This is the Accepted Manuscript Version

Mingming Cheng*
Lecturer
Department of Tourism
University of Otago, New Zealand
Email: mingming.cheng@otago.ac.nz
+64 3 479 3486

Deborah Edwards
Associate Professor and Research Director
UTS Business School
University of Technology Sydney, Australia
Deborah.Edwards-1@uts.edu.au
+61 2 9514 5424

*Corresponding author

A comparative automated content analysis approach on the review of the sharing economy discourse in tourism and hospitality

Abstract

Using the sharing economy (SE) as the context, this article provides a coherent and nuanced methodological understanding of automated content analysis (ACA) in tourism and hospitality (TH) field. By adopting a comparative automated content analysis approach, the paper compares the current TH Western academic literature of the SE with news media discourse in tourism and hospitality from the period 2011-2016 (August) (inclusive). The emerging issues from the news media discourse, such as mobility, SE companies and the role of government are absent in current tourism academic research. Findings reveal that ACA can facilitate a more systematic comparison between different sources of data. This paper offers a starting point for tourism scholars to methodologically engage with ACA that can draw useful insights on a particular context.

Keywords: sharing economy; automated content analysis; Leximancer; comparative analysis; news articles; sample size;

Introduction

Driven by growing shared values of the public and increasing technological advancement of Internet platforms, the sharing economy (SE) has enjoyed remarkable growth in the last five years with networked actors who are simultaneously coordinating, distributing and utilizing under-used resources either for a fee or for free (OECD, 2016; PwC, 2015). The tourism and hospitality (TH) sector is at the forefront of this disruptive business model, exemplified by the explosive growth of Airbnb a peer-to-peer accommodation sharing platform, and other new business models in transport, restaurants, and tour guiding (OECD, 2016). In TH, the SE is believed to have the potential to create better monetary value for tourists, encourage sustainable tourism consumption and facilitate authentic host-guest encounters (Forno & Garibaldi, 2015; Lyons & Wearing, 2015; OECD, 2016). However, there are on-going concerns about the rapid growth of the SE, particularly in terms of competitive fairness between SE companies and traditional tourism service providers (e.g. international hotel chains), bypass of the government's TH industry standards, tendency towards SE's companies' monopoly and increase of TH's labor casualization (Juul, 2015; OECD, 2016; Queensland Tourism Industry Council, 2014). Despite controversial views towards SE, the general public, many investors and TH researchers hold a positive vision of the transformative effect of SE on the current tourism system at various levels, including individual tourists, business owners, communities, tour operators and tourism destinations (Cheng, 2016b; Guttentag, 2015; Sigala, 2015).

The SE has only recently become the subject of scientific research in TH with two recent special issue calls from leading tourism (Journal of Travel and Tourism Marketing) and hospitality journals (International Journal of Contemporary Hospitality Management). With limited academic literature to offer scientific evidence and an increasing number of industry articles presenting various and contradictory claims, the current discourse confuses various stakeholders' understanding of the future development of SE (Cheng, 2016b). This confusion also means that various stakeholders (e.g. local communities, traditional service providers, regulatory authorities) are uncertain about their strategies and actions with the widespread feeling that "the problems are too big for individuals to deal with" (Low & Eagle, 2016, p. 643). Many might be reluctant to respond to the future growth of the SE or they might act in an irrational and ineffective way to just protect their own "turf" (Cheng, 2016a). This increased skepticism may undermine society's coping strategies to address the various challenges and harness the potential of the SE. As such, an opportunity arises to understand the status-quo of current researchers' knowledge in TH and what is actually happening in the real world. Specifically, a comparative study between the current literature and news media discourse could offer insights into the SE, from which future research directions can be identified to inform TH researchers' and practitioners' contributions to the theoretical and practical advancement of the SE. However, a simple search of the topic "sharing economy" on google news can return hundreds of articles in less than a few seconds. It raises the question as to how the general public and even researchers become effectively acquainted with the current knowledge of the SE - a phenomenon growing at a rapid rate. Thus, an alternative approach that can efficiently and effectively help classify massive amounts of data into categories would help readers gain clarity in understanding the relevant issues concerning the SE and formulate an informed decision. Therefore, this article addresses this need by undertaking a comparative automated content analysis approach to uncover key themes and concepts of concern to these interested groups.

Methods on text analysis

The academic community has long been engaged with the analysis of text. Traditional text analysis is manually based, where the researchers examine the text and make effective references to produce relevant insights from text. A traditional approach relies largely on manual text/content analysis. That is, researchers manually examine various sources (e.g. academic articles and news articles) to identify ideas and themes. As such, scholars tend to confine the sample size of the text, which results in a series of related limitations (e.g. under-sampling, biased estimates) (Nunez‑Mir, Iannone, Pijanowski, Kong, & Fei, 2016), time-consuming, and reduced efficiency and effectiveness. For example, with an increasing number of journal articles published in TH journals, many review papers are limited to analyzing the relevant articles in the top three tourism journals, which can miss important contributions from other TH journals, and lead to incorrect conclusions. The other limitation related to manual text analysis is subject to a certain degree of priori (e.g pre-conception bias) (Smith & Humphreys, 2006). As such, researchers might end up in inadvertently using exemplary studies when conducting reviews or unconsciously paying attention to what interests them while overlooking other important insights (Nunez‑Mir et al., 2016). Hence, there is need for academics to embrace new methods that could help tackle these existing issues and conduct text analysis in a more efficient, objective and robust manner. Indeed there are an increasing number of scholars engaged in an alternative approach to analyze text utilizing computer assisted programs.

