Reverse Engineering Chinese Censorship

Gary King, Jennifer Pan, Molly Roberts

Institute for Quantitative Social Science
Harvard University

(talk at Pennsylvania State University, 9/23/2013)
Papers

- An Observational Study:
  

- Experimental and Participatory Studies:
  
  *A Randomized Experimental Study of Censorship in China*

Copies at GaryKing.org
Chinese Censorship

The largest selective suppression of human expression in history
Chinese Censorship

The largest selective suppression of human expression in history:
- implemented *manually*,
- by $\approx 200,000$ workers,
- located in government and inside social media firms
Chinese Censorship

The largest selective suppression of human expression in history:

- implemented \textit{manually},
- by \(\approx 200,000\) workers,
- located in government and inside social media firms

Theories of the Goal of Censorship
Chinese Censorship

The largest selective suppression of human expression in history:
- implemented *manually*,
- by \( \approx 200,000 \) workers,
- located in government and inside social media firms

Theories of the Goal of Censorship
1. Stop criticism of the state
Chinese Censorship

The largest selective suppression of human expression in history:

- implemented *manually*,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. Stop criticism of the state
2. Stop collective action
Chinese Censorship

The largest selective suppression of human expression in history:

- implemented *manually*,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. Stop criticism of the state
2. Stop collective action

Either or both could be right or wrong.
Chinese Censorship

The largest selective suppression of human expression in history:

- implemented *manually*,
- by ≈ 200,000 workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. Stop criticism of the state *Wrong*
2. Stop collective action
Chinese Censorship

The largest selective suppression of human expression in history:
- implemented *manually*,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. Stop criticism of the state *Wrong*
2. Stop collective action *Right*
Chinese Censorship

The largest selective suppression of human expression in history:
- implemented *manually*,
- by \( \approx 200,000 \) workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Theories of the Goal of Censorship</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wrong</strong></td>
<td>1. Stop criticism of the state</td>
</tr>
<tr>
<td><strong>Right</strong></td>
<td>2. Stop collective action</td>
</tr>
</tbody>
</table>

1. Stop criticism of the state *Wrong*
2. Stop collective action *Right*
Chinese Censorship

The largest selective suppression of human expression in history:

- implemented *manually*,
- by \( \approx 200,000 \) workers,
- located in government and inside social media firms

<table>
<thead>
<tr>
<th>Theories of the Goal of Censorship</th>
<th>Benefit</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stop criticism of the state</td>
<td>Wrong</td>
<td></td>
</tr>
<tr>
<td>2. Stop collective action</td>
<td>Right</td>
<td></td>
</tr>
</tbody>
</table>
Chinese Censorship

The largest selective suppression of human expression in history:
- implemented *manually*,
- by \( \approx 200,000 \) workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop criticism of the state</td>
<td>Wrong</td>
</tr>
<tr>
<td>Stop collective action</td>
<td>Right</td>
</tr>
</tbody>
</table>
Chinese Censorship

The largest selective suppression of human expression in history:
- implemented *manually*,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

<table>
<thead>
<tr>
<th>Theory</th>
<th>Benefit</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stop criticism of the state</td>
<td><em>Wrong</em></td>
<td>?</td>
</tr>
<tr>
<td>2. Stop collective action</td>
<td><em>Right</em></td>
<td>Huge</td>
</tr>
</tbody>
</table>

(THEY ALSO CENSOR 2 OTHER SMALLER CATEGORIES WE'LL TALK ABOUT)
Chinese Censorship

The largest selective suppression of human expression in history:
- implemented *manually*,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop criticism of the state <strong>Wrong</strong></td>
<td>?</td>
</tr>
<tr>
<td>Stop collective action <strong>Right</strong></td>
<td>Huge</td>
</tr>
</tbody>
</table>

Either or both could be right or wrong. (They also censor 2 other smaller categories we'll talk about)
Chinese Censorship

The largest selective suppression of human expression in history:

- implemented *manually*,
- by \(\approx 200,000\) workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

