Contently. 2016. "Fixing native advertising: What consumers want from brands, publishers, and the FTC.".

Cook, Sarah. 2017. "Chinese government influence on the U.S. media landscape." Testimony before the U.S.-China Economic and Security Review Commission Hearing on China's Information Controls, Global Media Influence, and Cyber Warfare Strategy, May 4.

Dai, Yaoyao and Luwei Rose Luqiu. 2018. "Soft propaganda, misinformed citizens, and source credibility: A survey experiment on foreign political advertising." Unpublished manuscript, Pennsylvania State University.

Di Tella, Rafael and Ignacio Franceschelli. 2011. "Government advertising and media coverage of corruption scandals." *American Economic Journal: Applied Economics* 3(4):119–51.

Edmonds, Rick. 2017. "Native advertising grows up fast, shedding its rogue image." *Poynter*, November 27.

Edney, Kingsley. 2012. "Soft power and the Chinese propaganda system." *Journal of Contemporary China* 21(78):899–914.

Einstein, Mara. 2016. *Black Ops Advertising: Native Ads, Content Marketing and the Covert World of the Digital Sell*. New York: OR Books.

Fallows, James. 2011. "Official Chinese propaganda: Now online from the WaPo!" *The Atlantic*, February 3.

Ferrer Conill, Raul. 2016. "Camouflaging church as state: An exploratory study of journalism's native advertising." *Journalism Studies* 17(7):904–914.

Gehlbach, Scott and Konstantin Sonin. 2014. "Government control of the media." *Journal of Public Economics* 118:163–171.

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

Huang, Haifeng. 2018. "The pathology of hard propaganda." *Journal of Politics* 80(3):1034–1038.

Iversen, Magnus Hoem and Erik Knudsen. 2017. "When politicians go native: The consequences of political native advertising for citizens' trust in news." *Journalism* .

Jamieson, Kathleen Hall et al. 2000. *Everything You Think You Know About Politics — And Why You're Wrong*. New York: Basic Books.

Johnson, Thomas J. and Barbara K. Kaye. 1998. "Cruising is believing?: Comparing Internet and traditional sources on media credibility measures." *Journalism & Mass Communication Quarterly* 75(2):325–340.

Kaye, Jeff and Stephen Quinn. 2010. *Funding Journalism in the Digital Age: Business Models, Strategies, Issues and Trends*. New York: Peter Lang.

Kendrick, Alice and Jami A. Fullerton. 2004. "Advertising as public diplomacy: Attitude change among international audiences." *Journal of Advertising Research* 44(3):297–311.

Krugler, David F. 2000. *The Voice of America and the Domestic Propaganda Battles, 1945-1953*. Columbia: University of Missouri Press.

Little, Andrew T. 2017. "Propaganda and credulity." *Games and Economic Behavior* 102:224–232.

Misyuk, Iryna. 2013. "Propaganda and public diplomacy: The problem of differentiation." Unpublished manuscript, Lviv Polytechnic National University.

Mullin, Benjamin. 2017. "Vox Media pitches signature 'Explainer' format to advertisers." *Wall Street Journal*, October 13.

Pornpitakpan, Chanthika. 2004. "The persuasiveness of source credibility: A critical review of five decades' evidence." *Journal of Applied Social Psychology* 34(2):243–281.

Rawnsley, Gary D. 2015. "To know us is to love us: Public diplomacy and international broadcasting in contemporary Russia and China." *Politics* 35(3-4):273–286.

Schauster, Erin E., Patrick Ferrucci and Marlene S. Neill. 2016. "Native advertising is the new journalism: How deception affects social responsibility." *American Behavioral Scientist* 60(12):1408–1424.

Washington Free Beacon. 2014. "WaPo still publishing Russian propaganda." September 3.

Wojdynski, Bartosz W. and Nathaniel J. Evans. 2016. "Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising." *Journal of Advertising* 45(2):157–168.

Zaharna, Rhonda S. 2010. *Battles to Bridges: US Strategic Communication and Public Diplomacy after 9/11*. London: Palgrave Macmillan.

