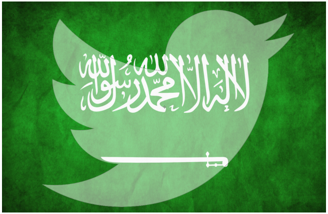
How Saudi Crackdowns Fail to Silence Online Dissent

Alexandra Siegel

Stanford University Immigration Policy Lab
NYU SMaPP Lab



Pan, Jennifer, and Siegel, Alexandra. 2020. "How Saudi Crackdowns Fail to Silence Online Dissent." The American Political Science Review.

Loujain al-Hathloul

#Free_Loujain

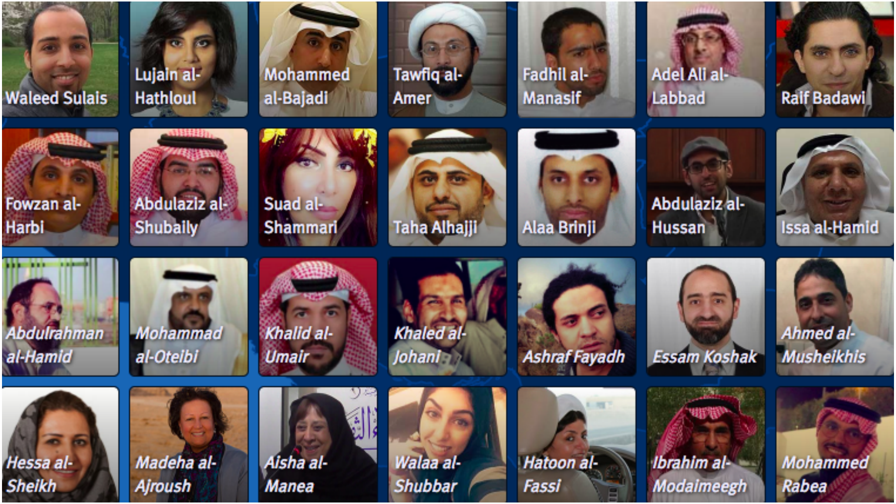
#الحرية_للجين

Mohamed al-Arefe

#Free_alArefe

#الحرية_للعريفي

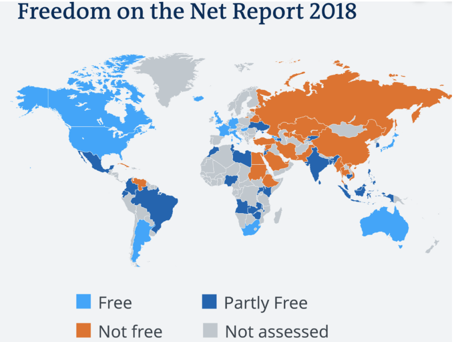
Well-Known Saudis Imprisoned for Online Dissent



Global Rise

Global Rise in Offline Repression of Online Speech

- $\sim$ 4 billion Internet users worldwide.
- 71% live in countries where users have been imprisoned for online activity.
- 48% live in countries where users have tortured or killed for online activity.

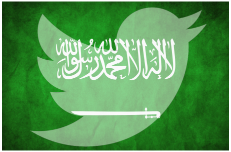


Research Question

- What are the consequences of offline repression for online dissent?
- What does this tell us about the relationship between repression and dissent more broadly?

Overview

- Goal: Evaluate the effects of offline repression on online dissent.
  - Imprisonment of well-known Saudis (2010-2017)
  - Dissent in Saudi Twittersphere (300M tweets)

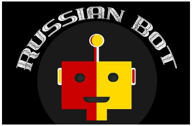


- Do we observe deterrence or backlash?
  - Arrested, Followers, Non-Arrested, Saudi Public
- Main Finding: While arrests deterred those who were directly targeted, **dissent did not decrease overall**.
- Real-time networked data offers new tests in a “black box” context.

Theoretical Motivation

- Offline repression and offline dissent
  (Feierabend et al., 1972; Hibbs, 1973; Gurr and Duvall, 1973; Rasler, 1996; Goldstone and Tilly, 2001; Davenport, 2007; Sullivan and Davenport, 2017; Davenport, Armstrong and Zeitzoff, 2019)
  - Mixed Findings: Deterrence, Backlash, U-Shaped, No Effect

- Online responses to online dissent
  (King, Pan and Roberts, 2013, 2014; Simon, 2014; Stukal et al., 2017; Hobbs and Roberts, 2018*a*; Roberts, 2018).



- But what about the effect of offline repression on online dissent?

Theoretical Expectations

- How might repression deter or mobilize online dissent?
  - Direct deterrence (Oberschall, 1973; Jenkins and Perrow, 1977; Tilly, 1978)
  - Indirect deterrence (Walter, 1969; Durkheim, 1984)
  - Backlash (Sullivan and Davenport, 2017; Young, 2017; Hassanpour, 2014; Hobbs and Roberts, 2018*b*; Jansen and Martin, 2015; Nabi, 2014; Roberts, 2018)
- Effects might vary by:
  - Actor
  - Behavior
  - Time

The Saudi Twittersphere

- Saudi Arabia is an absolute monarchy and theocracy.

  "Saudis live under repression, in fact we breathe repression with the air; it haunts us in our dreams. It is our hell before we encounter hell. Even our appearance, streets, and houses are designed by repression. Repression has shaped the media, religion, security services, universities, and institutions."

- Social media has provided an alternative space for political expression and civil society organizing.

The Saudi Twittersphere

- High Twitter penetration.
  - 70% of the population is under 30.
- Popular venue for political discussion and dissent.
- BUT Saudi Arabia carries out severe physical repression of Internet users.
- Anti-Cybercrime and counterterrorism laws.

- Posting content that “unsettles the social and national fabric...or any actions that touch the unity and stability of the Kingdom under any reason and in any form.”

Twitter Data

- Over 300 million tweets produced between 2010 and 2017.
  - Arrested Opinion Leaders' Tweets
  - Engaged Followers' Retweets, Replies, and Mentions
  - Engaged Followers' Tweets
  - Non-Arrested Opinion Leaders' Tweets

Arrested Opinion Leader Tweets

- Saudis whose arrests for online dissent were reported in press.
- 500K tweets from 36 arrested opinion leaders.
- Enables us to test direct effects of repression on online dissent.
- Changes in volume and content pre-arrest and post-release.

Engaged Follower Tweets

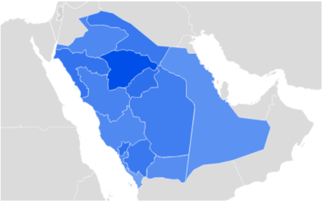
- All retweets, replies, or mentions of arrested opinion leaders
  - 32 million tweets by 8 million users
  - Indirect effects on engagement
- All tweets from sample of engaged followers
  - Latest 3200 tweets from 30K engaged followers (48 million tweets)
  - Indirect effects on dissenting content

Non-Arrested Opinion Leader “Match” Tweets

- Tweets by any Saudi Twitter user with $>$10,000 followers (230 million tweets).
- Compare **cosine similarity** of their tweets to tweets by arrested opinion leaders.
- Find top “matches” who weren’t also arrested.
- Indirect effects of repression on volume and content.

