A Study of Comparison between Cruise Tours in China and U.S.A through Big Data Analytics

Tao Shuting and Hak-Seon Kim[†]

School of Hospitality & Tourism Management, Kyungsung University

KEYWORDS	ABSTRACT
China cruise tour, U.S.A cruise tour, Big data, Web crawling, Text mining, Semantic network analysis.	The purpose of this study was to compare the cruise tours between China and U.S.A. through the semantic network analysis of big data by collecting online data with SCTM (Smart crawling & Text mining), a data collecting and processing program. The data analysis period was from January 1st, 2015 to August 15th, 2017, meanwhile, "cruise tour, china", "cruise tour, usa" were conducted to be as keywords to collet related data and packaged Netdraw along with UCINET 6.0 were utilized for data analysis. Currently, Chinese cruisers concern on the cruising destinations while American cruisers pay more attention on the onboard experience and cruising expenditure. After performing CONCOR (convergence of iterated correlation) analysis, for Chinese cruise tour, there were three clusters created with domestic destinations, international destinations and hospitality tourism. As for American cruise tour, four groups have been segmented with cruise expenditure, onboard experience, cruise brand and destinations. Since the cruise tourism of America was greatly developed, this study also was supposed to provide significant and social network-oriented suggestions for Chinese cruise tourism.

1. INTRODUCTION

Originally, the main function of the cruise ship was to carry mail and immigration, and later the cruise ship developed into a new way of tourism. However, due to its special time, place and cost, this way of tourism was not very common in many areas, and the understanding of consumers for cruise is limited. However, in recent years, with the rapid development of the economy and the development of the port industry around the world, cruise tourism has gradually been familiar and accepted by the common all over the world (Wang, 2015). Cruise tour is an advanced tourism activity which takes the large-scale luxury cruise as the carrier, considers the sea parade as the main touring form and treats the activities on the ship and port city tours as the main touring content (Chin, 2008). Meanwhile, cruise tour offers everything that people need for vacation - romance, excitement, relaxation, adventure, escape, discovery and luxury. With the economic development and globalization, the increasing number of people with different lifestyles, interests, hobbies participated into the cruise tour. The world's cruise market grew at an average annual rate of 7.2% growing in North America (60.5%) and Europe (27%) since 1990 (Klein, 2011). It is forecasted to be 20.98 million in 2013 and then 37 million in 2020 (Ministry

of Maritime Affairs & Fisheries, 2015). In addition, there are 292 cruise ships around the world, including 4 cruise ship companies (105 Carnivals, 40 Royal Caribbean, 12 MSC, 16 Star Cruises) occupying 60% (173) of the world market. 43 vessels are operating on three global carriers, including Carnival, Royal Caribbean and Star Cruise (Ministry of Maritime Affairs and Fisheries, 2014). The cruise industry develops and distributes a variety of products linked to aviation, shipping, ship-building, port, and tourism. Over the past two decades, more than 60 million people have enjoyed safe, comfortable, and exciting cruise tours (Chin, 2008).

Recently, big data has rapidly developed into a hotspot that attracts great attention from academia, industry, and even governments around the world (Waller, & Fawcett, 2013). Big data, especially big network data, contains a wealth of societal information and can thus be viewed as a network mapped to society (Kim, 2017b). To some extent, analyzing big data and further summarizing and finding clues such as semantic network analysis about the collected keywords can help us better perceive the present. Furthermore, through effective integration and accurate semantic network analysis on multi-source heterogeneous big data, better predictions of future trend and current comparison between two subjects can be achieved. Using and mining big data heralds a new wave of productivity growth and consumer impetus. O'Reilly Media even asserted that "the future belongs to the companies and people that turn data into products" (Schutt & O'Neil, 2013).

According to statistics, in 2001, the number of Chinese cruise consumers was only 8,325 people, whereas, in 2009, the number of Chinese consumers of cruise tour who are cruising from the mainland to outbound travel was about 200,000 people, combined with Hong Kong, Singapore, the United States, Europe Boarding tourists, China's domestic cruise tourism tourists bloomed to 350,000 to 380,000 (Dowling, 2006). The United States occupies the core of the international cruise tourism market, is the most important beneficiaries of the global cruise economy (Choi, Kwak, & Kim, 2017), North America has always been the main cruise tourism market and meanwhile, the Caribbean is the world's largest cruise tourists' destination with nearly half of world's cruise passenger capacity (Sun & Feng, 2012).

