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

Ke Jiang
University of
California, Davis
kejiang@ucdavis.edu

Grace A. Benefield
University of
California, Davis
grbenefield@ucdavis.edu

Junfei Yang
University of Science and
Technology of China
jfyang2010@gmail.com

George A. Barnett
University of
California, Davis
gabarnett@ucdavis.edu

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. It explores the differences in the content of Wikipedia through (a) correlation analysis of semantic networks and (b) the salience of semantic concepts through their network centralities. The results suggest there is high dissimilarity between the semantic content of the English and Chinese versions of articles on China. While both pages focused on government, population, language, character, diplomatic relations, development of the economy, and science and technology, the Chinese-speaking and English-speaking contributors framed the article on China differently—according to dissimilarities in cultures, values, interests, situations, and emotions of different language groups. This research contributes to the literature and understanding of how culture of different language groups influences the process of crowdsourcing knowledge on online collaboration platforms.

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 [1]. People with different cultural, social, national and linguistic backgrounds contribute to various 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 because millions of people read, contribute, and edit these articles. Cultural values and meaning are transmitted through this collaboration platform, allowing scholars

to analyze rich and detailed digital trace data. Fitting with its role as an open source collaboration platform, Wikipedia's repositories of various cultural values can also be easily obtained from different language versions of Wikipedia articles.

Previous research on inter-language analysis of Wikipedia articles mainly studied the geographic focus [2, 3], famous and prominent people [4-6], historical figures [7], editing behavior [1], user interaction [8], and article structure [9] of Wikipedia editions in different languages. Additionally, concept overlap [10], inter-language links [11] between Wikipedia versions in multiple languages have been examined. Many of these studies confirmed the existence of regional bias and differences in different language editions of Wikipedia. However, insufficient attention has addressed how contributors speaking different languages frame the content on Wikipedia according to their cultural preference and interests. This paper conducts a computer-assisted semantic network analysis: which examines and visualizes concept centralities and association to determine the salience of concepts in a text [12]. Here, semantic network analysis is used to explore and map the similarities and differences of the Wikipedia articles on China in both Chinese and English, thus providing an example of studying how language groups frame the same topic differently.

2. Inter-language Analysis of Wikipedia Article

Articles in Wikipedia are the outcome of a continuous collaborative effort of many volunteer contributors who speak different languages. Previous studies have analyzed 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 [13]. Also, only a small proportion of users produced most of the content of Wikipedia articles [14]. 90% of users contributed less than 10% of the overall

HI CSS

contributions in all language editions [15]. The core group of contributors does not consist of experts, but people who are heavily involved in Wikipedia and interested in open collaboration and free knowledge exchange [16].

Besides the similar patterns of contributors' activities, scholars have also studied the bias in Wikipedia across different language editions. Hecht and Gergle [17] found that across 25 different Wikipedia language versions, only 0.12% of all concepts (6,966) are covered in all 25 language editions. Rask [18] examined the connection between the activity of Wikipedia contributions and a country's development, finding 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 [3] studied how biased a particular Wikipedia page is toward speakers of its language, and confirmed the Steinberg Hypothesis that "everyone has a localized fish-eye view of the world." In particular, they found the English Wikipedia was much more biased than the Chinese Wikipedia, and attributed the small bias in Chinese Wikipedia to the fact that Wikipedia has been blocked in China at various degrees since 2004. When studying the 30 most popular articles about individuals in each language edition, Eom and Shepelyansky [5] found that local heroes are dominant, occurring 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 [4] found systematic differences related to cultural differences, histories, and values of Poland and the United States.

Network Analysis of Wikipedia Articles. Scholars have used network analysis methods to address the differences among 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. found the networks are more similar for geographically or linguistically closer communities [7]. Similarly, Hecht and Gergle [17] used in-degree sums and PageRank sums as indicators to show the existence of self-focus of 15 language editions. Warncke-Wang et al. used the inter-language link network to measure the similarities and differences among all of Wikipedia language editions and found that language similarity was positively correlated with the similarity between editions [10]. Also, they found the concept overlaps across different language platforms were about general topics, mainly countries, cities, and lists of events. Nemoto and Gloor [8] 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 showed a less stable collaboration network than their Western counterparts.