The Relationship between State and Corporate Censorship

by Anita R. Gohdes
University of Zurich

In the fall of 2018, the 20 most frequently accessed websites worldwide almost exclusively included search engines, such as Google, Baidu, and Yandex, and social media platforms, such as YouTube, Facebook, Twitter, and Weibo.[1] As more people are connecting to the world wide web across the globe, more users are spending longer hours on a smaller number of websites. In countries that have only recently experienced rising levels of digitalization, many users access the Internet through zero-rated services, such as Facebook's Free Basics, that allow them to surf a small set of websites (including a lightweight version of Facebook) for free while having to pay for further access (Global Voices

[1] An exception is Wikipedia, which is ranked fifth, according to Alexa.

Netizen Report Team, 2016). It is therefore not surprising that Facebook has become synonymous with the Internet in some parts of the world (Mirani, 2015).

The increasing amount of time spent on just a small set of websites is indicative of the increasing centralization of digital content on the Internet, and the rising power of a small number of technology companies. Governments intent on blocking access to certain content are finding themselves confronted with shifting challenges. One of these challenges includes having to negotiate terms and regulations with these tech giants. In some cases, these companies are partially or even fully controlled by state interests, allowing governments to continuously develop their censorship methods to match the changing modes of Internet usage. Scholars researching digital censorship in China have expertly demonstrated this trend (Roberts, 2018; MacKinnon, 2011; Xu, Mao and Halderman, 2011). Many countries, however, lack the capacity to build their own functioning Internet ecosystem. As a result, most of the popular content sharing platforms used by citizens in repressive states tends to be controlled by foreign companies.[2]

I. Government Censorship

Governments all across the world use a variety of tools to block access to specific websites that have been classified as threatening (Deibert et al., 2008, 2010). Figure 1 shows Internet censorship efforts by governments in 2017, where "censorship attempts include Internet filtering (blocking access to certain websites or browsers), denial-of-service attacks, and partial or total Internet shutdowns" (Coppedge et al., 2018, 181). Some countries maintain so-called 'blocklists' that are given to Internet service providers to block user requests to websites specified on the lists. In Russia, the government set up a blocklist in 2012, and has since then added thousands of banned websites to this list.[3] Cyberattacks aimed against websites' servers (for example, through denial of service attacks) can take down websites for certain periods of time while allowing state actors to deny involvement (Zuckerman et al., 2010). For example, Burmese activists have been confronted with a wide range of cyberattacks for the past decade, many of which are "consistent with government and military interest in information control and censorship" (Villeneuve and Crete-Nishihata, 2011, 235).

Censorship methods such as blocking websites or ordering cyberattacks are generally assumed to work effectively when targeted websites provide fringe content — content that is not necessarily in high demand with a large part of the population — and when the techniques chosen to block access are reasonably reliable. The centralization of content on certain websites poses a risk to both of these conditions. First, censoring popular platforms is a lot harder than censoring the websites of individual groups. If activists mostly publish their information on personal websites, the majority of a country may not be particularly concerned if access to that website is blocked. But when such content is distributed via highly popular platforms such as Twitter or Facebook, shutting down access to said websites can quickly provoke national outrage, as was the case when Twitter and Facebook were briefly blocked in Egypt in 2011 (Arthur, 2011). When activists therefore rely on platforms built for mass communication — such as mainstream social media sites — they ultimately make it harder for governments to censor them without alarming the broader population (Zuckerman, 2015). While the rising importance of platforms used by the majority of a country's population means that shutting down one or two popular services can effectively remove access for large parts of the population, many governments have experienced how such blocks can backfire and provoke protest and unrest from previously 'unpolitical' citizens.

Second, in relying on the infrastructure of large companies, activists and other opposition groups find themselves less vulnerable to conventional cyberattacks.[4] Larger social media platforms employ their own cybersecurity teams and usually pay for extensive security measures to protect their servers from such attacks. While there have been attempts in recent years that have managed to successfully attack the servers of larger on-