A number of researchers propose automated content analysis (ACA) – a technique that involves text mining algorithms from computer science to recognize patterns and make probabilistic predictions of data (Evans, McIntosh, Lin, & Cates, 2007; Nunez‑Mir et al., 2016). ACA as an alternative approach by means of computer assisted programs was introduced in the last decade, although earlier versions can be dated back to 1990s in the use of Latent Semantic Indexing, a method that retrieves information utilizing linear algebra techniques (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). ACA deploys a series of statistical algorithms to uncover hidden patterns and thematic compositions in a body of text. In other ways, ACA with its statistical algorithms is able to manage semantic and linguistic complexity of the text to produce concepts and themes referenced to the text. The most influential algorithm model is Laten Dirichlet Allocation based on a three-level hierarchical Bayesian theory (Blei, Ng, & Jordan, 2003). Fundamentally, ACA is conducted through three stages "identification, definition and text classification". A detailed explanation of ACA is in the work of Nunez‑Mir et al. (2016). Several studies have confirmed the reliability and validity of ACA. Leximancer, for example an ACA software has become widely accepted across various disciplines, with recent application in analyzing tourism academic publications (Cheng, 2016b; Cheng, Edwards, Darcy, & Redfern, 2016; Jin & Wang, 2016), Australian news articles to understand the representations of homelessness (MacKinnon, 2015), and social media data (Tseng, Wu, Morrison, Zhang, & Chen, 2015). This is because as an ACA tool - Leximancer has been found to generate a more objective and text-driven review of documents with reproducible and reliable concept extractions and thematic clustering (Biesenthal & Wilden, 2014; Randhawa, Wilden, & Hohberger, 2016; Smith & Humphreys, 2006), even with large chunks of text (Angus, Rintel, & Wiles, 2013; Edwards, Cheng, Wong, Zhang, & Wu, 2017). Newman, Noh, Talley, Karimi, and Baldwin (2010) confirm that the ACA model even

performed better for semantic coherence than human inter-rater correlation. Nunez‑Mir et al. (2016) in comparing a manual review with an ACA review, of the effects of land use on exotic plant invasion, found that ACA is similar to manual analysis but is able to identify trends that might be over-looked.

While the effectiveness and efficiency of ACA have been confirmed in the extant literature, ACA's use in TH is limited. More importantly, on a closer review of the extant literature using ACA, we realized that many researchers in TH have taken ACA for granted when applying the method to their research, with sometimes contradictory suggestions and they focus on a single source of data (e.g. Jin & Wang, 2016). As such, the aim of this study is to provide a coherent and nuanced methodological understanding of automated content analysis (ACA) in TH field. In particular, a comparison between news articles and academic journal articles utilizing ACA could facilitate a clear assessment of the generalization of certain findings and elucidate the relationships between social phenomena (Esser & Hanitzsch, 2013).

Research Design

Data collection

Two sources of data were used for this analysis: news articles and academic journal articles. As the SE is an alternative to conventional practice, its wide implications for the TH system are not yet fully understood. Thus, the general public's knowledge at the current stage is predominantly distributed through public communication (Schmidt, Ivanova, & Schäfer, 2013). As a dominant agent of the interpretative system of our modern society, news media serves as a key platform to raise awareness and disseminate information (Schmidt et al., 2013). As such, news articles reflect wider trends of "what is going on" (Bednarek, 2006) and convey a strong message to the public about "hot" topics in a field and produce "an agenda setting" effect (McCombs, 2013). In particular, news articles signal the relevance of an issue to the general public and might potentially influence the priority given to it by regulatory authorities (Schmidt et al., 2013; Schweinsberg, Darcy, & Cheng, 2017).

Academic journal articles are used as they are the primary outlets to advance the scholarly development of knowledge in a specific field (Pfeffer, 2007). In particular major TH journals were the sources for selecting the journal articles as these journals signify the current discipline-relevant theoretical knowledge base (Biesenthal & Wilden, 2014). The authors used Scopus to start the initial search within the TH journals by using the keywords "sharing economy" and "collaborative economy". Relevant references concerning the SE were also traced. The initial search ended with 40 journal articles prior to August 2015. Then the authors examined the articles to assess whether the SE was a focus within the articles. Three articles published by Cheng (2016a) and Heo (2016) (review articles) as well as Cheng (2016b) (media analysis) were excluded, as these would potentially crowd the analysis. News articles were retrieved from ProQuest Newsstand database, which is claimed by ProQuest as the world's most comprehensive collection of news content (ProQuest, 2014). ProQuest Newsstand contains indexed and fully searchable electronic copies of articles from over 1,500 state, regional, national, and international English newspapers (ProQuest, 2014). Duplicated news content was excluded. Key information was recorded, including the news source (newspaper), date of publication and full content of the news articles.

As a result of this selection, the authors retrieved the full content of 18 journal articles and 547 news articles in English from 01/01/2011 to 31/08/2016. The top five newspaper

outlets were The Mercury, The Times, The Wall Street Journal, The Australian Financial Review, and The New York Times. A number of journals contained content relevant articles of which the top two were Current Issues in Tourism (7) and Annals of Tourism Research (3). These journal articles were selected using the Scopus database as journal articles indexed in Scopus are peer-reviewed and have a publicly available publication ethics statement, deemed to reach a suitable standard for academic publication and readership (Scopus, 2017).

Data analysis
ACA was performed using Leximancer software (edition 4.0). As high level natural language processing software, Leximancer starts with no pre-conceptions and the analysis emerges from the data (Smith & Humphreys, 2006). Underpinned by Bayesian theory, it quantifies the texts by identifying the frequency of concepts and their relationships through an emergent iterative and unsupervised process (Smith & Humphreys, 2006). As such, "fragmented pieces of evidence" in documents "can be used to predict what is actually happening in a system" (Watson, Smith, & Watter, 2005, p. 1233). Leximancer produces a heat map that visually demonstrates the end results. Themes are color-coded, where brightness presents the theme's prominence (Angus et al., 2013). Concepts being mapped closely to each other means a strong semantic relationship (Campbell, Pitt, Parent, & Berthon, 2011; Smith & Humphreys, 2006).

The process of analysis involved three stages. The first stage is the data cleaning process of removing 'noise'. This includes author information and affiliation, and acknowledgement and references in journal articles (Cheng et al., 2016). An additional algorithm was developed in R to clean the news articles and prepare them in a format that aligns with the system requirements of Leximancer. In the second stage, while the concept map generated by Leximancer presents clear thematic clustering, it is important for the researcher to map complex themes to even higher theoretical/thematic constructs to assess their usefulness in terms of the research objectives (Povey et al., 2013) (e.g. broad area of foci in this study). This higher order theorizing process was undertaken based on 1) co-occurrence frequency, 2) inspection of the statistical results, and 3) a conceptual map generated during iterations.