Although this research has been significant in using advanced 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 meaning of the semantics. Semantic network analysis describes a wide variety of "computer supported solutions" that enable scholars to "extract network of concepts" from texts and discern the meaning represented [19]. Semantic social network analyses have been used to identify trends in online content on YouTube [20], and predict popular concepts based on content from blogs and forums [21]. Jiang, Barnett, and Taylor used semantic network analysis to analyze the patterns of keyword associations in the coverage of the Arab Spring in the United States and in China to examine the cultural characteristics reflected in news frames [12]. This paper uses 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 on 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 [22] and the tradition of frame semantics in linguistics [23] scholars have argued that words are hierarchically clustered in memory [24]. Thus, spatial models that illustrate the relations among words are representative of meaning [25]. The structured semantic representations of multiple connections between various concepts are regarded as semantic networks [26]. Therefore, semantic network analysis (SMA) is a form of content analysis that identifies the network of associations between concepts expressed in a text [27, 28].

This paper focuses on analyzing the salience of concepts in textual mediums. The word salience means "making a piece of information more noticeable, meaningful, or memorable to audiences" [29]. A textual medium can make a piece of information more salient through repetition or by placing it in a prominent position in the semantic structure of the content [29]. A textual medium may also "make bits of information more salient by associating them with culturally familiar symbols" [29].

SMA is a useful method for analyzing the salience of a concept. It can generate a visual map of the semantic structure of the content, illustrating both the position of individual concepts appearing in a text and the complex associations among them. Specifically, the salience of a concept can be measured by analyzing its centrality in the text. For example, Jiang et al. [12] compared the salience of the concepts related to democracy in coverage of the Arab Spring in the United States and China by analyzing the centralities of concepts, such as *democracy*, *free*, *rights*, *election*, and *vote* in semantic networks of coverage of the Arab Spring from *The Associated Press* and *Xinhua News Agency*.

Measures of centrality reflect the location and the importance of a concept in relation to other concepts in the network [30, 31]. SMA also examines the concept associations by looking at the communities or the concept clusters that compose the semantic networks and the frequency with which concepts co-occur nearby. Additionally, through analyzing the relationships between different semantic networks, the similarity of various texts can be determined.

This research uses semantic network analysis to examine the differences and similarities of the content of articles on *China* in Chinese and English Wikipedia pages.

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

4.1. Articles of China in Wikipedia and Research Questions

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 increased dialogue between cultures.

This discourse helps to mix international cultural values and preferences into a dominant worldview. China—with its emergence as an economic power—may also have increased its influence on this worldview, on or off the Internet.

Because the concept of "China" carries a rich connotation of the Chinese culture globally, an analysis of the content about China in Chinese and English Wikipedia provides a convenient and efficient way to reveal cultural bias of Chinese and English speakers.

According to the Wikimedia traffic analysis report, the edits of English Wikipedia mainly come from English-speaking countries including the United States, United Kingdom, Canada, and Australia [32]; and the edits of Chinese Wikipedia mainly come from Chinese-speaking countries and regions including Mainland China, Hong Kong, and Taiwan [33]. Although Wikipedia has been sporadically blocked in China, many editors from Mainland China reported that they could access Wikipedia using proxy servers. From November 2011 to March 2014, the percentage of edits of Chinese Wikipedia from Mainland China increased from 20.9% to 43.4% [33]. This paper assumes that articles about China in Chinese and English Wikipedia reflects how Chinese and English language groups illustrate the image of China according to different cultures, values, and interests.

The articles about China, in both Chinese and English, were retrieved on January 13, 2016 from Wikipedia. Table 1 describes the general information on the editing history of each article. The English page has more contributors and 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	English
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

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 to identify different framing strategies adopted by the Chinese-speaking and English-speaking contributors:

R$_1$: What is the relationship between *SCC* and *SEC*?

R₂: How many concept clusters are there in *SCC* and *SEC*?

R₃: What are the most central concepts in each cluster?

R₄: What are the most frequent concept associations in each cluster?

4.2. Procedures

The first step in the research process was to clean 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 stemmed, 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. Miller argued that people's working memory had a capacity of "seven plus-or-minus two" chunks, indicating people can process seven meaningful units, plus or minus two, at a time [34]. Based on this argument, 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 [35]. The analysis was restricted to 150 words to keep it parsimonious and at the same time provide sufficient depth to clearly identify the similarities and differences among the semantic networks between groups. The first two steps were conducted using the *ConText* software [36] and Python.