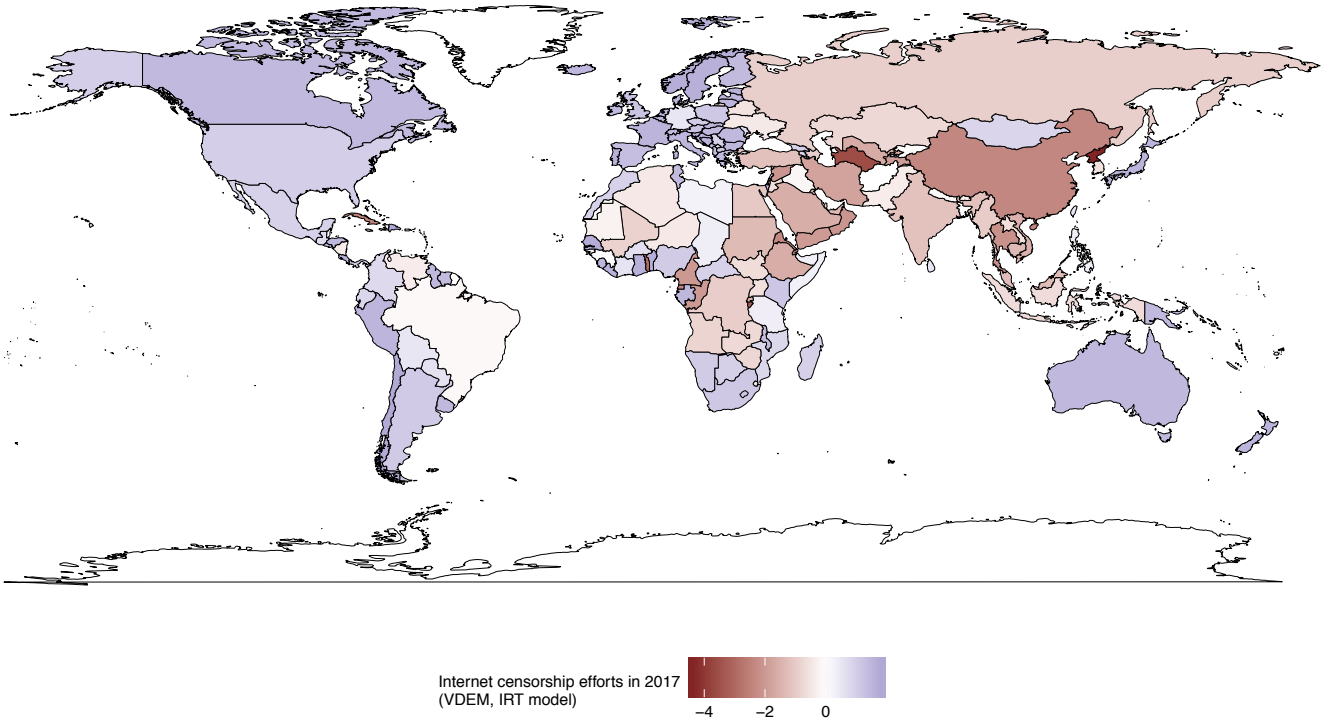
[2] For an excellent overview of corporate ethics and decision-making in the context of government-requested Internet filtering, see Zittrain and Palfrey (2008).

[3] Digital activists have been collecting a list of the banned websites at https://reestr.rublacklist.net/.

[4] Cyberattacks against the websites of activists and independent media outlets nevertheless continue to occur all over the world, even if many groups are now less vulnerable than they used to be. The continuously high number of attacks indicates that censorship may not be the only rationale for these attacks.

[5] In relying on the 'protection' of large social media platforms, activists are, however, frequently subjected to issues related to the companies' own regulatory agendas. Community guidelines, as well as platform-specific rules and regulations, frequently target minorities and

Figure 1: Internet Censorship Efforts in 2017



line platforms, they are usually mitigated within a short period of time.[5] Effectively and reliably censoring content through the use of cyberattacks has therefore become increasingly challenging for governments.

Repressive states now face the choice of shutting down access to entire websites, thereby potentially antagonizing their own support base, or granting access to it, thereby potentially allowing opposition groups to increase their public support.[6] Technology companies on the other hand have a strong incentive to increase their user base in as many countries as possible, which means that country-wide shutdowns should be avoided at all costs. Against rising pressure from governments across the world, including the pressure of democratic governments legally demanding the removal of content that violates domestic law,[7] many social media companies have put in place systems to deal with content removal requests.