Google Search Data

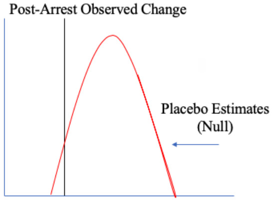
- Daily and Weekly Saudi Google Search Data
- Measure of private interest in arrested opinion leaders.
- Indirect effects of repression on the Saudi public.



Empirical Strategy: Volume Analysis

- Do we observe deterrence or backlash in tweet volume, engagement, or Google searches?

- Non-parametric test: Compare observed change to null distribution generated by placebo date estimates.



- ITSA: Measure immediate and longer term linear changes.

- Event Count Models: Robustness test.

Empirical Strategy: Crowd-Sourced Content Analysis

- Do we observe deterrence or backlash in the content of tweets?
- 3 native Arabic speakers on Figure8 coded each tweet according to whether it:
  - Criticized or supported the regime
  - Criticized or supported government policies
  - Criticized or supported Saudi society
  - Called for collective action
- Coded large stratified random samples of tweets from one-month pre-arrest, one month post-release, and one year post-release.

Intercoder Agreement Tweet Distribution

Tweet Examples

Critical (Regime/Society):

A society that denies and condemns and calls for the killing of all who do not agree with its ideology, religion and beliefs, is a society that is intellectually sick and the people suffer.

المجتمع الذي يُكفر ويُزندق ويدعوا لقتل كل من لا يوافق فكره ودينه
ومعتقداته، هو مجتمع مريض فكريا وتعاني الشعوب

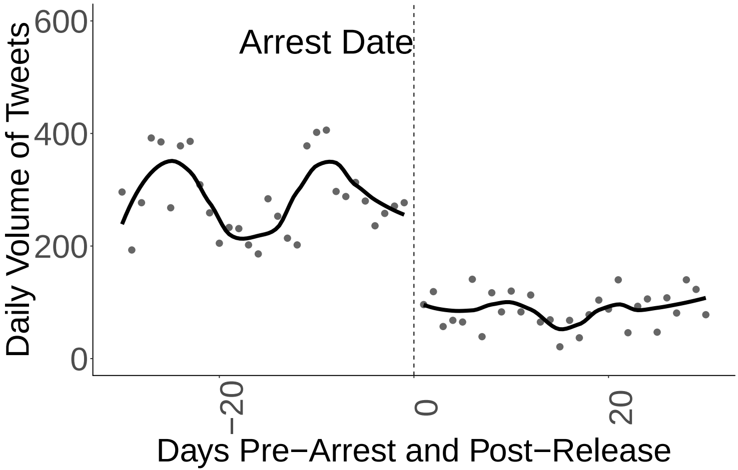
Supportive (Regime):

I followed the al-Arabiya interview with Prince Mohammed bin Salman. The man truly impressed me: fluent in his speech, transparent and practical, knows what he is talking about, his vision is clear, the future is his focus.

تابعت مقابلة العربية مع الأمير محمد بن سلمان. حقيقة أبهرني الرجل:
متمكن في حديثه، شفاف وعملي،يعرف ما يتحدث عنه،صاحب رؤية
واضحة المستقبل هاجسه

Arrested Opinion Leaders Tweet Less

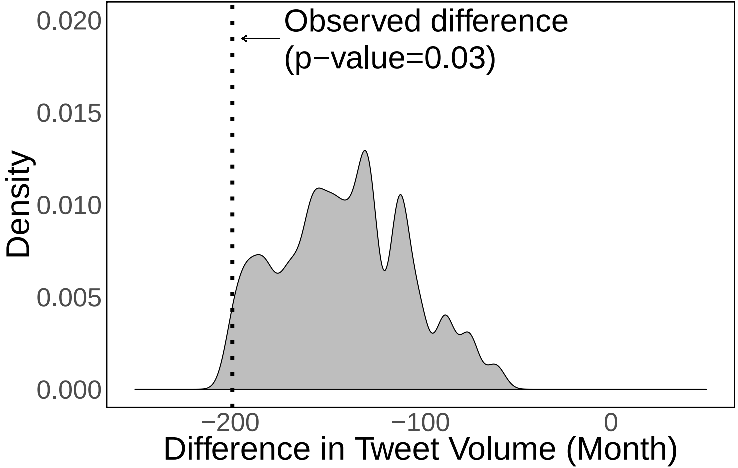
Figure 1: Daily Tweet Volume Month Pre-Arrest vs. Post-Release



Disaggregated ITSA

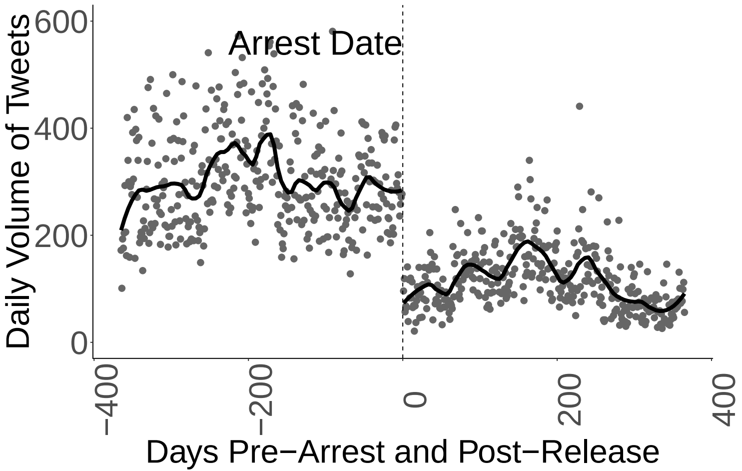
Arrested Opinion Leaders Tweet Less

Figure 2: Change in Tweet Volume Month Pre-Arrest vs. Post-Release



Arrested Opinion Leaders Tweet Less

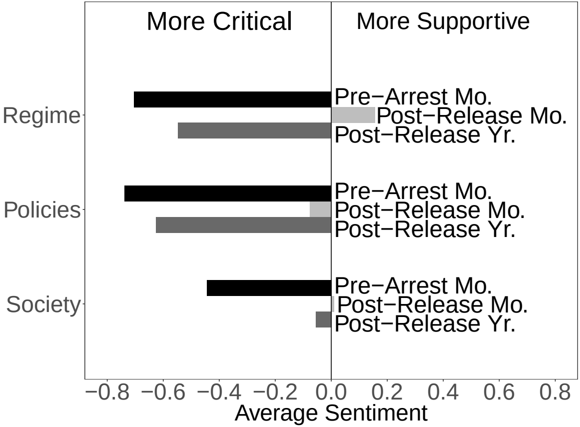
Figure 3: Daily Tweet Volume Year Pre-Arrest vs. Post-Release



