

# What do Chinese-language microblog users do with Baidu Baike and Chinese Wikipedia? A case study of information engagement

Han-Teng Liao  
Oxford Internet Institute  
University of Oxford  
Oxford, United Kingdom  
hanteng@gmail.com

## ABSTRACT

This paper presents a case study of information engagement based on microblog posts gathered from Sina Weibo and Twitter that mentioned the two major Chinese-language user-generated encyclopaedias. The content analysis shows that microblog users not only engaged in public discussions by using and citing both encyclopaedias, but also shared their perceptions and experiences more generally with various online platforms and China's filtering/censorship regime to which user-generated content and activities are subjected. This exploratory study thus raises several research and practice questions on the links between public discussions and information engagement on user-generated platforms.

## Categories and Subject Descriptors

[Human-centered computing]: Collaborative and social computing systems and tools—*Blogs, Wikis, Empirical studies in collaborative and social computing*

## Keywords

Censorship, Chinese Internet, Information use, Information engagement, Internet filtering.

## 1. INTRODUCTION

The microblogging services Sina Weibo and Tencent Weibo were launched in 2009 and 2010 respectively and have contributed to the rapid growth of online advertising markets in China [5, 11]. The term “Weibo” (微博) is the transliteration of the term “microblog” in Mandarin Chinese. Weibo allow users to publish short messages like tweets with a 140-character limit per post. Like other major companies in China and elsewhere whose business is user-generated content (hereafter UGC), the amount of content and traffic a platform can garner is crucial for its survival.

Microblogs can be seen as part of the historical development of the rise of user-generated content since 2005[18]. A recent survey has shown that user-generated content is more popular in Asia

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

*OpenSym '14*, August 27 - 29 2014, Berlin, Germany  
Copyright is held by the owner/author(s). Publication rights licensed to ACM.  
ACM 978-1-4503-3016-9/14/08\$15.00.  
<http://dx.doi.org/10.1145/2641580.2641611>

than in the West [3]. By the first half year of 2010, the content produced by amateur Chinese Internet users surpassed that produced by professional websites [25]. According to a survey report based on 2012 data, around 66.1% of mainland Chinese Internet users used a blog/personal space, 54.7% used microblogs, and 48.8% used social networking websites [5]. One US market report found that 47% of Chinese broadband users had contributed to UGC in China, including online review sites, forums, blogs, etc., and that this UGC content had influenced 58% of purchase decisions in China. Both numbers are much higher than the corresponding statistics in the US [23].

The growth of UGC in mainland China may have contributed to the creation of a so-called “public opinion monitoring sector” (輿情監測服務市場), where managers (mostly from the public sector) buy up-to-date information about the latest online public opinion in mainland China. The online public opinion data relies heavily on processing and mining UGC, especially from microblogs. One of the most successful example is the Online Public Opinion Monitoring and Measuring Unit of the People's Net (人民網輿情監測室), which alone has an annual revenue of close to 200 million RMB [20, 28], or about £20 million. It is noted that the Chinese government and corporations have been developing and using “Internet public opinion monitoring systems,” which involve intensive collection and analysis of web data [20, 28]. As of 2014, there is little Internet research in English about this development of public opinion monitoring, except for a panel discussion [14], which I organized during the 2013 Chinese Internet Research Conference. Further work will be needed to compare the political and commercial uses of microblog data in China and elsewhere.

As expected, microblog platforms that are hosted in mainland China are subject to Beijing's filtering and censorship regime. Several recent research papers have reverse-engineered the censorship mechanism to understand how censorship works on Weibo platforms. Analysing sensitive terms on Sina Weibo and Twitter, researchers found initial evidence that content is filtered or censored on Sina Weibo[1]. Researchers at Hong Kong University also conducted censorship research on Weibo [8, 9]. Another group of researchers found that individual criticisms of the government are permitted but that collective expression (which may lead to protests) is censored [12]. Yet another study tracked and quantified sensitive topics on Weibo and found that the posts were often short-lived and confined to a small core of users who posted sensitive content, suggesting that multiple layers of filtering produced effective censorship to contain the spread of such messages [37].

According to an industry report, in 2012, about 88% of Internet users in mainland China are Sina Weibo users [7]. Weibo

platforms are thus important sites of observation to see how user-generated encyclopaedias like Chinese Wikipedia, Baidu Baike and Hudong Baike[18] are mentioned and used. (Baike means “encyclopedia” in Chinese. Baidu is the name of China’s major search engine company. Hudong means “interactive” or “interaction” in Chinese.) First, the microblog posts are likely to contain the views and experiences of Chinese Internet users with the encyclopaedias. Second, since microblog platforms are also user-generated content platforms, it is possible that the users of microblogs are more likely to use user-generated encyclopaedias. Third, researchers can observe how Beijing’s censorship influences the use or perception of user-generated encyclopaedias. Especially in the case of Sina Weibo, which reaches a large enough number of mainland Chinese users, the posts can provide important indications about how mainland Chinese users are receptive to the two online encyclopaedias. Fourth, users are likely to use microblogs (both reading and writing, which can be regarded as one amongst several online information engagement practices [2]) in conjunction with search engines and user-generated encyclopaedias. Researchers can thus observe how users use different platforms in combination.

The concept of “information engagement” is used in two disciplines with different audience and research concerns. In the context of civic engagement including youth engagement, the focus is to engage citizens for more and better political participation, as proposed by W. Lance Bennett and Chris Wells to build “civic learning repertoires” for “information engagement” [2]. In the context of warfare, including information warfare, the practical concern is to provide actionable guidelines to inform foreign and domestic audiences [31], which has been discussed as part of the redefinition of information warfare [24]. In the Chinese context, the Internet is central for civic engagement [6, 36] and the state’s cyberwarfare doctrine [10], and the study of the use and reception of user-generated encyclopaedias expressed on microblogs can be conducted as a case study of information engagement in both senses.

This paper aims to explore what Chinese-language microblog users do with major user-generated encyclopedias such as Baidu Baike (Baidu’s encyclopedia website) and Chinese Wikipedia[18] by analysing the content of their posts. As user-generated encyclopaedias are expected to solve information problems for users, microblog content about users’ experiences and perceptions should indicate the status (and thus also the gaps) of information engagement experiences with the two major online Chinese encyclopaedias as expressed in a major type of platform that constitute a major part of Chinese public opinion.

## 2. METHODS

Computer-assisted quantitative methods (including several Chinese natural language processing and text-mining techniques) are first used to identify patterns and posts relevant to the research question, and then qualitative methods are used to interpret the findings. Similar to how Zhu, Phipps, Pridgen, Crandall, & Wallach [37] tracked and quantified censorship by selecting topics for research, I developed a means of filtering relevant posts that mention Baidu Baike or Chinese Wikipedia, which produced a dataset that combines data from Sina Weibo and Twitter.

This exploratory method mimics the tools that are likely used for public opinion monitoring. Relevant posts are first screened from designated online platforms using keywords based on the analyst’s judgement. Although there are no guidelines for how

many and which terms yield sufficiently comprehensive results, it can be expected a few iterations will yield certain heuristics for gauging the results. As an exploratory study, the combination of qualitative and quantitative methods is bound to be refined in the future as the analysis of microblogs advances. For example, one must consider the issues of sampling when claiming the observations on microblog platforms are valid public opinion data, which is outside the scope of this exploratory study. Nevertheless, this research should provide indications for future research and practices on monitoring and analysing microblog content. I now examine how the Chinese term ‘wiki’ in a number of variants, which are mentioned in microblog posts.

### 2.1 Data selection and data sets

For data collection, the Chinese term (both simplified and traditional characters) “wiki” was used instead of the more specific term “Wikipedia”, in addition to specific Chinese terms such as “Baidu Baike” and “Chinese Wikipedia” (see Table 1).

**Table 1. Wikipedia-related and Wiki-similar Chinese terms**

	Simplified Chinese	Traditional Chinese	Notes
Baidu Baike	百度百科	百度百科	unambiguous term
Chinese Wikipedia	维基百科	维基百科	unambiguous term
Hudong Baike	互动百科	互動百科	unambiguous term
wiki	维基	維基	ambiguous term
wikileak	维基解密	維基解密	
wikileak	维基揭密	維基揭密	alternative name
wikileak	维基泄密	維基泄密	alternative name
"Wong Wiki"	王维基	王維基	Ricky Wong, unrelated to Wiki

Two unrelated entities were expected to yield false positives since they contain exactly the same characters of popular Chinese term of “wiki”: One is “WikiLeaks” with its variant Chinese terms and another is a Hong Kong businessperson Ricky Wong, whose first name happens to share the exact same Chinese characters as “wiki” (see the bottom row in Table 1). The keyword filter used here can also introduce other false positive outcomes since the Chinese term “wiki” can also refer to general wiki technologies and practices, yet in this case the resulting dataset can be used to see how dominant Wikipedia or wiki-implemented encyclopaedias are on the subject of wiki.

