Erschienen in: Journal of Information Technology & Politics ; 20 (2023), 1. - S. 60-73
https://dx.doi.org/10.1080/19331681.2022.2037118

Attack or Block? Repertoires of Digital Censorship in Autocracies

Lukas Kawerau, Nils B. Weidmann, and Alberto Dainotti

ABSTRACT
Online censorship has become a common feature in autocracies. Previous work has investigated different online censorship tactics such as website blocking or cyberattacks independently. In reality, however, autocratic governments rely on a *repertoire* of censorship techniques to control online communication, which they are likely to use depending on the respective political situation on the ground. In this article, we study the interplay of different online censorship techniques empirically. Focusing on new Internet measurement techniques and large existing datasets, we study the relationship between *website blocking* and *cyberattacks* (Denial-of-Service). Our results provide evidence that autocrats select tactics from their censorship repertoire depending on the current level of contention. During quiet times, we find some evidence that governments rely on different censorship tactics in parallel. In weeks with protest, however, website blocking is negatively associated with Denial-of-Service attacks against opposition websites. This shows that when the stakes are high, autocrats become more selective in their use of censorship.

KEYWORDS
Autocracy; censorship; denial of service; cyberattack; ICT

Introduction

Without a doubt, the Internet is one of the technologies that has had the most profound impact on humankind in recent decades. The ability to communicate with many others around the globe, or to post messages that can instantly be seen by millions of people has changed the lives of many. Not surprisingly, much of political discussion and debate is also happening online, which is why research is increasingly focusing on the political repercussions of online news and social media. Early research along these lines has emphasized that online communication and social media may have "liberating effects" and empower oppressed groups in illiberal societies.

More recently, however, scholars have examined the use of digital communication by governmental actors, in particular with regards to their desire to curtail freedom of expression and exert control over information. Governments across the world have increased their efforts to control information, often under the guise of preventing the spread of "fake news" or misinformation. Governments can use digital technology for several different purposes: to identify and track citizens for the purpose of surveillance, to spread governmental propaganda online, but also to censor online content that is deemed problematic by the government. This latter aspect is the focus of this paper.

Online censorship can take a variety of forms. The most drastic way to disable online communication is a complete Internet shutdown that stops all traffic in and out of a country (Dainotti et al., 2014; Gohdes, 2015). However, besides the extensive disruption this causes to all citizens, Internet shutdowns can have negative repercussions on a variety of other outcomes, for example economic activity. This is why they are employed by governments primarily in particularly severe situations, for example when massive protests were expected in Cairo in January 2011 (Hassanpour, 2014). More common, every-day attempts to control information and the freedom of expression typically employ much less pervasive and more targeted measures.

In this paper, we compare two different types of censorship: through (1) the *blocking* of particular websites (Deibert et al., 2008; Filastò & Appelbaum, 2012), and (2) attempts to shut down particular servers with *cyberattacks* (Lutscher et al., 2020). For a government, the former is easier to implement e.g. through filtering at the gateways in and out of a country. However, Internet filtering can be bypassed by citizens with simple technical tools (such as

CONTACT Lukas Kawerau lukas@kawerau.org University of Konstanz

Virtual Private Networks, VPNs), which is why governments may prefer to disable the entire server that distributes the content to be censored. This can be done with the second type of censorship through Denial-of-service attacks, a brute-force type of cyberattack that requires little technical expertise to be carried out.

Existing research has produced insights into each of these types of censorship independently. As argued by Keremoğlu and Weidmann (2020), this approach has been able to generate new and interesting results on the use and effects of particular censorship technologies. However, governments rarely ever rely on a single type of censorship alone; rather, they employ different ones in combination with each other. This *repertoire* of censorship technologies used by governments is something we know less about, and it is the main focus of this paper. We study how blocking and attacks are used in combination with each other, and how this relationship may be affected depending on the political situation on the ground. Because this has important consequences for the work of activists and dissidents in non-democratic regimes in particular, we focus our analysis on autocracies.

Our analysis uses new and refined data on the use of censorship in autocracies. We rely on Internet measurement techniques and large existing datasets to observe attacks and blockings and to compare them to each other. Our results show that autocrats seems to adjust their repertoire of censorship depending on the political situation. There is some evidence that outside of political protest, blockings and attacks seem to be positively related, such that they are both used in tandem. However, when protest occurs, they are negatively correlated, meaning that autocrats rely on one, but not the other. This suggests that autocratic governments tread more carefully during times of ongoing contention, and use a more cautious strategy in an effort to not add fuel to the flames. In the following, we briefly review the literature and present our theoretical argument, before describing the datasets and our analysis in detail.

Related literature and theoretical argument

From the early days of the Internet in the 1990s until today, its meteoric rise in importance for everyday life has also been connected to hopes for political change (Barlow, 1996). Its potential as a "liberation technology" (Diamond, 2010) has been highlighted in connection to political upheaval and social movements around the world such as the "Arab Spring" in 2011 (El-Baradei, 2011; Hassanpour, 2014). Others, however, have cautioned against this idea and have argued that the Internet has the potential to further increase repression and censorship (Lessig, 1997; Morozov, 2011). These diverging predictions have resulted in a broad body of research investigating how governmental actors use the Internet and for what purposes. Beyond providing government services online and using the Internet to gauge citizens preferences, there is extensive evidence that governments also use the Internet for less benevolent reasons. This includes identifying and tracking citizens for surveillance, using the Internet to broadcast government propaganda at home and abroad and censoring online content that is critical of the government.

The usual assumption is that governments employ these strategies to secure their power. If governments do not live up to the promises they have made to win support, they may be held accountable and their support erodes. For this reason, governments try to shape and restrict what information reaches their citizenry to prevent negative outcomes for those in power (Roberts, 2018, p. 21 f). While efforts to control information are on the rise in both democracies and autocracies (Sundara Raman et al., 2020), these incentives are still held in check in democracies by strong institutions and the notion that the right to communicate and organize freely is the very core of what constitutes a democracy (Merkel, 2004). In contrast, autocracies have weaker institutions, and the regime's entire survival often depends on keeping the threat of public mobilization in check (Friedrich & Brzezinski, 1965; Geddes, Wright, & Frantz, 2018; Wintrobe, 2000).