Disaggregated ITSA

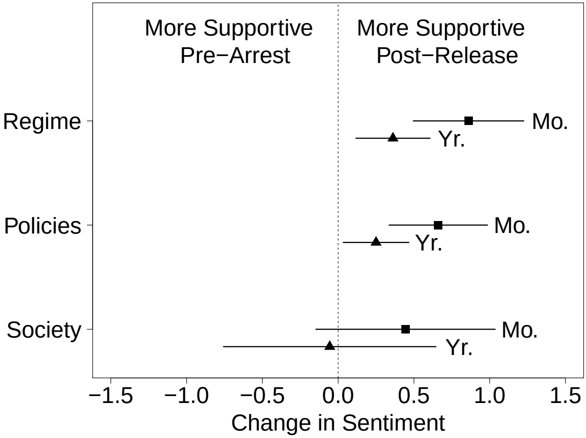
Arrested Opinion Leaders Dissent Less

Figure 4: Change in Arrested Elite Tweet Sentiment



Arrested Opinion Leaders Dissent Less

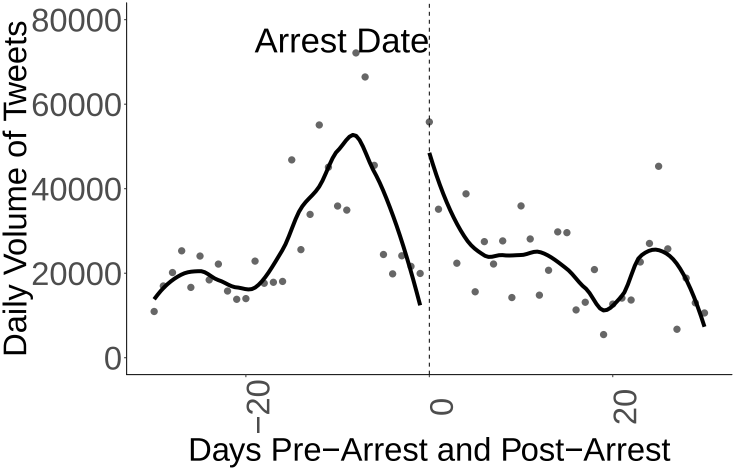
Figure 5: Change in Arrested Opinion Leaders Tweet Sentiment



Collective Action Tweet Example

Engaged Followers Keep Engaging

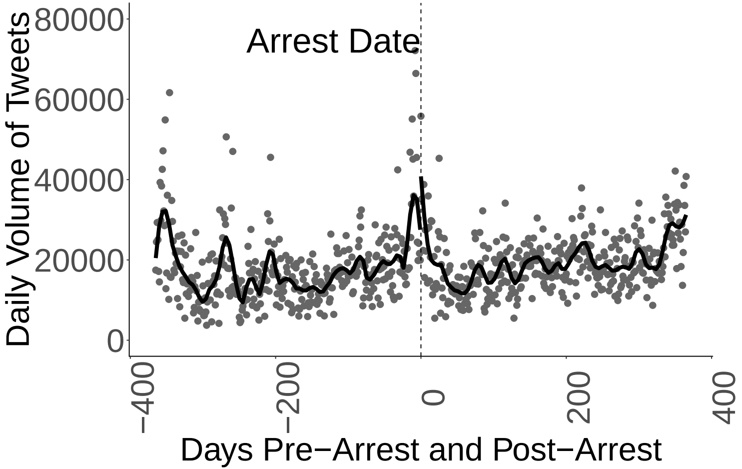
Figure 6: Daily Tweet Volume Month Pre-Arrest vs. Post-Arrest



Placebo Disaggregated Retweet Ratio

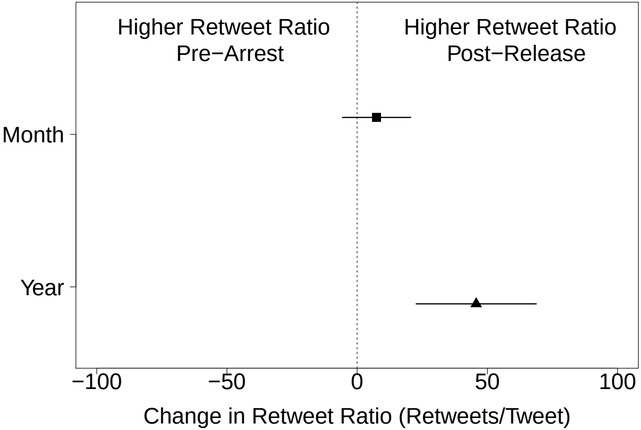
Engaged Followers Keep Engaging

Figure 7: Daily Tweet Volume Year Pre-Arrest vs. Post-Arrest



Engaged Followers Retweet More

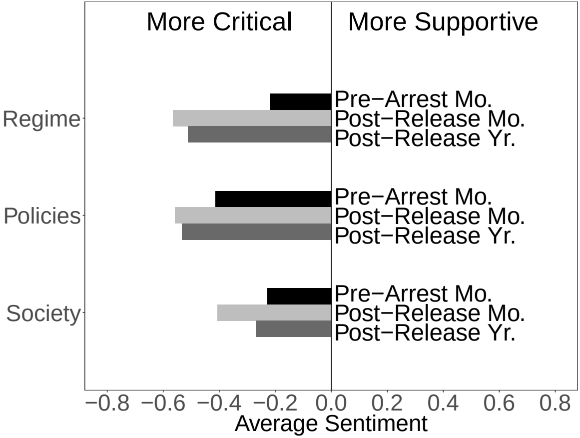
Figure 8: Change in Retweet Ratios of Arrested Opinion Leader Tweets