II. States versus Twitter

The micro-blogging platform Twitter has published limited information about the types of removal requests they receive by governments since 2012. The company reports the number of legal demands directed at content (or active accounts) posted on Twitter in six-month intervals.[8] Starting in December 2017, the company, following its 'Country Withheld Content Policy', now informs users when a tweet is withheld in a certain country due to local laws or other legal demands (Kessel, 2017). Twitter may also withhold entire accounts in response to legal demands. The data is available for download and distinguishes between requests that are based on court orders and requests that are based on other legal demands brought forward by specific government agencies or the police.[9] Together with the received removal requests, it also publishes the percentage of requests where at least some content was subsequently

may lead to content being deleted or hidden without any government involvement (York, 2017*a*,*b*).

[6] Another strategy used by governments is to overwhelm platforms with pro-government messages (Roberts, 2018; Gunitsky, 2015)

[7] Examples of this include hate speech laws in both Germany and France.

[8] It has also recently started to report requests to remove or withhold content or accounts to Lumen, a platform created by the Berkman Klein Center at Harvard to study requests to remove content from the web.

[9] See Twitter's report on removal requests at https://transparency.twitter.com/en/removal-requests.html.

withheld by the company.

In what follows, I present a short descriptive analysis of government-requested removals and Twitter's subsequent compliance with said requests. Analyzing both the number of requests and the company's willingness to comply offers a first insight into the complex and evolving tension between the censorship ambitions of governments and the risk calculations of tech companies.

III. Removal Requests

In Figure 2, I plot the number of removal requests for the countries with the most removal requests between 2012 and 2017.[10] Following Twitter's bi-annual reporting, the requests are presented for six-month intervals. The top left panel shows the overall trend for this time period. Between 2012 and 2014 few countries made requests to remove content. The number of requests then jumps from roughly $1,000$ overall requests at the beginning of 2015 to over $6,400$ removal requests at the end of the same year. While the number of court orders slightly increases in 2016 (from 486 at the end of 2015 to 761 in the first half of 2016, and 894 in the second half), the rise in removal requests is mostly due to the almost ten-fold growth in other types of legal demands. Looking at the most frequent requesting countries, we see that this increase is mostly due to rising requests coming from Turkey and Russia. In 2017 alone, over $2,500$ removal requests were filed by agencies of the Russian government, and Twitter complied with removing at least some content in over 50% of these cases.

IV. Compliance

In Figure 3, I show the level of compliance displayed by the company for the same subset of countries. Compliance with removal requests varies substantially both across countries and within countries over time. This variation indicates that the process of complying with requests involves some assessment of both the *type* of request made and the possible *repercussions* that the company expects to face if it rejects a request. Twitter fails to comply with all legal demands brought forward by the governments of Indonesia, the United States, and South Korea, and only complies with a very small percentage of requests made by the United Kingdom. For requests made by Japan, France, India, and Germany, compliance seems to be, albeit weakly, inversely correlated with the number of requests issued: the more content that governments request be removed, the less Twitter seems to comply with the requests. In the case of Russian requests, the pattern looks quite different. Even though the number of takedown requests has dramatically increased in the past few years, Twitter has frequently complied with almost half of those requests. Overall Twitter displays some of the highest levels of compliance both in relative and absolute terms to requests issued by Russian government agencies.

> While the rising importance of platforms used by the majority of a country's population means that shutting down one or two popular services can effectively remove access for large parts of the population, many governments have experienced how such blocks can backfire and provoke protest and unrest from previously 'unpolitical' citizens.

