Mapping Articles on China in Wikipedia: An Inter-Language Semantic Network Analysis

Ke Jiang[1]

kejiang@ucdavis.edu

Grace A. Benefield[1]

grbenefield@ucdavis.edu

Junfei Yang[2]

jfyang2010@gmail.com

George A. Barnett[1]

(Corresponding Author)

gabarnett@ucdavis.edu

1. University of California, Davis
2. University of Science and Technology of China, Hefei

Draft

January 2015

Abstract

This article describes an inter-language semantic network analysis examining the differences between articles about China in the Chinese and English versions of Wikipedia. The results not only confirmed previous findings of inter-language Wikipedia studies but also extended the research by providing an example of how exactly Chinese and English speaking groups frame the same topic differently in Wikipedia. Specifically, while both versions had the common focuses on government, population, language, and character, diplomatic relations, development of the economy, science and technology, the Chinese-speaking contributors and English-speaking contributors framed the article of China differently according to dissimilarities in national cultures, values, interests, situations, and emotions.

Keywords: Wikipedia, Article of China, Inter-Language, Semantic Network Analysis

Mapping Articles on China in Wikipedia: An Inter-Language Semantic Network Analysis

1. Introduction

Wikipedia is the largest free online encyclopedia that is globally and collaboratively generated. Although Wikipedia is a compilation of facts, it is not a culturally neutral space (Pfeil, Zaphiris, & Ang, 2006). People with different cultural, social, national and lingual backgrounds create different language editions of Wikipedia. Wikipedia provides a gigantic virtual arena for the public to negotiate the definition and meaning of social reality according to their personal preferences, which are enmeshed in specific cultural and linguistic contexts. The cross-language analysis of Wikipedia articles is valuable for intercultural studies since repositories of various cultural values can be easily obtained from the different language versions of Wikipedia articles. Previous research on inter-language analysis of Wikipedia articles mainly studied the geographic focus (Hecht & Gergle, 2009; Overell & Rüger, 2011), famous and prominent people (Callahan & Herring, 2011; Eom & Shepelyansky, 2013; Kolbitsch & Maurer, 2006), historical figures (Aragón, Laniado, Kaltenbrunner, & Volkovich, 2012), editing behavior (Pfeil, Zaphiris, & Ang, 2006), user interaction (Nemoto & Gloor, 2011), and article structure (Laufer, 2014) of Wikipedia editions in different languages. Additionally, concept overlap (Warncke-Wang, Uduwage, Dong, & Riedl, 2012), inter-language links (Hassan & Mihalcea, 2009) between Wikipedia versions in multiple languages have been examined. However, insufficient attention has been paid to the detailed content of Wikipedia articles. This paper conducts a computer-assisted semantic network analysis to explore and map the similarities and differences of the Wikipedia articles on China in both the Chinese and English versions providing an example of studying how different language groups frame the same topic differently.

2. Inter-language Analysis of Wikipedia Article

Articles in Wikipedia represent the outcome of a continuous collaborative effort of a large number of volunteer contributors speaking different languages. Previous studies have shown strong interests in analyzing the similarities and differences of Wikipedia articles in different languages. Among the activity of all 287 language editions of Wikipedia, approximately 15% of edits are made by bots (Steiner, 2014). Also, only a small proportion of user produced most of the content of Wikipedia article (Panciera, Masli, & Terveen, 2014). 90% of users contributed less than 10% of the overall contributions in all language editions (Ortega, Gonzalez-Barahona, & Robles, 2008). The core group of contributors does not consist of experts, but people heavily involved in Wikipedia and interested in open collaboration and free knowledge exchange (Bryant, Forte, & Bruckman, 2005).

Besides the similar patterns of contributors' activities, scholars also studied the bias in Wikipedia across different language editions. Hecht and Gergle (2009) found that across 25 different Wikipedia language versions, only one-tenth of one percent is comprised of common concepts. Rask (2008) examined the connection between the activity of Wikipedia contributions and a country's development and found that Wikipedia activity at a national level is correlated with the score on the United Nations Human Development Index (HDI). By classifying Wikipedia articles as locations and calculating the ratio between locations where the respective language is spoken and locations where it is not, Overell and Ruger (2011) studied how biased a particular Wikipedia is toward speakers of its language, and confirmed the Steinberg Hypothesis that "everyone has a localized fish-eye view of the world (p. 9)." In particular, they found the English Wikipedia (bias: 13.15) was much more biased than the Chinese Wikipedia (bias: 0.79), and attributed the small bias in Chinese Wikipedia to the fact that Wikipedia has been blocked in

China to various degrees from 2004, and many users editing Wikipedia in Chinese are expatriates. When studying the 30 most popular articles about individuals in each language edition, Eom and Shepelyansky (2013) found that local heroes are dominant among the top 30 people in each language version. Through comparing articles of famous people in the Polish and English editions of Wikipedia, Callahan and Herring (2011) found systematic differences related to cultural differences, histories, and values of Poland and the United States.

Scholars have used network analysis methods to address the differences of Wikipedia articles across different languages. For example, by looking at the networks of links between a set of biographical articles on the 15 largest language Wikipedias, Aragon et al. (2012) found the networks are more similar for geographically or linguistically closer communities. Similarly, Hecht and Gergle (2009) used in-degree sums and PageRank sums as indicators to show the existence of self-focus of 15 language editions. Warncke-Wang et al. (2012) used the inter-language link network to measure the similarities and differences between all of Wikipedia language editions and found that language similarity was positively correlated with the similarity between editions. Also, they found the concept overlaps across different language platforms were about general topics, mainly countries, cities, and lists of events. Nemoto and Gloor (2011) studied the networks of users talk and editing behavior in the English, German, Japanese, Korean, and Finnish editions and found that the Japanese and Korean editions show a less stable collaboration network than their Western counterparts.

Although this research was significant in using advancing algorithm models and network analysis methods to measure the differences among a vast number of Wikipedia articles in multiple language editions, it reduced the inter-language differences to dissimilarities among people, location, topics, categories, article structures, cross-language links, and user behaviors,

without paying sufficient attention to the detailed meaning from a particular semantic context. This paper extends the previous research by using semantic network analysis methods to examine the specific semantic differences that emerged from articles in different language editions of Wikipedia, as well as to map how different language speakers illustrate the meaning of a particular concept in various ways in Wikipedia.