In the third step, the two semantic networks were examined through *UCINET* [37] and *Gephi* [38], which are software for network analysis, graphics, and statistical computing.

UCINET uses Quadratic Assignment Procedure (QAP) correlation analysis to calculate the relationship between *SCC* and *SCE*. QAP correlation is similar to the traditional correlation analysis. The only difference is that QAP uses a nonparametric technique that does not rely on assumptions of independence to determine significance. The algorithm proceeds in two steps [37]. 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 the first step. A small proportion (< .05) suggests that there is 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. Degree is the total number of direct links to or from a concept. Eigenvector centrality indicates a concept's overall network centrality [39]. A concept's eigenvector centrality increases relatively if it is tied to more central concepts.

Gephi [38] calculates the clusters of networks by conducting modularity analysis [40] that measures how well a network is compartmentalized into sub-networks, and creates visual maps of semantic networks. In the visual maps, the size of each word's label depends on its degree centralities (the total number of direct links), such that the larger the object, the more central a word is in 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.

4.3. Results

Table 2 illustrates the overview of *SCC* and *SCE*. *SCE* is denser and with stronger concept links. This is probably because the English version of article on China is longer and has more contributors and edits. 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 a non-significant correlation (r = .02, p = .055) between the two semantic networks, indicating great dissimilarity between the semantic content of the English and Chinese versions of article on China.

For **R₂**, from Table 2, there are six concept clusters in *SCC*, and there are five clusters in *SCE*. Graphic representations of *SCC* and *SEC* are presented in Figure 1a and 1b. Different colors represent the various clusters of the semantic networks.

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	Food	leader	power
Asia	Free	level	reform

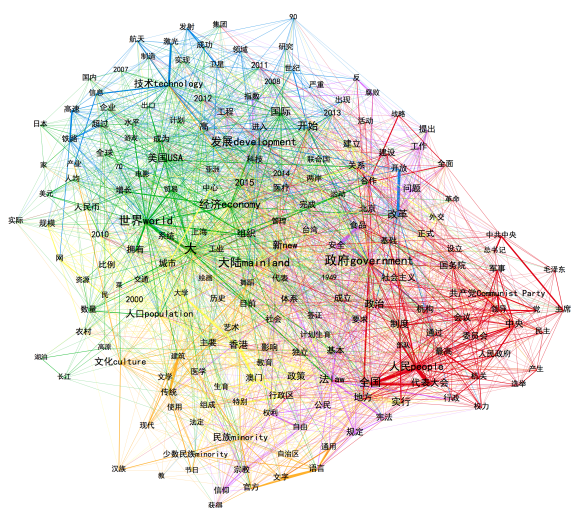|			
begin	Global	mainland	relation
Beijing	government	Mao	religion
central	Group	Macao	research
century	Han	military	revolution
character	Health	minority	Security
citizen	High	modern	Shanghai
CommunistParty	HK	movement	social
congress	increase	network	system
culture	influence	official	Taiwan
development	international	organization	technology
economy	issue	plan	trade
education	Japan	policy	USA
establish	language	politics	
family	largest	population	

Figure 1a. Graphic representations of SCC
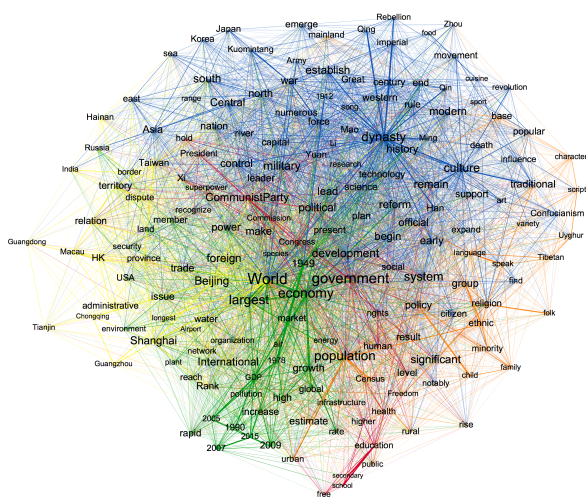
Figure 1b. Graphic representations of SEC

To answer **R₃**, Table 4 illustrates the five most central words in each cluster and their eigenvector centralities in the whole networks, as well as the percentage that the number of words in each colored cluster out of the total number of words (150) in *SCC* and *SEC*.