The tools autocrats might use to censor online content can roughly be distinguished as belonging to one of three layers of Internet technology, as Keremoğlu and Weidmann (2020) describe. The first layer is the infrastructure layer, where governments control whether the Internet is accessible at all (Keremoğlu & Weidmann, 2020, p. 3). Here,

governments have been shown to interfere in access provision to politically excluded ethnic groups (Weidmann et al., 2016) or shut down access to the Internet for the whole country during times of contention (Belarus and Iran are recent examples, but also Egypt and Libya during the Arab Spring, see Dainotti et al., 2014). Second, where infrastructure is in place and accessible, the network layer allows governments to impose restrictions at a more granular level (Keremoğlu & Weidmann, 2020, p. 3). Through tools based on keyword filters, governments can not only control access to information on a continuing basis (Hellmeier, 2016), but also carry out surveillance over their population and identify dissidents (Deibert et al., 2008; Okunoye et al., 2018). Where installing such censorship measures is too costly or citizens frequently circumvent the censorship by using tools such as VPNs, governments can use blunter tools like so-called denial-of-service (DoS) attacks to react quickly to emerging threats. DoS attacks overwhelm a website's server with network packets or data requests, in order to prevent citizens from reaching a website (Lutscher et al., 2020; Nazario, 2009). As a third and most researched layer, Keremoğlu and Weidmann (2020) identify the application layer, which is where network applications such as browsers, e-mail programs or social media clients operate. Here, autocrats choose between four tactics to use the Internet to their advantage: explicit content controls, information manipulation, surveillance and the provision of "controlled venues of preference articulation" (Keremoğlu & Weidmann, 2020, p. 5).

Each of the governmental tactics for information control is associated with its own set of costs. The most costly choice is shutting down the Internet completely within a country. This choice comes with severe repercussions such as reduced economic activity (Khrennikov, Kudrytski & Sazonov, 2020) and is therefore only employed in particularly severe situations, for example when massive protests were expected in Cairo in January 2011 (Hassanpour, 2014). Usually, censorship in autocratic countries is therefore often done through less pervasive and more targeted measures. Two of these measures are (1) censorship through blocking of particular websites (Deibert et al., 2008; Filastò & Appelbaum, 2012) and (2) the targeted shutdown of servers hosting potentially objectionable content through cyberattacks. On a technical level, the former is achieved by installing tools in the infrastructure of Internet Service Providers in a country which inspect each website a user wants to visit, and by blocking access if the website is on a government-controlled list. This is particularly effective at the gateways that connect one country to another, since these gateways create centralized points of control that all traffic has to flow through. However, this type of filtering can be circumvented relatively easily by technologically proficient users through the use of simple technical tools like Virtual Private Networks. There is ample anecdotal evidence both for the use of blocking as a censorship measure (Deibert et al., 2008; Okunoye, Xynou, Evdokimov, Alabi, & Okoli, 2018) and for the use of VPNs and other means to circumvent this (Hobbs & Roberts, 2018).

While blocking is generally sufficient to dissuade the average citizen from accessing information critical of or dangerous for the regime (Roberts, 2018), the fact that politically engaged people can circumvent this measure relatively easily leads to the need for tools that can prevent access even to those people. In this case, governments can use targeted cyberattacks to take information entirely offline, instead of merely blocking access for their citizens. Denial-of-service attacks are a brute-force tool to take information offline by overwhelming the server with data requests (or other network traffic) through for example the use of large collections of hacked computers and devices connected to the Internet called "botnets." Attacks via botnets can be purchased cheaply on the darkweb, and have the additional benefits that there is no visible indication of censorship to the end user and that the attacks themselves are not attributable to a specific actor. Despite this attribution problem, there is extensive anecdotal as well as some quantitative evidence for the use of DoS attacks as a tool for censorship. Examples include DoS attacks against independent news websites before the 2011 election in Russia (Jagannathan, 2012) and in Turkey in 2015 (Kelly, Truong, Shahbaz, Earp, & White 2017). These are not isolated events; a study by

Lutscher et al. (2020) shows increased levels of DoS activity during politically contentious periods across a large sample of autocratic countries.

If governments want to censor online, should they resort to website blocking or DoS attacks? The two tactics have different characteristics and implications. For one, they vary in the extent to which there is a visible indication of censorship. Website blocking is much more obvious to the end user, and is oftentimes even clearly indicated with a message alerting the user that they intended to visit a forbidden website. A DoS attack, if successful, temporarily disables a server, which results in an error message that would be the same in the case of a technical malfunction that is not politically motivated. Also, the attribution of the censorship intervention is much more obvious in the case of website blocking, where the end user knows that blocking happened at the level of a regional or national ISP. A DoS attack, in contrast, usually cannot be attributed to a particular actor and offers plausible deniability because "patriotic hackers" can also carry out such attacks without direct links to the government. Finally, the two types of tactics differ with respect to their probability of success. Website blocking can be implemented such that ordinary users cannot circumvent it (while the more technically skilled can). At the same time, DoS attacks sometimes fail to reach their goal of disabling a server, but if they do, nobody has access. Overall, blocking is a relatively overt type of censorship, visibly preventing access to particular sites with guaranteed success for large, but not all, groups of users. DoS attacks is a covert strategy that has a certain probability of failure, but if successful, can prevent access to everyone.

The question of whether governments choose one tactic over the other is one about their censorship *repertoire*, i.e. the combination of different ways to interfere in online communication. Similar to repertoires of conventional repression (Bagozzi, Berliner, & Welch, 2021), autocrats likely combine different online tactics depending on the current situation and on the goals they would like to achieve. For researchers, this means that studying one of these tactics alone can lead to misleading results: do we see low levels of website blocking, because (i) the government sees no need to censor, or because (ii) they have opted for other tactics, such as DoS? The focus on single tactics in empirical work is a common shortcoming of the literature on online censorship (Keremoğlu & Weidmann, 2020), and we make a first step to remedy it in this article.

In particular, there are two possible relationships between website blocking and DoS attacks. The first one is what we call "reinforcement," where governments rely on different tactics at the same time. This is an approach of limited selectivity; if governments see the need to censor, they employ different tactics in parallel. This approach can be used if governments see the need to double down when it comes to digital repression, in order to maximize the overall effect. For example, existing research has shown that the impact of physical repression depends on the information environment. If citizens have access to alternative sources of information, violence by the government can backfire and increase opposition support (Pop-Eleches & Way, 2021), while the opposite holds if access to information is restricted. Given this finding, autocratic governments will likely restrict information as much as possible with tactical reinforcement in online censorship. If this holds, we should see that

> the number of DoS attacks and the number of observed blockings in a country are *positively* correlated (H1).