Therefore, comparing cruise tour differences between China and the United States with the semantic network analysis through big data made it possible for us to capture the main elements that actual consumers are interested in. Moreover, by employing network analysis, we can develop novel and meaningful insights into cruise tour in cruise-developed America and cruise-developing China and then Chinese cruise industry can adopt the advanced merits and experience from America. Due to the volume and unstructured content generated by Google research and news, one important form of the content analysis, semantic network analysis, plays a significant role in big data analysis.

2. Literature Review

2.1. Cruise Tour

A cruise is defined as "to make a trip by sea in a liner for pleasure, usually calling at a number of ports" (Collins English Dictionary, 2010). It is characterized by the ship being similar to a mobile resort, which transports passengers (guests) from place to place (Dowling, 2006). Key cruising areas are the Caribbean, Europe and Alaska through the whole world. Cruise tourism is increasingly recognized as a successful and dynamic subsector of the global tourism industry, with the major cruise lines occupying '⋯the highest ranks of the tourism and leisure sector, as measured by shareholder capital and annual profits' (Weeden, Lester, & Thyne, 2011). The cruise industry is the fast-growing category in the leisure travel market. In 1980, the industry has experienced an average annual passenger growth rate of approximately 7.2% per annum. A record of just about 20 million passengers in the world cruised in 2011, with 11.6 million North American guests. The cruise industry is the largest growing segment of the leisure travel industry. In the years from 1970 to 2004, the number of passengers carried increased by 2200% from 500,000 to close to 11 million North Americans per year (Klein, 2006).

According to the North American Cruise Association (CLIA) and the European Cruise Council (ECC) statistics showed that North America and the European market accounted for 88% of the world cruise tourist market in 2010. Over the past decade, there has been some new trend in the cruise tourism market, with the market sharing of tourists in North America falling from 90% in 2001 to 73% in 2010, while the market sharing of international tourists in other regions has risen from 10% to 27%. Emerging market areas showed significant growth, including the rise of the Asian market (Testa & Sullivan, 2002). Especially in China, the year of 2014 is the Chinese cruise tourism "blowout" development year. Data showed that in 2014, China operated cruise 466 times, an increase of

14.78%, entry and exit cruise tourists was up to 1,723,400 person-times, an increase of 43.36%. Furthermore, Shanghai will become the Home Port as the "Asian cruise center" and for China, the number of cruise passengers will reach 4.5 million in 2020, an average annual growth of 38% compared with 2014, and China will become the most dynamic and largest cruise market in the Asia-Pacific region (Wang, 2015).

2.2. Big Data

There is an abundance of challenges in harnessing the potential of big data nowadays, ranging from the design of processing systems at the lower layer to analysis means at the higher layer, as well as a series of open problems in social problem research (Trnka, 2014). Many marketing researchers believe that social media analytics presents a unique opportunity for businesses to treat the market as a "conversation" between businesses and customers instead of the traditional business-to-customer, one-way marketing (Lusch, Liu, & Chen, 2010).The tremendous growth of data-generating sources has inspired the development of new approaches to understand social/economic phenomena in a variety of disciplines (George, Haas, & Pentland, 2014).Big data is being generated through many sources including Internet traffic (e.g., clickstreams), mobile transactions, user-generated content, and social media as well as purposefully captured content through sensor networks, business transactions, and many other operational domains such as bioinformatics, healthcare, an finance(George et al., 2014).

Big Data is a term for datasets that are so large, complex or diverse that traditional data processing applications are inadequate to collect, manage, and analyze them. Although the term is not clearly defined, a widely accepted definition of big data is the "3Vs" model of Doug Laney, analyst at Gartner: "Big data is high volume, high velocity, and high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery, and process optimization (Hu, Wen, Chua, & Li, 2014)." The first definition of big data comes from Merv Adrian: "Big data exceeds the reach of commonly used hardware environments and software tools to capture, manage, and process it within a tolerable elapsed time for its user population (Trnka, 2014). The classic example of big data analytics is the pioneer study using Google search queries to detect epidemic diseases in the society (Ginsberg et al., 2009).