Based on the keyword list, two datasets were retrieved from two data intermediaries. The WeiboScope dataset, provided by Dr. Fu at the Hong Kong University, randomly sampled results from Sina Weibo between January 2011 and April 2012 (see Table 2 for details). Second, the DiscoverText dataset contains publicly available microblog posts from Sina Weibo and Twitter for a time period of 23 days in 2012. Table 2 summarizes the number (N) and proportion (percentage) of posts from the two platforms and the two intermediary sources..

**Table 2 Posts collected: including false positives**

Microblog platform	WeiboScope		DiscoverText		Total	
	N	%	N	%	N	%
Twitter	0	0%	11640	63%	11640	19%
Weibo	42992	100%	6865	37%	49857	81%
<b>Total</b>	42992		18505		61497	
<b>Timestamps</b>						
Earliest	03 Jan 2011 00:21		20 Mar 2012 18:08		03 Jan 2011 00:21	
Latest	05 Apr 2012 13:05		12 Apr 2012 12:09		12 Apr 2012 12:09	
<b>Time Duration</b>	458 days		23 days		465 days	

The first step then was to remove the false positive data (unrelated to the research topic) included in Table 2, and then also to remove another set of “jamming” posts targeted at Wen Yunchao, resulting in a reduced dataset as summarized in Table 3. Wen Yunchao is a mainland Chinese activist who launched a series of online campaigns against Internet censorship in China, through writing messages on blogs and other social media platforms.

**Table 3 Posts collected: false positives removed**

Microblog platform	WeiboScope		DiscoverText		Total	
	N	%	N	%	N	%
Twitter	0	0%	1152	15%	1152	3%
Weibo	37185	100%	6330	85%	43515	97%
<b>Total</b>	37185		7482		44667	
<b>Timestamps</b>						
Earliest	03 Jan 2011 00:21		20 Mar 2012 18:08		03 Jan 2011 00:21	
Latest	05 Apr 2012 13:05		12 Apr 2012 12:09		12 Apr 2012 12:09	
<b>Time Duration</b>						
	458 days		23 days		465 days	

The removal of posts targeted at Wen Yunchao resulted in a notable reduction (more than 10,000 posts) in the Twitter dataset; so, it is worth explaining why these were removed. I examined these posts and noticed that almost all of them followed the format: “@wenyunchao Some Entry or Web Page Title – Baidu Baike or Wikipedia” (for example, “@wenyunchao 李明博 - 维基百科” and “@wenyunchao 山海关\_百度百科”). These posts were often written by users with ID names that are likely to be randomly generated by machines (e.g. 5sx63cyd, bd8wwkp). All these machine-generated posts were posted to Twitter within the 23 days covered by DiscoverText as shown in the first row of Table 4. These posts mostly cited Baidu Baike, but a few posts cited Chinese Wikipedia. The two posts containing “wenyunchao” from the WeiboScope dataset, in contrast, were sent by human users.

**Table 4. Posts that mention wenyunchao (the set "WYC")**

Microblog platform	Dataset	BB	CW	HD	"wiki"	wikileaks	"Wong"
Twitter	DiscoverText	8755	1426	0	1524	10	4
Weibo	HKU	0	0	0	2	2	0
<b>Total</b>		8755	1426	0	1526	12	4

As these posts would skew any content analysis, I decide not to include them, but rather to consider how the cyber-attacks against Wen Yunchao relate to the research here.

The remaining 44667 posts are then processed for Chinese-language word-tokenization, a necessary procedure to analyse Chinese-language texts because, unlike languages such as English, Chinese is written without space.

## 2.2 From overview to detailed analysis.

I first ran a word frequency analysis to obtain an overview about hot topics and frequently associated words. Then I compared the posts that mention the two encyclopaedias individually. In the process, I grouped and differentiated microblog posts using computer-assisted techniques (including Latent Semantic Analysis) so as to break down the large number of posts into manageable clusters. Originally developed for word-document relationship analysis in the field of natural language processing and information retrieval, these techniques proved suitable for the following reasons. First, they can process a large amount of documents, assisting researchers in clustering and differentiating documents and words for further analysis. Second, they mimic or even simulate the way the online texts are processed by search

engines and social media analysts (including China’s online public opinion monitors). Third, techniques such as Latent Semantic Analysis has been applied to Chinese-language texts with noticeable success [4, 34], solving effectively the matching problems caused by synonymy or polysemy [35]. Finally, by repurposing the quantitative techniques for cultural research, researchers can ground their analyses not only on specific relevant words, but also on the latent semantic relationships amongst them that are not immediately apparent to casual readers.

For this study, I implemented such analysis using a combination of NLTK (Natural Language Toolkit), scikit-learn (a set of machine learning tools) and gensim (“generate similar” documents).

After the posts were clustered, I categorized the relevant posts under the two main themes (filtering/censorship regime and information engagement) to elicit how the two encyclopaedias were discussed by microblog users.

## 3. RESULTS

The findings are now presented, proceeding from overall patterns to specific individual posts. Table 5 lists the most frequently occurring words. The most mentioned term is “encyclopaedia” (百科), after which the Chinese names of the three encyclopaedias follow in this order: Baidu Baike, Wikipedia and Hudong Baike. This paper focuses on the posts mentioning the first two of these.

**Table 5. Most frequently-occurring words**

Ranking	Word	Counts	Ranking	Word	Counts	Ranking	Word	Counts
1	Encyclopedia 百科	39951	31	Company 公司	1601	61	Sprinkle over 撒过	1346
2	Baidu 百度	22672	32	Harm 危害	1600	62	Test 检验	1344
3	Wiki 维基	21495	33	(English) wikipedia	1570	63	Initiation 引发	1342
4	Baidu Baike 百度百科	18623	34	Addition 增加	1569	64	Kai-fu Lee 李开复	1338
5	Wikipedia 维基百科	10780	35	Ha ha ha 哈哈	1561	65	Nerve 神经	1331
6	Interactive 互动	6948	36	Cause 导致	1536	66	Times 万倍	1324
7	Hudong Baike 互动百科	6583	37	Reference 参考	1522	67	Notes 笔记	1312
8	China 中国	4428	38	Life 生命	1518	68	Jokes 恶搞	1308
9	Floating Spirit 浮灵	3954	39	The book 全书	1513	69	100,000 times 十万倍	1307
10	Share 分享	3291	40	One kind 一种	1495	70	Bad 不良	1290
11	Reply 回复	3023	41	Not work 不行	1474	71	Necessary 就要	1286
12	Activity 活动	2732	42	Network 网络	1453	72	Search 搜索	1283
13	Excessive 超标	2682	43	Occur 发生	1444	73	Scientific name 学名	1268
14	Wallet 钱包	2608	44	Community 社会	1427	74	Serious harm 严重危害	1265
15	Wiki 维基	2310	45	In vivo 体内	1422	75	Ridicule 调侃	1264
16	Microblogging 微博	2290	46	Similar 类似	1421	76	Notebook 笔记本	1263
17	United States 美国	2239	47	University 大学	1418	77	Syndrome 综合症	1247
18	Recommend 推荐	2222	48	Knowledge 知识	1411	78	Sitting 坐姿	1241
19	Discovery 发现	2217	49	Oxygenated 含氧	1400	79	(English) syndrome	1235
20	World 世界	2181	50	Oxygen 含氧量	1399	80	(English) piriformis	1232
21	Introduction 介绍	2130	51	Temporarily 暂时	1399	81	Ischial 坐骨	1225
22	Website 网站	2008	52	Oxygen 氧量	1399	82	Intelligence 智力	1225
23	Forwarding 转发	2008	53	Canadian fish 加鱼	1395	83	Sciatic 坐骨神经	1222
24	Free 自由	1938	54	Originally 原本	1380	84	Country 国家	1211
25	Health 健康	1795	55	Complex 综合	1367	85	History 历史	1208
26	First 第一	1752	56	Time 时间	1366	86	Information 信息	1197
27	Shanghai 上海	1714	57	Soon 眼看	1365	87	Beijing 北京	1193
28	Click 点击	1701	58	Encyclopedia 百科全书	1353	88	Face 正视	1191
29	Entry 词条	1669	59	Hypoxia 缺氧	1352	89	Because fish 因鱼	1181
30	Call 称为	1642	60	Extend 延长	1350	90	Really 真的	1177