3. Semantic Network Analysis

Computer-assisted semantic network analysis developed from examining the visibility and co-occurrence of specific vocabularies in texts. Rooted in the cognitive paradigm (D'Angelo, 2002) and the tradition of frame semantics in linguistics (e.g. Fillmore, 1982), scholars have argued that words are hierarchically clustered in memory (Collins & Quillian, 1972). Thus spatial models that illustrate the relations among words are representative of meaning (Barnett & Woelfel, 1988). The structured semantic representations of multiple connections between various concepts are regarded as semantic networks (Schultz et al., 2012). Therefore, semantic network analysis (SMA) is a form of content analysis that identifies the network of associations between concepts expressed in a text (Carley & Palmquist, 1992; Doerfel, 1998).

This paper focuses on analyzing the salience of the concepts in texts. The word salience means "making a piece of information more noticeable, meaningful, or memorable to audiences" (Entman, 1993, p.53). On the one hand, the text can make a piece of information more salient through repetition or putting it in a prominent position in the semantic structure of the content (Entman, 1993). On the other hand, it also can make bits of information more salient by associating them with culturally familiar symbols" (Entman, 1993, p. 53). SMA has an advantage in analyzing the salience of the concept. It can generate the visual map of the semantic structure of the content, which illustrates not only the position of individual concepts appearing

in the text but also the complex associations among them. Specifically, the salience of the concept can be measured through the analysis of concept centrality that reflects the location and the importance of a concept in relation to other concepts in the network (Freeman, 1979; Wasserman & Faust, 1994). Since the meaning of a concept depends on its surrounding context, SMA also focuses on examining the concept associations by looking at the communities or the concept clusters that composing the semantic networks and the frequency with which concepts co-occur nearby. Also, through analyzing the correlations between different semantic networks, the similarity of various texts can be determined. This paper uses semantic network analysis to examine the differences and similarities of the semantic content of articles on *China* in Chinese and English Wikipedias.

3. A Semantic Network Analysis of the Articles of China in Wikipedia.

3.1. Articles of China in Wikipedia and Research Questions

China has a continuous history of civilization of nearly 5000 years. The cultural values of China (e.g. Taoism, Confucianism) have made tremendous impacts on the Eastern sphere of human society. To a certain extent, the cultures of English-speaking countries and the Chinese culture are somewhat polarized, with the former emphasizing individualism and the latter emphasizing holism. Globalization, international trade and the development of advanced information technology have intensified dialogue between the two cultures. In particular, Chinese economic growth has provided the country more power in discourse in the global arena, including the World Wide Web, to negotiate the dominant worldview according to its cultural preferences. Because the concept of "China" carries a rich connotation of the Chinese culture globally, the analysis of the content of an article about China in Chinese and English Wikipedia provides a convenient and efficient way to reveal the bias of Chinese and English speakers.

Thus, its analysis can demonstrate exactly how Chinese and English speakers frame the content of Wikipedia in different ways.

The articles of China on Wikipedia in both Chinese and English were retrieved on January 13, 2016. Both the language content and the editing contribution history were collected. The editing contribution history includes all the users that have added or subtracted content to the Wikipedia pages. Table 1 describes the general information on the editing history of each article, while Figure 1 graphs the log-log degree distribution of the users contributing to each of the Wikipedia pages. As expected, the shape of the graph for both pages suggests that a few users tend to contribute the most edits, while many users contribute only one or two edits. The English page has more users in the "tail" of the graph, indicating that the English page has more people who contribute fewer number of edits than the Chinese page. The Chinese-speaking editors have a higher average number of edits per user compared to the English-speaking editors.

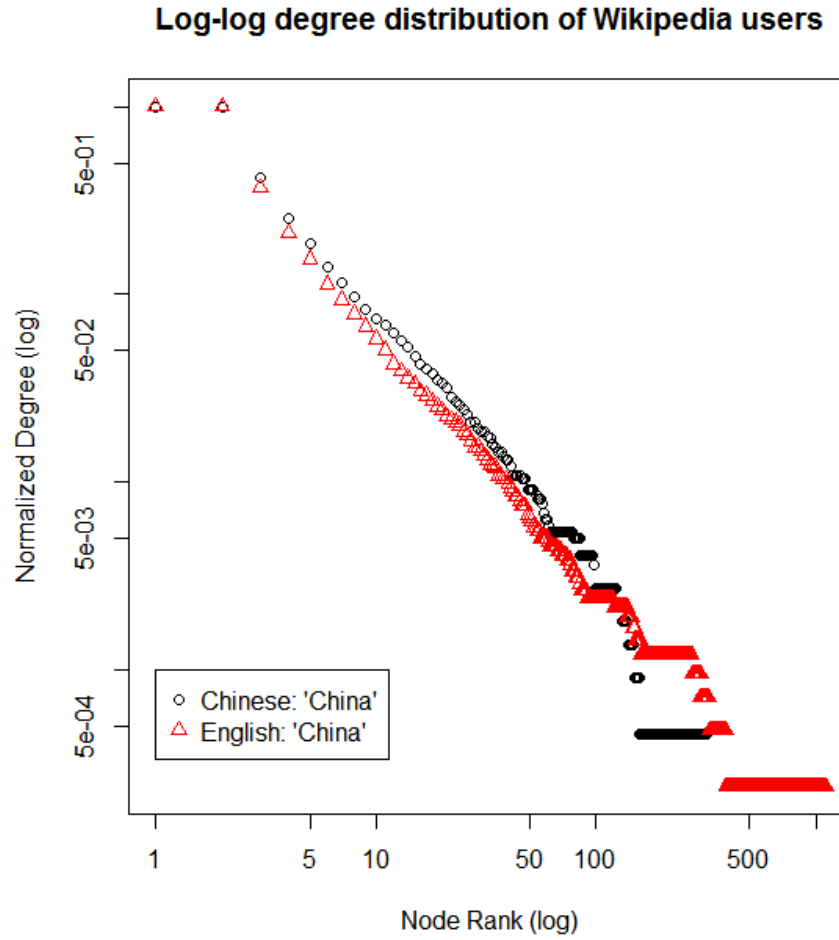| Table 1 Edit History of Wikipedia Articles		
Article of China	**Chinese Wikipedia**	**English Wikipedia**
Number of Bytes	192,220	230,512
Number of Contributors	2,218	4.192
Number of Edits	8,409	14,960
Average Edits per Person	3.79	3.54

Log-log degree distribution of Wikipedia users

Figure 1. The log-log degree distribution of Wikipedia contributors for the Chinese and English versions of the "China" page.

