Tourism Management 74 (2019) 276–289

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis

Zhiping Hou[a], Fasheng Cui[a], Yongheng Meng[a,*], Tonghui Lian[b], Caihua Yu[c]

[a] *Business School of Guilin University of Technology, Guilin, 541004, China*
[b] *School of Business Administration, Nanjing University of Finance and Economics, Nanjing, 210023, China*
[c] *School of Economics and Management, Nanjing University of Information Science and Technology, Nanjing, 210044, China*

ARTICLE INFO

Keywords:
Online travel reviews
Semantic association analysis
Opinion mining
Social network analysis

ABSTRACT

Online tourism reviews provide a crucial source of information for the tourism industry, and determining whether they can be effectively identified is key to influencing tourism decision-making. The purpose of this paper is to identify themes and compare differences in online travel reviews. A semantic association analysis was applied to extract thematic words and construct a semantic association network from 165,429 reviews obtained from three major online travel agencies (OTAs) in China. The findings show that there are apparent discrepancies on these platforms in terms of thematic words, the distribution of topics, structural properties and community relationships. In particular, the results of network visualization can clearly identify hot topics and the social network relationships of thematic words. The proposed analytical framework expands our understanding of the methodological challenges and offers novel insights for mining the opinions for the benefit of tourists, hotels and tourism enterprises and OTAs.

1. Introduction

With the rapid development of Web 2.0 technologies, online user-generated content (UGC) such as online travel reviews, has been widely used in the tourism and hospitality industry. Tourists tend to share their travel experiences through online travel agencies (OTAs) such as TripAdvisor (Guo, Barnes, & Jia, 2017; Liu, Schuckert, & Law, 2018), Expedia (Xiang, Schwartz, Gerdes, & Uysal, 2015), Yelp (Papathanassis & Knolle, 2011), Lvmama (Lian & Yu, 2017), Ctrip (Ye, Law, & Gu, 2009), Qunar (Zhang et al., 2016a, b), etc. Such descriptions of tourists' experiences, which are actively shared by users, have widely been regarded as a typical type of online travel review. Online travel reviews include hotel reviews, restaurant reviews, and attraction reviews (Xiang et al., 2015), which are the most popular sources of information for tourists in obtaining travel information and making travel plans as well as booking tickets and hotels (Li & Yi, 2014; Zhang et al., 2016a, b). Unlike ordinary consumer products, the adoption of travel involves more than a simple purchase decision, and tourism consumption is based on public willingness. Consumers are easily influenced by other people's opinions, and their willingness to search for the opinions and experiences of peer consumers before purchasing a product has been found to be relevant (García-Pablos, Cuadros, & Linaza, 2016; Guo et al., 2017; Hu, Chen, & Chou, 2017; Schuckert, Liu, & Law, 2015a). Industry data indicate that approximately 77% of prospective travellers will either "always" or "usually" not make decisions until they read online reviews (Ye et al., 2009). Travellers can minimize travel costs and obtain indirect purchasing experiences by browsing information nodes (e.g., government tourism websites, online travel portals), thereby reducing their perceived uncertainty and achieving an enjoyable psychological experience (Lian & Yu, 2017; Xiang et al., 2015; Ye, Law, Gu, & Chen, 2011).

Online reviews have the characteristics of noise. Tourists post deceptive reviews when they want to achieve some purpose (Min, Lim, & Magnini, 2015); for example, visitors might sometimes post a fake positive review to avoid unnecessary trouble or to receive kickbacks (Schuckert, Liu, & Law, 2015b), resulting in consumers' inability to quickly gain access to useful information (Liu & Park, 2015). Therefore, the question of whether online travel reviews can be identified and adopted by latent consumers is the key to influencing tourism decisions. Online reviews not only provide convenience to consumers in the search for information but also increase consumers' cognitive costs. Consumers can become confused and lost when faced with a massive quantity of online reviews, resulting in poorer decision-making and intangible pressure (Bellman, Johnson, Lohse, & Mandel, 2010;

* Corresponding author.
E-mail addresses: hzp@glut.edu.cn (Z. Hou), fashengcui@glut.edu.cn (F. Cui), yhmeng@glut.edu.cn (Y. Meng), liantonghui@njue.edu.cn (T. Lian), yucaihua@nuist.edu.cn (C. Yu).

https://doi.org/10.1016/j.tourman.2019.03.009
Received 11 November 2018; Received in revised form 31 January 2019; Accepted 18 March 2019

旅游管理74（2019）276-289

内容列表可在 ScienceDirect

旅游管理

期刊主页：www.elsevier.com/locate/tourman

基于在线旅游评论的意见挖掘：基于语义关联分析的中国主要在线旅行社对比分析

侯志平[a]、崔发生[a]、孟永恒[a,*]、联同辉[b]、余彩华[c]

〔0〕桂林工学院商学院，桂林，541004,中国
〔0〕南京大学财经学院工商管理学院，南京，210023,中国
〔0〕南京信息工程大学经济管理学院，南京，210044,中国

文章信息

关键词:
在线旅游评论 语义关联分析
意见挖掘
社会网络分析

摘要

在线旅游评论为旅游业提供了一个重要的信息来源，确定它们是否能被有效识别是影响旅游决策的关键。本文的目的是识别主题并比较在线旅游评论中的差异。通过语义关联分析，从中国三大在线旅行社的165429篇评论中提取主题词并构建语义关联网络。研究结果表明，这些平台在主题词、主题分布、结构属性和社区关系方面存在明显差异。特别是，网络可视化的结果可以清晰地识别热门话题和主题词的社会网络关系。提出的分析框架扩大了我们对方法学挑战的理解，并为挖掘游客、酒店、旅游企业和在线旅行社的利益观点提供了新的见解。

1. 简介

随着Web 2.0技术的快速发展，在线用户生成内容（UGC）如在线旅游评论已被广泛应用于旅游业和酒店业。游客倾向于通过在线旅行社（OTA）分享他们的旅行体验，如TripAdvisor（郭、巴恩斯和贾，2017；刘、舒克特和罗，2018）、Expedia（项、施瓦茨、格德斯和乌萨尔，2015）、Yelp（Papathanassis和Knole，2011）、Lwmama（Lian和Yu, 2017）、携程（叶、罗和顾，2009）、去哪儿（Zhang等人，2016a, b），这种对游客体验的描述被用户积极分享，被广泛认为是一种典型的在线旅游评论。在线旅游评论包括酒店评论、餐厅评论和景点评论（Xiang et al., 2015），这是游客获取旅游信息、制定旅游计划以及预订机票和酒店时最受欢迎的信息来源（Li & Yi, 2014；Zhang et al., 2016a, b）。与普通消费品不同，旅游产品的采用不仅仅是一个简单的购买决定，旅游消费基于公众意愿。消费者很容易受到其他人意见的影响，他们在购买产品之前搜索同行消费者意见和体验的意愿被发现是相关的（García-Pablos、Cuadros和Linaza, 2016；Guo等人，2017年；胡陈周，2017；舒克特、刘和洛，2015a）。行业数据表明，大约77%的潜在旅行者在阅读在线评论之前"总是"或"通常"不会做出决定（Ye等人，2009年）。旅行者可以通过浏览信息节点（如政府旅游网站、在线旅游门户网站）来最小化旅行成本并获得间接购买体验，从而减少他们感知的不确定性，并获得愉快的心理体验（Lian & Yu, 2017；Xiang et al., 2015；Ye, Law, Gu, & Chen, 2011）。

在线评论具有噪音的特点。游客为了达到某种目的而发布虚假评论（Min、Lim和Magnini, 2015）；例如，访客有时可能会发布虚假正面评论，以避免不必要的麻烦或收取回扣（Schuckert、Liu和Law, 2015b），导致消费者无法快速获取有用信息（Liu和Park, 2015）。因此，在线旅游评论能否被潜在消费者识别和采纳的问题是影响旅游决策的关键。在线评论不仅为消费者搜索信息提供了便利，还增加了消费者的认知成本。消费者在面对大量在线评论时可能会感到困惑和迷茫，从而导致决策能力下降和无形压力（Bellman、Johnson、Lohse和Mandel, 2010；

* 通讯作者。
电子邮件地址：hzp@glut.edu.cn（侯志强），fashengcui@glut.edu.cn（F.崔），yhmeng@glut.edu.cn（孟永康），liantonghui@njue.edu.cn（T. Lian），yucaihua@nuist.edu.cn（余振中）。

https://doi.org/10.1016/j.tourman.2019.03.009
2018年11月11日收到；2019年1月31日收到修订版表格；2019年3月18日接受

Schuckert et al., 2015a). The quality of online review information affects consumers' attitudes towards adopting information sources and making tourism decisions (Filieri & Mcleay, 2014; Schuckert et al., 2015a). The length, social characteristics, readability, accuracy of information, perceived value, and language style of reviews are the main factors that affect the quality of information acquisition (Filieri & Mcleay, 2014; Hodac, Carson, & Moore, 2013; Li, Xu, Tang, Wang, & Li, 2018; Papathanassis & Knolle, 2011; Racherla & Friske, 2012; Schuckert et al., 2015b). Papathanassis and Knolle (2011) argued that when the readability of online reviews is too low, consumers will show lower reservation intention, which affects the competitiveness of tourism enterprises (e.g., in terms of reputation and revenue). Therefore, quickly obtaining high-quality information from a large number of online reviews has become a critical issue in the field of practice and academia.

Prior studies on online travel reviews have been conducted using questionnaire surveys (Min et al., 2015), mathematical models (Liu & Park, 2015), statistical analysis (Racherla & Friske, 2012; Schuckert et al., 2015b), and grounded theory (Filieri, 2016; Papathanassis & Knolle, 2011). Online reviews are text-based and often comprise large information repositories that go beyond the analytical capabilities of traditional econometric and statistical methods (Guo et al., 2017). It is difficult to mine statistically meaningful differences among different types of groups within the review data (Boo & Busser, 2018; Xiang, Du, Ma, & Fan, 2017). Text analysis technology based on natural language processing (NLP) can automatically address large amounts of customer-generated reviews and comments from the perspective of word granularity, and this method is widely used in topic identification and opinion mining (García-Pablos et al., 2016). By utilizing the rich concept hierarchy structure and semantic knowledge provided by ontology, semantic association analysis can quickly extract significative topic words from a large number of texts and achieve business intelligence association analysis at a semantic level (Boo & Busser, 2018; Xiang et al., 2017). The approach is more reliable than traditional text analysis technology (e.g., latent Dirichlet analysis, fuzzy domain ontology, and support vector machine), and the analysis results have more reference value and applicability (Filieri, 2016). From a semantic point of view, the association of two documents in online travel reviews is often decided not only by literal repetition but also by the semantic logic behind the words. The combination of semantic idea and association analysis methods can eliminate sampling bias and refine the observation granularity of topic identification to the word level in the research on online travel reviews. As a result, it can significantly improve the efficiency and quality of online travel review analysis by accurately mining the centre of different topic groups and the degree of association of keywords, and it can achieve value added from information to knowledge to intelligence (He, 2013; Li et al., 2018).

With this in mind, we attempt to mine the opinions within online travel reviews on the three platforms by using semantic association analysis, providing practical help for understanding tourists' behaviour and leading to improvements in tourism industry marketing.

Specifically, the main aim of this study is to answer the following questions:

(1) What topics are discussed in online travel reviews?
(2) How can bigrams of co-occurrence phrases of semantic association be constructed?
(3) What are the structural properties of semantic association networks?
(4) How can we identify the focus of a discussion by visualizing semantic association networks?
(5) What are the differences among different OTAs?

The remainder of this paper is organized as follows. Section 2 provides background for the research, which reviews the status of the research on online travel reviews, opinion mining in online travel, and semantic association analysis and SNA. Section 3 provides a methodological description of our study accompanied by an introduction to how to collect and analyse data. Section 4 conducts an analysis of the current study's results, including statistical analysis of thematic words, the construction of bigrams of the co-occurrence of semantic association, analysis of the structural properties of semantic association networks, and visualizations. Section 5 presents discussion, conclusions, limitations and future directions.

2. Background

2.1. Online travel reviews

Online reviews are posted by consumers who have purchased and used a product, and they include consumers' experiences, evaluations, and opinions (Litvin, Goldsmith, & Pan, 2008). Online reviews are a typical type of eWOM (electronic word-of-mouth) and have important value for consumers, enterprises and sellers (Litvin et al., 2008). For consumers, online reviews are not only a channel for obtaining product information but also the basis for making tourism decisions. The opinions expressed by commentators on products or services affect other consumers' purchase intention (Wang & Wang, 2018). For enterprises and sellers, the features of online reviews, such as comprehensiveness (Zhao, Liang, Xiao, & Law, 2015), professionalism (Ladhari & Michaud, 2015), quality (Wenchin, Mingtsang, Liwen, & Lin, 2015), and reputation (Casaló, Flavián, Guinalíu, & Ekinci, 2015), have a significant impact on product sales (Hodac et al., 2013).

Online travel reviews are rich, complex bundles of information that reflect travellers' experiences and evaluations of products (Duverger, 2013; Hemmatian & Sohrabi, 2017; Litvin et al., 2008; Xiang et al., 2015), which are important sources of information for tourists to facilitate the making of travel arrangements (Bucur, 2015). Many studies have found significant effects of such online travel reviews on the tourism industry, such as consumer travel decision behaviour, tourism product sales, and tourism destination image. For instance, Ye et al. (2011), Xiang et al. (2015), and Lian and Yu (2017) explored the relationship between online travel reviews and consumer purchase intention and found that consumers can reduce the cost of information search and indirectly obtain travel experience, which may promote their willingness to purchase. Li and Yi (2014) showed that online travel reviews are closely related to the financial status of tourism products, and hotels with a higher number of positive reviews obtain more reservations. Online travel reviews, which reflect the reputation and satisfaction of tourism destinations, are an important part of a tourism destination network image and have a direct impact on tourists' perceived quality, satisfaction, and behavioural intention (Lian & Yu, 2017; Yuan & Wu, 2016). Somabhai, Varma, and Somabhai (2015) argued that consumers can obtain an intuitive impression of tourism destinations by reading other travellers' online travel reviews, helping them to reduce risk uncertainty and effectively make travel plans. Although online travel reviews provide convenience for consumers' travel decisions, massive reviews have also exacerbated confusion with regard to information overload in the big data era (Fang, Ye, Kucukusta, & Law, 2016; Ren & Hong, 2017; Zhang, 2014, pp. 1–46). Extracting key points from online textual reviews is complex and challenging, but it is crucial for predicting, interpreting and responding to customer behaviour (Wang, Yang, Sun, & Jiang, 2017).