The rise of digital and mobile communication has made the world become more connected, networked, and traceable and has typical led to the availability of such large scale datasets (Palen, Vieweg, & Hughes, 2009). Big data and big data analysis have been used to describe that data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies (Chen, Chiang, & Storey, 2012).

2.3. Semantic Network Analysis

The formation of social issues through SNS is accelerating from daily life to shopping, games, and business. The real-time transmission and sharing of information through SNS is a key to social issues and interest. Thus, there is a high interest in network analysis in order to extract the various meanings and concepts inherent in the message expressed in text by the development of SNS, and to grasp the relational properties formed between them (Esposito, Ferilli, Basile, & Di Mauro, 2012). The correlation between nodes and nodes composing a network is dynamically changing according to the external environment or time, rather than fixed (Kim, 2012). Therefore, analyzing new values and information through SNS data helps us to understand the changes and trends of various markets and to predict individual behavior and future by deriving useful values from vast amounts of data (Behrendt, Richter, & Trier, 2014).

Network analytics is a nascent research area that has evolved from the earlier citation-based bibliometric analytics to include new computational models for online community and social network analysis (Chen et al., 2012). Since the early 2000s, network science has begun to advance rapidly with contributions from sociologists, mathematicians, and computer scientists. Various social network theories, network metrics, topology, and mathematical models have been developed that help understand network properties and relationships (e.g., centrality, betweenness, cliques, paths, ties, structural holes, structural balance; random network, small-world network, scale-free network) (Arney, 2009; Barabási & Watts, 2002).

In the social sciences, content analysis is the general name for the methodology and techniques to analyze the content of (media) messages (Krippendorff, 2004). An alternative content analysis method is called semantic network analysis or relational content analysis (Van Atteveldt, 2008). Semantic network analysis analyzes the semantic pattern of a message

through the relationship between the frequency of words and words used simultaneously in one sentence without assuming a specific nominal (Shim, Kim, Shon, & Lim, 2011). Shim et al. (2011) conducted semantic network analysis based on interview texts in analyzing media usage behavior of smartphone and social media users. Semantic network analysis is based on the frequency of usage of the main words on the web, the link status between the main words, and the structure of the network (Kim, 2017a). Semantic network analysis can be used as a useful methodology for understanding the flow of web materials (Kim, 2017b).

3. METHODOLOGY

In order to investigate the characteristics of cruise tours in China and U.S.A and compare their differences through the semantic network analysis of the collected big data, this study collected data from Google Search and Google News since January 1st, 2015 to August 15th, 2017. Data collection was conducted by utilizing SCTM which is a user-friendly solution which is capable of simultaneously collecting and processing data on the web by using web crawling and text mining method. As for the data analysis, the refined data was extracted and the main words 60 words (each one for 30 words) were selected for analysis. Finally, the frequency of words was calculated and categorized into a matrix for the semantic network analysis and UCINET 6.0 was used to analyze the network centrality and CONCOR. Analysis results were visualized by Netdraw to illustrate a more intuitionistic visualization of the top frequency words.

4. RESULTS

Cruise Tour, China and Cruise Tour, U.S.A were crawled on the web with Google research and Google news, resulting in 145KB data with 9045 words and 166KB data with 10372 words, respectively. A total of 60 words for data analysis were extracted from the collected data with deleting redundant, repeated or unnecessary words. Table 1 summarizes 30 top keywords for each one (the total of 60 words) which were extracted from the collected data with deleting redundant or unnecessary words.

After filtering the data we have collected, this study concluded a total of 60 words (for each one is 30 words) to analyze. The results showed by Table 1 Top keywords fre-