Engagement

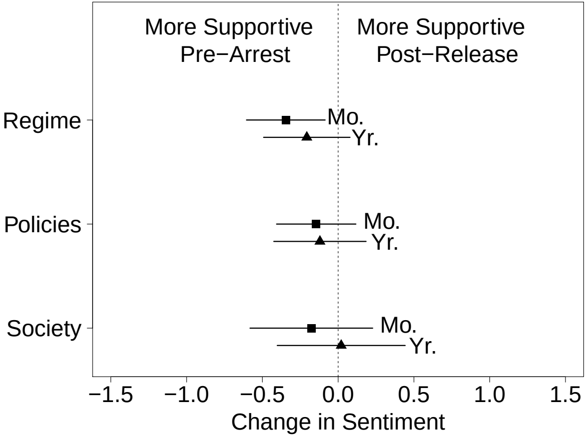
Engaged Followers Dissent More

Figure 9: Change in Engaged Followers Tweet Sentiment



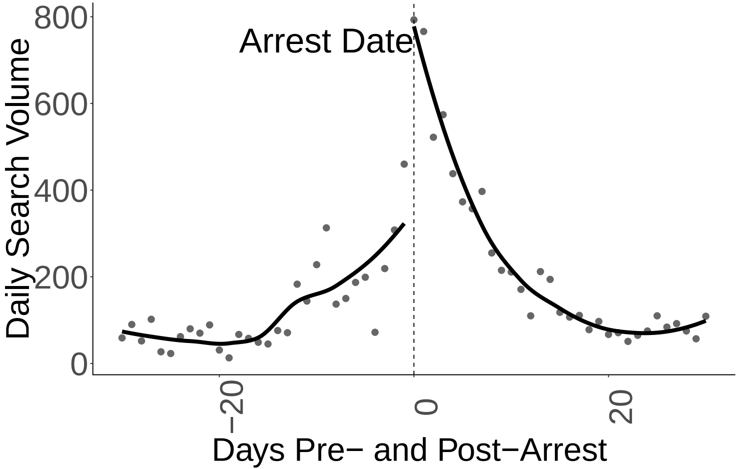
Engaged Followers Dissent More

Figure 10: Change in Engaged Followers Tweet Sentiment



Public Keeps Searching

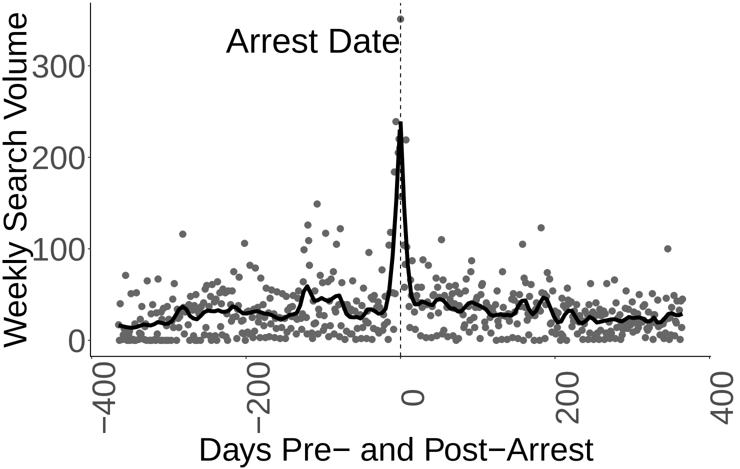
Figure 11: Daily Search Volume Month Pre-Arrest vs. Post-Arrest



Disaggregated Placebo

Public Keeps Searching

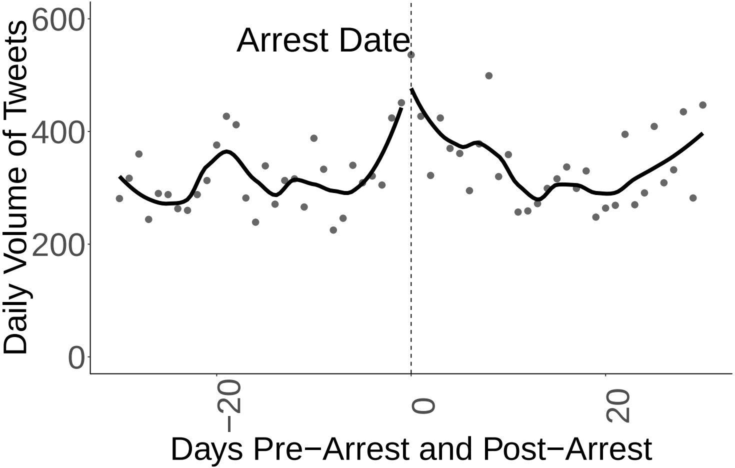
Figure 12: Weekly Search Volume Year Pre-Arrest vs. Post-Arrest



Disaggregated Placebo

Non-Arrested Keep Tweeting

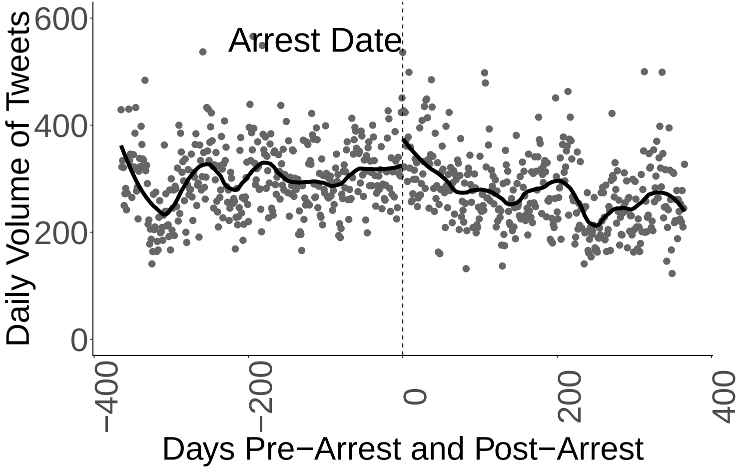
Figure 13: Daily Tweet Volume Month Pre-Arrest vs. Post-Arrest



Disaggregated Placebo

Non-Arrested Keep Tweeting

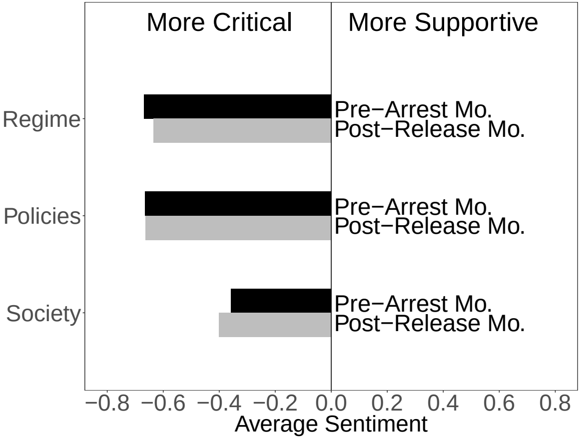
Figure 14: Daily Tweet Volume Year Pre-Arrest vs. Post-Arrest



Disaggregated Placebo

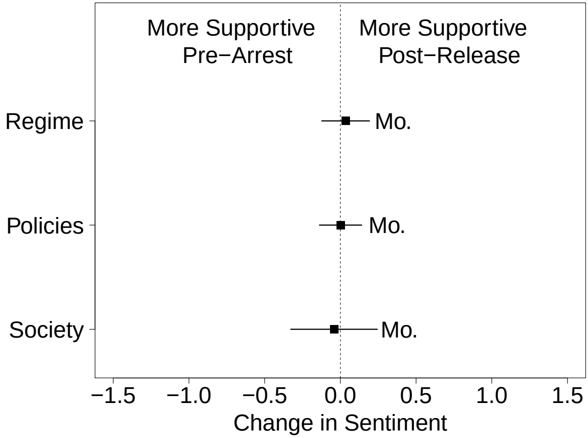
Non-Arrested Keep Dissenting

Figure 15: Change in Non-Arrested Opinion Leaders Tweet Sentiment



Non-Arrested Keep Dissenting

Figure 16: Change in Non-Arrested Opinion Leaders Tweet Sentiment



Results Summary

Effect Type	Actor	Behavior	Deterrent	Backlash
Direct	Arrested Opinion Leaders	Tweet Volume	↓	
Direct	Arrested Opinion Leaders	Tweet Content	↓	
Indirect	Engaged Followers	Engagement Volume		↑
Indirect	Engaged Followers	Tweet Content		↑
Indirect	Saudi Public	Search Volume		↑
Indirect	Non-Arrested Opinion Leaders	Tweet Volume		
Indirect	Non-Arrested Opinion Leaders	Tweet Content		

What's the counterfactual?

