Journal of International Logistics and Trade, Vol. 17, No. 3, 2019, 67-76

Journal of
International Logistics and Trade

A Text Mining Analysis of US-Chinese Leaders on Trade Policy

Jun Yeop Lee[a*], Juhyeon Lee[b]

[a] Department of International Trade and Regional Studies, Inha University, Korea
[b] Department of Computer Engineering, Inha University, Korea

ARTICLE INFO

Article history:
Received 11 September 2019
Revision received 22 September 2019
Accepted 24 September 2019

Keywords:
Official speeches
Text mining
TF-IDF
Topic analysis
Word hierarchy analysis
Trade

ABSTRACT

Using the methodologies of text mining, this paper examines the implications of US and Chinese policies on bilateral trade. Official speeches by political leaders of the U.S. and China on the issues of trade were collected and analytically examined for US-China gaps in major foreign policies, such as bilateral trade and the Belt and Road Initiative. In this paper, a term frequency-inverse document frequency word cloud, a network similarities index, machine learning-processed latent Dirichlet allocation (LDA), and structural equivalence are applied to examine the meanings of the speeches. The main arguments in this paper are as follows. First, the document similarity between the speeches of Chinese and US leaders appears to be completely different. Also, while the documents from Chinese leaders are considerably similar, the documents from US leaders differ by far. Secondly, LDA topic analysis indicates that China concentrates more on international and collaborative relationships, while the U.S. has more focus on domestic and economic interests. Third, from a word hierarchy analysis, the basic words used by American and Chinese leaders are also completely different. Agriculture, farmers, automobiles, and negotiations are the basic words for American leaders, but for Chinese leaders, the basic words are planning, markets, and education.

1. Introduction

US-China trade disputes have been escalating, and are expected to be long-lasting, which will influence the current orders of the global economy and geopolitics (Freund et al., 2018; National Bureau of Economic Research, 2019). One potential explanation for worsening collisions between the two superpowers is their cultural gap, which has been accepted recently by international politics experts[1]. The cultural gap hypothesis assumes that the leaders of the two superpowers have fundamentally different ways of thinking (Haynes, 2019). If the disputes originated from different mindsets, the negotiation strategies and goals of trade negotiations between the U.S. and China would have been fundamentally different from the beginning. Hence, it is expected that the disputes will last longer and cannot be resolved harmoniously (Allison, 2017; Foreign Affairs, 2017; Goldstein, 2013; Lieberthal and Jisi, 2012). The wider the cultural gap, the less the possibility of peacefully settling the disputes.

* Corresponding author: Department of International Trade and Regional Studies, Inha University, 100 Inha Ro, Nam-gu, Incheon 22212, Incheon, Korea; Email: jylee@inha.ac.kr

1 Kiron Skinner, the director of policy planning at the State Department said on “Think Tank New America”: “This is a fight with a really different civilization and a different ideology, and the United States hasn’t had that before.” Adam Taylor, World View, Washington Post, May 2, 2019. Keyu Jin, “Why China’s current trade war strategy is to keep calm and make new friends,” South China Morning Post, 12 July 2019. Available at: https://www.scmp.com/news/world/united-states-canada/article/3021286/us-state-dept-fires-kiron-skinner-who-made-clash.

https://doi.org/10.24006/jilt.2019.17.3.001

Disagreements surrounding US-China trade and the Belt and Road Initiative (BRI) might be not only due to economic conflicts, but also due to the mindset gaps between the two countries. The BRI has been under way since 2013, several years after the 2009 global financial crisis. And the BRI was coincidently implemented with the Chinese Dream slogan and the rejuvenation of the Chinese people, which was first proposed by President Xi Jinping in the 19th plenum of the CPCC held in 2013. The Chinese Dream perspective of the BRI will inescapably challenge the U.S.-dominated global system. Hence, for the implementation of the BRI, it is indispensable to reconcile cultural and policy clashes with U.S. (The World Bank, 2017; Forbes, 2017; U.S. Congress, 2018; Reuters, 2017).

This paper tries to propose an alternative viewpoint on the background for why and how the U.S. and China have collided in such an abrupt way. In this context, the aim of this paper is to systematically analyze the essence of the mindset gaps between US and Chinese political leaders, focusing on the differences in the notion of trade policy and the Belt and Road Initiative. Mindset hunting of the political leaders in our research is from text-mining analysis. Various text-mining methodologies and social network analysis are applied, in which word-word networks are the basic component. If the cultural gap hypothesis is significantly constructed, it implies that the US and Chinese disputes might even be escalating.

Text-mining techniques have developed from simple calculations of word frequencies to the identification of word-word networks and document-document network analysis and are applied to deep learning procedures in order to capture the topic of the text. The accumulation of large sets of documents from a specific speaker would make it easier and more accurate to identify patterns in word networks. Recently developed topic analysis applies deep learning methodologies, calculates the importance of words stochastically, and sorts out the topic of the document.