In the third stage, to reveal the differences in concepts from these two different sources, a comparative approach was employed. Two issues emerged. This first one is comparative equivalence (Esser & Hanitzsch, 2013). As news and academic journal articles are written for different audiences (i.e. the general public vs academics), the concepts generated by Leximancer are context-dependent with variations in their definitions and scopes from both sources. Hence, the authors examined each concept and linked it to the original text to understand its context and re-word the concepts if necessary. The second issue is the different sizes of the samples. Differences in sample size can affect the final comparison of the data. In ACA, concepts are automatically selected to characterize the entire set of data. That is to say, if one data source contributes much more data, then this particular source will also dominate the automatic selection of concepts. Similarly, when the thesaurus learning discovers the characteristic vocabulary for a concept, then if that concept appears in more than one data source, the larger source will contribute more to its vocabulary. This difference in sample size can make comparisons difficult. There are two solutions to this problem. One is to create a separate project for each data source, and compare the results, as ACA tools (Leximancer) will avoid mixing two different mental models. The second solution is to increase the number of automatically discovered concepts and as such there are enough concepts discovered to characterize the minority data components. It was decided to compare both approaches. The results are discussed in the following section.

Findings

SE in TH literature

The analysis of the TH journal articles revealed two major areas of foci: (1) the nature of SE and its relationship to TH and (2) SE's accommodation sector (Figure 1).
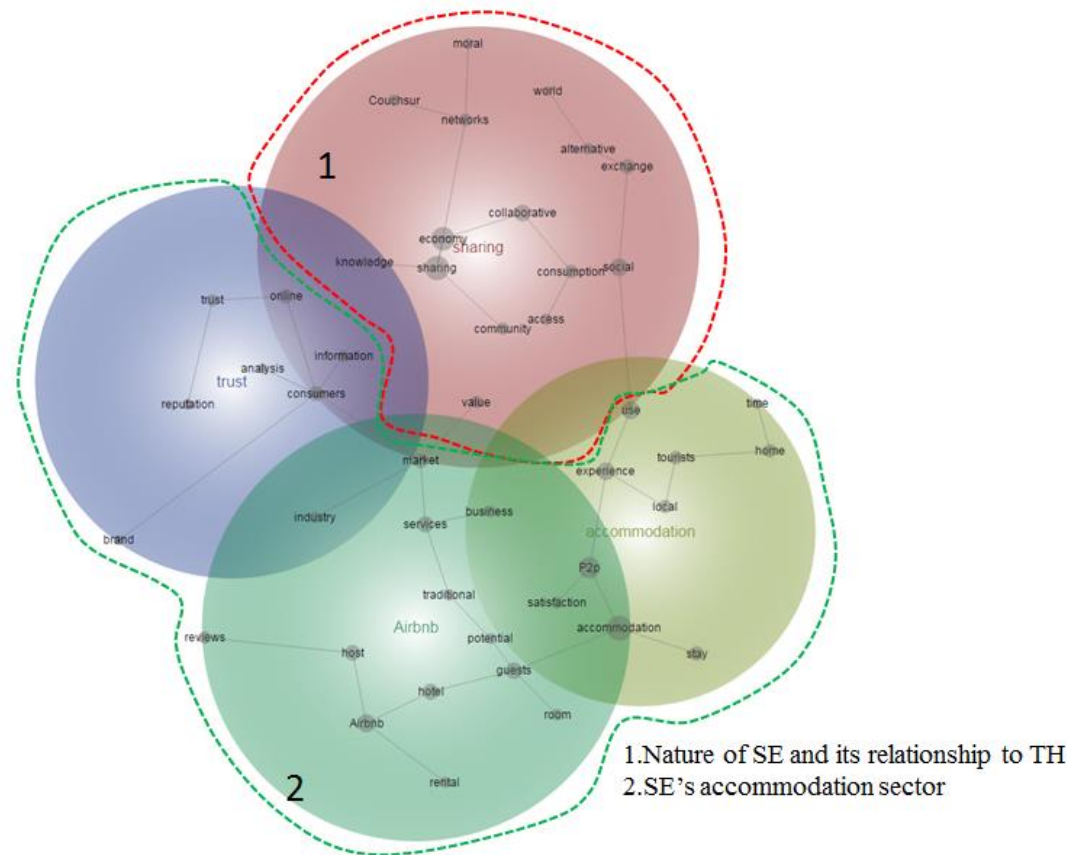

Figure 1: Conceptual map of SE literature in TH

In foci 1 the theme "sharing" represents a definitional discussion of the SE, its impacts on value, role in community engagement, access to under-utilized tourism resources and people's initial understanding of the SE. It highlights the social changes caused by the SE (Forno & Garibaldi, 2015; Guttentag, 2015; Tussyadiah & Pesonen, 2015) through a network as it reflects the importance of networks in connecting users in the online and offline space characterized by the moral economy, such as couchsurfing (Germann Molz, 2013).

In foci 2 there are three interrelated themes. Airbnb is the largest theme signifying an academic focus of the impacts of Airbnb on various aspects of the TH industry that challenge *traditional* tourism service providers and destinations. The connection concept *accommodation* with "Airbnb" indicates the significant influence of Airbnb on tourist experience including satisfaction, length of stay and guest-host relationships (Guttentag, 2015). The third theme is trust (in purple) reflecting the importance of SE business's reputation as a currency for consumers (Sigala, 2015) and highlights ways to refine both the SE and traditional tourism service providers' brands (Richard & Cleveland, 2016). It is clear that the accommodation

sector, particularly Airbnb, is treated as a representative of SE businesses, in the TH academic discourse.

The SE concerning TH in the news articles

Figure 2 indicates that there are three broad areas within the news articles including (1) SE and its impacts; (2) growth of the SE; (3) SE companies.