2.2. Opinion mining of online travel reviews

Opinion mining is a technology that automatically extracts online comment information by using textual analysis, including computer language and natural language processing. It analyses people's opinions, appraisals, attitudes, and emotions towards organizations, entities, person, issues, actions, topics and their attributes (García-Pablos et al., 2016; Hemmatian & Sohrabi, 2017). The main task of opinion

Schuckert等人，2015a）。在线评论信息的质量会影响消费者对采用信息源和做出旅游决策的态度（Filieri & Mcleay, 2014；Schuckert等人，2015a）。信息的长度、社会特征、可读性、准确性、感知价值、，评论的语言风格是影响信息获取质量的主要因素（Filieri & Mcleay, 2014；Hodac, Carson和Moore, 2013；Li, Xu, Tang, Wang和Li, 2018；Papathanassis和Knole，2011；Rachella和Friske, 2012；Schuckert等人，2015b）。Papathanassis和Knolle（2011）认为，当在线评论的可读性太低时，消费者会表现出较低的预订意愿，这会影响旅游企业的竞争力（例如声誉和收入）。因此，从大量在线评论中快速获取高质量信息已成为实践和学术界的一个关键问题。

此前对在线旅游评论的研究采用了问卷调查（Min等人，2015年）、数学模型（Liu & Park, 2015年）、统计分析（Racherla & Friske, 2012年；Schuckert等人，2015b）和扎根理论（Filieri，2016年；Papathanassis&Knole，2011年）。在线评论基于文本，通常包含大型信息库，超出了传统计量经济学和统计方法的分析能力（郭等人，2017年）。在审查数据中，很难挖掘不同类型群体之间具有统计意义的差异（Boo & Busser, 2018；Xiang, Du, Ma和Fan, 2017）。基于自然语言处理（natural language processing，NLP）的文本分析技术可以从单词粒度的角度自动处理大量客户生成的评论和评论，这种方法广泛应用于主题识别和观点挖掘（García-Pablos et al., 2016）。语义关联分析利用本体提供的丰富概念层次结构和语义知识，可以从大量文本中快速提取有意义的主题词，实现语义层面的商务智能关联分析（Boo & Busser, 2018；Xiang等人，2017）。该方法比传统的文本分析技术（如潜在Dirichlet分析、模糊领域本体和支持向量机）更可靠，分析结果更具参考价值和适用性（Filieri，2016）。从语义的角度来看，在线旅游评论中两个文档的关联通常不仅取决于字面重复，还取决于单词背后的语义逻辑。在在线旅游评论研究中，语义思想和关联分析方法的结合可以消除抽样偏差，将主题识别的观察粒度细化到词级。因此，通过准确挖掘不同主题组的中心和关键词的关联度，可以显著提高在线旅游评论分析的效率和质量，实现从信息到知识再到智能的增值（何，2013；李等，2018）。

有鉴于此，我们试图通过语义关联分析挖掘三个平台上在线旅游评论中的观点，为理解游客行为提供实际帮助，并促进旅游业营销的改进。

具体而言，本研究的主要目的是回答以下问题:

(1) 在线旅游评论中讨论了哪些主题？
(2) 如何构建语义关联共现短语的双格图？
(3) 语义关联网络的结构特征是什么？
(4) 我们如何通过可视化语义关联网络来确定讨论的焦点？
(5) 不同的在线旅行社有什么不同？

本文的其余部分组织如下。第二部分介绍了本研究的背景，回顾了在线旅游评论、在线旅游中的意见挖掘和在线旅游的研究现状语义关联分析和SNA。第3节对我们的研究进行了方法学描述，并介绍了如何收集和分析数据。第四部分对本研究的结果进行了分析，包括主题词的统计分析、语义关联共现双格图的构建、语义关联网络的结构特性分析和可视化。第5节介绍了讨论、结论、局限性和未来方向。

2. 背景

2.1. *在线旅游评论*

在线评论由购买和使用过产品的消费者发布，包括消费者的体验、评估和意见（Litvin、Goldsmith和Pan, 2008）。在线评论是一种典型的电子口碑，对消费者、企业和卖家都有重要价值（Litvin等人，2008）。对于消费者来说，在线评论不仅是获取产品信息的渠道，也是做出旅游决策的基础。评论员对产品或服务发表的意见会影响其他消费者的购买意愿（Wang & Wang, 2018）。对于企业和卖家来说，在线评论的特点，如全面性（赵亮尚和罗，2015年）、专业性（拉德哈里和米肖，2015年）、质量（文钦、明曾、李文和林，2015年）和声誉（卡萨洛、弗拉维安、吉尼亚卢和埃金奇，2015年），对产品销售有重大影响（Hodac等人，2013年）。

在线旅游评论是一组丰富而复杂的信息，反映了旅行者的体验和对产品的评价（Duverger，2013；Hemmatian & Sohrabi, 2017；Litvin等人，2008；Xiang等人，2015），是游客促进旅游安排的重要信息来源（Bucur，2015）。许多研究发现，此类在线旅游评论对旅游业产生了重大影响，如消费者旅游决策行为、旅游产品销售和旅游目的地形象。例如，叶等人（2011年）、项等人（2015年）、廉和余（2017年）研究了在线旅游评论与消费者购买意愿之间的关系，发现消费者可以降低信息搜索成本，间接获得旅游体验，这可能会提升他们的购买意愿。Li and Yi（2014）表明，在线旅游评论与旅游产品的财务状况密切相关，正面评论越多的酒店获得的预订越多。在线旅游评论反映了旅游目的地的声誉和满意度，是旅游目的地网络形象的重要组成部分，直接影响游客的感知质量、满意度和行为意向（Lian & Yu, 2017；Yuan & Wu, 2016）。Somabhai、Varma和Somabhai（2015）认为，消费者可以通过阅读其他旅行者的在线旅行评论，获得对旅游目的地的直观印象，帮助他们减少风险不确定性，并有效地制定旅行计划。尽管在线旅游评论为消费者的出行决策提供了便利，但海量评论也加剧了大数据时代信息过载的困惑（方、叶、库库库斯塔和罗，2016；任和洪，2017；张，2014,第1–46页）。从在线文本评论中提取关键点既复杂又具有挑战性，但这对于预测、解释和回应客户行为至关重要（Wang、Yang、Sun和Jiang, 2017）。

2.2. *在线旅游评论的意见挖掘*

意见挖掘是一种通过文本分析（包括计算机语言和自然语言处理）自动提取在线评论信息的技术。它分析了人们对组织、实体、个人、问题、行动、话题及其属性的看法、评价、态度和情绪（García-Pablos等人，2016年；Hemmatian & Sohrabi, 2017年）。舆论的主要任务

mining can be divided into six categories: sentiment analysis, opinion extraction, sentiment mining, subjection analysis, affect or emotion analysis, and review mining (Kim & Park, 2017; Rahimi & Kozak, 2016; Sirgy, 2010; Xu & Mcgehee, 2016). All techniques used for opinion mining can be categorized into two main classes: lexicon-based approaches (Brandes, 2001; Daud, Khan, & Che, 2017; Telesford, Joyce, Hayasaka, Burdette, & Laurienti, 2011) and machine learning (Leclerc & Martin, 2004; Weiler, 2002; Weiler & Walker, 2014). The lexicon-based method classifies text sentiment polarity by relying on a sentiment dictionary and linguistic knowledge approach, which includes a corpus-based approach and a dictionary-based approach. The machine learning approach, which benefits from machine learning algorithms, can be divided into three groups: supervised learning, semi-supervised learning and unsupervised learning (García-Pablos et al., 2016; Hemmatian & Sohrabi, 2017). The approach extracts the text information from product reviews by feature construction technology (such as the bag-of-words) and uses a classification method to analyse online reviews (Raisi, Baggio, Barratt-Pugh, & Willson, 2017). Within the online travel review context, the proposed opinion mining method has significant advantages of acceptable accuracy and resource savings in dealing with unstructured review texts (Bucur, 2015). Guo et al. (2017) analysed 266,544 online reviews extracted from 25,670 hotels located in 16 countries to identify key dimensions of customer service based on latent Dirichlet analysis (LDA) data mining methods and uncovered 19 controllable dimensions that are key for hotels to manage their interactions with visitors. Ali, Kwak, and Kim (2016) proposed a fuzzy domain ontology (FDO) and support vector machine (SVM) opinion mining system to automatically classify online reviews.

2.3. Semantic association analysis and SNA

Semantic association analysis, which is an important text analysis method (Alemanmeza, Halaschekwiener, Sahoo, Sheth, & Arpinar, 2005; Xiang, Tian, & Huang, 2007), was first mentioned in the study of the brain's response to expected words (Kutas & Hillyard, 1984). The core idea of semantic association analysis is that semantics are defined by the co-occurrence of two words in a sentence with high-frequency words as a node, taking into consideration the frequency of high-frequency phrase co-occurrence as a link between nodes (Alemanmeza et al., 2005; García-Pablos et al., 2016; Schuckert et al., 2015a; Xiang et al., 2007). Common methods include natural language processing technology, topic models, ontology, etc. Bigram co-occurrence can prevent information loss and distortion in linguistic evaluation information aggregation and computation and make the calculation results more accurate (Herrera & Martinez, 2000). In particular, semantic association analysis is based on external knowledge and a semantic lexicon to construct a model of feature words, which enables better text classification than that found in prior studies. This approach has been widely developed in research pertaining to business intelligence, medical image association, and human behaviour. For example, Yin and Peng (2010) designed a method that builds semantic associations between product features and sentiment words to identify the sentiment expressed regarding each product feature from product reviews in Chinese. Zhang et al. (2016a, b) presented a CCA-PairLDA feature representation method for similarity computation between medical images with high-level semantics. The image similarity can be calculated based on local feature sets, word frequency histograms, latent topic distributions, and semantic association coefficients. Kuhlmann, Hofmann, and Jacobs (2017) proposed that emotion valence can be regarded as a semantic super-feature in human forced-choice evaluations, and semantic association networks can be constructed to judge the polarity of words.

The semantic web formed by semantic association is a type of presentation of social networks. Therefore, the opinion mining of online travel reviews from the perspective of social network theory is one of the research contents of this study. Social network analysis emphasizes that each individual has ties to other individuals (Wasserman & Faust, 1994), and such ties are a means of identifying potential relationships in the data (Asiedu, 2014). Density, modularity, and network diameter are often regarded as main analysis indicators. Raisi et al. (2017) conducted research on a hyperlink network for the Australian tourism industry using indicators such as the density, modularity, network diameter, and number of groups. They found that the hyperlink network of a destination is extremely sparse. This finding has critical implications for improving the effectiveness of information flow between tourism organizations and enterprises on the Internet. Wehbe used modularity and density to indicate that a company's cultural security requires more connectivity and frequent quality communication (Wehbe, Hattab, & Hamzeh, 2016). These indicators also have a strong explanatory role in mining the potential information of tourism texts (Casanueva, Gallego, & Garcíasánchez, 2014; Chen, Liang, Hong, & Gu, 2015). Using the social network analysis method, Ji, Li, and Chen (2016) revealed that the spatial structure of self-service travel in Yunnan province is characterized by "closely contacting with each other locally, although the overall link is loose". However, there are still some gaps in the research on identifying the potential needs of consumers based on word granularity. Therefore, this study conducts a semantic association analysis of online travel reviews based on the perspective of social network theory to explore the potential needs of tourists and discover the connections among these needs.

3. Method

3.1. Research design

To identify tourists' potential demand from online travel reviews and improve customer satisfaction, we propose a semantic association analysis approach for practical guidelines in fields related to opinion mining, as shown in Fig. 1, which summarizes the process used in this study. In this paper, web crawlers are first used to extract online travel reviews from three major OTA websites in China. Then, the data are pre-processed to build a review dataset using natural language processing (NLP), such as Jieba and NLTK, in Python programming language, including data cleaning and tokenization. Finally, statistical analysis of thematic words, semantic association analysis and visualization are performed on the data.

3.2. Data collection

In previous studies, most articles only retrieved online travel review data from a single platform (Banerjee & Chua, 2016; Cheng, Fu, Sun, Bilgihan, & Okumus, 2019; Fang et al., 2016; Guo et al., 2017; Kim, Park, Yun, & Yun, 2017; Liu et al., 2018; Schuckert et al., 2015b; Ye, Li, & Wang, 2014; Zhang et al., 2016a, b) or even from only a single destination (Lui, Bartosiak, Piccoli, & Sadhya, 2018). To obtain the most representative data in this study, we choose three major OTAs platforms (Ctrip, Tuniu and Tongcheng) as the data source and extract online travel reviews using Python programming language. Ctrip (http://www.ctrip.com/), Tuniu (http://www.tuniu.com) and Tongcheng (http://www.ly.com) are the top three Chinese OTAs and lead new business to customer (B2C) tourism e-commerce websites. As of the end of December 2017, their market shares were 43.6%, 22.7% and 11.1%, respectively. Ctrip, which was founded in 1999, is one of the largest integrated online travel service companies. By the end of May 2018, it had more than 300 million registered users. Tuniu provides more than one million tourism products for consumers, covering self-help, self-driving, cruise, hotels, visas, tickets for scenic spots, company tours, etc. Tongcheng is the most professional online service platform for leisure tourism in China. According to a report from iResearch.cn, in May 2018, the active monthly users of Ctrip, Tuniu and Tongcheng were 68.55 million, 8.3 million and 22.2 million, respectively (iResearch, 2018). The platforms of Ctrip (Lu & Liu, 2016; Qi, Li, Zhu, &

挖掘可分为六类：情绪分析、意见提取、情绪挖掘、隶属分析、情感或情绪分析和评论挖掘（Kim & Park, 2017；Rahimi & Kozak, 2016；Sirgy, 2010；Xu & Mcgehee, 2016）。所有用于观点挖掘的技术可分为两大类：基于词典的方法（Brandes，2001；Daud, Khan和Che, 2017；Telesford, Joyce, Hayasaka, Burdette和Laurienti, 2011）和机器学习（Leclerc和Martin, 2004；Weiler, 2002；Weiler和Walker, 2014）。基于词典的方法依靠情感词典和语言知识方法对文本情感极性进行分类，其中包括基于语料库的方法和基于词典的方法。得益于机器学习算法的机器学习方法可分为三组：有监督学习、半监督学习和无监督学习（García-Pablos等人，2016年；Hemmatian & Sohrabi, 2017年）。该方法通过特征构建技术（如单词袋）从产品评论中提取文本信息，并使用分类方法分析在线评论（Raisi、Baggio、Barratt Pugh和Willson, 2017）。在在线旅行评论环境中，建议的意见挖掘方法在处理非结构化评论文本时具有可接受的准确性和资源节约的显著优势（Bucur，2015）。郭等人（2017年）分析了从16个国家的25670家酒店中提取的266544条在线评论，以基于潜在狄利克莱分析（LDA）数据挖掘方法确定客户服务的关键维度，并发现了19个可控维度，这些维度对于酒店管理与游客的互动至关重要。Ali、Kwak和Kim（2016）提出了一个模糊领域本体（FDO）和支持向量机（SVM）意见挖掘系统来自动对在线评论进行分类。

2. 3. 语义关联分析与SNA

语义关联分析是一种重要的文本分析方法（Alemanmeza、Halaschekwier、Sahoo、Sheth和Arpinar, 2005；Xiang、Tian和Huang, 2007），在大脑对预期单词的反应的研究中首次被提到（Kutas和Hillyard, 1984）。语义关联分析的核心思想是，语义定义为一个句子中两个词的共现，以高频词为节点，考虑到作为节点之间链接的高频短语共现的频率（Alemanmeza等人，2005年；García-Pablos等人，2016年；Schuckert等人，2015a；Xiang等人，2007年）。常用的方法包括自然语言处理技术、主题模型、本体等。Bigram共生可以防止语言评价信息聚合和计算中的信息丢失和失真，使计算结果更准确（Herrera & Martinez, 2000）。特别是，语义关联分析基于外部知识和语义词典来构建特征词模型，这使得文本分类比以前的研究更好。这种方法在商业智能、医学图像关联和人类行为的研究中得到了广泛发展。例如，尹和彭（2010）设计了一种方法，在产品特征和情感词之间建立语义关联，以识别中文产品评论中对每个产品特征表达的情感。Zhang等人（2016a，b）提出了一种用于具有高级语义的医学图像之间相似度计算的CCA PairLDA特征表示方法。图像相似度可以根据局部特征集、词频直方图、潜在主题分布和语义关联系数来计算。Kuhlmann、Hofmann和Jacobs（2017）提出，情感配价可以被视为人类强迫选择评估中的一个语义超级特征，可以构建语义关联网络来判断单词的极性。

语义关联形成的语义网是社会网络的一种表现形式。因此，从社会网络理论的角度对在线旅游评论进行意见挖掘是本研究的研究内容之一。社交网络分析强调每个人都与其他人有联系（Wasserman & Faust, 1994），这种联系是在数据中识别潜在关系的一种手段（Asiedu，2014）。密度、模块化和网络直径通常被视为主要分析指标。Raisi等人（2017年）利用密度、模块化、网络直径和群体数量等指标，对澳大利亚旅游业的超链接网络进行了研究。他们发现目的地的超链接网络非常稀疏。这一发现对提高互联网上旅游组织和企业之间信息流动的有效性具有重要意义。Wehbe使用模块化和密度来表示公司的文化安全需要更多的连通性和频繁的高质量沟通（Wehbe，Hattab和Hamzeh, 2016）。这些指标在挖掘旅游文本的潜在信息方面也有很强的解释作用（Casanueva、Gallego和Garcíasánchez, 2014；Chen、Liang、Hong和Gu, 2015）。纪、李和陈（2016）利用社会网络分析方法揭示，云南省自助旅游的空间结构具有"局部联系紧密，整体联系松散"的特点。然而，基于词粒度识别消费者潜在需求的研究仍存在一些空白。因此，本研究基于社会网络理论的视角对在线旅游评论进行语义关联分析，以探索游客的潜在需求，并发现这些需求之间的联系。

3. 方法

3.1. 研究设计

为了从在线旅游评论中识别游客的潜在需求并提高客户满意度，我们提出了一种语义关联分析方法，用于与意见挖掘相关领域的实用指南，如图1所示，它总结了本研究中使用的过程。本文首次使用网络爬虫从中国三大在线旅行社网站中提取在线旅游评论。然后，使用Python编程语言中的自然语言处理（如Jieba和NLTK）对数据进行预处理，以构建审查数据集，包括数据清理和标记化。最后，对数据进行主题词统计分析、语义关联分析和可视化。

3.2. 数据收集

在之前的研究中，大多数文章仅从单一平台（班纳吉和蔡，2016；郑、傅、孙、比尔吉汉和奥库姆斯，2019；方等人，2016；郭等人，2017；金、朴、云和云，2017；刘等人，2018；舒克特等人，2015b；叶、李和王，2014；张等人，2016a, b）或甚至仅从单一目的地检索在线旅游评论数据（Lui、Bartosiak、Piccoli和Sadhya, 2018年）。为了获得本研究中最具代表性的数据，我们选择三大在线旅行社平台（携程、途牛和桐城）作为数据源，使用Python编程语言提取在线旅行评论。携程(http://www.ctrip.com/), Tuniu (http://www.tuniu.com)以及童诚(http://www.ly.com)是中国前三大在线旅行社，并领导新的B2C旅游电子商务网站。截至2017年12月底，其市场份额分别为43.6%、22.7%和11.1%。携程网成立于1999年，是全球最大的综合在线旅游服务公司之一。截至2018年5月底，其注册用户已超过3亿。途牛为消费者提供100多万种旅游产品，涵盖自助、自驾、邮轮、酒店、签证、景点门票、公司旅游等。桐城是中国最专业的休闲旅游在线服务平台。根据iResearch的报道。2018年5月，携程、途牛和桐城的月活跃用户分别为6855万、830万和2220万（艾瑞咨询，2018）。携程的平台（陆和刘，2016；齐，李，朱&

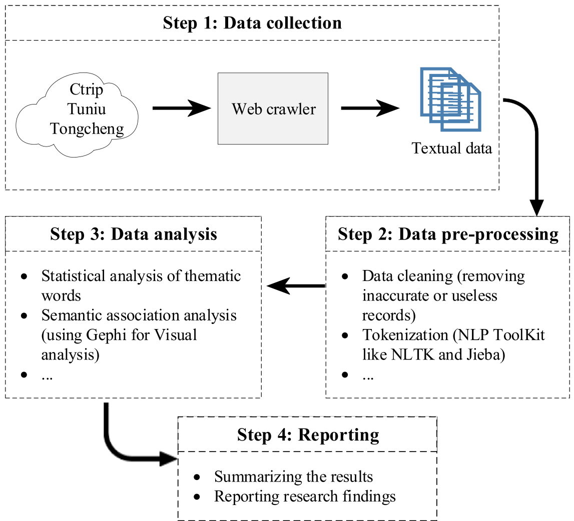


Fig. 1. Framework for sematic association analysis based on online travel reviews.

