



# Harnessing the Power of Collaborative Filtering

## Comparing the Network Gatekeeping of Baidu Baike and Chinese Wikipedia

Han-Teng Liao

---



**Electronic version**

URL: <http://journals.openedition.org/chinaperspectives/6854>  
DOI: 10.4000/chinaperspectives.6854  
ISSN: 1996-4617

**Publisher**

Centre d'étude français sur la Chine contemporaine

**Printed version**

Date of publication: 1 December 2015  
Number of pages: 35-49  
ISSN: 2070-3449

**Electronic reference**

Han-Teng Liao, « Harnessing the Power of Collaborative Filtering », *China Perspectives* [Online], 2015/4 | 2015, Online since 01 January 2017, connection on 28 October 2019. URL : <http://journals.openedition.org/chinaperspectives/6854> ; DOI : 10.4000/chinaperspectives.6854

---

# Harnessing the Power of Collaborative Filtering

Comparing the Network Gatekeeping of Baidu Baike and Chinese Wikipedia

HAN-TENG LIAO

**ABSTRACT:** China's Internet development raises questions on the interplay of user autonomy and Internet governance in the larger Chinese-language online cultural sphere. To what extent has user autonomy been established to conduct information gatekeeping collaboratively online? To examine how the power of Chinese-language Internet users has been harnessed in relation to the filtering and censorship regime, the article applies the concept of "network gatekeeping" to analyse two major user-generated websites, Chinese Wikipedia and Baidu Baike, as different examples of collaborative filtering projects. Effectively they share "word-of-mouth" recommendations on encyclopaedic knowledge and information by user-contributors to edit content from different parts of the world. Two salient network gatekeeping mechanisms have emerged: the censorship mechanisms in Baidu Baike and internationalisation/localisation mechanisms in Chinese Wikipedia. The findings show a contrast between the two interaction patterns, indicating mainland-centric versus transnational Chinese gatekeeping processes employing different kinds of collaborative filtering and different levels of user autonomy.

**KEYWORDS:** network gatekeeping, Chinese cultural sphere, online encyclopaedias, Internet politics, Internet censorship.

During an Oxford Union Debate, Jimmy Wales, the co-founder of Wikipedia, cited a Taiwanese Wikipedian as evidence that "the Internet is the greatest force for Democratisation in the World" because she "is talking to the mainland Wikipedians (...) bringing a bit of democracy and information freedom to China."<sup>(1)</sup> While Jimmy Wales was using language in the debate for rhetorical effect, the actual impact of the Internet in democratising authoritarian governments is likely to be very limited. Despite Wales's rhetorical optimism, the possibilities are almost non-existent for a group of Wikipedians to bring down an authoritarian government. Nonetheless, the rise of "collaborative filtering and accreditation," which allows individuals to decide "whom to trust and whose words to question"<sup>(2)</sup> online, points to an information process that may challenge or reinforce Internet filtering and censorship regimes.

The prospect of harnessing the potential of online users to work collaboratively has attracted much research attention, generating concepts such as "peer production,"<sup>(3)</sup> "wikinomics,"<sup>(4)</sup> and "cognitive surplus."<sup>(5)</sup> For user-generated content, users must be recruited and editorial practices must meet certain quality standards. How has the power of Chinese-language Internet users been harnessed in relation to the filtering and censorship regime imposed by the People's Republic of China (hereafter Beijing)?

Users from mainland China have been blocked by Beijing from many websites. In the case of Chinese-language user-generated encyclopaedias, Chinese Wikipedia was blocked in mainland China when it faced no substantial competition, and Baidu Baike was later launched as the only viable alternative for mainland users unable to circumvent the filtering/censorship regime. Table 1 summarises some basic information about Baidu Baike and Chinese Wikipedia. First, Baidu Baike is hosted by a for-profit company in Beijing, Baidu.com, whereas Chinese Wikipedia is hosted by a non-profit

charitable organisation in San Francisco. Although Chinese Wikipedia started as the first and only Chinese-language user-generated encyclopaedia, Baidu Baike nevertheless has a much larger number of entry articles than does Chinese Wikipedia (about ten times more in September 2012). Baidu Baike surpassed Chinese Wikipedia in the number of articles on roughly the third day after its initial launch on 20 April 2006. Third, Baidu Baike supports web pages encoded in the standard of GB-2312, a Chinese national standard that permits only simplified Chinese. In contrast, like all other language versions of Wikipedia, Chinese Wikipedia uses Unicode, an international standard that supports all the languages in the world.

To examine how the power of Chinese-language Internet users has been harnessed in relation to the filtering and censorship regime, this article applies the concept of "network gatekeeping" to analyse Chinese Wikipedia and Baidu Baike as different examples of collaborative filtering projects. The two online encyclopaedias can be regarded as collaborative filtering projects in the sense that they effectively share "word-of-mouth" recommendations on encyclopaedic knowledge and information, which may indicate a shift from knowledge authority (a traditional encyclopaedia publisher) to knowledge authorities (several online encyclopaedia user-con-

1. Oxford Internet Institute, *Oxford Union Debate: Internet and Democratisation*, Part 6, Oxford, 2007, [http://webcast.oii.ox.ac.uk/?view=Webcast&ID=20070518\\_194](http://webcast.oii.ox.ac.uk/?view=Webcast&ID=20070518_194) (accessed on 2 October 2015).
2. Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, Yale University Press, 2006, p. 465.
3. *Ibid.*
4. Don Tapscott and Anthony D. Williams, *Wikinomics: How Mass Collaboration Changes Everything*, New York, Portfolio, 2008.
5. Clay Shirky, *Cognitive Surplus: Creativity and Generosity in a Connected Age*, New York, Penguin, 2010.

**Table 1 – Basic information about Baidu Baike and Chinese Wikipedia**

Features	Baidu Baike	Chinese Wikipedia
Operator	Baidu.com, Inc. (a NASDAQ-listed web services company known for its search engine)	Wikimedia Foundation inc. (a nonprofit charitable organisation)
Physical location	Beijing, China	San Francisco, USA
Number of entries (September 2012)	over 5,3 million	around 527,000
Date of 1 <sup>st</sup> article	20 April 2006	17 November 2002
Date of 100,000 <sup>th</sup> article	23 April 2006*	12 November 2006
Date of 1,000,000 <sup>th</sup> article	16 January 2008	N/A
Language policy	Simplified Chinese Only	Both simplified and traditional Chinese

Note: Based on self- and mutual-description of Baidu Baike and Chinese Wikipedia as entry articles in both encyclopedias.

\*Interpolated value based on Baidu Baike's articles' sequential serial identification numbers.

tributors).<sup>(6)</sup> The term "collaborative filtering" can refer to its original and more technical definition of recommendation algorithms or methods when information recommendation systems are developed.<sup>(7)</sup> The term was later expanded by authors such as Yochai Benkler to refer to the broader enhanced agency of Internet users in public discourse. Since the connotations of the term "collaborative filtering" are thus more celebratory than, say, "Internet filtering" or even "censorship," for the purposes of this article, I decide to apply the concept of network gatekeeping so that both kinds of filtering can be analysed.

In the context of networks, network gatekeeping is given a broader definition of information control process. Unlike traditional gatekeeping, which relies on mainly editorial mechanisms that are often manual, network gatekeeping relies on many other mechanisms that could be open or automated. The general concept of gatekeepers or gatekeeping has been used in various fields, including political science, communication, sociology, management, information science, and law.<sup>(8)</sup> The more updated theory of network gatekeeping has been developed to assist researchers in the identification and analysis of the gatekeeping mechanisms.<sup>(9)</sup> Through the theoretical lens of network gatekeeping, the comparison of Baidu Baike and Chinese Wikipedia should include not only traditional gatekeeping through direct editorial control, but also other mechanisms such as censorship and internationalisation/localisation. Indeed, encyclopaedia editors serve as gatekeepers of knowledge to their readers, while user-generated encyclopaedia editors function as part of the network gatekeeping processes of providing information and knowledge to users.

Thanks to the global connectivity of the Internet, network gatekeeping activities such as editing an encyclopaedia article can be transnational. Yang Guobin has argued that the Internet has contributed to a transnational Chinese cultural sphere that fulfils political functions within and beyond mainland China.<sup>(10)</sup> Building on Tu Wei-ming's notion of "cultural China," Yang has found that the discourses produced online by users in Taiwan, Hong Kong, and North America have to some extent explored the meaning of being Chinese in the global context. Similarly, the network gatekeeping of user-generated encyclopaedias can be transnational. The empirical focus should cover not just mainland China, but also any potential Chinese-language users online who can contribute.

Still, local acceptance of network gatekeeping is also important, and some research has suggested that cultural factors have prevented wider domestic acceptance of Wikipedia in national markets, especially the Chi-

nese and Korean versions.<sup>(11)</sup> This article thus aims to examine how the two websites allow readers of Chinese-language user-generated encyclopaedias to decide "whom to trust and whose words to question."<sup>(12)</sup>

To answer the main question regarding how the power of Chinese-language Internet users has been harnessed in relation to the filtering and censorship regime imposed by Beijing, the article seeks to answer the following more specific questions: How, starting from a pool of Chinese-language users and content online, did Baidu Baike and Chinese Wikipedia begin to develop their core users as editors? What kind of Chinese-language users are implicitly and explicitly included over the Internet? How do their editorial policies and technical practices reflect different gatekeeping mechanisms (defined as "a tool, technology, or methodology used to carry out the process of gatekeeping"<sup>(13)</sup>)?

By answering these empirical questions, researchers can better examine how Chinese-language users are empowered to foster what kind of interactions. I argue that the contrast between the two encyclopaedias can be understood better as an instance of the ongoing cultural political struggles between mainland-centric versus transnational Chinese gatekeeping processes. In other words, different kinds of collaborative filtering and different levels of user autonomy contribute to different Chinese network gatekeeping on the two websites.

- Andrew J. Flanigan, Craig Flanigan, and Jon Flanigan, "Technical Code and the Social Construction of the Internet," in Morton Winston and Ralph Edlbach (eds), *Society, Ethics, and Technology*, Boston, MA, Cengage Learning, 2013, pp. 292-306.
- Michael D. Ekstrand, John T. Riedl, and Joseph A. Konstan, *Collaborative Filtering Recommender Systems*, Boston, Delft, Now Publishers Inc, 2011.
- Monique Jucquois-Delpierre, "Fictional Reality or Real Fiction: How Can One Decide?: The Strengths and Weaknesses of Information Science Concepts and Methods in the Media World," *Journal of Information, Communication & Ethics in Society*, Vol. 5, No. 2/3, 2007, pp. 235-252.
- Karine Barzilai-Nahon, "Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control," *Journal of the American Society for Information Science and Technology*, Vol. 59, No. 9, 2008, pp. 1493-1512.
- Guobin Yang, "The Internet and the Rise of a Transnational Chinese Cultural Sphere," *Media Culture Society*, Vol. 25, No. 4, 2003, pp. 469-490.
- J. P. Shim and J. Yang, "Why Is Wikipedia Not More Widely Accepted in Korea and China? Factors Affecting Knowledge-Sharing Adoption," *Decision Line*, Vol. 40, No. 2, 2009, pp. 12-15; Huijun Suo, "Zhongguo Weiji wangzhan fazhan yuanyin zhi chutan" (The Development of Wiki Sites in China: An Exploratory Study), in *Hexie shehui, gongmin shehui yu dazhong meijie* (Harmonious Society, Civic Society, and Mass Media), Beijing, China University Communication Press, 2007.
- Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, op. cit., p. 465.
- Karine Barzilai-Nahon, "Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control," art. cit.

The focus on network gatekeeping by both users and governments leads to the more general normative question on the role of user autonomy in different online environments. To what extent has user autonomy been established to conduct information gatekeeping collaboratively online? In this regard, user-generated websites provide site observation to examine the interplay of user autonomy and Internet governance because the gatekeeping processes include editorial practices of keeping and removing content and the emerging practices of soliciting user contributions from different parts of the world on the Internet.

