Using digital trace data to study online censorship and repression

Philipp M. Lutscher

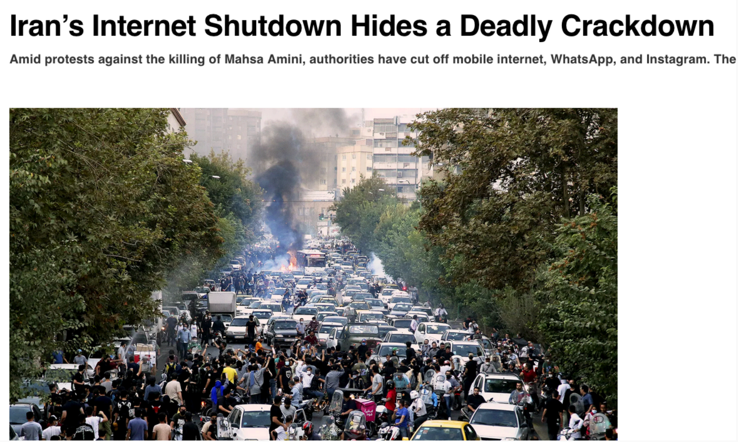
November 23, 2022

Prepared for the HON1000 lecture series

What we will talk about today:

- Repression and censorship in the digital age
- Let's zoom in: How to do social science research on this topic
    - My article on Denial-of-Service attacks in Venezuela
    - King et al. article on online censorship in China
- Q & A and brainstorming

Repression and censorship in the digital age



Iran’s Internet Shutdown Hides a Deadly Crackdown

Amid protests against the killing of Mahsa Amini, authorities have cut off mobile internet, WhatsApp, and Instagram. The