Shi, 2017), Tuniu (Lian & Yu, 2017; Zhang & Zhou, 2018) and Tongcheng (Li & Yi, 2014; Wang & Wang, 2018) have been used many times as samples in prior studies; therefore, the selection of these data platforms for online travel reviews is reasonable.

The data collection took place in June 2018. We collected review data on the Top ten tourist cities in China released by the *2018 Global Destination Marketing Summit and World Culture and Tourism Forum* (ShaanxiDaily, 2018), including Shanghai, Beijing, Xi'an, Chengdu, Hangzhou, Sanya, Hongkong, Guangzhou, Xiamen and Nanjing. Web crawlers (Xiang et al., 2017) designed in the Python programming language were used to mimic a user's access to the three OTA websites for all the tourist products of the ten cities, and the user ID, product name, review text and review time of the search results were downloaded, as shown in Fig. 2. To maintain the timeliness of the data, we only crawled online review information from January 2015 to June 2018.

3.3. Data pre-processing

All reviews collected from the three OTA platforms were pre-processed by four operations: data cleaning, tokenization, stop word removal and the translation of Chinese words to English. Data cleaning was used to detect and remove inaccurate or useless records from online textual data, such as misspellings and non-target language (Guo et al., 2017; Xiang et al., 2017), leaving only valuable tourism-related information. In this study, we first deleted reviews with a textual length of less than 15 words. The length of the textual portion of the review positively affects "helpfulness" perceptions (Racherla & Friske, 2012). Cai, Xu, and Wu (2014) also found that information value was poor when the length of Chinese language-based online reviews was less than 15 words. Second, because multiple repeated reviews posted by a single user may lead to statistical bias, we kept only one record in a user's duplicate record. Finally, we deleted the reviews that were advertisements to ensure the authenticity and accuracy of the data samples. Table 1 presents the results after the data cleaning of the reviews in Ctrip, Tuniu and Tongcheng. We collected approximately 47,000 reviews from Ctrip, 51,000 reviews from Tuniu, and 67,000 reviews from Tongcheng. Although Tongcheng had the largest number of reviews, Tuniu had the highest number of Chinese words (approx. 4,084,000) and the highest average length of reviews (80.32 words).

The purpose of tokenization is to divide travel review texts into keywords, phrases or other meaningful elements, such as tourist spots and travel feelings (Guo et al., 2017; Xiang et al., 2017). For this study, we applied the existing open source tool LTP (see http://ir.hit.edu.cn/ltp/) provided by the Research Centre for Social Computing and Information Retrieval of Harbin Institute of Technology (Che, Li, & Liu, 2010) and Python 3.6.4 (see https://www.python.org) to implement tokenization and stop word removal for all effective reviews. The stop word list, which consisted of 1893 Chinese words, came from Harbin Institute of Technology and has been widely applied in opinion mining and analytics.

Because the reviews were posted in Chinese, we adopted the following translation method to ensure the accuracy of the translation. First, we divided the translators into two groups, with two persons in each group. The first group was led by the first author of this article, and the second group was led by an English teacher. All participants possessed an excellent level of English and were able to use native English. After dividing the group, we selected the top 2000 words in thematic word tables for each platform for translation. The translation work of each group was performed independently and simultaneously, and no discussion was held until the formal translation was completed. If uncertain thematic words were found in the translation, the participants marked them as "uncertain" thematic words and discussed them in the next step. After completing the first translation, we compared the translation results of the two groups one by one to avoid translation bias, especially with regard to the "uncertain" thematic words.

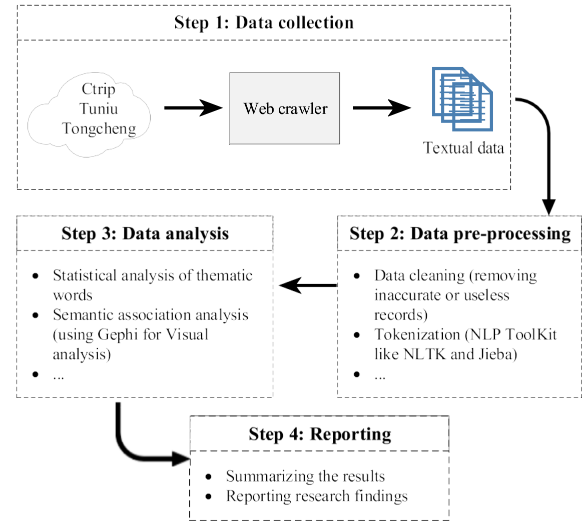


图1. 基于在线旅游评论的语义关联分析框架。

在之前的研究中，图努（Lian & Yu, 2017；Zhang & Zhou, 2018）和童成（Li & Yi, 2014；Wang & Wang, 2018）多次被用作样本；因此，选择这些数据平台进行在线旅游评论是合理的。

数据收集于2018. 6. 进行。我们收集了2018. 全球旅游目的地营销峰会和世界文化与旅游论坛（陕西日报，2018. 发布的中国十大旅游城市的回顾数据，包括上海、北京、西安、成都、杭州、三亚、香港、广州、厦门和南京。使用Python编程语言设计的网络爬虫（Xiang et al., 2017. 模拟用户访问这十个城市所有旅游产品的三个OTA网站，并下载搜索结果的用户ID、产品名称、评论文本和评论时间，如图2. 示。为了保持数据的及时性，我们只收集了2015. 1. 至2018. 6. 的在线审查信息。

3. 3. 数据预处理

从三个OTA平台收集的所有评论都经过了四个操作的预处理：数据清理、标记化、停止词删除和中文词英译。数据清理用于从在线文本数据中检测和删除不准确或无用的记录，例如拼写错误和非目标语言（郭等人，2017；项等人，2017），只留下有价值的旅游相关信息。在这项研究中，我们首先删除了文本长度小于15个单词的评论。评论文本部分的长度对“帮助性”的认知有积极影响（Racherla & Friske, 2012）。蔡、徐和吴（2014）还发现，当基于中文的在线评论长度小于15个单词时，信息价值很低。第二，由于单个用户发布的多次重复评论可能会导致统计偏差，我们在用户的重复记录中只保留了一条记录。最后，我们删除了广告评论，以确保数据样本的真实性和准确性。表1显示了审查数据清理后的结果在携程、途牛和桐城。我们从携程网收集了大约47000条评论，从途牛网收集了51000条评论，从桐城网收集了67000条评论。虽然桐城的评论数量最多，但Tuniu的中文单词数量最多（约4084000个），平均评论长度最高（80.32个单词）。

标记化的目的是将旅游评论文本分为关键词、短语或其他有意义的元素，如旅游景点和旅游感受（郭等人，2017；项等人，2017）。在本研究中，我们应用了现有的开源工具ltp（参见http://ir.hit.edu.cn/ltp/）由哈尔滨工业大学社会计算与信息检索研究中心（Che，Li和Liu, 2010）和python 3.6.4提供（参见https://www.python.org)对所有有效的评论实施标记化并停止删除单词。“停止词列表”由1893个中文单词组成，来自哈尔滨工业大学，已广泛应用于意见挖掘和分析。

由于评论是用中文发布的，我们采用以下翻译方法来确保翻译的准确性。首先，我们将翻译人员分为两组，每组两人。第一组由本文第一作者领导，第二组由英语老师领导。所有参与者都拥有优秀的英语水平，能够使用母语英语。分组后，我们为每个翻译平台选择了主题词表中的前2000个单词。每个小组的翻译工作是独立同时进行的，在正式翻译完成之前没有进行讨论。如果在翻译中发现不确定的主题词，参与者将其标记为“不确定”主题词，并在下一步进行讨论。在完成第一次翻译后，我们逐一比较了两组的翻译结果，以避免翻译偏差，尤其是在“不确定”主题词方面。

Fig. 2. Example of collected reviews.

Table 1
The results after data cleaning.

Review Platform	N of Reviews	N of Chinese Words	Avg. Length of Reviews
Ctrip	47,446	3,428,609	72.26
Tuniu	50,844	4,083,948	80.32
Tongcheng	67,139	3,484,343	51.90

3.4. Data analysis

3.4.1. Analysis of thematic words

A thematic word is a word with a definite meaning and the characteristics of conciseness and timeliness as well a large amount of information. An analysis of thematic words aims to remove words that are meaningless and thereby affect the research results of the keywords extracted from the online review text through data pre-processing. It also aims to calculate the frequency of keywords in the text (Yuan & Wu, 2016). The Jieba toolkit was used for the tokenization and stop word removal of all the review data. A Chinese word segmentation module was that developed by Chinese programmers in Python. This approach was mostly used in the text mining and the Chinese word segmentation in search engines. Additionally, the NLTK was used to calculate the frequency of the thematic words. The calculation formula is as shown in Equation (1):

$$F_i = R_i \times \frac{L_i}{L_t} \tag{1}$$

where F_i is the frequency of a thematic word i in each platform, R_i is the occurrence number of i in the review text, L_i is the length of i, and L_t is the length of all words in each platform. Therefore, in this study, the range of t is from 1 to 3, representing the three OTAs platforms, respectively.

3.4.2. Semantic association analysis

Semantic association is an essential element of human language from which people can infer valuable information in communication (Kim, Karunanayaka, Privitera, Holland, & Szaflarski, 2011). In this study, we first constructed semantic association bigram co-occurrence phrases of thematic words using NTLK and Jieba. For example, an original review appeared as follows:

整体不错。住宿和餐饮挺好的,行程安排也好。就是导游服务能再提高就更好了。

The translation is as follows:

Overall good. The accommodations and dining are fine, and the schedule is also good. It would be even better if the tour guide service could be further improved.

After pre-processing of the review, the text was as follows:

整体/不错/。/住宿/餐饮/挺好/,/行程/安排/也好/。/就是/导游/服务/能/再/提高/就/更好/了/。

The results of the semantic association bigrams co-occurrence of thematic words are shown in Table 2.

Then, we accumulated the frequency of the bigram co-occurrence phrases to generate co-occurrence phrase lists for the three platforms for social network analysis. Gephi, a popular open-source software for graph and network analysis developed by the research institutions of SciencesPo and Linkfluence in France, which is widely applied in the fields of social network analysis, biology, and genomics (Bastian, Heymann, & Jacomy, 2009; Jacomy, Venturini, Heymann, & Bastian, 2014), was used in this study for network structure feature analysis and visualization.

4. Analysis of the results

4.1. Statistical analysis of thematic words

If only simple word frequency analysis is performed on thematic words, the meaning of the context of the words itself will not be explained. For example, Boo and Busser (2018) argued that hotel topics include thematic words such as guestroom (e.g., room amenities, view,

Table 2
A sample of semantic association bigram co-occurrence.

Thematic Word	Thematic Word	Frequency
整体	**住宿**	1
住宿	**餐饮**	1
餐饮	**行程**	1
行程	**安排**	1
安排	**导游**	1
导游	**服务**	1
服务	**提高**	1

图2. 收集评论的例子。

表1
数据清理后的结果。

审查平台	N个评论	汉语词汇的分类	评论的平均长度
Ctrip	47,446	3,428,609	72.26
Tuniu	50,844	4,083,948	80.32
桐城	67,139	3,484,343	51.90

3.4. *数据分析*

3.4.1. *主题词分析*

主题词是一个具有明确意义的词，具有简洁性、及时性和信息量大的特点。主题词分析的目的是通过数据预处理去除无意义的词，从而影响从在线评论文本中提取的关键词的研究结果。它还旨在计算文本中关键词的频率（Yuan & Wu, 2016）。Jieba工具包用于标记化和删除所有审查数据的停止字。使用了由中国程序员用Python开发的中文分词模块。这种方法主要用于搜索引擎中的文本挖掘和中文分词。此外，NLTK被用来计算主题词的频率。计算公式如式（1）所示:

$$F_i = R_i \times \frac{L_i}{L_t} \tag{1}$$

式中，F_i是每个平台中主题词i的出现频率，R_i是评论文本中i的出现次数，L_i是i的长度，L_t是每个平台中所有单词的长度。因此，在本研究中，t的范围为1到3,分别代表三种OTA平台。

3.4.2. *语义关联分析*

语义关联是人类语言的一个基本要素，人们可以从中推断出有价值的信息（Kim、Karunanayaka、Privitera、Holland和Szaflarski, 2011）。在这项研究中，我们首先使用NTLK和界把构建了主题词的语义关联双格共现短语。例如，最初的评论如下:

整体不错。住宿和餐饮挺好的,行程安排也好。就是导游服务能再提高就更好了。

翻译如下:

总的来说不错。住宿和餐饮都很好，日程安排也很好。如果导游服务能进一步改善，那就更好了。

在对审查进行预处理后，案文如下:

整体/不错/。/住宿/餐饮/挺好/,/行程/安排/也好/。/就是/导游/服务/能/再/提高/就/更好/了/。

表2显示了主题词的语义关联双格图共现结果。

然后，我们累积二元共现短语的频率，为三个社交网络分析平台生成共现短语列表。Gephi是一款流行的图形和网络分析开源软件，由法国的SciencesPo和Linkfluence研究机构开发，广泛应用于社会网络分析、生物学和基因组学领域（Bastian、Heymann和Jacomy, 2009；Jacomy、Venturini、Heymann和Bastian, 2014），用于网络结构特征分析和可视化。

4. 结果分析

4.1. *主题词的统计分析*

如果只对主题词进行简单的词频分析，就不会解释词本身的上下文含义。例如，Boo and Busser（2018）认为，酒店主题包括主题词，如客房（如客房设施、景观、，

表2
语义关联二元共现样本。

主题词	主题词	频率
整体	住宿	1
住宿	餐饮	1
餐饮	行程	1
行程	安排	1
安排	导游	1
导游	服务	1
服务	提高	1

comfortable), location (e.g., atmosphere, close airport), and transportation (e.g., shuttle, parking). Tsai, Wang, and Tseng (2015) believed that the physical attractiveness, sense of humour and seniority of tour guides were important factors affecting their interactions with tourists, and the context should be considered in the classification of thematic words. Therefore, the classification statistics of thematic words are significant and allow us to better identify topics and focuses of attention. In prior studies, the topics of online travel reviews were classified into aspects such as hotel location, cleanliness, basic service, value, landmarks and attractions, dining and experiences, core product, satisfaction, and so on (Boo & Busser, 2018; Guo et al., 2017; Xiang et al., 2017).

