Measuring Decentralization of Chinese Keyword Censorship via Mobile Games

Jeffrey Knockel[1,2], Lotus Ruan[1], and Masashi Crete-Nishihata[1]

[1] *Citizen Lab, Munk School of Global Affairs, University of Toronto*
[2] *Dept. of Computer Science, University of New Mexico*

Abstract

China has the world's largest mobile gaming market. Like other technology and Internet companies operating in the country, the gaming industry must follow strict content control policies including submitting lists of blacklisted keywords to regulators. In this paper we provide a first look at how content regulations over the gaming industry are implemented in practice by analyzing over 180,000 unique blacklisted keywords collected across over 200 games from app stores in China.

Internet censorship in China is often presented as a uniformly enforced, top-down system. However, we find content control responsibilities are pushed down to companies resulting in varied implementations. We find that, among the hypotheses we tested, the only consistent predictor of keyword list similarity is whether games have the same publisher and developer, which suggests there is no central state or provincial authority controlling the generation of keyword lists and companies have a degree of flexibility in implementing controls. These results suggest a decentralized and fractured regime of control.

1 Introduction

Internet companies operating in China are required to comply with government content regulations. They are held responsible for content on their platforms and are expected to dedicate resources to ensure compliance with relevant laws and regulations. Failure to do so can lead to fines or revocation of business licenses [26]. This system is a form of intermediary liability or "self-discipline", which allows the government to push responsibility for information control to the private sector [12].

Previous work has provided windows into how this system works. It is well documented that Chinese government authorities issue directives to media and Internet companies on how to handle certain sensitive content [13]. However, analyses of blacklisted keywords collected from chat applications and social media platforms [19, 16, 31, 20] have consistently found limited overlap between keyword lists across companies suggesting that Chinese authorities do not provide a centralized list of keywords to companies. What is the role of companies and government authorities in the development of blacklisted keyword lists?

To further probe this question we analyzed the Chinese mobile gaming industry, which has recently come under increased government pressure [15, 36]. We found a large number of games implement keyword censorship client-side, which provided the opportunity to collect hundreds of keyword blacklists. Facilitated by the large number of keyword lists, we analyzed the similarity between them and assessed four new hypotheses for how the lists are created: (1) Content directives are determined at the city or provincial level and may vary depending on where companies are based; (2) Content directives are determined for specific genres of games; (3) Content directives are related to the date that games are approved by regulators; (4) Companies are under general regulatory pressures, but have a degree of flexibility in determining which specific content to block.

We found that the only consistently significant predictors of similarity between keyword lists were whether games had the same publishers and developers (common developers had the highest correlation) validating hypothesis 4. Our findings show that the system of "self-discipline" in China for content regulation creates a more distributed form of governance than traditional top-down models of authoritarianism.

This paper makes the following contributions:

- We analyze over 250 keyword lists collected from over 200 games together comprising 183,111 unique keywords.
- We present a novel application of a statistical technique and use it to test for correlation between keyword list similarity and other features in games such

as whether they share the same publisher or developer.

- We show that Chinese companies generating keyword lists themselves is the only plausible explanation for similarity between certain keyword lists among all of the hypotheses we tested.

2 Background

With an estimated value of over 27.5 billion US$ in 2017 [29], China represents the largest gaming market in the world. Lucrative as it is, China's market presents unique challenges to companies due to its strict regulatory environment. In 2010 China's Ministry of Culture (MOC) published a regulation [15] listing a number of prohibited topics for online games:

> "violating basic principles set by the Constitution; jeopardizing national unity, state sovereignty and territorial integrity; leaking state secretes, endangering state security or damaging state honor and interests; instigating ethnic hatred or discrimination, jeopardizing ethnic unity, and infringing ethnic rituals or customs; promoting heretical or superstitious idea; spreading rumors, disrupting social order and stability; disseminating obscenity, pornography, gambling, violence or abetting crime; humiliating or slandering others, infringing the lawful rights of others; transgressing social morality; and other contents forbidden by laws and administrative regulations".

The vague definitions of prohibited topics make it unclear how to stay within the line and have been called "pocket crimes" (口袋罪), because anything can fit into them [36]. This regulatory environment can push companies to over-censor, a phenomenon Perry Link described as the "anaconda in the chandelier" [23].