quency of cruise tour were as follows. Firstly, for China cruise tour, the most frequent keywords mainly focused on the destinations, both national and international. They are all the popular and famous ports or destinations for cruise tour, the web visibility of South Korea, Beijing, Shanghai, Hong Kong and other port cities is high at the same time, tours and cruise demonstrated the highest web visibility rate at 12.97% and 12.68% and frequency are 958 times and 684 times. Secondly, for U.S.A cruise tour, the word "offer "showed the highest web visibility rate and the frequency was up to 6605 times. The word "offer " implied that cruisers focused on what cruise could provide to them and searched for the conditions, services and special experiences onboard. In addition, the expenditure of cruise related contents such as coupon, price, fee, and discount were presented in details. Meanwhile, the frequency of the cruise experience related words such as condition, service, stateroom, etc. are correspondingly high. Consequently, according to the web visibility rate and words frequency, it demonstrated that for the cruise-developed U.S.A, cruise consumers pay more attention to the experiences and leisure activities, services onboard and the expenditure of the cruise based on their mature knowledge and understanding for cruise tour. Nonetheless, for cruise-developing China, cruise consumers mainly focus on the destinations and departure issues since they possess limited even no perceptions for cruise tour without previous taking experience. Interestingly, Caribbean as the world's largest cruise tourists' destination with nearly half of world's cruise passenger capacity has exposed on the list of top keywords both in China and U.S.A. The results of visualizing the network reflecting the frequency of top keywords are shown in Fig. 1.

Freeman's degree centrality is an index that measures how much one node has a connection with other nodes on the network and eigenvector centrality or so-called chain centrality, is a useful index to find the most influential central node in the network (Kim, 2017a). Table 2 and Table 3 compares frequency, freeman's degree centrality and eigenvector of the high-frequency words of cruise tour in China and U.S.A, synchronously and respectively. In Table 2, for cruise tour in China, as a result of confirming the degree centrality and eigenvector centrality of key words connection to "cruise tour, china", it showed that they shared a similar pattern to the main words frequency results. Nonetheless, the word "sea" with high frequency is of low ranking of centrality which means that "sea" possesses limited connection with other
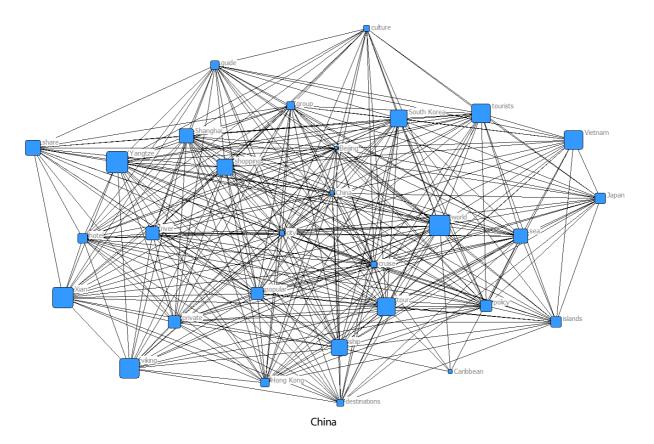
Table 1. Top keywords frequency of cruise tour (China and U.S.A)

	Cruise Tour, China				Cruise Tour, U.S.A		
Rank	Word	Frequency	%	Rank	Word	Frequency	%
1	Tour	958	12.97%	1	Offer	6605	18.65%
2	China	937	12.68%	2	Cruise	5240	14.80%
3	Cruise	684	9.26%	3	Shore	2568	7.25%
4	Beijing	381	5.16%	4	Booking	1793	5.06%
5	Sea	374	5.06%	5	New York	1544	4.36%
6	Yangtze	363	4.91%	6	Excursion	2085	5.89%
7	Ship	327	4.43%	7	Online	1180	3.33%
8	SouthKorea	321	4.34%	8	Booking	1163	3.28%
9	River	316	4.28%	9	Guests	1049	2.96%
10	Hotel	311	4.21%	10	Ship	981	2.77%
11	Private	293	3.97%	11	Qualify	954	2.69%
12	Xian	222	3.00%	12	Coupon	948	2.68%
13	Shanghai	215	2.91%	13	Stateroom	945	2.67%
14	Hong Kong	210	2.84%	14	Condition	842	2.38%
15	City	185	2.50%	15	Sailing	742	2.10%
16	Islands	184	2.49%	16	Onboard	595	1.68%
17	Group	143	1.94%	17	Carnival	556	1.57%
18	Share	118	1.60%	18	USA	550	1.55%
19	World	109	1.48%	19	Service	473	1.34%
20	Guide	104	1.41%	20	Reservation	433	1.22%
21	Shopping	82	1.11%	21	Discount	679	1.92%
22	Policy	80	1.08%	22	Caribbean	417	1.18%
23	Destinations	78	1.06%	23	Activity	719	2.03%
24	Caribbean	74	1.00%	24	Fee	373	1.05%
25	Culture	67	0.91%	25	Price	358	1.01%
26	Tourists	63	0.85%	26	Canada	353	1.00%
27	Popular	52	0.70%	27	Norwegian	353	1.00%
28	Japan	51	0.69%	28	Participate	329	0.93%
29	Vietnam	44	0.60%	29	Port	308	0.87%
30	Viking	42	0.57%	30	Agent	282	0.80%

nodes.