Two semantic networks were created based on the analysis of word co-occurrence. One is the Chinese semantic network of China (*SCC*); the other is the English semantic network of China (*SEC*).

The research raises the following questions:

R_1: What is the correlations between *SCC* and *SEC*?

R_2: What are the most central concepts in *SCC* and *SEC*?

R₃: How many clusters are there in *SCC* and *SEC*?

R₄: What are the most central concepts in each cluster?

R₅: What are the most frequent concept associations in each cluster?

3.2. Procedures

The first step in the research process was editing the texts. Syntactically functional words (e.g. a, an, the, 的, 地，得) were removed, different forms of the same word (e.g. *China/Chinese, 法/法律*) were adjusted, and some bigram words were combined as one concept (e.g. Communist Party). The second step was to generate the semantic matrices from the edited texts. The principle of producing the links between words in semantic network was based on the measurement of word co-occurrence. Based on Miller (1956)'s argument that people's working memory can process seven meaningful units at a time, besides the word *China*, 150 most frequent words that occurred within seven concepts of each other in an article were considered connected regardless of the number of words separating the terms (Danowski, 1993). The first two steps were conducted using the ConText software (Diesner, Aleyasen, Kim, Mishra, & Soltani, 2013) and Python.

In the third step, the two semantic networks were examined through UCINET (Borgatti, Everett, & Freeman, 2002), Gephi (Bastian, Heymann, & Jacomy, 2009), which are software for network analysis, graphics, and statistical computing.

UCINET uses QAP correlation analysis to calculate the correlation between two semantic networks. QAP correlation is similar to the traditional correlation analysis. The only difference is that QAP is a nonparametric technique that does not rely on assumptions of independence. The algorithm proceeds in two steps (Borgatti, et al., 2002). First, it computes Pearson's correlation between the corresponding cells of the two networks. Second, it randomly permutes the rows and

columns of the matrix and re-computes the correlation hundreds of times to determine the proportion that is larger than or equal to the measure calculated in step 1. A small proportion (< 0.05) suggests a strong relationship between networks that is unlikely to have occurred by chance. UCINET also calculates the normalized degree and eigenvector centralities of each concept in the two semantic networks. The degree is the total number of direct links. Eigenvector centrality indicates a concept's overall network centrality (Bonacich, 1972).

Gephi (Bastian, et al., 2009) calculates the clusters of networks and creates visual maps of semantic networks. In the visual maps, the size of the label of each word depends on its degree centralities, such that the larger the object, the more central a word is to the description of China. Lines on the maps indicate the presence of a relationship between each pair of words. The thicker lines represent a stronger relationship between two words. Also, the shorter distance between two words, the closer relationship there is between them. For visual maps of *SEC*, the English translations of the Chinese words were noted next to the Chinese labels.

3.3. Results

Table 2 illustrates the overview of *SCC* and *SCE*. The 150 Chinese words were translated into English words. Among the 150 most frequent English words in the two networks, there are 62 common words (Table 3). For **R₁**, the results of QAP correlations revealed non-significant correlations ($r = .02$, $p = .055$) between the two semantic networks. To answer **R₂**, table 4 demonstrates the 20 most frequent words with the greatest normalized eigenvector centralities. Among these 20 central words, *law*, *nationality*, and *new* are unique in *SCC*; *Dynasty*, *foreign*, *remain*, *early*, and *history* are unique in *SEC*. For **R₃**, from table 2, there are six clusters in *SCC*, and there are five clusters in *SCE*.

Table 2 Overview of SCC and SCE

	SCC	SCE
Density	.141	.297
Mean Link Strength	.285	.77
SD of Link Strength	1.047	1.907
Number of Clusters	6	5

Table 3 Common Words between SCC and SEC

art	development	health	level	plan	Shanghai
Asia	economy	high	mainland	policy	social
begin	education	HK	Mao	politics	system
Beijing	establish	increase	Macao	population	Taiwan
central	family	influence	military	power	technology
century	food	international	minority	reform	trade
character	free	issue	modern	relation	USA
citizen	global	Japan	movement	religion	
CommunistParty	government	language	network	research	
congress	group	largest	official	revolution	
culture	Han	leader	organization	security	

Table 4 The 20 Most Frequent Words with Greatest Normalized Eigenvector Centralities

SCC		SCE	
government	.3	world	.239
largest	.287	government	.223
mainland	.266	economy	.213
people	.233	dynasty	.195
economy	.222	largest	.192
world	.219	population	.184
development	.206	system	.181
politics	.199	military	.173
law	.199	development	.171
begin	.185	culture	.17
reform	.179	foreign	.156
HK	.173	modern	0.155
international	.171	remain	0.154
system	.167	reform	0.154
congress	.166	CommunistParty	0.154
USA	.163	early	.15
nationality	.159	begin	.148
new	.158	Bejing	.145
technology	.156	history	.144