### 3.1 When mentioning either

Table 6 provides an overview for each of the two portions of posts mentioning the two encyclopaedias, and ranks the most frequently co-occurring words. Instead of commenting on the two encyclopaedias per se, many posts use or cite the encyclopaedias to comment on certain popular incidents. The left-hand side of the table shows the predominance of the Yúfúlíng (鱼浮灵) incident for posts mentioning Baidu Baike, whereas the right-hand side shows the predominance of the ‘Wallet’ incident for posts mentioning Chinese Wikipedia. The findings mainly demonstrate the topicality of microblogs in responding to the latest events. Baidu Baike was mentioned because of the background information it provided concerning an online rumour that a

chemical used to revive dying fish called Yúfúling is carcinogenic. Chinese Wikipedia was referred to for its content on piriformis syndrome, a topic around which the ex-Google China Chief Lee Kaifu was ridiculed by many and defended by some for recommending that people not sit with their wallets in their rear pockets. To gauge their impact of these events on the datasets, I used clustering and keyword search independently, and found about 1,400 posts for the Yúfúling incident and about 1,200 posts for the ‘Wallet’ incident. I also identified the words predominantly associated with the incidents (see the “Memo” columns in Table 6), which included ‘Shanghai.’ The findings highlight the potential of using how online encyclopaedias are referenced and linked in topical microblog public discussions when authoritative information is needed.

**Table 6. Most frequently-occurring words when mentioning...**

Baidu Baike				Chinese Wikipedia			
Word	Counts	Memo	Word	Counts	Memo		
1 Floating Spirit 浮灵	39951	Yúfúling	Wallet 钱包	2600	Wallet		
2 Excessive 超标	22672	Yúfúling	Introduction 介绍	1451	Wallet		
3 Baidu 百度	21495		Reference 参考	1317	Wallet		
4 Share 分享	18623		Call 称为	1307	Wallet		
5 China 中国	10780		Kai-fu Lee 李开复	1297	Wallet		
6 Discovery 发现	6948	Yúfúling	Health 健康	1293	Wallet		
7 Addition 增加	6583	Yúfúling	Similar 类似	1292	Wallet		
8 Life 生命	4428	Yúfúling	Kuso 恶搞	1258	Wallet		
9 In vivo 体内	3954	Yúfúling	Cause 引发	1251	Wallet		
10 Oxygen level 含氧量	3291	Yúfúling	Scientific name 学名	1239	Wallet		
11 Canadian fish 加鱼	2732	Yúfúling	Notebook 笔记本	1238	Wallet		
12 Temporarily 暂时	2682	Yúfúling	Sitting 坐姿	1237	Wallet		
13 Hypoxia 缺氧	2608	Yúfúling	piriformis piriformis	1232	Wallet		
14 Extend 延长	2321	Yúfúling	syndrome syndrome	1228	Wallet		
15 Sprinkle over 撒过	2310	Yúfúling	Ridicule 调侃	1224	Wallet		
16 Originally 原本	2290	Yúfúling	Sciatic 坐骨神经	1222	Wallet		
17 Test 检验	2239	Yúfúling	Syndrome 综合症	1217	Wallet		
18 Shanghai 上海	2222	Yúfúling	Face 正视	1184	Wallet		
19 Serious harm 严重危害	2217	Yúfúling	Wikipedia 维基百科	1096			
20 Necessary 就要	2181	Yúfúling	Website 网站	768			

Note that some word segmentation errors are present in Table 6: The term Yúfúling (鱼浮灵) was segmented into two separate words: Yú (fish) and fúling (floating spirit). Also, the first two characters from the Chinese term for using Yúfúling was wrongly segmented into “Canadian fish”. This reveals the challenges in dealing with Chinese texts containing seldom-used terms. At the same time, these can be used to detect new terms.

While the above findings indicate the “hot topics” on which microblogs *cite* the two encyclopaedias, they say little about the comments *about* the encyclopaedias. Thus the posts about the two incidents are removed, producing the results listed in Table 7.

**Table 7. Most frequently-occurring words when mentioning... (two major incidents removed)**

Baidu Baike			Chinese Wikipedia		
Word	Counts		Word	Counts	
1 Baidu 百度	2552		Website 网站	768	
2 Share 分享	2418		Free(dom) 自由	669	
3 China 中国	1784		China 中国	663	
4 “a bit” 一下	1135		United States 美国	660	
5 Entry 词条	983		Forwarding 转发	608	
6 Microblogging 微博	930		Encyclopedia 百科全书	533	
7 Ency. 百科	683		Wiki 维基	498	
8 One kind 一种	624		Google 谷歌	447	
9 World 世界	618		“a bit” 一下	444	
10 Quote 引用	587		Baidu 百度	430	
11 United States 美国	567		Microblogging 微博	423	
12 Forwarding 转发	562		Foundation 基金会	401	
13 Company 公司	551		google google	394	
14 Search 搜索	536		Event 事件	382	
15 Beijing 北京	507		Community 社会	377	
16 Support 支持	501		Entry 词条	374	
17 Introduction 介绍	494		World 世界	374	
18 Activity 活动	485		Knowledge 知识	371	
19 Network 网络	453		English 英文	366	
20 Website 网站	441		Data 资料	359	

Some of the words most commonly associated with Baidu Baike included the verbs “share” and “quote,” the nouns “China,” “entry,” and “Weibo,” and some geographic entities such as the “United States,” “Beijing,” and “Japan.” With Chinese Wikipedia, the list of words included “website,” “free(dom),” “China,” the “United States,” “wiki,” “Google” (both in Chinese and in English), “a bit,” “Baidu,” and “Weibo.” As will be discussed later, the term “a bit” (一下) points to the phenomena where the term “wiki” and other wiki-based encyclopaedias become verbs that describe the action of using online encyclopaedias (e.g. to wiki a bit”).

The same procedure that was used to extract the major discussion threads and “hot topics” could be deployed again for obtaining high-frequency words listed in Table 7. The words on the left provide clues that lead to a series of prize-winning campaigns initiated by Baidu Baike to promote itself on Weibo, accounting for the high frequency of the term “share”. The words on the right, including the terms “website” and “free(dom),” point to Internet filtering and censorship in China.

The high frequency of the verb “share” for Baidu Baike is the direct result of a 5-year anniversary campaign by Baidu Baike that started in early April 2011. More than 85% of the instances were simply sharing a certain entry article on Sina Weibo, with an explicit note: “shared from @Baidu Baike”. Weibo users were incentivized to promote Baidu Baike in this way by a chance to win a prize in addition to gaining points in Baidu Baike. With the help of machine-assisted clustering of the texts, I also identified another set of more than 2600 posts that promoted Hudong Baike and Baidu Baike with mentions of awards. (Like Baidu Baike, Hudong Baike is another commercial user-generated encyclopedia website hosted in mainland China[18]). In fact, more than 1800 posts that came from Hudong Baike constituted more than 30 different prizes from smartphones to chocolate. This phenomenon also indicates certain forms of cross-spherical activities between the platforms of Weibo and those of other user-generated encyclopaedias.

The high frequency of the terms “free(dom)” and “websites” for Chinese Wikipedia, in contrast, is largely the outcome of the filtering and censorship experienced by Weibo users across several websites or topics. Using terms such as “sensitive” (敏感), “the Great Firewall” (GFW), “walled” (被墙), “climbed over the wall” (翻墙), I found that many Weibo users share their experiences on a range of websites and topics to gauge the boundaries of filtering/censorship

Table 8 lists some of these posts. (Each ID number refers to a particular post; negative numbers signal that the post is collected from the WeiboScope; positive ones from the DiscoverText). Hereafter I refer to a post by using the format of ID123 or ID-45. For instance, one Weibo user questioned why her/his post from Wikipedia to the Sina blog was deleted but survived in Sina Weibo (ID4414) and another user claimed, “Wikipedia has become classified as sensitive words in Sina Weibo” (ID-29805). For others, Wikipedia become one of the tests of censorship, with various access issues to its website and specific pages. In post ID2667, a Weibo user shared her/his experience of being blocked from visiting Wikipedia. Post ID-5667 described another users’ observation that the article on the Chinese Cultural Revolution was unblocked but the article on a Chinese official was not. Post ID1303 even suggested that, her/his experience of being blocked probably because of several visits to sensitive articles on Wikipedia, could somehow be resolved by visiting Baidu.