Figure 1: Source: Wired

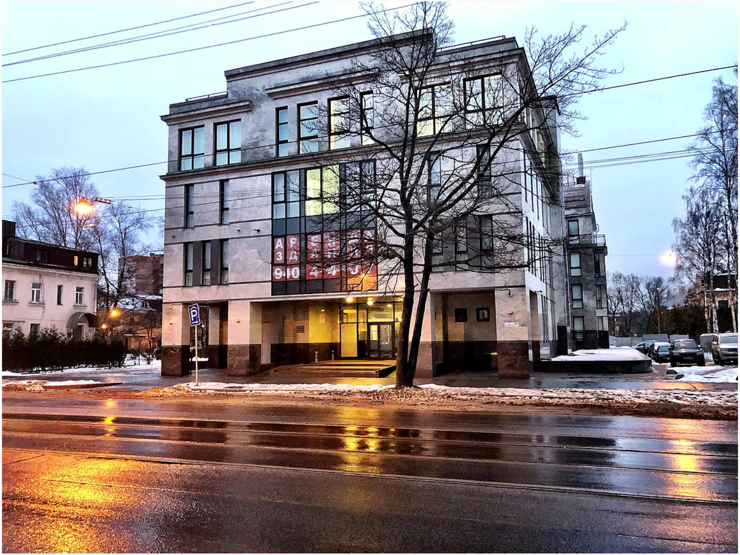


Figure 2: Source: Wikipedia

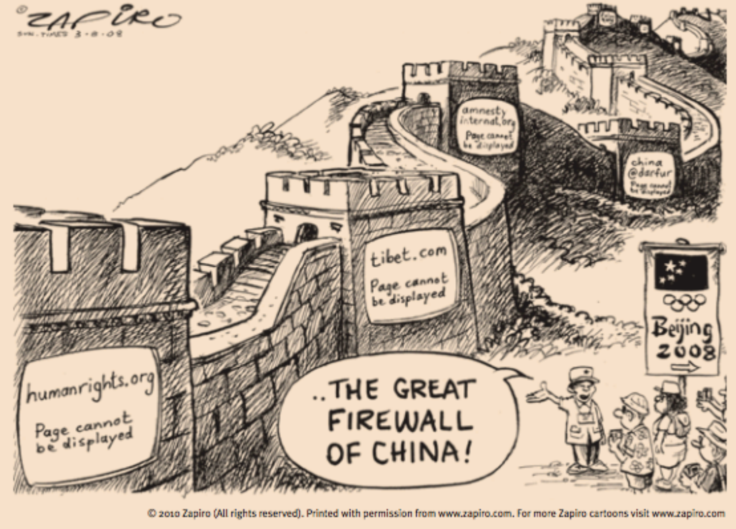

© ZAPIRO
SUN.TIMES 3-8-08
amnesty internat.org Page cannot be displayed
china @darfur Page cannot be displayed
tibet.com Page cannot be displayed
humanrights.org Page cannot be displayed
Beijing 2008
..THE GREAT FIREWALL OF CHINA!

Figure 3: Source: Zapiro

Defining state repression:

> *"actual or threatened use of physical sanctions against an individual or organization, within the territorial jurisdiction of the state, for the purpose of imposing a cost on the target as well as deterring specific activities" (Davenport 2007)*

Do you agree with this definition? If not, what parts could be challenged?

What part of digital repression draws on “traditional” tactics?

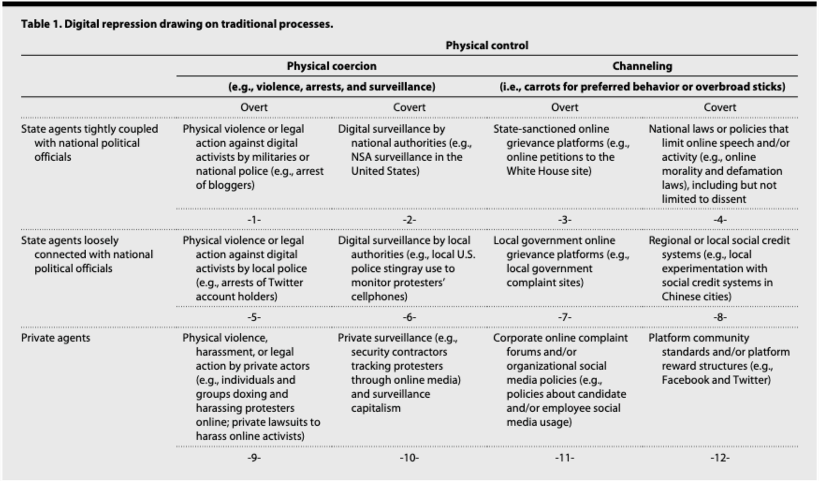


Table 1. Digital repression drawing on traditional processes.

	Physical control			
	Physical coercion		**Channeling**	
	(e.g., violence, arrests, and surveillance)		**(i.e., carrots for preferred behavior or overbroad sticks)**	
	Overt	Covert	Overt	Covert
State agents tightly coupled with national political officials	Physical violence or legal action against digital activists by militaries or national police (e.g., arrest of bloggers)	Digital surveillance by national authorities (e.g., NSA surveillance in the United States)	State-sanctioned online grievance platforms (e.g., online petitions to the White House site)	National laws or policies that limit online speech and/or activity (e.g., online morality and defamation laws), including but not limited to dissent
	-1-	-2-	-3-	-4-
State agents loosely connected with national political officials	Physical violence or legal action against digital activists by local police (e.g., arrests of Twitter account holders)	Digital surveillance by local authorities (e.g., local U.S. police stingray use to monitor protesters' cellphones)	Local government online grievance platforms (e.g., local government complaint sites)	Regional or local social credit systems (e.g., local experimentation with social credit systems in Chinese cities)
	-5-	-6-	-7-	-8-
Private agents	Physical violence, harassment, or legal action by private actors (e.g., individuals and groups doxing and harassing protesters online; private lawsuits to harass online activists)	Private surveillance (e.g., security contractors tracking protesters through online media) and surveillance capitalism	Corporate online complaint forums and/or organizational social media policies (e.g., policies about candidate and/or employee social media usage)	Platform community standards and/or platform reward structures (e.g., Facebook and Twitter)
	-9-	-10-	-11-	-12-

Figure 4: Table 1 in Earl et al. (2022)

What is new?