Figure 2: Conceptual map of news articles

The theme "Airbnb" is the dominant theme with various concepts reflecting the impact Airbnb has on the SE, including *host-guest* relationships, *payment* concerning tax, people's *use* of their *home*s, impacts on the *properties*, *locals, short-term rental market, hotel, owners* and visitor's length of *stay* as well as competition with the traditional accommodation sector. The theme *city* suggests that these impacts occur mainly in cities. Also, the closely connected concepts –*New York* and *Paris* suggest the ongoing debates around the SE in these two cities.

8

The impacts of the SE on cities are on short-term housing, property market, rentals, residents as well as investments concerning the rental market as a direct consequence of tourists' holidays. The theme government highlights the important role of the government in this process to deal with the impacts on community and owners of SE providers in cities. Because of the impacts created by the SE products, many stakeholders even voiced to ban or impose restrictions on Airbnb and Uber in cities. The theme experience indicates the SE's impacts on people's mobility and holiday experience. It highlights that the SE products are treated as an alternative means to facilitate tourist experiences.

The second focus is on the growth of the SE highlighting its disruptive role in the tourism economic system providing both challenges and opportunities that are driven by technological advancement. The results show that the SE is a billion dollar field. The concept growth also indicates the explosive growth of the SE that leads to global social changes, particularly with the SE's contributions to sustainable consumption practices and innovation in TH. The theme China highlights the rapid development of the Chinese SE models.

Included within the broad area of the SE company, are issues related to the growth of SE *companies*, different services offered by SE companies, SE companies' impacts on the work and time arrangements of people, the management of these companies, the business potential of these companies. Many of these companies operate their business through the format of an App. The concept *Uber* points out the service model in some SE companies and highlights the disruptive nature of Uber on the *car* industry as well as *drivers*. The concept *work* also describes new employment issues in TH under the SE. The growth of many SE companies highlights their co-existence with traditional TH service providers including restaurants, transport and tour guide services by creating challenges to existing businesses.

Comparison between journal and news articles

In Solution 1, we created a separate project for each data source (Journal and News Articles), and compare the results to avoid mixing two different mental models. Figures 3.1 and 3.2 present the concepts that appear in both sources with different relevance scores. Relevance is defined as "the percentage frequency of text segments which are coded with that concept, relative to the frequency of the most frequent concept in the list" (Leximancer, 2011). It indicates the relative strength of a concept's frequency of occurrence. However, with more than 100 concepts to compare, it is difficult to visualize them in one graph and as such it was necessary to present the analysis in two separate graphs. Figure 3.1 present concepts that appear in both sources (line indicates the same concepts), and Figure 3.2 presents concepts that are unique to each source. In Solution 2, we increased the size in the number of automatically discovered concepts to characterize the minority data components i.e. from 75 to 100. Then the concepts were compared using prominence scores (Figure 4).

After comparing the outcomes of both solutions, it was found that while most of the major concepts did appear in both solutions, Solution 1 worked better by avoiding the inflation of a certain concept if they exist in both sources and generating more emerging concepts for each individual source. As such, the results of Solution 1 were used. Figure 3.1 indicates that news articles tend to a wider discussion of the sharing economy around Airbnb and its market, host relationship with locals, and reputation with locals, while the literature tends to focus on the SE's impacts on accommodation, guests/tourists, people's online behavior and hospitality and services in general. Figure 3.2 shows that many concepts have not been covered by the current literature indicating that the academic research on this area still lags behind news