This paper adopted a method of manual content analysis similar to (Xiang et al., 2017) and (Boo & Busser, 2018). This method of manual content analysis can transform unsystematic and qualitative reviews into systematic and quantitative data and includes quantitative analysis and the factual judgment of text content (Jennings, 2001). In the study of online travel reviews, manual content analysis is necessary to allow the authors to be immersed in the text to better identify topics and improve the rigor and flexibility (Boo & Busser, 2018; Sotiriadou, Brouwers, & Le, 2014), which has a wide range of applications in research on topic classification (Boo & Busser, 2018; Cheng et al., 2019; Lian & Yu, 2017; Luo, Gu, Jie, & Phang, 2017; Xiang et al., 2017).

The process of manual content analysis in this paper was as follows. First, we merged the thematic words of three platforms and selected the top 500. The selected thematic words were categorized by three experts in e-commerce according to the criterion that thematic words with similar attributions were grouped into respective categories by analysing their interrelation and logical order. As a result, we identified five topics: tour guide, hotel, service, scenic spot and experience. In addition, we defined each topic to reduce the understanding bias in the next classification (Cheng et al., 2019). Second, we invited six research assistants to participate in a manual content analysis process. To ensure the validity and reliability of the classification, the assistants were divided into two contrasting groups. They classified the thematic words extracted in the first stage according to the classification rules in step (1). The work of the two groups was performed independently and simultaneously. If there were disputes within the group, they marked them as "uncertain" thematic words and discussed them in the next step. Finally, the classification results of the two groups were compared one by one, and the thematic words that were marked as "uncertain" were again discussed with the experts. At last, five topics were proposed, as shown in Table 3.

The frequency of thematic words was calculated based on formula (1), which is explained in the last section of the data analysis. Note that the five topics represent common characteristics of the three platforms, which appear to be more generic. To save space, Table 3 shows only the top 10 thematic words for each topic. Obviously, "tour guide" (Topic 1) and "hotel" (Topic 2) are the two topics about which users were most concerned, with a particular focus on "guide" (1.22%), "hotel" (0.65%), "itinerary" (0.63%), and "scheduling" (0.55%).

Table 4
Results of high-frequency word threshold values.

Review Platform	N of Thematic words	I_1	T	N of High-Frequency Words
Ctrip	21,420	10,189	141.75	473
Tuniu	24,089	11,422	150.14	530
Tongcheng	26,924	12,454	156.82	596

To accurately understand the difference in the distribution of the five topics on the three review platforms, this paper conducted a statistical analysis of high-frequency words. We calculated the threshold value of high-frequency words based on the demarcation method of high-frequency words and low-frequency words proposed by Donohue (1973). The calculation formula is shown in Equation (2):

$$T = -1 + \frac{\sqrt{1 + 8I_1}}{2} \tag{2}$$

where T is the threshold value of high-frequency words, which means that the lowest value of frequency in thematic words is calculated by this formula (Donohue, 1973), that is, the critical value between high-frequency words and low-frequency words. I_1 is the number of thematic words with frequencies equal to 1. For example, we first extracted 21420 thematic words from Ctrip, in which the number of thematic words with a frequency of 1 was 10,189 (that is, $I_1 = 10189$). Table 4 shows the calculation results of the high-frequency word threshold value for the three platforms. The threshold values of Ctrip, Tuniu and Tongcheng are 141.75, 150.14 and 156.82, respectively, and the corresponding high-frequency words are 473, 530 and 596. Considering that the inconsistency of the number of high-frequency words may affect the analysis results, this paper appropriately extends the number of high-frequency words and selects the top 700 thematic words to show the distribution of the 5 topics. Currently, the frequencies of words are 85, 105 and 129, respectively.

Fig. 3 shows the relative differences of five topics on the three OTAs. The X-axis represents the percentage of high-frequency words in five topics on each platform. A higher percentage indicates more attention paid to the topic. The Y-axis represents the five topics. In general, the distribution of the five topics is quite different among the three platforms; the percentages of "tour guide", "hotel" and "experience" are higher, and those for "service" and "scenic spot" are relatively low. The distribution of Topic 3 (i.e., service) and Topic 5 (i.e., experience) is almost equal among the three platforms, suggesting that tourists have the same level of concern about them in online reviews. For all platforms, users are more concerned about their own tourism experience and service (Chang, Kivela, & Mak, 2011). However, the manifestations of the other three topics are quite different. Topic 1 (i.e., tour guide) appeared to be more prominent in Ctrip (37.8%) and Tuniu (31.3%) than Tongcheng (8.3%), which shows that tourists are more concerned about the itinerary, scheduling and explanations made by the guide. Song and Wang (2013) and Tsai et al. (2015) argued that guides' service

Table 3
Topic identification of thematic words.

Topic 1: Tour guide		Topic 2: Hotel		Topic 3: Service		Topic 4: Scenic Spot		Topic 5: Experience	
Guide	1.22%	Hotel	0.65%	Enthusiasm	0.16%	Ticket	0.43%	Satisfaction	0.43%
Itinerary	0.63%	Accommodation	0.19%	Enjoy	0.14%	Spot	0.35%	Happy	0.35%
Scheduling	0.55%	Overall	0.17%	Driver	0.14%	Location	0.22%	Feeling	0.27%
Explanation	0.36%	Environment	0.12%	Recommendation	0.13%	Queue	0.21%	Kids	0.21%
Shopping	0.14%	Room	0.11%	Gratitude	0.12%	Entertainment	0.17%	Time	0.16%
Always	0.12%	Breakfast	0.09%	Thoughtful	0.09%	Attraction	0.15%	Play	0.14%
Responsible	0.10%	Big	0.09%	Consultation	0.09%	View	0.13%	Travel	0.14%
Patience	0.10%	Clean	0.07%	Free	0.09%	Weather	0.11%	Suggest	0.12%
Humour	0.08%	Comfortable	0.06%	Id Card	0.13%	Fun	0.11%	Pleasant	0.11%
Considerate	0.07%	Aircraft	0.06%	Care For	0.08%	Scenery	0.11%	Whole	0.10%

舒适）、位置（如大气、机场附近）和交通（如班车、停车场）。蔡、王和曾（2015）认为，导游的外表吸引力、幽默感和资历是影响他们与游客互动的重要因素，在主题词的分类中应考虑上下文。因此，主题词的分类统计具有重要意义，可以让我们更好地识别主题和关注焦点。在之前的研究中，在线旅游评论的主题被分为酒店位置、清洁度、基本服务、价值、地标和景点、餐饮和体验、核心产品、满意度等方面（Boo & Busser, 2018；郭等人，2017；项等人，2017）。

本文采用了类似于（Xiang等人，2017）和（Boo & Busser, 2018）的手动内容分析方法。这种手动内容分析方法可以将非系统性和定性的评论转化为系统性和定量的数据，包括定量分析和对文本内容的事实判断（Jennings，2001）。在研究中

在在线旅游评论中，手动内容分析是必要的
作者应该沉浸在文本中，以便更好地确定主题和主题
提高严谨性和灵活性（Boo & Busser, 2018；Sotiriadou, Brouwers和Le, 2014），这在主题分类研究中有广泛的应用（Boo & Busser, 2018；Cheng等人，2019；Lian & Yu, 2017；Roo，Gu, Jie和Phang, 2017；Xiang等人，2017）。

本文中的手动内容分析过程如下。首先，我们合并了三个平台的主题词，并选择了前500. 。三位电子商务专家根据相似属性的主题词通过分析它们之间的相互关系和逻辑顺序被分为不同的类别的标准对所选主题词进行了分类。因此，我们确定了五个主题：导游、酒店、服务、景点和体验。此外，我们定义了每个主题，以减少下一个分类中的理解偏差（Cheng等人，2019. ）。其次，我们邀请了六名研究助理参与手动内容分析过程。为了确保分类的有效性和可靠性，助理被分为两个对照组。他们根据步骤（1. 中的分类规则对第一阶段提取的主题词进行分类。两组的工作是同时独立进行的。如果小组内部存在争议，他们会将其标记为"不确定"主题词，并在下一步进行讨论。最后，将两组的分类结果逐一进行比较，并再次与专家讨论标记为"不确定"的主题词。最后，提出了五个主题，如表3. 示。

主题词的频率是根据公式（1）计算出来的，这在数据分析的最后一部分进行了解释。请注意，这五个主题代表了这三个平台的共同特征，它们似乎更通用。为了节省空间，表3只列出了每个主题的前10个主题词。显然，"导游"（主题1）和"酒店"（主题2）是用户最关心的两个主题，尤其是"导游"（1.22%）、"酒店"（0.65%）、行程（0.63%）和"日程安排"（0.55%）。

表4
高频词阈值的结果。

审查平台	主题词的分类	I_1	T	高频词的分类
Ctrip	21,420	10,189	141.75	473
Tuniu	24,089	11,422	150.14	530
桐城	26,924	12,454	156.82	596

为了准确理解这五个主题在三个评论平台上的分布差异，本文对高频词进行了统计分析。我们根据Donohue（1973）提出的高频词和低频词的划分方法，计算了高频词的阈值。计算公式如式（2）所示:

$$T = -1 + \frac{\sqrt{1 + 8I_1}}{2} \tag{2}$$

其中T是高频词的阈值，这意味着主题词中频率的最低值是通过这个公式计算的（Donohue，1973. ，即高频词和低频词之间的临界值。I_1是频率等于1. 主题词的数量。例如，我们首先从携程提取了21420. 主题词，其中频率为1. 主题词数量为10189. 即I[1. = 10189. 。表4. 示了三种平台的高频词阈值的计算结果。携程、途牛和桐城的阈值分别为141. 75. 150. 14. 156. 82. 对应的高频词分别为473. 530. 596. 考虑到高频词数量的不一致可能会影响分析结果，本文对高频词的数量进行了适当的扩展，并选取了前700. 主题词来展示这5. 主题的分布。目前，单词的频率分别为85. 105. 129.

图3显示了三个OTA上五个主题的相对差异。X轴代表每个平台上五个主题中高频词的百分比。百分比越高，表示对该主题的关注度越高。y轴代表五个主题。总的来说，这五个主题在三个平台上的分布是非常不同的；"导游"、"酒店"和"体验"的比例较高，"服务"和"景点"的比例相对较低。在三个平台中，主题3（即服务）和主题5（即体验）的分布几乎相等，这表明游客在在线评论中对它们的关注程度相同。对于所有平台，用户更关心自己的旅游体验和服务（Chang、Kivela和Mak, 2011）。然而，其他三个主题的表现形式却截然不同。主题1（即导游）在携程（37.8%）和途牛（31.3%）比桐城（8.3%）更突出，这表明游客更关心导游的行程安排、日程安排和解释。宋和王（2013）和蔡等人（2015）认为导游服务

表3
主题词的主题识别。

主题1：导游		话题2：酒店		话题3：服务		主题4：风景区		话题5：经验	
导游	1.22%	酒店	0.65%	热情	0.16%	票	0.43%	满足感	0.43%
行程	0.63%	住处	0.19%	享受	0.14%	斑点	0.35%	快乐的	0.35%
行程安排	0.55%	总体而言	0.17%	驱动程序	0.14%	位置	0.22%	感觉	0.27%
解释	0.36%	环境	0.12%	推荐	0.13%	队列	0.21%	孩子们	0.21%
购物	0.14%	房间	0.11%	感激	0.12%	娱乐	0.17%	时间	0.16%
总是	0.12%	早餐	0.09%	周到的	0.09%	吸引力	0.15%	玩	0.14%
负责任的	0.10%	大的	0.09%	咨询	0.09%	视图	0.13%	长途行走	0.14%
耐心	0.10%	清洁的	0.07%	免费	0.09%	天气	0.11%	建议	0.12%
幽默	0.08%	使人舒服的	0.06%	身份证	0.13%	享乐	0.11%	令人愉快的	0.11%
考虑周到的	0.07%	飞机	0.06%	照顾	0.08%	风景	0.11%	整体	0.10%

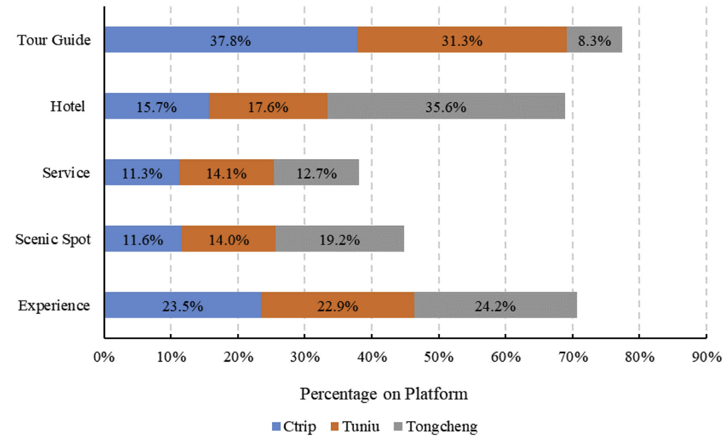


Fig. 3. Distribution of five topics on the three OTAs.

level (such as responsibility, professionalism and humour) was an important factor affecting tourist satisfaction. Compared to Ctrip (15.7%) and Tuniu (17.6%), Topic 2 (i.e., hotel) was much more prominent in Tongcheng (35.6%), which may be related to its strategy of leisure travel. The leisure travellers were solo groups, and they were more sensitive to the environment around the hotel (Kim & Park, 2017; Mccain, Jang, & Hu, 2005; Radojevic, Stanisic, & Stanic, 2015). For Topic 4 (i.e., scenic spot), Ctrip (11.6%) was lower compared to Tuniu (14.0%) and Tongcheng (19.2%).

4.2. Constructing bigram co-occurrence of semantic associations

To explore the relationship between thematic words, we constructed a bigram co-occurrence of semantic association using bigram tools in the NLTK of the Chinese corpus (Bird, Klein, & Loper, 2009). Because the bigram phrases generated in this paper were undirected data, we added the frequency of the two bigram phrases that contained thematic words repeatedly in different locations, such as A-B and B-A. Consequently, we obtained more than 200 thousand bigram co-occurrence phrases from each platform. According to Equation (2) formulated in the above statistical analysis of thematic words, the value of I_1 is over 150 thousand, and the value of T is more than 550; therefore, the number of phrases that can be used is not more than 100. Therefore, to better reveal the relationship between thematic words, we chose the top 2000 bigram co-occurrence phrases for semantic association analysis. The value of co-occurrence frequency on the three platforms is approximately 10, covering most high-frequency words.

Table 5 presents the top 10 bigram co-occurrence phrases of each OTA. The weight represents the accumulated value of the co-occurrence of two thematic words. As shown in Table 5, the top 10 bigram co-occurrence phrases on Ctrip and Tuniu are very similar, and *itinerary–scheduling*, *guide–explanation*, *guide–satisfaction*, *guide–scheduling* are most frequently mentioned by users, meaning that the travel itinerary and the service quality of the guide were extremely important factors for the users of Ctrip and Tuniu. The overall service quality of tour guides or leaders is very important to the satisfaction of tourists and influence consumers' decisions in the selection of all-inclusive tours (Caber & Albayrak, 2016; Heung, 2008; Mossberg, 1995). However, the top 10 bigram co-occurrence phrases shown on Tongcheng are obviously different compared to Ctrip and Tuniu, and the hotel-related topics are the focus of attention, such as *room–clean*, *hotel–location*, *hotel–room*, *hotel–environment*. To some extent, this result further verifies the distribution of the five topics mentioned above.

4.3. Structural properties of the semantic association network

To analyse the network structural properties of each platform, we imported the top 2000 bigram co-occurrence phrases into Gephi in a CSV file format to construct an undirected network. The results of the social network analysis show that the three networks were non-fully connected networks, and there were a certain number of isolated nodes with no links to others. Isolated nodes were excluded from our analysis (Raisi et al., 2017), including the visualizations in the next section, unless otherwise stated. The isolated nodes of Ctrip, Tuniu, and Tongcheng were 18, 46, and 298, respectively.

Table 6 shows the results of the measurement in this paper, including nodes, density, average degree, etc. The three networks are obviously divided into two categories: Ctrip and Tuniu are quite similar, and Tongcheng is quite different. There are 1,462 nodes in the Tongcheng network, which is the largest of the three platforms; Ctrip and Tuniu have almost the same number of nodes. The density of the three networks is less than 0.02, which indicates that the networks are extremely sparse. In particular, the network density of Tongcheng is only 0.002, which means that out of 1,000 possible links, only 2 of them actually exist in the network. Compared to those of Ctrip and Tuniu, the average degree of Tongcheng is the smallest at 2.488. The three platforms, especially Tongcheng, are low-density networks characterized by more scattered topics, which may reduce the impact of reviews on potential tourists. After excluding the isolated nodes, all the bigram co-occurrence phrases of the three platforms were connected together in one component. The network diameter refers to the maximum distance between any two nodes in the network, representing the closeness of connection among the nodes. The diameter of Tongcheng is 13, which is far greater than the values for Ctrip and Tuniu, meaning that information is passed from one node to another 13 times at most in the Tongcheng network.