For this study, officially released speeches of US and Chinese leaders were collected, including those from former-president Barack Obama, US President Donald Trump, China's President Xi Jinping, former Chinese president Hu Jintao, former premier Wen Jiabao, and Premier Li Keqiang. The main contents of this paper are as follows. In Section 2, the documents of the US-Chinese leaders and the process of making speech documents into word-word networks is explained. Also, the similarity in the documents from US and Chinese leaders is presented. In Section 3, frequency-based and term frequency-inverse document frequency (TF-IDF)-based word clouds are presented. In Section 4, results of latent Dirichlet allocation (LDA) topic modeling and their implications are discussed. In Section 5, using structural equivalence, differences in the basic word components between the U.S. and China are presented.

2. Data and document similarity

2.1 Data

The data in this paper are from official presidential and prime ministerial speeches that include 50 from President Xi Jinping, 44 from Li Keqiang, 18 from Wen Jiabao, 25 from Hu Jintao, 40 from President Obama, and 41 from President Trump[2]. The years for the distribution of the speeches are from 2003 to 2018. For Chinese leaders, speeches that were officially translated into English were included in our text analysis.

From a social network analysis perspective, each speech forms a document, and a document is treated as a node. A word that is used in a document is also treated as a node. Hence, our document data basically constitute a two-mode network of document-words; that is to say, a two-mode network that is composed of a specific document and the frequency of each word in the document. To transform from a two-mode document-word network to a one-mode word-word network or document-document network, the Jaccard similarity coefficient[3] was used, by which a two-mode network is transformed into a document-document or word-word network based on the similarity of words between documents. It is one of the salient contributions to apply the social network approach in analyzing the text data. In similar presidential speeches, Cho et al. (2015) used clustering methodologies and demonstrated the changing agendas of the US inaugural addresses by clustering movements. Crockett and Lee (2012) also basically applied clustering methodologies to analyze the patterns of the State of the Union addresses from four US presidents.

In Figure 1, the original word-word network is visualized. The original data are too complex to be analyzed. Therefore, for actual analysis, it is necessary to build condensed data by link reduction. In this paper, top 10% link weight data are used.

[2] Official speeches that are included in this paper include President Trump's "weekly address on April 12, 2018", President Trump "attends an afternoon Chanukah Reception at the White House on December 1, 2018", President Xi's "Keynote address at the APEC CEO Summit Da Nang on November 10, 2017", P.M. Li's "10th East Asia Summit remarks on November 22, 2015", President Obama's "Inaugural address on January 20", President Hu's "speech at the CPC 90th anniversary gathering on July 1, 2011", and P.M. Wen's "Expo 2010 Shanghai China Summit Forum On October 31, 2010."

[3] For selected two-node row profiles R=(R_1, R_2, ..., R_n) and S=(S_1, S_2, ..., S_n), where a is the number for i with R_i =1 and S_i = 1, b is the number for i with R_i =1 and S_i = 0, c is the number for i with R_i =0 and S_i = 1, and d is the number for i with R_i =0 and S_i = 0. The Jaccard coefficient equals a/(a+b+c+d).

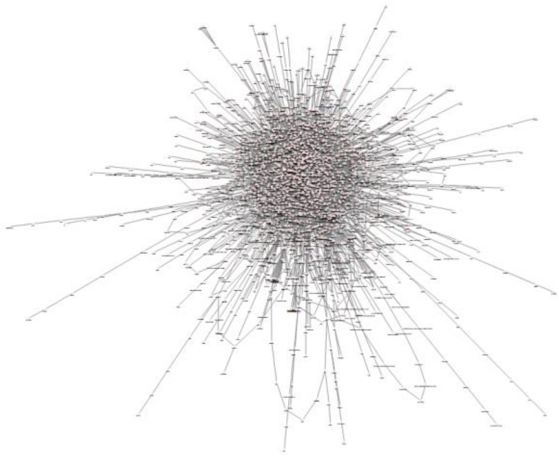


Figure 1. Original word-word network of speech documents

2.2 Document similarity

After constructing social network data from the speeches of political leaders, one immediate question is how much the speech data reveal differences in thought among the leaders. In other words, are the speeches of the leaders from the same country similar, compared to the leaders from the other country? The similarity is presented using a Kamada and Kawai (1989) drawing graph algorithm[4] as seen in Figure 2. Each node in Figure 2 represents a speech document of a specific leader, and a different node type means a document from a different leader. The closer the distance, the more likely it can be interpreted as a document with similar word characteristics.

In Figure 3, based on a partitioning around medoids (PAM) clustering algorithm[5], clustering of speech documents is presented. In Table 1, the average silhouette coefficient and the average Rand index are also presented, where a large average silhouette coefficient can be interpreted as meaning that clusters in our document data are clearly divided. In our document data, the average silhouette coefficient of the U.S. is larger than that of China, which means that speeches by American leaders are more clearly distinguished from each other than the speeches of Chinese leaders. And the large adjusted Rand index implies that similar documents are properly grouped in the same cluster. Hence, a smaller adjusted Rand index for the documents of Chinese leaders compared to US leaders means that the documents from the former are not differentiated and mingled together, compared to US leaders' speeches.