Graphic representations of *SCC* and *SEC* are presented in Figure 2a and 2b. Different colors represent different clusters of the semantic networks. To answer **R₄**, table 5 illustrates the most central word in each cluster and the proportion of each colored cluster in *SCC* and *SEC*.
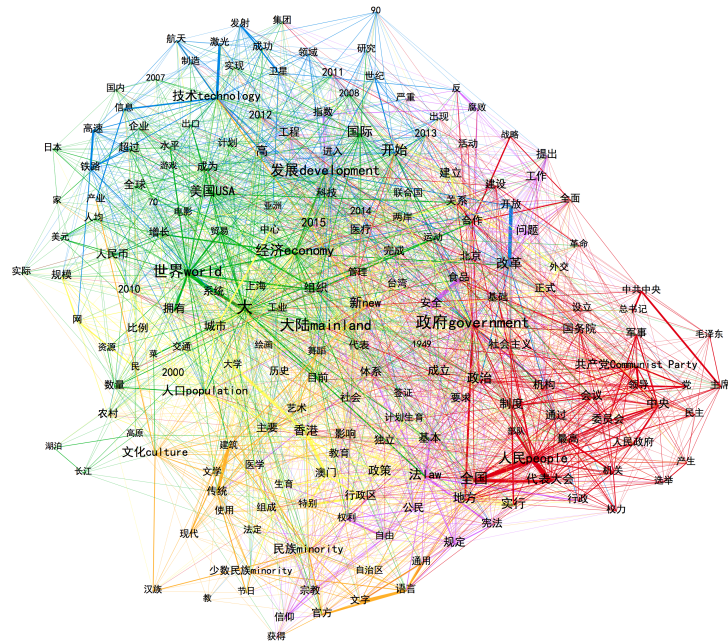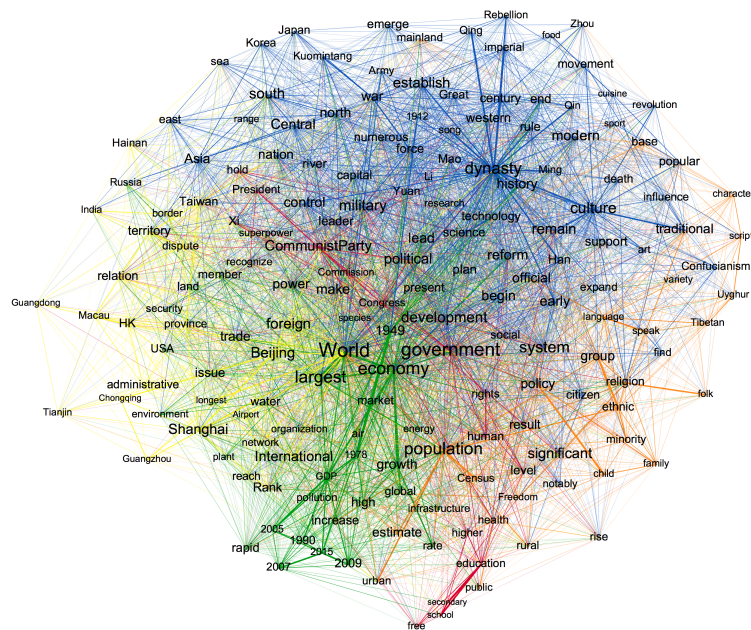

Figure 2a Graphic representations of *SCC*

Figure 2b Graphic representations of *SEC*

Table 5 Top 5 Central Words and Proportion of Each Cluster in SCC and SEC

	SCC		SEC	
Red	government people Communist Party central system	21.94%	government politics Communist Party education rights	10.16%
Orange	nationality language new culture character	11.73%	population policy religion official ethnic	17.32%
Yellow	mainland HK population Macao relation	16.84%	Beijing Shanghai foreign territory international	13.97%
Blue	technology development begin opening high	13.27%	dynasty culture military war development	38.55%
Green	world largest economy organization USA	21.94%	world economy largest growth GDP	19.55%
Purple	law food security provision right	14.29%		

Graphic representations of each cluster in *SCC* and *SEC* are presented in Figure 3a-3f. To answer **R₅**, table 6a-6f illustrated the top 5 word associations in each cluster.
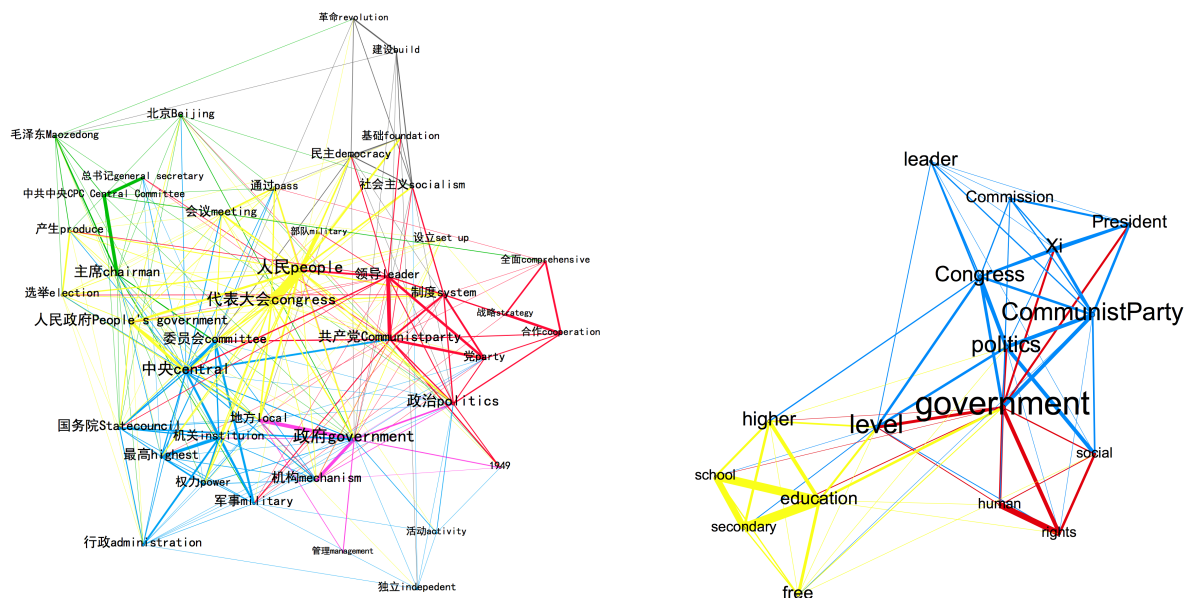
Figure 3a Red Cluster in *SCC* and S*EC*: Centered about Government