discussion. This includes other business models (e.g. company, app), property markets, the role of government, payment transactions and tax issues.
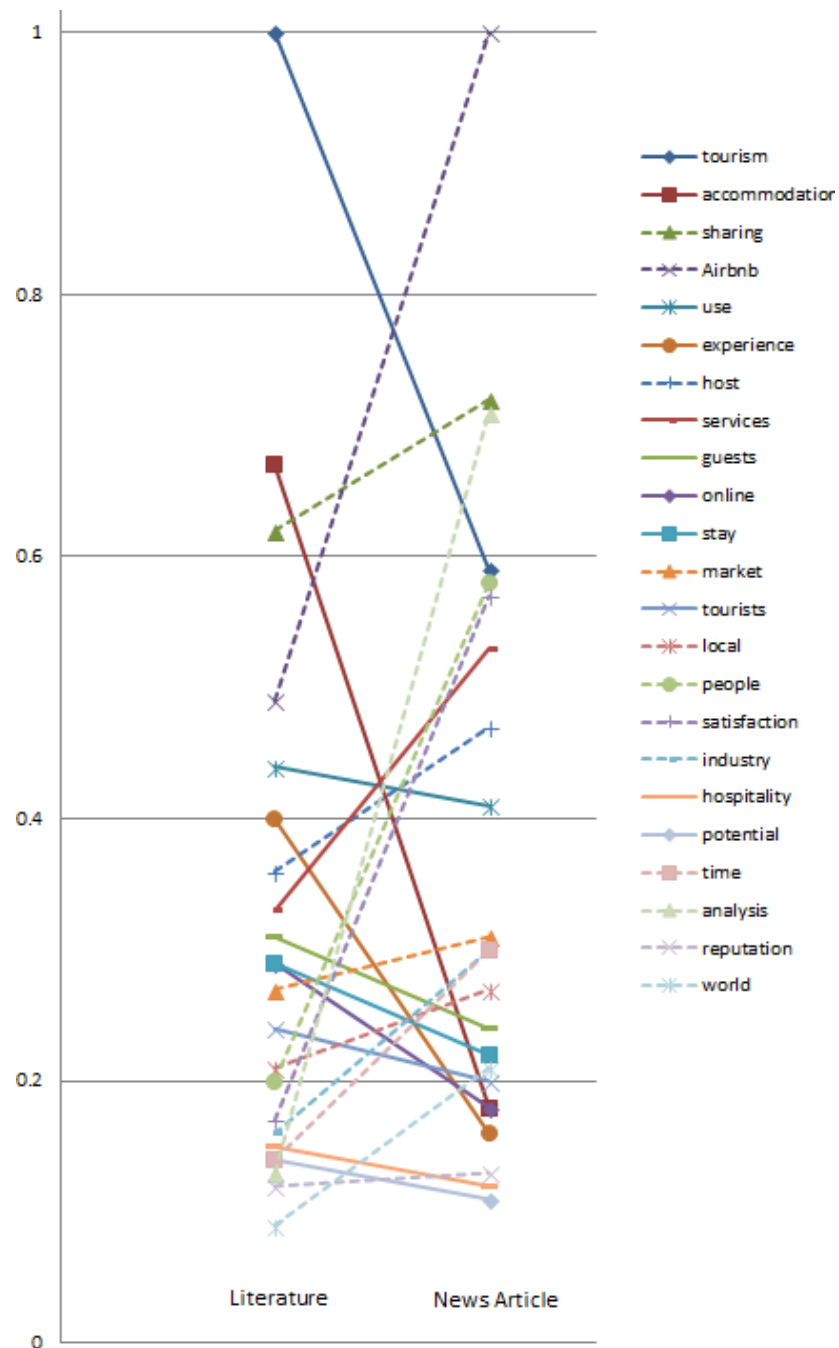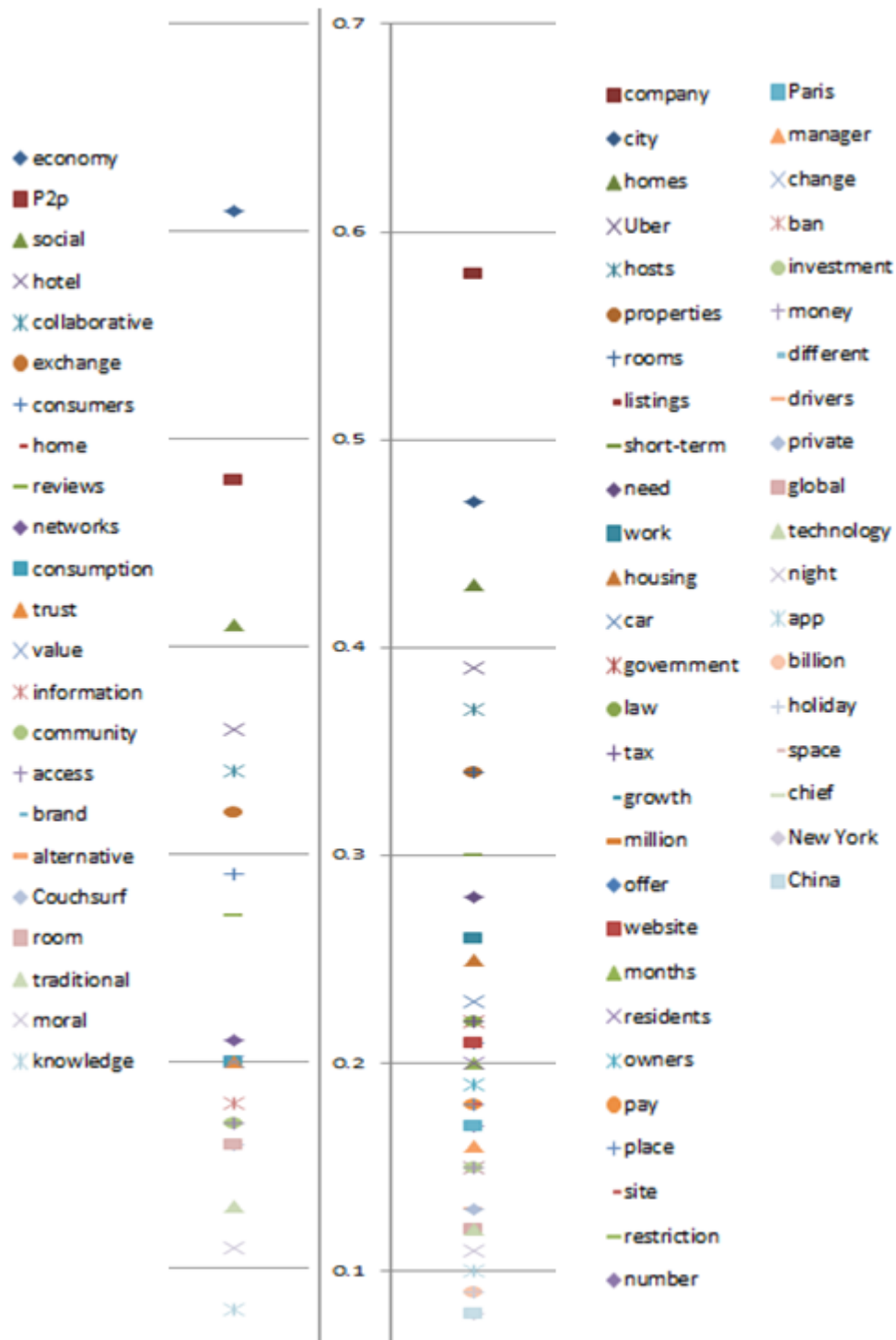