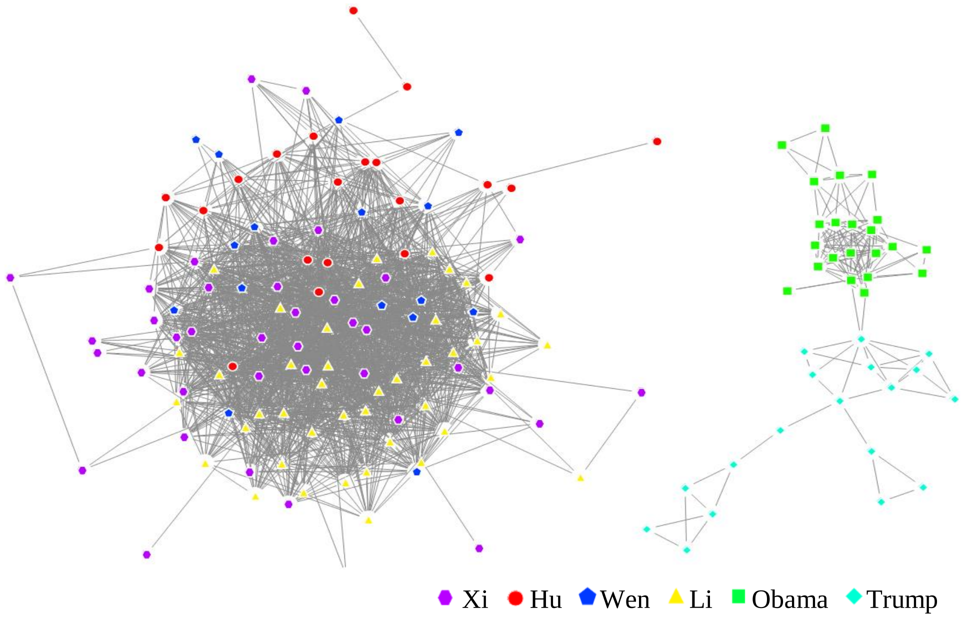


Figure 2. Similarity of speech documents

[4] In Kamada and Kawai's 1989 spring embedding algorithm, the aim is to find coordinates in which, for each pair of nodes, the Euclidean distance is approximately proportional to the geodesic distance between two nodes. That is, this algorithm tries to represent an ideal distance between nodes that are not adjacent to each other.

[5] The partitioning around medoids (PAM) algorithm is the most common realization of a k-medoids algorithm.