<table>
<thead>
<tr>
<th>Theories of the Goal of Censorship</th>
<th>Benefit</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Stop criticism of the state  <em>Wrong</em></td>
<td>?</td>
<td>Huge</td>
</tr>
<tr>
<td>2 Stop collective action  <em>Right</em></td>
<td>Huge</td>
<td>None</td>
</tr>
</tbody>
</table>
Chinese Censorship

The largest selective suppression of human expression in history:
- implemented *manually*,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

<table>
<thead>
<tr>
<th>Theorization</th>
<th>Benefit</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop criticism of the state</td>
<td>Wrong</td>
<td>Huge</td>
</tr>
<tr>
<td>Stop collective action</td>
<td>Right</td>
<td>Huge</td>
</tr>
</tbody>
</table>

(They also censor 2 other smaller categories we’ll talk about)
Chinese Social Media: Fractured over 1,400+ sites
Chinese Social Media: Fractured over 1,400+ sites

(Plus 59% from Sina Blog)
Observational Study

Collect 3,674,698 social media posts in 85 topic areas over 6 months.

Random sample: 127,283

(Repeat design; Total analyzed: 11,382,221)

For each post (on a timeline in one of 85 content areas):
- Download content the instant it appears.
- Carefully revisit each later to determine if it was censored.
- Use computer-assisted methods of text analysis (some existing, some new, all adapted to Chinese).
Observational Study

Collect 3,674,698 social media posts in 85 topic areas over 6 months
Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months
- Random sample: 127,283
Observational Study

- Collect **3,674,698** social media posts in 85 topic areas over 6 months
- Random sample: **127,283**
- (Repeat design; Total analyzed: **11,382,221**)
Collect 3,674,698 social media posts in 85 topic areas over 6 months

Random sample: 127,283

(Repeat design; Total analyzed: 11,382,221)

For each post (on a timeline in one of 85 content areas):
Collect 3,674,698 social media posts in 85 topic areas over 6 months
Random sample: 127,283
(Repeat design; Total analyzed: 11,382,221)
For each post (on a timeline in one of 85 content areas):
- Download content the instant it appears
Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months
- Random sample: 127,283
- (Repeat design; Total analyzed: 11,382,221)
- For each post (on a timeline in one of 85 content areas):
  - Download content the instant it appears
  - (Carefully) revisit each later to determine if it was censored
Collect **3,674,698** social media posts in 85 topic areas over 6 months

Random sample: **127,283**

(Repeat design; Total analyzed: **11,382,221**)

⇝ For each post (on a timeline in one of 85 content areas):
  
  - Download content the instant it appears
  - (Carefully) revisit each later to determine if it was censored
  - Use computer-assisted methods of text analysis (some existing, some new, all adapted to Chinese)
Censorship is not Ambiguous: Example Error Page

Sorry, the host you were looking for does not exist, has been deleted, or is being investigated.
Censorship is not Ambiguous: BBS Error Page

The page you requested is temporarily down. How about you go look at another page.

Jingjing, one of China's cartoon internet police
The Censors are Fast; Our Automated Methods are Faster
The Censors are Fast; Our Automated Methods are Faster
Example: Shanghai Subway Crash
The Censors are Fast; Our Automated Methods are Faster
Example: Shanghai Subway Crash
For 2 Unusual Topics: Constant Censorship Effort
For 2 Unusual Topics: Constant Censorship Effort

**Pornography**

**Criticism of the Censors**
All other topics: Censorship & Post Volume are “Bursty”
All other topics: Censorship & Post Volume are “Bursty”

Unit of analysis: volume burst (≈ 3 SDs greater than baseline volume)

We monitored 85 topic areas (Jan–July 2011)

Found 87 volume bursts in total

Identified real world events associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)
All other topics: Censorship & Post Volume are “Bursty”

- Unit of analysis:

![Graph showing count published vs count censored from January to July with a spike in June indicating riots in Zengcheng.]

- **Unit of analysis:**
  - Volume burst (≈ 3 SDs greater than baseline volume)
  - We monitored 85 topic areas (Jan–July 2011)
  - Found 87 volume bursts in total
  - Identified real world events associated with each burst

- Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state).
All other topics: Censorship & Post Volume are “Bursty”

Unit of analysis:
- volume burst
All other topics: Censorship & Post Volume are “Bursty”

- Unit of analysis:
  - volume burst
  - (≈ 3 SDs greater than baseline volume)

We monitored 85 topic areas (Jan–July 2011) Found 87 volume bursts in total. Identified real world events associated with each burst.