Figure 3.1: Concepts appear in both sources

*Dash Line indicates the relevance scores of literature are smaller than the one from News
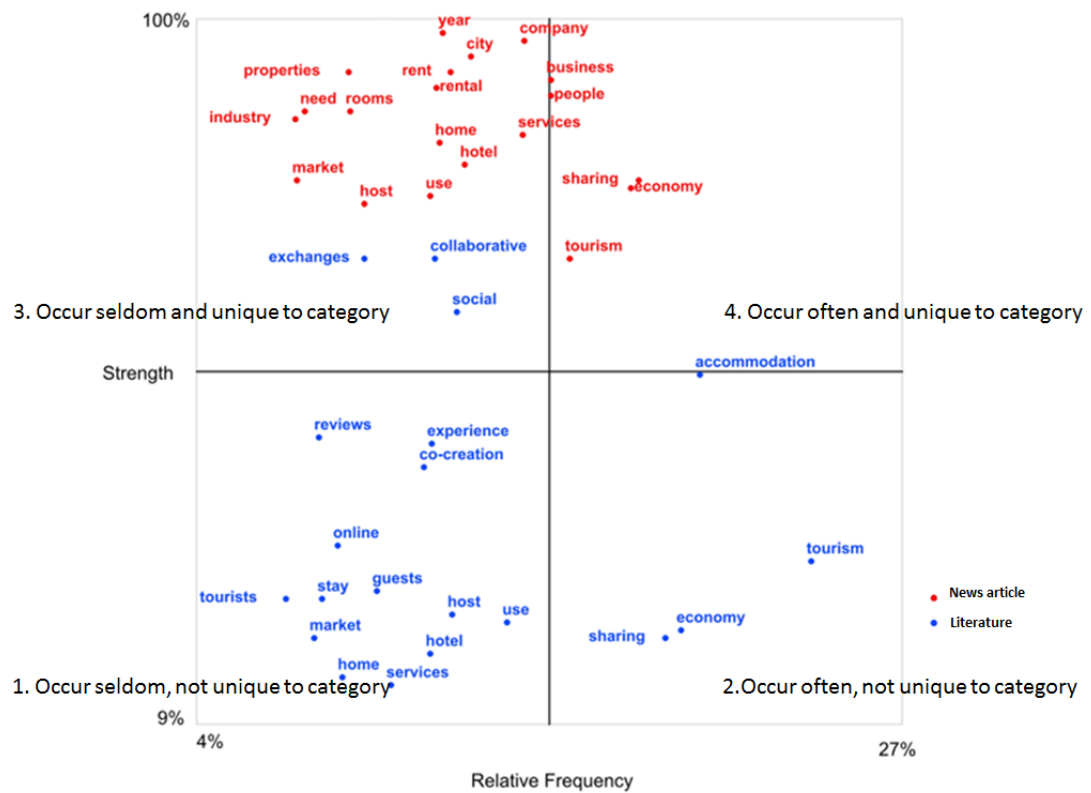
Figure 3.2: Concepts appear in each individual source

Figure 4: Comparison between concepts from literature and news articles using Solution 2

Discussion

The comparative ACA approach through Leximancer offers a visual representation of the building blocks of the current SE literature and news articles by revealing key concepts and themes to identity emerging research areas. The analysis of both sources of information suggests that two broad areas exist in TH literature (1) the nature of the SE and its relationship to TH; and (2) SE's accommodation sector and three broad areas in news articles (1) SE and its impacts; (2) growth of the SE; and (3) SE companies. The comparison between both sources of information indicates that TH literature is not yet fully addressing the emerging issues in TH and the SE is growing more rapid than the researchers can comprehend. The areas of mobility, sustainability, and the SE's business models are clearly under-represented or missing in the existing TH academic literature. Various business models and confusing approaches taken by regulatory authorities highlight the urgent need for academics to investigate the SE at various levels. While this research has chosen the SE as a topic of inquiry, its main purpose is to provide a coherent and nuanced methodological understanding of ACA in TH field. As such, only a brief discussion of the findings is presented.

News articles demonstrate that there exists various types of SE companies dependent on their positions in the spectrum of transactions, business approaches, and governance models (Cohen, 2016). For example, Airbnb represents market pricing, whereas alternative transaction models use "alternative dollars" (e.g. Bliive a barter website in Brazil) for exchange, which users can later use for other services/goods (Cohen, 2016). It is unclear how these different types of SE companies operate and their associated impacts on the current tourism social-economic system, in particular with new SE models emerging from non-traditional markets, such as China.

While news articles acknowledge the SE's contribution to a new form of sustainable tourism consumption by utilizing under-used resources, they also question the actual contribution of the SE to reduce consumption as the SE reduces costs (e.g. Uber vs taxi) and improves convenience and might actually boost tourists' consumption (OECD, 2016). As a result, this may lead to serious effects on infrastructure, such as traffic congestion and increased housing demand in cities. For example, the relatively cheaper price of taking Uber than a taxi for a short distance might monetarily encourage tourists to take Uber rather than use public transportation. It highlights that considerations of the SE being universally "green" may be overly optimistic. It raises the issue of how to identify conditions that harness the potential of the SE to achieve sustainability in tourism.

From a host perspective, news articles show that sharing one's house for a fee is operated largely in a self-regulated environment (Lyons & Wearing, 2015). How can standards be applied for the benefits of the host and tourists (e.g. safety), as the SE is self-regulated and limited complaint reporting mechanisms raises on-going unresolved concerns over public safety and government regulations (Cannon & Summers, 2014). The results show that regulatory authorities have been taking various and, sometimes, contradictory measures in regulating the SE (Frankfurt against Uber while Amsterdam promotes Airbnb). A further examination of news articles concerning policy implications rising from the SE reveals different strategic policy responses. While these policy implications present opportunities, it is important for authorities to be flexible in regulatory frameworks in order to balance various stakeholders' interests by drawing the line between peer-to-peer sharing and conventional commercial tourism activities (Johal & Zon, 2015; OECD, 2016). Governments should benefit from establishing a threshold that allows SE businesses to thrive while protecting workers and consumers and balancing the interests from traditional tourism service providers (Lyons &

Wearing, 2015). More importantly, it clearly raises the question as to how government can develop effective strategies to integrate SE companies into existing TH economic systems given the continued resistance from traditional TH players.

The concept of people's mobility indicates that Airbnb helps remote tourism destinations develop tourism industries by increasing tourists' length of stay. Additionally, the creation of business opportunities in these remote areas can stimulate the whole supply chain, particularly creating employment opportunities (e.g. the need for extra cleaners). Further, this concept is also related to shared mobility – shared use of a vehicle (e.g. bicycle sharing) to enhance transportation accessibility at tourism destinations.