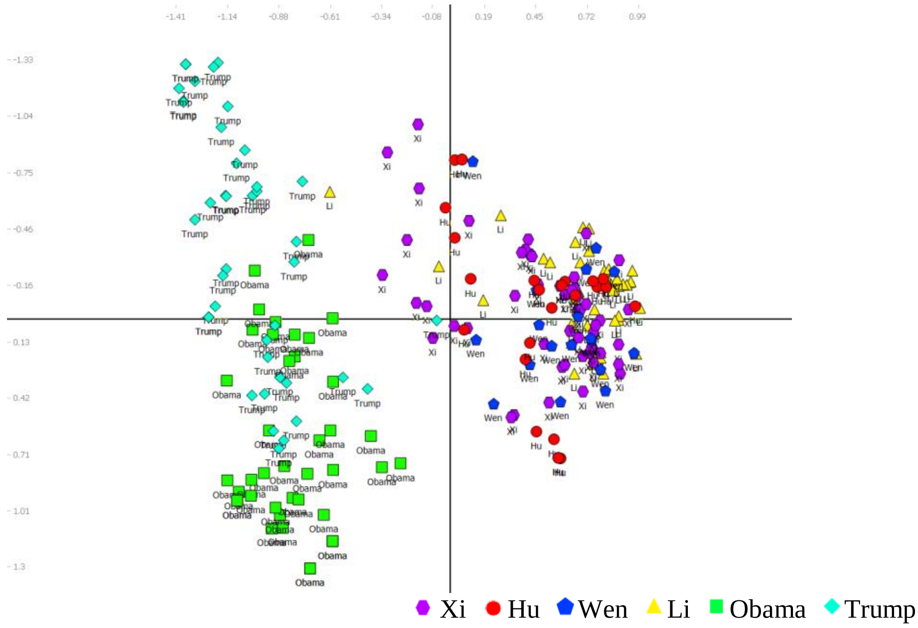


Figure 3. PAM clustering visualization

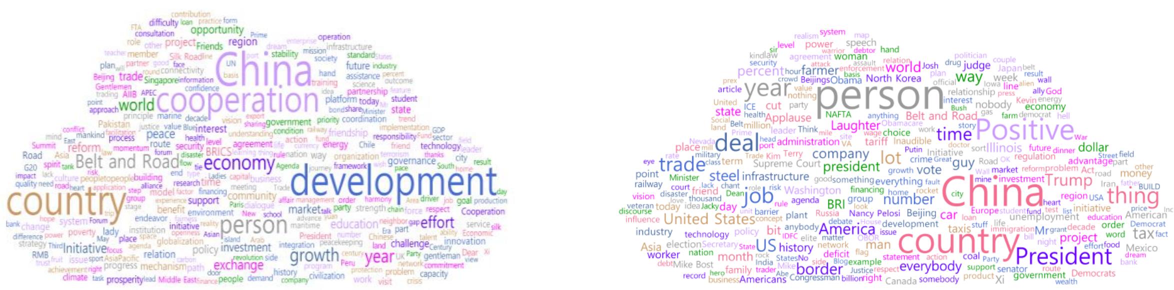
Table1. Clustering summary

	No. of documents	No. of clusters	Average silhouette coefficient	Adjusted rand index
All	218	6	0.023	0.368
China	137	4	0.019	0.103
U.S.	81	2	0.045	0.359

3. Word cloud

A widely used methodology to grasp the key contents of a document is to extract the important words from the document. In this context, the simplest method of word selection is to extract words with a high frequency. However, TF-IDF has been used recently as a useful indicator to measure how words emerge as the key semantic meaning, compared with ordinary frequency-based word clouds (Aizawa, 2003). If a particular word is used intensively in a document, it means that the word is comparatively more important. Conversely, a word appearing repeatedly in every document has a smaller TF-IDF value and is removed as meaningless. Hence, the higher the TF-IDF value, the more important the word[6].

An ordinary frequency-based word cloud and a word cloud in which TF-IDF is higher than 0.2 are presented in Figure 4 and Figure 5, respectively. Words such as dollars, power, and tariffs appear in the TF-IDF-based word cloud of the American leaders, while words such as culture, contribution, and coordination emerged as key words in the TF-IDF word cloud of the Chinese leaders. These words are not detected as important words by the standards of word frequencies.



China U.S.

Figure 4. Word frequency-based word cloud

[6] In most commonly used words (e.g. do), the TF value for it will be high, even though it is not important. To prevent this, document frequency (DF) is used to measure how frequently a word appears in other documents. As a result, TF-IDF gives tips for judging whether or not a word is important in a specific document based on word frequency and document frequency. The TF-IDF value of a word in document j is calculated as TF((number of times a word appears in document j/total number of word I in the documents)) multiplied by IDF(log(total number of documents/number of documents with the word in it)).

China U.S.

Figure 5. TF-IDF based word cloud

An ordinary frequency-based word cloud and a word cloud in which TF-IDF is higher than 0.2 are presented in Figure 4 and Figure 5, respectively. Words such as dollars, power, and tariffs appear in the TF-IDF-based word cloud of the American leaders, while words such as culture, contribution, and coordination emerged as key words in the TF-IDF word cloud of the Chinese leaders. These words are not detected as important words by the standards of word frequencies.

Compared with frequency-based key words, clear differences between US and Chinese leaders' speeches are shown in the TF-IDF key words. Higher TF-IDF words from U.S. leaders appear far more targeted and more economic interest-oriented than those from the Chinese. In this sense, American and Chinese political leaders seem to have contrasting viewpoints in their foreign policies.

4. Topic analysis and ego network

4.1 Topic modelling

Topic modeling is the statistical process of estimating the word distribution in topics, and the topic distribution in documents, by using the observed words in the documents. The basic assumption of topic modeling is that the words in a document are not chosen randomly. The topics of the documents are expressed in terms of the words. Therefore, it is assumed that the combination of the word allocation in the topic and the topic distribution in the document creates the words in the document. For example, if family love and success on the job are the key topics, the actual document is composed of words well-associated with the topic. So, through the statistical inference process, words will be arranged to reflect these two topics well.

In this paper, topic modeling of the LDA methodology[7] is applied. LDA is applied to categorize the topics of the political speeches. LDA is a machine learning process to optimally allocate documents and words to the specific topic by satisfying two trade-off conditions[8]. On the topic model used for political science, concepts of the LDA technique, which was proposed by Blei et al. (2003), have been extensively used due to hard-to-observe underlying topics, indirect political expressions, and webs of interests (Yu et al., 2008). Su (2016), using the LDA technique, analyzed President Xi Jinping's speech delivered in Singapore to investigate several positive discourses.

Table 2. Keywords from topic analysis

			President Xi		
	1st keyword	2nd keyword	3rd keyword	4th keyword	5th keyword
Topic 1	cooperation	BRI	development	exchange	Initiative
Topic 2	development	world	economy	growth	reform
			President Trump		
	1st keyword	2nd keyword	3rd keyword	4th keyword	5th keyword
Topic 1	positive	deal	China	job	tax
Topic 2	border	wall	job	security	Democrats

for j^{th} word in the i^{th} document:
choose a topic $z_{i,j} \sim$ Multinomial($p(z|d_i)$)
choose a topic $w_{i,j} \sim$ Multinomial($p(w|z_{i,j})$)

[7] Suppose that p(z|d_i), p(w|z_{i,j}) is the topic distribution for each document i for a topic allocated to the j^{th} word of document i, respectively. In the machine learning process, LDA fits p(z|d_i), p(w|z_{i,j}) to a set of documents. Given these distributions, the LDA can generate a new document with the following generative process:

[8] The first trade-off condition is that each document contains a minimum number of topics, and the second condition is that each topic contains a minimum number of words.