On June 2, 2016, China Audio-video and Digital Publishing Association, a non-governmental organization established in 1994, suggested a list of topics to be filtered on mobile games, including "attacking leaders of the Community Party of China (CPC)", "opposing Maoism", "referring to Taiwan, Hong Kong and Macau as an independent country". No specific keywords are listed in the document [10].

Games released in China require registration approval from MOC and a publication license from China's State Administration of Press, Publication, Radio, Film and Television (SAPPRFT). In May 2016, SAPPRFT extended the requirement to China's fast-growing mobile gaming industry. Games without a license will be removed from app stores. Previously, authorization from SAPPRFT was not mandated for individual games as long as the operators registered the games with the MOC within 30 days of publication [35]. In line with the new regulations, Apple Inc. notified all developers in mainland China that they would need to submit the approval number issued by SAPPRFT to list a game in the iTunes App Store.

Both high-level game content such as the story line and low-level content including all scripts and conversations are subject to content controls [18, 14]. To ensure compliance with China's laws and regulations, gaming companies manage content by enabling keyword filtering systems on their products. Keyword filtering may be enabled to block content in chat features or to prevent players from creating user names and scoreboards containing sensitive keywords. To obtain a publication number for a game, the applicant must submit a list of blocked keywords enabled in the game and administrative accounts for regulators to test and review all scripts and features in the game [34]. In addition to content censorship, since August 1, 2016, Chinese mobile application developers are required to give the Cyberspace Administration of China (CAC) access to basic user information, which must be verified with a user's mobile phone number or real identification.

3 Methodology

Our analysis of keyword censorship in Chinese mobile games consists of two experiments. First, we collected the top games from a popular Chinese app store to evaluate which variables best predict the similarity between any two games' censored keyword lists. In the second experiment, we analyzed games from top Chinese game publishers and developers to further explore how well the "same publisher" and "same developer" variables predict keyword list similarity.

3.1 Analyzing highly downloaded games

To understand which commonalities between games best predict the similarity of those games' keyword lists, we collected a variety of popular Chinese games. First, in December 2016, we acquired games from a Top 20 list of games for November 2016 [5] compiled by Newzoo, a gaming market analytics company. We also acquired games from the earliest month Newzoo had a Top 20 list for, October 2014 [4]. In March 2017, to expand the collection of games, we collected highly downloaded games from Hi Market (安卓市场) [9], a popular [6] app store owned by Baidu [17]. We downloaded games from the site by scraping the first 500 search results from a search query we designed to restrict the results to "Chinese" games with over two million downloads and a rating of 4 to 5 stars. We repeated the process again for what the site termed "English" games, which consisted more broadly of all international games that generally had been adapted to meet Chinese regulations. (Our search queries were not designed to include games with exactly five stars as those tended to be games with very few reviews.) Together these methods produced a total of 836 unique games.

We then analyzed the games. Since there were too many games to reverse engineer each by hand, we first automatically searched for strings in the games to narrow down which warranted closer analysis and reverse engineering effort. We searched for sensitive words including "falun", "法轮" (Falun), "fuck", and "肏" (fuck), and general words that based on experience in our previous work [19, 16, 31, 20] we expected to find in program code implementing censorship specifically, `blacklist`, `censor`, `dirty`, `filter`, `forbid`, `illegal`, `keyword`, `profan`, and `sensitiv`. We then manually reviewed the search results, collecting obvious blacklists, and manually reverse engineering games that appeared to implement censorship, but (*e.g.*) may have encrypted their blacklists. Using this method, we found keyword lists in a variety of formats, including text formats such as plain text, XML, JSON; binary formats such as compiled Lua or C++ code; and in files encrypted with a variety of different algorithms that required reverse engineering to decrypt.

We next tested which commonalities between two games best explained the similarity of those games' keyword lists. The commonalities we tested were whether two games had the same publisher city, same developer city, same genre, same publisher, same developer, or similar approval date. For the publisher and developer city, if a publisher or developer had offices in multiple cities, we used the city that they were primarily headquartered in. For genre and approval date, we used the genre and approval date under which the game was approved by the MOC [7].