- Alternative Interpretation
  - Repression deterred everyone.
- Evidence of backlash
  - Critical tweets reference arrests.
  - Google Search co-occurring terms reference arrests.
  - For some arrested opinion leaders there is a long lag between dissenting tweets and arrests.

Discussion

- Where should we expect to see direct deterrence?
  - Consequences of dissent are severe.
  - Social movement organizations are largely absent.
- Where should we expect to see indirect backlash?
  - Repression is public and overt.
  - Observing repression does not change risk calculus.
  - State lacks control over the online sphere.

Contributions

- First large-scale systematic study of the relationship between offline repression and online dissent.
- Disentangles conflicting findings in the repression-dissent literature:
  - Adjudicates between direct, indirect, and downstream effects.
  - Disaggregates the effects of repression by actor, behavior, and time.
- Provides evidence from "black box" context.
- Substantive implications for how information is controlled in the digital age.

Thank You!

References

Blaydes, Lisa. 2018. *State of Repression: Iraq Under Saddam Hussein*. Princeton University Press.

Davenport, Christian. 2007. "State repression and political order." *Annual Review of Politcal Science* 10:1–23.

Davenport, Christian, Dave Armstrong and Thomas Zeitzoff. 2019. "Acceptable or Out of Bounds? Scaling Perceptions of Challenger and Government Tactics as well as Understanding Contentious Interaction." *Workshop on Human Rights and Repression: Latin America in Comparative Perspective* p. Working paper.

Durkheim, Emile. 1984. *The Division of Labor in Society*. Free Press. [Translated by W.D. Halls.].

Feierabend, Ivo K, Rosalind L Feierabend, Ted Robert Gurr et al. 1972. *Anger, violence, and politics: theories and research*. Prentice-Hall Englewood Cliffs, NJ.

Goldstone, Jack A and Charles Tilly. 2001. "Threat (and opportunity): Popular action and state response in the dynamics of contentious action." *Silence and voice in the study of contentious politics* pp. 179–94.

Gurr, Ted Robert. 1988. "War, revolution, and the growth of the coercive state." *Comparative Political Studies* 21(1):45–65.

Gurr, Ted Robert and Raymond Duvall. 1973. "Civil conflict in the 1960s: A reciprocal theoretical system with parameter estimates." *Comparative Political Studies* 6(2):135–169.

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

Hibbs, Douglas A. 1973. *Mass political violence: A cross-national causal analysis*. Vol. 253 Wiley New York.

Hobbs, William R and Margaret E Roberts. 2018*a*. "How sudden censorship can increase access to information." *American Political Science Review* pp. 1–16.

Hobbs, William R and Margaret E Roberts. 2018*b*. "How sudden censorship can increase access to information." *American Political Science Review* pp. 1–16.

Jansen, Sue Curry and Brian Martin. 2015. "The Streisand effect and censorship backfire." *International Journal of Communication* 9:656–671.

Jenkins, J Craig and Charles Perrow. 1977. "Insurgency of the powerless: Farm worker movements (1946-1972)." *American sociological review* pp. 249–268.

King, Gary, Jennifer Pan and Margaret E Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(2):1–18.

King, Gary, Jennifer Pan and Margaret E Roberts. 2014. "Reverse-engineering censorship in China: Randomized experimentation and participant observation." *Science* 345(6199):1–10.

Nabi, Zubair. 2014. "Censorship is futile." *First Monday* .
URL: *https://firstmonday.org/article/view/5525/4155*

Oberschall, Anthony. 1973. *Social conflict and social movements*. Prentice-Hall Englewood Cliffs, NJ.

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

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

Appendix

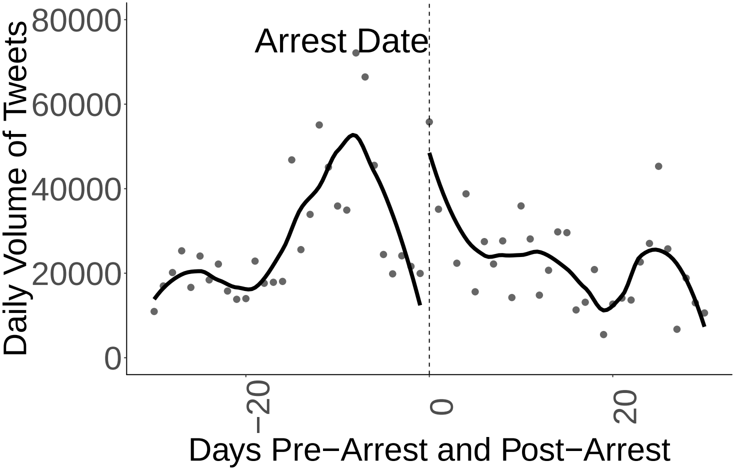
Why repress?

- Institutional culture (Blaydes, 2018; Gurr, 1988)
- Online strategies are infeasible or can backfire (Hassanpour, 2014; Hobbs and Roberts, 2018*a*; Jansen and Martin, 2015; Nabi, 2014; Roberts, 2018).
- Targeted repression might induce broader self-censorship (Stern and Hassid, 2012).
- The goal might not actually be to curtail online dissent!

Appendix

Engaged Followers Keep Engaging (No Indirect Deterrent)

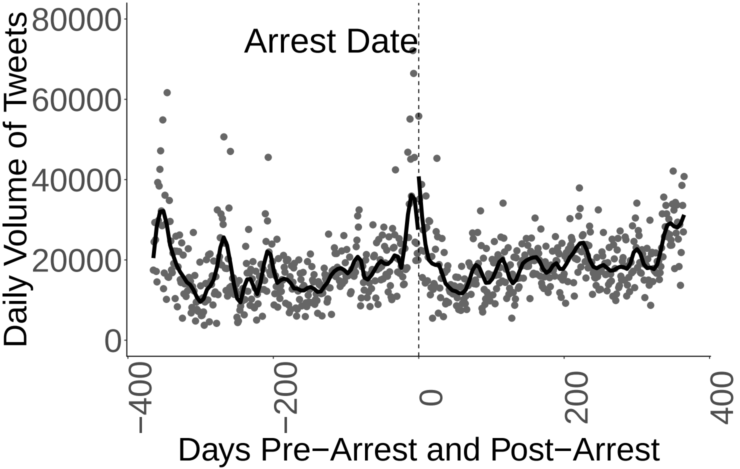
Figure 17: Daily Tweet Volume Month Pre-Arrest vs. Post-Arrest



Placebo Disaggregated Retweet Ratio Appendix.