The question regarding how user autonomy is framed by varied practices of Internet governance is central to user-generated projects more generally because of their dependence on user contributions. For instance, while both encyclopaedia websites claim that “anyone can edit” freely, further analysis is required on the actual network boundaries of “anyone” and the level of editorial freedom permitted. Some comparative studies on their editorial practices exist,<sup>(14)</sup> but detailed in-depth analysis remains lacking. This article seeks to fill this gap. Direct evidence is drawn from both explicit and implicit rules, which provide editorial and stylistic guidance that serves to structure the content and collaboration of user-generated encyclopaedias.<sup>(15)</sup> To avoid arbitrary interpretation of rules, user testimonies and media commentaries are included whenever needed.

### Power users who are empowered

User-generated encyclopaedias need gatekeepers to process user contributions. Consider the following two observations made by online users:

Any submission by the users of Baidu Baike must go through an internal review process conducted by Baidu employees before it can take effect on the content outcome.<sup>(16)</sup>

There is not really any editorial team inside [Chinese] Wikipedia. The knowledge level of its editors (not teams) is not much better than that of Baidu Baike's. The main reason why the quality of articles appears to be higher in Wikipedia is that low-quality edits are filtered at the first moment.<sup>(17)</sup>

Rejecting bad edits while keeping good ones, as explained in the quotes above, is the primary task for the two major Chinese-language encyclopaedias.<sup>(18)</sup> As quoted above, Baidu Baike uses “an internal review process conducted by Baidu employees” and “low-quality edits are filtered” by Chinese Wikipedia's editors. Such observations seem to confirm what has been reported about Wikipedia projects: for wikis to be run as successful collaborative platforms, certain policies, norms, and technological support are required.<sup>(19)</sup> Editorial powers to contribute and review new edits need to be delegated from website owners/managers to users.

A central contrast between Baidu Baike and Chinese Wikipedia is that salaried Baidu employees do the bulk of the Baike internal review work, whereas the same work for Chinese Wikipedia is done by volunteer user-contributors, a process in which Wikimedia employees rarely intervene.

According to the earliest and subsequent versions of “Baike's Basic Rules,” users must “subordinate completely to the unified management of Baidu Baike” to avoid deletion of their edits and commentaries.<sup>(20)</sup> Thus, by default and in practice, new edits must go through Baidu's internal review, a process that is not transparent to other users. In direct contrast, for Chinese Wikipedia any new edit will, by default, take effect transparently and

go through the open editorial process where decisions of reversion (undoing new edits) or deletion are made. Thus, although both websites support basic functions such as editing, commenting, and edit-history tracking, the actual editorial practices differ in processing new user contributions: Baidu Baike adopts a model of largely internal review (by Baidu employees), and Chinese Wikipedia uses a model of open review (by users).

The contrast is important in the ways in which “power users” are arranged between website employees and normal users, in which the website locations are geopolitically positioned, and in which power users are recruited from different Chinese-speaking regions.

### Power users: Between employees and normal users

Pivotal to the internal hierarchies of the two sites are the “power users”: Baike Kedou 百科蝌蚪 for Baidu Baike and “administrators” for Chinese Wikipedia. Here I use the generic term “power users” loosely to describe a group of users who are given extra powers and privileges beyond what “normal users” have. Note that I do not intend to argue that Baidu Kedou and Chinese Wikipedia administrators play similar or equivalent roles (in fact they do not). I merely aim to contrast how these “privileged” users operate in the power hierarchy of their respective editorial settings.

The power to change rules and select power users is limited to Baidu employees, who exercise several managerial powers. Baidu employees select around 100 users, based on their performance and application, to become members of the “Baike Kedou Group” (*Baike Kedou tuan* 百科蝌蚪团). Although any Baidu Baike user can apply, the selection and review processes are not open. The Kedou-specific rules outline the extra rights and obligations for these power users (see Figure 1): they enjoy an exclusive group platform, a status icon in front of their user ID, individual editing platforms, and other undisclosed powers and benefits,<sup>(21)</sup> including the privilege of

- Han-Teng Liao, “A Webometric Comparison of Chinese Wikipedia and Baidu Baike and Its Implications for Understanding the Chinese-Speaking Internet,” in 9<sup>th</sup> Annual Internet Research Conference: Rethinking Community, Rethinking Place, Copenhagen, 2008; Huijun Suo, “Zhongguo Weiji wangzhan fazhan yuanyin zhi chutan” (The Development of Wiki Sites in China: An Exploratory Study), *art. cit.*
- Brian Butler, Elisabeth Joyce, and Jacqueline Pike, “Don't Look Now, but We've Created a Bureaucracy: The Nature and Roles of Policies and Rules in Wikipedia,” in *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, New York, ACM, 2008, pp. 1101-1110.
- NetEase, “Bi Baidu Baike kaopu de Weiji Baike” (More reliable than Baidu Baike: Chinese Wikipedia), *Wangyi xinwen zhongxin: wangyi shu du* (NetEase News: Data Blog), 7 April 2012, <http://data.163.com/12/0407/03/7UF6PBB00014MTN.html> (accessed on 12 October 2015).
- Chen Zetian, “Baidu Baike ci tiao yu zhongwen Weiji baike tiaomu ge you shenme tedian? Yuanyin shi shenme?” (What are the Distinct Features of Entries of Baidu Baike and those of Chinese Wikipedia? What Makes Them Different?), *Zhihu.com*, 2012, [www.zhihu.com/question/19710565](http://www.zhihu.com/question/19710565) (accessed on 12 October 2015).
- Han-Teng Liao, “Special Speech Zones and Diversity in the Chinese-Written Internet,” in 7<sup>th</sup> Annual Chinese Internet Research Conference (CIRC 2009), Annenberg School for Communication, University of Pennsylvania, Philadelphia, 2009.
- Andrea Forte and Amy Bruckman, “Scaling Consensus: Increasing Decentralization in Wikipedia Governance,” in *Hawaii International Conference on System Sciences, Proceedings of the 41<sup>st</sup> Annual*, 2008, p. 157; Piotr Konieczny, “Governance, Organization, and Democracy on the Internet: The Iron Law and the Evolution of Wikipedia,” *Sociological Forum*, Vol. 24, No. 1, 2009, pp. 162-192.
- Baidu Baike (hereafter in footnotes BB), “Baidu Baike bangzhu” (Baidu Baike Help), *Baidu Search Help Center*, 2006, [http://web.archive.org/web/20060424182803/http://www.baidu.com/search/baike\\_help.html](http://web.archive.org/web/20060424182803/http://www.baidu.com/search/baike_help.html) (accessed on 28 September 2012); BB, “Baidu Baike bangzhu” (Baidu Baike Help), *Baidu Search Help Center*, 2012, [www.baidu.com/search/baike\\_help.html](http://www.baidu.com/search/baike_help.html) (accessed on 12 October 2015).
- BB, “Kedou tuan zongze” (Kedou Group Rules), *Baidu Baike: Policies and Guidelines*, [www.baidu.com/search/baike/odp/main.html](http://www.baidu.com/search/baike/odp/main.html) (accessed on 14 September 2009); BB, “Baike Kedou tuan zhangcheng (2012/04/26)” (Baike Kedou Group Rules [26 April 2012]), *Baike Kedou tuan ba* (Baike Kedou Chatroom), 2012, <http://tieba.baidu.com/p/1556813452> (accessed on 12 October 2015).

**Figure 1 – Baidu Baike team help: Rules outlined for Baïke Kedou members**

? 百科团队帮助	? Baïke Team Help
● 蝌蚪团章程	● Kedou Group Rules
→ 总则	→ General provisions
→ 权利与义务	→ Rights and obligations
● 附录	● Appendix
→ 管理与考核	→ Management and evaluation
→ 资格管理	→ Eligibility management
→ 考核制度	→ Appraisal system
→ 请假制度	→ Leave system
→ 奖惩制度	→ Reward and penalty system
● 团员操作手册	● Manual for team members

"green channels" and receiving monthly and seasonal gifts.<sup>(22)</sup> Thus, the power of the Kedou group of users is still subordinate to Baidu employees. While Kedou members' new edits can normally go through the "green channels" without review by Baidu employees, the rules clearly indicate the hierarchical power relationship.

The Kedou Group Rules (*Kedou tuan zhangcheng* 蝌蚪团章程) detail how Kedou members are evaluated by Baidu employees, who manage the "eligibility management" (*zige guanli* 资格管理), "appraisal system" (*kaoheshidu* 考核制度), "leave system" (*qingjia zhidu* 请假制度), and "reward and penalty system" (*jiangcheng zhidu* 奖惩制度) – as if Kedou members were employees managed by Baidu the company.

In contrast, normal Chinese Wikipedia users can, among themselves, change rules and select power users in a transparent fashion. These power users, or "administrators" for Chinese Wikipedia, once nominated and elected by users themselves via open discussions, have extra powers, including the powers to temporarily ban other users from contributing.<sup>(23)</sup> Unlike the pivotal role played by Baidu employees in Baidu Baike, the employees of Wikimedia Foundation rarely intervene in the day-to-day editorial processes and have no say in electing administrators for Chinese Wikipedia. Hence, acting between the employees of the hosting organisation and fellow users, the power users of Baidu Baike and Chinese Wikipedia differ in the extent, level of transparency, and internal relationship of their given extra powers.

By analysing how an elite group of users is recruited and managed to facilitate "relevance filtration and accreditation,"<sup>(24)</sup> the discussion above shows that the power of Baidu Baike's elite users is limited by Baidu employees. In contrast, normal Chinese Wikipedia users can, among themselves, change rules and select power users in a transparent fashion. The contrasting features of those users with extra privileges, also called "power users," demonstrate different network gatekeeping relationships between gatekeepers and gated. Users of Chinese Wikipedia are the main gatekeepers and gated at the same time, whereas Baidu employees (as the main gatekeepers) have a relationship to user-contributors (as gated) that is mostly unidirectional. Thus, the gatekeeping authority is kept mostly inside Baidu the company, whereas for Chinese Wikipedia the authority is kept among the user-contributors of Chinese Wikipedia rather than by the Wikimedia employees in San Francisco.

The next subsection will discuss the role of the hosting organisations (Baidu and Wikimedia) and their geopolitical context.

## Beijing and San Francisco: the offline context

On the level of hosting organisations, Baidu Baike is hosted by a for-profit company, Baidu.com, in Beijing, China, whereas Chinese Wikipedia is hosted by a non-profit charitable organisation, the Wikimedia Foundation, in San Francisco, USA.

With servers located in Virginia and Florida, Wikimedia Foundation believes that there are "specific advantages to US laws that do not exist anywhere else in the world" and cites the First Amendment of the US Constitution as evidence of strong protection of its projects.<sup>(25)</sup> Thus, even when one US Supreme Court ruling limited the scope of public domain work that frustrated many members of Wikimedia community, the Wikimedia Foundation rejected proposed plans to move some of the data outside of the US.

Since an ICP (Internet Content Provider) license/certificate is required for any website to operate physically in mainland China,<sup>(26)</sup> Baidu has its ICP license under Beijing City. Previous comparative research has shown that Baidu's filtering of search results is much subtler than China's blocking of its users from accessing certain Google results.<sup>(27)</sup> Baidu Baike is one of many services found under the web domain name Baidu.com, whereas Chinese Wikipedia is one of many different language versions of Wikipedia.org. Since Baidu Baike is a service hosted by a search engine company and Chinese Wikipedia is a service hosted by a global encyclopaedia project, they differ in many respects. Financially, although Baidu the company is a profitable NASDAQ-listed company, whether Baidu Baike by itself (whose finance cannot be externally assessed separately from the larger company Baidu) is financially sustainable remains a mystery; Wikipedia is funded by annual fund-raising efforts run by the Wikimedia Foundation. Encyclopaedia pages in Baidu Baike do have ads promoting products and services provided by other companies, whereas Chinese Wikipedia has no ads. Politically, as to be discussed later, Baidu the company has editorial policies and practices of purging content that is deemed politically sensitive by Beijing. In contrast, although users of Chinese Wikipedia must comply with the "Terms of use" drafted by the Wikimedia Foundation<sup>(28)</sup> for all language versions of Wikipedia, the foundation does not engage in similar practices of purging politically sensitive content. Also, the footprint of the hosting organisations on editorial politics differs. In an opaque fashion, Baidu the company unilaterally sets the policies, which expanded from seven subsections on a single page in 2006 to 17 subsections in 2008, not to mention its model of internal editorial review

22. BB, "Baïke Kedou tuan zhangcheng (2012/04/26)" (Baïke Kedou Group Rules [26 April 2012]); BB contributors, "Baïke Kedou tuan" (Baïke Kedou Group), *Baidu Baïke*, <http://baïke.baidu.com/view/881001.htm> (accessed on 30 September 2012); NetEase, "Bi Baidu Baïke kaopu de Weiji Baïke" (More Reliable than Baidu Baïke: Chinese Wikipedia), *art. cit.*

23. Brian Butler, Elisabeth Joyce, and Jacqueline Pike, "Don't Look Now, but We've Created a Bureaucracy," *art. cit.*; Chinese Wikipedia (hereafter in footnotes zhWP), "Weiji Baïke: Guanliyuan" (Wikipedia: Administrators), Official policy page, *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Wikipedia:%E7%AE%A1%E7%90%86%E5%91%98&oldid=24077941> (accessed on 13 December 2012).

24. Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, *op. cit.*, p. 465.

25. Wikimedia Legal Team, "Legal/URAA Statement," *Wikimedia Meta*, 6 February 2013, [https://meta.wikimedia.org/wiki/Legal/URAA\\_Statement](https://meta.wikimedia.org/wiki/Legal/URAA_Statement) (accessed on 12 October 2015).

26. Gabriela Taylor, *Build & Market Your Business with Google: A Step-By-Step Guide to Unlocking the Power of Google and Maximizing Your Online Potential*, Global & Digital, 2013.

27. Min Jiang, "A Comparative Study of Baidu and Google's Search Results of Internet Events from China," *New Media & Society*, Vol. 16, No. 2, 2013.

28. "Shiyong tiaokuan – Meta" (Terms of Use – Meta), *Meta-Wiki*, [https://meta.wikimedia.org/w/index.php?title=Terms\\_of\\_use/zh-hant&oldid=4627551](https://meta.wikimedia.org/w/index.php?title=Terms_of_use/zh-hant&oldid=4627551) (accessed on 22 November 2012).

by its salaried staff. In Chinese Wikipedia, it is the user-contributors themselves who set and enforce editorial policies. Even copyright-violating materials are purged by user-contributors. Hence, Baidu Baike operates in a corporate setting with political oversight by Beijing, whereas Chinese Wikipedia governs itself among its user-contributors.

On the level of the wider geopolitical context, a contrast exists as to whether and how political authorities influence or are involved in the operations of the hosting organisations. As mentioned earlier, the political influence of Beijing on Baidu Baike is evident from multiple sources as regards “internal monitoring and censorship” routines.<sup>(29)</sup> In contrast, there is little or no evidence that the US authorities have exercised influence over the Wikimedia Foundation so as to shape the content on Chinese Wikipedia. On the contrary, there are a few cases where certain US agencies, politicians, and corporations have been exposed (and thus embarrassed) for contributing edits as normal editors.<sup>(30)</sup> Some leaked documents regarding Baidu’s daily internal monitoring and censorship indicate editorial influence on Baidu’s content (including Baidu Baike) by the government authorities in Beijing,<sup>(31)</sup> while there is less of a likelihood that government officials in Washington DC have sought to shape the development of Chinese Wikipedia through the operations of the Wikimedia Foundation. In fact, the foundation has joined the Wikipedia community to protest against US legislation for the Stop Online Piracy Act (SOPA) and the PROTECT IP Act (PIPA). In addition, some Western observers suspect that certain articles in Chinese Wikipedia are undermined by self-censorship, thereby presenting viewpoints sympathetic to Beijing,<sup>(32)</sup> suggesting that the editorial processes in Chinese Wikipedia may be more easily influenced by Beijing than by San Francisco. There are also unverified reports that some Chinese Wikipedia articles are edited by paid online commentators, often known as the “50 Cent Party,” in order to defame pro-democracy opinion leaders.<sup>(33)</sup>

To sum up, the above assessment of the hosting organisations in the network gatekeeping processes presents many contrasts. Baidu Baike allows much less user autonomy than Chinese Wikipedia. Baidu the company has a much larger footprint on Baidu Baike than the Wikimedia Foundation has on Chinese Wikipedia. Also, several sources of evidence indicate the direct influence of the authorities in Beijing on Baidu Baike’s editorial processes, whereas this link of direct institutional influence is missing between the authorities in Washington and the Wikimedia Foundation in San Francisco, and their influence on and intervention in Chinese Wikipedia is found to be limited and inconsequential. In sum, Baidu Baike is influenced by and embedded in the national legal, regulatory, and cultural framework of the People’s Republic of China, while Chinese Wikipedia is more self-governed and transnationally-structured. We can now examine how these geopolitical contexts are reflected in the geographic distribution of power users.

### **Power users from Hong Kong and Taiwan: The major difference**

I further collected and coded self-disclosed data for the years of 2009 and 2012. The disclosure rate for Chinese Wikipedia is a bit higher (over 90%) than for Baidu Baike (over 79%), suggesting that the majority of power users do disclose their place of residence. The findings in Table 2 show how power users are distributed across various regions, including mainland Chinese regions (zh-cn), regions with ethnic majority Chinese populations (zh), and the rest of the world.

As a proxy for active contributors, the regional distribution of Chinese Wikipedia administrators in 2012 also reflects a similar diversity: 25 are from mainland China, 15 from Hong Kong/Macau, 2 from Singapore, and 16 from Taiwan. In contrast, all disclosed power users of Baidu Baike come from mainland China. Figure 2 shows the world map for 2012 data. It shows Baidu Baike’s focus on mainland China and Chinese Wikipedia’s spread across the world, including North America, Western Europe, and Asia Pacific.

Figure 3 further shows greater detail across the Chinese and East Asian regions. Chinese Wikipedia clearly has a significant number of power users not in only the major cities of mainland China, such as Beijing, Guangdong, and Shanghai, but also in Hong Kong and Taiwan (and two in Singapore). In contrast, Baidu Baike’s power users are bounded within mainland China. The absence (versus the presence) of power users from Taiwan, Hong Kong, Macau, and Singapore distinguishes Baidu Baike from Chinese Wikipedia. The reason why Hong Kong and Taiwan constitute the major difference between the two encyclopaedias may be closely linked to the organisation of local chapters. Both Wikimedia Taiwan and Wikimedia Hong Kong were founded and then recognised by the Wikimedia foundation around 2007 and 2008. Both have been active and organised enough in 2007 and 2013 to host Wikimania, an international event held annually since 2005 with all Wikipedians in the world. This is in direct contrast to Wikimedia’s efforts in mainland China. The founder of Wikimedia, Jimmy Wales, visited with China’s censors in Beijing in 2008 and in 2009 with Hudong’s CEO to discuss issues of censorship and Chinese trademark of Wikipedia, showing the tricky and adverse environment Chinese Wikipedia faces in order to grow in mainland China.<sup>(34)</sup>

Different network gatekeeping patterns seem to have different networking effects across Chinese-speaking regions. The findings above suggest that Chinese Wikipedia is an example of what Yang Guobin described as the rise of a “transnational Chinese cultural sphere,”<sup>(35)</sup> whereas Baidu Baike has gained some domestic acceptance within the national market of mainland China by operating in a corporate setting with political oversight by Beijing. The transnational versus national contrast is not unexpected, because Wikipedia offers the space for different language versions and al-

29. User:430072, “Talk: Baidu Baike” (Talk:Baidu Baike), Discussion page, *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Talk:%E7%99%BE%E5%BA%A6%E7%99%BE%E7%A7%91&oldid=22100016> (accessed on 23 July 2012); Eva Woo, “Baidu’s Censored Answer to Wikipedia,” *BusinessWeek*, 13 November 2007, [www.businessweek.com/globalbiz/content/nov2007/gb20071113\\_725400.htm](http://www.businessweek.com/globalbiz/content/nov2007/gb20071113_725400.htm) (accessed on 12 October 2015); Xiao Qiang, “Baidu’s Internal Monitoring and Censorship Document Leaked (1) (Updated),” *China Digital Times*, 30 April 2009, <http://chinadigitaltimes.net/2009/04/baidu-internal-monitoring-and-censorship-document-leaked> (accessed on 12 October 2015).
30. Katie Hafner, “Seeing Corporate Fingerprints in Wikipedia Edits,” *The New York Times*, 19 August 2007, sec. Technology, [www.nytimes.com/2007/08/19/technology/19wikipedia.html](http://www.nytimes.com/2007/08/19/technology/19wikipedia.html) (accessed on 12 October 2015); Jonathan Fildes, “Wikipedia ‘Shows CIA Page Edits,’” *BBC*, 15 August 2007, sec. Technology, <http://news.bbc.co.uk/2/hi/technology/6947532.stm> (accessed on 12 October 2015).
31. Xiao Qiang, “Baidu’s Internal Monitoring and Censorship Document Leaked (1) (Updated),” *art. cit.*
32. Howard W. French, “China News: Chinese-Language Wikipedia Presents Different View of History,” *The New York Times*, 29 November 2006, <http://chinadigitaltimes.net/2006/11/chinese-language-wikipedia-presents-different-view-of-history-howard-w-french> (accessed on 12 October 2015).
33. Oiwan Lam, “Hong Kong: Battle against 50 Cents at Wikipedia,” *Global Voices*, 27 October 2012, <http://advocacy.globalvoicesonline.org/2012/10/27/hong-kong-battle-against-50-cents-at-wikipedia> (accessed on 12 October 2015).
34. Lara Farrar, “It’s Tricky for Wikis and Online Encyclopedias in China,” *CNN.com/technology*, 15 October 2009, <http://edition.cnn.com/2009/TECH/10/14/wiki.china>, (accessed on 12 October 2015).
35. Guobin Yang, “The Internet and the Rise of a Transnational Chinese Cultural Sphere,” *op. cit.*

Table 2 – Geographic distribution of the power users

	Baidu Baike		Chinese Wikipedia	
	2012	2009	2012	2009
<b>Total</b>	97	99	78	78
<b>Disclosure ratio</b>	79.38%	70.71%	92.31%	97.44%
<b>Mainland China (Zh=cn)</b>	77	69	25	24
Beijing	10	3	7	6
Shanghai	4	6	5	3
Guangdong	8	7	4	4
Other provinces	55	53	9	11
<b>Zh Regions (Zh)</b>	0	0	33	37
Hong Kong	0	0	13	16
Macau	0	0	2	2
Taiwan	0	0	16	18
Singapore	0	0	2	1
<b>Asia Pacific</b>	0	1	1	2
Japan	0	1	0	0
Australia	0	0	1	2
<b>Americas</b>	0	0	10	11
United States	0	0	9	8
Canada	0	0	1	3
<b>Europe</b>	0	0	3	2
Germany	0	0	1	1
United Kingdom	0	0	1	0
Sweden	0	0	1	1
<b>Missing or undisclosed</b>	20	29	6	2

Source: data collected by the author.

lows users to switch from one language version to another. Nonetheless, it is noted that Hong Kong and Taiwan have been somehow excluded from Baidu Baike. The differentiating factor of Hong Kong and Taiwan is further reinforced by the following findings on the configuration of geolinguistic factors, which in many ways relate to the scope and definition of Chinese network gatekeeping.

### Geolinguistic arrangements

Baidu Baike effectively excludes users of traditional Chinese characters from contributing by using simplified Chinese only with the encoding standard GB-2312, whereas Chinese Wikipedia uses not only the Unicode standard that accommodates all world languages, but has also developed a multiple script-writing system and set of policies, including the "Avoid-Region-Centrism Policy,"<sup>(36)</sup> to avoid any Chinese regional-centric, Han-centric, and Chinese-centric statements.

Localisation (L10n), the process of adapting computer software or information systems for a group of users usually defined by national boundaries or geolinguistic profiles,<sup>(37)</sup> can also be discussed under the rubric of contributing to the "internationalisation mechanisms" of network gatekeep-

ing.<sup>(38)</sup> Since what constitutes legitimate "localisation" in a Chinese-language context is highly contested among media scholars,<sup>(39)</sup> I have chosen the term "geolinguistic" factors and their configuration as a relatively neutral term for this part of the geolinguistic analysis of media and Internet research.<sup>(40)</sup> The findings below provide plausible answers as to why the major differences in the geolinguistic profile of power users lie in the pres-

36. *Ibid*; User:Lorenzarius and other contributors, "Weiji Baike: bimian diyu zhongxin" (Wikipedia: Avoid-Regional-Centrism), Official policy page. *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Wikipedia:%E9%81%BF%E5%85%8D%E5%9C%B0%E5%9F%9F%E4%B8%AD%E5%BF%83&oldid=14594312> (accessed on 8 October 2010).