To test for correlation between each of these commonalities and keyword list similarity, we used a statistical test called a partial Mantel test [33, 22], an extension of the Mantel Test [27]. A Mantel test is a statistical test for the presence of Pearson correlation between two similarity or distance matrices, a response variable Y and a potential explanatory variable X. The result of this test is the Mantel r statistic. The r statistic is related to the Mantel z statistic,

$$z = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} Y_{ij}$$

except normalized to be between -1 and 1, where the closer r is to -1, the more negative correlation exists, and the closer to 1, the more positive correlation exists (see [22]). It also calculates the test statistic's p value, the probability that at least as extreme of correlation could have occurred by chance, *via* a Monte Carlo test method. A partial Mantel test extends the Mantel test by being able to control for additional matrices $Z_1, Z_2, \ldots$.

To create our response variable Y representing keyword list similarity, we first vectorize each list by creating a vector v with dimension equal to the total # of unique keywords we have seen across all lists and where v_k is the # of times the kth word in our dataset appears in that list, which in our dataset could be greater than one. Then Y_{ij} is the cosine similarity between lists i and j, where the cosine similarity is defined as the cosine of the angle between two vectors, *i.e.*,

$$\text{similarity}(u, v) = \frac{u \cdot v}{\|u\|_2 \|v\|_2}.$$

Depending on the variable that we wish to test for correlation with Y, we define X differently. For all but the test of nearby approval dates, X is a binary matrix where X_{ij} is one if and only if the ith list has a property in common with (*e.g.*, if common publishers is being tested, then has the same publisher as) the jth. For testing nearby approval dates, we first construct X as the matrix such that X_{ij} is the distance between i and j's approval dates as measured in days. We then normalize the matrix by dividing by X's largest value, and then add -1 to each value to turn it into a similarity matrix.

We often observed multiple lists included in each game. Some games used separate lists for censoring different features of the game. Other games seemed to accidentally include an older, outdated version of a list. As the partial Mantel test allows us to control for variables, in each of our tests we control for a binary matrix Z where Z_{ij} is one if and only if the ith list is from the same game than the jth.

3.2 Analyzing games from popular publishers and developers

During April 2017, after observing the results from the last experiment (see Section 4.1), we noticed that many games did not share a publisher or developer with any other game. We decided to perform another experiment focusing on further testing the correlation between games' keyword list similarity and games with the same developers and publishers. In this experiment, we specifically look at the games from five popular publishers and seven popular developers to increase the number of games that share the same publisher or developer.

To determine the list of publishers, we returned to the November 2016 Top 20 list [5] and looked at all publishers from this list who both appeared in the Top

20 list and in whose games we had previously found at least one blacklist. This search resulted in us looking at Giant Interactive Group Inc. (巨人网络科技有限公司), Happy Elements (乐元素游戏), iDreamSky Technology Ltd. (乐逗游戏), NetEase (网易), and Tencent (腾讯).

To find a similar list of developers, we could not immediately use the same Top 20 list as it only listed publishers and not developers. Instead, we used the data from our previous analysis of highly downloaded games to collect the five developers who had the highest number of games containing keyword blacklists in our analysis. Since there was a four-way tie for fourth place, this yielded seven developers: CatCap Studio (达唯科技股份有限公司), Chukong Technologies (触控科技), Joymeng (乐堂动漫), Ourpalm Co. Ltd. (掌趣科技), Smile Games (乐人游戏), Ultralisk (雷兽互动), and Xiao Ao (小奥游戏).

We compiled a list of games for each publisher from the comprehensive list of mobile games for which each publisher had ever successfully obtained approval from the MOC, information available on their website [7]. We compiled a preliminary list of games for each developer from each developer's website and again from the MOC website [7]. However, often it was unclear whether a game on a developer's website or attributed to them by the MOC was published or developed by that company, as the MOC only provides data concerning the publisher and/or operator of a game. To exclude such games, we obtained copyright and ownership information for games using a tool called Tianyancha, a privately-owned platform that consolidates all public information of companies registered with China's State Administration for Industry and Commerce and relevant regulators [8].

After downloading these games for each publisher and developer, we found 341 and 240 games, respectively, and 574 unique games combined. We used the same techniques as before to find lists in these games, and we again performed partial Mantel tests to assess correlation between keyword list similarity and a variety of different commonalities between games (again controlling for lists from the same game). However, in this case, we only tested keyword list similarity with correlation between same publisher, same developer, and nearby approval dates.

4 Results

In this section we describe our results from both of our experiments.