Table 6a Top 5 Word Associations in the Red Cluster			
SCC		SCE	
people congress	30	secondary education	23
local government	12	education school	20
people military	11	secondary school	15
highest institution	10	human rights	14
government mechanism	10	government rights	9
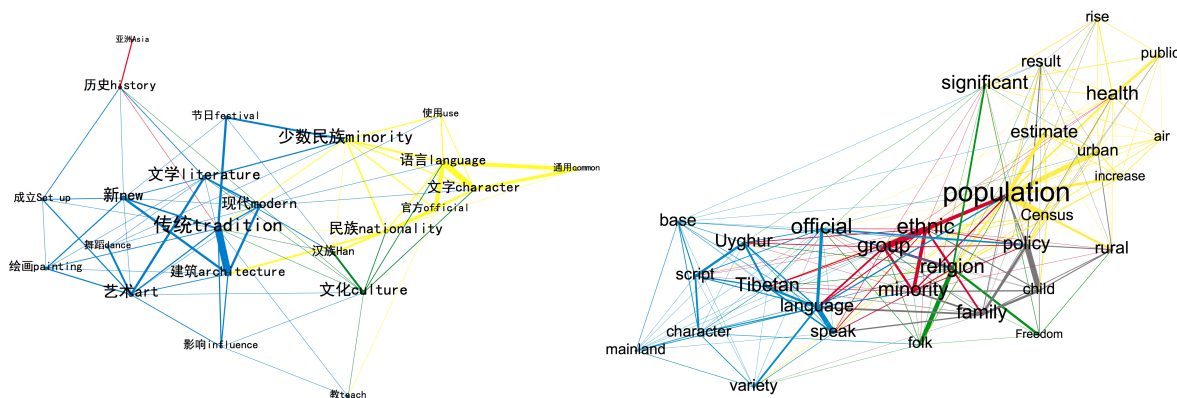
Figure 3b Orange Cluster in *SCC* and *SEC*: Common or Diverse Language

Table 6b Top 5 Word Associations in the Orange Cluster			
SCC		SCE	
architecture tradition	19	population census	25
language character	16	ethnic group	24

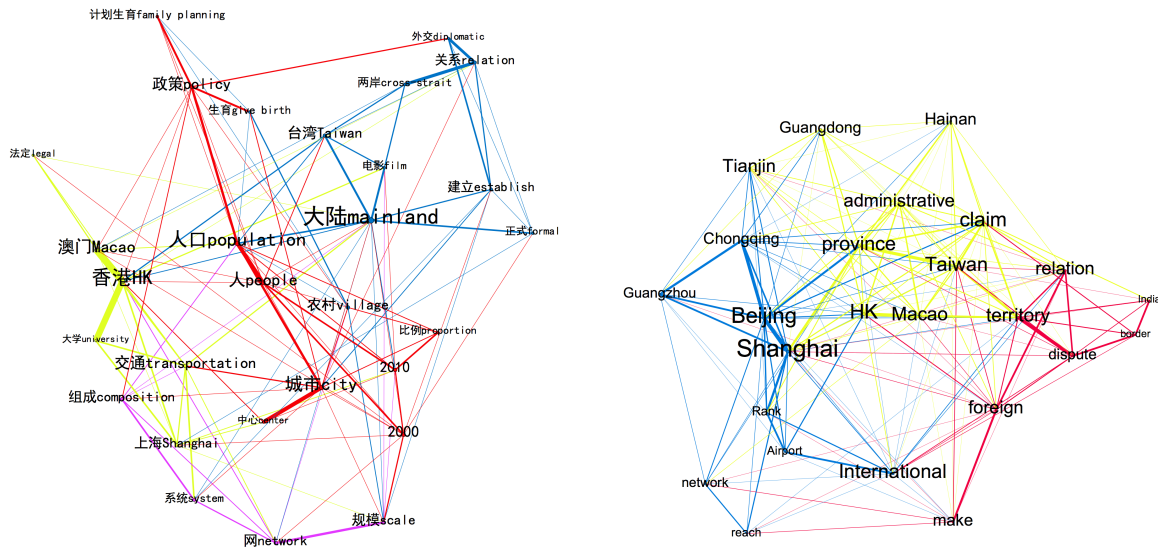| official language	11	child policy	22
common language	10	language speak	19
common character	8	urban population	18

Figure 3c Yellow Cluster in *SCC* and *SEC*: Different Relations with Taiwan

Table 6c Top 5 Word Associations in the Yellow Cluster			
SCC		SCE	
Macao HK	18	Beijing Shanghai	17
HK university	16	HK Macao	15
center city	10	territory dispute	15
cross-strait relation	8	Beijing Chongqing	14
diplomatic relation	7	Shanghai rank	10
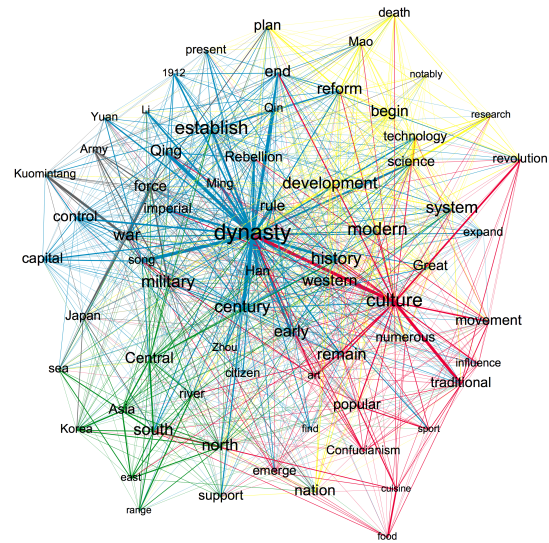
Figure 3d Blue Cluster in *SCC* and *SEC*: New Technology VS History