Engaged Followers Keep Engaging (No Indirect Deterrent)

Figure 18: Daily Tweet Volume Year Pre-Arrest vs. Post-Arrest



Placebo Disaggregated Retweet Ratio Appendix.

Arrested Opinion Leaders Tweet Less (Direct Deterrent)

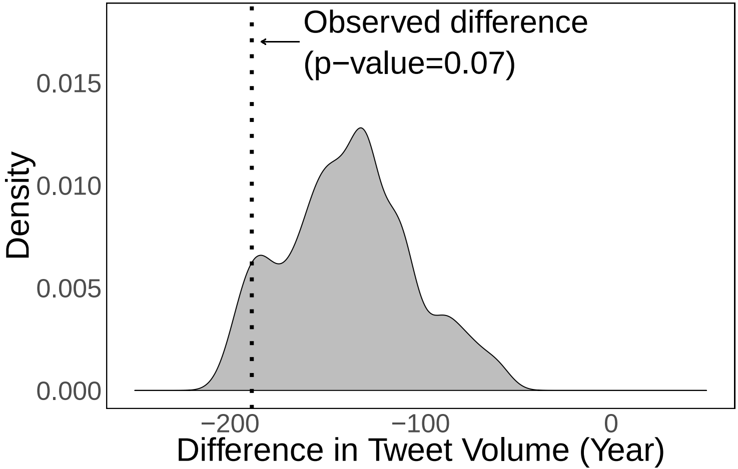
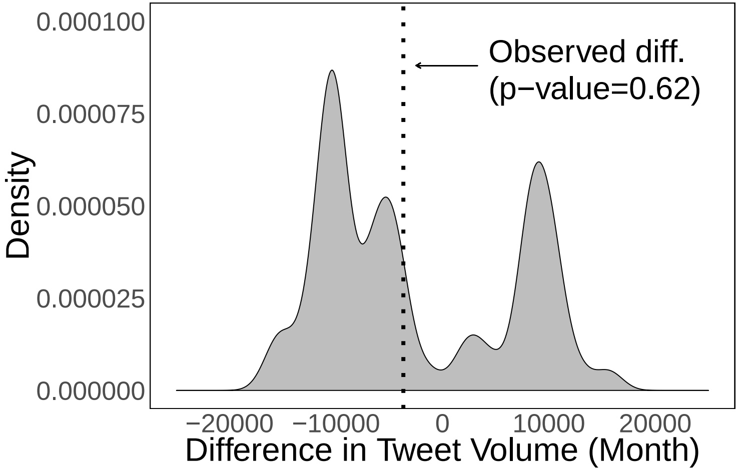


Figure 19: Change in Tweet Volume Year Pre-Arrest vs. Post-Release

Back to main or appendix.

Engaged Followers Keep Engaging (No Indirect Deterrent)

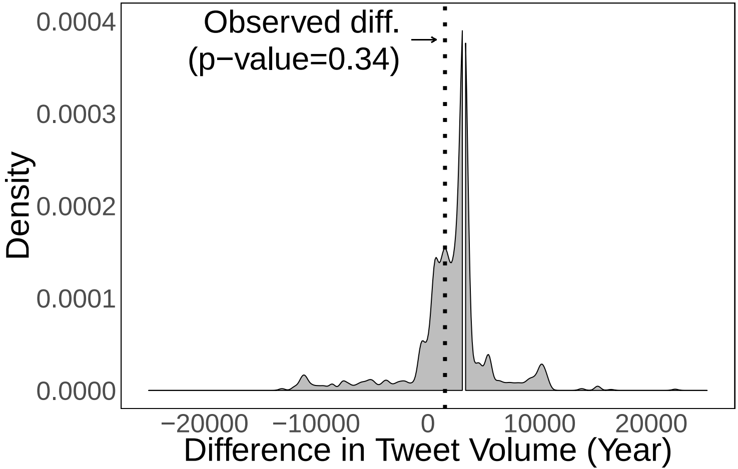
Figure 20: Daily Tweet Volume Month Pre-Arrest vs. Post-Arrest



Back to main or appendix.

Engaged Followers Keep Engaging (No Indirect Deterrent)

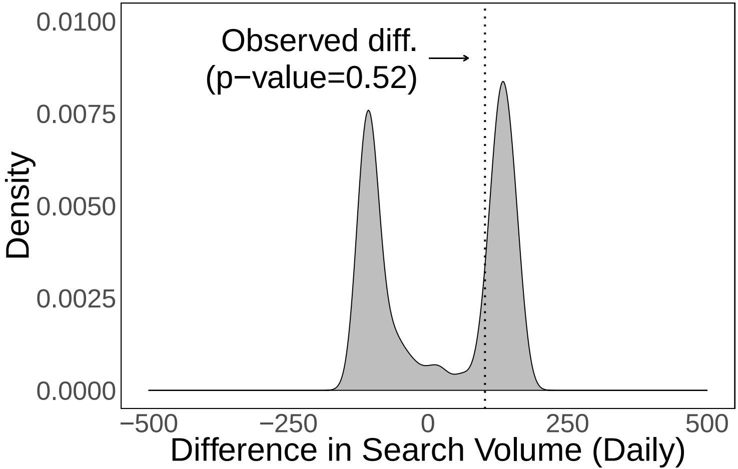
Figure 21: Daily Tweet Volume Year Pre-Arrest vs. Post-Arrest



Back to main or appendix.

Public Keeps Searching (No Indirect Deterrent)

Figure 22: Daily Search Volume Month Pre-Arrest vs. Post-Arrest



Back to main or appendix.