While this study has confirmed the effectiveness and efficiency of using ACA, it also reveals that ACA can facilitate a systematic comparison between different sources of data. However, such comparisons should be undertaken with caution as different sources are written in different styles and context. As a result, concepts are context-dependent with variations in their definitions and scope. However, on a closer review of the existing literature using ACA, we realized that many researchers have taken ACA for granted when applying to their research, with sometimes contradictory suggestions. Thus, based on the use of Leximancer in this study and previous studies which have analyzed interview transcripts (Povey et al., 2013), social media (e.g. review comments, blogs) (Pearce & Wu, 2015), academic literature (Jin & Wang, 2016), and news articles (Cheng, 2016a), a series of iterative steps and considerations for its use are presented.

Table 1: Steps for undertaking automated content analysis by using Leximancer

	Steps	Reasons
Step 1	Understand the nature of the data	Documents are written in different styles and contexts. As such, concepts are context-dependent with variations in their definitions and scopes. Examples: Social Media (e.g. twitter) – Messy, short and lots of abbreviations Interview data: Casual and extended style of discussion
Step 2	Prepare the data	Unnecessary information that can potentially crowd the analysis and prepare the data in a format that feeds into the Leximancer operating system Examples: Literature review: delete - abstract, references, author name and affiliation, acknowledgements
Step 3	Adjust concepts seeds/text classification	• Increase sentence blocks to three or more to identify emerging concepts. • Using "tag" function to identity the relationship between category and attributes (interview: who talks about what)
Step 4	Interpret the concepts	• Higher level of reasoning in synthesizing the conceptual and relational insights • Comparative approach within and across different sources of data • Visual representation of the results • Return to the original text with associated concepts to construct narratives

In addition, this paper contributes to the ACA literature by presenting a solution for attending to the issue of sample size where comparative measures are performed between different data sources. As concepts are automatically selected in ACA, it is recommended that the researcher use Solution 1 so that they can avoid mixing two different mental models which can inflate the comparative results. More importantly, Solution 1 effectively characterizes the minority data components and identifies emerging concepts. Further, this paper highlights a visual analytics approach (Figures 1, 2, 3.1 and 3.2) to analyze and present the research findings (Cheng & Edwards, 2015; Edwards et al., 2017). With increasing volumes of data, data visualization will become an important component of ACA. However, this study demonstrates that interpretation of the visual representation takes considerable time and effort as using ACA requires both macro and micro insights of the phenomena to make insightful inferences (McAbee, Landis, & Burke, 2017). Thus, this research demonstrates the need for TH researchers to be comfortable in interpreting visual representations when using ACA.

Conclusion

In summary, this study extends current TH literature in two aspects. First, academically, the visual representation of the results offers a clearer and richer picture of the themes and concepts of the SE discourse in TH. It offers some insights that have been overlooked in recent reviews of the SE discourse (see Cheng, 2016a; Heo, 2016). Secondly, it contributes to methodological literature in TH by using an ACA approach in an integrated and comparative manner. Importantly, it presents a series of clear and transparent steps of ACA using the tool - Leximancer - that could be used for future TH research. More importantly, the approach used here differs from the conventional use of ACA in TH that focuses primarily on a single source of information by bringing together two separate sources of knowledge. Thus, this comparative ACA approach advances the methodological literature concerning ACA.

While this article provides valuable insights through its comparative approach, there are two areas for further research. First, while the issues raised in news articles on the SE are considered relevant and important, they may be influenced by gate-keepers – media professionals. Media professionals' select news topics based on their perceived appeal and importance to their target audience (Koopmans, 2004). This competitive selection process signals social construction, which invites further research. Second, this study offers a starting-point for researchers to re-evaluate how news articles influence academics' research agenda. Increasingly, researchers are required to demonstrate the research impacts of their research beyond academic publication and engagement in addressing issues revealed in news articles could be an approach.

References

Angus, D., Rintel, S., & Wiles, J. (2013). Making sense of big text: a visual-first approach for analysing text data using Leximancer and Discursis. *International Journal of Social Research Methodology, 16*(3), 261-267.

Bednarek, M. (2006). *Evaluation in media discourse: Analysis of a newspaper corpus*. London: A&C Black.

Biesenthal, C., & Wilden, R. (2014). Multi-level project governance: Trends and opportunities. *International Journal of Project Management, 32*(8), 1291-1308.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993-1022.

Campbell, C., Pitt, L., Parent, M., & Berthon, P. (2011). Understanding Consumer Conversations around Ads in a Web 2.0 World. *Journal of Advertising, 40*(1), 87-102.

Cannon, S., & Summers, L. H. (2014). How Uber and the Sharing Economy Can Win Over Regulators. *Harvard business review, 2015*(13).

Cheng, M. (2016a). Current sharing economy media discourse in tourism. *Annals of Tourism Research, 60*, 111-114. doi:10.1016/j.annals.2016.07.001

Cheng, M. (2016b). Sharing economy: A review and agenda for future research. *International Journal of Hospitality Management, 57*, 60-70. doi:http://dx.doi.org/10.1016/j.ijhm.2016.06.003

Cheng, M., & Edwards, D. (2015). Social media in tourism: a visual analytic approach. *Current Issues in Tourism, 18*(11), 1080-1087. doi:10.1080/13683500.2015.1036009

Cheng, M., Edwards, D., Darcy, S., & Redfern, K. (2016). A tri-method approach to a review of adventure tourism literature: bibliometric analysis, content analysis and a quantitative systematic literature review. *Journal of Hospitality & Tourism Research*. doi:10.1177/1096348016640588

Cohen, B. (2016). Making Sense Of The Many Business Models In The Sharing Economy. Retrieved from http://www.fastcoexist.com/3058203/making-sense-of-the-many-business-models-in-the-sharing-economy

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for information Science, 41*(6), 391.

Edwards, D., Cheng, M., Wong, A., Zhang, J., & Wu, Q. (2017). Ambassadors of Knowledge Sharing: Co-produced travel information through tourist-local social media exchange. *International Journal of Contemporary Hospitality Management, 29*(2), 690-708. doi:10.1108/IJCHM-10-2015-0607

Esser, F., & Hanitzsch, T. (2013). *The handbook of comparative communication research*. London: Routledge.

Evans, M., McIntosh, W., Lin, J., & Cates, C. (2007). Recounting the courts? Applying automated content analysis to enhance empirical legal research. *Journal of Empirical Legal Studies, 4*(4), 1007-1039.