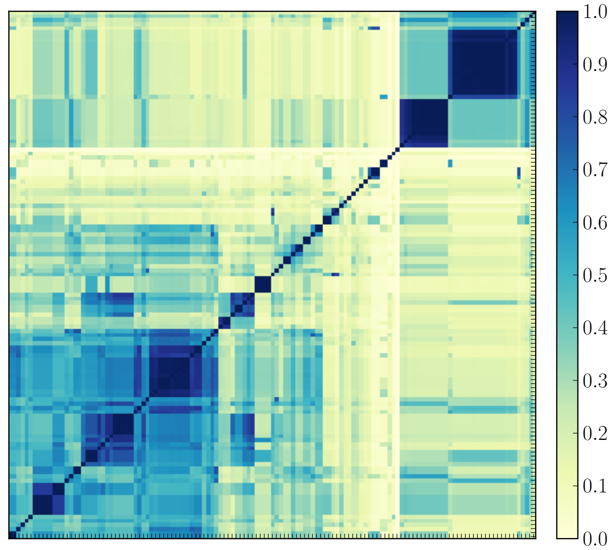


Figure 1: **Cosine similarity of blacklists from highly downloaded games hierarchically clustered according to the Farthest Point Algorithm.**

4.1 Results from analyzing highly downloaded games

From the 836 games that we analyzed in this experiment, we found 132 lists in 113 different games together containing 152,114 unique keywords. (This should not be interpreted to mean that other games had no censorship, as their client-side censorship may not have been discovered by our methods or as they may have performed censorship server-side.) These lists came from a number of popular Chinese games such as 天天酷跑3D (Tiantian Dash), and popular international games adapted to comply with Chinese regulations such as Ski Safari 2 and Candy Crush Saga. Figure 1 shows a heatmap of the pairwise cosine similarity of each blacklist hierarchically clustered according to the Farthest Point Algorithm.

We performed the partial Mantel test as described in Section 3 testing for correlation between keyword list similarity and a number of different commonalities between games. The results of this test are the Mantel r statistic, a measurement of correlation between -1 and 1, and its corresponding p value, the probability that at least as extreme of correlation could have occurred by chance. Three game commonalities tested did not show significant correlation: **publisher city**, -0.014 ($p = 0.65$); **developer city**, -0.0069 ($p = 0.58$); and **genre**, -0.013 ($p = 0.65$). Three game commonalities tested did show significant correlation: **approval date**, 0.16 ($p = 0.0067$); **publisher**, 0.15 ($p < 0.001$); and **developer**, 0.17 ($p < 0.001$).

Since the game commonalities tested may be correlated among themselves (*e.g.*, games with the same de-

Publisher	**Downloaded**	**Lists Found**
Giant	26	11
Happy Elements	11	3
iDreamSky	29	8
Netease	87	22
Tencent	188	34

Table 1: **The # of games downloaded and lists found for each publisher.**

Developer	**Downloaded**	**Lists Found**
CatCap	23	7
Chukong	38	10
Joymeng	63	9
Ourpalm	38	11
Smile	19	4
Ultralisk	21	16
Xiao Ao	38	32

Table 2: **The # of games downloaded and lists found for each developer.**

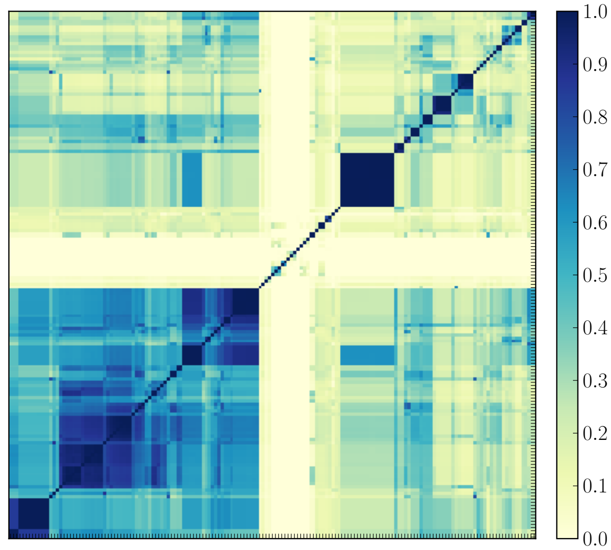


Figure 2: **Cosine similarity of blacklists from games from top publishers and developers hierarchically clustered according to the Farthest Point Algorithm.**

veloper tend to have the same publisher), we repeated each of the three tests two times for approval date, developer, and publisher, controlling for first the one and then the other of whichever of the two we were not presently testing. The greatest changes in results happened when testing same developers controlling for same publishers, where the r statistic reduced to 0.095 ($p < 0.001$), and same publishers controlling for same developers, where the r statistic reduced to only 0.047 ($p = 0.0015$). This suggests that, after controlling for each other as confounding variables, similar developers better predicts keyword similarity than similar publishers.