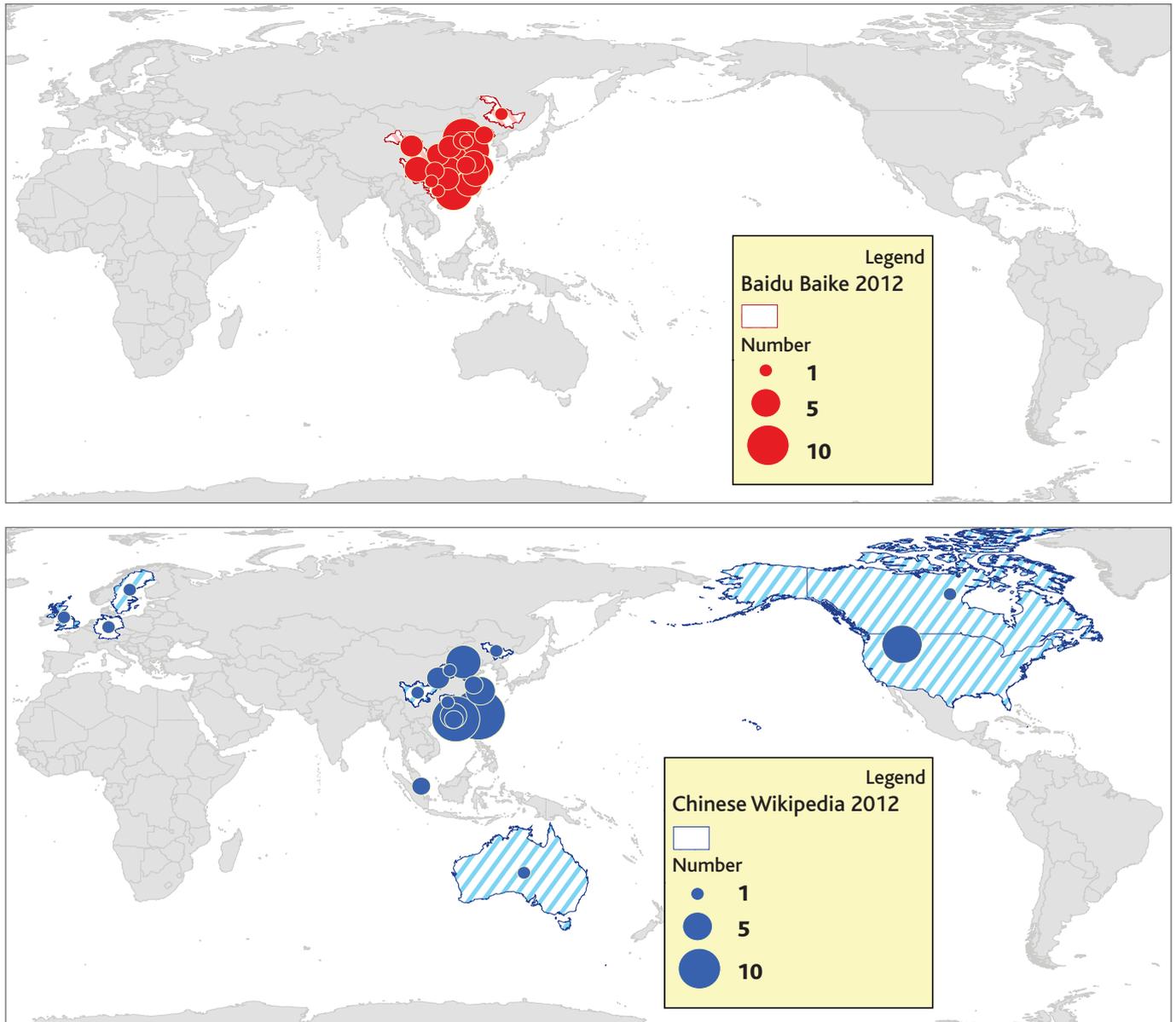
37. Han-Teng Liao, *Needing to Have a Voice: Linguistic Grouping in the Digital Networked Environment*, ISD Working Papers in New Diplomacy, Washington, DC, Institute for the Study of Diplomacy, Georgetown University, July 2011; Michael G. McKenna and Henry Naftulin, "Challenges in the Multicultural HCI Development Environment," in *CHI '00 Extended Abstracts*, New York, ACM, 2000, p. 362.

38. Barzilai-Nahon, "Toward a Theory of Network Gatekeeping," *art. cit.*

39. Huang Xingmin, "Huaxia chuanbo yanjiu chuyi" (An Introduction to Huaxia Communication Studies), *Xinwen yu chuanbo yanjiu* (Journalism & Communication), No. 4, 2002, <http://dspace.xmu.edu.cn/dspace/handle/2288/7838> (accessed on 15 October 2015).

40. Han-Teng Liao and Thomas Petzold, "Analysing Geo-Linguistic Dynamics of the World Wide Web: The Use of Cartograms and Network Analysis to Understand Linguistic Development in Wikipedia," *Cultural Science*, Vol. 3, No. 2, 2010; John Sinclair, "Culture and Trade: Some Theoretical and Practical Considerations," in Emile G. McAnany and Kenton Todd Wilkinson (eds), *Mass Media and Free Trade*, Austin, University of Texas Press, 1996, p. 444.

Figure 2 – Distribution of power users of Baidu Baike and Chinese Wikipedia in 2012



Source: data collected by the author. © Han-Teng Liao

ence versus the absence of Hong Kong and Taiwan: this difference is mainly caused by practical technical arrangements for collaborative editing.

### **Language encoding: Baidu Baike excludes traditional Chinese users**

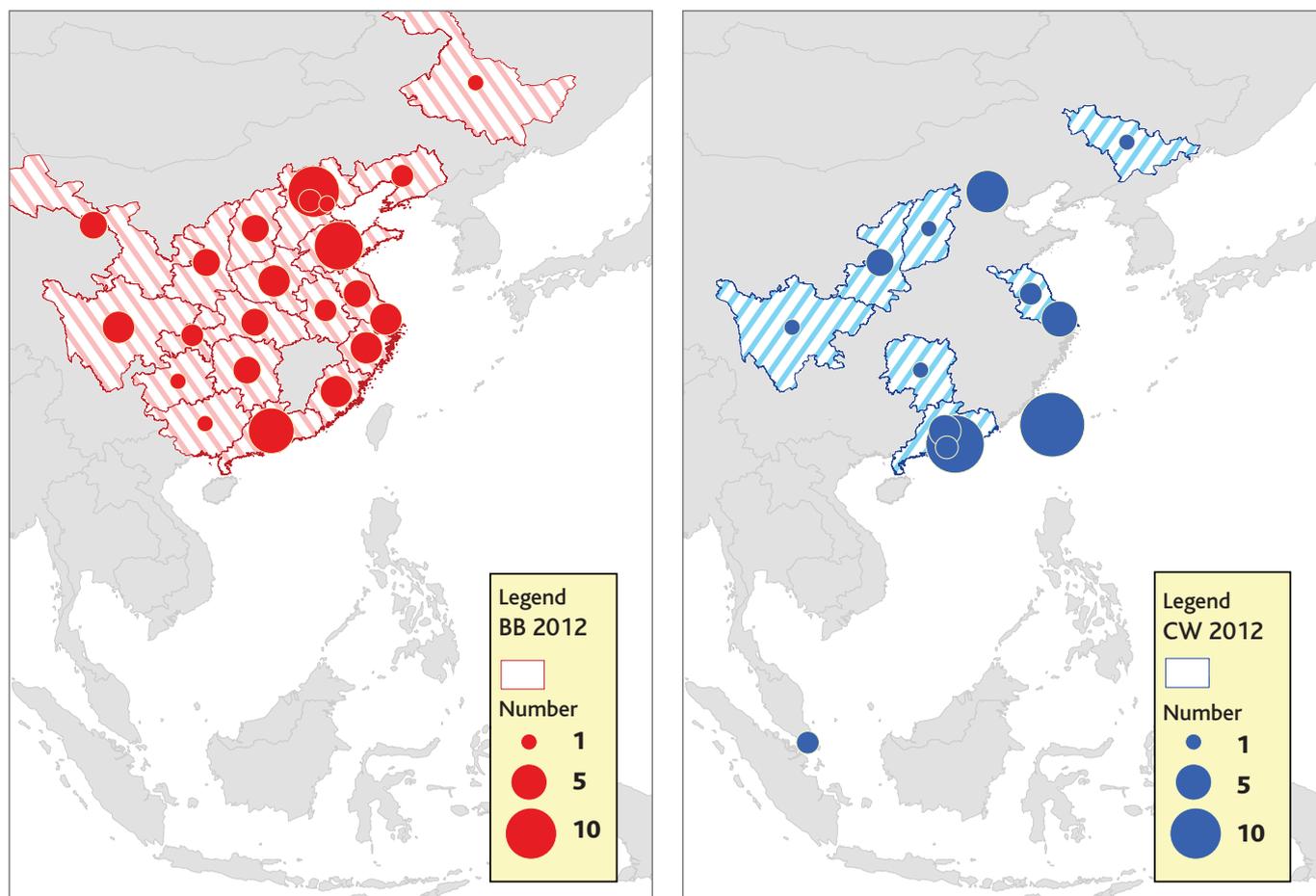
By using simplified Chinese only with the encoding standard GB-2312, Baidu Baike effectively excludes users of traditional Chinese characters from contributing, whereas Chinese Wikipedia and other websites (including Baidu Japan) use the Unicode encoding standard, which can accommodate all the languages of the world.

The choice of language encoding is significant for both the wider context of Chinese-language Internet and the specific context of these two competing online encyclopaedias. The development of language encoding standards and the later widespread adoption of the Unicode encoding standard have been essential elements for digital networking technologies to be diffused in East Asian regions, including Japan and China. Indeed, be-

fore the all-inclusive Unicode was developed fully enough for wide adoption, traditional Chinese content was often encoded in Big5, a standard set by the computing industry in Taiwan and Hong Kong, and simplified Chinese content as encoded in GB-2312, a national standard set by authorities in Beijing. These two standards are not compatible, and thus for both traditional and simplified Chinese content to coexist, the global Unicode standard or compatible Chinese national standard of GB18030 must be used. In other words, the wider context of Chinese-language Internet has experienced a history of competing incompatible language standards before Unicode and Unicode-compatible standards were widely adopted. Thus, the decision to use a certain language-encoding standard has geopolitical significance, cultural, and political significance.

For the specific context of the two competing online encyclopaedias, a brief historical review shows that Chinese Wikipedia was among the early adopters of Unicode, while Baidu Baike adopted the simplified-Chinese-only standard of GB-2312 even though Baidu the company has both capacity for and experience with using Unicode.

**Figure 3 – Distribution of power users of Baidu Baike (BB) and Chinese Wikipedia (CW) in 2012: East and South-East Asia**



Source: data collected by the author. © Han-Teng Liao

The early decision by Chinese Wikipedia to implement Unicode support can be traced back to the year 2002. In contrast, since its launch for a time, Baidu Baike used only GB-2312 for all of its article pages. For Chinese Wikipedia, the first documented user from mainland China, user:Mountain, used Unicode standard in 2002 to solve the issue of processing Chinese characters for the then-one-year-old global Wikipedia project, thereby creating the first Chinese-language article. It took around ten months for user:Mountain and other users to localise and translate the English-based software and documents, thereby providing fundamental components for Chinese Wikipedia at the domain name of zh.wikipedia.org. It was a significant cutting-edge move by the volunteers to globalise Wikipedia: this becomes clear when we consider that at the same time, in 2002, even popular commercial software such as Microsoft Outlook 2002 was struggling to provide Unicode support for multilingual solutions. Judging from my own computer science training based in Taiwan, Chinese Wikipedia has been a leading website in Unicode support in the Chinese-language digital world, where most of the other websites use either Big5 or GB2312.<sup>(41)</sup> As a result, as early as 2002, Chinese Wikipedia had the capacity to process both types of Chinese writing scripts on the same page.