**Table 8. Selected posts that mentioned CW in relation to blockages or censorship**

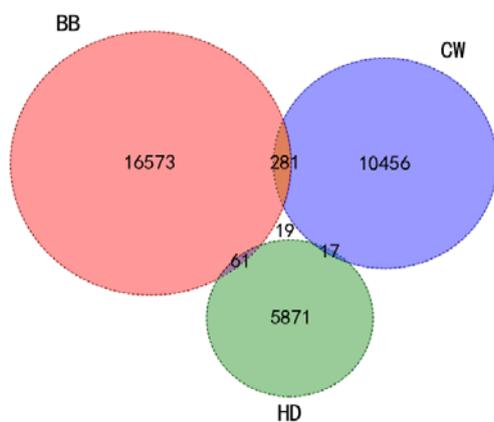
Category	ID	Main text translated
Sina as censors (blog or Weibo)	4414	• Sina blog deleted my post copied from Wikipedia; is Sina blog more sensitive than Sina Weibo??? The same post survive Weibo.
	15243	• [citing CW] ... after the Wang Lijun incident, the name "Jiang Weiping" was no longer blocked by Baidu or Sina Weibo.
	-18987	• Well, examples such as the "June Fourth" is blocked by Sina and Baidu as sensitive historical events, but Wikipedia dares to tell the truth on historical facts.
Access Issues	-29805	• Wikipedia has been listed as sensitive words in Sina Weibo [emoticon: pity]
	2667	• Can any one open Wikipedia websites? Why after several pages of visit, I saw "this page cannot be displayed". Nothing sensitive really ah.
	-5667	• The ban on Wikipedia is lifted, and the term "mosquito cells" [referring to China's Cultural Revolution] is also removed from the list of sensitive terms. But the recent ex-leader of Chongqing remains inaccessible.
Views on the issue of access	1303	• On Wikipedia, I visited a relatively sensitive topic, ended up with no access to all pages. Later I entered Baidu's address by accident. As a result, I can go online and search for information. Thank all-powerful "Baidu", my grateful tears of
	2609	• Just finished the exam on Marxism, and then you will not let me to visit wiki now?! You crazy ah! I hate the Great Firewall!!
	2997	• After searching some politically sensitive terms as well as several major events using Wikipedia, it feel really good, and should be closer to the truth. I have a new perspective on the historical events. The same words on Baidu Search I find "Search results are not displayable because they may not comply with relevant laws, regulations and policies." Better use foreign websites, for they may be closer to the truth. My mourning for the innocent people who were persecuted.
	3270	• What triggers the sensitive nerve system as Wikipedia becomes inaccessible? Life is like a farce at times. Can never get the intention of the "superstructure" [high officials]. Better to live in the moment and avoid commenting. Time is the best encyclopaedia, older when we revisit, we'll realize other stories exist.

Still other users voiced their opinions on their encounters with blockages or censorship when using Wikipedia. Post ID2609 seemed to express a student's angry response for not being able to access Wikipedia after a Marxism exam. Posts ID 2997 and ID-18987 expressed their belief that Wikipedia was more likely to be true for sensitive topics, whereas post ID3270 expressed a more passive attitude about the futility of commenting.

This way of using high frequency words to identify major discussion threads and "hot topics" can be iterated after removing found sets of posts. The procedure can be useful to identify major topics one by one. We can thus now move to the posts that explicitly mention both Baidu Baike and Chinese Wikipedia.

### 3.2 When mentioning both

To zero in on the posts that explicitly compare Baidu Baike and Chinese Wikipedia, we can now examine the posts that mention both Baidu Baike (BB) and Chinese Wikipedia (CW). Figure 1 shows that there are 281 posts mention both.



**Figure 1. The mentions of the three encyclopaedias**

Users' comparisons are most likely to occur in these posts, which can highlight contrasting experiences and attitudes about the two encyclopaedias. Thus, focusing on the 281 posts that mention both Baidu Baike and Chinese Wikipedia, two set of analysis can be conducted to see (1) the relationship of users to the censorship/filtering regime and (2) how Weibo and Twitter users compare. I address these two avenues of analysis separately because it is likely that the differences in censorship/filtering between Weibo and Twitter influences users. To do so, I used Chinese terms such as "wall," "harmony," and "river crabs" that refer to the filtering and censorship regime to derive the first sample. Then I manually coded the remaining posts for the second analysis.

At least 21 posts specifically mentioned the keywords referring to the censorship/filtering regime, of which 15 posts are listed in Table 9 (the unlisted ones contained irrelevant content).

**Table 9. Selected posts that mentioned both encyclopaedias (in relation to filtering/censorship)**

Category	ID	Main text translated
Baidu Baike is better	-13491	• On "Socialism with Chinese characteristics", I responsibly vouch for the correct version in Baidu Baike, Wikipedia's entry is simply nonsense! How could such an unreliable site have a place in China? It should be thrown over the Wall.
	-26194	• For the people of China, visiting Baidu Baike is enough. Consider the risk of being "Walled" when visiting Wikipedia.
Baidu Baike as alternative	-28952	• Having no choice, Wikipedia is "Walled". Please use Baidu Baike as the foundation.
Chinese Wikipedia is better	4388	• Wikipedia is more fair and comprehensive! I advise everyone not to use that "harmonized" Baidu Baike!
	5323	• Some people died on foreign soils, but remained alive in the country. Some died in Google, but still alive in Baidu. Some died in Weibo, but still alive on the CCTV [Chinese Central TV]. Some died in Wikipedia, but still alive in Baidu Baike. Some are dead outside the Wall, but inside the Wall still alive.
	-5667	• Some people died in a foreign country, but still alive in the country. Some died in Google, but still alive in Baidu. Some died in Weibo, but still alive on the CCTV. Some died in Wikipedia, but still alive in Baidu Baike. Some is dead outside the Wall, but still alive inside.
	1303	• [On certain botany entry..] Opposite plants are treated as the same species in Baidu Baike. What a miracle in this harmonious society ah! While accumulating your wealth, Du-girl [feminized name for Baidu], would you please do something substantial a bit? Access to Wikipedia blocked, Baidu Baike cannot be trusted. No wonder many brain-damaged in this Heavenly Dynasty!
	2609	• Ever since on the other side of the Wall, I become used to using Google and clicking on Wikipedia.. [to other user]:To be honest, I have now subconsciously use Baidu because Google is often off due to some unknown forces. Besides, Baidu Baike, Baidu Zhidao and Baidu Wenku indeed make a great contribution to the diffusion of knowledge to the ignorant Netizens like me. [to yet another user] I choose to wait another ten minutes when Google is not accessible.
	2997	• Baidu Baike not only copied but also harmonized the content from Wikipedia. Much to see in Wikipedia once over the Wall.
	3270	• I always thought Baidu Baike is the best, because it understand Chinese language very well! I was wrong! Completely wrong! In fact, Wikipedia is the best! And truly the best when over the Wall!
	3301	• Better to visit Wikipedia over the wall ... All lies in Baidu Baike.
	3403	• This "river-crabbed" [meaning censored] entry existed in Baidu Baike long time ago, which tells us nothing about the Great Firewall. Whether we can visit certain entries in Wikipedia is the standard test whether the "river-crab" (censorship) is loosening.

From the top to the bottom, the first category of posts showed minority opinions of the selected dataset: one claimed that Wikipedia's specific entry on "Socialism with Chinese characteristics" is nonsense when compared to the Baidu Baike's correct version (ID-13491). The other argued that Baidu Baike was enough for people of China, citing the risk of being "Walled" (ID-26194). The post in the second category shared a similar opinion that Baidu Baike is an alternative when Chinese Wikipedia is blocked. Of the majority of the posts that mention both encyclopaedias in relation to filtering/censorship, the third and most representative category consisted of posts arguing that Chinese Wikipedia is better because it is not censored: Wikipedia

is "more fair and comprehensive" (ID4388), "the best" (ID-33992), "have much to see" (ID-33991) and can serve as "the standard test" of censorship (ID-40579), whereas Baidu Baike is censored (ID5323) or "harmonized"(ID4388, ID-33991 and ID-40579) or full of "lies" (ID-40570), repetition (ID2790) and errors (ID-3473). Among them, clear anger was also expressed in posts ID2790 and ID-24937 against access blockages.