Table 4. The Percentage of the Number of Words and Eigenvector Centralities of the 5 Most Central Words of Each Cluster in SCC and SEC

	SCC 22%	SEC 10%
Red	government .30	government .22
	people .23	politics .15
	Communist Party .14	Communist Party .15
	central .14	rights .10
	system .11	education .08
	SCC 12%	**SEC 18%**
Orange	nationality .16	policy .13
	new .16	official .13
	culture .13	population .11
	language .11	religion .10
	character .10	ethnic .10
	SCC 17%	**SEC 14%**
Yellow	mainland .27	foreign .16
	HK .17	Beijing .14
	population .14	international .14
	relation .13	Shanghai .13
	Macao .11	territory .11
	SCC 13%	**SEC 39%**
Blue	development .21	Dynasty .19
	begin .19	culture .17
	technology .16	military .17
	opening .10	development .17
	high .14	war .14
	SCC 22%	**SEC 19%**
Green	largest .29	world .24
	economy .22	economy .21
	world .22	largest .19
	USA .16	growth .14
	organization .15	GDP .05
	SCC 14%	
Purple	law .20	food .10
	security .11	right .07
	provision .11	

To answer **R₄**, Tables 5a-5f illustrated the 5 strongest word associations in each cluster.

Table 5a. The 5 Strongest 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

Table 5b. The 5 Strongest 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

Table 5c. The 5 Strongest 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

Table 5d. The 5 Strongest 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

Table 5e. The 5 Strongest 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 5f. The 5 Strongest Word Associations in the Yellow Cluster in SCC

food security	24	anti corruption	8
free right	11	religion faith	7
law constitution	10		

4.4. Implications

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 are substantially different.

4.4.1. Authority VS Democracy. "*Government*" is the first and second most central word in *SCC* and *SEC*. Also, it is the most central concept of the red clusters in both *SCC* and *SEC*. But the number of words of the red cluster in *SCC* is twice as large as in *SEC* (Table 4), indicating that the Chinese contributors allocate much more attention to government issues. The government clusters include words that different language groups tend to associate together on these pages. Interestingly, these associations are culturally distinct, replicating political-cultural values of the respective cultures. While the Chinese contributors portray "government" as the central and most important institution indicating the cultural value of respecting authority, the English contributors interpret it from the perspective of human rights, which is one of the key elements of democracy.

The orange clusters have a slightly different focuses in each of the pages, with words about nationality in *SCC* and population in *SEC*.

Both the clusters discuss language and character but with a different emphasis. While language and character are closely associated with *common* in *SCC* (Table 5b), they have close associations with *variety*, *minority* and *ethnic groups*, such as Uyghur and Tibet in *SEC* (Figure 2a). Each of these emphases reflects Chinese values of communality/respecting authority and Western values of diversity, respectively.
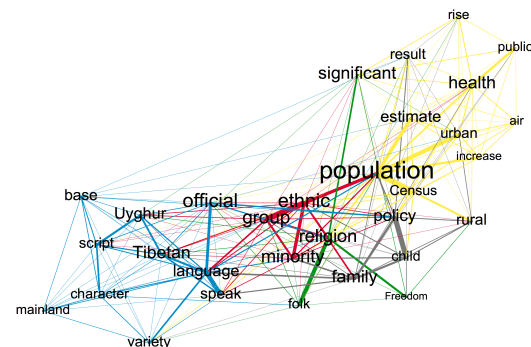

Figure 2a. Orange Cluster in SEC

In the Chinese page, there is a discussion of how the Han language and character has become the dominant and official language and character in Chinese society after thousands of years. The contributors seem to assume that respecting a

common language becomes the responsibility of the Chinese citizens.

In the English page, although the contributors acknowledged the official role of Han language, they also valued the existence of diversity, emphasizing that "there are as many as 292 living languages in China," including ethnic minority languages in Tibet, Xinjiang, and also southwest, northeast, northwest of China [41].