In contrast, when Baidu launched Baidu Baike in 2006, its choice of the GB-2312 encoding standard effectively precluded traditional Chinese users (mainly from Hong Kong and Taiwan) from participating. Since Baidu provides search services that include traditional Chinese content on the Web, it is unlikely that Baidu lacks the technical capacity to accommodate tra-

ditional Chinese characters in its encyclopaedia. In fact, it remains a telling decision that Baidu supports Unicode for its search engine and input method services for Japanese users,<sup>(42)</sup> but it does not support Unicode for Baidu Baike users.

In addition to the choice of encoding standard, Chinese Wikipedia has established a conversion mechanism and editorial rules to ensure it does not preclude certain groups of Chinese-language users as Baidu Baike does, as will be detailed in the following paragraphs.

### **Language platform: Chinese Wikipedia develops a multiple script-writing system**

Collaborative editing implies the integration of different writing systems used by contributors. Although the Wikipedia platform is able to support all the languages in the world with the Unicode standard, there are multiple separate language versions of Wikipedia that are run by different communities. Thus, an issue arises: should there be a single version or multiple separate versions of Chinese Wikipedia? For the modern Chinese language, the gradual decision to merge the two originally separate versions (one

41. Han-Teng Liao, "Conflict and Consensus in the Chinese Version of Wikipedia," *IEEE Technology and Society Magazine* 28, No. 2, 2009, pp. 49-56; John Yunker, *Beyond Borders: Web Globalization Strategies*, Indianapolis, New Riders, 2002.

42. MasPoster, "Baidu Japan Releases Japanese Input Method Beta," *Eming.com*, 17 December 2009, <https://web.archive.org/web/20100123172340/http://eming.com/en/baidu-japan-releases-japanese-input-method-beta> (accessed on 15 October 2015).

Figure 4 – Screenshot of the entry “Taxi” without the popup box



© Wikipedia Authors, <https://zh.wikipedia.org/wiki/出租車> (accessed on 21 November 2015).

traditional, the other simplified) into a single version was reached around 2003 or 2004.<sup>(43)</sup> Users of different Chinese language-scripts now collaborate on the same version of articles.

After the decision to merge, a language platform emerged to accommodate simplified Chinese users and traditional Chinese users on an equal footing. It mixed different contributed texts on the back stage but presented the overall content in several options on the front stage. This process began as a working automatic Chinese-Chinese conversion system in late 2004, which converted the content between the two language scripts: traditional Chinese and simplified Chinese. It subsequently evolved into a system that formally recognised four Chinese geolinguistic regions: Mainland-simplified, Hong Kong-Macau-traditional, Taiwan-orthodox, and Singapore-Malaysia-simplified. Basically, the platform's front end provided user-readers with several geolinguistic options so as to present the content consistently in a fashion that matched readers' geolinguistic settings or choices. At its back end, since contributed texts came from individual editors' preferred language scripts, the diversity of scripts was preserved. Thus, it was common for an article to contain mixed Chinese-script content. Contributing and reading content were thus two different experiences: the front-end minimised the chances of annoying user-readers with unfamiliar characters or terms, while the back end showed basic respect for original user-contributors' choices. This way, the platform bridged the gap among Chinese language variants. It can be seen that the processability of language is intimately bound up with the editorial platform: editing is not just about content, but also about how the content in each script is expressed and represented. This is why the discussion of network gatekeeping mechanisms of internationalisation/localisation is crucial.

Figure 4 shows a typical Chinese Wikipedia page where readers can choose among five available options (from the third tab downwards): Mainland-Simplified, Hong Kong-Traditional, Macau-Traditional, Singapore/Malaysia-Simplified, and Taiwan-Orthodox. These options correspond to different geo-linguistic codes or identifiers: zh-cn (Mainland-Simplified), zh-hk (Hong Kong-Traditional), zh-mo (Macau-Traditional), zh-sg (Singapore-simplified), zh-my (Malaysia-Simplified), and zh-tw (Taiwan-Ortho-

dox). Note that the actual entry title stored in the Wikipedia database is the term “出租车” (zh-cn). Nevertheless, the content will be presented differently according to the preference or choice of the user-readers: *chuzuche* 出租车 (mainland China), *jichengche* 計程車 (Taiwan), *dishi* 的士 (Hong Kong and Macau), or *deshi* 德士 (Singapore and Malaysia), as displayed in Figure 4. For user-readers, a familiar linguistic environment is thus provided, and they also have the option to switch between formats.

The experience of dealing with multiple scripts in the same language version of Wikipedia has been shared and discussed globally since the Wikimania conference in 2005 in Frankfurt, the global annual event for Wiki-projects including Wikipedia, where it has been reported that Arabic, Malaysian, and Indonesian language users have also expressed interest in this issue.<sup>(44)</sup> These Chinese-language users who helped to develop the solution, mostly from mainland China, have contributed their computer codes back to the free software project on which Wikipedia has been based. At the time of this writing, there are several ongoing implementations and discussions on adopting this platform for other language projects in languages such as Serbian, Kazakh, Kurdish, Tajik, Uzbek, Gan Chinese, Kyrgyz, Uyghur, Chechen, etc., which use more than one writing script or system.<sup>(45)</sup> This successful

43. zhWP, "Weiji Baike: guanyu zhongwen Weiji Baike" (Wikipedia: About Chinese Wikipedia), *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Wikipedia:%E5%85%B3%E4%BA%8E%E4%B8%AD%E6%96%87%E7%BB%B4%E5%9F%BA%E7%99%BE%E7%A7%91&oldid=2265206> (accessed on 17 July 2006); zhWP, "Weiji Baiketoolun qu: fanjian chuli/dang'an 6" (Wikipedia Talk: Processing Traditional-Simplified/Archive6), *Chinese Wikipedia*, [http://zh.wikipedia.org/w/index.php?title=Wikipedia\\_talk:%E7%B9%81%E7%AE%80%E5%A4%84%E7%90%86/%E6%A1%A3%E6%A1%88&oldid=11582876](http://zh.wikipedia.org/w/index.php?title=Wikipedia_talk:%E7%B9%81%E7%AE%80%E5%A4%84%E7%90%86/%E6%A1%A3%E6%A1%88&oldid=11582876) (accessed on 28 November 2012); enWP, "Chinese Wikipedia," *English Wikipedia*, [https://en.wikipedia.org/w/index.php?title=Chinese\\_Wikipedia&oldid=582421264](https://en.wikipedia.org/w/index.php?title=Chinese_Wikipedia&oldid=582421264) (accessed on 19 November 2013); zhWP, "Zhongwen Weiji Baike" (Chinese Wikipedia), *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=%E4%B8%AD%E6%96%87%E7%BB%B4%E5%9F%BA%E7%99%BE%E7%A7%91&oldid=8919479> (accessed on 28 December 2008).

44. zhWP, "Weiji Baike: 2005 nian Weiji Meiti guoji dahui" (Wikipedia: Wikimania 2005 Frankfurt), *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Wikipedia:2005%E5%B9%B4%E7%B6%AD%E5%9F%BA%E5%AA%92%E9%AB%94%E5%9C%8B%E9%9A%9B%E5%A4%A7%E6%9C%83&oldid=4625510> (accessed on 11 July 2007).

45. Wikimedia Foundation, "Wikipedias in Multiple Writing Systems," *Meta-Wiki, the Global Community Site for the Wikimedia Foundation's Projects and Related Projects*, [https://meta.wikimedia.org/w/index.php?title=Wikipedias\\_in\\_multiple\\_writing\\_systems&oldid=4131910](https://meta.wikimedia.org/w/index.php?title=Wikipedias_in_multiple_writing_systems&oldid=4131910) (accessed on 15 September 2012).

story mirrors that of Unicode, whereby the global development to accommodate Chinese characters led the way to accommodating other languages.<sup>(46)</sup> In sum, the language platform developed by users of Chinese Wikipedia, which supports multiple writing scripts with inter-script conversion capacity, has become a model of emulation and adoption for other language versions of Wikipedia.

Chinese characters are bound up with questions about Chinese modernity, the Cold War, and ongoing geolinguistic differences across Chinese-speaking regions.<sup>(47)</sup> Such tensions are alleviated by the multi-script platform of Chinese Wikipedia. No Chinese characters are "better" or "more correct" than others. I have reported elsewhere that it is arguably one of the most advanced platforms among Chinese-language websites in tackling the inter-script conversion thanks to the fact that conversion mapping tables are constantly updated by the user-contributors of Chinese Wikipedia.<sup>(48)</sup> While other types of conversion software, such as the one provided with the Chinese version of Microsoft Office, are limited to the pre-determined orthographic and lexical conversion tables, the conversion tables maintained in Chinese Wikipedia evolve through ongoing user contributions. The language platform is more than just the usual simplified-complex conversion; it feeds on the collective intelligence from user-contributors everywhere, making it possible for them to work together on the same article. No specific script is treated as superior to any other.

To sum up, the two encyclopaedias deal with geolinguistic variants of Chinese differently: the language encoding and platform of Baidu Baike implicitly accepts only simplified Chinese characters, whereas those of Chinese Wikipedia explicitly accommodate various Chinese variants on equal footings using various internationalisation mechanisms of localisation.

### Regional considerations: editorial policies

To further maintain a balance so as not to favour one writing system over the other, several editorial policies and norms have been adjusted to take regional differences into account. Among these, the "Avoid-Region-Centrism Policy" (*bimian diyu zhongxin* 避免地域中心) is arguably the most significant.<sup>(49)</sup> It essentially mandates that mainland China-centric (also Hong Kong-centric, Macau-centric, Taiwan-centric, Singapore-centric, etc.), Han-centric, and Chinese-centric statements should be avoided by emphasising that Chinese Wikipedia is a project written in Chinese, and not just a project of, by, and for Chinese people. Other policies, such as "Wikipedia: Vandalism," also take regional variants into account because naming, translation, transliteration, and writing conventions may differ from region to region. For instance, if a user-contributor changes others' contributions from one language script to another, it is counted as an act of vandalism. Inclusiveness is thus ensured.

In contrast, the absence of such geolinguistic considerations in Baidu Baike is expected to discourage participation by user-contributors from Hong Kong and Taiwan. This helps to explain the aforementioned differences in the geolinguistic distribution of power users of Baidu Baike and Chinese Wikipedia. Although actual data on the geolinguistic profiles of all user-contributors of Baidu Baike and Chinese Wikipedia remains a research challenge, the data on the top 100 user-contributors<sup>(50)</sup> shows consistently that non-mainland user-contributors are rare in Baidu Baike and user-contributors in Chinese Wikipedia are mixed. Thus, the links are clear between the geolinguistic outcome and the ways in which Chinese characters are configured with encoding standards by Web platforms: Baidu Baike's sim-

plified-Chinese-only configuration precludes participation by traditional Chinese users, which happen to be the majority of users in Taiwan and Hong Kong, while Chinese Wikipedia's multi-script support allows contribution from different regions.

As an ongoing discussion in Chinese Studies, the transnational aspects of Chinese people, media, and political institutions often surrounding the term "Greater China" after the Cold War are essential to understanding the modern dynamics of Chinese nationalism, regionalism, and transnationalism.<sup>(51)</sup> Instead of accepting the notion that Baidu Baike is Chinese and Chinese Wikipedia is American or Western, the evidence presented above suggests two notions of Chineseness: Baidu Baike embodies the Beijing or mainland China-centric point of view, whereas Chinese Wikipedia reflects an instance of multi-centric integration of Chinese transnationalism.

Since it has been reported that Chinese authorities in Beijing have rejected the adoption of the term "Greater China" on the grounds that it would put Hong Kong and Taiwan on equal footings with mainland China,<sup>(52)</sup> it can be speculated that the way Chinese Wikipedia recognises four separate but equal geolinguistic regions will not be well-liked by Chinese authorities. Here, the ambiguity of what counts as transnational and what counts as national is evident. The Chinese authorities continue to claim that Hong Kong and Taiwan are "part of China" and thus belong to the Chinese nation, whereas Hong Kong and Taiwan seem to have gained and maintained some equal status with mainland China in cultural-political aspects. For example, their preferred language script (orthodox or traditional Chinese) is on par with Beijing's simplified Chinese.