Forno, F., & Garibaldi, R. (2015). Sharing Economy in Travel and Tourism: The Case of Home-Swapping in Italy. *Journal of Quality Assurance in Hospitality & Tourism, 16*(2), 202-220. doi:10.1080/1528008x.2015.1013409

Germann Molz, J. (2013). Social Networking Technologies and the Moral Economy of Alternative Tourism: The Case of Couchsurfing.Org. *Annals of Tourism Research, 43*, 210-230. doi:10.1016/j.annals.2013.08.001

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism, 18*(12), 1192-1217. doi:10.1080/13683500.2013.827159

Heo, C. Y. (2016). Sharing economy and prospects in tourism research. *Annals of Tourism Research*. doi:http://dx.doi.org/10.1016/j.annals.2016.02.002

Jin, X., & Wang, Y. (2016). Chinese outbound tourism research: A review. *Journal of Travel Research, 55*(4), 440-453. doi:10.1177/0047287515608504

Johal, S., & Zon, N. (2015). *Policymaking for the Sharing Economy*. Retrieved from Toronto:

Juul, M. (2015). The sharing economy and tourism.   Retrieved from http://www.europarl.europa.eu/RegData/etudes/BRIE/2015/568345/EPRS_BRI(2015)568345_EN.pdf

Koopmans, R. (2004). Movements and media: Selection processes and evolutionary dynamics in the public sphere. *Theory and Society, 33*(3-4), 367-391.

Leximancer. (2011). *Leximancer Manual Version 4*. Retrieved from Brisbane:

Low, D. R., & Eagle, L. (2016). Climate Change Science Versus Climate Sceptics: Is the World Really Flat? In C. Campbell & J. J. Ma (Eds.), *Looking Forward, Looking Back: Drawing on the Past to Shape the Future of Marketing: Proceedings of the 2013 World Marketing Congress* (pp. 643-653). Cham: Springer International Publishing.

Lyons, K., & Wearing, S. (2015). *The sharing economy issues, impacts, and regulatory responses in the context of the NSW visitor economy*. Retrieved from Sydney:

MacKinnon, M. (2015). Representations of homelessness in'The Australian'newspaper, 2008-2012. *Australian Journalism Review, 37*(1), 165.

McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review, 27*(2), 277-290.

McCombs, M. (2013). *Setting the agenda: The mass media and public opinion*. Cambridge: Polity Press.

Newman, D., Noh, Y., Talley, E., Karimi, S., & Baldwin, T. (2010). *Evaluating topic models for digital libraries.* Paper presented at the Proceedings of the 10th annual joint conference on Digital libraries, Gold Coast, Australia.

Nunez-Mir, G. C., Iannone, B. V., Pijanowski, B. C., Kong, N., & Fei, S. (2016). Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods in Ecology and Evolution, 7*(11), 1262-1272.

OECD. (2016). *Policies for the tourism sharing economy*. Retrieved from Paris:

Pearce, P. L., & Wu, M.-Y. (2015). Entertaining International Tourists An Empirical Study of an Iconic Site in China. *Journal of Hospitality & Tourism Research*, 1096348015598202.

Pfeffer, J. (2007). A modest proposal: How we might change the process and product of managerial research. *Academy of Management Journal, 50*(6), 1334-1345.

Povey, J., Shaw, E., Head, B., Cherney, A., Boreham, P., & Ferguson, M. (2013). *The functionality of Leximancer software as an analytic tool for semi-structured interview data exploring evidence-based policy in practice from the perspectives of academic social scientists and policymakers.* Paper presented at the AMSRS Conference, Sydney.

ProQuest. (2014). The World's Most Comprehensive Collection of News Content.   Retrieved from http://www.proquest.com/libraries/academic/news-newspapers/

PwC. (2015). *The Sharing Economy*. Retrieved from London:

Queensland Tourism Industry Council. (2014). *THE SHARING ECONOMY: HOW IT WILL IMPACT THE TOURISM LANDSCAPE AND WHAT BUSINESSES CAN DO*. Retrieved from Brisbane: https://www.qtic.com.au/sites/default/files/140714_draft_sharing_economy_paper.pdf

Randhawa, K., Wilden, R., & Hohberger, J. (2016). A bibliometric review of open innovation: Setting a research agenda. *Journal of Product Innovation Management, 33*(6), 750-772.

Richard, B., & Cleveland, S. (2016). The future of hotel chains Branded marketplaces driven by the sharing economy. *Journal of Vacation Marketing*, 1356766715623827.

Schmidt, A., Ivanova, A., & Schäfer, M. S. (2013). Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries. *Global Environmental Change, 23*(5), 1233-1248.

Schweinsberg, S., Darcy, S., & Cheng, M. (2017). The agenda setting power of news media in framing the future role of tourism in protected areas. *Tourism Management, 62*, 241-252.

Scopus. (2017). Content Policy and Selection.   Retrieved from https://www.elsevier.com/solutions/scopus/content/content-policy-and-selection

Sigala, M. (2015). Collaborative commerce in tourism: implications for research and industry. *Current Issues in Tourism*, 1-10. doi:10.1080/13683500.2014.982522

Smith, A. E., & Humphreys, M. S. (2006). Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavior Research Methods, 38*(2), 262-279.

Tseng, C., Wu, B., Morrison, A. M., Zhang, J., & Chen, Y. (2015). Travel blogs on China as a destination image formation agent: A qualitative analysis using Leximancer. *Tourism Management, 46*, 347-358. doi:10.1016/j.tourman.2014.07.012

Tussyadiah, I. P., & Pesonen, J. (2015). Impacts of Peer-to-Peer Accommodation Use on Travel Patterns. *Journal of Travel Research*. doi:10.1177/0047287515608505

Watson, M., Smith, A., & Watter, S. (2005). Leximancer concept mapping of patient case studies. In R. Khosla, R. J. Howlett, & L. C. Jain (Eds.), *Knowledge-based intelligent information and engineering systems* (pp. 1232-1238). Berlin: Springer.