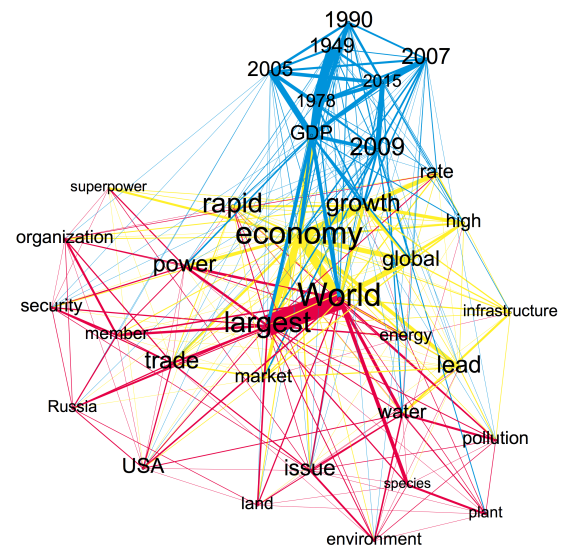| Table 6d Top 5 Word Associations in the Blue Cluster			
SCC		SCE	
opening reform	22	Qing dynasty	21
laser technology	16	Song dynasty	21
high-speed railway	12	science technology	20
information technology	9	dynasty culture	20
aerospace technology	8	Han dynasty	18

Figure 3e Green Cluster in *SCC* and *SEC*: Centered about World Economy

Table 3e Top 5 Word Associations in the Green Cluster			
SCC		SCE	
world largest	19	World largest	70
world trade	9	world economy	29
economy increase	8	economy growth	26
become USA	6	economy GDP	22
exceed USA	4	global economy	16
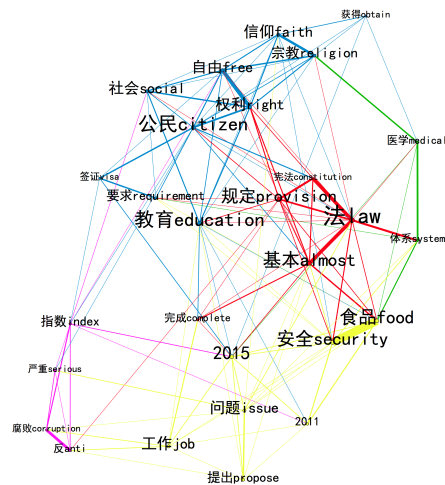
Table 6f Top 5 Word Associations in the Yellow Cluster in SCC	
food security	24
free right	11
law constitution	10
anti corruption	8
religion faith	7

Figure 3f Purple Cluster in SCC: Law and Right

As explained in the next section, the results suggest that when discussing knowledge about China, although both the Chinese and English speakers focused on describing aspects of politics, economy, culture, and development, the framing strategies they used were substantially different.

4. Discussion – Similarity and Differences between Articles of China on Wikipedia

4.1. Authority VS Democracy

The *government* is the first and second most central word in *SCC* and *SEC*. It was also the most central concept of the red clusters in both *SCC* and *SEC* (Figure 3a). But the proportion of the red cluster in *SCC* is twice as large as in *SEC*, indicating that the Chinese contributors have allocated much more attention to government issues. Moreover, while the Chinese contributors portrayed government as the central and highest institution, the English contributors

interpreted it from the perspective of human rights. This difference resembles the political-cultural differences between China and Western countries. Specifically, the Chinese civilization's fundamental themes of social and political life emphasize the importance of collectivity and individuals' responsibilities to the collective (Jiang, Barnett, & Taylor, 2015). One of the most important responsibilities can be described as obeying authority (Pye, 1992). Respecting authority is one of the central concerns in Chinese political culture. This traditional political culture of submission to authority indicates that the Chinese believe that power lies external to the self (Hua, 2001), and belongs to the government, the highest institution. On the contrary, the quality and scope of democratic values are crucial to Western society, especially English-speaking countries, including the United States. For example, more than nine out of ten American citizens support the idea that individuals, including all minorities, have rights to free speech, press, expression, and religion, and that public officials should be chosen by majority votes in regular elections (Miroff, Seidelman, Swanstrom, & Luca 2009). Facing the rapid economic growth of the centralized China led by the Communist Party, people in Western countries also have critical concerns of the threat of Chinese government on human rights and freedom. For example, in English Wikipedia, it was written: "foreign press agencies and NGOs also routinely criticize China's human rights record…The Chinese state is regularly accused of large-scale repression and human rights abuses in Tibet and Xinjiang, including violent police crackdowns and religious suppression." ("China," 2016)

The orange clusters (Figure 3b) that centered about nationality in *SCC* and population in *SEC* also demonstrated this cultural difference. Both the two clusters discussed language and character but a different emphasis. While language and character were closely associated with *common* in *SCC*, they had close associations with *variety*, *minority* and *ethnic groups*, such as

Uyghur and Tibet in *SEC*. After thousands of years of evolution, the Han language and character have become the dominant and official language and character in Chinese society. When the culture of respecting authority extended to the area of language, respecting common language becomes the responsibility of the Chinese citizens. However, in the English Wikipedia, although the contributors acknowledged the official role of Han language, they also valued the existence of diversity, and emphasized, "there are as many as 292 living languages in China," including ethnic minority languages in Tibet, Xinjiang, and also southwest, northeast, northwest of China ("China," 2016).

4.2. Cooperation VS Dispute

In the yellow cluster (Figure 3c), *mainland* and *Beijing* were the most central words in *SCC* and *SEC* respectively. Both the two yellow clusters mapped the network that linking China's big cities, two Special Administrative Regions (Hong Kong, Macao), and Taiwan. However, different relations between China and Taiwan in *SCC* and *SEC* were demonstrated. The Chinese contributors emphasized the establishment of formal diplomatic cross-strait relations; the English contributors focused on the territory dispute between China and Taiwan. This difference illustrates the different foreign policies in China and the United States. On the one hand, from the beginning of the 21st century, starting with the Hu-Wen leadership, the harmony-based Confucian rhetoric, which claims harmony facilitates common development and growth, has been reiterated to project a pacifist cultural image of China (Cao, 2007). On the other hand, the foreign policies of the English-speaking countries value the importance of democracy. In particular, the United States is regarded as the global leader and guardian of democracy and freedom (Marsella, 2011). The commitment to democracy also makes the United States views itself as "a righter of wrongs around the world, in pursuit of tyranny, in defense of

freedom no matter the place or cost" (Said, 1993, p. 5). The Chinese and English contributors'

descriptions of the relations between China and Taiwan reflected the ideas of two different

foreign policies. The Chinese Wikipedia describes the free trade agreement and the integration

and convergence of cultural industry between China and Taiwan. However, the English

Wikipedia writes: "The People's Republic of China has administrative control over 22 provinces

and considers Taiwan to be its 23rd province, although Taiwan is currently and independently

governed by the Republic of China, which disputes the PRC's claim." ("China," 2016)

4.3. Nationalism

The different semantic contents between *SCC* and *SEC*'s blue clusters (Figure 3d) and

green clusters (Figure 3e) expressed the nationalism and patriotic emotions of Chinese

contributors.