Altogether, Chinese Wikipedia has more "internationalisation mechanisms" of network gatekeeping<sup>(53)</sup> to allow for collaborative editing across

46. Christoph Burgmer, "Python, Unicode and the Digital Divide," *Christoph's CJK-Centered Concerns*, <http://cбургmer.nfshost.com/content/python-unicode-and-digital-divide> (accessed on 6 August 2009); Kam-Fai Wong, Wenji Li, and Ruifeng Xu, *Introduction to Chinese Natural Language Processing*, San Rafael, Morgan & Claypool, 2009.
47. Han-Teng Liao, "Conflict and Consensus in the Chinese Version of Wikipedia," *art. cit.*; John Yunker, *Beyond Borders: Web Globalization Strategies*, Indianapolis, New Riders, 2002.
48. Han-Teng Liao, "Conflict and Consensus in the Chinese Version of Wikipedia," *art. cit.*
49. User:Lorenzarius and other contributors, "Weiji Baike: bimian diyu zhongxin" (Wikipedia: Avoid-Regional-Centrism), *art. cit.*
50. BB, "Gongxian bang" (Ranking of User Contribution), *Baidu Baike Contribution*, <http://baike.baidu.com/star/contribution/grow.html?func=congrow> (accessed on 5 December 2012); User:Emijrp, "List of Wikipedians by number of edits," *Chinese Wikipedia*, [http://zh.wikipedia.org/w/index.php?title=User:Emijrp/List\\_of\\_Wikipedians\\_by\\_number\\_of\\_edits&oldid=22141797](http://zh.wikipedia.org/w/index.php?title=User:Emijrp/List_of_Wikipedians_by_number_of_edits&oldid=22141797) (accessed on 4 December 2012).
51. William A. Callahan, "Nationalism, Civilization and Transnational Relations: The Discourse of Greater China," *Journal of Contemporary China*, Vol. 14, No. 43, 2005, pp. 269-289; Joseph Man Chan, "Toward TV Regionalization in Greater China and Beyond. TV China: Institutions, Programming, and Audiences in Greater China and the Chinese Diaspora," in Ying Zhu and Christopher Berry (eds), *TV China*, Bloomington, Indiana University Press, 2009; Guo Sujian and Guo Baogang, *Greater China in an Era of Globalization*, Lanham, Rowman & Littlefield, 2010; Christopher Hughes, "Controlling the Internet Architecture Within Greater China," in Françoise Mengin (ed.), *Cyber China: Reshaping National Identities in the Age of Information*, Basingstoke, Palgrave Macmillan, 2004, pp. 71-90; Chin-Chuan Lee, "Rethinking Political Economy: Implications for Media and Democracy in Greater China," *Javnost-The Public*, Vol. 8, No. 4, 2001, pp. 81-102; Rex Li, "China in Transition: Nationalism, Regionalism and Transnationalism," *Contemporary Politics*, Vol. 3, No. 4, 1997, pp. 365-380; Françoise Mengin, "The Changing Role of the State in Greater China in the Age of Information," in Françoise Mengin (ed.), *Cyber China: Reshaping National Identities in the Age of Information*, *op. cit.*; Ngai-Ling Sum, "Cyber-Capitalism and the Remaking of Greater China," in Françoise Mengin (ed.), *Cyber China: Reshaping National Identities in the Age of Information*, *op. cit.*, pp. 205-236; Kenneth C.C. Yang, "A Comparative Study of Internet Regulatory Policies in the Greater China Region: Emerging Regulatory Models and Issues in China, Hong-Kong SAR, and Taiwan," *Telematics and Informatics*, Vol. 24, No. 1, February 2007, pp. 30-40.
52. Henry Yuhuai He, "Da Zhonghua" (Greater China), in *Zhonghua renmin gongheguo zhengzhi wenhua yongyu dadian* (Dictionary of the Political Thought of the People's Republic of China), Armonk, M.E. Sharpe, 2001.
53. Barzilai-Nahon, "Toward a Theory of Network Gatekeeping," *art. cit.*

major Chinese-speaking regions beyond mainland China. In addition, the fact that users of Chinese Wikipedia have developed, on top of the international technical encoding standards, a multiple script-writing system, and an editorial "Avoid-Region-Centrism Policy," highlights the high level of user autonomy of Chinese Wikipedia. Users themselves can develop both new editorial and technical means to facilitate network gatekeeping processes that are more international.

### Editorial priorities

Additional evidence suggests that the contrast between the editorial priorities of Baidu Baike and Chinese Wikipedia may result from the absence versus presence of users from Hong Kong and Taiwan.

### Copyright-dubious materials

Based on evidence from various sources, Baidu Baike's internal review is lax in screening copyright-dubious materials, whereas Chinese Wikipedia's open review is stricter. Indeed, since its launch in 2006, Baidu Baike has been reported to be the "worst copyright violator" of Wikipedia's copyright.<sup>(54)</sup> This scale of violation is also corroborated by evidence gathered by Chinese Wikipedia users, including obvious searchable digital traces of English and Chinese signature phrases, such as finding "From Wikipedia, the free encyclopaedia" and "citation needed" in the content of Baidu Baike, and a sample copyright complaint letter has been used to send complaints to Baidu.<sup>(55)</sup> Based on copying content from Chinese Wikipedia, Baidu Baike quickly (within three days) overtook Chinese Wikipedia in terms of the number of entries, raising the issue of unfair competition.<sup>(56)</sup>

The fact that Baidu Baike copies Chinese Wikipedia highlights Wikipedia's copyleft norm for Wikipedia in content sharing. The copyleft licenses demand that any derived works must be released under compatible copyleft licenses to ensure further sharing. Since I was myself one of the copyleft advocates in Taiwan and the person who came up with a Chinese translation of the term "copyleft," I agree with the legal opinions given by some Chinese Wikipedians that Baidu Baike has violated the copyright of Wikipedia by refusing to remove the copyright symbol (e.g., "©2012 Baidu") found at the bottom of each article page.<sup>(57)</sup> Baidu Baike can legally copy any content, in full or selectively, from Wikipedia, but when doing so it must give attribution to Wikipedia and also release the final content of Baidu Baike with a copyleft license instead of claiming ownership as if it were the property of Baidu. This violation prompted some Chinese Wikipedia contributors to declare, as a gesture of protest, that "any use of this image by Baidu Baike is not permitted" on their contributed images.<sup>(58)</sup>

Still, it should be noted that "copy and paste" activities are more common among users around the globe than a basic knowledge about copyleft practices. As reported by user-contributors, at least 10% of newly added entries to Chinese Wikipedia also include copyright-dubious materials, mostly contributed by new users. A task force has been organised to prevent such copyright violations.<sup>(59)</sup> In addition, copyright violation constitutes one possible basis for article deletion (in an open review by user-contributors), along with promotional content and advertisements.<sup>(60)</sup> In contrast, for Baidu Baike to review copyright violation complaints, copyright owners must send the complaint via snail mail to Baidu's address in Beijing

for internal review.<sup>(61)</sup> It has been reported by contributors of Chinese Wikipedia that Baidu produces no substantial reply to the copyright-infringement complaints that Chinese Wikipedia has filed against Baidu.<sup>(62)</sup>

### Politically sensitive materials and comments

This lax/strict contrast is reversed when the two encyclopaedias deal with politically sensitive materials and comments. Baidu Baike has a record of censoring content, while Chinese Wikipedia has an explicitly stated policy that "Wikipedia is not censored."<sup>(63)</sup> Effectively in collaborative editorial decisions, user autonomy in Chinese Wikipedia is protected from political censorship.

In terms of politically sensitive materials and comments, Baidu Baike rejects any content that may offend the Chinese authorities in Beijing, whereas, again, Wikipedia's policy is not to censor anything. Baidu the company has defined seven categories for deletion from Baidu Baike.<sup>(64)</sup> Among them, the third category refers to "reactionary content" (*fandong neirong* 反动内容), an ideologically and politically loaded term in the context of mainland China, which means any content that "maliciously criticises the current system of the state," "disrupts social and public order," "provokes disputes over nationalities, ethnicities, religions, and regions," "maliciously attacks state agencies and officials," "promotes superstition and cults," or "provides any hyperlinks to the aforementioned content," all of which are in line with Beijing's priorities in policing and censoring online content (though not specifically explained in the website's rules, the "state" refers to Beijing). In contrast, Chinese Wikipedia has no such policies. Instead, it maintains that it is a "Chinese-language" encyclopaedia that serves users around the world beyond just mainland China and Taiwan, and user-contributors must be vigilant against any specific regional centric or Han-Chinese-centric viewpoints when paraphrasing sources.<sup>(65)</sup>

54. Dan Nystedt, "Baidu May Be Worst Wikipedia Copyright Violator," *IDG News Service*, <http://www.pcworld.com/article/id,135550-page,1/article.html> (accessed on 6 August 2007).
55. zhWP, "Weiji Baike: Baidu Baike dui Weiji Baike de qinquan" (Wikipedia: Baidu Baike Violates Wikipedia's Copyright), *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Wikipedia:%E7%99%BE%E5%BA%A6%E7%99%BE%E7%A7%91%E5%B0%8D%E7%B6%AD%E5%9F%BA%E7%99%BE%E7%A7%91%E7%9A%84%E4%BE%B5%E6%AC%8A&oldid=23751469> (accessed on 19 November 2012).
56. User:430072, "Talk: Baidu Baike", *art. cit.*
57. zhWP, "Weiji Baike: Baidu Baike dui Weiji Baike de qinquan" (Wikipedia: Baidu Baike Violates Wikipedia's Copyright), *art. cit.*
58. User:ljq513, "Zhe yuannian you duo shen a: 'ben tupian bu yunxu Baidu Baike shiyong'" (How Deep the Resentment is This: "Any Use of This Image by Baidu Baike is Not Permitted"), *Baidu Tieba*, <http://tieba.baidu.com/p/1742518344> (accessed on 22 July 2012).
59. zhWP, "Weiji Baike: qinquan zhengjiu gongzuo xiaozu" (Wikipedia: Copyright Violation Remedy Taskforce), *Chinese Wikipedia*, <http://zh.wikipedia.org/w/index.php?title=Wikipedia:%E4%BE%B5%E6%9D%83%E6%8B%AF%E6%95%91%E5%B7%A5%E4%BD%9C%E5%B0%8F%E7%BB%84&oldid=23675358> (accessed on 13 November 2012).
60. zhWP, "Weiji Baike: yemian cun fei taolun/yisi qinquan," (Wikipedia: Articles for Deletion-Copyright Violation), *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Wikipedia:%E9%A0%81%E9%9D%A2%E5%AD%98%E5%BB%A2%E8%A8%E8%8E%E8%AB%96%96%91%E4%BC%BC%E4%BE%B5%E6%AC%8A&oldid=23762964> (accessed on 20 November 2012).
61. BB, "Baidu Baike Bangzhu" (Baidu Baike Help), *art. cit.*
62. zhWP, "Weiji Baike: Baidu Baike dui Weiji Baike de qinquan" (Wikipedia: Baidu Baike Violates Wikipedia's Copyright), *art. cit.*
63. zhWP, "Weiji Baike: Weiji Baike bushi shenme" (Wikipedia: What Wikipedia is Not) *Chinese Wikipedia* <https://zh.wikipedia.org/w/index.php?title=Wikipedia:%E7%BB%B4%E5%9F%BA%E7%99%BE%E7%A7%91%E4%B8%8D%E6%98%AF%E4%BB%80%E4%B9%88&oldid=24104466>, (accessed on 25 December 2012).
64. BB, "Baidu Baike Bangzhu" (Baidu Baike Help), *art. cit.*
65. User:Lorenzarius and other contributors, "Weiji Baike: bimian diyu zhongxin" (Wikipedia: Avoid-Regional-Centrism), *art. cit.*

Furthermore, many politically sensitive comments exist on Chinese Wikipedia, mainly because contributors do not need to go through the internal review process required by Baidu Baike when editing articles. Based on my own observations of the discussion pages on both encyclopaedias, politically sensitive comments are fairly common in Chinese Wikipedia but not in Baidu Baike. Some users openly declare their political leanings and attitudes in user pages, ranging from pro-Communists to pro-Falun Gong, and from Han-Supremacists to East-Turkestan independence supporters. I have yet to find another Chinese-language website that hosts users as politically diverse as those of Chinese Wikipedia.