Note that post ID-32406 discussed the decisions users face when they are blocked. One user reflected on the gradual change to the use of Google and Wikipedia that occurred when living outside the Wall, and another comment reflected on the subconscious shift to Baidu because of the Wall, while another made the conscious decision to wait for ten minutes when "Walled". These comments clearly suggest that not all users of mainland China are oblivious to the existence of Wikipedia and that their use and reception of the websites inside and outside the Wall, while diverse, is more or less conditioned by it. Concerning Baidu Baike and Chinese Wikipedia, users clearly expressed contrasting attitudes to the Great Wall, which also provides additional user testimony on the impacts of the censorship/filtering regime imposed on the two: Baidu Baike is often censored and Chinese Wikipedia is blocked.

To compare how user opinions differ apart from in relation to the censorship/filtering regime, I examined the remaining posts for users' evaluative terms about the two encyclopaedias, and found 39 posts (listed in Table 10). Of these, the majority found Chinese Wikipedia to be better. Baidu Baike was reported to be less reliable (e.g. post ID3502, 5733, 5917, 6210, -7323, -7371), less objective (e.g. post ID3570, 6190, -1457, -10079, -13965, -19690, -25687, -34458) and prone to government's influence or censorship (e.g. post ID5587, -19547, -19690, -21512, -25182, -25221). In contrast, Chinese Wikipedia was perceived to be more useful (e.g. post ID6306), better referenced (e.g. post ID5733, -3755) and to have an elegant interface and style (e.g. post ID3664). Nonetheless, some posts exist stated Baidu Baike was better. Post ID-14431, for example, claims that Baidu Baike is more detailed on R&B (rhythm and blues music). Post ID-26631 suggested that Baidu Baike was richer in content because it could "freely" copy content from other websites, and thus had more comprehensive content that suits the needs of users in China. Post ID-19336 seems to be ironic by agreeing that Baidu Baike is thus "freerer" (I am not be certain whether it is meant to be ironic).

**Table 10. Selected posts that mentioned both (apart from those concerning filtering/censorship)**

Category	ID	Main text translated
Baidu	-14431	• On R&B, Baidu Baike is more detailed than Wikipedia.
Baike is better	-19336	• It turns out Baidu Baike is freerer than Baidu Baike [Haha]
	-26631	• Users of China do not like Wikipedia. Why? For users who only read without editing, they do not care whether the content is copied or original. All they want is plenty and complete information. Because Wikipedia cannot copy but Baidu can copy Wikipedia and others. Baidu Baike's content is richer than Wikipedia's.
	-27322	• You are welcome to read the essay penned by "Between Men&God", an administrator of Chinese Wikipedia, published in Douban.com: "Why is (Chinese) Wikipedia encyclopaedia worse than Baidu Baike?" (Chinese Wikipedia indeed needs constant self-inspections. Wiki-girl approves this article [emotion:greivance])
	16123	• Chinese Wikipedia recently did something stupid. Competing with Vietnamese Wikipedia on the number of entries. It appears that, outnumbered by Baidu Baike and Hudong Baike, Chinese Wikipedia starts to enjoy itself by comparing with minor languages. After Chinese Wikipedia criticized its competitors' value of quantity over quality, a rather ironic development that it enters the number game.
Both good	-5667	• Occasionally Baidu Baike is useful and very comprehensive. Wikipedia is better when looking up English words.

(Table 10 continued)

- Chinese Wikipedia is better
- 1303 • [On "Great Chinese Famine"] Better read more detailed Wikipedia. ... Just also checked Baidu Baike's, which is almost consistent with Wikipedia's. It is a rare situation where Baidu Baike shows some progress.
  - 2609 • [On Zhang Zhixin] Huge difference between Baidu Baike and Chinese Wikipedia. Which lies?
  - 2997 • The difference between Wikipedia and Baidu Baike? Can tell from one simple entry on Fang Zhouzi..
  - 3270 • Stronger than Baidu Baike: Wikipedia
  - 3301 • My foreign teacher asked us to consult Wikipedia and share found information in class. Finally realized that Baidu Baike is the epitome of the new online "Sakoku" (isolationism). Although having less entries, Chinese Wikipedia has absolute essentials, totally unlike Baidu Baike which copies and pastes stuff everywhere. Better learn English well to enjoy the fun reading English Wikipedia!
  - 3403 • Wikipedia better than Baidu Baike. Checked the history of Haifeng, Wikipedia has the same info as the county government website, but Baidu Baike is way off!
  - 3502 • Baidu Baike's content unreliable. Contributors often promote companies, books, and so on. The entries on the Internet of things, Smart Industries, Industry clusters, etc. mislead readers. I checked; none is comparable to English Wikipedia.
  - 3570 • For "Wukan incident", now we can visit Wikipedia's description, which is objective, very detailed, encyclopaedic, and thus much better than Baidu Baike's
  - 3664 • Wow the Web interface is elegant and language is stylish in Wikipedia. No more Baidu Baike for me from now on.
  - 4021 • Our history class assignment has the Red Guards as its theme, a bit heavy and difficult. Everyone is advised to take a look at Wikipedia's and Baidu Baike's entry.
  - 5587 • Very fun contrasting Wikipedia and Baidu Baike. Cannot open Wikipedia's page because it contains the phrases such as "... regarded as a dissident for her opposition to Mao's personality cult and extreme leftism". Can open it in Baidu Baike, because its description goes like this. ...
  - 5733 • Baidu Baike not reliable. It appears to have complete information but all of them unreferenced or self-referenced... Better use Wikipedia. At very least, each Wikipedia's entry has clear references, respecting original authors' copyright.
  - 5917 • After reading the linked article, I search for information on "Pan Jinfu". With the help of @hischild, I found this article. Baidu Baike's entry proved to be "watery"... Wikipedia is more reliable. Of course, what is the most reliable is our habits to look for original sources, which can help understand the facts.
  - 6190 • The description on Milton Friedman in Baidu Baike, in contrast to Chinese Wikipedia, removed the parts where he was challenged and questioned by others.
  - 6210 • Scholars with expertise have conducted research and found the precision of Wikipedia is higher than encyclopaedia Britannica, and thus can be treated as reliable sources. Note: Do not ever use Baidu Baike and Hudong Baike, where false information is distributed, otherwise you will hurt yourself and others.
  - 6306 • Now I realized that Wikipedia is more useful than Baidu Baike.
  - 1457 • Better use Baidu Baike less. Like Baidu search, it is castrated and inflicted by viruses...Wikipedia has verification and coordination mechanisms and Baidu Baike has not. With a political leaning, Baidu Baike is wildly unfair and chaotic.
  - 3755 • Wikipedia first demands info sources, second requires footnotes in scholarly formats, third prohibits [editors'] original research, and fourth says no to vandals and ads. If Baidu Baike can do any of the above, it is no longer Baidu Baike
  - 7323 • [Comments on the errors in Baidu Baike's prazepam entry...] Everyone please use Wikipedia when consulting entries in professional fields such as medicine, architecture, law, etc. Baidu Baike is far behind Wikipedia in terms of professionalism and precision.
  - 7371 • [On making a mistake regarding the age of an actor.] I recognized my mistakes after googling. ... I realized this: Wikipedia is indeed better than Baidu Baike. It is not about the word count, but about correct information
  - 7622 • Fortunately Wikipedia is not wrong on this. When can Baidu Baike correct itself to avoid repeating baseless assertions?
  - 10079 • Today I bumped into a new term "neoliberalism". Not knowing what it is, I looked it up in Baidu Baike and Wikipedia. After comparison, the orientation of the two sites is clearly different. The former contains largely negative comments, with emphasis on the opposition of socialism and capitalism... The latter does not contain comments with ideological orientations, except for explaining all knowledge backgrounds. ... From now on wiki only ... Baidu you garbage.
  - 13965 • Encyclopaedias are supposed to document facts. Have a look at the content in Wikipedia and Baidu Baike on TVXQ (a South Korean pop group) and its five members. I seldom visit Baidu, and today at a casual look I found it contains weird stuff in it. Speechless. An encyclopaedia is used as a fan site, with all its biased comments. Only Baidu is brain-damaged enough to allow such content. No chance for such content to survive in Wikipedia's standard.
  - 16600 • In practice, Wikipedia is proved to be more cultured than Baidu Baike.
  - 19547 • Wikipedia is better than Baidu Baike. Baidu is a dog for the government.
  - 19690 • On many topics, the results out of using Wikipedia is more objective than Baidu Baike.
  - 21512 • Both as encyclopaedias, why such a big different between the content from Baidu Baike and that from Wikipedia? ...[sh] [emotion:depressed]
  - 25182 • Baidu Baike does not have and Chinese Wikipedia has it. Immediately I know why.
  - 25221 • No such paragraph in Baidu Baike. So much content is different from Wikipedia.
  - 25470 • Ha ha ha ha ha Said the teacher. Wikipedia is more authoritative than Baidu Baike!
  - 25687 • Consult Wikipedia and Baidu Baike respectively on the topic of the "White Sea - Baltic Canal". Contrast and see the obvious evil of Baidu.