While in Table 3, for cruise tour in U.S.A, the words "booking" and "cruise" showed high frequency, nonetheless, they were relatively low in connection with centrality and it also does not possess a meaningful connection with others. Although the degree of link accuracy is basically based on the frequency of the words, it can be different from the result of frequency analysis because it is structured around connectivity based on completeness in one line (Guarino, Masolo, & Vetere, 1999). Additionally, words like "discount" and "coupon" which are of low frequency while the centrality is high, which implies that these two words are the significantly
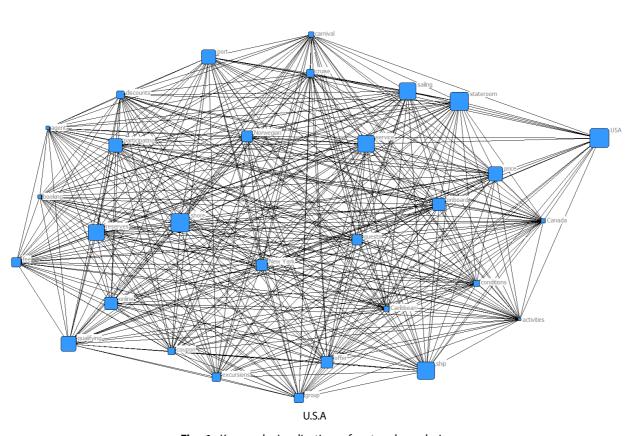
China

U.S.A

Fig. 1. Keyword visualization of network analysis.

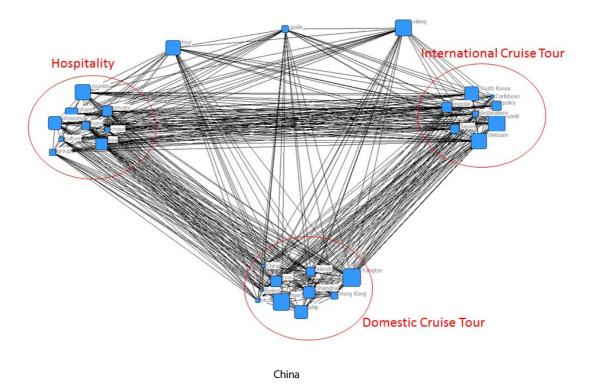**Table 2.** Comparison of keywords frequency and centrality for cruise tour in China

	Frequency		Freeman's degree centrality		Eigenvector centrality	
	Freq.	Rank	Coefficient	Rank	Coefficient	Rank
Tour	958	1	0.114	2	62.186	1
China	937	2	0.123	1	57.51	2
Cruise	684	3	0.058	8	28.014	9
Beijing	381	4	0.099	3	56.367	3
Sea	374	5	0.017	16	8.206	16
Yangtze	363	6	0.073	4	34.736	6
Ship	327	7	0.064	6	29.593	8
SouthKorea	321	8	0.035	12	16.601	13
River	316	9	0.044	10	21.682	11
Hotel	311	10	0.069	5	43.824	4
Private	293	11	0.059	7	40.334	5
Xian	222	12	0.039	11	24.738	10
Shanghai	215	13	0.022	14	11.93	15
Hong Kong	210	14	0.008	21	4.00	21
City	185	15	0.047	9	29.723	7
Islands	184	16	0.011	19	4.965	19
Group	143	17	0.032	13	20.206	12
Share	118	18	0.001	29	0.728	29
World	109	19	0.014	17	8.068	17
Guide	104	20	0.012	18	7.829	18
Shopping	82	21	0.008	20	4.907	20
Policy	80	22	0.002	26	0.794	26
Destinations	78	23	0.002	28	0.785	27
Caribbean	74	24	0.002	27	0.766	28
Culture	67	25	0.001	30	0.382	30
Tourists	63	26	0.007	22	2.794	23
Popular	52	27	0.02	15	12.414	14
Japan	51	28	0.006	24	2.266	24
Vietnam	44	29	0.005	25	2.109	25
Viking	42	30	0.006	23	3.083	22