Modularity is one of the most common methods for community detection and can help to illuminate the intermediate structure of the network (Raisi et al., 2017). The value of the modularity index ranges from 0 to 1. It is generally acknowledged that when the index of

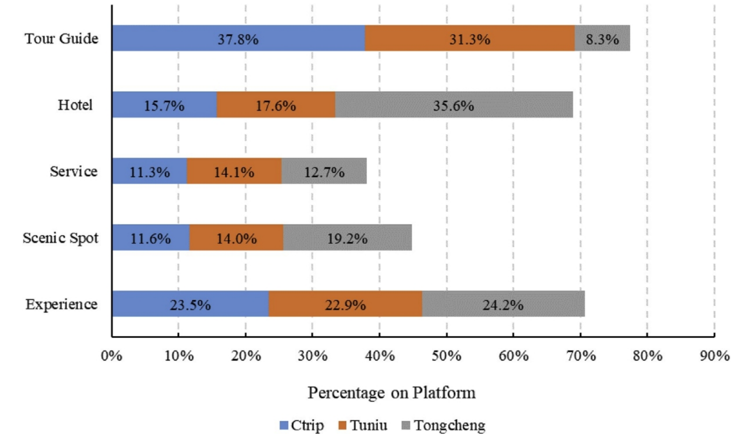


图3. 三个在线旅行社五个主题的分布。

水平（如责任感、专业性和幽默感）是影响游客满意度的重要因素。与携程（15.7%）和途牛（17.6%）相比，主题2（即酒店）在桐城（35.6%）更为突出，这可能与其休闲旅游战略有关。休闲旅行者是单独的群体，他们对酒店周围的环境更敏感（Kim & Park, 2017；Mccain, Jang和Hu, 2005；Radojevic, Stanisic和Stanic, 2015）。在主题4（即景点）方面，携程（11.6%）低于途牛（14.0%）和桐城（19.2%）。

4.2. 构建语义联想的二元共现

为了探索主题词之间的关系，我们在中文语料库的NLTK（Bird、Klein和Loper, 2009. 中使用二元图工具构建了语义关联的二元图共现。由于本文中生成的二元短语是无向数据，我们添加了在不同位置重复包含主题词的两个二元短语的频率，例如A-B和B-A。因此，我们从每个平台获得了20. 万个二元短语共现。根据上述主题词统计分析公式（2. ，I_1的值超过15. ，T的值超过550. 因此，可以使用的短语数量不超过100. 。因此，为了更好地揭示主题词之间的关系，我们选择了前2000. 二元共现短语进行语义关联分析。三个平台上的共现频率值约为10. 涵盖了大多数高频词。

表5列出了每个OTA的前10个双字共现短语。权重代表两个主题词共现的累积值。如表5所示，携程和途牛的前10个bigram共现短语非常相似，用户最常提到的是Trainer-y-日程安排、guide–解释、guide–满意度、guide–日程安排，这意味着旅行路线和导游的服务质量对携程和途牛的用户来说是极其重要的因素。导游或领队的整体服务质量对游客的满意度非常重要，并影响消费者在选择全包旅游时的决策（Caber & Albayrak, 2016；Heung, 2008；Mossberg, 1995）。然而，与携程和途牛以及酒店相关的网站相比，桐城网站上显示的前10个双关语共现短语明显不同

主题是关注的焦点，如房间–清洁、酒店–位置、酒店–房间、酒店–环境。这一结果在一定程度上进一步验证了上述五个主题的分布。

4.3. 语义关联网络的结构性质

为了分析每个平台的网络结构属性，我们以CSV文件格式将前2000个bigram共现短语导入Gephi，以构建无向网络。社交网络分析的结果表明，这三个网络是非完全连接的网络，并且存在一定数量的孤立节点，与其他节点没有链接。我们的分析中排除了孤立的节点（Raisi等人，2017），包括下一节中的可视化，除非另有说明。携程、途牛和桐城的孤立节点分别为18个、46个和298个。

表6. 示了本文的测量结果，包括节点、密度、平均度等。三个网络明显分为两类：携程和途牛非常相似，而桐城非常不同。桐城网络有1462. 节点，是三个平台中最大的一个；携程和途牛的节点数量几乎相同。三个网络的密度小于0.02. 这表明网络非常稀疏。尤其是桐城的网络密度仅为0.002. 这意味着在1000. 可能的链路中，只有2. 链路实际存在于网络中。与携程和途牛相比，桐城的平均学位最低，为2. 488. 这三个平台，尤其是桐城，是低密度的网络，主题更加分散，这可能会减少评论对潜在游客的影响。在排除孤立节点后，三个平台的所有二元共现短语被连接在一个组件中。网络直径是指网络中任意两个节点之间的最大距离，表示节点之间连接的紧密程度。桐城的直径为13. 远大于携程和途牛的直径，这意味着桐城网络中的信息从一个节点传递到另一个节点最多13. 。

模块化是社区检测最常用的方法之一，有助于阐明网络的中间结构（Raisi等人，2017. ）。模块化索引的值在0到1. 间。人们普遍认为

Table 5
Bigram co-occurrence phrases on the three platforms.

Ctrip		Tuniu		Tongcheng	
Bigram phrase	Weight	Bigram phrase	Weight	Bigram phrase	Weight
itinerary–scheduling	7001	itinerary–scheduling	6032	room–clean	3226
guide–explanation	3917	guide–explanation	2221	hotel–location	1927
guide–satisfaction	2370	guide–satisfaction	2061	hotel–room	1911
guide–scheduling	2278	guide–scheduling	1814	hotel–environment	1686
guide–enthusiasm	2098	guide–itinerary	1790	traffic–convenience	1561
guide–itinerary	1824	guide–enthusiasm	1454	hotel–clean	1457
itinerary–satisfaction	1553	guide–responsible	1280	clean–tidy	1245
explanation–thoughtful	1458	itinerary–satisfaction	1275	hotel–worthy	1239
guide–happy	1391	guide–happy	1020	room–facility	1234
guide–responsible	1339	hotel–environment	981	check in–hotel	1155

Table 6
Network statistics for bigram co-occurrence phrases.

Index	Ctrip	Tuniu	Tongcheng
Nodes	428	452	1462
Edges	1991	1976	1819
Density	0.017	0.015	0.002
Avg. degree	9.304	8.743	2.488
No. of connected components	1	1	1
Diameter	7	7	13
Modularity	0.289	0.329	0.514
No. of communities	18	13	53
Avg. path length	2.688	2.674	4.233
Avg. clustering coefficient	0.647	0.659	0.113

modularity is greater than 0.44 (Liu & Du, 2017), the independence of the network community will be relatively high. Using the algorithm provided by Blondel, Guillaume, Lambiotte, and Lefebvre (2008), we obtained the modularity and number of communities of the three platforms. As the results in Table 6 indicate, the modularity indexes of Ctrip and Tuniu are relatively low. It is worth noting that Tongcheng has a modularity index of 0.518 and a community number of 53. These results indicate that the topics within each community are relatively concentrated; however, the relationship among the communities is relatively loose.

In addition, the three platforms show meaningful phenomena in the properties of small-world networks, which are characterized by highly clustered and small characteristic path lengths (Watts & Strogatz, 1998). Telesford et al. (2011) proposed that a small-world network can be identified by a clustering coefficient and average path length. Using the algorithm of Brandes (2001) and Latapy (2008), we obtain the average path lengths of Ctrip, Tuniu and Tongcheng as 2.688, 2.674, and 4.233, with average clustering coefficients of 0.647, 0.659, and 0.113, respectively. According to the research of (Raisi et al., 2017), Ctrip and Tuniu show discernible but not excessive small-world network properties. However, Tongcheng, which is characterized by low clustering and large characteristic path lengths, does not have small-world network properties, resulting in a lower efficiency of information dissemination than the previous two networks.

4.4. Semantic association network visualization

Visualization is the transformation of text data into a network graph that shows the links among thematic words in the form of nodes and lines (Zhao, Gao, Guo, & Tao, 2009). To make the semantic relationship of thematic words more intuitive, Gephi is used for the network visualization in this paper. We use the layout algorithm of ForceAltas2 to draw a network graph. Compared to other layout algorithms, ForceAltas2 has a better measured quality (Jacomy et al., 2014). The balanced state graphs of Ctrip, Tuniu and Tongcheng are shown in Fig. 4, Fig. 5 and Fig. 6, respectively. The nodes represent the corresponding thematic words. Larger nodes indicate greater concern with the thematic words. The links represent the association between thematic words, and the thickness of the connection lines represents the strength of the relationship.

As shown in Fig. 4, the core nodes in the Ctrip network can be divided into two major communities: guides and hotels. The hotel community is relatively small in scale, and the main association phrases include *hotel–breakfast, hotel–environment*, *hotel–room, hotel–comfortable, hotel–clean, hotel–location,* and *hotel–shuttle*, which represent the main factors that tourists consider when choosing a hotel. The internal relationship within the guide community is quite close and includes six core nodes: guide, hotel, satisfaction, itinerary, scheduling, and explanation. The internal links of the guide community are close, including six core nodes: tour guide, satisfaction, travel, schedule, arrangement and explanation, among which *scheduling–itinerary, guide–explanation, guide–satisfaction, guide–itinerary, guide–scheduling,* and *guide-happy* have strong associations.

The guide community can be further divided into five sub-communities: guide, itinerary, satisfaction, explanation, and happiness. The guide-centred sub-community is mainly concerned with the personal characteristics of the guide, who is described by thematic words such as "expectation," "handsome," "hard," and "excellent." The guide's organization is the focus of the itinerary-centred sub-community, including thematic words such as "scheduling," "spot," "play," "route," and "tight." The third sub-community, which is centred on satisfaction, represents the effect of guide services, such as "enthusiasm," "overall," "gratitude," and "considerate." The sub-community that is centred on explanation refers to the quality of the evaluation of guides at scenic spots and comprises core thematic words such as "knowledge," "humour," "patience," "fun," "professional," and "history." The happiness-centred sub-community reflects travel experiences with words such as "pleasant," "feeling," "thoughtful," "family," "parents," "kids," and "friend."

Apart from slight differences in link strength, the manifestations of Figs. 4 and 5 are almost the same, indicating that users of Ctrip and Tuniu have no essential differences and that they have common concerns. However, compared to Ctrip and Tuniu, Fig. 6 has an entirely different manifestation, which shows the visualization of Tongcheng in a preview ratio of 80%. As shown in Fig. 6, the hotel is the most important node. Contrary to Figs. 4 and 5, the hotel-centred community occupies more than half of the whole network, with high-intensity association phrases: *hotel–room, hotel–location, hotel–reception, hotel–satisfaction, hotel–environment, hotel–breakfast, room–clean,* etc. However, despite the larger scale of the hotel community, there is no clear distinction between its sub-communities, which can be roughly divided into hotels and rooms. The distribution of other nodes is very scattered and includes words such as time, queue, entertainment, ticket, performance, kids, happy, animal, etc., and the relationship between each word is quite loose, which indicates that the topics of online travel reviews are not clear and that consumers have no common concerns. The emergence of this phenomenon may be related to the platform strategy and user characteristics of Tongcheng, which make it difficult

表5
三个平台上的双格词共现短语。

Ctrip		Tuniu		桐城	
双字短语	重量	双字短语	重量	双字短语	重量
itinerary–scheduling	7001	itinerary–scheduling	6032	room–clean	3226
guide–explanation	3917	guide–explanation	2221	hotel–location	1927
guide–satisfaction	2370	guide–satisfaction	2061	hotel–room	1911
guide–scheduling	2278	guide–scheduling	1814	hotel–environment	1686
guide–enthusiasm	2098	guide–itinerary	1790	traffic–convenience	1561
guide–itinerary	1824	guide–enthusiasm	1454	hotel–clean	1457
itinerary–satisfaction	1553	guide–responsible	1280	clean–tidy	1245
explanation–thoughtful	1458	itinerary–satisfaction	1275	hotel–worthy	1239
guide–happy	1391	guide–happy	1020	room–facility	1234
guide–responsible	1339	hotel–environment	981	入住–酒店	1155

表6
二元共现短语的网络统计。

索引	Ctrip	Tuniu	桐城
节点	428	452	1462
边缘	1991	1976	1819
密度	0.017	0.015	0.002
平均学位	9.304	8.743	2.488
连接组件的数量	1	1	1
直径	7	7	13
模块化	0.289	0.329	0.514
社区数目	18	13	53
平均路径长度	2.688	2.674	4.233
平均聚类系数	0.647	0.659	0.113

模块化大于0.44. Liu & Du, 2017. ，网络社区的独立性将相对较高。使用Blondel、Guillaume、Lambiotte和Lefebvre（2008. 提供的算法，我们获得了三个平台的模块化程度和社区数量。如表6. 示，携程和途牛的模块化指数相对较低。值得注意的是，桐城的模块化指数为0.518. 社区数量为53. 这些结果表明，每个社区内的主题相对集中；然而，社区之间的关系相对松散。

此外，这三个平台在小世界网络的特性中表现出了有意义的现象，这些网络的特点是高度聚集且特征路径长度较小（Watts & Strogatz, 1998）。Telesford等人（2011年）提出，可以通过聚类系数和平均路径长度来识别小世界网络。利用Brandes（2001）和Latapy（2008）的算法，我们得出携程、途牛和桐城的平均路径长度分别为2.688、2.674和4.233,平均聚类系数分别为0.647、0.659和0.113。根据（Raisi等人，2017年）的研究，携程和途牛的小世界网络特性明显，但并不过度。然而，桐城的特点是低聚类和大特征路径长度，不具有小世界网络特性，导致信息传播效率低于前两个网络。

4. 4. 语义关联网络可视化

可视化是将文本数据转换成网络图，以节点和线条的形式显示主题词之间的联系（赵、高、郭和陶，2009）。为了使主题词之间的语义关系更加直观，本文将Gephi应用于网络可视化。我们使用ForceAltas2的布局算法绘制网络图。与其他布局算法相比，For- ceAltas2具有更好的测量质量（Jacomy等人，2014）。携程、途牛和桐城的平衡状态图分别如图4、图5和图6所示。节点代表相应的主题词。节点越大，表示对节点的关注度越高主题词。链接代表了主题词之间的关联，而连接线的厚度代表了这种关系的强度。

如图4所示，携程网的核心节点可分为两大社区：导游和酒店。酒店的社区规模相对较小，主要的关联短语包括酒店-早餐，酒店-环境，酒店房间，酒店-舒适，酒店-清洁，酒店-位置，和酒店穿梭胡，这是游客选择酒店时考虑的主要因素。导游社区内部的关系非常密切，包括六个核心节点：导游、酒店、满意度、行程、日程安排和解释。导游社区的内部联系非常紧密，包括六个核心节点：导游、满意度、旅行、日程安排、安排和解释，其中日程安排-行程安排、导游-解释、导游-满意度、导游-行程安排、导游-日程安排和导游快乐有很强的关联性。

导游社区可进一步分为五个子社区：导游、行程、满意度、解释和幸福。以导游为中心的亚社区主要关注导游的个人特征，他们用"期望"、"英俊"、"努力"和"优秀"等主题词来描述导游的组织是以行程为中心的次社区的重点，包括"日程安排"、"现场"、"玩耍"、"路线"和"紧密"等主题词第三个亚社区以满意度为中心，代表了导游服务的效果，如"热情"、"全面"、"感激"和"体贴"以讲解为中心的子社区指的是对景点导游的评估质量，包括核心主题词，如"知识"、"胡摩"、"耐心"、"乐趣"、"专业"和"历史"以快乐为中心的亚社区用"愉快"、"感觉"、"体贴"、"家庭"、"父母"、"孩子"和"朋友"等词来反映旅行体验

除了链接强度的细微差异外，无花果的表现。4和5几乎相同，表明携程和途牛的用户没有本质区别，他们有共同的担忧。然而，与携程和途牛相比，图6有一个完全不同的表现形式，它显示了桐城在预览率为80%的可视化。如图6所示，酒店是最重要的节点。与无花果相反。4和5,以酒店为中心的社区占据了整个网络的一半以上，具有高强度的关联短语：酒店-房间、酒店-位置、酒店-接待、酒店-服务、酒店-环境、酒店-早餐、房间-清洁等。然而，尽管酒店社区规模较大，它的子社区之间没有明确的区别，可以大致分为酒店和客房。其他节点的分布非常分散，包括时间、队列、娱乐、门票、表演、孩子、快乐、动物等词，每个词之间的关系非常松散，这表明在线旅游评论的主题不明确，消费者没有共同的担忧。这一现象的出现可能与桐城的平台策略和用户特点有关，这使得通城的运营变得困难

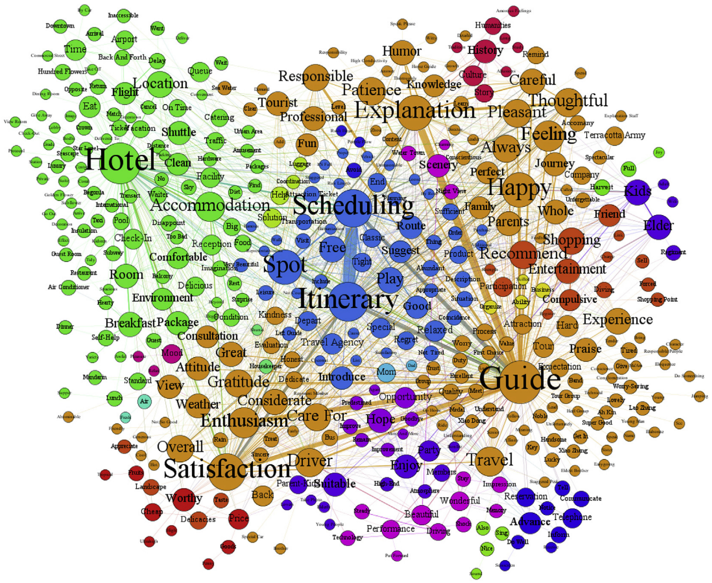


Fig. 4. Network visualization for Ctrip.