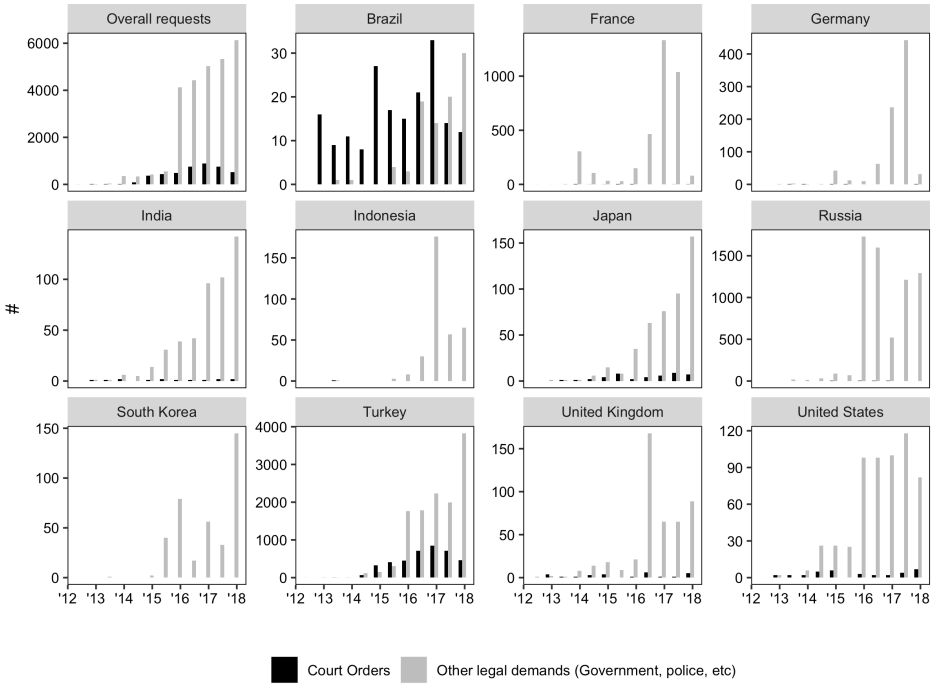
The Turkish government's relationship with social media platforms is an instructive case to study the intricate relationship between corporate and state interests. Turkey's citizens are extremely active on social media, and Twitter in particular enjoys high popularity, not just from voices critical of the government, but also from Erdogan supporters (Dogramaci and Radcliffe, 2015). Studies of the Gezi Park protests that dominated Turkish politics in 2013 have repeatedly indicated the key role that social media played in light of the high levels of traditional media censorship (Barberá et al., 2015). And as Zeynep Tufekci writes,

> Twitter is not just a protest tool in Turkey, nor just a place where the growing corruption scandal [in 2014 was] discussed. It's also used by large numbers of government supporters, including almost all of the leading officials and parliamentarians from the ruling party. It's an entrenched part of the networked public sphere (Tufekci, 2014).

In February 2014, the Turkish government introduced far-reaching amendments to an existing Internet

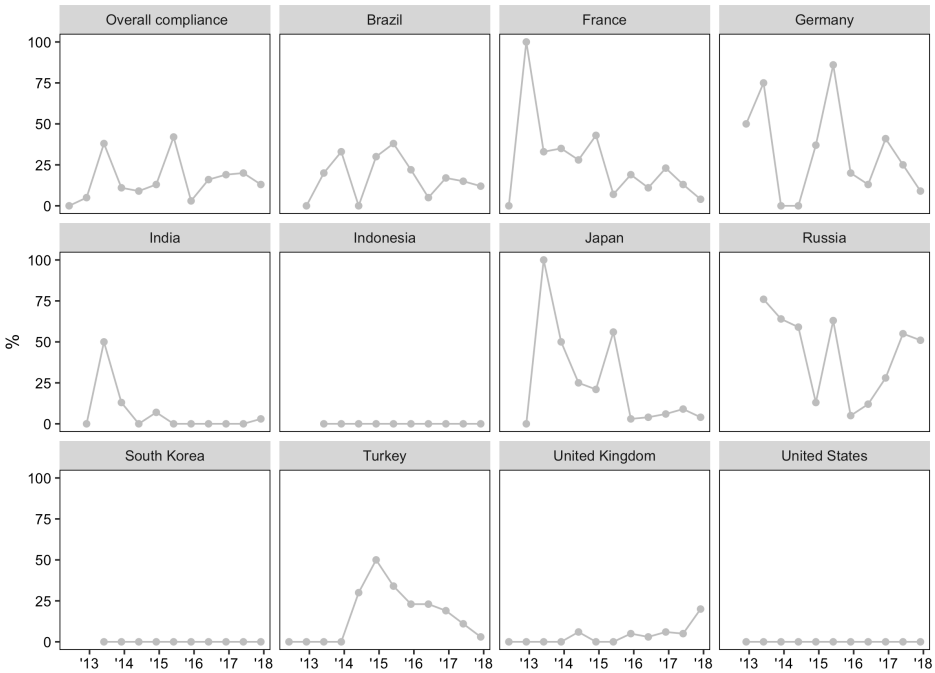
[10] The graph includes ten countries with the most legal removal requests, as well as Brazil, which has few overall requests but the second highest number of court-ordered removals.