Although researchers cannot directly examine Baidu's non-transparent internal review processes, several leaked internal documents provide convincing details of the day-to-day "internal monitoring and censorship" operations, with the help of using keyword-filtering software. For example, there are documents allegedly leaked from Baidu the company by Chinese overseas political activists that include employee logs, performance evaluations, and directives issued by the authorities.<sup>(66)</sup> These documents reveal the constant censorship efforts of many of Baidu's services, including Baidu Baike and its discussion forums. This evidence is consistent with other reports of Chinese censorship of user-contributed content.<sup>(67)</sup>

In contrast, while Chinese Wikipedia's open review model does not guarantee that all political views will be included, the politically sensitive edits and discussions are recorded by default for public review. Formal policies on "edit wars" are established to guide normal users and administrators.<sup>(68)</sup> In fact, a non-factual satirical article called "The Political Edit War in Chinese Wikipedia" describes the history of Chinese Wikipedia as a war between supporters of the People's Republic of China and supporters of the Republic of China (Taiwan), using the "weapons of vandalism" to impose what they perceive to be a "neutral point of view."<sup>(69)</sup> The fact that such a wide array of politically sensitive articles exists on Chinese Wikipedia allowed an independent researcher<sup>(70)</sup> to conduct a study testing China's Internet filtering regime by accessing all of its entry articles, thereby differentiating between blocked and unblocked articles and reverse-engineering the filtering mechanisms.

In 2005 and 2007, there were proposals discussed on Chinese Wikipedia to practice self-censorship, which generated lengthy and heated debates.<sup>(71)</sup> The community rejected these proposals, but they sought to ensure that a self-censored "apolitical" encyclopaedia would make Chinese Wikipedia accessible again for users in mainland China. A group of editors who supported these self-censorship proposals reportedly migrated to Baidu Baike after the proposals were rejected. Chu has praised the public nature of discussion on Wikipedia as an "extremely valuable resource to study the cultural phenomenon of self-censorship in China."<sup>(72)</sup> Thus, the explicit "no censorship" policy has faced some challenges from the self-censorship culture brought by users who are sympathetic to the Chinese authorities, but their actions on Wikipedia were publicly recorded and scrutinised by users who hold different opinions about editorial policies. The transparency of edits and editorial discussion that explains content-deletions could be construed as documenting "censorship," but otherwise, edits can be reverted to restore the deleted content on the grounds of the "Wikipedia is not censored" policy.<sup>(73)</sup>

Hence, the different ways in which the two encyclopaedias handle politically sensitive materials and comments present a clear contrast, suggesting two distinct approaches in processing user contributions: Baidu Baike employees purge politically sensitive content based on Beijing's directives,

whereas Chinese Wikipedia contributors engage one another openly, sometimes sparking heated edits and discussions among themselves.

Indeed, the contrast between the two websites in their attitudes towards copyright and politics was also observed by a user-contributor of Chinese Wikipedia around the time when Baidu Baike was launched:

I took a walk there when Baidu Baike was first launched. (...) Today [5 May 2006, shortly after Baidu Baike's launch on 20 April] I checked again. It has now more than 82,000 articles. The growth speed is astonishing. Some contributors there have no sense of copyright, pushing copy/paste to an extreme, without citing where the content comes from. Since [that time] much content has been moved to Baidu Baike from other websites, Wikipedia included. (...) In particular, many high-quality non-political articles and images of Chinese Wikipedia became part of Baidu Baike without attribution. (Content written in traditional Chinese was converted to simplified characters). (...) If Baidu continues to let this happen, Baidu will declare itself the biggest Chinese-language encyclopaedia in the world.<sup>(74)</sup>

This quote not only encapsulates what has been discussed on different attitudes in processing copyright-dubious and politically sensitive materials, but also points out the conversion of traditional Chinese character content to simplified Chinese, a point discussed earlier.

Altogether, the findings above show the different processing of copyright-dubious materials and politically sensitive materials. Baidu Baike is more restrictive of new user contributions on the basis of political content while Chinese Wikipedia is more restrictive on the grounds of copyright and quality. Baidu Baike has a record of censoring content while Chinese Wikipedia has an explicitly stated policy that "Wikipedia is not censored."

## Offline environment

The editorial priorities of the two websites discussed above also reflect some differences in the offline environment between mainland China on one hand, and Hong Kong and Taiwan on the other.

Respect for and protection of copyright is generally better in Hong Kong

66. Xiao Qiang, "Baidu's Internal Monitoring and Censorship Document Leaked (1) (Updated)," *art. cit.*
67. Rebecca MacKinnon, "Flatter World and Thicker Walls? Blogs, Censorship and Civic Discourse in China," *Public Choice*, Vol. 134, No. 1, 2008, pp. 31-46; Eva Woo, "Baidu's Censored Answer to Wikipedia," *art. cit.*
68. zhWP, "Weiji Baike: chelun zhan" (Wikipedia: Wheel\_war), *Chinese Wikipedia*, <http://zh.wikipedia.org/w/index.php?title=Wikipedia:%E8%BB%8A%E8%BC%AA%E6%88%B0&oldid=23171133> (accessed on 11 October 2012); zhWP, "Weiji Baike: bianji zhan" (Wikipedia: Edit War), *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Wikipedia:%E7%B7%A8%E8%BC%AF%E6%88%B0&oldid=24095466> (accessed on 14 December 2012).
69. zhWP, "Weiji Baike: Weiji zhengzhi bianji zhan" (Wikipedia: Wikipedia Political Edit War), *Chinese Wikipedia*, <https://zh.wikipedia.org/w/index.php?title=Wikipedia:%E5%9D%8F%E7%AC%91%E8%AF%9D%E5%92%8C%E5%88%A0%E9%99%A4%E7%9A%84%E8%83%A1%E8%AF%9D/%E7%BB%B4%E5%9F%BA%E6%94%BF%E6%B2%BB%E7%BC%96%E8%BE%91%E6%88%98&oldid=16172807> (accessed on 4 April 2011).
70. Xia Chu, "Appendix B. Self Censorship Efforts on Chinese Wikipedia," *Complete GFW Rulebook for Wikipedia*, 25 December 2013, [https://docs.google.com/file/d/0B8ztBERE\\_FUwLWxUX0laeWF3aE0](https://docs.google.com/file/d/0B8ztBERE_FUwLWxUX0laeWF3aE0) (accessed on 12 October 2015).
71. *Ibid.*
72. *Ibid.*
73. zhWP, "Weiji Baike: Weiji Baike bushi shenme" (Wikipedia: What Wikipedia is Not), *art. cit.*
74. User:430072, "Talk: Baidu Baike," *art. cit.*

and Taiwan than in mainland China. For instance, the US Trade Representative (USTR) has been delisting Hong Kong and Taiwan from the priority and general watch list of the Special 301 Report while China remained on the priority watch list in 2015. It is thus likely that the offline copyright protection practices in Hong Kong and Taiwan may have contributed to the stronger gatekeeping process of Chinese Wikipedia on copyright, in addition to the universal norm of copyright protection across global Wikipedia projects.

Practices of free speech and democratic development are also better in Taiwan and Hong Kong than in mainland China, as succinctly described by historian Rana Mitter:

China itself is neither fully free nor democratic. Taiwan, since the 1990s, has been both free and democratic. (...) Most intriguing is Hong Kong, which is little more democratic than it was under the British. Yet it is a very free society: although there is political pressure and a certain level of self-censorship, it has a lively press, it is easy to publish books attacking the Chinese government, and it supports a variety of political parties (although the legislature is arranged to prevent any such party ever coming to power). There are few, if any, other such free, undemocratic societies.<sup>(75)</sup>

The contrast of the editorial priorities of Baidu Baike and Chinese Wikipedia correspond to the offline environments in the areas of copyright and political censorship. It is plausible that the absence versus presence of user contributions from Hong Kong and Taiwan contributes to the different editorial priorities of the two websites.

## Conclusion

As major collaborative filtering projects for Chinese-language information done by Chinese-language users, Baidu Baike and Chinese Wikipedia present different network gatekeeping patterns. First, the level of user autonomy in Chinese Wikipedia is consistently higher than in Baidu Baike. Baidu Baike employs salaried internal reviewers to manage user contribution, whereas Wikimedia Foundation in San Francisco allows user-contributor themselves to filter information. The users in Chinese Wikipedia have more power to develop editorial policies, build language platforms, and contribute political-sensitive content. Second, the gatekeeping mechanism in Chinese Wikipedia has been more open to input from Hong Kong, Taiwan, and other regions beyond mainland China. Third, Beijing's involvement in Baidu Baike's editorial policies/routines and its blocking of mainland Chinese users from accessing Chinese Wikipedia from time to time have also contributed to the differences in network gatekeeping patterns. Altogether, the network gatekeeping processes have been making Chinese Wikipedia more transnational and Baidu Baike more mainland China-centric, a main empirical finding that requires the application of network gatekeeping theory.

## Network gatekeeping

As the theoretical framework of network gatekeeping allows for a broader definition of information control, the article highlights relevant political, cultural, and technical contexts that have contributed to differences between the two. When gatekeeping information online, Baidu Baike has been more accommodating to the needs of Beijing and mainland Chinese users,

whereas Chinese Wikipedia has to accommodate the needs of Chinese-language users from different regions. Two salient network gatekeeping mechanisms have emerged: the censorship mechanisms in Baidu Baike and internationalisation/localisation mechanisms in Chinese Wikipedia.

Censorship mechanisms in Baidu Baike have various effects. First, Baidu Baike removes politically sensitive content under Beijing's directives with the help of salaried internal reviewers and keyword-filtering programs. Effectively, the Beijing government implicitly becomes one of the main but indirect user-contributors that intervene in the internal review routines of Baidu Baike. Second, the contrast in editorial priorities further confirms that Baidu Baike responds more to the censorship of content that is sensitive to Beijing, and less so to the removal of copyright-infringing content. Third, Beijing blocks the access of mainland Chinese users to Chinese Wikipedia from time to time, effectively channelling users to Baidu Baike.

Internationalisation/localisation mechanisms in Chinese Wikipedia have other contrasting effects. First, self-governed Chinese Wikipedia has reached a working consensus to put several Chinese-speaking regions (including the fourth, Singapore/Malaysia) on equal footings in terms of reading and editing articles. Such geolinguistic arrangements constitute the basis for system and content localisation, which then influences the intake of specific groups of users and information. In contrast, Baidu Baike's hosting company has decided not to do so even though it offers Unicode support for Japanese-language search engine users in Japan. Second, users of Chinese Wikipedia have developed an "Avoid-Region-Centrism Policy" to avoid Chinese regional-centric, Han-centric, and Chinese-centric statements. In contrast, Baidu Baike's policies have been set under the national framework of the People's Republic of China. Altogether, the norm of the global Wikipedia movement empowers user-contributors from Hong Kong, Taiwan, and the rest of the world to work with user-contributors from mainland China. The technical default of Baidu Baike disregards edits from non-mainland Chinese characters by design.

Through the analysis of gatekeeping mechanisms, the article has also enumerated multiple factors that contribute to the transnational character of Chinese Wikipedia, including its host, editorial policies, technical design, and participants. In addition, the national (i.e. mainland China-centric) character of Baidu Baike can be explained by the contrasting evidence. Still, further research is required to further tease out which factors are more critical than the others in shaping the transnational or national character of the websites.

More generally, I further argue that such network gatekeeping mechanisms (i.e., mechanisms of censorship and localisation) have been shaping national and transnational dynamics online, an area of research that both Chinese studies and Internet studies need to systematically explore. The different network gatekeeping of new edits and users suggests that Chinese Wikipedia is an example of what Yang Guobin described as the rise of a "transnational Chinese cultural sphere,"<sup>(76)</sup> while Baidu Baike has gained domestic acceptance within the national market of mainland China by operating in a corporate setting with political oversight by Beijing. They present two distinct patterns of engaging Chinese-language users and content: one is directly shaped by the political and market practices of mainland China, and the other is developed by and negotiated among Chinese-language users across regions.

75. Rana Mitter, *Modern China: A Very Short Introduction*, Oxford, Oxford University Press, 2008, p. 132.

76. Guobin Yang, "The Internet and the Rise of a Transnational Chinese Cultural Sphere," *art. cit.*

## Implications for interactions on Chinese Internet

Different gatekeeping patterns suggest different power configurations and interactions among Chinese-language users. Managed by the oversight of Baidu employees who report to Beijing, Baidu Baike contributes mainly to interactions within mainland China. In contrast, self-governed by active editors and administrators (elected mutually among the editors themselves), Chinese Wikipedia amounts to an interesting experiment in governing interactions across major Chinese-speaking regions and cultural-political divisions, which may require more research.