Tactical reinforcement, however, comes with certain costs. The literature on repression has already examined backlash effects, where excessive use of force generates an even larger counter-mobilization in the population (Curtice & Behlendorf, 2021; Siegel, 2011). Dictators are in need to balance similar effects when deciding on which portfolio of censorship tactics to employ. Roberts (2020) discusses different ways in which online censorship can backfire. In particular, censorship can lead to information becoming more popular and sought after, which is exactly the opposite of what governments want to achieve. What is essential here is that users become aware of online censorship caused by active state intervention, as this determines whether and how they respond to it. If this is true, overly high and visible levels of censorship may indeed be counterproductive. If we assume that governments carefully balance strengths and weaknesses of these tactics, but at the same are afraid of a potential backlash effect, we should expect that they become

more selective in their use of censorship tactics, and employ the one that is considered to be the most useful. This is what we call "substitution," and it means that the use of one of them goes along with a reduced reliance on the other:

> The number of DoS attacks and the number of observed anomalies are *negatively* correlated (H2).

In addition, tactical reinforcement or substitution may not be constant over time. In fact, governments may choose to opt for an approach of sweeping censorship (reinforcement) at certain times, while favoring a more selective one (substitution) at other times. As the repression literature has found out, governments tend to adjust their repressive strategies depending on the threats they face (Keremoğlu, Hellmeier, & Weidmann, 2022). At a general level, autocratic governments encounter two types of threats: from within the ruling elite, and from the citizenry (Svolik, 2012). Censorship is typically used to address the latter. Periods of mass protest are situations where the threat from the population is highest. At the same time, it is in these situations that censorship can make a difference, by preventing unwanted information from spreading, therefore reducing the likelihood of further escalation. Thus, the periods of popular protest against the regime should be those where autocrats should carefully consider which ones of their digital tactics to deploy, and in particular, in what combination. During other types of contention, for example during civil war, censorship is unlikely to be effective, since the conflict has already escalated to the level of a full military confrontation. Therefore, we study the relationship between different censorship tactics depending on the political situation on the ground, distinguishing between periods with mass protest and those without.

Data on censorship tactics

For our analysis, we require systematically measured data of two censorship tactics: website blocking and DoS attacks. In order to keep our measures of these tactics comparable, we keep the list of censorship targets constant, and then observe whether they were affected by either type of censorship. We rely on lists maintained by Citizenlab, which publishes categorized lists of potential censorship targets for 141 countries. For our analysis, we restrict potential targets to websites categorized as belonging to one of the following categories: religion, political criticism, human rights, militants or terrorism, news media, host and blogging platforms, or intergovernmental organizations.[1] For the websites in these categories, we measure both DoS attacks and website blocking, as described in the following sections.

Website blocking

For website blocking, we rely on data gathered by the Open Observatory of Network Interference (OONI) web connectivity tests (OONI, 2020*b*). OONI collects measurements of potential Internet censorship around the world through crowdsourced network measurements. Residents of a country can download an app on their phone or computer, a so-called "OONI probe," and run network measurement tests. Because OONI measurements are only collected when a user decides to manually initiate such a measurement, measurements are not carried out at regular intervals, but depend on the number of active participants in each country.

Each time a user initiates their probe, the application samples sites from the Citizenlab list of potentially censored websites for the respective country and performs automated visits to these websites (OONI, 2020b). The list of potentially censored websites is compiled from background research by Citizenlab (Deibert et al., 2008) and community contributions following criteria defined by OONI (2020a).

The probe then records whether it received data for each website and sends these measurements as a "report" back to the OONI infrastructure. By comparing both measurements, OONI determines whether there is potential censorship occurring: if both results match, there is likely no censorship but if the user's results differ, the website being tested is likely censored. Because the OONI probe only *samples* from the list of potentially censored websites, it is difficult to investigate changes in blocking behavior of individual websites. Since there is no deterministic interval in which a censored website appears in a report, observing particular websites over time or creating a measure of changes in blocking is not possible.

On a technical level, the web connectivity test performs three checks for potential censorship (OONI, 2020b). As a first step, the test checks whether a requested website is blocked via DNS tampering, i.e. the user's Internet service provider maps the request to a website to the wrong Internet Protocol (IP) address. Once the probe has received the IP address of the website, it tries to connect to that IP address through a TCP/IP request. Finally, the probe will send an HTTP GET request to the website, to which websites usually respond with their web content. If the HTTP request fails or the HTTP status codes do not match between the probe and the test run by the OONI infrastructure, this can indicate censorship. If any of the three checks fail, the report will record an "anomaly" for the particular website, which is used in our statistical analysis below to measure blocking. Importantly, this methodology addresses concerns about potential sources of failed requests other than through censorship. Because every website is requested twice, once from the user and once from OONI, damage to the Internet infrastructure in a country during contentious periods would not result in an "anomaly." Either the datacenter hosting the website is directly impacted, in which case the request by the OONI servers also fails, or the infrastructure at the user end is impacted, in which case carrying out a probe becomes impossible and OONI does not initiate a request. Given the dynamic routing of the Internet, it is exceedingly unlikely that the user is able to reach OONI but only OONI is able to reach the server unless there is censoring interference by means of blocking – which is what we assume.

DoS attacks

For DoS attacks, we contribute new data that allows us to compare the use of this tactic to blocking in a way that was previously not possible. In contrast to website blocking, which is amenable to active testing as carried out by OONI, DoS attacks cannot be observed by such active probing. This means that previous studies had to rely on media reports (Asal et al., 2016; Jagannathan, 2012) even though media reports come with the significant risk of potential reporting bias: only successful and highly salient attacks are reported, and the targets of cyberattacks vary substantially in their salience for reporting, with attacks on human rights and other non-governmental organizations being under-reported (Hardy et al., 2014, p. 527).

To overcome reporting bias in the study of DoS attacks and allow detailed comparison to website blocking, we leverage passively measured data from the Center for Applied Internet Data Analysis (CAIDA, UC San Diego, 2019). This data permits a high-resolution perspective on denial-of-service attacks targeted at victims all around the world, irrespective of their salience and coverage in English speaking media. Through the UCSD network telescope, it is possible to capture one of the most frequently used DoS attack types, which are the so-called "randomly spoofed" attacks. Following the approach presented in Moore et al. (2006), it is possible to detect the IP addresses of attacked systems worldwide. Using this data, previous work by Lutscher et al. (2020) was able to show that DoS attacks are used for political purposes, with more DoS attacks recorded during election periods in autocracies.