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state).
All other topics: Censorship & Post Volume are “Bursty”

- Unit of analysis:
  - Volume burst
  - (≈ 3 SDs greater than baseline volume)

- We monitored 85 topic areas (Jan–July 2011)
All other topics: Censorship & Post Volume are “Bursty”

- Unit of analysis:
  - volume burst
  - (≈ 3 SDs greater than baseline volume)

- We monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
All other topics: Censorship & Post Volume are “Bursty”

- **Unit of analysis:**
  - volume burst
  - \((\approx 3 \text{ SDs greater than baseline volume})\)
- **We monitored 85 topic areas** (Jan–July 2011)
- **Found 87 volume bursts in total**
- **Identified real world events** associated with each burst
All other topics: Censorship & Post Volume are “Bursty”

- **Unit of analysis:**
  - volume burst
  - ($\approx$ 3 SDs greater than baseline volume)

- We monitored **85 topic areas** (Jan–July 2011)
- Found **87 volume bursts** in total
- Identified real world events associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)
All other topics: Censorship & Post Volume are “Bursty”

- Unit of analysis:
  - volume burst
  - ($\approx 3$ SDs greater than baseline volume)

- We monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
- Identified real world events associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)
All other topics: Censorship & Post Volume are “Bursty”

- Unit of analysis:
  - volume burst
  - ($\approx 3$ SDs greater than baseline volume)

- We monitored 85 topic areas (Jan–July 2011)

- Found 87 volume bursts in total

- Identified real world events associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)
All other topics: Censorship & Post Volume are “Bursty”

- Unit of analysis:
  - volume burst
  - ($\approx 3$ SDs greater than baseline volume)

- We monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
- Identified real world events associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)
All other topics: Censorship & Post Volume are “Bursty”

- **Unit of analysis:**
  - volume burst
  - ($\approx 3$ SDs greater than baseline volume)

- We monitored **85 topic areas** (Jan–July 2011)
- Found **87 volume bursts** in total
- Identified real world events associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)
Observational Test 1: Post Volume

Begin with our 87 volume bursts in 85 topic areas. For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude.

If the goal of censorship is to stop collective action, we expect:

1. On average, % censored should increase during volume bursts.
2. Some bursts (associated with politically relevant events) should have much higher censorship.
Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas
Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
Observational Test 1: Post Volume

- Begin with our **87 volume bursts** in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – **censorship magnitude**
- If goal of censorship is to stop collective action, **we expect:**
Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
- If goal of censorship is to stop collective action, we expect:

On average, % censored should increase during volume bursts.
Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
- If goal of censorship is to stop collective action, we expect:

1. On average, % censored should increase during volume bursts
2. Some bursts (associated with politically relevant events) should have much higher censorship
Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
- If goal of censorship is to stop collective action, we expect:

1. On average, % censored should increase during volume bursts
2. Some bursts (associated with politically relevant events) should have much higher censorship
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. Protest or organized crowd formation outside the Internet
2. Individuals who have organized or incited collective action on the ground in the past;
3. Topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
4. Other News
5. Government Policies

13/29
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. Collective Action Potential
Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. **Collective Action Potential**
   - protest or organized crowd formation outside the Internet
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. Criticism of censors
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. Criticism of censors

3. Pornography
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. **Collective Action Potential**
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. **Criticism of censors**

3. **Pornography**

4. **(Other) News**
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. Criticism of censors

3. Pornography

4. (Other) News

5. Government Policies
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. **Collective Action Potential**
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. **Criticism of censors**
3. **Pornography**
4. **(Other) News**
5. **Government Policies**
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. **Collective Action Potential**
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. **Criticism of censors**

3. **Pornography**

4. **(Other) News**

5. **Government Policies**
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. **Collective Action Potential**
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. **Criticism of censors**

3. **Pornography**

4. **(Other) News**

5. **Government Policies**
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. **Collective Action Potential**
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. **Criticism of censors**

3. **Pornography**

4. **(Other) News**

5. **Government Policies**
Observational Test 2: The Events Generating the Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. **Collective Action Potential**
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

2. **Criticism of censors**

3. **Pornography**

4. **(Other) News**

5. **Government Policies**
What Types of Events Are Censored?
What Types of Events Are Censored?