Over the past decade, Chinese-language users have rapidly adopted Internet technologies beyond just Baidu Baike and Chinese Wikipedia. How and where these users can become part of the ongoing mutually-reinforcing cycle of increased participation, content, and readership are important questions for many websites. Based on the findings, I further argue that Baidu Baike and Chinese Wikipedia represent two emerging collaborative ecosystems<sup>(77)</sup> that filter information for Chinese-language Internet users. Baidu Baike exemplifies a national ecosystem that delimits such collaborative participation cycles within mainland China, while Chinese Wikipedia embodies a transnational one that bridges collaborative participation across major Chinese-speaking regions.

The contrast is historically significant in the context of the growth and control of Internet users in mainland China. Their policies differ greatly in terms of whether they censor what Beijing considers politically sensitive content, a contrast that is typical between websites hosted inside versus outside mainland China. Although the more integrative system in Chinese Wikipedia has matured since late 2004, its development in recruiting users from mainland China has been blocked by the Internet filtering/censorship regime of Beijing. Despite being ultimately rejected by the community, proposals for Chinese Wikipedia to practice self-censorship in 2005 and 2007 imply a link between censorship and blocking: self-censor yourself or face blocking by Beijing to access users from mainland. The policy message is similar to the one sent by Beijing to corporations such as Google, Yahoo!, and other Internet companies around that time. The timing of Baidu Baike's launch in 2006 and its ensuing growth also reinforce this policy message. From this perspective, Internet filtering by Beijing constitutes a form of network gatekeeping of user-contributors by preventing users in mainland China from participating.

All these differences indicate two distinct approaches: one defines the centrality or the centre of gravity with a political core (i.e. mainland China), and another integrates diversity of contribution. Ironically for Beijing, which seeks political integration with Hong Kong and Taiwan, by alienating if not outright excluding user participation outside the core, the first approach reinforces the boundary between the mainland and regions outside of mainland China. By avoiding region-centric bias, the second approach seems to have relatively more integrative effects, particularly for users from Hong Kong, Taiwan, and mainland China.

The findings tell neither a story of mere US-centric Americanisation nor of a mere "Chinese local-area network." It is rather a story of two different network gatekeeping efforts mainly across and within Chinese-speaking regions. Ultimately, it is a struggle to "centre" the online Chinese cultural sphere by promoting certain directions, intensities, and normative values via network gatekeeping processes.

## Chinese gatekeeping and gatekeeping Chineseness

The different extent and centres of the two gatekeeping patterns further show the cultural-political tensions across the regions, thereby manifesting different "Chinese" character. Arguably, Beijing's censorship/filtering regime has successfully avoided outside influence from Hong Kong and Taiwan at the expense of potential interactions between the regions inside and outside of mainland China. Were it not for Beijing's intervention and gatekeeping of mainland Chinese users, the centre of gravity would likely be spatially located among Hong Kong, Taiwan, and other coastal urban areas in mainland China, as shown by the geographic distribution of power users in Figure 3 and Figure 4. Baidu Baike's dominance in mainland China but limited outreach to traditional Chinese users indicates the re-centering efforts that have aimed to make Beijing *the* cultural-political centre. These cultural-political struggles are inherently *Chinese*. The main struggles are about whether Beijing should be the sole centre and source of collaborative filtering (and network gatekeeping) processes of Chinese-language information and whether Hong Kong and Taiwan could be the other main sources or even alternative centres.

Serving as major examples of Chinese network gatekeeping, Baidu Baike mirrors the offline cultural-political environments of mainland China, while Chinese Wikipedia features a system that integrates diverse cultural-political experiences from major Chinese-speaking regions. The media environments and user experiences influence how user contribution is processed online. Online communicative spaces, no matter how new or universal, are shaped by its participants' prior and local experiences. Baidu Baike's internal communicative space, despite its claim to open participation, excludes specific geo-linguistic variants of Chinese language and certain aspects of editorial freedom. This is in direct contrast to Chinese Wikipedia, where users are given significantly more autonomy to communicate and collaborate in a more open and inclusive fashion. Thus, Baidu Baike's communicative space seems to fall in line with Beijing's actions and attitudes towards Internet users within mainland China, whereas that of Chinese Wikipedia reflects both the initial consensus and ongoing challenges in integrating participants' experiences and attitudes towards diverse modern Chinese media environments that are transnational.

Although the theories of "network gatekeeping"<sup>(78)</sup> and "collaborative filtering and accreditation"<sup>(79)</sup> anticipate new information gatekeeping activities contributed by individuals online, offline environments remain important. I further argued that the offline environments of Hong Kong and Taiwan, with friendlier attitudes towards free speech and copyright protection, have shaped the network gatekeeping processes in Chinese Wikipedia. Indeed, many of the resulting differences can be explained by the offline environments in mainland China on one side and Hong Kong/Taiwan on the other.

While the core idea of "collaborative filtering and accreditation" in deciding "whom to trust and whose words to question"<sup>(80)</sup> remains useful, the authoritarian state-led or sanctioned measures may shape or even alter such network gatekeeping processes. Collaborative filtering and accreditation thus do not necessarily reproduce liberal democratic norms, and can

77. Chitu Okoli et al., "The People's Encyclopedia under the Gaze of the Sages: A Systematic Review of Scholarly Research on Wikipedia," working paper, 2012, doi:10.2139/ssrn.2021326.

78. Karine Barzilai-Nahon, "Toward a Theory of Network Gatekeeping," *art. cit.*

79. Yochai Benkler, *The Wealth of Networks*, *op. cit.*, p. 465.

80. *Ibid.*

be used to re-establish other social orders (a set of social structures and practices that reinforce or change social norms). For example, Baidu Baike's network gatekeeping shares the same extent and compatible order with mainland China, the main jurisdiction of the People's Republic of China. Chinese Wikipedia has managed to integrate the varied practices and expectations of users across regions, with a set of norms of no censorship and respect for copyright.

Beijing's filtering and censorship regime thus reinforces a larger network-gatekeeping pattern within the geolinguistic identifier of "zh-cn." By blocking Chinese Wikipedia from mainland Chinese users, Beijing channelled some of the early and crucial popularity away from Chinese Wikipedia to Baidu Baike, thereby fortifying its cultural-political power in ways that are possible with Baidu Baike but much less so with Chinese Wikipedia. Such practices led to the collaborative filtering patterns desired by Beijing. Indeed, it has been reported that the Chinese authorities in Beijing have "guided" public opinion by encouraging some voices while discouraging others on various media platforms.<sup>(81)</sup> What we can see here is the importance of user autonomy, which cannot be overstated; hence it is necessary to examine the ways in which user contributions are encouraged or discouraged through website designs and policies, including the internationalisation/localisation mechanisms discussed as "geolinguistic" arrangements here. The overall network gatekeeping outcome thus depends on the host, design, and participants of user-generated encyclopaedias. Any conclusion, then, must address the question of how authoritarian constraints imposed on the Internet segment inside mainland China may produce different network gatekeeping effects on user autonomy. In particular, the roles of Hong Kong and Taiwan are expected to be relevant and controversial if Beijing's Internet sovereignty agenda is to be extended to cover Hong Kong and Taiwan, over both of which China claims sovereignty.

As with the recent development concerning citizenship in Hong Kong, Taiwan, and mainland China,<sup>(82)</sup> researchers are bound to confront the controversy of defining "Chinese" information and communication processes. I further propose a tentative hypothesis as a way of conclusion.

A pair of geographic metaphors used by historians of modern China such as Fairbank<sup>(83)</sup> and Wang<sup>(84)</sup> can be instructive for us to understand the online dynamics as part of the longer Chinese modernisation process: "Maritime China" refers to modern maritime activities that connect China to the world, whereas "Continental China" refers to the Chinese agrarian-bureaucratic tradition. Despite being peripheral to geographical China, Maritime China has grown into a richer and bigger segment of Chinese society as various leading economic and cultural centres, whereas Continental China views maritime interests as "narrow and self-serving," with Hong Kong and Taiwan serving as "extreme manifestations."<sup>(85)</sup> A linguistic dimension can be added: Maritime China tends to be more tolerant of language variations and exchanges (Han Chinese southern dialects are concentrated along the eastern coast of southern China). Continental China tends to be less tolerant because of its bureaucratic tradition that insists on one unified Chinese national language.

Baidu Baike and Chinese Wikipedia can thus be understood as the latest incarnations of Continental China and Maritime China, at least in filtering information and knowledge for Chinese-language users. Geolinguistically, Baidu Baike sticks to simplified Chinese in mainland China, whereas Chinese Wikipedia integrates mainland China alongside the major regions that constitute Maritime China, which even includes Singapore and Malaysia. It is worth mentioning that, although not thoroughly examined here,

Wikipedia has hosted other southern Chinese dialect versions such as Cantonese, Wu, Min Nan, and Hakkanese.

Seen from the longer historical context, China's current Internet filtering and censorship regime, known as the Great Firewall of China, should be viewed as a cultural-political project that Continental China has imposed on Maritime China as part of the latest development of Chinese modernisation. At least two notions of Chineseness emerge: Baidu Baike embodies the Beijing or mainland China-centric version of Continental China, while Chinese Wikipedia reflects an instance of multi-centric integration of Chinese transnationalism as the latest version of Maritime China. The proposed hypothesis needs further research beyond the limited evidence here.

To sum up, the theoretical lens of network gatekeeping has been successfully applied to show how the two encyclopaedia websites have engaged Chinese regions differently. Original and in-depth comparisons contribute to the current understanding of the Chinese Internet, showing two distinct approaches in harnessing the power of Chinese-language users that generate distinct patterns of network gatekeeping. Part of the contribution to this distinction comes from Beijing's intervention in the exchanging dynamics of different media and information environments across Chinese-speaking regions. The authoritarian constraints imposed by Beijing on Chinese Wikipedia could thus be interpreted as a particular form of cultural-political policy that aims to promote Beijing-centric activities and values. Baidu Baike's Beijing-centric practices are in direct contrast to the region-neutral, multi-script, open-ended practices of Chinese Wikipedia that put the various Chinese-speaking regions on an equal footing. When researchers conceptualise the power of Chinese Internet users in relation to the power that wants to harness them, they should consider not only the emerging mechanisms of network gatekeeping, but also the implications for representing and reproducing Chineseness.

■ Han-Teng Liao (DPhil, 2014, Oxford) works as a research fellow with the United Nations University Institute on Computing and Society (UNU-CS), a research institute at the intersections of ICTs and international development. Estrada do Engenheiro Trigo, No. 2-4, Macau (hanteng@unu.edu).

81. Anne-Marie Brady, *Marketing Dictatorship: Propaganda and Thought Work in Contemporary China*, Lanham, Rowman & Littlefield, 2009; Phil Muncaster, "Chinese Micro-Blogs a Hit with Police: But Sina Users Urged to Snitch on Each Other," *The Register*, 30 May 2012, [www.theregister.co.uk/2012/05/30/police\\_china\\_weibo\\_censorship](http://www.theregister.co.uk/2012/05/30/police_china_weibo_censorship) (accessed on 12 October 2015); Guobin Yang, "Internet and Civil Society," in William S. Tay and Alvin Y. So (eds), *Handbook of Contemporary China*, Singapore, World Scientific, 2011.
82. Jun Xing, Pak-Sheung Ng, and Chloe Cheng, *General Education and the Development of Global Citizenship in Hong Kong, Taiwan and Mainland China: Not Merely Icing on the Cake*, London, Routledge, 2012.
83. John Fairbank, "Introduction: Maritime and Continental in China's History," in John Fairbank (ed.), *Republican China 1912–1949, Part 1*, The Cambridge History of China, Cambridge, Cambridge University Press, 1983, pp. 1–27.
84. Gungwu Wang, *The Revival of Chinese Nationalism*, IAS Lecture Series, Leiden, International Institute for Asian Studies, 1996; Gungwu Wang, "Maritime China in Transition," in Gungwu Wang and Chin-Keong Ng (eds), *Maritime China in Transition 1750–1850*, Wiesbaden, Harrassowitz Verlag, 2004.
85. Gungwu Wang, *The Revival of Chinese Nationalism*, op. cit., p. 17.