We develop a new measurement to better map Denial-of-service Attacks to possible targets to address the question who the actual targets of DoS attacks are. Because CAIDA can only capture the IP address that was attacked but not the intended host/website, it is not directly possible to identify attacks against websites on the Citizenlab list of hosts: we do not know which IP address a particular host was using at a particular point in time, since the mapping of host names to addresses via the Domain Name System (DNS) changes frequently. To solve this, we combine the data collected by CAIDA with historical data on websites and IP addresses collected by CommonCrawl (https://commoncrawl.org). CommonCrawl is a US-based non-governmental organization that crawls a large portion of the public Internet at four week intervals, scraping the content and IP addresses of the websites it comes across, and makes this content publicly available for research purposes. When a given host is visited by CommonCrawl multiple times and has the same IP address in two adjacent observations, we assume that the host had the same address on every day between those two observations. We call this period

a "stable IP period" (SIP) and only use DoS attacks that fall within such a period when counting DoS attacks against the hosts in our sample.[2]

Our new measurement improves on existing efforts to measure censorship in several ways. First, our measurement provides event-level data on the use of a specific censorship tactic. Existing research often relies on aggregate measures based on expert surveys, coded on the level of country-years (Coppedge et al., 2020). Second, our measurement is not biased by the salience of targets often present in news coverage (Hardy et al., 2014, p. 527) or constrained by language barriers in reporting. Finally, the global coverage of our measurement allows comparative research that goes beyond individual case studies.

Research design

We set up regression models to test whether the presence of one type of censorship (blocking) is statistically related to the occurrence of the other (attacks). It is important to mention here that these models do not serve to test a causal relationship; we do *not* claim that one type of censorship *causes* an increase (or decrease) in the use of the other. Rather, in line with our hypotheses, our regression models allows us to estimate partial correlations between the two types of censorship, removing country and time-specific trends, and allowing us to check whether correlations change depending on the political situation on the ground. Our analysis covers the period from 2015 to 2019, since this is the maximum period for which we have data from the OONI database that our analysis relies on. In line with our focus on autocracies, we include those countries with a Polity IV score of less than six in 2015, the first year of our analysis period (Marshall, Gurr,& Jaggers, 2019). All countries are observed in weekly intervals.

In our main models, we require weekly observations of either type of censorship. For attacks, we use the occurrence (0/1) of DoS attacks during SIPs for the websites in the categories listed above. The DoS measurement is obtained by constructing SIPs for all websites in these categories. For each of these periods, we then record whether at least one attack occurred during that week. The second main indicator in our analysis captures the occurrence of website blockings. For this indicator, we use all OONI Probe reports of web connectivity tests that fall into our analysis period. The reports are generated when a user of the OONI Probe software in a given country manually initiates a test, which means there is no consistent interval at which websites are tested. From these reports, we create a binary variable that takes the value of 1 if there is at least one "anomaly" in the reports for the selected website categories during a particular week, which indicates potential censorship.

Both our main variables – the occurrence of attacks and censorship – depend on whether we are actually able to observe anything in a given country and week: We can only observe potential attacks if we have at least one SIP in a country, and we can only observe anomalies when users actually run the probe and a report is generated. This creates two issues for our analysis. First, we can only include those time periods in our analysis where we are able to potentially observe anything, i.e. those weeks where we have at least one SIP *and* at least one report. The intersection of the coverage of the two measurements reduces the size of our final dataset massively. While the complete dataset with 70 autocracies and 260 weeks has about 18,000 observations, we only have at least one SIP for about 16,000 country-weeks, and OONI reports for less than 3,500 weeks in 52 countries. This is why we are left with a very sparse dataset. Since reports are often generated in single weeks only, we also cannot treat this dataset as a time series, and therefore do not include lagged dependent variables as they would reduce the size even further. In Figure 1, we provide a map of the countries included in our study. Additional details about the sample can be found in Appendix A2.

The second issue related to our measurement methods is that even if we can observe censorship (because SIPs and reports are available), the number of potential attacks and blockings is affected by the number of "probes" (SIPs and reports) we have. For example, if we have data about many SIPs, the potential number of attacks we can see is high. We address this problem in two ways. First, as introduced above, we use binary indicators of attacks and blockings throughout, to make these outcomes more comparable in light of the different numbers of probes we have. Second, we include the number

Figure 1. Countries included in the sample (in gray).

of SIPs and the number of reports as control variables in additional analyses, in order to make sure that our results do not depend on the number of probes we have available.

The goal of our analysis is to assess whether the two types of censorship are used in combination or whether they complement each other. To estimate correlations between the occurrence of attacks and the occurrence of blockings, we estimate regression models using either as the dependent variable. As stated above, we compare periods of high political contention (protest) to those without, to see if there is a potential switch in the governmental strategy toward censorship. For this, we rely on protest data collected in the ICEWS dataset (Boschee, Lautenschlager et al., 2015). From ICEWS, we select all protest events directed at the government[3] to code whether a given country-week had any anti-regime protest. ICEWS is a suitable data source for this project, since it not only allows us to record protest at a fine-grained level (weeks) with global coverage, but unlike other protest datasets, includes also other, less intense types of activity, which allows us to conduct placebo tests with other variables from the same data source. Still, we may be concerned that the coverage of the OONI data is systematically related to the occurrence of protest, since people may be launching more probes during times of contention. We test this in two simple regressions in Appendix A3. The results show that there is no discernible effect of protest on coverage, which indicates that this is not a severe problem in our analysis.

Analysis

Case illustrations

In Figure 2, we start with a descriptive look at four case examples for the use of website blocking and DoS Attacks. The top row displays the number of reports with anomalies per week in each country, while the bottom row shows the number of DoS attacks per week. The gray vertical lines indicate weeks in which the ICEWS data records at least one protest event. For all countries, we see substantial variation of anomalies and attacks over time that suggest long-term trade-offs as well as more tactical interactions. As a general pattern, we see that some countries rely predominantly on one tactic over another. Iran, Ukraine and Venezuela, for example, seem to shift most of their censorship toward blocking, beginning in 2017 with Iran. Malaysia, in contrast, seems to use both tactics consistently. At the same time, Malaysia in particular shows patterns consistent with short-term tactical choices. Both in 2017 and 2018, we see short drops in reported anomalies in Malaysia, followed by quick rises in such reports that coincide with an increase in DoS attacks and the occurrence of protest. This pattern points toward tactical reinforcement and is repeated in early 2019: after a period without any recorded DoS attacks at the end of 2018, DoS attacks increase for a few weeks after weeks of protest, adding to existing censorship efforts through blocking. Similar short term dynamics are visible in Iran in the summer of 2019 or Ukraine in the first half of the same year. Ukraine also shows indicators for possible tactical substitution as well, however. At the end of 2017, a short