Because many of the games did not share the same publisher (50%) or developer (62%) with any of the other games we found lists from, and many others shared very few of each, we decided to perform a second experiment looking at games from popular publishers and developers in order to increase the number of games that share the same publisher or developer and in order to generally confirm our results.

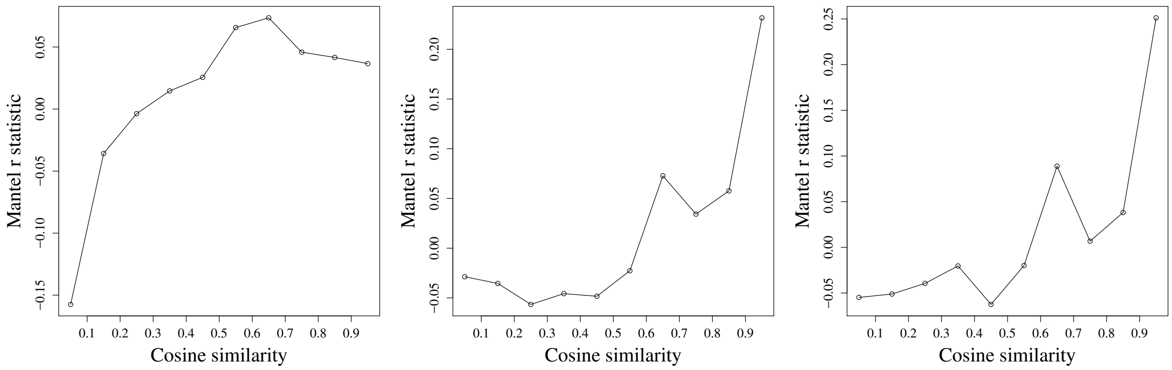
4.2 Results from analyzing games from popular publishers and developers

From the 574 unique games that we analyzed in this experiment, we found 167 lists in 129 different games together containing 171,150 unique keywords. This list again includes popular Chinese (*e.g.*, 开心消消乐 or Anipop) and international (*e.g.*, Fruit Ninja and Temple Run 2) games. See Tables 1 and 2 for a breakdown of lists found from each publisher and developer. Figure 2 shows a heatmap of the pairwise cosine similarity of each blacklist hierarchically clustered according to the Farthest Point Algorithm.

Our partial Mantel test results for the tested game commonalities are as follows: **approval date**, -0.056 ($p = 0.83$); **publisher**, 0.21 ($p < 0.001$); **developer**, 0.23 ($p < 0.001$).

Compared to the previous experiment, we saw even more correlation when testing same publishers and same developers (see Figures 3b and 3c for corresponding correlograms); however, we were surprised to see the correlation that we had originally seen in the first experiment when testing similar approval dates (see Figure 3a) disappear in the second, as we expected to see some correlation due to (*e.g.*) developers adding to lists over time. The second experiment's result may be because the experiment was specifically designed to further test correlation in games with the same publishers and developers by decreasing publisher and developer variety. However, by focusing on these variables and decreasing variety, we may have limited our ability to observe the effect of other explanatory variables.

We also again tested same publisher controlling for same developer and same developer controlling for same publisher. The same publisher controlling for same developer reduced the correlation to 0.064 ($p = 0.015$). The same developer controlling for same publisher only reduced the correlation to 0.13 ($p < 0.001$). This indicates that, after controlling for each other as confounding variables, games having the same developers better predicts keyword list similarity than similar publishers.



(a) **Similar approval dates** (b) **Common publisher** (c) **Common developer**

Figure 3: **Shape of correlation: selected partial Mantel correlograms for three different game commonalities showing nonlinear shape; (a) is from the first experiment; (b) and (c) are from the second experiment**

5 Keyword Analysis

In this section we discuss the provenance of keyword lists and provide a preliminary content analysis.