Table 3. Comparison of keywords frequency and centrality for cruise tour in U.S.A

	Frequency		Freeman's degree centrality		Eigenvector centrality	
	Freq.	Rank	Coefficient	Rank	Coefficient	Rank
Offer	6605	1	0.174	1	78.062	1
Cruise	5240	2	0.019	17	11.464	17
Shore	2568	3	0.085	3	49.332	3
Excursion	2085	4	0.014	20	8.714	20
Booking	1793	5	0.012	23	7.514	23
New York	1544	6	0.042	6	25.568	6
Online	1180	7	0.041	7	24.624	7
Booking	1163	8	0.067	4	40.184	4
Guests	1049	9	0.038	8	23.403	8
Ship	981	10	0.027	14	16.170	14
Qualify	954	11	0.034	10	20.338	10
Coupon	948	12	0.117	2	63.649	2
Stateroom	945	13	0.034	9	20.612	9
Condition	842	14	0.034	11	20.300	11
Sailing	742	15	0.025	15	14.967	15
Activity	719	23	0.013	22	7.651	22
Discount	679	21	0.045	5	27.003	5
Onboard	595	16	0.020	16	11.790	16
Carnival	556	17	0.029	12	17.786	12
USA	550	18	0.000	30	0.001	30
Service	473	19	0.015	19	9.001	19
Reservation	433	20	0.016	18	9.839	18
Caribbean	417	22	0.013	21	8.236	21
Fee	373	24	0.029	13	17.245	13
Price	358	25	0.012	26	7.050	26
Norwegian	353	27	0.006	29	3.412	29
Canada	353	26	0.009	27	5.294	27
Participate	329	28	0.012	25	7.106	25
Port	308	29	0.007	28	4.278	28
Agent	282	30	0.012	24	7.404	24

influential nodes among all nodes and the share the relatively high connections with other nodes in the network.

As a result of checking the degree of centrality of main words related to "cruise tour, usa", it was confirmed that words such as cruise, shore, ship were seized of strong connection with the keyword.
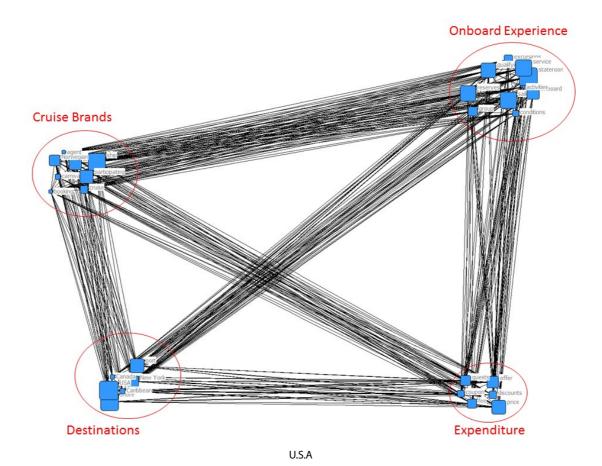
China

U.S.A

Fig. 2. Visualization with CONCOR analysis.

CONCOR (convergence of iterated correlation) is a method of repeatedly performing correlation analysis to find an appropriate level of similarity groups by performing correlation analysis repeatedly. In this study, we conducted CONCOR analysis based on semantic network analysis results and it was performed in Fig. 2.

As a result, for China, three clusters were created and named "domestic cruise tour", "international cruise tour"and the last segment was a cluster of words related to the hospitality which included shopping, hotel, policy, group and Viking which is relevant to the cruise brand. The cluster of domestic cruise tour consisted of Shanghai, Beijing, Xian, Yangtze, Hong Kong, etc. and they are the well-known or popular port cities, destinations or main river for the Chinese domestic cruise tour. The cluster of international cruise tour mainly included South Korea, Vietnam, Japan, and Caribbean. CONCOR analysis for U.S.A, the generated data was divided into four clusters with an appropriate level of similarity. Cruise destinations, cruise expenditure, onboard experience and cruise brand, these four segments were formed. The 'onboard experience' community with the largest number of semantic networks was composed of words such as conditions, service, activities, stateroom and it suggested that onboard offering services and conditions are strongly relevant to the cruisers experience. The segment "expenditure" consisted of price, fee, discount, coupon, showing the expenditure of the cruise concerning the cruisers cruise choice and perception. Furthermore, several words of cruise brand were categorized such as Carnival, Norwegian.