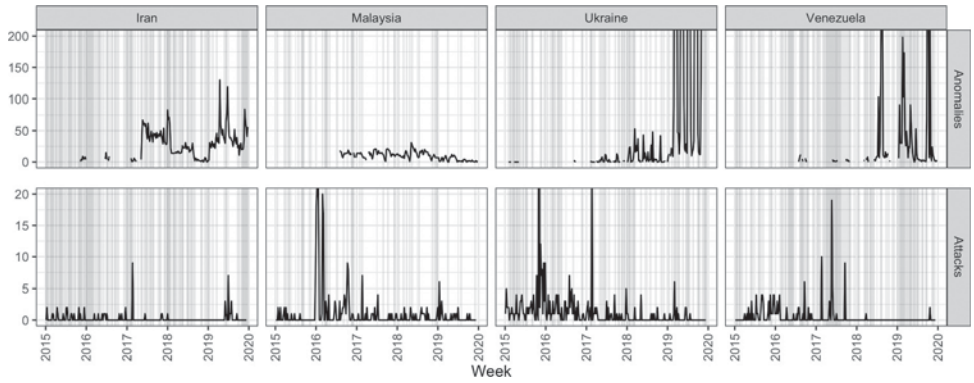


Figure 2. Blocking and DoS Attacks in Iran, Malaysia, Ukraine and Venezuela, ICEWS protest weeks in gray (y-axis truncated to improve readability).

spike in anomalies is followed by a short spike in DoS attacks, and followed again by a larger spike in reported anomalies in the first weeks of 2018 during a period of sustained protest.

These patterns provide anecdotal evidence for our hypotheses but do not yet allow us to draw any firm conclusions, in particular because Figure 2 also shows a generally high variance in the numbers of attacks and anomalies both within a country over time as well as between countries. We therefore proceed to a more systematic comparison by means of statistical analysis, which is able to separate out country-specific and time-specific levels of censorship.

Regression analysis

For our main analyses, we use linear probability models. We are interested in substantive correlations *within* countries rather than between them. In addition, the global level of DoS attacks or blockings is likely to vary between years differently across countries, we include country year fixed effects to net out country-specific trends in the overall number of DoS attacks per year. We use OLS to estimate our models, which in the case of the binary dependent variables correspond to linear probability models (LPMs).

Our main analysis is concerned with a possible tactical interaction between blocking and DoS attacks. As discussed above, we do not posit a causal relationship between the two; rather, we are interested in finding out if and when one coincides with the other (positively or negatively). For this reason, we run two kinds of analysis, for the first one using attacks and the second using blockings as the dependent variable. This approach makes sure that our main finding does not depend on the distribution of one of these variables and the distribution of the missing cases. For each set of models, we proceed in three steps. First, we include a model that only contains the respective independent variable as well as the protest indicator, in addition to the fixed effects. Second, we use the same model, but add an interaction effect between the main independent variable and the presence of protest. Third, to eliminate concerns about the measurement of the respective dependent variable, we add the number of "probes" as an additional independent variable (i.e. the number of SIPs for the attack models, and the number of reports for the blocking models). Table 1 shows the results when using attacks as the dependent variable (Models 1–3), and for blockings as the dependent variable (Models 4–6).

In Models 1 and 4 in Tables 1, we see no correlation between presence of anomalies and the occurrence of DoS attacks. However, when separating the effect between times of protest and those without, we see that there is evidence for a positive relationship outside of protest, and a negative relationship between both types of censorship during times of protest (Models 2 and 5). In Models 3 and 6, we additionally control for the number of "probes" we have (SIPs in Model 3 for attacks, and reports in Model 6 for blockings), to address the measurement problem discussed above. The substantive results remain unchanged.

In Figure 3, we present marginal effect plots to facilitate the interpretation of the relationships in our models. In Models 2 and 3, outside of protest the occurrence of blocking is positively related to the occurrence an attack by about 2.3%. During protest, the occurrence of blocking goes along with a decrease in the probability of an attack by about 7%. When switching dependent and independent variables in Model 5 and 6, the magnitude of the relationship is similar. Here, outside of protest, the occurrence of blocking increases by about 6% outside of protest, and decreases by about 10% during protest. Again, we emphasize that these are not estimates of causal *effects* – anomalies do not *cause* more or fewer DoS attacks. Rather, they reflect decisions about tactical choices that governments make when it comes to using different ways of information control.