Still, these comments do not contradict the observation that overall, Chinese Wikipedia values content quality and respects copyright more [16]. In fact, the essay mentioned by post ID-237322 is originally titled: “What kind of improvements does (Chinese) Wikipedia need when compared with Baidu Baike.” In the essay mentioned, before summarizing his opinions that Chinese Wikipedia’s content is too limited, the page editor too complex to use, and it is too unfriendly for newcomers, the writer (an administrator of Chinese Wikipedia) acknowledged that much has already been said about where Chinese Wikipedia is better.

Altogether, these posts represent a sample of opinions voiced in both Weibo and Twitter, suggesting that Chinese Wikipedia is perceived as more reliable and less censored than Baidu Baike. In particular, posts ID3301 and ID-25470 use teachers’ opinions, and the former shares her/his class experience in using Wikipedia as that “Baidu Baike is the epitome of the new online ‘Sakoku’” (isolationism). The two posts are amongst many posts that mentioned Baidu Baike or Chinese Wikipedia as one of the important digital tools for improving one’s understanding of the world. The idea that online encyclopaedias are important tools for gaining information is further supported by the fact that the Chinese term “wiki” has become a verb similar to the verb “to google”.

### 3.3 When “wiki” becomes a verb.

Indeed in the dataset, like “google” in English, “Baidu” and the Chinese term “wiki” (维基) are used as verbs that refer to visiting these encyclopaedias for information. I found nearly 200 posts for each of the Chinese-language phrases “Baidu a bit” and “Wiki a bit.” To find out what kinds of “verbification” are expressed, I categorize these posts according to the preceding words because they provide contexts for such verbification. Table 11 lists the categorized results.

Table 11 Baidu/Wiki a bit

Category	"Baidu a bit" (百度一下)		"Wiki a bit" (维基一下)	
	Precedent	Word context	Precedent	Word context
Action done	刚(2) ...just (recently) Baidu a bit...		刚(11) ...just (recently) Wiki a bit...	
			刚(3) ...just (recently) Wiki a bit...	
			刚/刚才(2) ...just (recently) Wiki a bit...	
	于是(4) ...then Baidu a bit...		顺手(2) ...casually Wiki a bit...	
	特意(3) ...purposely Baidu a bit...		重新(2) ...again Wiki a bit...	
			再(2) ...again Wiki a bit...	
Action	去(17) ...go Baidu a bit...		去(10) ...go Wiki a bit...	
Action possible			可(10) ...can Wiki a bit...	
			能(10) ...can Wiki a bit...	
	可以(8) ...can Baidu a bit...		可以(3) ...can Wiki a bit...	
Action better required	先(6) ...first Baidu a bit...		不会(4) ...why not Wiki a bit?...	
	要(2) ...must Baidu a bit...		为啥不(2) ...why not Wiki a bit?...	
	只要(2) ...simply Baidu a bit...		最好是(1) ...better Wiki a bit...	
	请(2) ...please Baidu a bit...			
Pronoun and noun	大家(4) ...everyone Baidu a bit...		大家(2) ...everyone Wiki a bit...	
	我(3) ...I Baidu a bit...			
	你(3) ...you Baidu a bit...		你(3) ...you Wiki a bit...	
	同学(2) ...classmate, Baidu a bit...		@(5) ...@[a Weibo and Twitter user]...	
Promotion	都(23) ...all Baidu a bit...			
Action combined			百度(8) ...Baidu/Wiki a bit...	
			或者(6) ...or Wiki a bit...	
			和(3) ...Baidu and Wiki a bit...	

The four categories in Table 11 distinguish between when an action is done, will be done, can be done, or should be done. More than 20 posts refer to the posters’ past actions (the category “Action done” in the table) in using one of the encyclopaedias for information, and there are more such posts for Chinese Wikipedia: e.g. “... just recently wikied a bit”. This suggests that some users share their experiences using online encyclopaedias on microblogs. In addition, the categories “Action possible” and “Action better required” refers to when users make the suggestion that online encyclopaedias should be consulted.

Note that there are eight posts that use the combined verb “Baidu/Wiki a bit” (百度维基一下), with an example post quoted below (ID-31873):

*The masses should take advantage of this opportunity to popularize this universal knowledge ah! If one does not believe in experts, one can Baidu/Wiki a bit.*

One instance that distinguishes Baidu Baike and Chinese Wikipedia is when the verb “Baidu a bit” is used by Baidu to promote itself in Weibo. At least 23 posts, most likely generated by some Weibo users who participated in promotion campaigns, followed the format:

*Got news? Brush [meaning "refresh" or "update" in English] Weibo. Got new knowledge? Brush Weibo too. Because what matters most for communication is brush something anew. It is most foreign flavour (meaning "exotically fashionable" in English) to all Baidu a bit on everything. What you do not know, @BaiduZhidao let you know. I am @UserID [Weibo user] at Weibo.com. Got something to ask? Don't ask me. Ask @Baidu Baike.*

What the above post demonstrates is exactly the crossover of communicative and information spaces, where the information spaces of Baidu Baike and Baidu Zhidao [a Baidu user-generated questions and answers website,] can be used to update new information in the communicative spaces of Weibo. Although such cross-over or cross-spherical activities may in this case not be completely voluntary, there is ample evidence that Weibo or Twitter users share their experiences in using search engines, user-generated encyclopaedias and microblogs across platforms. Some of them have also noticed the relationship between the SERPs and UGEs:

*...As matter of fact, you will know by simply Baidu a bit. Key in three-character “Yinbingshi” to search, the first item should be Baidu Baike’s entry on Yinbingshi.... (ID-33642)*

*Google is sensitive...search “lie group” one can find Lie group-Wikipedia, the free encyclopaedia as the top result. (ID-23435)*

*...Simply Baidu a bit, Baidu Baike has a specific entry on him.... (ID-34850)*

*...Amazing that through Baidu I found the “sensitive”-men (i.e. Tiananmen) incident twenty years ago on a sensitive day and sensitive month in a sensitive year. The first SERP item is Wikipedia’s entry page on this incident, and it can be opened. (ID-42437)*

Thus in general, what is shared between “Baidu a bit” and “Wiki a bit” as a verb is the specific use of Baidu Baike and Chinese Wikipedia: looking for information. I found no cases where the verb is used to refer to writing or contributing content to them.

Thus, in the Chinese-language context, as shown by the text analysed here, when used as verbs, the terms “Baidu a bit” and “Wiki a bit” refer to general information-seeking actions, not to the original meaning of wiki for collaborative writing. The posts clearly indicate the prominence of using Wikipedia and/or Baidu Baike as information practices on par with Google or Baidu.

#### 4. DISCUSSION

Ample evidence exists to show the existence of cross-platform activities. Weibo and Twitter users are often aware of both encyclopaedias. Research must take such cross-platform activities into account, as suggested by some digital method researchers [27]. Such cross-platform activities also include promotion campaigns initiated by Baidu Baike and Hudong Baike, and even cyber-attacks against Wen Yunchao’s Twitter account by flooding tweets using Baidu Baike’s entry, as also testified by Wen Yunchao himself at a US Senate hearing:

*....For about a year starting April, 2011, unidentified persons “tweet bombed” me on Twitter with trash information. Using a software [program] called Tween to filter the trash, I found the heaviest attack took place on April 25, 2012 – with a staggering 590,000 spam posts within 24 hours. Unidentified persons also posted viciously defaming information about me online at the rate of over 10,000 times per day. As far as I know, artist Ai Weiwei has been similarly attacked ....*

Thus, the Twitter data collected for this research (from March 20, 2012 to April 12, 2012) contains the early phase of the spam post attack against Wen Yunchao on Twitter. Encyclopaedia pages’ titles and links, mostly from Baidu Baike, can be used to jam a Chinese activist’s account in Twitter that is outside Beijing’s filtering and censorship regime. This is evidence of information engagement in the sense of information warfare. The availability of user-generated encyclopaedia data is repurposed to jam a microblog account by introduced introducing noise to keep certain Chinese activists from engaging their intended audience.