4.4.2. Cooperation VS Dispute. In the yellow cluster, *mainland* and *foreign* are the most central words in *SCC* and *SEC* in terms of eigenvector centralities, respectively. Both the two yellow clusters map the network linking China's big cities, two Special Administrative Regions (Hong Kong, Macao), and Taiwan. However, different relations are expressed between China and Taiwan in *SCC* and *SEC*. The Chinese-speaking contributors emphasize the establishment of formal diplomatic cross-strait relations (Table 5c). Also, this focus demonstrates the main body of Chinese Wikipedia editors that come from Mainland and Taiwan had engaged in editing conflicts over political topics related to Cross-Straits relations. Slightly different from establishing cross-strait relations, the English contributors focus on the territory dispute in Taiwan and they associate Taiwan closely with the concept foreign (Table 5c, Figure 2b).

The Chinese and English contributors' descriptions of the relations between China and Taiwan reflect the ideas of two different foreign policies. The Chinese page describes the free trade agreement and the integration and convergence of cultural industry between China and Taiwan. In contrast, 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." [41].

These contextual differences are reflective of the common ideologies in Chinese-speaking countries and the English-speaking western countries. Starting in the 21st century, Confucian rhetoric has emphasized pacifism as part of the image of Chinese culture. Confucian rhetoric proposes that harmony facilitates common development and growth [42]. However, in many English-speaking countries, democracy frame foreign policies and is highly valued. For example, the United States is regarded as the global leader and guardian of democracy and freedom [43]. Also, the commitment to democracy 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" [44]. These differences may be affecting how Chinese-speaking contributors and English-speaking contributors frame the relations between China and Taiwan.

4.4.3. Nationalism. The different semantic content between *SCC* and *SEC*'s blue clusters and green clusters express the nationalism and patriotic emotions of Chinese contributors.
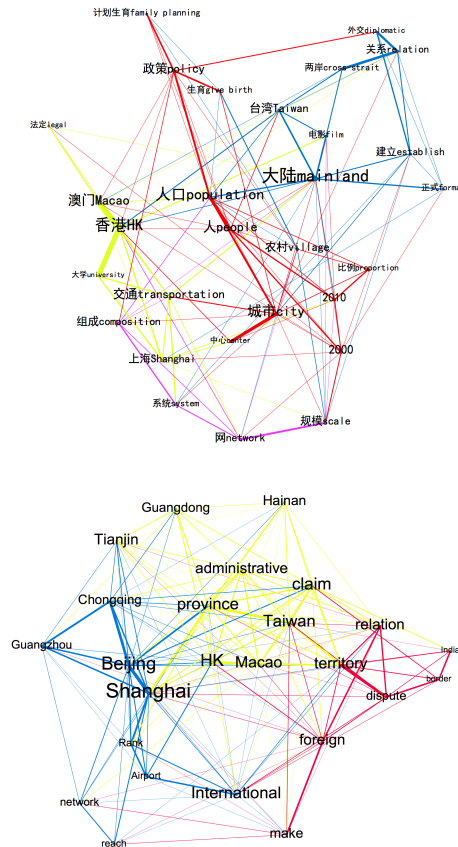

Figure 2b. Yellow Cluster in SCC and SEC

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 to the world. However, the English contributors tend to portray history, culture, and science and technology of ancient China, while the modern technological development is at the peripheral of the blue cluster in *SEC* (Table 5d).

Exceed USA VS Economic Issue. Moreover, the green clusters in *SCC* and *SEC* both emphasize the