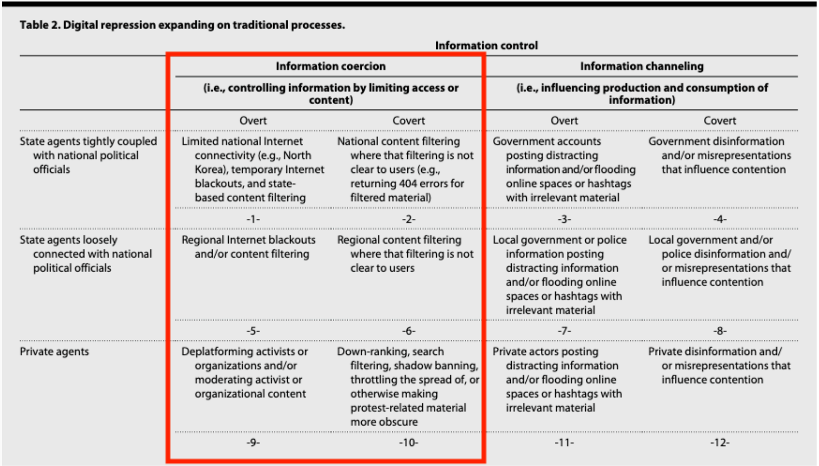


Table 2. Digital repression expanding on traditional processes.

	Information control			
	Information coercion		**Information channeling**	
	(i.e., controlling information by limiting access or content)		**(i.e., influencing production and consumption of information)**	
	Overt	Covert	Overt	Covert
State agents tightly coupled with national political officials	Limited national Internet connectivity (e.g., North Korea), temporary Internet blackouts, and state-based content filtering -1-	National content filtering where that filtering is not clear to users (e.g., returning 404 errors for filtered material) -2-	Government accounts posting distracting information and/or flooding online spaces or hashtags with irrelevant material -3-	Government disinformation and/or misrepresentations that influence contention -4-
State agents loosely connected with national political officials	Regional Internet blackouts and/or content filtering -5-	Regional content filtering where that filtering is not clear to users -6-	Local government or police information posting distracting information and/or flooding online spaces or hashtags with irrelevant material -7-	Local government and/or police disinformation and/or misrepresentations that influence contention -8-
Private agents	Deplatforming activists or organizations and/or moderating activist or organizational content -9-	Down-ranking, search filtering, shadow banning, throttling the spread of, or otherwise making protest-related material more obscure -10-	Private actors posting distracting information and/or flooding online spaces or hashtags with irrelevant material -11-	Private disinformation and/or misrepresentations that influence contention -12-

Figure 5: Table 2 in Earl et al. (2022)

Lutscher (2021): Hot Topics: Denial-of-Service Attacks on News Websites in Autocracies

Motivation

- Denial-of-Service attacks (DoS)[1] often used against news websites worldwide
- Explore the reason for why and when these attacks are used
- Focus on Venezuela in 2017/18 as there is qualitative evidence for the use of such attacks by state (-affiliated) authorities

[1] Often called DDoS (Distributed Denial-of-Service) attacks.

Why to use DoS attacks?

Table 1: DoS attacks and friction costs

	Temporary costs	Long-term costs
Consumer	(a) No access to information	(b) Unreliable information
Provider	(c) No provision of information/economic costs	(d) Self-censoring of information

When to use DoS attacks (a.k.a. my hypotheses)

Likelihood of DoS attacks increases...

- when news outlets report on protests, repressive events
- when news outlets report on negative economic development
- when news outlets are generally critical about the government

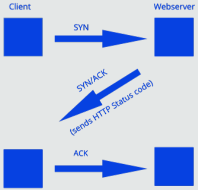
Presumption: The more widespread coverage of a topic will lead to a greater likelihood of DoS attacks since the topic is more salient and encountered by more readers.