While this study had not yet applied a systematic method to detect and remove such jamming data points, it indicates how future research might identify such abuse on microblogs. Posts that keep mentioning user-generated encyclopaedias’ different articles in a short timeframe targeting individuals are likely to be jamming.

Also, the systematic method of first clustering posts (by computers programs) and then categorizing them (by analysts) one by one through identifying and then removing the hot topics should also help detect other spamming or en-masse promoting activities, as demonstrated by the self-promotion campaign initiated by Baidu Baike and Hudong Baike. Although the approach of the study may be questioned on the grounds of accuracy and validity, it has several merits that traditional methods such as user interviews cannot rival. First, this approach is feasible and reproducible enough for constant monitoring. Dynamic reports could be quickly generated with a workflow consisting of computer programs and social media analysts. Second, this approach may be better to anticipate or even predict the hot topics of public discussions that use or comment on user-generated encyclopaedias. For instance, to predict trendy topics, future research may use the time series scanning method[30]. Third, the accuracy and validity of findings can be further enhanced through better programs and more experienced analysts after several iterations.

The use and reception of the two major user-generated encyclopaedias on Weibo and Twitter yielded several findings.

First, it confirmed that some Weibo and Twitter users use Chinese Wikipedia and/or Baidu Baike, and that Beijing’s filtering and censorship regime has different effects on these uses, thereby contributing to the existing research on online Chinese-language encyclopaedias [15, 19, 21]. Several users were also aware of the differences between the two encyclopaedias. In fact, several posts suggest the contrast between Baidu Baike and Chinese Wikipedia as a test for tracking what counts as a sensitive topic. While users did sometimes get frustrated when they were denied access to Chinese Wikipedia, many of them preferred Chinese Wikipedia precisely because it was not censored. Second, overall, users had positive experiences using online encyclopaedias, and more so in using Chinese Wikipedia than in using Baidu Baike. For example, Chinese Wikipedia was perceived to value content quality and respect copyright more, in contrast to Baidu Baike, which was known for copying and pasting content from other websites. Explicit comparisons were also made by Weibo and Twitter users in short posts, and more users perceived Chinese Wikipedia to be of higher quality. Third, like the new English word “google” being used as verb, the Chinese terms “Baidu” and “Wiki” are also used as verbs to describe looking up content in Baidu Baike or Wikipedia. This suggests that the use of online user-generated encyclopaedias have become an important part of users’ online experience, including when users cite information to discuss topical issues, request that others find information, or verify information seen elsewhere. For these users, user-generated encyclopaedias were resources they relied on.

Moreover, the findings show that many users view the two resources as an established part of everyday information activities, like using the search engines Google and Baidu. These activities are often linked with information or digital literacy, as exemplified by comments from students using Wikipedia in classroom settings. Here we can see the expectation that user-generated encyclopaedias lead to the betterment of self and society through gaining knowledge, partly expressed by a post suggesting that “[t]he masses should take advantage of this opportunity to popularize this universal knowledge ah! If one does not believe in experts, one can Baidu/Wiki a bit” (ID-31873). The new experiences with information and communication spaces are increasingly associated with these websites, as are cultural and social norms about information access, knowledge sharing, and content filtering and censorship. The online practices of information engagement are thus expected to be influenced by users’ experiences with different platforms, which are of cultural-political significance [17].

The findings also have implications for youth engagement in mainland China since Weibo users in mainland China are more likely to be younger. More than 50% of Sina Weibo users were 26-35 years old according to an industry report in 2012 [7] and more than 80% of the visiting traffic came from 10-39 year old users according to an annual report released by Chinese Academy of Social Sciences in 2014 [29]. In light of the rough user profile, it can be argued that state and market players are expected to engage the younger generation in mainland China through microblog platforms, which may explain why the information warfare kind of information engagement is observed in the case of jamming microblog accounts of Chinese activists such as Wen Yunchao. It may also explain why the learning aspect of citing, using, and sharing experiences about user-generated encyclopaedias is also important for several microblog users.

## 5. CONCLUSIONS AND IMPLICATIONS

In terms of information engagement, this study finds that microblog users cite major user-generated encyclopaedias to engage in information-critical or information-intensive public discussions. In a broadened sense of open collaboration, the open discussion on microblog platforms use information resources from user-generated encyclopaedia platforms.

The dynamics of the information engagement concerns the largest non-English segment of the world's Internet population, and this paper has examined how it is practised and expressed on two major Chinese-language microblog platforms about two major Chinese-language encyclopaedias. The findings demonstrate the potentials and gaps towards more active and sound collaborative information engagement beyond mere information seeking and consumption. The microblog posts were still predominantly about using encyclopaedias (in a way contributing to better informed users and public discussions), not about contributing back to the content of user-generated encyclopaedias.

Still, both Baidu Baike and Chinese Wikipedia enjoy mentions from Twitter and Weibo users, and both have become verbs in the Chinese language just like the new English verb "google." The Chinese verbification of the two encyclopaedias thus indicates what has been achieved and what has yet to be achieved. More research and practice is needed to advance the understanding of the potentials and the challenges to improve information engagement on both platforms, possibly using microblog discussion to improve encyclopaedia platforms and their content as well as other open collaboration activities that are more "egalitarian," "meritocratic" and "self-organizing" [26].

With regards to the recent efforts by Chinese Wikipedians to engage users on Sina Weibo [32], the findings point to a direction in which users engage in public discussions on microblog using and enhancing encyclopaedia articles. The research methods here can be repurposed and streamlined to build a public opinion monitoring system. First, robust jam- or spam-detection software can be developed and implemented to highlight the abuse and/or misuse of user-generated encyclopaedias. In addition, it might be possible to show (or even predict) which incidents may become hot discussion topics that require quality and authoritative information sources that may or may not be found in major user-generated encyclopaedias. Future work can be conducted to explain, predict, or even plan interventions to the information-related activities that can be publicly observed across the microblog platforms and user-generated encyclopaedia platforms. Overall, the development of certain event monitoring devices, information engagement rules and new encyclopaedia articles may result in better public discussion and better human knowledge. This may in turn translate into more kinds of civic information engagement and less information warfare.

A division of users and user experiences currently exists in mainland China characterized by the choice between Baidu Baike and Chinese Wikipedia. While the frequent occurrences of the verbs such as "Google a bit", "Baidu a bit", and "Wiki a bit" suggest that user-generated encyclopaedias are part of users' overall information-seeking activities, the choice between Baidu Baike and Chinese Wikipedia entails different norms and attitudes. Baidu Baike was often associated with the Beijing-censored version of information sources and Beijing-sanctioned ways of information engagement. On the other hand, Chinese Wikipedia is not censored and thus was perceived by some to be more reliable and was associated with the freedom of information.

Access blockages to Chinese Wikipedia were also perceived as signals of Beijing's political "sensitivity" towards information.

The perspective of learning to find and cite information in order to participate in public discussions shows the potential of social media platforms as "civic learning repertoires" for "information engagement" [2]. For users in mainland China, at least two major learning strategies can be identified: one takes the censorship and filtering regime for granted, and the other seeks information and communication platforms that are outside the influence of the regime. It is likely that the "civic learning repertoires" for "information engagement" are limited to the bounds of the regime (i.e. mainland China) for those who use Baidu and Baidu Baike. Indeed, both Baidu Baike and Chinese Wikipedia become important reference points and even everyday communicative practices (e.g. "Baidu a bit" or "Wiki a bit"). Still, the different uses and receptions suggest two set of learning experiences and information choices. For contributors and readers, the choice of one over the other reflects the different experiences in civic learning made possible by the two websites.

The findings add to the cross-spherical analysis of web spaces proposed by Richard Rogers[27]. Because of the citing and commenting activities, Web spaces are indeed porous between Chinese-language microblog and encyclopaedia platforms. On the other hand, the distinct choice between Baidu Baike and Chinese Wikipedia suggests not only different information engagement practices because of the censorship/filtering regime, but also different civic learning repertoires. The censorship/filtering regime clearly has impacts on shaping web spaces and their information engagement activities.