Protests in Inner Mongolia
Pornography Disguised as News
Baidu Copyright Lawsuit
Zengcheng Protests
Pornography Mentioning Popular Book
Ai Weiwei Arrested
Collective Anger At Lead Poisoning in Jiangsu
Google is Hacked
Localized Advocacy for Environment Lottery
Fuzhou Bombing
Students Throw Shoes at Fang BinXing
Rush to Buy Salt After Earthquake
New Laws on Fifty Cent Party

U.S. Military Intervention in Libya
Food Prices Rise
Education Reform for Migrant Children
Popular Video Game Released
Indoor Smoking Ban Takes Effect
News About Iran Nuclear Program
Jon Huntsman Steps Down as Ambassador to China
Gov't Increases Power Prices
China Puts Nuclear Program on Hold
Chinese Solar Company Announces Earnings
EPA Issues New Rules on Lead
Disney Announced Theme Park
Popular Book Published in Audio Format

Censorship Magnitude

-0.2 0 0.1 0.3 0.5 0.7

Collective Action
Criticism of Censors
Pornography

Policies
News
Censoring Collective Action: Ai Weiwei’s Arrest

![Graph showing the count of published and censored articles mentioning Ai Weiwei's arrest over time. The graph highlights a significant spike in censored articles in April.]
Censoring Collective Action: Protests in Inner Mongolia
Low Censorship on One Child Policy

Speculation of Policy Reversal at NPC

Count Published

Count Censored
Low Censorship on News: Power Prices

Power shortages
Gov't raises
power prices
to curb demand
Observational Test 3: Content of Posts

We adapted "ReadMe" (Hopkins-King) methodology to the content of censored posts in Chinese.

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)

Code a validation set

ReadMe estimates: % of posts in each category
Quantity of interest: % of posts censored in a category

\[
P(Censored|\text{Category}) = P(\text{Category}|Censored) \cdot P(Censored) / P(\text{Category})\]
Observational Test 3: Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese
We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
Observational Test 3: Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
Observational Test 3: Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
Observational Test 3: Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (\% censored by category)
Observational Test 3: Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)
  - Code a validation set
Observational Test 3: Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)
  - Code a validation set
  - ReadMe estimates: % of posts in each category
Observational Test 3: Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)
  - Code a validation set
  - ReadMe estimates: % of posts in each category
  - Quantity of interest: % of posts censored in a category
Observational Test 3: Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)
  - Code a validation set
  - ReadMe estimates: % of posts in each category
  - Quantity of interest: % of posts censored in a category

\[
P(\text{Censored}|\text{Category}) = \frac{P(\text{Category}|\text{Censored})P(\text{Censored})}{P(\text{Category})}
\]
“ReadMe” Algorithm Validated in Chinese
“ReadMe” Algorithm Validated in Chinese

Example: Labor Strikes, 2010 (Training set: 100; Test set: 900)
“ReadMe” Algorithm Validated in Chinese

Example: Labor Strikes, 2010 (Training set: 100; Test set: 900)
“ReadMe” Algorithm Validated in Chinese

Example: Labor Strikes, 2010 (Training set: 100; Test set: 900)
Uncensored: Non-Collective Action Posts (that Support or Oppose the State)
Uncensored: Non-Collective Action Posts (that Support or Oppose the State)

One Child Policy  Corruption Policy  Food Prices Rise

Percent Censored

0.0  0.2  0.4  0.6  0.8  1.0
Criticize  Support  Criticize  Support  Criticize  Support
Censored: Collective Action Posts (that Support or Criticize the State)

<table>
<thead>
<tr>
<th></th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>