Research design: sample and measurement

Monitoring of 19 Venezuelan independent news outlets from November 2017 - June 2018

Measurement task to proxy (potential) DoS attacks

Contact news outlet servers every 30 minutes and retrieve their status code:



Exploit the fact that servers return standardized codes, where a “503” code means unavailable to code attacks (Complication: Cloudflare protected servers). *Question*: Can you think about a different measurement strategy?

Excursion: different measurement strategy

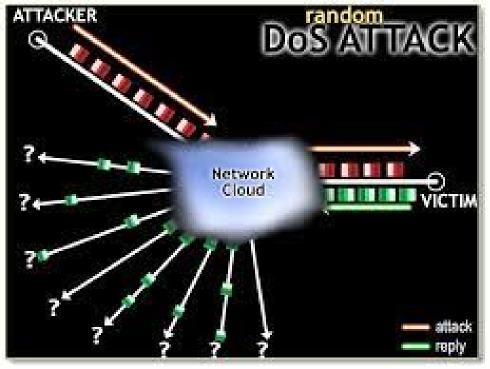


Figure 6: Passively measured data on DoS attacks from the Center of Applied Internet Data Analysis (CAIDA)

Using these data, Lutscher et al. (2020) show that DoS attacks increase around election periods in non-democratic countries

Research design: descriptives

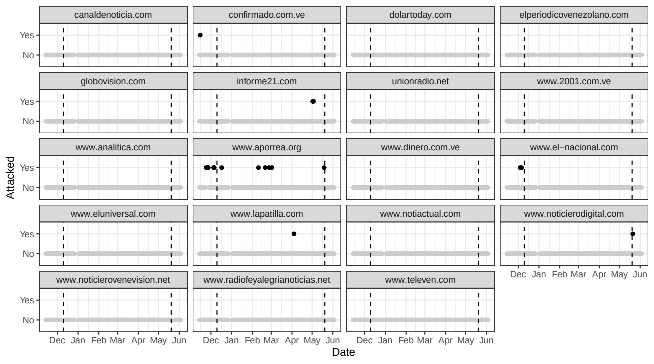


Figure 7: 19 recorded events

Research design: news topics I

- Download main page once each day
- Extract headlines and first paragraphs (this I had to manually adjust for each website after the data collection…)
- Challenge: How to systematically analyse these headlines now
- *Question:* How would you do this?

Research design: news topics II

I used an unsupervised approach back then:

- Use of so-called BTM topic models (modeling word-word co-occurrences patterns in the whole corpus)
- Much better for short text as compared to standard LDA models that look for word co-occurrences within the document (in this case headline)
- Requirement to decide on topics in advance (K = 50)
- Top 5 topics: General opinion, sanctions, national assembly, Maduro, opposition candidate

Research design: news topics II

I used an unsupervised approach back then:

- Use of so-called BTM topic models (modeling word-word co-occurrences patterns in the whole corpus)
- Much better for short text as compared to standard LDA models that look for word co-occurrences within the document (in this case headline)
- Requirement to decide on topics in advance (K = 50)
- Top 5 topics: General opinion, sanctions, national assembly, Maduro, opposition candidate

What would I do differently today? Probably rather use some word embedding models.

Research design: news topics III

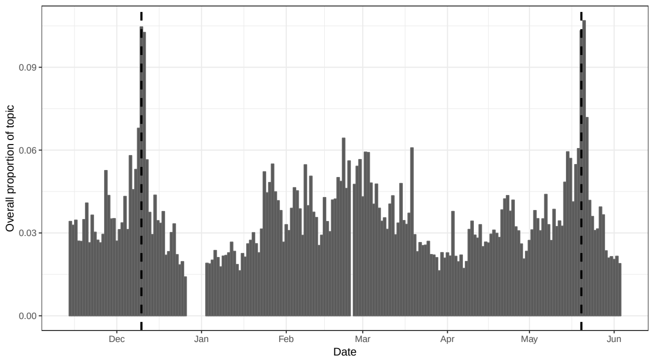


Figure 8: Example topic to test for face validity: Elections

Analysis

Penalized logistic regression on the newspaper/day level aggregating generated topics to their mean value per newspaper/day

$$Logit(DoS_{i,t}) = \beta_0 + \beta_1 topic_{i,t} + \beta_2 DoS_{i,t-1} + \gamma_i + \delta_t + \epsilon_{i,t} \qquad (1)$$

In the end, I look for deviation within websites, taking into consideration also how other websites reporting on attack days. Since I run the model for each topic separately, I adjust for multiple comparison.