2239

prominent role of China in the global economy. However, the green cluster in *SCC* emphasizes the competitive relations between China, the United States, and other developed countries, such as Japan. Specifically, in *SCC*, the verbs *become* and *exceed* are closely associated with the *USA* and *Japan* (Table 5e, Figure 2c). In contrast, in *SEC*, when describing China's relations with the United States and Japan, the English contributors uses the phrase *remain behind*, "China's investment in basic and applied scientific research remains behind that of leading technological powers such as the United States and Japan" [41]. Furthermore, in *SEC*, while the economic growth of China is mentioned, the English contributors also raised concerns about the rapid development of the economy, including air, water, and other environmental pollution (Figure 2c).
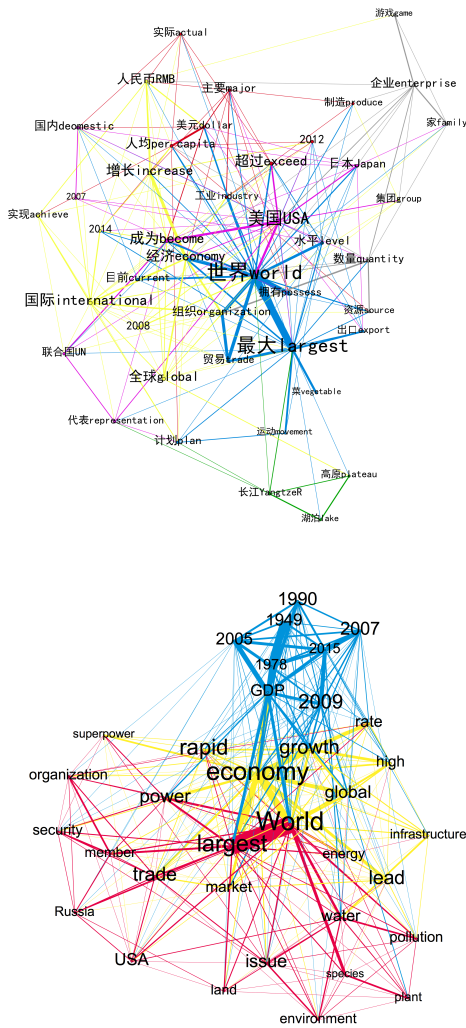

Figure 2c. Green Cluster in SCC and SEC

4.4.4. Food Security, Education Law, Anti-corruption, and Citizen Right. The concept of *law* is unique in *SCC*. In addition to the five clusters mentioned above, the *SCC* has a unique cluster centered about the concept *law*. In Figure 2d, the concept of *law* is closely associated with aspects of food security, education, anti-corruption, constitution, and citizen rights. Western media are often critical of these issues in China.

To some extent, this unique cluster reflects how the Chinese contributors defend their country against common critiques from citizens from other cultures. For example, Transparency International has published the Corruption Perceptions Index (CPI) since 1995, and China was recognized as having the highest perceived levels of corruption in 2014 [45]. However, Chinese contributors wrote 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. [46]" The strong associations between *law*, *food*, and *security* also reflected the public concern about Chinese food security that due to the environmental and agricultural pollution.
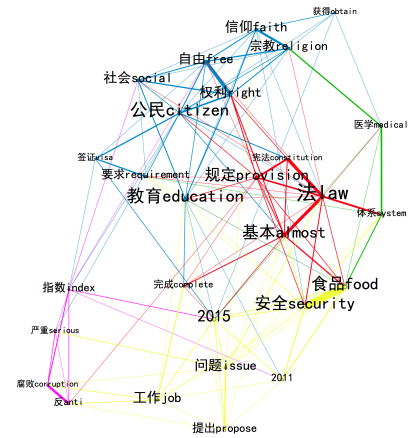

Figure 2d. Purple Cluster in SCC

5. Conclusion

In conclusion, this paper conducted an inter-language semantic network analysis to examine and map how Chinese and English speakers illustrate the meaning and image of China in different ways on

Wikipedia pages. Both pages are focused on government, population, language and character, diplomatic relations, development of economy, science and technology topics. However, the Chinese-speaking contributors and English-speaking contributors frame the article on China in different and often opposing ways.

The findings of this study not only confirms the existence of differences in Wikipedia articles in multi-language editions found in previous research, but also extent the knowledge of the regional differences of Wikipedia articles to the meanings of the content that are influenced by cultures, values, interests, situations, and emotions of different language groups. In particular, the Chinese-speaking contributors of Wikipedia framed the article from the perspectives of respecting authority and emphasizing harmony, which are aspects of cultural essence of Chinese-speaking countries and regions. The Chinese contributors also demonstrated the emotion of patriotism that might due to the rapid development of economy and science and technology in China. The English-speaking contributors expressed their emphasis on the Western-society's core value of democracy and thus demonstrated critical attitudes toward the authority of the Chinese central government and Communist Party in terms of human rights and territorial dispute.