High Technology VS History

Among the 20 most frequent concepts with greatest normalized eigenvector centralities,

the appearance of *new* and *technology* in *SCC* and *dynasty* and *history* in *SEC* has demonstrated

the distinct contrast between the framing strategies of contributors in the Chinese and English

Wikipedia. The visual maps of the blue cluster (Figure 3d) illustrate this difference more clearly.

Specifically, the blue cluster was centered about *technology* and *dynasty* in *SCC* and *SEC*

respectively. On the one hand, the Chinese contributors emphasize the development and

achievements of advanced lasers, aerospace, the high-speed railway, and information technology

after China's reform and opening up. On the other hand, the English contributors showed great

interests in portraying the history, culture, and science and technology of ancient China but the

description of its development in modern China was at the peripheral of the blue cluster in *SEC*.

Exceed USA VS Economic Issue

Moreover, the green clusters (Figure 3e) in *SCC* and *SEC* all emphasized the prominent role of China in the global economy. However, the green cluster in *SCC* expressed inclinations to emphasize the competitive relations between China, the United States and other developed countries, such as Japan. Specifically, in *SCC*, the verbs *become* and *exceed* were closely associated with the *USA* and *Japan*. In contrast, in *SEC*, when describing the relations of China with the United States and Japan, the English contributors used the phrase "*remain behind*." They wrote: "Science and technology are seen as vital for achieving China's economic and political goals, and are held as a source of national pride to a degree sometimes described as 'techno-nationalism.' Nonetheless, China's investment in basic and applied scientific research remains behind that of leading technological powers such as the United States and Japan." ("China," 2016) Furthermore, in *SEC*, while the economic growth of China is mentioned, the English contributors also raised issues resulting from the rapid development of the economy, such as issues of air, water, and other environmental pollutions.

4.4. Food Security, Education Law, Anti-corruption, and Citizen Right

The concept *law* ranked number nine among the 20 most frequent concepts with the greatest normalized eigenvector centralities and was unique in *SCC*. Besides the above five clusters, the *SCC* also has a unique cluster centered about the concept *law*. In Figure 3f, the concept *law* had a close association with aspects of food security, education, anti-corruption, constitution and citizen rights. These concepts are often criticized by foreign media. To some extent, the existence of this unique cluster reflects how the Chinese contributors address their attitudes and opinions toward critiques from the outside. For example, Transparency International (TI) has published the Corruption Perceptions Index (CPI) since 1995, and China was recognized as the highest perceived levels of corruption and ranks one hundred in 2014.

However, in the Chinese Wikipedia, it was written that: "the CPI ranks of China in 2014 was seriously inconsistent with the facts that China has achieved significant progress in anti-corruptions…To solve the food security related problems, the Chinese government will integrate multiple institutions to accelerate the establishment of new food standards and industrial system of food security, and meanwhile start to promote the scientific knowledge of food security." ("China," 2016)

5. Conclusion

In conclusion, this paper conducted an inter-language semantic network analysis to examine the semantic differences that emerged from articles on China in the Chinese and English editions of Wikipedia. It mapped how Chinese and English speakers illustrate the meaning of China in different ways on Wikipedia. Besides the common focuses on topics of government, population, language and character, diplomatic relations, development of economy, science and technology, the Chinese-speaking contributors and English-speaking contributors framed the article on China in different and even opposite ways according to different national cultures, values, interests, situations, and emotions.

The Chinese Wikipedia framed China as a country centered around the Communist Party and the related governmental institutions, featuring in the rapid rate of economic growth and advanced science and technology, using common language and characters, and making efforts to exceed the United States and establish formal diplomatic relations with surrounding regions, such as Taiwan. The English Wikipedia framed China as a country controlled by the Communist Party, threatening the human rights and the development of democratic society, speaking a variety of languages, with many environmental issues emerged from the rapid economic growth,

remaining behind the United States and other developed countries in the field of science and

technology, and evolving into the territory dispute with countries, such as India and Taiwan.

Methodologically, this study used QAP correlation analysis to identify the overall

similarity of the semantic networks of the articles on China in Chinese and English Wikipedia

pages. Also, it examined the salience of the concepts in the semantic networks of articles of

China based on generating the visual maps of the semantic contents and measuring the concept

centralities and associations. This paper not only confirms the existence of differences in

Wikipedia articles in multi-language editions found in previous research, but also extended the

previous studies by providing an example of studying exactly how Chinese and English-speaking

groups frame the same topic in different ways. Future research should determine if the results of

this study can be generalized to other topics between the Chinese and English Wikipedias. Future

scholars also should address the connections between the semantic networks of articles and the

user interaction networks of Wikipedia.

References

Aragón, P., Laniado, D., Kaltenbrunner, A., & Volkovich, Y. (2012, August). Biographical

 social networks on Wikipedia: a cross-cultural study of links that made history.

 In Proceedings of *the Eighth Annual International Symposium on Wikis and Open*

 Collaboration (p. 19). ACM.

Barnett, G. A., & Woelfel, J. (Eds.). (1988). *Readings the Galileo system: Theory, methods and*

 applications. Dubuque, IA: Kendall/Hunt.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for

 exploring and manipulating networks. In *ICWSM* (pp. 361-362).

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique

 identification. *Journal of Mathematical Sociology, 2,* 113–20.

 http://dx.doi.org/10.1080/0022250X.1972.9989806

Borgatti, S., Everett, M., & Freeman, L. C. (2002). *UCINET: Network analysis software.*

 Columbia, SC: Analytic Technologies.

Bryant, S. L., Forte, A., & Bruckman, A. (2005, November). Becoming Wikipedian:

 transformation of participation in a collaborative online encyclopedia. In Proceedings of

 the 2005 international ACM SIGGROUP conference on Supporting group work (pp. 1-

 10). ACM. http://dx.doi.org/10.1145/1099203.1099205.