Results I

Table 2: Categorization of topics.

Category	P(z)	Share of pos. related topics ($\alpha = \frac{0.05}{281}$)
Social/economic crisis	0.311	16.7%
Domestic politics	0.213	20%
Political legitimacy	0.212	57.1%
Protest/repression	0.097	75.0%
International politics	0.088	20%
Nonpolitical topics	0.079	0.0%

Results II

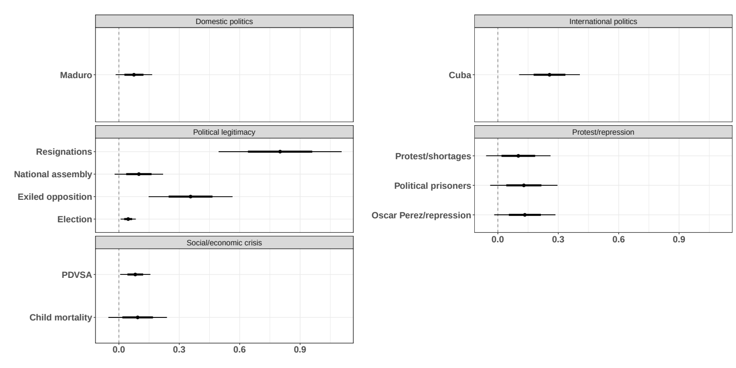


Figure 9: Results (changing topic distribution from min–max)

Limitations

- Potential measurement errors for the DoS measurement and topic modeling approach
- Attribution problem: Unclear who actually conducted the potential attacks
- In the end, my study can show which broader topics are associated with a higher likelihood of attacks but not what exact piece of news was responsible for the attack

King et al. (2014): Reverse-engineering censorship in China: Randomized experimentation and participant observation

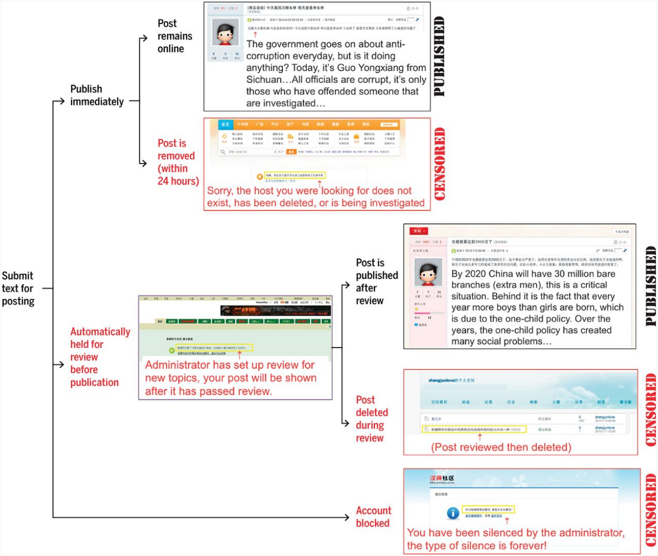
Objective

- Research question: What social media posts are censored in China?
- Main hypothesis: Posts that refer to offline collective action events are more likely to be censored
- Overall, there is a lot going on in this article and we focus on the experimental approach

Objective

- Research question: What social media posts are censored in China?
- Main hypothesis: Posts that refer to offline collective action events are more likely to be censored
- Overall, there is a lot going on in this article and we focus on the experimental approach
- Small remark: The authors have/had clearly more resources as I had for my study

How does social media censorship in China work?