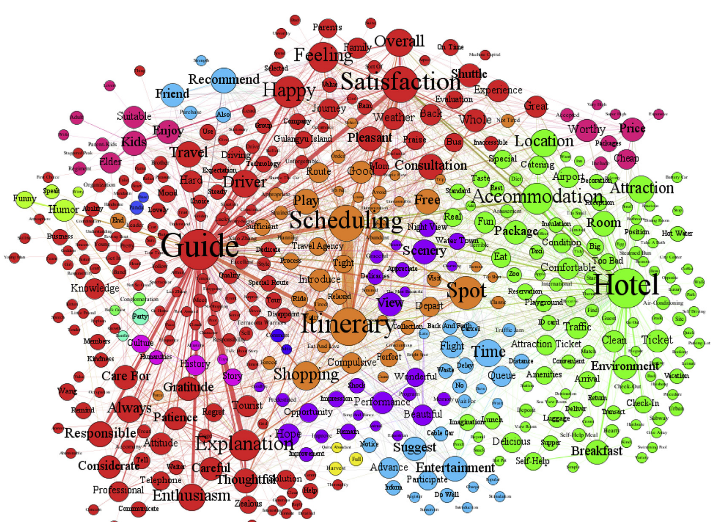


Fig. 5. Network visualization for Tuniu.

Z. Hou, et al. Tourism Management 74 (2019) 276–289

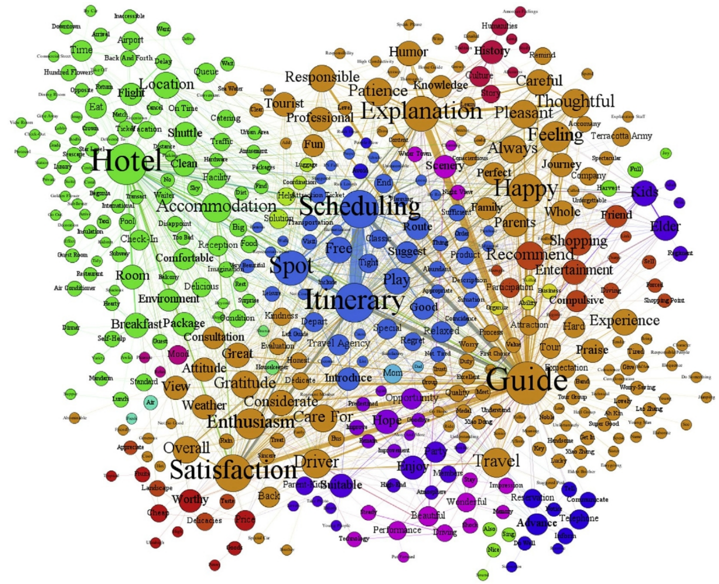


图4. 携程的网络可视化。

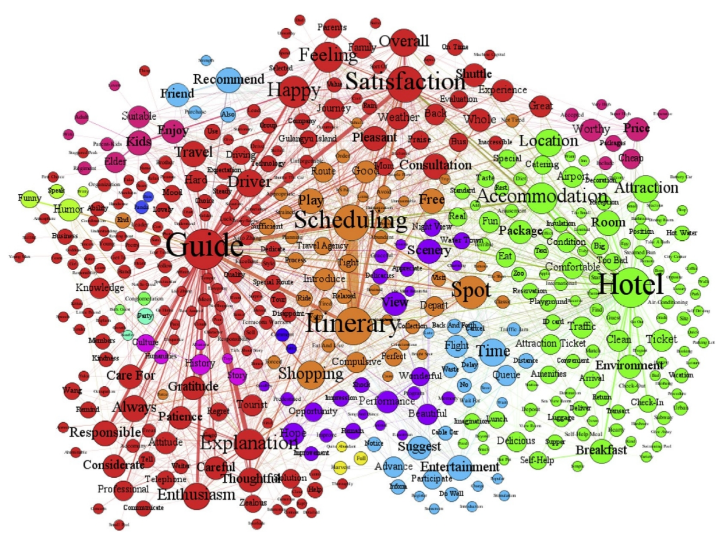


图5. Tuniu的网络可视化。

284

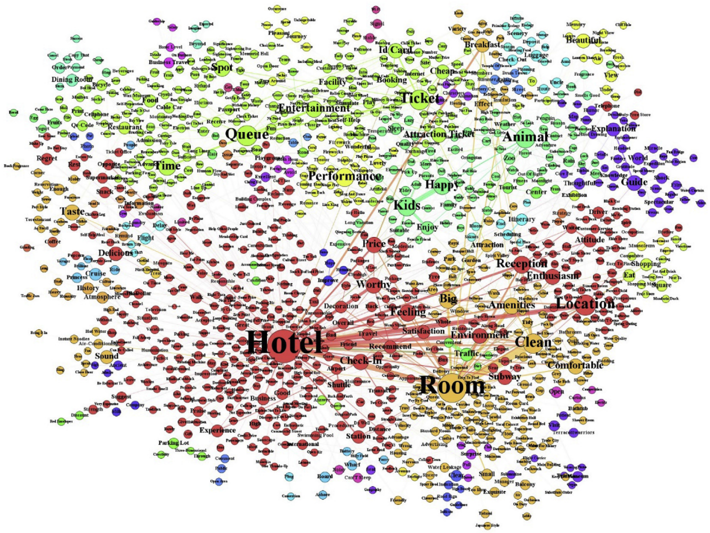


Fig. 6. Network visualization for Tongcheng (preview ratio of 80%).

for consumers to obtain valuable information from online travel reviews.

5. Discussion and conclusion

5.1. Main findings

First, this study found that there are differences in the core concerns of users on the three platforms studied. The users of Ctrip and Tuniu focus on guides and experience, while Tongcheng users are more concerned about hotels and experience. The results of the classification statistical analysis of thematic words show that the users of Ctrip and Tuniu pay 37.8% and 31.3% attention to guides, respectively, far exceeding Tongcheng's 8.3%. This finding shows that the service level of tour guides has a great influence on the satisfaction of tourists on the two platforms, which supports the results of recent research (Tsai et al., 2015). However, Tongcheng users pay more attention to hotel-related information such as the environment and facilities (Guo et al., 2017; Xiang et al., 2017), with attention reaching 35.6%. We argue that the thematic tendencies discussed by users reflect the core competitiveness and marketing strategy of each platform. For example, Ctrip and Tuniu employ mature tourism product development and design systems, which can provide personalized travel itineraries to tourists. In personalized tourism, the service of tour guides is key to ensuring tourists' satisfaction. Leisure tourism is the main business in Tongcheng, and leisure travellers are more sensitive to hotels and their environment (Kim & Park, 2017). It is not surprising that such users are mainly concerned about hotel information.

Second, we obtained meaningful discoveries on network structure properties. On the one hand, we found that the modularity indexes of Ctrip and Tuniu are very low, 0.289 and 0.329, respectively, indicating that the network community is scattered. It is noteworthy that the modularity index value of Tongcheng exceeds the threshold of 0.44 but the number of modular groups is large, indicating that although the topics within the Tongcheng group are relatively centralized, communication among the groups is very loose. This loose layout can lead to low connectivity among networks and a low degree of association among groups. This may be due to low user acceptance of a tourism product, which affects the quality of the reviews they post (Chatterjee, 2001). On the other hand, we find that not all of the platforms have the characteristics of small-world networks. In our study, Tongcheng has a larger average path length (4.233) and a smaller clustering coefficient (0.113); thus, it does not have small-world network characteristics (Telesford et al., 2011). In the network, a short average path length can quickly transmit information and reduce costs (Zhang & Guo, 2014). Obviously, in this study, the efficiency of information transmission in Tongcheng is lower than in Ctrip and Tuniu. This inefficient transmission of information reduces Tongcheng users' perception of tourism products, which may reduce their willingness to make purchase decisions.

Third, our research can accurately identify the hot topics of online travel reviews and the social network relationships formed by hot topics. Our study found that the Ctrip and Tunniu platforms are composed of two communities: guides and hotels. There are several sub-communities in the tour community, such as the sub-community of satisfaction, the sub-community of explanation, and the sub-community of travel itinerary. Among these, the connection strength between the guide community and the explanation sub-community is greater, indicating that users usually consider the information of the guide and the guide's explanation at the same time. Therefore, this study suggests that improving the service level of tour guides based on the guide's explanation ability and itinerary is the key to increasing the satisfaction of consumers on Ctrip and Tuniu. On the Tongcheng platform, the hotel community is the largest. It is worth noting that the hotel community occupies more than half of the whole network; however, the strength of its associated sub-communities, such as the room sub-community, the location sub-community, and the reception sub-community, is not very high. Other communities in the network have a lower degree of

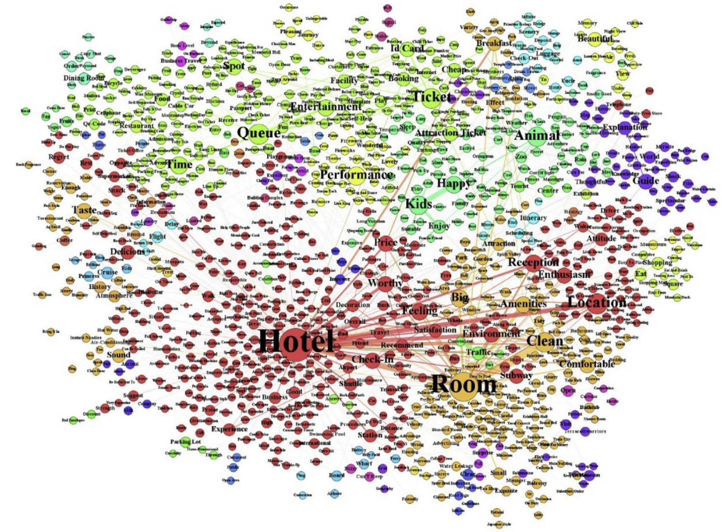


图6. 桐城的网络可视化（预览率为80. ）。

让消费者从在线旅游评论中获得有价值的信息。

5. 讨论和结论

5.1. 主要发现

首先，这项研究发现，在三个研究平台上，用户的核心关注点存在差异。携程和途牛的用户关注指南和体验，而桐城的用户更关注酒店和体验。主题词分类统计分析结果显示，携程和途牛用户对导游的关注度分别为37.8%和31.3%，远远超过桐城的8.3%。这一发现表明，导游的服务水平对两个平台上游客的满意度有很大影响，这支持了最近的研究结果（Tsai等人，2015）。然而，桐城用户更关注酒店相关信息，如环境和设施（郭等人，2017；项等人，2017），关注率达到35.6%。我们认为，用户讨论的主题倾向反映了每个平台的核心竞争力和营销策略。例如，携程和途牛采用成熟的旅游产品开发和设计系统，可以为游客提供个性化的旅行路线。在个性化旅游中，导游服务是保证游客满意度的关键。休闲旅游是桐城的主要业务，休闲游客对酒店及其环境更为敏感（Kim & Park, 2017）。这类用户主要关注酒店信息也就不足为奇了。

第二，我们在网络结构特性方面获得了有意义的发现。一方面，我们发现携程和途牛的模块化指数非常低，分别为0.289和0.329,表明网络社区是分散的。值得注意的是，桐城的模块化指数值超过了0.44的阈值，但模块化组的数量很大，这表明尽管桐城集团内部的话题相对集中，集团之间的沟通非常松散。这种松散的布局可能会导致网络之间的低连通性和组之间的低关联度。这可能是因为用户对旅游产品的接受度较低，这影响了他们发布评论的质量（Chatterjee，2001）。另一方面，我们发现并非所有平台都具有小世界网络的特征。在我们的研究中，桐城具有较大的平均路径长度（4.233）和较小的聚类系数（0.113）；因此，它没有小世界网络特征（Telesford等人，2011年）。在网络中，较短的平均路径长度可以快速传输信息并降低成本（Zhang & Guo, 2014）。显然，在本研究中，桐城的信息传输效率低于携程和途牛。这种低效的信息传输降低了桐城用户对旅游产品的感知，这可能会降低他们做出购买决定的意愿。

第三，我们的研究能够准确识别在线旅游评论的热点话题，以及热点话题形成的社会网络关系。我们的研究发现，携程和屯牛平台由两个社区组成：导游和酒店。旅游社区中有几个亚社区，如满意度亚社区、解释亚社区和旅游行程亚社区。其中，引导社区与解释子社区之间的连接强度较大，表明用户通常同时考虑导游信息和导游说明。因此，本研究表明，基于导游的讲解能力和行程安排，提高导游的服务水平是提高携程和途牛消费者满意度的关键。在桐城平台上，酒店社区是最大的。值得注意的是，酒店社区占据了整个网络的一半以上；然而，其相关子社区（如房间子社区、位置子社区和接待子社区）的强度不是很高。网络中的其他社区具有较低的

association with the hotel community. This low-association network indicates that Tongcheng's reviews are not sufficiently focused, which leads to the deviation of information obtained by potential consumers. Accordingly, there is important practical value in identifying ways to guide users to post reviews with high information quality.

5.2. Implications for research

First, this study provides a unique research perspective on online travel reviews. We move past the limitations of prior studies and introduce social network theory into our research. This study holds that online travel reviews reflect the social interaction between reviewers and tourists, which represents a type of social relationship. Social networks, which are understood as social relationships, refer to all formal and informal social relationships within a group of specific people, including the indirect social relationships linked by the physical environment, cultural sharing and direct social relationships (Mitchell, 2010). Therefore, the thematic words of online travel reviews can be regarded as nodes, and the semantic associations among them can be regarded as connections. Together, nodes and connections construct the social networks of online travel reviews. This finding not only considers a new research perspective on online travel reviews but also expands the scope of the application of social network theory.

Second, this study proposes a novel analytical framework to extract important topics from a large number of online travel reviews. This analytical framework integrates various analysis methods, including web crawls, semantic association analysis, social network analysis, and visual analysis. Compared to the previous research methods of online travel reviews, such as questionnaires, LDAs, sentiment analysis, and statistical analysis (Guo et al., 2017; Hu et al., 2017; Min et al., 2015; Ren & Hong, 2017; Schuckert et al., 2015b), we can obtain more reliable conclusions. On the one hand, tourists' needs can be identified based on word granularity by using semantic association analysis, which provides strong support for mining potential customer value in online travel reviews. Objective data can be obtained through the use of web crawlers to eliminate the influence of data deviation in subjective surveys. On the other hand, social network analysis provides a quantitative representation of the relationship between subject words in online travel reviews, reveals the structural characteristics of the relationship, and builds a bridge between macro- and micro research on online travel reviews. In particular, visualization turns boring textual information into interesting pictures to help us deepen our understanding and quickly identify important information in online travel reviews.