5. CONCLUSIONS & IMPLICATIONS

This study analyzed a total of 60 words related to cruise tour in China and U.S.A from January 1st, 2015 to August 15th, 2017, using the data collecting and processing program SCTM (Smart crawling & Text mining). The collected data was analyzed with Freeman's degree centrality, eigenvector centrality by using UCINET 6.0 packaged with Netdraw.

For cruise tour both in China and U.S.A, the web visibility of Shanghai, Beijing, South Korea, and New York, Caribbean was high, which represented major port cities, places or destinations for cruise tour. This result showed that to some extent, no matter how developed the cruise tour is, cruisers are all concerning where they will go and where their cruising destination is. And for U.S.A cruise tour specially, it formed a group of words such as fee, price, discount, coupon and

all of these words were relevant to the cruise expenditure and a group of cruise destination, Caribbean, Canada, and all of them were oversea destinations for cruise tour. It suggested that since the cruise tour in U.S.A is well-developed and American cruisers take cruise tour more frequently, they more concerns about their expenditure during the cruise tour, at the same time, they also concerned about the onboard conditions, services and brand choices since they possessed perceptions on cruise tour based on their previous cruise taking experience. As for the segment of China cruise tour, it conceived two clusters for destinations, domestic and exotic, and mainly focused on the domestic and Asian area, and this illustrated that the cruise industry in China is just emerging and Chinese cruisers cares more about their destinations and are not willing to take long journey voyage, hence, they searched the information about the short-distance destinations on the Internet.

Concluding the above mentioned materials, this study indicated that characteristics of cruise tour in China and U.S.A. For the cruise-developing China, limited by finite perceptions and previous experience for cruise tour, Chinese cruisers mostly concerns about the destinations of the cruise tour, however, for the cruise-developed U.S.A, since American cruisers harnessed abundant previous experience and in-depth understanding for cruise tour, they mostly focus on the service and conditions onboard and the brand of the cruise ships. In addition, American cruisers pursed a beneficial and economical way to take their cruise tour so that they surfed on the Internet to figure out the price, fee, discount, etc. of the cruise tour. Most importantly, the characteristics of cruise tour in U.S.A provided significant and social network-oriented suggestions for Chinese cruise tourism.

The present study provided both theoretical and managerial implications for academicians and practitioners. From a theoretical point of view, this study provided certain guidance on how to utilize big data generated from tremendous online sources in cruise tourism, furthermore, it set an example for utilizing online information to interpret the current market direction with timeliness. As for the managerial implications, the current circumstance of American cruise market will be the facing-situation for future Chinese cruise market, hence, it is meaningful and significant for Chinese cruise market to improve their onboard conditions to enhance cruisers' cruising experience and formulate effective pricing programme to stimulate and maintain the development for Chinese cruise

tourism. And for the American cruise tourism, it is supposed to set up the market strategy focusing on the pricing programme and the improvement of the onboard experience. To some extent, it is better to expand its market to international destinations since American cruisers prefer the exotic shipping ports and cities.

This study has several limitations and the findings should be interpreted with caution. While big data analysis has been utilized as a new paradigm in many disciplines, we have seen very few applications in the field of hospitality that fully explore its capabilities. Eighty percent of the world's data is unstructured, and most businesses do not even attempt to use this data to their advantages. The trend of big data is generating new opportunities and new challenges for business across industries. Problems of big data is mainly concerned with non-user data (not inputted by the user), but loaded by various automated means (e.g. data from security cameras).

As for the future research, we will overcome these obstacles and may consider applying methods of combination of big data analysis with questionnaire survey to validate the cruise tour differences between China and U.S.A with being quantitative and qualitative so as to develop a more comprehensive knowledge about cruise tour.

REFERENCES

Arney, C. (2009). Linked: How everything is connected to everything else and what it means for business, science, and everyday life. *Mathematics and Computer Education, 43*(3), 271.