Submit text for posting
Publish immediately
Post remains online
PUBLISHED
Post is removed (within 24 hours)
CENSORED
Automatically held for review before publication
Post is published after review
PUBLISHED
Post deleted during review
CENSORED
Account blocked
CENSORED

The government goes on about anti-corruption everyday, but is it doing anything? Today, it's Guo Yongxiang from Sichuan…All officials are corrupt, it's only those who have offended someone that are investigated…

Sorry, the host you were looking for does not exist, has been deleted, or is being investigated

Administrator has set up review for new topics, your post will be shown after it has passed review.

By 2020 China will have 30 million bare branches (extra men), this is a critical situation. Behind it is the fact that every year more boys than girls are born, which is due to the one-child policy. Over the years, the one-child policy has created many social problems…

(Post reviewed then deleted)

You have been silenced by the administrator, the type of silence is forever!

Figure 10: Social media censorship in China (Fig. 1 in King et al.)

How did they find this out?

- Automatically collected media posts when they are published and looked them up later (King et al. 2013)
- Creation of own social media bulletin platform in China
- Gained access in how censorship works and how to the website acts in accordance with government requirements
- Most platforms conduct automatic review do so via a version of keyword matching, probably using hand-curated sets of keywords

Quote

> *“We found employees of the software application company to be forthcoming when we asked for recommendations as to which technologies have been most useful to their other clients in following government information management guidelines.”*

Research design: Experimental approach

- Created two accounts in over 100 social media platforms in China
- Three rounds of experiment with real-world treatment conditions referring to recent events (1) with collective action potential or (2) without
- Plus the post could be either (a) in favor; or (b) critical of the government
- Accounts submitted each two posts referring to different events
- Overall 1,200 hand-written posts were posted
- Checked whether they were put under review and/or censored

Research design: Results I

40% of all posts go into (automatic) review first!

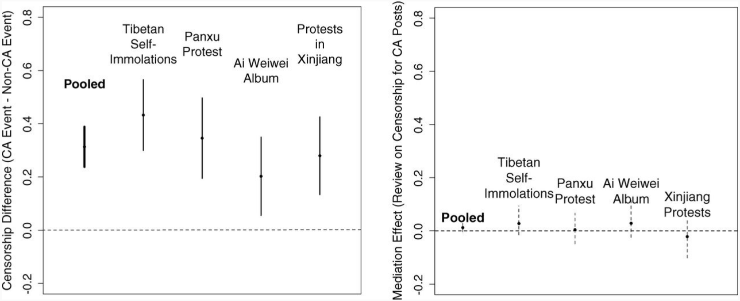


Figure 11: Causal effects of what is being censored: CA vs. Non-CA (Fig. 2 in King et al.)

Research design: Results II

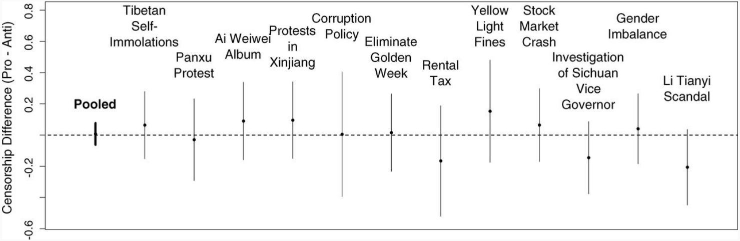


Figure 12: Causal effects of what is being censored: Pro vs. anti-gov (Fig. 3 in King et al.)

