deter all citizens from engaging in off limits behavior – like crime, some of the population will find it worth it to evade censorship. However, even small costs of evasion can keep many people who have low benefits of evading censorship from doing so. In equilibrium, we would expect that individuals who have lower costs of participating in off-limits behavior would be more likely

to do so. Those who have never participated in the restricted behavior before, and therefore those who have to pay a fixed costs to do so initially, would be less likely to participate in the restricted behavior. Individuals who are savvy, wealthy, and well-connected will be more likely to engage in banned behavior if these traits allow them to more easily evade restrictions.

Second, we would expect that those who have a higher benefit from participating in the banned behavior would be more likely to do so. These could be political benefits, such as political expression or organization. But there could also be non-political benefits to repression. For example, it might be that individuals' jobs or socializing with friends are tied to participating in the restricted behavior, which would increase their probability of participating in the restricted behavior.

How increased censorship impacts evasion behavior in this model will depend on how censorship increases. When censorship increases by banning more activities or types of information that were not already off-limits, if the person derives any utility from the newly banned activity, W_i will decrease and the magnitude of I_i will increase. If the government adds a new religious organization, a new book, or a new website to banned activities, any utility derived from those activities will move from in-bounds utility W_i to out-of-bounds utility I_i. This will increase participation in censorship evasion. It may also make participation in restricted behavior more likely in the long-run, as it will increase the number of people who have learned how to evade censorship and therefore decrease the sum of F_i across individuals in the population.

Alternatively, if the direct costs for participating in evasion increase, it should reduce the likelihood that people will participate in the restricted behavior. This could be an increase in the magnitude of punishment P_i, an increase in the probability of punishment p, an increase in the variable costs C_i, or an increase in fixed costs F_i. Increasing the cost of getting a VPN or cracking down on banned book sellers will make those interested in the material less likely to evade censorship to access it.

Description of users

When we subset users on Twitter and Weibo to include only those that indicate that their primary language is Chinese, we still see important differences between Twitter and Weibo users. To compare the locations of the two groups, we collected all geo-located Sina Weibo posts in Beijing and its surrounding areas during September of 2014. In Figure 6, we compare the distribution of Twitter and Weibo users in this area by plotting a point for each unique geo-located social media posts. We highlight highly populated areas using two-dimensional kernel density estimation. We see that even among Chinese users, Twitter are much more likely to be clustered in the major cities in this area, such as Beijing and Tianjin, whereas Weibo users are spread out across the entire area, including rural areas.

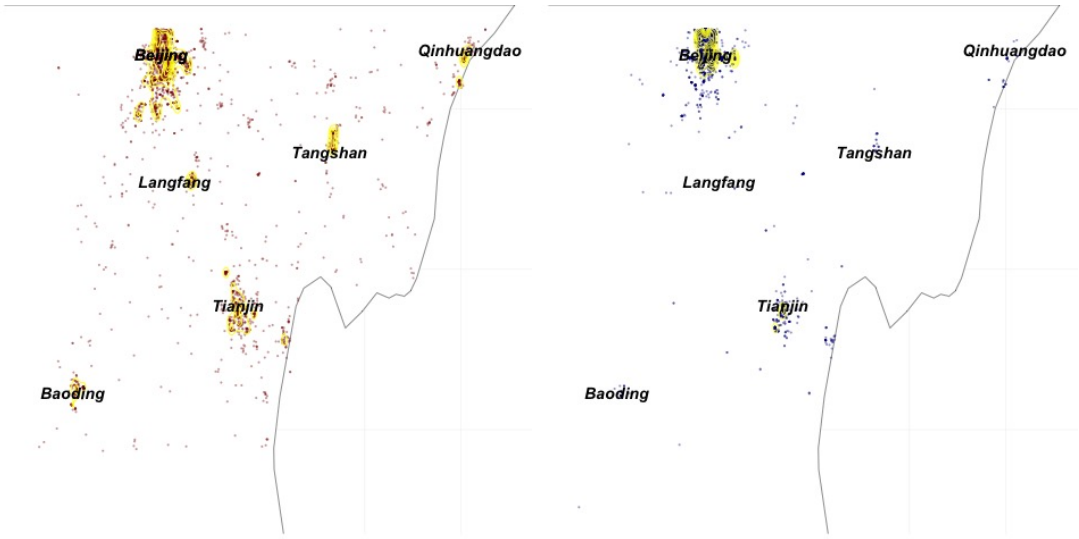

Figure 6: *Geo-located Weibo users (left) and Chinese language Twitter users (right) in Beijing and surrounding areas during September 2014.* Weibo users were more geographically dispersed than Twitter users. Most Chinese language Twitter users were concentrated in urban centers.

Who stays?