5.1 Keyword list provenance

Our results suggest that developers, and, to a lesser extent, publishers are largely responsible for providing the keyword lists that come with their games; however, while our results do not speak directly to how they accumulate these lists, there are clues in our keyword data set that point to possible directions. Text-based data formats typically have escape sequences used to escape otherwise special characters that are used to structure the data. While we took care to unescape keywords from whatever file format it was stored in, some developers did not always do so, leaving the escape sequences in and providing clues as to where the lists were taken from. For instance, we saw C-style escapes in other file formats, *e.g.*, 私\\服 which we found in an XML document with a C-style escape of the \ by preceding it with a second \ (the keyword appears as 私\服 in many lists). Conversely, we saw XML escapes in keywords in non-XML files such as the ampersand in 冠西&艳照 having been replaced with & resulting in 冠西&艳照.

Some keywords may even trace back to originally being shared on old web apps. For instance, old PHP web apps escaped database input by manually calling the *addslashes* [1] function or by enabling “magic quotes” [2] which called it automatically. However, this functionality has long been deprecated due to it being insecure and not multibyte character encoding aware [32]. In 2004, when a leaked keyword list had been published on a bulletin board [11], the published list contained many erroneous \ escapes due the bulletin board’s improper escaping. For instance, the list features 胡錦\濤, since the second character (錦) in the Chinese GBK encoding encodes to the bytes `0xe5` and `0x5c`, the latter byte being \ in ASCII, which the web app erroneously interpreted as a byte needing to be escaped and preceded it with another \. As keywords such as and including this one are featured in many of our lists, this suggests that they and potentially many others had been at one time been shared on old web apps that had improperly escaped database input.

Theme	Examples
Event	Anniversaries, Current Events
Political	Communist Party of China, Religious Groups
People	Government officials, Dissidents
Social	Gambling, Prurient Interests
Technology	Online Games, URLs
Misc	No Clear Context

Table 3: **Content themes and related categories**

5.2 Content Analysis

Previous research [19, 16, 31, 20] performed content analysis of keyword lists by manually grouping keywords into content categories based on contextual information. Using similar methods as previous work, we conducted a preliminary content analysis to provide a high level description of the data set. We analyzed a random sample of 7,000 keywords from our data set of 183,111 keywords. A native Chinese speaker reviewed the random sample and based on the context of the keywords assigned high level themes according to a code book developed in [16]. See Table 3 for a description of each theme and Figure 4 for the distribution of keywords by theme. We describe and provide examples of each theme in the remainder of this section.

Social: This theme accounted for the largest percent-

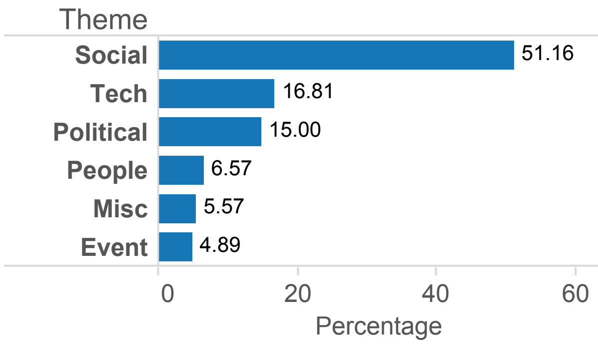


Figure 4: **Breakdown by theme of 7,000 randomly sampled keywords** (±1.1 **error at 95% confidence)**

age of the keywords we analyzed. Examples include references to illicit goods such as "出售业主信息数据" (selling data information of property owners) and gambling (*e.g.*, "六合彩" Mark Six, a lottery betting system organized by Hong Kong Jockey Club). A similar focus on Social theme keywords was found in past work on live streaming platforms [20].

Political: This theme includes general references to the CPC and government bodies (*e.g.*, "人民检察院" People's Procuratorate, the agency responsible for both prosecution and investigation under China's legal system), criticism of state policy (*e.g.*, "敏感词屏蔽的社会", meaning "a society where sensitive keywords are blocked"), religion (*e.g.*, Falun Gong and Christianity), Sino-Japanese relations, and ethnic groups in China. Homoglyphs or homonyms were often used when referring to the CPC (共产党) (*e.g.*, "哄铲挡", hǒng chǎn dǎng) and Falun Gong (*e.g.*, "發圇", pronounced fā lún, as opposed to 法轮功, fǎ lún gōng).

