Measuring Decentralization of Chinese Keyword Censorship via Mobile Games

Jeffrey Knockel, Lotus Ruan, and Masashi Crete-Nishihata

Citizen Lab, Munk School of Global Affairs, University of Toronto
Dept. of Computer Science, University of New Mexico
“1989年民运”
(1989 Year Democracy Movement)
“习近平时代”
（Xi Jinping Era）
“Baby Mama Drama”
“Baby Mama Drama”

(A keyword appearing in a chat client)
Who determines what’s censored in Chinese apps?
Centralized and Monolithic?

- Implementations are uniform
- What is censored necessarily reflects CPC strategies
- e.g., collective action targeted, government criticism permitted (King, Pan, Roberts; 2013, 2014)
Decentralized and Fragmented?

- Intermediary liability
- Censorship laws and policy can be intentionally vague
- Responsibility for implementing censorship pushed down to companies
- “Anaconda in the Chandelier” (Perry Link)
How can we understand which is right?

- Analyzing censorship in apps used in China
- Client-side censorship offers research opportunities
- Extract entire keyword lists used to trigger censorship
- Compare across apps and industries
Previous work

Chat (IM) clients

- TOM-Skype
- Sina UC
- LINE

Found no central blacklist among lists

\[ n = 3 \]

(Knockel et al, 2011, Crandall et al 2013, Hardy 2013)
Previous work

Live streaming platforms

- YY
- Sina Show
- 9158
- GuaGua

Keyword similarities explained by developer similarities

\[ n = 4 \text{ (or 7)} \]

(Knockel et al, 2015)
China has the world’s largest and most lucrative mobile gaming market

Estimated value of over 27.5 billion US$ in 2017

Registration Approval

→ Ministry of Culture

Publication License

→ State Administration of Press, Publication, Radio, Film and Television
Prohibited Content in Online Games

1. violating basic principles set by the Constitution;
2. jeopardizing national unity, state sovereignty and territorial integrity;
3. leaking state secrets, endangering state security or damaging state honor and interests;
4. instigating ethnic hatred or discrimination, jeopardizing ethnic unity, and infringing ethnic rituals or customs;
5. promoting heretical or superstitious idea;
6. spreading rumors, disrupting social order and stability;
7. disseminating obscenity, pornography, gambling, violence or abetting crime;
8. humiliating or slandering others, infringing the lawful rights of others;
9. transgressing social morality;
10. other contents forbidden by laws and administrative regulations.
Mobile Games in China

There are a lot more Chinese games than Chinese chat platforms!

\[ n > 200 \]

Allows us to test new hypotheses.

Commonly censor in game chat and usernames.

Many of these games are international games adapted for the Chinese market.
“Initiating banned keywords data~!”
Please enter your user name: Xi Jinping

User name does not comply with regulations, please re-enter.
Sampling methodology

- Collected first 500 results from Hi Market using search query that only returned highly downloaded Chinese-developed games
- Same for internationally developed games
- Searched APKs for sensitive words
  - falun, 法轮 (falun), fuck, 狗 (fuck)
- Searched for censorship-related strings
  - blacklist, censor, dirty, filter, forbid, illegal, keyword, profan, sensitiv
Hypotheses

Censorship keyword lists are:

1) Determined at the city or provincial level
2) Determined for specific genres of games
3) Related to the date that games are released
4) Largely determined by the publisher or developer
Keyword Lists

From 836 games, found 132 lists from 113 games (152,114 unique keywords)

- XML, JSON, CSV
- Compiled Lua, C++
- Encrypted files

Turned each list into a vector of word counts

\[
\text{similarity}(u, v) = \frac{u \cdot v}{\|u\|_2 \|v\|_2}
\]
Hypotheses

Censorship keyword lists are:

1) Determined at the city or provincial level
2) Determined for specific genres of games
3) Related to the date that games are released
4) Largely determined by the publisher or developer
Statistical testing

Mantel test – a test for statistical correlation between similarity matrices $X$ and $Y$

$r$ statistic

a correlation statistic between -1 and 1

$p$ value

probability that at least as extreme correlation would arise from chance
Statistical testing

Mantel test – a test for statistical correlation between similarity matrices $X$ and $Y$

$Y$ is the matrix of cosine similarities

$X$ is different depending on what we want to test

- same genre
- same publisher city
- same developer city
- similarity in approval dates
- same publisher
- same developer
## Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>$r$ statistic</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same publisher city</td>
<td>−0.014</td>
<td>0.65</td>
</tr>
<tr>
<td>Same developer city</td>
<td>−0.0069</td>
<td>0.58</td>
</tr>
<tr>
<td>Same genre</td>
<td>−0.013</td>
<td>0.65</td>
</tr>
<tr>
<td>Similar approval date</td>
<td>0.16</td>
<td>0.0067</td>
</tr>
<tr>
<td>Same publisher</td>
<td>0.15</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Same developer</td>
<td>0.17</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Repeated experiment

Different sampling methodology this time

Many didn’t share the same publisher (50%) or developer (62%) with any other

Selected from five popular publishers

    Giant, Happy Elements, iDreamSky, Netease, Tencent

And from eight popular developers

    CatCap, Chukong, Joymeng, Ourpalm, Smile, Ultralisk, Xiao Ao
Keyword Lists

From 574 unique games, we found

- 167 lists from 129 games
- 171,150 unique keywords

We compared the lists in the same way as before.
## Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>$r$ statistic</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similar approval date</td>
<td>-0.056</td>
<td>0.83</td>
</tr>
<tr>
<td>Same publisher</td>
<td>0.21</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Same developer</td>
<td>0.23</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Hypotheses

Censorship keyword lists are:

✗ Determined at the city or provincial level
✗ Determined for specific genres of games
❓ Related to the date that games are released
✔ Largely determined by the publisher or developer

This suggests that the responsibility of determining what to censor is pushed down as far as possible.
## Content analysis

Sampled 7,000 keywords from 183,111 (1.1% margin with 95% confidence)

<table>
<thead>
<tr>
<th>Theme</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event</td>
<td>Anniversaries, Current Events</td>
</tr>
<tr>
<td>Political</td>
<td>Communist Party of China, Religious Groups</td>
</tr>
<tr>
<td>People</td>
<td>Government officials, Dissidents</td>
</tr>
<tr>
<td>Social</td>
<td>Gambling, Prurient Interests</td>
</tr>
<tr>
<td>Technology</td>
<td>Online Games, URLs</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>No Clear Context</td>
</tr>
</tbody>
</table>
Interesting Keywords

Criticism of Censorship Policies

● 敏感词屏蔽的社会 (a society where sensitive keywords are blocked)

Multilingual Keywords

● 일진회 (Iljinhoe), a nationwide pro-Japan organization that operated in Korea in the 1900s
**Interesting Keywords**

**Coded Language**

刁净瓶 (diāo jìng píng), referencing state leader 习近平 (xí jīnpíng)

无法领奖的人 (a person who is unable to receive the award), referring to China’s Nobel Laureate and dissident Liu Xiaobo

**Competitor Names**

侠客天下 (World of Knights)

仙境传说 (Ragnarok Online)
Future Work

● Explore application of other statistical techniques
● Complete keyword content analysis (manual / machine learning techniques)
● Compare keyword list content across games and industry segments
Acknowledgments

This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. #1314297, #1420716, #1518523, and #1518878. We thank Professor Ron Deibert and Professor Jedidiah Crandall for supervision and guidance. We are also grateful to the anonymous FOCL reviewers for valuable feedback.
Questions?

Keyword data available at

https://github.com/citizenlab/chat-censorship/