### Percent Censored

- 0.0
- 0.2
- 0.4
- 0.6
- 0.8
- 1.0

### Censorship Categories

- Support
- Criticize

**Examples:**

- Ai Weiwei
- Inner Mongolia
- Fuzhou Bombing

22/29
Censored: Collective Action Posts (that Support or Criticize the State)
Additional Research Designs

Randomized Experiment (for causal inferences)

Selected 100 top social media sites (\(\sim 87\%\) of blogs, >500M Users, geographically diverse)
Created 2 accounts on each (from inside China all over the country)
Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)
Submitted posts randomly assigned to type
Checked on censorship (from computers all over the world)

Participatory Study (for descriptive inferences)

Current method of learning how they censor: ask (carefully!)
Our goal: change our sources’ incentives
Procedure: create our own social media website in China
Bought URL; contracted with firms for servers & software; posted and censored ourselves
To learn: we tried every software option, read the documentation, and called customer support(!)
1 Randomized Experiment (for causal inferences)

Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
Created 2 accounts on each (from inside China all over the country)
Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)
Submitted posts randomly assigned to type
Checked on censorship (from computers all over the world)

2 Participatory Study (for descriptive inferences)

Current method of learning how they censor: ask (carefully!)
Our goal: change our sources' incentives
Procedure: create our own social media website in China
Bought URL; contracted with firms for servers & software; posted and censored ourselves
To learn: we tried every software option, read the documentation, and called customer support(!)
Additional Research Designs

1. Randomized Experiment (for causal inferences)

2. Participatory Study (for descriptive inferences)
1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (≈87% of blogs, >500M Users, geographically diverse)

2. Participatory Study (for descriptive inferences)
Additional Research Designs

1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (∼87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)

2. Participatory Study (for descriptive inferences)
Additional Research Designs

1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)
   - Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)

2. Participatory Study (for descriptive inferences)
Additional Research Designs

1 Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)
   - Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)
   - Submitted posts randomly assigned to type

2 Participatory Study (for descriptive inferences)
Additional Research Designs

1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (∼87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)
   - Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)
   - Submitted posts randomly assigned to type
   - Checked on censorship (from computers all over the world)

2. Participatory Study (for descriptive inferences)

   Current method of learning how they censor: ask (carefully!)
   Our goal: change our sources' incentives
   Procedure: create our own social media website in China
   Bought URL; contracted with firms for servers & software; posted and censored ourselves
   To learn: we tried every software option, read the documentation, and called customer support(!)
Additional Research Designs

1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)
   - Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)
   - Submitted posts randomly assigned to type
   - Checked on censorship (from computers all over the world)

2. Participatory Study (for descriptive inferences)
   - Current method of learning how they censor: ask (carefully!)
Additional Research Designs

1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (∼87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)
   - Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)
   - Submitted posts randomly assigned to type
   - Checked on censorship (from computers all over the world)

2. Participatory Study (for descriptive inferences)
   - Current method of learning how they censor: ask (carefully!)
   - Our goal: change our sources’ incentives
Additional Research Designs

1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)
   - Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
   - Submitted posts randomly assigned to type
   - Checked on censorship (from computers all over the world)

2. Participatory Study (for descriptive inferences)
   - Current method of learning how they censor: ask (carefully!)
   - Our goal: change our sources’ incentives
   - Procedure: create our own social media website in China
Additional Research Designs

1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)
   - Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)
   - Submitted posts randomly assigned to type
   - Checked on censorship (from computers all over the world)

2. Participatory Study (for descriptive inferences)
   - Current method of learning how they censor: ask (carefully!)
   - Our goal: change our sources’ incentives
   - Procedure: create our own social media website in China
   - Bought URL; contracted with firms for servers & software; posted and censored ourselves
Additional Research Designs

1. Randomized Experiment (for causal inferences)
   - Selected 100 top social media sites (≈87% of blogs, >500M Users, geographically diverse)
   - Created 2 accounts on each (from inside China all over the country)
   - Wrote 1,200 unique social media posts (CA/not CA, Pro/Anti)
   - Submitted posts randomly assigned to type
   - Checked on censorship (from computers all over the world)