Public Keeps Searching (No Indirect Deterrent)

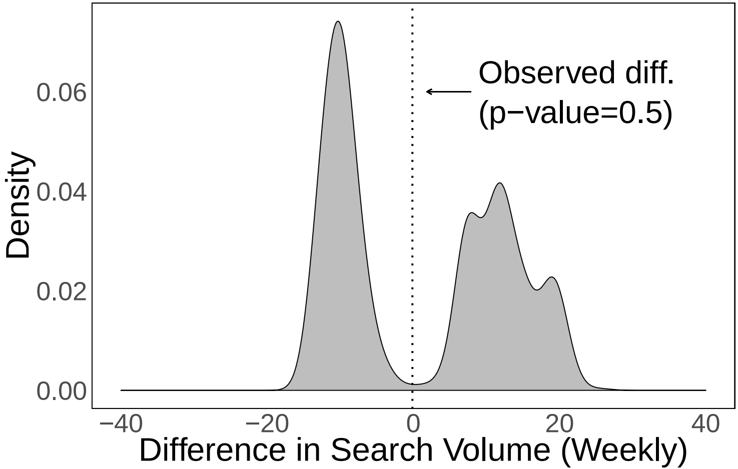
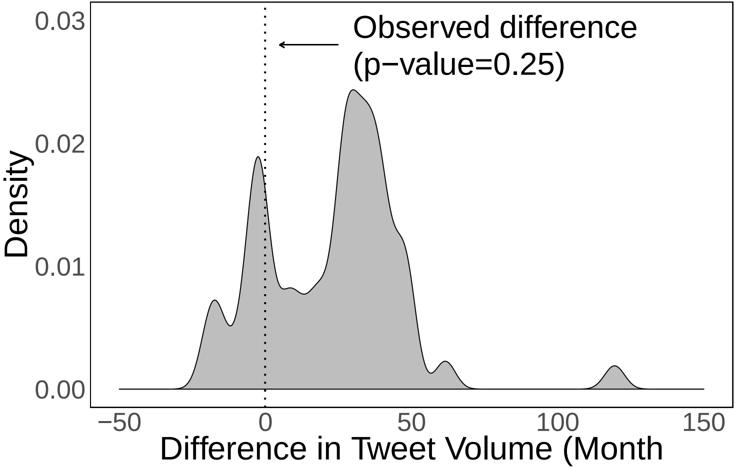


Figure 23: Weekly Search Volume Year Pre-Arrest vs. Post-Arrest

Back to main or appendix.

Non-Arrested Opinion Leaders Keep Tweeting (No Indirect Deterrent)

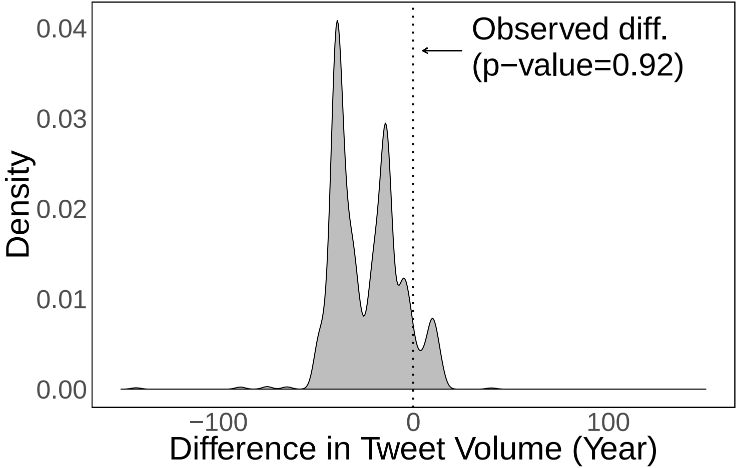
Figure 24: Daily Tweet Volume Month Pre-Arrest vs. Post-Arrest



Back to main or appendix..

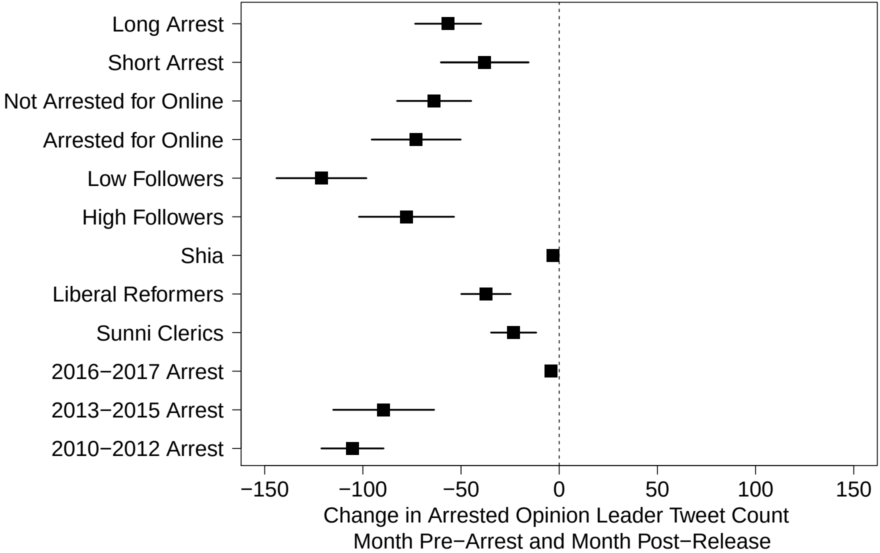
Non-Arrested Opinion Leaders Keep Tweeting (No Indirect Deterrent)

Figure 25: Daily Tweet Volume Year Pre-Arrest vs. Post-Arrest



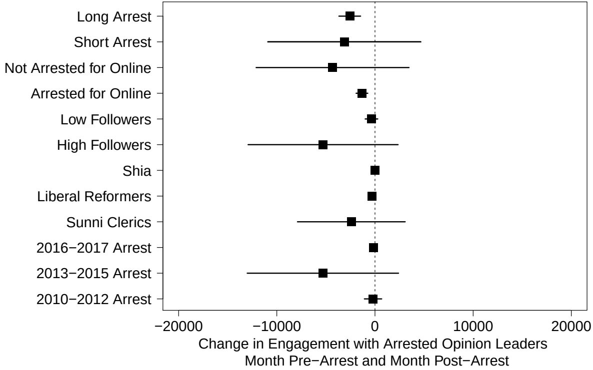
Back to main or appendix..

Arrested Opinion Leaders Tweet Less: Dissaggregated



Engagement: Dissaggregated

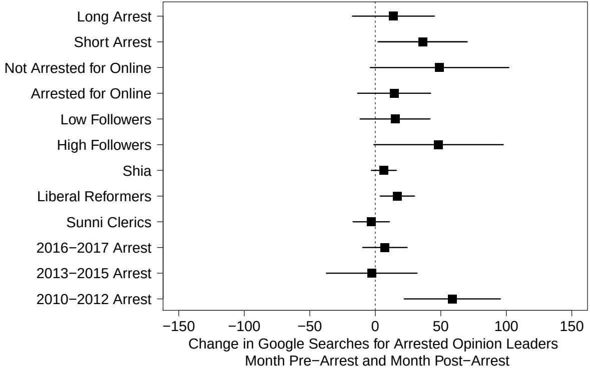
Figure 26: Disaggregated Effect of Political Imprisonment on Volume of Mentions and Retweets of Imprisoned Opinion Leaders



Back to main or appendix..