We also found keywords in Korean and Japanese related to international relations issues in Asia. Examples include "일진회", which refers to Iljinhoe, a nationwide pro-Japan organization that operated in Korea in the 1900s. The presence of these keywords may be because some games we analyzed were imported from foreign gaming companies or were aimed at foreign markets. It may also have to do with the unique characteristics of gaming platforms where in-group identity and nationalism is often shared and championed by players [3, 30].

People: Similar to content found in other keyword lists [20, 19, 24], we found references to people including government officials, relatives of officials, dissidents, and names without clear context. Keywords included use of homonyms, homoglyphs, coded messages, and pinyin to refer to party leaders and dissidents. For example, "刁净瓶" contains a homoglyph (刁) and a homonym (净瓶, jìngpíng) for President Xi Jinping (习近平, xí jìnpíng). In another example a coded reference is made to China's first Nobel Peace Prize winner Liu Xiaobo who received the award in absentia due to imprisonment ("无法领奖的人", a person who is unable to receive the award).

Event: Past work has shown sensitive and critical events can act as catalysts for censorship of social media in China [16, 31, 20]. We tested the prevalence of event-related keywords by searching for terms related to current events that occurred within 2016 to 2017 and that had been found censored on Chinese chat applications [24, 28, 25] and microblogs [25] or that were known to have related government censorship directives [13]. We did not find references to these up-to-date news events in the gaming keyword data set. However, we did find keywords referring to other politically sensitive events (*e.g.*, "1989年民运", or 1989 Year Democracy Movement, a reference to the 1989 Tiananmen Square movement; "茉莉花活动", and the Jasmine Revolution, a reference to a series of pro-democracy protests in China inspired by the Jasmine Revolution in Tunisia in 2011).

Chat apps and microblogs provide means to communicate and spread content, whereas games are primarily for entertainment. Therefore, there may be increased concern and scrutiny of how apps used for news consumption manage content related to current events. It is also possible there are references to events we did not find due to the limitation of the scope of our current project. In our analysis we did not analyze historical versions of each mobile game, and we did not track any downloaded updates, if there were any, to its keyword lists.

Technology: This theme includes references to identifiers such as phone numbers and emails as well as names of websites and various technology services. Keywords also included references to the names of competing game companies. We suspect the inclusion of these keywords were driven by commercial interests to prevent users from criticizing a platform or being lured away to its competitors, a pattern that was seen in previous work [20].

6 Future Work

Mantel and partial Mantel tests are useful for testing the significance of a single similarity- or distance-based explanatory variable at a time. One avenue for future work is to explore other statistical methods such as distance-based redundancy analysis [21] that can model and compare the effects of multiple distance-based explanatory variables simultaneously.

Another avenue for future work is a complete keyword content analysis through manual or machine learning methods. A high level semantic analysis would permit the comparison of lists thematically and further reveal implications of game censorship in China.

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 FOCI reviewers for valuable feedback.

References

[1] Php: addslashes. Available at https://secure.php.net/manual/en/function.addslashes.php.

[2] Php: What are magic quotes. Available at https://secure.php.net/manual/en/security.magicquotes.what.php.

[3] Racial and ethnic hate speech thrives in online games. Available at http://www.japantimes.co.jp/news/2017/05/14/business/tech/racial-ethnic-hate-speech-thrives-online-games/.

[4] Top 20 android games in china: Time & money spend in a stable october. Available at https://newzoo.com/insights/articles/top-20-android-games-in-china-time-is-money-stable-october/.

[5] Top android games china. Available at https://newzoo.com/insights/rankings/top-20-android-games-china/.

[6] Top chinese android app stores. Available at https://newzoo.com/insights/rankings/top-10-android-app-stores-china/.

[7] 中国文化市场网. Available at http://ccm.gov.cn/.

[8] 关于我们. Available at http://www.tianyancha.com/property/1/.

[9] 安卓网-中国最大android手机垂直门户. Available at http://www.hiapk.com/.

[10] 移动游戏内容规范(2016年版). Available at http://www.ce.cn/culture/gd/201606/02/t20160602_12463013.shtml.

[11] Qq过滤词列表 zt. Available at https://web.archive.org/web/20040908030852/http://bbs.omnitalk.org/arts/messages/3824.html, 2004.