2. Participatory Study (for descriptive inferences)
   - Current method of learning how they censor: ask (carefully!)
   - Our goal: change our sources’ incentives
   - Procedure: create our own social media website in China
   - Bought URL; contracted with firms for servers & software; posted and censored ourselves
   - To learn: we tried every software option, read the documentation, and called customer support(!)
Mechanisms of Censorship
Mechanisms of Censorship

1. Submit text for posting
   - Immediately held for review before publication

2. Publish immediately
   - Post remains online

3. Post is removed (within 24 hours)
   - Sorry, the host you were looking for does not exist, has been deleted, or is being investigated

4. Post is published after review
   - Administrator has set up review for new topics, your post will be shown after it has passed review

5. Post deleted during review
   - (Post reviewed then deleted)

6. Account blocked
   - You have been silenced by the administrator, the type of silence is forever

7. Published
   - 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...

   - 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...
Mechanisms of Censorship

1. Submit text for posting
   - Immediately held for review before publication

2. Publish immediately
   - Post remains online
   - Post is removed (within 24 hours)

3. Post published after review
   - Administrator has set up review for new topics, your post will be shown after it has passed review.
   - Post is published after review

4. Post deleted during review
   - Post reviewed then deleted

5. Account blocked
   - You have been silenced by the administrator, the type of silence is forever!

Examples:
- 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.
- 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...
Mechanisms of Censorship

Post remains online

Post is removed (within 24 hours)

Publish immediately

Submit text for posting

Immediately held for review before publication

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

Post is published after review

Post deleted during review

Account blocked

PUBLISHED

CENSORED

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...

YOU HAVE BEEN SILENCED BY THE ADMINISTRATOR, THE TYPE OF SILENCE IS FOREVER!

66% OF WEBSITES REVIEW SOME CONTENT

41% OF SUBMISSIONS ARE REVIEWED

OBSERVATIONAL STUDY

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...
Posts For v. Against Government: Zero Causal Effect
Posts For v. Against Government: Zero Causal Effect
Posts For v. Against Government: Zero Causal Effect

Panxu Protest
Posts For v. Against Government: Zero Causal Effect

Panxu Protest

Censorship Difference (Pro − Anti)
Posts For v. Against Government: Zero Causal Effect
Posts For v. Against Government: Zero Causal Effect

Censorship Difference (Pro − Anti)

Panxu Protest
Tibetan Self-Immolations
Ai Weiwei Album
Posts For v. Against Government: Zero Causal Effect

Censorship Difference (Pro − Anti)

Panxu Protest
Tibetan Self-Immolations
Ai Weiwei Album
Protests in Xinjiang

25/29
Posts For v. Against Government: Zero Causal Effect
Posts For v. Against Government: Zero Causal Effect

Censorship Difference (Pro − Anti)

Panxu Protest
Ai Weiwei Album
Corruption Policy
Tibetan Self-Immolations
Protests in Xinjiang
Eliminate Golden Week

Panxu Protest
Ai Weiwei Album
Corruption Policy
Tibetan Self-Immolations
Protests in Xinjiang
Eliminate Golden Week
Posts For v. Against Government: Zero Causal Effect
Posts For v. Against Government: Zero Causal Effect

Censorship Difference (Pro − Anti)

-0.5 0.0 0.5 1.0

Panxu Protest
Ai Weiwei Album
Corruption Policy
Eliminate Golden Week
Rental Tax

Tibetan Self-Immolations
Protests in Xinjiang
Yellow Light Fines

Protests in Xinjiang
Corruption Policy
Eliminate Golden Week
Rental Tax

Panxu Protest
Ai Weiwei Album
Corruption Policy
Eliminate Golden Week
Rental Tax

25/29
Posts For v. Against Government: Zero Causal Effect

Censorship Difference (Pro − Anti)

Panxu Protest
Ai Weiwei Album
Corruption Policy
Tibetan Self-Immolations
Protests in Xinjiang
Eliminate Golden Week
Rental Tax
Yellow Light Fines
Stock Market Crash

-0.5 0.0 0.5 1.0
Posts For v. Against Government: Zero Causal Effect

Censorship Difference (Pro − Anti)