The findings also update our understanding of the online "Chinese cultural sphere" of which Chinese-language resources are important[33]. Chinese-language platforms such as microblogs and user-generated encyclopaedias constitute major instances where such resources can be indexed, searched, retrieved and viewed amongst mostly Chinese-language users. Such openly available resources permit potentially new information engagement and even open collaboration. However, such potentials may be limited by censorship (e.g. Baidu Baike) or blockage (e.g. Chinese Wikipedia), shaping the engagement dynamics with different sets of norms and knowledge sources.

Future research may also consider how the concept of information engagement intersect with the official and popular discourse of the Chinese term "*suzhi*" (quality) of people, a discourse that demands individual competition, responsibility and self-improvement—including the ability to use technology to do so. In the dataset, several posts discussing the lack of "quality" information and people can be further analysed. Since the notion of information engagement is relatively unknown to the general public in China, the notion of *suzhi* can be used as potential bridging concept to discuss which kind of information engagement is better in promoting knowledge for the betterment of Chinese individuals and societies. Some questions can also be formulated based on the existing Chinese studies literature on "*suzhi*" for online platforms. For instance, *Suzhi* discourse may have aimed to turn peasants into modern Chinese citizens [22], but how does it work, say, to turn migrant workers into modern online Chinese citizens? If *suzhi* discourse justified social hierarchies [13], how is it related to the choices of information platforms and the practices of information engagement? More empirical research is thus needed to examine and compare across different information platforms and groups of users.

Important links are shown between public discussions and information engagement in this study. Since both microblogs and encyclopaedias depend on users and their contributed content, users learn to use and choose platforms while platforms learn from users. Future research is needed to advance the quality of public discussions with better information engagement.

## 6. ACKNOWLEDGMENTS

I acknowledge the WeiboScope team and Dr. Fu King-wa at the University of Hong Kong for the provision of the WeiboScope dataset. I also acknowledge Texifter.com for the trial of their platform that allowed me create the DiscoverText-Weibo dataset.

## 7. REFERENCES

- [1] Bamman, D. et al. 2012. Censorship and deletion practices in Chinese social media. *First Monday*. 17, 3-5 (2012).
- [2] Bennett, W.L. and Wells, C. 2009. Civic Engagement: Bridging Differences to Build a Field of Civic Learning. *International Journal of Learning and Media*. 1, 3 (Aug. 2009), 1–10.
- [3] Bolsover, G. et al. 2013. Social Foundations of the Internet in China and the New Internet World: A Cross-National Comparative Perspective. (Oxford Internet Institute, University of Oxford, Jun. 2013).
- [4] Chen, B. et al. 2006. Chinese Spoken Document Summarization Using Probabilistic Latent Topical Information. *2006 IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings* (May 2006), I–I.
- [5] CNNIC 2013. *Statistical Report on Internet Development in China (January 2013)*. Technical Report #31st. China Internet Network Information Center.
- [6] Damm, J. 2008. Technological Empowerment: The Internet, State, and Society in China. By Yongnian Zheng. Stanford, Calif.: Stanford University Press, 2008. xviii, 272 pp. \$50.00 (cloth). *The Journal of Asian Studies*. 67, 03 (2008), 1080–1081.
- [7] DCCI 2012. *Blue Book of China MicroBlog*. Data Center of China Internet (DCCI).
- [8] Fu, K. et al. 2013. Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and Impact Evaluation of the “Real Name Registration” Policy. *IEEE Internet Computing*. 17, 3 (2013), 42–50.
- [9] Fu, K. and Chau, M. 2013. Reality Check for the Chinese Microblog Space: a random sampling approach. *PLOS ONE*. 8, 3 (2013), e58356.
- [10] Hagestad II, W. 2012. *21st Century Chinese Cyberwarfare*. IT Governance Ltd. Cambridgeshire.
- [11] iResearch 2013. *Sina’s Weibo New Advertising Model Marks Monetization of the Long-tail Marketing Strategy*. iResearch Consulting Group.
- [12] King, G. et al. 2012. How censorship in China allows government criticism but silences collective expression. *APSA 2012 Annual Meeting Paper* (2012).
- [13] Kipnis, A. 2006. Suzhi: A Keyword Approach. *The China Quarterly*. 186, (Jun. 2006), 295.
- [14] Liao, H.-T. et al. 2013. Chinese Web Data: Definition, Uses, and Scholarship.
- [15] Liao, H.-T. 2009. Conflict and Consensus in the Chinese version of Wikipedia. *IEEE Technology and Society Magazine*. 28, 2 (2009), 49–56.
- [16] Liao, H.-T. 2013. How do Baidu Baike and Chinese Wikipedia filter contribution? A case study of network gatekeeping. *Proceedings of the 9th International Symposium on Open Collaboration* (Hong Kong, Aug. 2013).
- [17] Liao, H.-T. 2014. *The Cultural Politics of User-Generated Encyclopedias: Comparing Chinese Wikipedia and Baidu*. University of Oxford.
- [18] Liao, H.-T. 2014. “User Generated Content” (用户生成内容). *The Internet in China*. Berkshire Publishing.
- [19] Lou Z.-C. and Fu Z. 2008. The Impact Analyze of Some External Forces on the Wikipedia Order Process (外部因素对维基百科序化过程的影响分析). *Document, Information & Knowledge (圖書情報知識)*. 123, (May 2008), 28–33.
- [20] Luo, T.(罗婷) and Li, C.(李成) 2010. How to develop public opinion monitoring products? - The modes of operation of the Public Opinion Monitoring Unit, People’s Net. [如何开发舆情监测产品?——人民网舆情监测室的运作模式]. *Chinese Journalist (中国记者)*. 6 (2010).
- [21] Ma, F. and Xia, Y. 2008. Mechanism of ordering in Wikipedia based on the CAS theory. *Library Tribune*. 28, 6 (2008), 85–92.
- [22] Murphy, R. 2004. Turning Peasants into Modern Chinese Citizens: “Population Quality” Discourse, Demographic Transition and Primary Education. *The China Quarterly*. 177, (Mar. 2004), 1–20.
- [23] Netpop Research 2007. *Chinese consumers have dramatically surpassed Americans in adopting Web 2.0 behavior, relying heavily on social media for guidance in purchase decisions*.
- [24] Porche, I. et al. 2013. *Redefining Information Warfare Boundaries for an Army in a Wireless World*. Rand Corporation.
- [25] Qiang, X. 2010. User-generated content online now 50.7% of total. *China Daily*.
- [26] Riehle, D. 2012. Definition of Open Collaboration. *The Joint International Symposium on Open Collaboration*.
- [27] Rogers, R. 2013. *Digital methods*. The MIT Press. Cambridge, Massachusetts; London.
- [28] Tan, Y.F. 2012. Online public opinion monitoring: an exploratory report [探照网络舆情监测]. *CAIJING (财经)*.
- [29] Tang, X. 2014. *Annual Report on Development of New Media in China (2014) 《中国新媒体发展报告》*. Technical Report #5. Institute of Journalism and Media, Chinese Academy of Social Sciences (CASS).
- [30] Thelwall, M. and Prabowo, R. 2007. Identifying and characterizing public science-related fears from RSS feeds. *Journal of the American Society for Information Science and Technology*. 58, 3 (Feb. 2007), 379–390.
- [31] US Army 2009. *Tactics in Counterinsurgency FM 3-24.2*. Headquarters, Department of the Army Head.
- [32] Wang, A. 2013. Using social media to engage Wikipedia readers and editors in China. *Wikimedia blog*.
- [33] Yang, G. 2003. The Internet and the Rise of a Transnational Chinese Cultural Sphere. *Media Culture Society*. 25, 4 (Jul. 2003), 469–490.
- [34] Yang, Y. and Xie, L. 2008. Subword Latent Semantic Analysis for Texttiling-Based Automatic Story Segmentation of Chinese Broadcast News. *6th International Symposium on Chinese Spoken Language Processing, 2008*. (Dec. 2008), 1–4.
- [35] Yu, Z.-T. et al. 2006. Answer Extracting for Chinese Question-Answering System Based on Latent Semantic Analysis. *Chinese Journal of Computers 《计算机学报》*. 29, 10 (2006), 1889.
- [36] Zheng, Y. 2007. The Internet, Civic Engagement, and Public Distrust. *Technological Empowerment: The Internet, State, and Society in China*. Stanford University Press.
- [37] Zhu, T. et al. 2012. Tracking and Quantifying Censorship on a Chinese Microblogging Site. *arXiv:1211.6166*. (Nov. 2012).