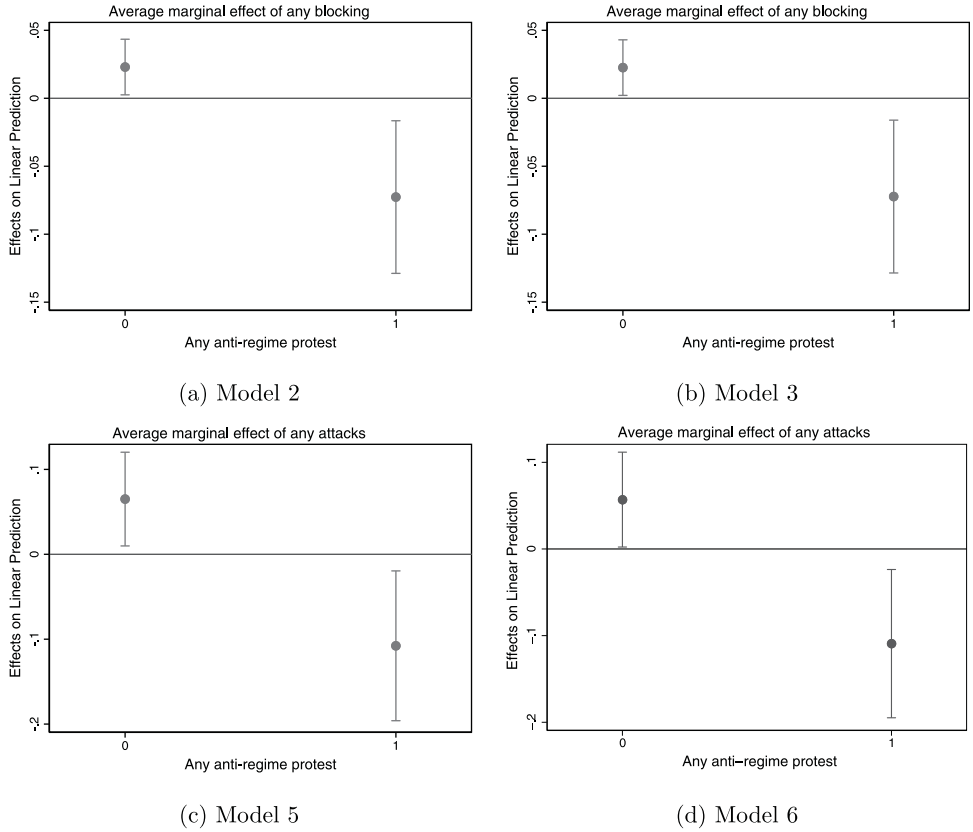
What can we learn about tactical choices from our analysis? First, tactical choices are not constant over time: in models without interaction effects, we see no consistent correlation between anomalies and attacks. Only if we include an indicator for political contention in our models as a moderator variable, we start to see statistically significant relationships between our variables of interest. Second, we do find evidence for both our hypotheses: in the absence of protest, observing any anomalies is associated with more DoS attacks, suggesting tactical reinforcement (H1). In other words, during "normal" times, autocrats do not seem to be very selective in their tactical choices. If they censor, they do so by different means, relying at the same time on blockings and attacks. During times of political contention, however, observing any anomalies is correlated with fewer DoS attacks, thus indicating tactical substitution (H2). In other words, when times politically threatening for autocratic governments, they become more selective in their censorship tactics and use one of them, but not both at the same time. Overall, this finding is counter-intuitive: one would naively assume that autocrats "double down" during periods of contention and "ease up" on censorship when there is no protest. Clearly, autocrats do *not* generally lower censorship efforts during protest periods – if they did, we would have seen this in the models with protest (Models 2 and 5). Rather, they seem to be temporarily reducing the tactical diversity in their censorship repertoire.

Robustness tests

In this section, we present a number of tests to check the robustness of our results. First, to address concerns about our use of linear models for binary dependent variables, we repeat our main models

Table 1. Relationship between the occurrence of DoS attacks and the presence of anomalies. Models 1–3 use attacks as the dependent variable, Models 4–6 use blockings as the dependent variable. Linear probability model with country year fixed effects and robust standard errors.

	Any attacks			Any blocking		
	(1)	(2)	(3)	(4)	(5)	(6)
	b/se	b/se	b/se	b/se	b/se	b/se
Any blocking = 1	0.007	0.023**	0.023**			
	(0.011)	(0.010)	(0.010)			
Any attacks = 1				0.017	0.065**	0.057**
				(0.025)	(0.028)	(0.028)
Any anti-regime protest = 1	−0.003	0.072***	0.072***	0.010	0.026*	0.027*
	(0.012)	(0.027)	(0.027)	(0.015)	(0.016)	(0.016)
Any blocking = 1 x Any anti-regime protest = 1		−0.096***	−0.095***			
		(0.029)	(0.029)			
Any attacks = 1 x Any anti-regime protest = 1					−0.173***	−0.166***
					(0.052)	(0.051)
Number of SIPs			0.000*			
			(0.000)			
Number of reports						0.000***
						(0.000)
Constant	0.073***	0.062***	0.019	0.701***	0.698***	0.678***
	(0.008)	(0.008)	(0.024)	(0.007)	(0.007)	(0.008)
N	3393	3393	3393	3393	3393	3393
Adjusted R2	0.248	0.251	0.252	0.402	0.404	0.407



(a) Model 2 (b) Model 3

(c) Model 5 (d) Model 6

Figure 3. Effect plots for the main regression models.

using conditional logit regressions. The results are presented in Appendix A4. These models show similar patterns as the LPMs above, with the exception that the positive relationship outside of protest is no longer significant. Second, one could object that our protest coding does not take into account protest severity, and therefore does not distinguish between large-scale protest and smaller ones. The ICEWS database does not include a measure of protest severity. However, we can re-run our analysis with different cutoffs for the binary protest indicator. In Appendix A5, we do this such that protest is coded as 1 if there are at least (a) 2 or (b) 4 protest events for a given week. In particular, the higher cutoff values may give reason to worry, as this reduces the overall frequency of anti-regime protest in the sample. Still, most of the additional analyses confirm our above results, with the exception of the negative relationship during protest becoming insignificant during protest when we use a cutoff value of 4 (which reduces the occurrence of protest to an extremely rare event, and therefore increases the standard error of the coefficient estimate).

To see whether ICEWS's relative broad definition of protest influences our results, we run additional models with the Mass Mobilization Protest Dataset by Clark and Regan (2021). Again, we code our protest indicator as 1 if at least one instance of protest occurs in a given country and week according to this dataset. The results are presented in Appendix A6. While the general relationship we find in the above analysis holds, we again see that the much less frequent occurrence of protest in the Mass Mobilization Protest Dataset leads to less precise estimations of the coefficients, which leads to the negative relationship during protest episodes becoming insignificant.

We also conduct a number of placebo tests to test whether our results are really driven by the actual occurrence of protest, and not by reporting bias and media attention (which is always an issue with media-based event data). The results of theses tests are reported in Appendix A7. The first of these tests replaces the anti-regime protest coding in our main analysis with an indicator of whether protest of any type (and not just anti-regime) occurs in a given country and week. This type of protest is

of course more frequent; while anti-government protest occurs in about 15% of all cases, any protest happens about twice as often (30%). As the results show, we no longer find significant coefficients for the main interaction effects. The second placebo test uses the "make public statement" event category from ICEWS rather than collective protest. This helps us find out whether it is really manifest protest action on the ground that leads a shift in the government's use of censorship, or whether public attention could be driving our results. The result shows that there is no evidence for the latter; the interaction plots show that the relationship is either insignificant or has the opposite direction compared to our main results.

In sum, most of our robustness tests confirm that autocrats adjust their tactics depending on the level of contention in a country. Still, the tests fail to paint a perfectly clear picture, and some of our main results become weaker once we use different data or estimation techniques. In particular, the tactical reinforcement outside of protest periods we have observed in our main models proves not to be robust in some of the additional models. Clearly, the limitations of our measurement contribute to this, which lead to a large reduction in the number of cases and an uneven coverage across countries.