Research design: Results IV

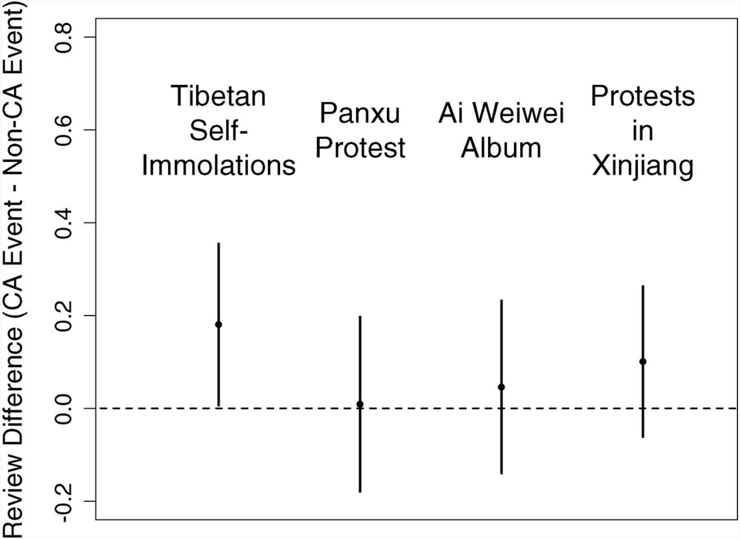
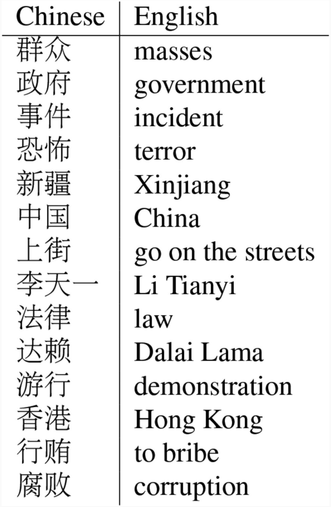


Figure 13: Causal effects of what is being reviewed (Fig. 5 in King et al.)

Research design: Results V



Chinese	English
群众	masses
政府	government
事件	incident
恐怖	terror
新疆	Xinjiang
中国	China
上街	go on the streets
李天一	Li Tianyi
法律	law
达赖	Dalai Lama
游行	demonstration
香港	Hong Kong
行贿	to bribe
腐败	corruption

Figure 14: Reversed engineered keywords that increase the likelihood to be reviewed—term frequency inverse document frequency comparing reviewed vs. non reviewed posts (Tab. 2 in King et al.)

Ethical considerations & potential shortcomings

- Since they had to rely on locals, could this potentially have endangered someone?
- Potential to influence social media discourses?
- Did everyone had the opportunity to consent?
- Can you think about limitations of this study?

Ethical considerations & potential shortcomings

- Since they had to rely on locals, could this potentially have endangered someone?
- Potential to influence social media discourses?
- Did everyone had the opportunity to consent?
- Can you think about limitations of this study? → main limitation: temporal validity
- Gueorguiev and Malesky (2019) point out that the study was run when the regime allowed citizens to criticize policy proposal on social media

Q & A

Please discuss in groups of five:

- What would be a research question in this field of study you would be interested in?
- How would you approach this question (empirically)?

Here, I some interesting data sources on online censorship, but feel free to come up with own ideas:

- *https://reestr.rublacklist.net*: leaked list of the Russian online censorship agency with more than 1 million observations from 2012 onwards
- *https://ioda.inetintel.cc.gatech.edu/*: data on network outages and DoS attacks
- Social media data to measure explore information campaigns etc.

References

C. Davenport. State repression and political order. *Annual Review Political Science*, 10:1–23, 2007.

J. Earl, T. V. Maher, and J. Pan. The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances*, 8(10): eabl8198, 2022.

D. D. Gueorguiev and E. J. Malesky. Consultation and selective censorship in china. *The Journal of Politics*, 81(4):1539–1545, 2019.

G. King, J. Pan, and M. E. Roberts. How censorship in china allows government criticism but silences collective expression. *American political science Review*, 107(2):326–343, 2013.

G. King, J. Pan, and M. E. Roberts. Reverse-engineering censorship in china: Randomized experimentation and participant observation. *Science*, 345(6199):1251722, 2014.

P. M. Lutscher. Hot topics: Denial-of-service attacks on news websites in autocracies. *Political Science Research and Methods*, pages 1–16, 2021.

P. M. Lutscher, N. B. Weidmann, M. E. Roberts, M. Jonker, A. King, and A. Dainotti. 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, 2020.