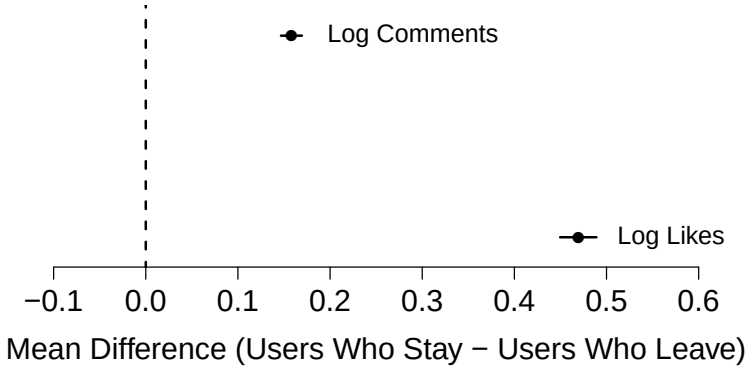

Figure 7: *T-tests of pre-block log likes and log comments of users who stay on Instagram after the block and users who left Instagram after the block.* Users who stay on Instagram tend to be more active on Instagram before the block.

Crackdown

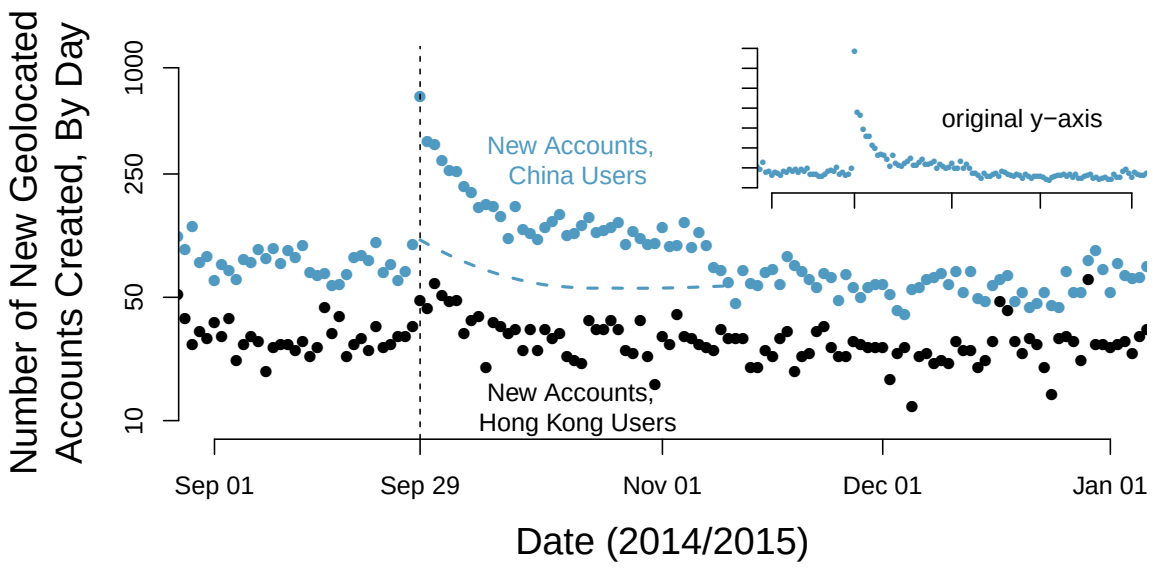

Figure 8: *The Instagram block's effect on* new *account creation Twitter users from mainland China within our sample.* In the days following the Instagram Block, new user account creation jumped over 600%. Note that this figure measures the marginal number of users joining Twitter per day, rather than cumulative number or levels of activity on the site. It is limited to geo-locating users who made up only 1% of the worldwide Twitter user population in 2014. The decline in new sign-ups roughly corresponds to reports of crackdowns on VPN access.

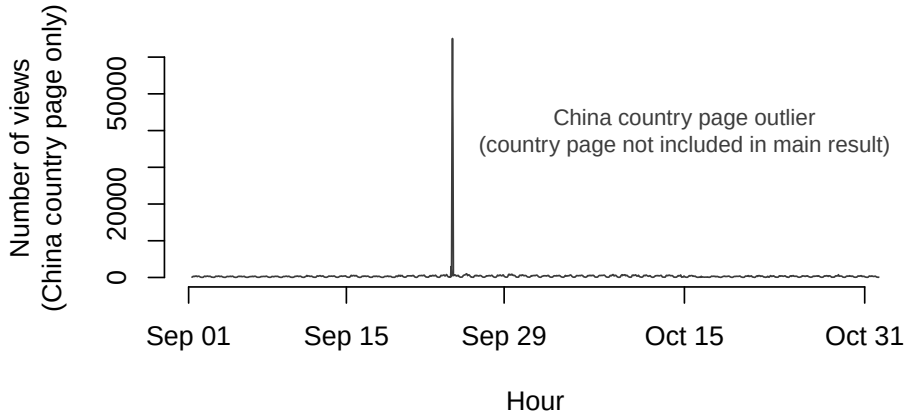

Figure 9: *Outlier in Wikipedia page view analysis.* We discovered one large outlier in our analysis of Wikipedia page views and excluded it from our analysis. This figure shows the number of views of the Chinese language Wikipedia (zh.wikipedia.org) page for "People's Republic of China". There is a massive spike in views to the page on September 24th. This spike was limited to 9am to 11am Beijing time and could be driven by bot activity.