-0.5 0.0 0.5 1.0

Panxu Protest
Ai Weiwei Album
Corruption Policy
Tibetan Self-Immolations
Protests in Xinjiang
Eliminate Golden Week
Yellow Light Fines
Rental Tax
Stock Market Crash
Investigation of Sichuan Vice Governor
Posts For v. Against Government: Zero Causal Effect
Posts For v. Against Government: Zero Causal Effect
Collective Action Events: Large Causal Effect
Collective Action Events: Large Causal Effect
Collective Action Events: Large Causal Effect
Collective Action Events: Large Causal Effect
Collective Action Events: Large Causal Effect

Censorship Difference (CA Event − Non-CA Event)

Panxu Protest
Tibetan Self-Immolations
Collective Action Events: Large Causal Effect

- Panxu Protest
- Tibetan Self-Immolations
- Ai Weiwei Album

Censorship Difference (CA Event - Non-CA Event)
Collective Action Events: Large Causal Effect

Graph showing the censorship difference (CA Event − Non-CA Event) for various events:
- Panxu Protest
- Tibetan Self-Immolations
- Ai Weiwei Album
- Protests in Xinjiang
Censorship to Preempt Collective Action: Ai Weiwei’s Arrest

<table>
<thead>
<tr>
<th>Date</th>
<th>% of Posts Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar. 19</td>
<td>0.0</td>
</tr>
<tr>
<td>Mar. 29</td>
<td>0.2</td>
</tr>
<tr>
<td>Apr. 08</td>
<td>0.4</td>
</tr>
<tr>
<td>Apr. 18</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Actual % censorship vs. Predicted % censorship trend based on 3/19-3/29 data.

Mar. 29, 5 days prior to Ai Weiwei’s arrest.

Placebo Test: Most extreme of all effects.
Censorship to Preempt Collective Action: Ai Weiwei’s Arrest

Mar. 29, 5 days prior

Apr. 3, Ai Weiwei Arrested

Actual % censorship

Predicted % censorship trend based on 3/19–3/29 data

Mar 19 Mar 29 Apr 08 Apr 18

% of Posts Censored

0.0 0.2 0.4 0.6 0.8 1.0

Mar. 29, 5 days prior

Apr. 3, Ai Weiwei Arrested

Actual % censorship

Predicted % censorship trend based on 3/19–3/29 data
Censorship to Preempt Collective Action: Ai Weiwei’s Arrest

Mar. 29, 5 days prior

Apr. 3, Ai Weiwei Arrested

Actual % censorship

Predicted % censorship trend based on 3/19–3/29 data

Placebo Test:
Censorship to Preempt Collective Action: Ai Weiwei’s Arrest

Mar. 29, 5 days prior

Apr. 3, Ai Weiwei Arrested

Actual % censorship

Predicted % censorship trend based on 3/19–3/29 data

Placebo Test: Most extreme of all effects
Concluding Remarks

Chinese people's speech: Individually Free, Collectively in Chains

The Chinese Censorship Program:
- Stops collective action, not criticism
- Reveals government intentions; predicts actions

Our methodology:
- Observational, Experimental, Participatory
- Automated text analysis methods validated in Chinese (without translation)
- Enables continuous time measurement of Chinese government censorship activities
- Reveals information about state action when traditional media is silent
- Possibly applicable to other countries
Chinese people’s speech: Individually Free, Collectively in Chains
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
  - Stops collective action, not criticism
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
  - Stops collective action, not criticism
  - Reveals government intentions; predicts actions
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
  - Stops collective action, not criticism
  - Reveals government intentions; predicts actions
- Our methodology:
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
  - Stops collective action, not criticism
  - Reveals government intentions; predicts actions
- Our methodology:
  - Observational, Experimental, Participatory
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
  - Stops collective action, not criticism
  - Reveals government intentions; predicts actions
- Our methodology:
  - Observational, Experimental, Participatory
  - Automated text analysis methods validated in Chinese (without translation)
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
  - Stops collective action, not criticism
  - Reveals government intentions; predicts actions
- Our methodology:
  - Observational, Experimental, Participatory
  - Automated text analysis methods validated in Chinese (without translation)
  - Enables continuous time measurement of Chinese government censorship activities
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
  - Stops collective action, not criticism
  - Reveals government intentions; predicts actions
- Our methodology:
  - Observational, Experimental, Participatory
  - Automated text analysis methods validated in Chinese (without translation)
  - Enables continuous time measurement of Chinese government censorship activities
  - Reveals information about state action when traditional media is silent
Concluding Remarks