Figure 2: Twitter Removal Requests from the Most Active Countries, 2012-2018



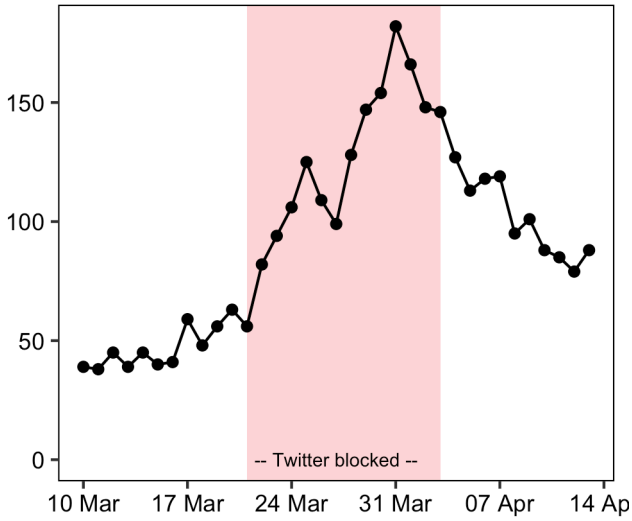
Note: Figure 2 plots the number of removal requests made to Twitter by the ten countries (and Brazil) with the most legal removal requests.

Figure 3: Compliance with Removal Requests from the Most Active Countries, 2012-2018



Note: Figure 3 plots the percentage of removal requests made to Twitter that are complied with for the ten countries (and Brazil) with the most legal removal requests.

Figure 4: Daily Tor Users in Turkey, 2014



Note: Figure 4 shows the change in the daily number of Tor users in Turkey after a court order was issued to completely block Twitter in 2014. The number of Tor users is calculated as the estimated number of clients connecting to Twitter via bridges.

law that would essentially allow the government to request the blocking of websites without obtaining a court order (Freedom House, 2014). One month later, against the backdrop of massive protests (Arsu, 2014), a court order was issued to block Twitter completely after the company failed to respond to the government's requests to remove certain accounts. Given the large number of active Twitter users in Turkey, the company was quick to condemn the shutdown of their site (Gadde, 2014).

Following the shutdown, the number of people using circumvention tools to access Twitter from within Turkey increased massively. Figure 4 shows how the estimated daily number of Tor users increased in the days following the Twitter ban.[11] The numbers reported here are quite small, but are indicative of the fact that the proportion of Turkish Internet users who accessed some type of circumvention tool increased rapidly in response to the shutdown.

At the beginning of April, the court order was overruled, and the following day Twitter was unblocked in Turkey. Following this brief episode of extreme blocking and the subsequent popular outrage, the Turkish government started making more extensive use of the platform's 'country withheld content' option. The requests start out quite sparingly in the first half of 2014 with a mere 186 requests, but by the second half of 2014 the number has already risen to 500. In 2015 the number further increases, going from 718 in the first half of the year, to 2,211 by the end of December.

Twitter's compliance pattern changed considerably throughout this time. After the company failed to remove content prior to being shut down in Turkey in March 2014, it complied with about half of all requests by the end of 2014. However, with massively increasing demands by the Turkish government starting in 2015, the company slowly moved towards noncompliance, essentially ignoring the government's requests by the end of 2017. Starting at 34% in 2015, compliance falls all the way to 3% in 2017. While these numbers merely describe the company's level of compliance, the patterns we see here are compatible with a development whereby the company decided to show goodwill and cooperate with the government's takedown requests in the aftermath of being banned, but then over time weighed the benefits of compliance with the Turkish government's preferences against its international image as a purveyor of freedom of expression. The company may also have calculated a diminishing risk of being fully shut down again.

The massive backlash in 2014 may have contributed to the fact that the Turkish government has been fairly restrained in its use of full shutdowns of Twitter since

[11] Tor software allows users to surf the Internet anonymously, thereby circumventing blocks directed at users in certain locations.

then. For governments intent on blocking access to specific content on large social media platforms, requesting specific content to be removed may be more effective than shutting it down entirely. Fewer people are likely to resort to circumvention tools when confronted with selective censorship than when confronted with indiscriminate censorship (Hobbs and Roberts, 2018). However, the large volume of requests submitted by the Turkish government to Twitter demonstrates the tension that exists between the company's corporate interests and the government's censorship ambitions.