LDA results assuming two topics in the speeches of President Xi and President Trump are shown in Table 2[9]. The word network in our LDA only includes words in which the TF-IDF value is higher than 0.2. It is inferred that the first topic derived from Xi Jinping's speech is related to cooperation with BRI countries, and the second topic is about China's economic development and reform. On the other hand, the two topics that appeared in President Trump's speech are trade negotiations with China and border closures in order to guarantee domestic job security.

In the topic analysis, each word can be assigned a probability under each topic. For example, the word family is allocated a contribution of 60% to the first topic, and 40% to the second topic. Based on these word-topic assignments, word clouds can be drawn with words that have a higher statistical assignment to the topic (Figure 6, Figure 7). The TF-IDF-based word cloud including higher topic assignment words is presented in Figure 8.

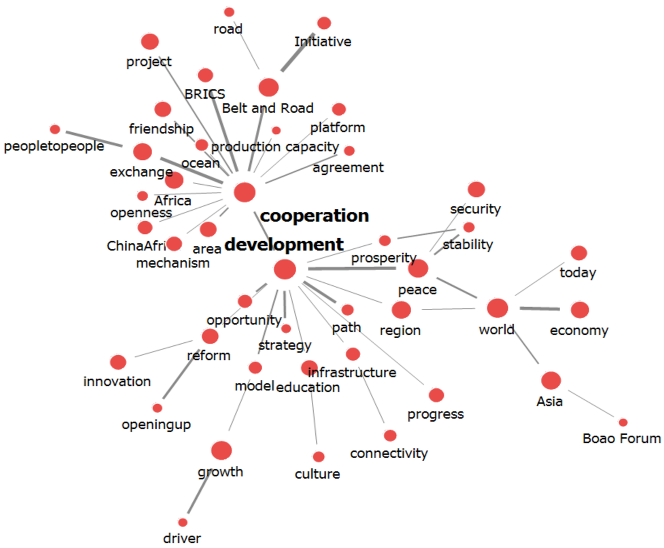


Figure 6. Words network: President Xi's topic modelling

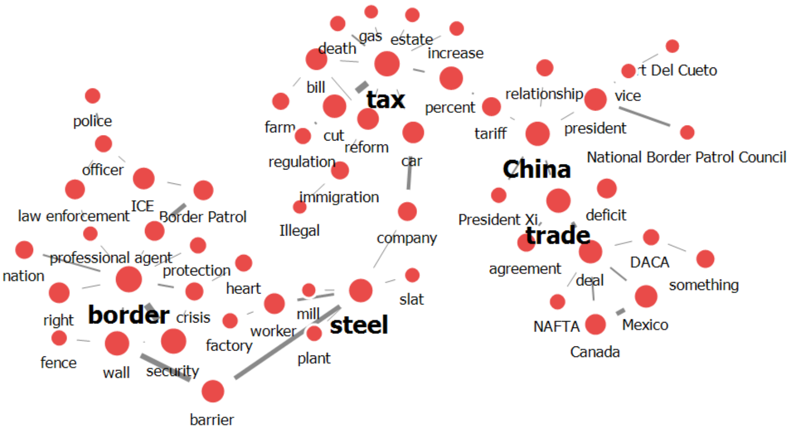


Figure 7. Words network: President Trump's topic modelling

President Xi's topic 1

President Trump's topic 1

Figure 8. TF-IDF word cloud

9 Researchers conduct topic modeling by arbitrarily setting the number of topics in advance of analysis. In this study, the number of topics was set at two.

4.2. Ego network of trade and BRI

To capture the thoughts of the leaders on the issues of trade and the Belt and Road Initiative, an ego network for trade and Belt and Road were constructed[10]. The visualization of the ego network for trade is shown in Figure 9. In China's ego network for trade, the words that have the strongest link to trade are investment and cooperation. The BRI ego network has strong links to words such as development and cooperation. On the other hand, in the US presidential speeches, the words deficit and negotiation have the strongest links to the word trade. Because there is little mention of the BRI, an ego network for BRI was not constructed from the US leaders' speeches.