On the one hand, the crowdsourcing knowledge emerged from online platform needs to be used with caution because of the regional cultural and ideological bias. On the other hand, online collaboration forums, such as Wikipedia, have provided an important channel for different language groups to negotiate the dominant world culture in the process of crowdsourcing knowledge according to their personal preferences, which are enmeshed in specific cultural and linguistic contexts. Because Wikipedia is continually updated, in the future, it would be meaningful to examine if the differences between the English and Chinese version of the Wikipedia articles become smaller due to globalization and cultural convergence [12]. It is also interesting for scholars to compare the content of English Wikipedia with the content of Baidu Baike, which is a Chinese-language collaborative web-based encyclopedia and more representative of mainland Chinese communities.

Methodologically, this study used QAP correlation analysis to identify the overall similarity of the semantic networks between the articles on China in Chinese and English Wikipedia pages. Also, it examined differences in the two Wikipedia pages by analyzing the salience of the concepts, which is operationalized through calculating concept

centralities and associations the above in semantic networks. Miller's argument about people's working memory and the related concept of chunk of information [34] has been applied to this study to establish the semantic networks, which is created by identifying 150 frequent words with co-occurrences within seven concepts. However, people's working memory differs due to different language and culture [47]. Future research of inter-language semantic network analysis also should test the results using different units of analysis, such as 3, 5, and 9, to construct the word co-occurrence matrix.

6. References

[1] U. Pfeil, P. Zaphiris, and C.S. Ang, "Cultural differences in collaborative authoring of Wikipedia", Journal of Computer-Mediated Communication, 12, 2006, pp. 88-113.

[2] B. Hecht, and D. Gergle, "Measuring self-focus bias in community-maintained knowledge repositories", In Proceedings of the fourth international conference on Communities and technologies, June 2009, pp. 11-20.

[3] S.E. Overell, and S. Rüger, "View of the world according to Wikipedia: Are we all little Steinbergs?", Journal of Computational Science, 2, 2011, pp. 193-197.

[4] E.S. Callahan, and S.C. Herring, "Cultural bias in Wikipedia content on famous persons", Journal of the American society for information science and technology, 62, 2011, pp. 1899-1915.

[5] Y.H. Eom, and D.L. Shepelyansky, "Highlighting entanglement of cultures via ranking of multilingual Wikipedia articles", PloS one, 8, 2013, pp. e74554.

[6] J. Kolbitsch, and H.A. Maurer, "The Transformation of the Web: How Emerging Communities Shape the Information we Consume". J. UCS, 12, 2006, pp. 187-213.

[7] P. Aragón, D. Laniado, A. Kaltenbrunner, and Y. Volkovich, "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, August 2012, pp. 9.

[8] K. Nemoto, and P.A. Gloor "Analyzing cultural differences in collaborative innovation networks by analyzing editing behavior in different-language Wikipedias", Procedia-Social and Behavioral Sciences, 26, 2011, pp. 180-190.

[9] P. Laufer, C. Wagner, F. Flöck, and M. Strohmaier, "Mining cross-cultural relations from Wikipedia-A study of 31 European food cultures", arXiv preprint arXiv: 1411.4484, 2014.

[10] M. Warncke-Wang, A. Uduwage, Z. Dong, and J. Riedl, "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, August 2012, pp. 20.

[11] S. Hassan, and R. Mihalcea, "Cross-lingual semantic relatedness using encyclopedic knowledge", In Proceedings

of the 2009 Conference on Empirical Methods in Natural Language Processing, August 2009, pp. 1192-1201.

[12] K. Jiang, G.A. Barnett, and L.D. Taylor, "Dynamics of Culture Frames in international News Coverage: A Semantic Network Analysis", International Journal of Communication, 10, 2016, pp. 3710-3736.

[13] T. Steiner, "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, August 2014, pp. 25.

[14] K. Panciera, M. Masli, and L. Terveen, "Cream of the crop: Elite contributors in an online community", In Proceedings of The International Symposium on Open Collaboration, August 2014, pp. 21

[15] F. Ortega, J.M. Gonzalez-Barahona, and G. Robles, "On the inequality of contributions to Wikipedia". In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, January 2008, pp. 304-304

[16] S.L. Bryant, A. Forte, and A. Bruckman, "Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia", In Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work, November 2005, pp. 1-10.