Overall the Turkish government's ambitions to censor the Internet seem to have increased. In 2017, the government started to block access to Wikipedia following the introduction of a new law that allows censorship of content deemed to be obscene or a threat to national security (Farid, 2017). Furthermore, according to journalists, throttling — slowing access to — the Internet has become a new popular strategy to keep critical voices, such as those of journalists or activists, from using the Internet for their work (Rios Yaguez, 2017).

With the increasing ambitions of governments to censor and regulate online content, and the increasing power concentrated in the hands of a few technology companies, investigating the tensions and balancing act that characterize the relationship between corporate and state interests will become vital for our understanding of censorship and control of the Internet.

References

Arsu, Sebnem. 2014. "Across Turkey, new unrest as teenage boy is buried." *New York Times*, March 12.

Arthur, Charles. 2011. "Egypt blocks social media websites in attempted clampdown on unrest." *Guardian*, January 26.

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):1–15.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish et al. 2018. "V-Dem Codebook v8.".

Deibert, Ronald, John Palfrey, Rafal Rohozinski, Jonathan Zittrain and Janice Gross Stein. 2008. *Access Denied: The Practice and Policy of Global Internet Filtering*. Cambridge: MIT Press.

Deibert, Ronald, John Palfrey, Rafal Rohozinski, Jonathan Zittrain and Miklos Haraszti. 2010. *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge: MIT Press.

Dogramaci, Esra and Damian Radcliffe. 2015. "How Turkey uses social media." Digital News Report.

Farid, Farid. 2017. "Turkey has blocked Wikipedia and is censoring Twitter." *Motherboard*, April 29.

Freedom House. 2014. "Freedom on the Net 2014.".

Gadde, Vijaya. 2014. "Challenging the access ban in Turkey." Twitter, March 26.

Global Voices Netizen Report Team. 2016. "Netizen report: Facebook's 'Free Basics' app is under fire in India, offline in Egypt." January 8.

Gunitsky, Seva. 2015. "Corrupting the cyber-commons: Social media as a tool of autocratic stability." *Perspectives on Politics* 13:42–54.

Hobbs, William R. and Margaret E. Roberts. 2018. "How sudden censorship can increase access to information." *American Political Science Review* 112(3):621–636.

Kessel, Jeremy. 2017. "Increased transparency on country withheld content." Twitter, December 20.

MacKinnon, Rebecca. 2011. "China's networked authoritarianism." *Journal of Democracy* 22(2):32–46.

Mirani, Leo. 2015. "Millions of Facebook users have no idea they're using the internet." February 9.

Rios Yaguez, Beatriz. 2017. "How throttling internet has become a way of censorship in Turkey." April 9.

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

Tufekci, Zeynep. 2014. "The day the Turkish government banned itself from Twitter.". April 2.

Villeneuve, Nart and Masashi Crete-Nishihata. 2011. Control and resistance: Attacks on Burmese opposition media. In *Access Contested: Security, Identity, and Resistance in Asian Cyberspace*. cambridge: MIT Press pp. 233–268.

Xu, Xueyang, Z. Morley Mao and J. Alex Halderman. 2011. Internet censorship in China: Where does the filtering occur? In *Passive and Active Measurement*, ed. N. Spring and G. Riley. Springer pp. 133–142.

York, Jillian. 2017*a*. "Twitter suspensions reveal the company's skewed views of 'extremism'." *Motherboard*, April 3.

York, Jillian. 2017*b*. "Facebook celebrates pride, except where homosexuality is illegal." *Motherboard*, June 20.

Zittrain, Jonathan and John Palfrey. 2008. Reluctant gatekeepers: Corporate ethics on a filtered Internet. In *Access Denied: The Practice and Policy of Global Internet Filtering*, ed. Ronald J. Deibert, John Palfrey, Rafal Rohozinski and Jonathan Zittrain. Cambridge: MIT Press pp. 103–122.

Zuckerman, Ethan. 2015. Cute cats to the rescue? Participatory media and political expression. In *From Voice to Influence: Understanding Citizenship in a Digital Ag*, ed. Danielle Allen and Jennifer Light. Chicago: University of Chicago Press.