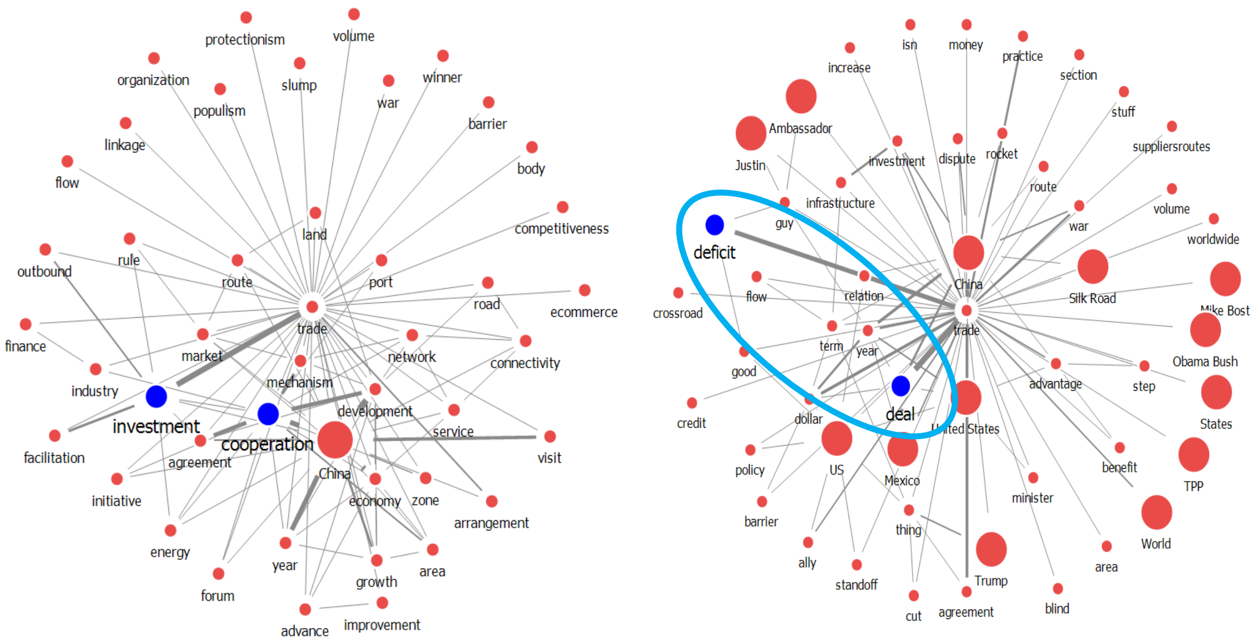


Figure 9. Ego network for *trade*

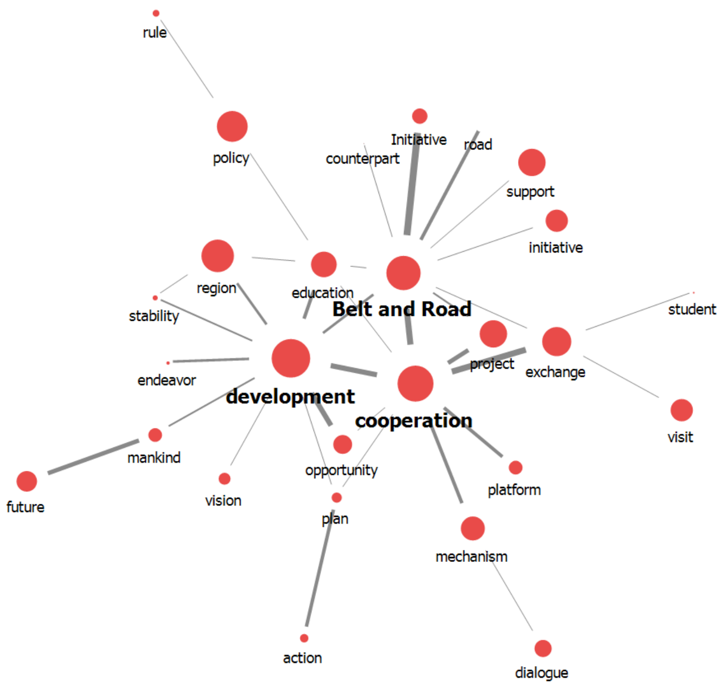


Figure 10. Ego network of *Belt and Road Initiative*

10 From the raw word network, nodes with two-step links in both directions with the word trade are extracted.

5. Word role analysis: Word hierarchy

The role of a particular node in a network is greatly affected by the node's position. In this context, if two nodes have similar hierarchies, then the behavior and role of the two nodes might be similar. Roles or hierarchies in the social network perspective identifies a structural behavior or functions in a network. Moreover, nodes that are connected to the same neighbor node and have the same connection pattern are defined as Structurally Equivalent (Lorrain and White, 1971). The basic idea of Structural Equivalence is to measure how much the in-links and out-links are matched between two nodes. As similar nodes are classified in the same group, clustering and the hierarchy structure can be built among nodes. The word hierarchies using Structural Equivalence from speeches of Chines and U.S. leaders are presented in Figure 11 and 12[11]. The hierarchy structure of words used by Chinese and American leaders in their speeches appears profoundly meaningful. The word *trade* is clustered with words such as *market growth* in Chinese leaders' speeches. On the other hand, in the speeches by United States leaders, *trade* is clustered with words such as *barrier, borders, term*. Moreover, the basic words, which lie in the lowest hierarchies, are fundamentally different between China and the U.S. In China, instead, institutional and normative words such as *education, planning, and models* constitute basic words in their speeches. In the United States, however, more practical and economic words, such as *automobiles, taxes, negotiations, and farmers* are basic in the word hierarchy.