Callahan, E. S., & Herring, S. C. (2011). Cultural bias in Wikipedia content on famous

 persons. *Journal of the American society for information science and technology*, *62*,

 1899-1915.

Cao, Q. (2007). Confucian vision of a new world order. Culturalist Discourse, Foreign Policy

 and the Press in Contemporary China. *The International communication Gazette, 69*,

431-450. http://dx.doi.org/10.1177/1748048507080868

Carley, K., & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social Forces, 70*, 601–636. http://dx.doi.org/10.1093/sf/70.3.601

China. (n.d.). In *Wikipedia*. Retrieved January 10, 2016, from http://en.wikipedia.org/wiki/China

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition in learning and memory* (pp. 117-138). New York Wiley.

Danowski, J.A. (1993). Network Analysis of Message Content. In Richards, W.D. & Barnett, G.A. (Eds.), *Progress in Communication Sciences 12* (pp. 197-220). Norwood, NJ: Albex.

D'Angelo, P. (2002). News framing as a multi-paradigmatic research program: A response to Entman. *Journal of Communication, 52*, 870–888. http://dx.doi.org/10.1111/j.1460-2466.2002.tb02578.x

Diesner, J., Aleyasen, A., Kim, J., Mishra, S., & Soltani, K. (2013). Using Socio-Semantic Network Analysis for Assessing the Impact of Documentaries. paper presented at *Workshop of Information in Networks 2013*, NYU.

Doerfel, Marya L. (1998). What constitutes semantic network analysis? A comparison of research and methodologies. *Connections 21*, 16-26.

Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication 43*, 51–68. http://dx.doi.org/10.1111/j.1460-2466.1993.tb01304.x

Eom, Y. H., & Shepelyansky, D. L. (2013). Highlighting entanglement of cultures via ranking of multilingual Wikipedia articles. *PloS one, 8*, e74554.

Fillmore, C. J. (1982). Frame semantics. In Linguistic Society of Korea (Ed.), *Linguistics in the morningcalm* (pp. 111–138). Seoul: Hanshin.

Freeman, L. C. (1979). Centrality in social networks. *Social Networks, 1*, 215-239. http://dx.doi.org/10.1016/0378-8733(78)90021-7

Hassan, S., & Mihalcea, R. (2009, August). Cross-lingual semantic relatedness using encyclopedic knowledge. In Proceedings of *the 2009 Conference on Empirical Methods in Natural Language Processing*: Volume 3-Volume 3 (pp. 1192-1201). Association for Computational Linguistics.

Hecht, B., & Gergle, D. (2009, June). Measuring self-focus bias in community-maintained knowledge repositories. In Proceedings of *the fourth international conference on Communities and technologies* (pp. 11-20). ACM.

Hecht, B., & Gergle, D. (2010, April). The tower of Babel meets web 2.0: user-generated content and its applications in a multilingual context. In*Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 291-300). ACM. http://dx.doi.org/10.1145/1753326.1753370.

Hua, S. (2001). *Chinese Political Culture, 1989-2000*. M E Sharpe Inc.

Jiang, K., Barnett, G. A., & Taylor, L. D. (2014, Dec). Culture Frames in International News Coverages: A Semantic Network Analysis, Paper to be presented at *the Daegu (Korea) Gyeongbuk International Social Network Conference*.

Kolbitsch, J., & Maurer, H. A. (2006). The Transformation of the Web: How Emerging Communities Shape the Information we Consume. *J. UCS, 12*, 187-213.

Laufer, P. (2014). Cultural Similarity, Understanding and Affinity on Wikipedia Cuisine Pages (Doctoral dissertation, Master Thesis in progress).

Marsella, A. J. (2011). The United State of America: "A culture of war". *International Journal of Intercultural Relations, 35*, 714-728. http://dx.doi.org/10.1016/j.ijintrel.2011.09.007

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review 63*: 81–97. http://dx.doi.org/10.1037/h0043158. PMID 13310704.

Miroff, B., Seidelman, R., Swanstrom, T., & Luca, T.D. (2009). *The Democratic Debate: American Politics in an Age of Change*. Cengage Learning.

Nemoto, K., & Gloor, P. A. (2011). Analyzing cultural differences in collaborative innovation networks by analyzing editing behavior in different-language Wikipedias. *Procedia-Social and Behavioral Sciences, 26*, 180-190.

Ortega, F., Gonzalez-Barahona, J. M., & Robles, G. (2008, January). On the inequality of contributions to Wikipedia. In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual (pp. 304-304). IEEE. http://dx.doi.org/10.1109/HICSS.2008.333.

Overell, S. E., & Rüger, S. (2011). View of the world according to Wikipedia: Are we all little Steinbergs?. *Journal of Computational Science, 2*, 193-197

Panciera, K., Masli, M., & Terveen, L. (2014, August). Cream of the crop: Elite contributors in an online community. In Proceedings of The International Symposium on Open Collaboration (p. 21). ACM. http://dx.doi.org/10.1145/2641580.2641609.

Pfeil, U., Zaphiris, P., & Ang, C. S. (2006). Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer‐Mediated Communication, 12*, 88-113.

Pye, L. (1992). *The Spirit of Chinese Politics*. Harvard University Press.

Rask, M. (2008). The reach and richness of Wikipedia: Is Wikinomics only for rich

countries?. *First Monday*, *13*.

Said, E. (1978). Orientalism. 1978. *New York: Vintage*, *1994*.

Steiner, T. (2014, August). Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global

Study of Edit Activity on Wikipedia and Wikidata. InProceedings of The International

Symposium on Open Collaboration (p. 25).

ACM. http://dx.doi.org/10.1145/2641580.2641613.

Warncke-Wang, M., Uduwage, A., Dong, Z., & Riedl, J. (2012, August). In search of the ur-

Wikipedia: universality, similarity, and translation in the Wikipedia inter-language link

network. In Proceedings of *the Eighth Annual International Symposium on Wikis and*

Open Collaboration (p. 20). ACM.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis*. Cambridge: Cambridge University

Press. http://dx.doi.org/10.1017/CBO9780511815478