Third, the results of this study contribute to the theoretical development of the online travel review-related research by using semantic association analysis. The method of semantic association analysis can reveal the hidden logical relationship behind the text, promote the understanding of the content of online travel reviews, and more effectively identify valuable topics. For example, the results of the bigram co-occurrence phrases of semantic association show that *itinerary–scheduling*, *guide–explanation*, *guide–satisfaction*, and *guide–scheduling* are the core topics on Ctrip and Tuniu, indicating that tour guide service quality is an important factor of tourist satisfaction, which is consistent with previous studies (Caber & Albayrak, 2016; Heung, 2008; Mossberg, 1995). In the analysis of the structural properties of the semantic association network, we reached a conclusion that differs from (Raisi et al., 2017). For example, the phenomenon of a small-world network on the Internet is not universal, and the semantic association network of online travel reviews constructed by Tongcheng does not have small-world network properties. Generally, semantic association analysis is not only an effective tool for the opinion mining of online travel reviews but also provides a new perspective for value discovery.

Finally, this study accurately reveals the intricate network relationship through visualization. We extracted thematic words from the texts of online travel reviews and then constructed bigram co-occurrence phrases using semantic association and formed a visual network graph using Gephi software. The graph depicts the intricate relationship among the core topics from the word granularity perspective, effectively resolving the shortcomings of accuracy of the previous studies (Boo & Busser, 2018; Xiang et al., 2017). For example, as seen from the visualization graph of Ctrip, the satisfaction sub-community includes many thematic words, such as attitude, consideration, consultation, and enthusiasm, which indicate the origin of tourist satisfaction. Furthermore, there is a complex relationship between the satisfaction sub-community and other sub-communities, such as the guide sub-community and the hotel sub-community. Thicker connection lines indicate greater association strength.

5.3. Implications for practice

The results of this study have important practical implications for consumer travel decision-making, for the improvement of hotel and tourism enterprises' service quality, and for the strategic development of online tourism platforms.

For consumers, online travel reviews are an important source of information for obtaining travel services (Blomberg-Nygard & Anderson, 2016). These reviews help them to improve the accuracy and relevance of their access to information by using our method, thereby enhancing willingness to make purchase decisions. Online travel reviews allow quick access to content and reduce the risk of making travel decisions (Lian & Yu, 2017; Xiang et al., 2015; Ye et al., 2011). For example, before choosing a specific tourism product, according to this research method, consumers can obtain tourism-related information such as travel schedules, tourists' satisfaction with tourism products or destinations, hotels and surrounding environments. Additionally, by analysing the differences among the core businesses of different tourism platforms, consumers can purchase tourism products or services from the appropriate OTA depending on their needs.

The findings of this study can help hotels and tourism enterprises quickly identify the hot topics of online tourism reviews and find correlations among topics. The potential information extracted from online reviews is important for improving the service management and competitive advantage of hotels and tourism firms (Fang et al., 2016; Lui et al., 2018; Miguéis & Nóvoa, 2017; Yang, Shin, Joun, & Koo, 2016). Using this approach, enterprises can better and more accurately understand the needs of users and improve their quality of service. Helpful topic information will attract consumers, thus enhancing their willingness to purchase. Therefore, managers should focus on topics that are precise or easy to understand because these topics are more influential than fuzzy reviews. For example, "guide" in Ctrip and Tuniu is the core thematic, and thematises with high relevance include itinerary, scheduling, explanation, satisfaction, enthusiasm, responsible, patience, etc. Thus, tourism enterprise managers should improve the service quality of tour guides and increase professional training based on aspects such as tour guides' explanatory ability, itinerary scheduling, and service attitude (Alani, Khan, & Manuel, 2017, pp. 2395–7654; Weiler & Walker, 2014). In particular, there is a high degree of correlation between itinerary and scheduling. Therefore, the results of this paper also provide a reference for managers to conduct differentiation strategies (Lui et al., 2018). In addition, opinion mining is a process of knowledge discovery with which managers can design brand advertisements and develop themes that meet consumer demand using our research method, which will provide a better experience to tourists.

For online travel platforms, there are three aspects of practical significance. They can be summarized as follows.

First, through the mining of online tourism reviews, the platform's positioning is clearly defined from the perspective of consumers, which provides a reference for the platform to develop strategic planning and operational strategies. As our research findings reveal, not all review websites have the same quality of service and focus. For example,

与酒店社区的联系。这种低关联网络表明桐城的评论不够集中，导致潜在消费者获得的信息出现偏差。因此，确定引导用户发布高信息质量评论的方法具有重要的实用价值。

5.2. 对研究的影响

首先，本研究为在线旅游评论提供了独特的研究视角。我们超越了以往研究的局限性，将社会网络理论引入到我们的研究中。本研究认为，在线旅游评论反映了评论者和游客之间的社会互动，这代表了一种社会关系。社交网络被理解为社会关系，指的是特定人群中所有正式和非正式的社会关系，包括由物理环境、文化共享和直接社会关系联系起来的间接社会关系（Mitchell，2010）。因此，在线旅游评论中的主题词可以被视为节点，它们之间的语义关联可以被视为连接。节点和连接共同构成了在线旅游评论的社交网络。这一发现不仅为在线旅游评论提供了新的研究视角，也拓展了社交网络理论的应用范围。

其次，本研究提出了一个新的分析框架，从大量在线旅游评论中提取重要主题。该分析框架集成了各种分析方法，包括网络爬网、语义关联分析、社交网络分析和可视化分析。与之前的在线旅游评论研究方法相比，如问卷调查、LDA、情绪分析和统计分析（郭等人，2017年；胡等人，2017年；闵等人，2015年；任和洪，2017年；舒克特等人，2015b），我们可以获得更可靠的结论。一方面，通过语义关联分析，基于词粒度识别游客需求，为在线旅游评论中挖掘潜在客户价值提供有力支持。通过使用网络爬虫可以获得客观数据，以消除主观调查中数据偏差的影响。另一方面，社交网络分析为在线旅游评论中主题词之间的关系提供了定量表示，揭示了这种关系的结构特征，并在在线旅游评论的宏观和微观研究之间搭建了一座桥梁。特别是，可视化将枯燥的文本信息转化为有趣的图片，帮助我们加深理解并快速识别在线旅游评论中的重要信息。

第三，本研究的结果有助于利用语义关联分析进行在线旅游评论相关研究的理论发展。语义关联分析方法可以揭示文本背后隐藏的逻辑关系，促进对在线旅游评论内容的理解，更有效地识别有价值的话题。例如，语义联想的双元共现短语的结果表明，在携程和途牛上，行程-y-安排、导游-解释、导游-满意度和导游-安排是核心话题，表明导游服务质量是游客满意度的重要因素，这与之前的研究一致（Caber & Albayrak, 2016；Heung, 2008；Mossberg, 1995）。在分析语义关联网络的结构属性时，我们得出了一个不同于（Raisi等人，2017年）的结论。例如，互联网上的小世界网络现象并不普遍，桐城构建的在线旅游评论语义关联网络不具备小世界网络属性。一般来说，语义关联分析不仅是在线旅游评论意见挖掘的有效工具，而且为价值发现提供了新的视角。

最后，本研究通过可视化准确揭示了复杂的网络关系。我们从文章中提取了主题词在线旅游评论的文本，然后使用语义关联构建二元共现短语，并使用Gephi软件形成可视化网络图。该图从单词粒度的角度描述了核心主题之间的复杂关系，有效地解决了之前研究的准确性缺陷（Boo & Busser, 2018；Xiang等人，2017）。例如，从携程的可视化图表中可以看出，满意度子社区包含许多主题词，如态度、考虑、咨询和热情，这些词表明了游客满意度的来源。此外，满意度子社区与其他子社区（如导游子社区和酒店子社区）之间存在复杂的关系。较粗的连接线表明关联强度更大。

5.3. 对实践的影响

研究结果对消费者旅游决策、酒店和旅游企业服务质量的提高以及在线旅游平台的战略发展具有重要的现实意义。

对于消费者来说，在线旅游评论是获得旅游服务的重要信息来源（Blomberg Nygard & Anderson, 2016）。这些审查有助于他们使用我们的方法提高获取信息的准确性和相关性，从而增强做出购买决策的意愿。在线旅游评论允许快速访问内容并降低做出旅游决策的风险（Lian & Yu, 2017；Xiang等人，2015；Ye等人，2011）。例如，在选择特定的旅游产品之前，根据这一研究方法，消费者可以获得与旅游相关的信息，如行程安排、游客对旅游产品或目的地、酒店和周边环境的满意度。此外，通过分析不同旅游平台的核心业务之间的差异，消费者可以根据自己的需求从适当的在线旅行社购买旅游产品或服务。

本研究的结果可以帮助酒店和旅游企业快速识别在线旅游评论的热点话题，并找到话题之间的相关性。从在线评论中提取的潜在信息对于改善酒店和旅游公司的服务管理和竞争优势非常重要（方等人，2016年；吕等人，2018年；米盖斯和诺沃阿，2017年；杨、申、俊和古，2016年）。使用这种方法，企业可以更好、更准确地了解用户的需求，提高他们的服务质量。有用的话题信息会吸引消费者，从而提高他们的购买意愿。因此，管理者应该关注精确或易于理解的主题，因为这些主题比模糊的评论更有影响力。例如携程和途牛的"指南"是核心主题，具有高度相关性的主题包括行程安排、日程安排、讲解、满意度、热情、负责、耐心等，旅游企业管理者应提高导游的服务质量，并根据导游的讲解能力、行程安排和服务态度等方面加强专业培训（Alani、Khan和Manuel, 2017年，第2395–7654页；Weiler和Walker, 2014年）。特别是，行程和日程安排之间有着高度的相关性。因此，本文的研究结果也为管理者实施差异化战略提供了参考（Lui等人，2018）。此外，意见挖掘是一个知识发现的过程，管理者可以利用我们的研究方法设计品牌广告和开发满足消费者需求的主题，这将为游客提供更好的体验。

对于在线旅游平台来说，有三个方面的现实意义。它们可以总结如下。

首先，通过对在线旅游评论的挖掘，从消费者的角度明确了平台的定位，为平台制定战略规划和运营策略提供了参考。正如我们的研究结果所揭示的，并不是所有的评论网站都有相同的服务质量和关注点。例如

Tongcheng users focus more heavily on hotel-related online reviews and less on scenic spots or guides. We suggest that this platform should focus on businesses that are related to scenic spots or guides when formulating a tourism product operation strategy.

Second, the results of this study can improve a platform's keyword matching and optimize its search engine. Our approach can identify keywords that affect users' travel decisions. The thematic words with a high correlation with the core keywords are long-tail keywords. According to these results, a platform can optimize the keyword settings of its tourism products based on core keywords and long-tail keywords.

Finally, the results lend strong support to the platform's ability to guide more consumers to post high-quality online travel reviews. As the last step of online travel, consumers often ignore online comment publishing. The question of how to guide consumers to actively post travel reviews with rich and high-quality information has become the focus of online tourism platforms. For example, this study found that users of Ctrip and Tuniu focused on information on guides. Consequently, the managers of these platforms can set up comment labels, such as Guide, Hotel, Service, Spot and Experience, in their review sections. In addition, we suggest that platforms introduce "sub-labels" under these labels. For example, under the Guide label, a "tree-like" evaluation system can be established with sub-labels such as itinerary scheduling, explanation level, and personal charm, which can effectively save consumers the cost of thinking when they post comments.

5.4. *Limitations and future research*

This study has some limitations and highlights future research directions. First, the results of this study are based on a dataset of three major OTAs in China. Under the circumstances of the rapid integration of e-commerce into tourism, online tourism platforms have emerged all over the world. The findings of this study may not be generalizable to other scenic spots and international tourism markets. Future research can improve the generalizability of the study's results by using international data and can explore the impact of destination differences and cultural differences. Second, to control analysis bias, we selected only online travel reviews published after 2015. Future research should consider obtaining more data to further validate the research conclusions. Third, this study analyses the structural properties of the semantic association networks of online travel reviews, such as density, modularity, and average clustering coefficient; however, it does not discuss the relationship between structural properties and tourism product performance. Determining whether a semantic association network affects the performance of online tourism products and identifying influence mechanisms are also valuable directions for future research.

Author Contribution

The following is the initial letter of the author's name that contributes to each category.

Guarantor of integrity of entire study: Hou, Cui, Meng, Lian, Yu.
Study concepts: Hou, Cui.
Study design: Hou, Cui, Meng.
Literature research: Hou, Cui, Meng.
Experimental studies: Hou, Cui, Lian.
Data acquisition: Cui, Lian, Yu.
Data analysis/interpretation: Cui, Meng, Lian.
Statistical analysis: Hou, Cui, Lian, Yu.
Manuscript preparation: Cui, Meng, Yu.
Manuscript definition of intellectual content: Hou, Meng, Lian, Yu.
Manuscript editing: Hou, Meng, Lian, Yu.
Manuscript revision/review: Hou, Cui, Meng, Lian, Yu.
Manuscript final version approval: Hou, Cui, Meng.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC No. 71562009, 71562008, 71602092), Key Research Institute of Philosophies and Social Sciences in Guangxi Universities (No. 17ZD002), Innovation Project of Guangxi Graduate Education (No. YCSW2018162) and Natural Science Foundation of Anhui Province (No. 1608085QG168).

References

iResearch (2018). *Research report on China's online tourism and holiday industry in 2018*http://report.iresearch.cn/report/201805/3214.shtml.
Alani, F., Khan, F. R., & Manuel, D. F. (2017). *Need for professionalism and quality service of the tourist guides in Oman. Vol. 4*Social Science Electronic Publishing 1.
Alemanmeza, B., Halaschekwiener, C., Sahoo, S. S., Sheth, A., & Arpinar, I. B. (2005). Template based semantic similarity for security applications. *IEEE International Conference on Intelligence and Security Informatics* (pp. 621–622). .
Ali, F., Kwak, K. S., & Kim, Y. G. (2016). Opinion mining based on fuzzy domain ontology and support vector machine: A proposal to automate online review classification. *Applied Soft Computing, 47*(1), 235–250.
Asiedu, K. (2014). Applying data mining and social network analysis techniques in a land tenure information system. *New Biotech, 31*(28), 1–136.
Banerjee, S., & Chua, A. Y. K. (2016). In search of patterns among travellers' hotel ratings in tripadvisor. *Tourism Management, 53*, 125–131.
Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the third international ICWSM conference* (pp. 361–362). .
Bellman, S., Johnson, E. J., Lohse, G. L., & Mandel, N. (2010). Designing marketplaces of the artificial with consumers in mind: Four approaches to understanding consumer behavior in electronic environments. *Journal of Interactive Marketing, 20*(1), 21–33.
Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit.* Sebastopol: O'Reilly Media.
Blomberg-Nygard, A., & Anderson, C. K. (2016). United nations world tourism organization study on online guest reviews and hotel classification systems: An integrated approach. *Service Science, 8*(2), 139–151.
Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics, 2008*(10), 155–168.
Boo, S., & Busser, J. A. (2018). Meeting planners' online reviews of destination hotels: A twofold content analysis approach. *Tourism Management, 66*(6), 287–301.
Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology, 25*(2), 163–177.
Bucur, C. (2015). Using opinion mining techniques in tourism. *Procedia Economics & Finance, 23*, 1666–1673.
Caber, M., & Albayrak, T. (2016). Assessing daily tour service quality: A proposal for a DAILYSERV scale. *Journal of Destination Marketing & Management, 7*, 18–25.
Cai, X., Xu, J., & Wu, S. (2014). Research on filter model of customer review for sentiment analysis. *New Technology of Library & Information Service,* (4), 58–64.
Casaló, L. V., Flavián, C., Guinalíu, M., & Ekinci, Y. (2015). Do online hotel rating schemes influence booking behaviors? *International Journal of Hospitality Management, 49*, 28–36.
Casanueva, C., Gallego, Á., & Garcíasánchez, M. (2014). Social network analysis in tourism. *Current Issues in Tourism, 19*(12), 1190–1209.
Chang, R. C. Y., Kivela, J., & Mak, A. H. N. (2011). Attributes that influence the evaluation of travel dining experience: When East meets West. *Tourism Management, 32*(2), 307–316.
Chatterjee, P. (2001). Online reviews: Do consumers use them? *Advances in Consumer Research, 28*, 133–139.
Che, W., Li, Z., & Liu, T. (2010). LTP: A Chinese language technology platform. *International Conference on Computational Linguistics: Demonstrations* (pp. 13–16). .
Cheng, X., Fu, S., Sun, J., Bilgihan, A., & Okumus, F. (2019). An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust. *Tourism Management, 71*, 366–377.
Chen, R., Liang, C. Y., Hong, W. C., & Gu, D. X. (2015). Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. *Applied Soft Computing Journal, 26*(C), 435–443.
Daud, A., Khan, W., & Che, D. (2017). *Urdu language processing: A survey. Alphen aan den rijn.* Kluwer Academic Publishers.
Donohue, J. C. (1973). *Understanding scientific literatures: A bibliometric approach.* Massachusetts: MIT Press.
Duverger, P. (2013). Curvilinear effects of user-generated content on hotels' market share: A dynamic panel-data analysis. *Journal of Travel Research, 52*(4), 465–478.
Fang, B., Ye, Q., Kucukusta, D., & Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management, 52*, 498–506.
Filieri, R. (2016). What makes an online consumer review trustworthy? *Annals of Tourism Research, 58*, 46–64.
Filieri, R., & Mcleay, F. (2014). E-wom and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews. *Journal of Travel Research, 53*(1), 44–57.
García-Pablos, A., Cuadros, M., & Linaza, M. T. (2016). Automatic analysis of textual hotel reviews. *Information Technology & Tourism, 16*(1), 45–69.