[17] B. Hecht, and D. Gergle, "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, April 2009, pp. 291-300.

[18] M. Rask, "The reach and richness of Wikipedia: Is Wikinomics only for rich countries?", First Monday, 13, 2008.

[19] J. Diesner, and K.M. Carley, "Revealing social structure from texts: Meta-matrix text analysis as a novel method for network text analysis". In V.K. Narayanan and D.J. Armstrong (Eds.), Causal mapping for research in information technology, Idea Group Publishing, Harrisburg, PA, 2005, pp. 81-108.

[20] G. Szabo, and B.A. Huberman, "Predicting the popularity of online content", Communications of the ACM, 53, 2010, pp. 80-88.

[21] P.A. Gloor, J. Krauss, S. Nann, K. Fischbach, and D. Schoder, "Web science 2.0: Identifying trends through semantic social network analysis", Computational Science and Engineering International Conference, Vol.4, 2009, pp215-222.

[22] P. D'Angelo, "News framing as a multi-paradigmatic research program: A response to Entman", Journal of Communication, 52, 2002, pp. 870–888.

[23] C.J. Fillmore, "Frame semantics". Linguistics in the morning calm, 1982. pp. 111-137.

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

[25] Barnett, G.A., and J. Woelfel, Readings the Galileo system: Theory, methods and applications, Kendall/Hunt, Dubuque, IA, 1988.

[26] F. Schultz, J. Kleinnijenhuis, D. Oegema, S. Utz, and W. Van Atteveldt, "Strategic framing in the BP crisis: A semantic network analysis of associative frames", Public Relations Review, 38(1), 2012, pp. 97-107.

[27] K. Carley, and M. Palmquist, "Extracting, representing, and analyzing mental models", Social Forces, 70, 1992, 601–636.

[28] M.L. Doerfel, "What constitutes semantic network analysis? A comparison of research and methodologies", Connections 21, 1998, pp. 16-26.

[29] R.M. Entman, "Framing: Toward Clarification of a Fractured Paradigm", Journal of Communication 43, 1993, pp. 51–68.

[30] L.C. Freeman, "Centrality in social networks", Social Networks, 1, 1979 215-239.

[31] Wasserman, S., and K. Faust, Social Network Analysis. Cambridge University Press, Cambridge, 1994.

[32] English_Wikipedia. Retrieved from https://en.wikipedia.org/wiki/English_Wikipedia

[33] Chinese Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Chinese_Wikipedia

[34] G.A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information", Psychological Review 63, 1956, pp. 81–97.

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

[36] J. Diesner, A. Aleyasen, J. Kim, S. Mishra, and K. Soltani, "Using Socio-Semantic Network Analysis for Assessing the Impact of Documentaries", Paper presented at Workshop of Information in Networks 2013, NYU, 2013.

[37] Borgatti, S., M. Everett, and L.C. Freeman, UCINET: Network analysis software, Analytic Technologies, Columbia, SC, 2002.

[38] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks", ICWSM, 2009, pp. 361-362.

[39] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification", Journal of Mathematical Sociology, 2, 1972, pp. 113-120.

[40] V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks", Journal of statistical mechanics: theory and experiment, 2008, pp. P10008.

[41] China. In Wikipedia. Retrieved on January 10, 2016, from http://en.wikipedia.org/wiki/China

[42] Q. Cao, "Confucian vision of a new world order. Culturalist Discourse, Foreign Policy and the Press in Contemporary China", The International communication Gazette, 69, 2007, pp. 431-450.

[43] A.J. Marsella, A. "The United State of America: "A culture of war", International Journal of Intercultural Relations, 35, 2011, pp. 714-728.

[44] Said, E., Orientalism. 1978. Vintage, New York, 1994.

[45] Corruption Perceptions Index 2014. Retrieved from https://www.transparency.org/cpi2014

[46] 中华人民共和国. In Wikipedia. Retrieved on January 10, 2016, from https://zh.wikipedia.org/wiki/中华人民共和国

[47] T. Vecchi, M.L. Monticellai, and C. Cornoldi, "Visuo-spatial working memory: Structures and variables affecting a capacity measure", Neuropsychologia, 33, 1995, pp. 1549-1564.