[12] Are China's demands for Internet 'self-discipline' spreading to the West? Available at http://www.mcclatchydc.com/opinion/article24570625.html, 2010.

[13] Directives from the Ministry of Truth. Available at http://chinadigitaltimes.net/china/directives-from-the-ministry-of-truth/, 2017.

[14] Ahmad, D. A look at china's stringent mobile game regulations half a year on. *Pocket Gamer* (Mar 2017).

[15] China's Ministry of Culture. Interim measures for the administration of online games, Jun 2010.

[16] Crandall, J., Crete-Nishihata, M., Knockel, J., McKune, S., Senft, A., Tseng, D., and Wiseman, G. Chat program censorship and surveillance in China: Tracking TOM-Skype and Sina UC. *First Monday 18*, 7 (2013).

[17] Custer, C. 5 reasons why baidu spent nearly $2 billion to acquire 91 wireless. Available at https://www.techinasia.com/5-reasons-baidu-spent-2-billion-acquire-91-wireless.

[18] Hancock, T. China game developers battle crackdown on content. *Financial Times* (Mar 2017).

[19] Knockel, J., Crandall, J. R., and Saia, J. Three researchers, five conjectures: An empirical analysis of tom-skype censorship and surveillance. In *FOCI 2011: Proceedings of the USENIX Workshop on Free and Open Communications on the Internet* (2011).

[20] Knockel, J., Crete-Nishihata, M., Ng, J. Q., Senft, A., and Crandall, J. R. Every Rose Has Its Thorn: Censorship and Surveillance on Social Video Platforms in China. In *5th USENIX Workshop on Free and Open Communications on the Internet* (2015).

[21] Legendre, P., and Anderson, M. J. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological monographs 69*, 1 (1999), 1–24.

[22] Legendre, P., and Legendre, L. F. *Numerical ecology*, vol. 24. Elsevier, 2012.

[23] Link, P. China: The Anaconda in the Chandelier. *The New York Review of Books* (2002).

[24] Lotus Ruan, Jeffrey Knockel, M. C.-N. One App, Two Systems: How WeChat uses one censorship policy in China and another internationally. Available at https://citizenlab.org/2016/11/wechat-china-censorship-one-app-two-systems/, 2016.

[25] Lotus Ruan, Jeffrey Knockel, M. C.-N. We (can't) Chat: "709 Crackdown" Discussions

Blocked on Weibo and WeChat. Available at https://citizenlab.org/2017/04/we-cant-chat-709-crackdown-discussions-blocked-on-weibo-and-wechat/, 2017.

[26] MacKinnon, R. China's Censorship 2.0: How companies censor bloggers. *First Monday; Volume 14, Number 2 - 2 February 2009* (2009).

[27] Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer research 27*, 2 Part 1 (1967), 209–220.

[28] Masashi Crete-Nishihata, Jeffrey Knockel, L. R. Tibetans blocked from Kalachakra at borders and on WeChat. Available at https://citizenlab.org/2017/01/tibetans-blocked-from-kalachakra-at-borders-and-on-wechat/, 2017.

[29] McDonald, E. The global games market will reach 108.9 billion usd in 2017. Available at https://newzoo.com/insights/articles/the-global-games-market-will-reach-108-9-billion-in-2017-with-mobile-taking-42/, Apr 2017.

[30] Nie, H. A. Gaming, Nationalism, and Ideological, Work in Contemporary China: online games based on the War of Resistance against Japan. *Journal of Contemporary China* (2013).

[31] Ruan, L., Knockel, J., Ng, J. Q., and Crete-Nishihata, M. One app, two systems: How wechat uses one censorship policy in china and another internationally. Available at https://citizenlab.org/2016/11/wechat-china-censorship-one-app-two-systems/.

[32] Shiflett, C. addslashes() versus mysql_real_escape_string(). Available at http://shiflett.org/blog/2006/jan/addslashes-versus-mysql-real-escape-string, 2006.

[33] Smouse, P. E., Long, J. C., and Sokal, R. R. Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Systematic zoology 35*, 4 (1986), 627–632.

[34] State Administration of Press, Publication, R. F., and of The People's Republic of China, T. Review and approval of online game publication.

[35] Tang, T. China sapprft issued implementation rules to streamline the approval process of mobile games, Jun 2016.

[36] 张训. 口袋罪视域下的寻衅滋事罪研究. 政治与法律, 3 (2013).