Conclusion

With this article, we have sought to investigate how autocratic governments use their *repertoire* of censorship techniques to control online communication and how this usage may be affected by the political situation on the ground. Previous work has investigated online censorship tactics independently, leading to a lack of insight into the interplay between these tactics (Keremoğlu & Weidmann, 2020). Relying on Internet measurement techniques and large existing datasets, we provide first evidence that autocrats select tactics from their censorship repertoire depending on the current situation. In weeks with protest, website blocking is negatively associated with DoS attacks against opposition websites, which means that autocrats are likely to rely on either – but not both – tactic. In weeks without protest, there is some (weaker) evidence that it is correlated with *more* DoS attacks. This confirms our theoretical expectation that autocrats choose between tactical reinforcement and tactical substitution when deciding how to employ the tactics in their repertoire of techniques.

Our finding also raises new questions. That autocrats seem to decide *against* tactical reinforcement during times of protest is counter-intuitive. One possible explanation might be that while DoS attacks are a covert censorship technique from the perspective of potential website visitors, the owners of a website receive evidence for interference. Do autocrats take this into account and employ DoS attacks less so that evidence can not be shared with media outlets to increase attention to the situation? Is there evidence for tactical choices being influenced by media coverage? Further, autocrats might learn from the success and failures of other autocrats in deploying specific tactics. Are there patterns of tactics or combinations of tactics increasing and decreasing in popularity among autocrats over time?

Despite improving upon the existing literature, our analysis also comes with limitations that further work can address. First, our analysis still relies on historical data collected by third parties. This results in relatively sparse data, and future work should aim to introduce continuous and active measurement of politically relevant websites. In particular, our measurement of SIPs can be improved upon by recording the IP address of relevant websites more frequently going forward. Second, the data we use to detect website blocking relies on infrequent measurements by individual users in autocratic countries. Future work should find ways to detect website blocking in a more automated way that allows better continuous monitoring and thus higher-resolution analyses. Third, the data we use to detect DoS attacks is based on only one – even if popular – way to carry out DoS attacks. Further analysis would benefit from supplementing such a dataset with data on attacks performed through other techniques, such as, e.g., amplification attacks. Finally, our analysis focuses on two censorship tactics while the repertoire of techniques available to autocrats contains other tactics as well. Frequent active measurement of website *content* could also allow researching additional tactics, such as website defacements (replacing the content of a website with a message that the website has been hacked).

Notes

1. The full list of categories can be found here: https://github.com/citizenlab/test-lists/blob/master/lists/00-LEGEND-new_category_codes.csv
2. We provide a more detailed description of our method in the Appendix A1.
3. We select events where the ICEWS variable *target sectors* mentions at least one of the following: "Executive", "Executive Office", "Government", "Government Major Party (In Government)", "Ministry", "Legislative / Parliamentary", "Lower House", "Municipal", "Police", "Upper House", "Cabinet", "Elite", "Legislative / Parliamentary", "Army", "Military"

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft [402127652] and the National Science Foundation [CNS-1730661, CNS-1705024]

Notes on contributors

Lukas Kawerau (PhD, University of Konstanz) is an independent researcher. His research focuses on the intersection of governments and the Internet, with a particular eye towards issues of cybersecurity.

Nils B. Weidmann (PhD, ETH Zurich) is Professor of Political Science at the Department of Politics and Public Administration and co-speaker of the Cluster of Excellence "The Politics of Inequality" at the University of Konstanz, Germany. His research focuses on the analysis of democratic and non-democratic regimes in a comparative perspective, and the effect of modern information technology on political mobilization and violence.

Alberto Dainotti is an Associate Professor in the School of Computer Science at the College of Computing at Georgia Tech. His research is at the intersection of Internet measurement, data science and cybersecurity and largely focuses on understanding when and how Internet infrastructure can fail and proposing remedies.

ORCID

Lukas Kawerau http://orcid.org/0000-0003-4725-2355
Nils B. Weidmann http://orcid.org/0000-0002-4791-4913

References

Asal, V., Mauslein, J., Murdie, A., Young, J., Cousins, K., & Bronk, C. (2016). Repression, Education, and Politically Motivated Cyberattacks. *Journal of Global Security Studies*, *1*(3), 235–247. doi:10.1093/jogss/ogw006

Bagozzi, B. E., Berliner, D., & Welch, R. M. (2021). The Diversity of Repression: Measuring State Repressive Repertoires with Events Data. *Journal of Peace Research*, *58*(5), 1126–1136. doi:10.1177/0022343320983424

Barlow, J. P. 1996. "A Declaration of the Independence of Cyberspace.". Available at https://www.eff.org/de/cyberspace-independence .

Boschee, E., Lautenschlager, J., O'Brien, S. Sean., Shellman, S., Starz, James, J., & Ward, M. (2015). "ICEWS Coded Event Data."

CAIDA, UC San Diego. 2019. "Historical and Near-Real-Time UCSD Network Telescope Traffic Dataset." Online resource, available at https://www.caida.org/data/passive/telescope-near-real-time_dataset.xml

Clark, D., & Regan, P. (2021). "Mass Mobilization Protest Data." Online resource, Available at 10.7910/DVN/HTTWYL.

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., Bernhard, M., Fish, M. S., Glynn, A., & Hicken, A. (2020). *V-Dem [Country-Year/Country-Date] Dataset V10*. Gothenburg: Varieties of Democracy (V-Dem) Project.

Curtice, T. B., & Behlendorf, B. (2021). Street-level Repression: Protest, Policing, and Dissent in Uganda. *Journal of Conflict Resolution*, *65*(1), 166–194. doi:10.1177/0022002720939304

Dainotti, A., Squarcella, C., Aben, E, Claffy, K. C., Chiesa, M., Russo, M., & Pescape, A. (2014). Analysis of Country-Wide Internet Outages Caused by Censorship. *IEEE/ACM Transactions on Networking (TON)*, *22*(6), 1964–1977. doi:10.1109/TNET.2013.2291244

Deibert, R., Palfrey, J., Rohozinski, R., & Zittrain, J. (2008). *Access Denied: The Practice and Policy of Global Internet Filtering*. Cambridge: MIT Press.

Diamond, L. (2010). Liberation Technology. *Journal of Democracy*, *21*(3), 69–83. doi:10.1353/jod.0.0190

El-Baradei, M. 2011. "Wael Ghonim – Spokesman for a Revolution." *TIME Magazine* April 21, 2011. Available at http://content.time.com/time/specials/packages/article/0,28804,2066367_2066369_2066437,00.html.