Barabási, A. L., & Watts, D. J. (2002). *Six degrees: the Science of a connected age.*

Behrendt, S., Richter, A. & Trier, M. (2014). Mixed methods analysis of enterprise social networks. *Computer Networks, 75*(24), 560-577.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly, 36*(4).

Chin, C. B. (2008). *Cruising in the global economy: Profits, pleasure and work at sea.* Ashgate Publishing, Ltd..

Choi, H. Y., Kwak, G. H., & Kim, H. S. (2007). A positioning study of national food: In perspective of Korean, American, Chinese food tourists. *Culinary Science & Hospitality Research, 23*(2), 86-94.

Collins English Dictionary (2010). Collins english dictionary.

Bishop Briggs, Glasgow: HarperCollins.

Dowling, R. K. (2006). Looking ahead: The future of cruising. *Cruise ship tourism*, 414.

Esposito, F., Ferilli, S., Basile, T. M., & Di Mauro, N. (2012). Social networks and statistical relational learning: A survey. *International Journal of Social Network Mining, 1*(2), 185-208.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal, 57*(2), 321-326.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457* (7232), 1012-1014.

Guarino, N., Masolo, C., & Vetere, G. (1999). Ontoseek: Content-based access to the web. *IEEE Intelligent Systems and their Applications, 14*(3), 70-80.

Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE access, 2*, 652-687.

Kim, H. S. (2017a). A semantic network analysis of big data regarding the food exhibition on convention center. *Culinary Science & Hospitality Research, 23*(3), 257-270.

Kim, H. S. (2017b). An exploratory study on the semantic network analysis of food tourism through the big data. *Culinary Science & Hospitality Research, 23*(4), 22-23.

Kim, H. Y., (2012). Analysis of dynamic change of the novel land and character network. *The Journal of the Korea Contents Association, 12* (11), 519-526.

Klein, R. A. (2006). 24 Turning water into money: The economics of the cruise industry. *Cruise ship tourism*, 261.

Klein, R. A., (2011). Responsible cruise tourism: Issues of cruise tourism and sustainability. *Journal of Hospitality and Tourism Management, 18*(1), 107-116.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology.* Sage.

Ministry of Maritime Affairs & Fisheries (2014). *The second marine tourism promotion basic plan.* Sejong: Ministry of Maritime Affairs and Fisheries.

Ministry of Maritime Affairs & Fisheries (2015). *Cruise industry activation measures.* Sejong: Ministry of Maritime Affairs and Fisheries.

Lusch, R., Liu, Y., & Chen, Y. (2010). The phase transition of markets and organizations: the new intelligence and entrepreneurial frontier.

Palen, L., Vieweg, S., Liu, S. B., & Hughes, A. L. (2009). Crisis in a networked world: Features of computer-mediated communication in the April 16, 2007, Virginia Tech event. *Social Science Computer Review*, *27*(4), 467-480.

Schutt, R., & O'Neil, C. (2013). *Doing data science: Straight talk from the frontline*. " O'Reilly Media, Inc.

Shim, H. J., Kim, Y. C., Shon, H. Y., & Lim, J. Y. (2011). An exploratory usage pattern research of smartphone and social media users through semantic network analysis: Gender and age differences in perception and evaluation of usage pattern. *Korean Journal of Broadcasting*, *25*(4), 28-138.

Sun, X. D., & Feng, X. G. (2012). China cruise tourism industry: Research status and prospects. *Journal of Tourism*, (2), 101-112.

Testa, M. R., & Sullivan, K. (2002). Customer satisfaction, quality in cruise industry. *Hospitality Review*, *20*(2), 1.

Trnka, A. (2014). Big data analysis. *European Journal of Science and Theology*, *10*(1), 143-148.

Van Atteveldt, W. H. (2008). Semantic network analysis: Techniques for extracting, representing, and querying media content.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), 77-84.

Wang, Y. (2015). The rise of the 'shareholding state': financialization of economic management in China. *Socio-Economic Review*, *13*(3), 603-625.

Weeden, C., Lester, J. A., & Thyne, M. (2011). Cruise tourism: Emerging issues and implications for a maturing industry. *Journal of Hospitality and Tourism Management*, *18*(1), 26-29.