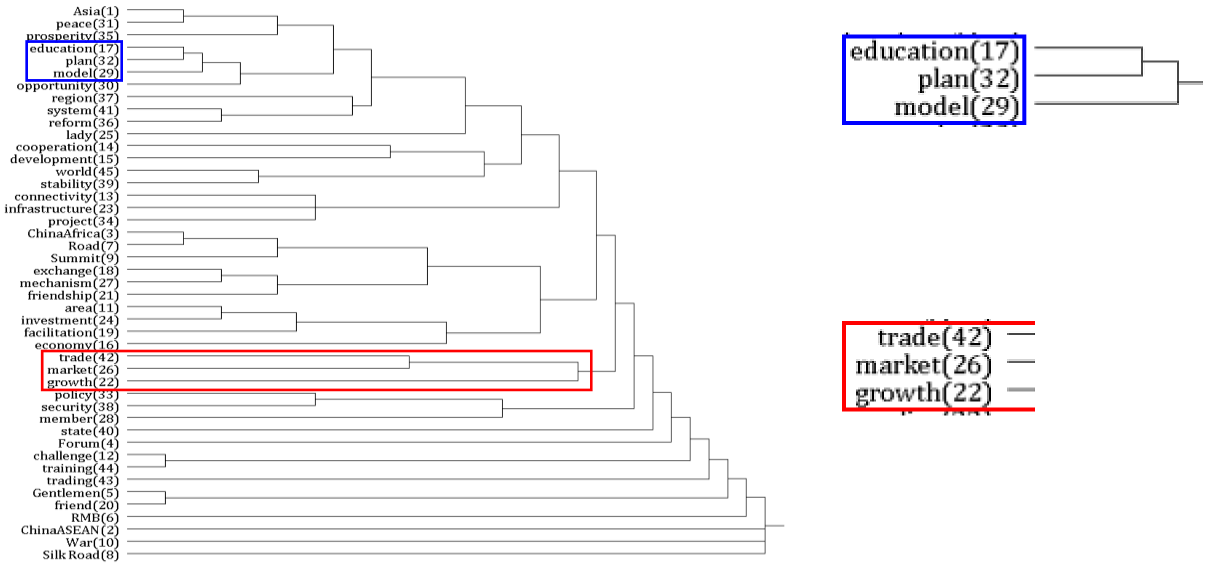


Figure 11. Word hierarchy for President Xi

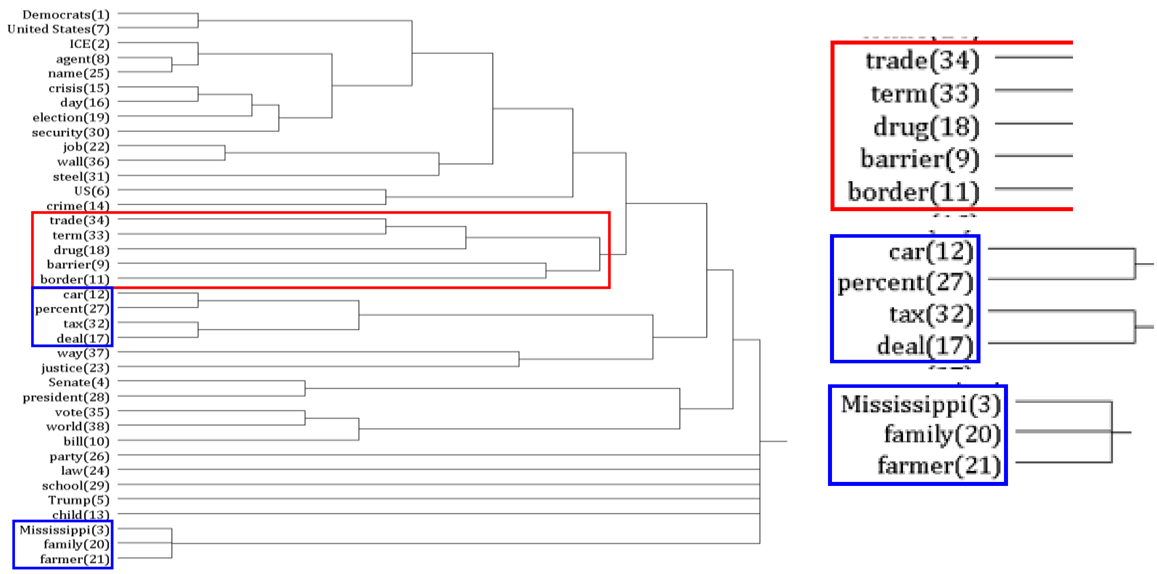


Figure 12. Word hierarchy for President Trump

[11] In our presidential speech data, more than 3,700 words have to be analyzed. For the Structural Equivalence analysis, it needs to proceed calculation of 3,700 x 3,700 matrix, too large to analyze. Hence node reduction is proceeded by choosing only non- isolated words, of which degree size is large than 1. The number of non-isolated words in our data is 264.