Filastò, A., & Appelbaum, J. (2012). OONI: Open Observatory of Network Interference 2nd USENIX Workshop on Free and Open Communications on the Internet (FOCI 12) . Bellevue, WA: USENIX Association https://www.usenix.org/conference/foci12/workshop-program/presentation/Filastò.

Friedrich, C. J., & Brzezinski, Z. K. (1965). *Totalitarian Dictatorship*. Cambridge, MA: Harvard UP.

Geddes, B., Wright, J. G., & Frantz, E. (2018). *How Dictatorships Work: Power, Personalization, and Collapse*. United Kingdom: Cambridge University Press.

Gohdes, A. R. (2015). Pulling the Plug: Network Disruptions and Violence in Civil Conflict. *Journal of Peace Research*, *52* (3), 352–367. doi:10.1177/0022343314551398

Hardy, S., Crete-Nishihata, M., Kleemola, K., Senft, A., Sonne, B., Wiseman, G., Gill, P., & Deibert, R. J. 2014. Targeted Threat Index: Characterizing and Quantifying Politically-Motivated Targeted Malware. In *23rd USENIXSecurity Symposium (USENIX Security 14)*, San Diego, CA. pp. 527–541.

Hassanpour, N. (2014). Media Disruption and Revolutionary Unrest: Evidence From Mubarak's Quasi-Experiment. *Political Communication*, *31*(1), 1–24. doi:10.1080/10584609.2012.737439

Hellmeier, S. (2016). The Dictator's Digital Toolkit: Explaining Variation in Internet Filtering in Authoritarian Regimes. *Politics & Policy*, *44*(6), 1158–1191. doi:10.1111/polp.12189

Hobbs, W. R., & Roberts, M. E. (2018). How Sudden Censorship Can Increase Access to Information. *American Political Science Review*, *112*(3), 621–636. doi:10.1017/S0003055418000084

Jagannathan, M. 2012. "DDoS Attacks Disable Independent News Sites during Russian Protests." Available at https://blogs.harvard.edu/herdict/2012/06/14/ddos-attacks-disable-independent-news-sites-during-russian-protests/.

Kelly, S., Truong, M., Shahbaz, A., Earp, M., & White, J. (2017). "Freedom on the Net 2017 - Manipulating Social Media to Undermine Democracy." Madeline Earp Jessica WhiteFreedom House.

Keremoğlu, E., Hellmeier, S. & Weidmann, N. B. (2022). Thin-skinned Leaders: Regime Legitimation, Protest Issues, and Repression in Autocracies. *Political Science Research and Methods*, *10*(1), 136–152.

Keremoğlu, E., & Weidmann, N. B. (2020). How Dictators Control the Internet: A Review Essay. *Comparative Political Studies*, *53*(10–11), 1690–1703. doi:10.1177/0010414020912278

Khrennikov, I., & Kudrytski, A. S. (2020). Tech Workers Flee Belarus As IT Haven Takes Authoritarian Turn. *Bloomberg.com*. Accessed 17 09 2020. https://www.bloomberg.com/news/articles/2020-09-05/belarus-protests-tech-workers-flee-as-country-takes-authoritarian-turn.

Lessig, L. 1997. "Tyranny in the Infrastructure." *Wired*. Available at https://www.wired.com/1997/07/cyber-rights-8/ .

Lutscher, P. M., Weidmann, N. B., Roberts, M. E., Jonker, M., Alistair King, A., & Dainotti, A. (2020). At Home and Abroad: The Use of Denial-of-Service Attacks during Elections in Nondemocratic Regimes. *Journal of Conflict Resolution*, *64* (2–3), 373–401. doi:10.1177/0022002719861676

Marshall, M. G., & Ted Robert, G. K. J. 2019. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2018." Online resource, Center for Systemic Peace.

Merkel, W. (2004). Embedded and Defective Democracies. *Democratization*, *11*(5), 33–58. doi:10.1080/13510340412331304598

Moore, D., Shannon, C., Brown, D. J., Voelker, G. M., & Savage, S. (2006). Inferring Internet Denial-of-Service Activity. *ACM Transactions on Computer Systems (TOCS)*, *24*(2), 115–139. doi:10.1145/1132026.1132027

Morozov, E. (2011). *The Net Delusion: The Dark Side of Internet Freedom*. New York: PublicAffairs.

Nazario, J. (2009). Politically Motivated Denial of Service Attacks. In C. Czosseck & K., Geers (Eds.), *The Virtual Battlefield: Perspectives on Cyber Warfare* (pp. 163–181). Amsterdam: IOS Press.

Okunoye, B., Xynou, M., Evdokimov, L., Alabi, S., & Okoli, C. (2018). "Measuring Internet Censorship in Nigeria." IEEE Internet Policy Newsletter. Available at https://internetinitiative.ieee.org/newsletter/december-2018/measuring-internet-censorship-in-nigeria.

OONI, Open Observatory of Network Interference. 2020a. "The Test List Methodology.". Available at https://ooni.org/get-involved/contribute-test-lists/.

OONI, Open Observatory of Network Interference. 2020b. "Web Connectivity Test.". Available at https://ooni.org/nettest/web-connectivity/ .

Pop-Eleches, G., & Way, L. A. (2021). Censorship and the Impact of Repression on Dissent. *American Journal of Political Science*. EarlyView. Available at doi: 10.1111/ajps.12633.

Roberts, M. E. (2018). *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton: Princeton University Press.

Roberts, M. E. (2020). Resilience to online censorship. *Annual Review of Political Science*, *23*(1), 401–419. doi:10.1146/annurev-polisci-050718-032837

Siegel, D. A. (2011). When Does Repression Work? Collective Action and Social Networks. *Journal of Politics 73*(4).

Sundara Raman, R., Shenoy, P., Kohls, K., & Ensafi, R. (2020). Censored Planet: An Internet-Wide, Longitudinal Censorship Observatory. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. CCS '20 New York, NY, USA: Association for Computing Machinery pp. 49–66.

Svolik, M. W. (2012). *The Politics of Authoritarian Rule*. New York: Cambridge University Press.

Weidmann, N. B., Benitez-Baleato, S., Hunziker, P., Glatz, E., & Dimitropoulos, X. (2016). Digital Discrimination: Political Bias in Internet Service Provision Across Ethnic Groups. *Science*, *353*(6304), 1151–1155. doi:10.1126/science.aaf5062

Wintrobe, R. (2000). *The Political Economy of Dictatorship*. Cambridge: Cambridge University Press.