Google Trends: Dissaggregated

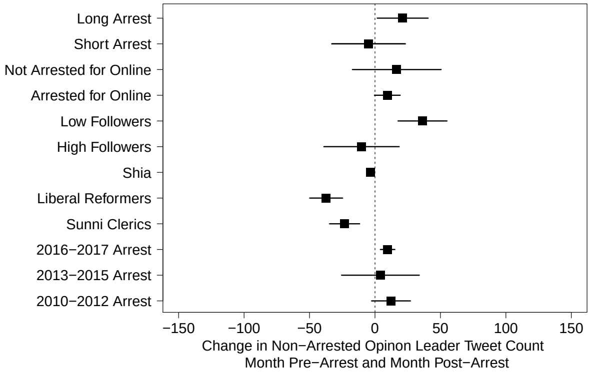
Figure 27: Disaggregated Effect of Political Imprisonment on Google Searches for Imprisoned Opinion Leaders



Back to main or appendix..

Non-Arrested Opinion Leader: Dissaggregated

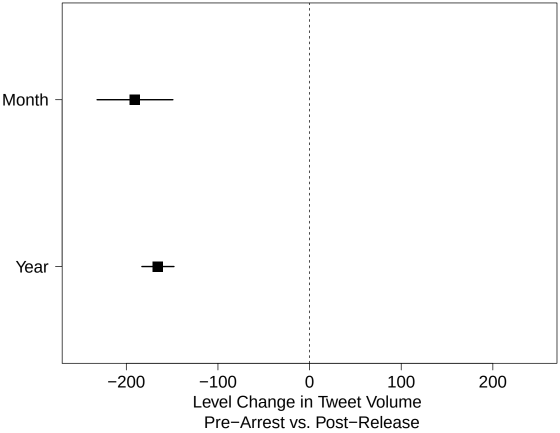
Figure 28: Disaggregated Effect of Political Imprisonment on Non-Imprisoned Opinion Leader Tweet Volume



Back to main or appendix..

Arrested Opinion Leaders: ITSA Tweet Volume

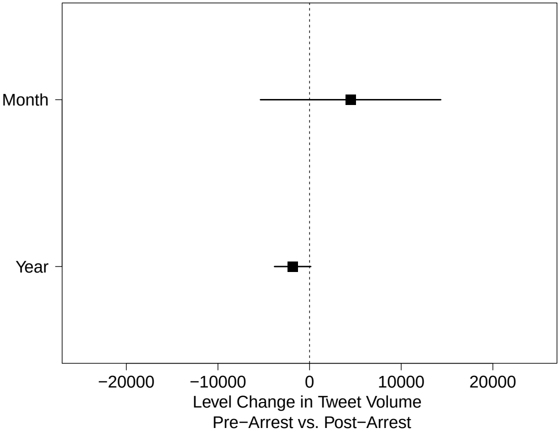
Figure 29: Effect of Imprisonment on Arrested Opinion Leaders Tweet Volume



Back to main or appendix..

Engaged Followers: ITSA Engagement Volume

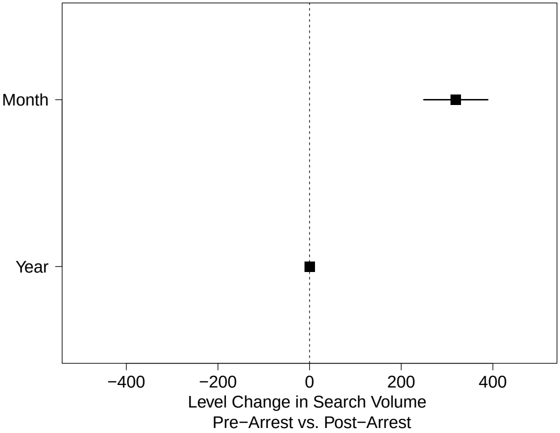
Figure 30: Effect of Imprisonment on Daily Engagement Volume



Back to main or appendix..

Saudi Public: ITSA Google Searches

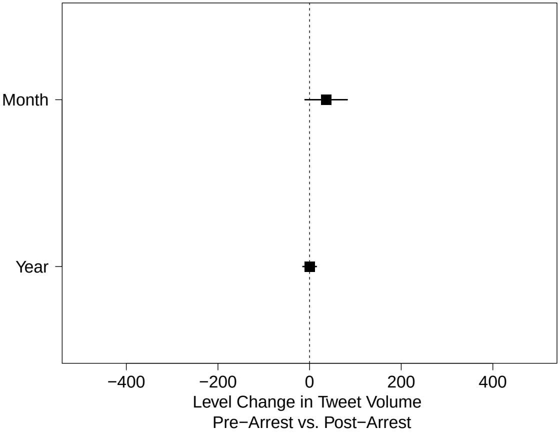
Figure 31: Effect of Imprisonment on Daily/Weekly Search Volume



Back to main or appendix..

Non-Arrested Opinion Leaders: ITSA Tweet Volume

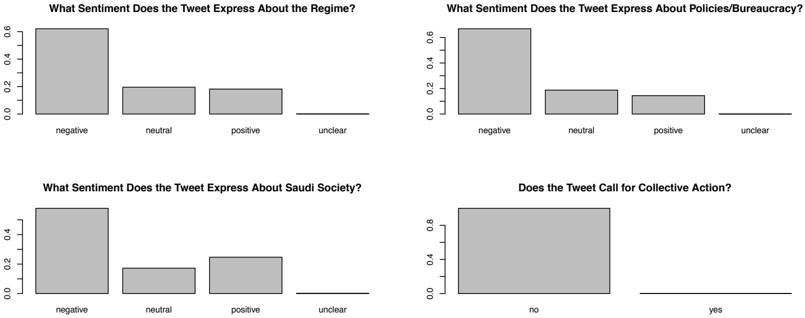
Figure 32: Effect of Imprisonment on Daily Tweet Volume



Back to main or appendix..

Distribution of Tweet Content

Figure 33: Distribution of Tweet Content



Coding distributions across human coded tweets excluding irrelevant tweets.

Back to main or appendix..

Intercoder Agreement

Table 1: Average Intercoder Agreement

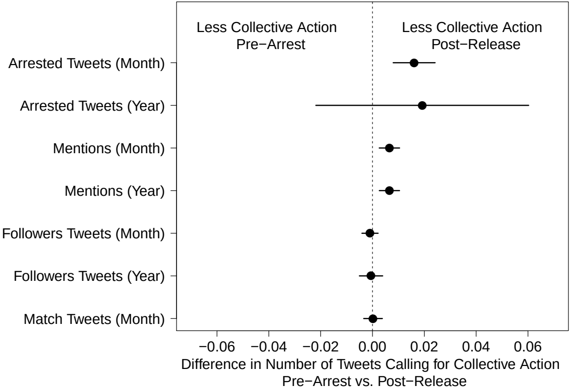
	mean	sd
policies	0.91	0.16
regime	0.91	0.17
society	0.93	0.15
collective action	0.99	0.05

Average intercoder agreement by category among the three human coders that coded each tweet on Figure8.

Back to main or appendix..

Collective Action Content Results

Figure 34: Change in Average Number of Tweets Calling for Collective Action



Back to main or appendix..