6. Conclusions

This study suggests that text mining analysis is a useful tool for extracting key words and grasping the word network structure from political leaders' speeches. Especially in an environment of escalating US-China rivalry, using various text mining methodologies, this research argues that there are clear differences in the thinking and policy priorities of leaders in China and the U.S.

First of all, the most striking difference in speeches by US and Chinese leaders is in their group similarity. In the case of China, even though leaders such as Wen Jiabao and Hu Jintao have been replaced, there is a very high degree of similarity. On the other hand, speeches by presidents Trump and Obama are so low-key that they are divided into completely different clusters. The implications of this analysis are for policy sustainability (as to whether the United States will maintain its current stance toward China after Trump). In the same context, the BRI pursued by the Xi Jinping administration is likely to be carried out over the long term. The reasoning of this argument is that the Chinese government's policies converge, regardless of the leader, from a text mining perspective.

Secondly, the characteristics of key words among Chinese and US leaders are clearly different. For US leaders, key words from topic modeling after TF-IDF processing are composed of words that are far more targeted and aggressive. Conversely, from topic modeling, the characteristics of the key words of Chinese leaders have more harmonious and collaborative tendencies.

Third, the ego word networks for trade and BRI show that President Xi delivers external and cooperative messages, whereas President Trump delivers domestic and economic messages in the sense that, in the US presidential speeches, the words deficit and negotiation have the strongest links to the word trade.

Lastly, from the word hierarchy analysis, the basic words in China and the U.S. are fundamentally different. The basic-word differences between the U.S. and China imply that there is a fundamental gap between the two countries in their ways of thought, where Chinese leaders prefer normative ways of thinking, while U.S. leaders think practically. And the differences in the ways of thinking might be a fundamental reason for the US-China trade war, and this might be a reason why it is difficult to find a harmonious solution to the US-China conflicts.

Acknowledgments

This work was supported by INHA University Research Grant.

References

Aizawa, A., 2003. An information-theoretic perspective of tf–idf measures. Information Processing & Management 39, 45-65.
Allison, G., 2017. Destined for War: Can America and China Escape Thucydides's Trap? Houghton Mifflin Harcourt, New York.
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning Research 3, 993-1022.
Cho, S.G., Cho, J., Kim, S.B., 2015. Discovering meaningful trends in the inaugural addresses of United States presidents via text mining. Journal of Korean Institute of Industrial Engineers 41, 453-460.
Crockett, S., Lee C., 2012. Does it matter what they say? A text mining analysis of the State of the Union addresses of USA presidents. Proceeding of the 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) 2012, Kyoto, Japan.
Forbes, 2017. China's Bridge to Nowhere. <https://www.forbes.com/sites/joelmoser/2017/09/07/chinas-bridge-to-nowhere/#7d24b60d1568>.
Foreign Affairs, 2017. China vs America: Managing the next clash of civilization. <https://www.foreignaffairs.com/articles/united-states/2017-08-15/china-vs-america>.
Freund, C., Ferrantino, M., Maliszewska, M., Ruta, M., 2018. Impacts on global trade and impacts of current trade disputes. Macroeconomis, Trade Investment (MIT) Practice Notes 2, 1-11.
Goldstein, A., 2013. First things first: The pressing danger of crisis instability in US-China relations. International Security 37, 49-89.
Haynes, J., 2019. From Huntington to Trump: Twenty-five years of the "Clash of civilizations". The Review of Faith & International Affairs 17, 11-23.
Kamada, T., Kawai, S., 1989. An algorithm for drawing general undirected graphs. International Processing Letters 31, 7-15.
Lieberthal, K., Jisi, W., 2012. Addressing US-China Strategic Distrust. Brookings, Washington, DC.
Lorrain, F., White, H.C., 1971. Structural equivalence of individuals in social networks. Journal of Mathematical Sociology 1, 49-80.
National Bureau of Economic Research, 2019. The return to Protectionism. Cambridge, MA.
Reuters, 2017. Behind China's Silk Road Vision: Cheap Funds, Heavy Debt, Growing Risk. <https://www.reuters.com/article/us-china-silkroad-finance/behind-chinas-silk-road-vision-cheap-funds-heavy-debt-growing-risk-idUSKCN18B0YS>.
South China Morning Post, 2019. The director of policy planning at the State Department answered on the Think Tank New America on Monday, <https://www.scmp.com/news/world/united-states-canada/article/3021286/us-state-dept-fires-kiron-skinner-who-made-clash>.
Su, T., 2016. Positive discourse analysis of Xi Jinping's speech at the National University of Singapore under appraisal theory. Journal of Language Teaching and Research 7, 796-801.
The World Bank, 2017. Remarks of World Bank Group President Jim Yong Kim at the Belt and Road Forum for International

Cooperation-Opening Plenary Session. <http://www.worldbank.org/en/news/speech/2017/05/14/remarks-of-world-bank-group-president-jim-yong-kim>.

Yu, B., Kaufmann, S., Diermeier, D., 2008. Classifying party affiliation from political speech. Journal of Information Technology & Politics 5, 33-48.