- Chinese people’s speech: Individually Free, Collectively in Chains
- The Chinese Censorship Program:
  - Stops collective action, not criticism
  - Reveals government intentions; predicts actions
- Our methodology:
  - Observational, Experimental, Participatory
  - Automated text analysis methods validated in Chinese (without translation)
  - Enables continuous time measurement of Chinese government censorship activities
  - Reveals information about state action when traditional media is silent
  - Possibly applicable to other countries
For more information

GaryKing.org

j.mp/JenPan

j.mp/MollyRoberts
### Predicting the South China Sea Peace Agreement

<table>
<thead>
<tr>
<th>Date</th>
<th>% of Posts Censored</th>
<th>Predicted % Censor Trend Based on 6/10−6/20 Data</th>
<th>Actual % Censorship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jun. 20, 5 days prior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jun. 25, Peace Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Placebo Test:** Most extreme of all effects

31/29
Predicting the South China Sea Peace Agreement

- Jun. 20, 5 days prior
- Jun. 25, Peace Agreement

Predicted % censor trend based on 6/10–6/20 data

Actual % censorship
Predicting the South China Sea Peace Agreement

Predicted % censorship trend based on 6/10–6/20 data

Actual % censorship

Placebo Test:
Predicting the South China Sea Peace Agreement

Jun. 20, 5 days prior

Jun. 25, Peace Agreement

Predicted % censor trend based on 6/10–6/20 data

Actual % censorship

Placebo Test: Most extreme of all effects
Censorship Pre-empting Collective Action: Wang Lijun’s Demotion
Censorship Pre-empting Collective Action: Wang Lijun’s Demotion

Jan. 28, 5 days prior
Feb. 2, Wang Lijun demoted

Actual % censorship

Predicted % censorship trend based on 1/18–1/28 data
Censorship Pre-empting Collective Action: Wang Lijun’s Demotion

Jan. 28, 5 days prior

Feb. 2, Wang Lijun demoted

Actual % censorship

Predicted % censorship trend based on 1/18–1/28 data

Placebo Test:

Most extreme of all effects

32/29
Censorship Pre-empting Collective Action: Wang Lijun’s Demotion

Jan. 28, 5 days prior
Feb. 2, Wang Lijun demoted

Actual % censorship

Jan. 23 Jan 30 Feb 06 Feb 13

% of Posts Censored

Predicted % censorship trend based on 1/18–1/28 data

Placebo Test: Most extreme of all effects
Uncensored Posts (w/o Collective Action Potential)
Critical of the State
Uncensored Posts (w/o Collective Action Potential)
Critical of the State

<table>
<thead>
<tr>
<th>State Critique</th>
<th>Collective Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
This is a city government [Yulin City, Shaanxi] that treats life with contempt, this is government officials run amuck, a city government without justice, a city government that delights in that which is vulgar, a place where officials all have mistresses, a city government that is shameless with greed, a government that trades dignity for power, a government without humanity, a government that has no limits on immorality, a government that goes back on its word, a government that treats kindness with ingratitude, a government that cares nothing for posterity...
Censored Post (with Collective Action Potential)
Supporting the State
Censored Post (with Collective Action Potential)
Supporting the State

<table>
<thead>
<tr>
<th>State Critique</th>
<th>Collective Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>CENSORED</td>
</tr>
<tr>
<td>no</td>
<td>Free</td>
</tr>
</tbody>
</table>

Free
The bombing led not only to the tragedy of his death but the death of many government workers. Even if we can verify what Qian Mingqi said on Weibo that the building demolition caused a great deal of personal damage, we should still condemn his extreme act of retribution.... The government has continually put forth measures and laws to protect the interests of citizens in building demolition. And the media has called attention to the plight of those experiencing housing demolition. The rate at which compensation for housing demolition has increased exceeds inflation. In many places, this compensation can change the fate of an entire family.