Z. Hou, et al.
Tourism Management 74 (2019) 276–289

桐城的用户更多地关注与酒店相关的在线评论，而较少关注景点或导游。我们建议该平台在制定旅游产品运营策略时，应关注与景区或导游相关的业务。

第二，本研究的结果可以改善平台的关键词匹配，优化其搜索引擎。我们的方法可以识别影响用户出行决策的关键词。与核心关键词高度相关的主题词是长尾关键词。根据这些结果，平台可以基于核心关键词和长尾关键词优化其旅游产品的关键词设置。最后，研究结果有力地支持了该平台引导更多消费者发布高质量在线旅游评论的能力。作为在线旅游的最后一步，消费者往往忽视在线评论发布。如何引导消费者主动发布具有丰富高质量信息的旅游评论已成为在线旅游平台关注的焦点。例如，这项研究发现携程和途牛的用户关注的是导游信息。因此，这些平台的经理可以在他们的评论部分设置评论标签，如指南、酒店、服务、地点和体验。此外，我们建议平台在这些标签下引入"子标签"。例如，在向导标签下，可以建立一个"树状"评估系统，包括行程安排、解释级别和个人魅力等子标签，这可以有效地节省消费者发布com时的思考成本-

是的。

5. 4. *局限性与未来研究*

这项研究有一些局限性，并强调了未来的研究方向。首先，本研究的结果基于中国三大在线旅行社的数据集。在电子商务迅速融入旅游业的背景下，在线旅游平台在世界各地应运而生。这项研究的结果可能无法推广到其他景点和国际旅游市场。未来的研究可以利用国际数据提高研究结果的普遍性，并探索目的地差异和文化差异的影响。其次，为了控制分析偏差，我们只选择了2015. 后发布的在线旅游评论。未来的研究应该考虑获取更多的数据来进一步验证研究结论。第三，本研究分析了在线旅游评论语义关联网络的结构特征，如密度、模块性和平均聚类系数；然而，它没有讨论结构属性和旅游产品性能之间的关系。确定语义关联网络是否会影响在线旅游产品的性能，并确定影响机制也是未来研究的重要方向。

作者贡献

以下是作者姓名的首字母，它构成了每个类别。
整个研究的完整性保证人：侯、崔、孟、连、余。学习理念：侯，崔。
研究设计：侯、崔、孟。文献研究：
侯、崔、孟。实验研究：侯、崔、廉
。数据采集：崔，连，于。
数据分析/解释：崔、孟、廉。统计分析：侯，
崔，连，于。
稿件准备：崔、孟、余。
书稿对知识分子内容的定义：侯、孟、廉、虞。稿件编辑：侯、孟、连、于。
审稿/审稿：侯、崔、孟、连、于。稿件最终版本批准：侯、崔、孟。

确认

本研究得到中国国家自然科学基金（NSFC 71562009, 71562008, 71602092）、广西大学哲学社会科学重点研究所（17ZD00）的资助。广西省研究生教育创新工程（YCSW2018162）和安徽省自然科学基金（1608085 QG168）。

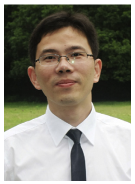
参考文献

艾瑞（2018）。2018年中国在线旅游度假产业研究报告 http://report.iresearch.cn/report/201805/3214.shtml.
阿拉尼·F.、汗·F·R.、曼努埃尔·D·F.（2017）。阿曼对导游专业化和优质服务的需求。第4卷社会科学电子出版1。
Alemanmeza, B., Halaschekwiener, C., Sahoo, S. S., Sheth, A.,和Arpinar, I. B.（2005）。基于模板的安全应用语义相似性。IEEE情报与安全信息学国际会议（第621–622页）。
阿里，F.,郭，K. S.,和金，Y. G.（2016）。基于模糊领域本体和支持向量机的观点挖掘：一种自动在线评论分类的建议。应用软件计算，47（1），235–250。
Asiedu, K. (2014). Applying data mining and social network analysis techniques in a land tenure information system. 新生物科技，31（28），1–136。
Banerjee, S., & Chua, A. Y. K. (2016). In search of patterns among travellers' hotel ratings in tripadvisor. 旅游管理，53125-131。
巴斯蒂安，M.,海曼，S.,和杰科米，M.（2009）。Gephi：用于探索和操作网络的开源软件。第三届ICWSM国际会议记录（第361–362页）。
贝尔曼，S.,约翰逊，E. J.,洛斯，G. L.,曼德尔，N.（2010）。以消费者为中心设计人工市场：理解电子环境中消费者行为的四种方法。互动营销杂志，20（1），21–33。
伯德，S.,克莱恩，E.,和洛珀，E.（2009）。使用Python进行自然语言处理：使用自然语言工具包分析文本。塞巴斯托波尔：奥莱利媒体。
Blomberg Nygard，A.和Anderson, C. K.（2016）。联合国世界旅游组织关于在线客人评论和酒店分类系统的研究：综合方法。服务科学，8（2），139–151。
Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. 统计力学杂志，2008（10），155–168。
Boo, S., & Busser, J. A. (2018). Meeting planners' online reviews of destination hotels: A twofold content analysis approach. 旅游管理，66（6），287–301。
美国布兰德斯（2001）。中间值中心的快速算法。《数学社会学杂志》，25（2），163–177。
Bucur, C.（2015）。在旅游业中使用意见挖掘技术。《经济学与金融百科全书》，231666-1673年。
Caber, M., & Albayrak, T. (2016). Assessing daily tour service quality: A proposal for a DAILYSERV scale. 目的地营销与管理杂志，7, 18–25。
蔡X，徐J，吴S.（2014）。用于情绪分析的客户评论过滤模型研究。图书情报服务新技术，（4），58–64。
Casaló, L. V., Flavián, C., Guinalíu, M.,和Ekinci, Y.（2015）。在线酒店评级方案是否会影响预订行为？《国际酒店管理杂志》，49、28–36。
Casanueva, C., Gallego, Á., & Garcíasánchez, M. (2014). Social network analysis in tourism. 《旅游业当前问题》，19（12），1190–1209。
张学友，基维拉，J.,和麦，A. H. N.（2011）。影响旅游餐饮体验评价的属性：东西方相遇。旅游管理，32（2），307–316。
Chatterjee, P.（2001年）。在线评论：消费者使用它们吗？《消费者研究进展》，28133-139。
车伟、李子和刘婷（2010）。LTP：一个中文技术平台。计算语言学国际会议：演示（第13–16页）。
郑，X.,傅，S.,孙，J.,比尔吉汉，A.,和奥库姆斯，F.（2019年）。共享经济驱动的酒店平台在线评论调查：信任的观点。旅游管理，71366-377。
陈若亮、梁春云、洪文昌、顾德新（2015）。基于季节性支持向量回归和自适应遗传算法的假日日游客流量预测。应用软件计算杂志，26（C），435–443。
Daud, A., Khan, W.,和Che, D.（2017）。乌尔都语语言处理：一项调查。阿尔芬·安登·莱恩。Kluwer学术出版社。
Donohue, J. C. (1973). 理解科学文献：文献计量学方法。马萨诸塞州：麻省理工学院出版社。
杜维杰，P.（2013）。用户生成内容对酒店市场份额的曲线效应：动态面板数据分析。旅游研究杂志，52（4），465–478。
方，B.,叶，Q.,库库斯塔，D.,罗，R.（2016）。在线旅游评论感知价值分析：可读性和评论人特征的影响。旅游管理，52498-506。
菲列里，R.（2016）。什么使在线消费者评论可信？《旅游研究年鉴》，58, 46–64。
菲列里·R.和麦克莱伊·F.（2014）。电子口碑和住宿：影响旅行者采纳在线评论信息的因素分析。旅游研究杂志，53（1），44–57。
García-Pablos, A., Cuadros, M., & Linaza, M. T. (2016). Automatic analysis of textual hotel reviews. 信息技术与旅游，16（1），45–69。

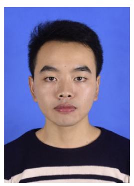
Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467–483.
He, C. (2013). Business intelligence analysis algorithm: A research based on semantic association analysis. *Journal of Intelligence, 32*(4), 134–137.
Hemmatian, F., & Sohrabi, M. K. (2017). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review,* (1), 1–51.
Herrera, F., & Martinez, L. (2000). A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems, 8*(6), 746–752.
Heung, V. C. S. (2008). Effects of tour leader's service quality on agency's reputation and customers word-of-mouth. *Journal of Vacation Marketing, 14*(4), 305–315.
Hodac, N. N., Carson, S. J., & Moore, W. L. (2013). The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *Journal of Marketing, 77*(6), 37–53.
Hu, Y. H., Chen, Y. L., & Chou, H. L. (2017). Opinion mining from online hotel reviews – a text summarization approach. *Information Processing & Management, 53*(2), 436–449.
Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, A continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS One, 9*(6), 1–12.
Jennings, G. (2001). *Qualitative methods and tourism research & qualitative methods of data analysis.* Sydney, Australia: John Wiley & Sons.
Ji, X., Li, J., & Chen, F. (2016). Spatial structure of self- help tourism in Yunnan Province based on social network analysis. *Journal of Arid Land Resources & Environment, 30*(6), 204–208.
Kim, K. K., Karunanayaka, P., Privitera, M. D., Holland, S. K., & Szaflarski, J. P. (2011). Semantic association investigated with functional MRI and independent component analysis. *Epilepsy and Behavior, 20*(4), 613–622.
Kim, D., & Park, B. J. (2017). The moderating role of context in the effects of choice attributes on hotel choice: A discrete choice experiment. *Tourism Management, 63*, 439–451.
Kim, K., Park, O., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change, 123*, 362–369.
Kuhlmann, M., Hofmann, M. J., & Jacobs, A. M. (2017). If you don't have valence, ask your neighbor: Evaluation of neutral words as a function of affective semantic associates. *Frontiers in Psychology, 8*(8), 1–7.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163.
Ladhari, R., & Michaud, M. (2015). eWOM effects on hotel booking intentions, attitudes, trust, and website perceptions. *International Journal of Hospitality Management, 46*(3), 36–45.
Latapy, M. (2008). Main-memory triangle computations for very large (sparse power-law) graphs. *Theoretical Computer Science, 407*(1), 458–473.
Leclerc, D., & Martin, J. N. (2004). Tour guide communication competence: French, German and American tourists' perceptions. *International Journal of Intercultural Relations, 28*(3–4), 181–200.
Lian, T., & Yu, C. (2017). Representation of online image of tourist destination: A content analysis of huangshan. *Asia Pacific Journal of Tourism Research, 22*(2), 1–20.
Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management, 29*(3), 458–468.
Liu, Y., & Du, Y. (2017). *Network data visualization and analysis tools: Gephi Chinese tutorial.* Beijing: Publishing House of Electronics Industry.
Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management, 47*, 140–151.
Liu, X., Schuckert, M., & Law, R. (2018). Utilitarianism and knowledge growth during status seeking: Evidence from text mining of online reviews. *Tourism Management, 66*, 38–46.
Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management, 68*, 301–323.
Li, M., & Yi, L. (2014). Application and optimization of customer relationship management in tongcheng website. *Business Economy, 36*(11), 136–141.
Lui, T. W., Bartosiak, M., Piccoli, G., & Sadhya, V. (2018). Online review response strategy and its effects on competitive performance. *Tourism Management, 67*, 180–190.
Lu, C., & Liu, S. (2016). Cultural tourism O2O business model innovation: A case study of ctrip. *Journal of Electronic Commerce in Organizations, 14*(2), 16–31.
Luo, X., Gu, B., Jie, Z., & Phang, C. W. (2017). *Expert blogs and consumer perceptions of competing brands. 41*, Social Science Electronic Publishing371–395 2.
Mccain, S. L. C., Jang, S. C., & Hu, C. (2005). Service quality gap analysis toward customer loyalty: Practical guidelines for casino hotels. *International Journal of Hospitality Management, 24*(3), 465–472.
Miguéis, V. L., & Nóvoa, H. (2017). Exploring online travel reviews using data analytics: An exploratory study. *Service Science, 9*(4), 315–323.
Min, H. N., Lim, Y. M., & Magnini, V. P. (2015). Factors affecting customer satisfaction in responses to negative online hotel reviews: The impact of empathy, paraphrasing, and speed. *Cornell Hospitality Quarterly, 56*(2), 223–231.
Mitchell, J. C. (2010). Social networks in urban situations: Analyses of personal relationships in Central African towns. *American Journal of Sociology, 22*(7), 19–21.
Mossberg, L. L. (1995). Tour leaders and their importance in charter tours. *Tourism Management, 16*(6), 437–445.
Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. *Tourism Management, 32*(2), 215–224.
Qi, M., Li, X., Zhu, E., & Shi, Y. (2017). Evaluation of perceived indoor environmental quality of five-star hotels in China: An application of online review analysis. *Building and Environment, 111*, 1–9.
Racherla, P., & Friske, W. (2012). Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications, 11*(6), 548–559.
Radojevic, T., Stanisic, N., & Stanic, N. (2015). Ensuring positive feedback: Factors that influence customer satisfaction in the contemporary hospitality industry. *Tourism Management, 51*, 13–21.
Rahimi, R., & Kozak, M. (2016). Impact of customer relationship management on customer satisfaction: The case of a budget hotel chain. *Journal of Travel & Tourism Marketing, 34*(1), 1–12.
Raisi, H., Baggio, R., Barratt-Pugh, L., & Willson, G. (2017). Hyperlink network analysis of a tourism destination. *Journal of Travel Research, 57*(2), 1–27.
Ren, G., & Hong, T. (2017). Investigating online destination images using a topic-based sentiment analysis approach. *Sustainability, 9*(10), 1–18.
Schuckert, M., Liu, X., & Law, R. (2015a). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing, 32*(5), 608–621.
Schuckert, M., Liu, X., & Law, R. (2015b). A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently. *International Journal of Hospitality Management, 48*, 143–149.
ShaanxiDaily (2018). *2018 global destination marketing Summit and world culture and tourism Forum*.
Sirgy, M. J. (2010). Toward a quality-of-life theory of leisure travel satisfaction. *Journal of Travel Research, 49*(2), 246–260.
Somabhai, P. B., Varma, T., & Somabhai, P. P. (2015). A survey on feature based opinion mining for tourism industry. *Journal of Engineering Computers & Applied Sciences, 4*(3), 83–86.
Song, Z., & Wang, Y. (2013). Analysis on the access system and management of tour guides. *Tourism Tribune, 28*(7), 57–63.
Sotiriadou, P., Brouwers, J., & Le, T. A. (2014). Choosing a qualitative data analysis tool: A comparison of NVivo and leximancer. *Annals of Leisure Research, 17*(2), 218–234.
Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. (2011). The ubiquity of small-world networks. *Brain Connectivity, 1*(5), 367–375.
Tsai, C. Y., Wang, M. T., & Tseng, H. T. (2015). The impact of tour guides' physical attractiveness, sense of humor, and seniority on guide attention and efficiency. *Journal of Travel & Tourism Marketing, 33*(6), 1–13.
Wang, W., & Wang, L. (2018). Merger case analysis of tongcheng and yilong, two online travel agents. *Business & Economy,* (5), 93–94 + 113.
Wang, Y., Yang, L., Sun, C., & Jiang, J. (2017). Comparative research on travel sharing of typical travel website based on text mining: Taking gansu province as an example. *Resource Development & Market, 33*(1), 100–104.
Wasserman, S., & Faust, K. (1994). Social network analysis. *Encyclopedia of Social Network Analysis & Mining, 22*(Suppl 1), 109–127.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.
Wehbe, F., Hattab, M. A., & Hamzeh, F. (2016). Exploring associations between resilience and construction safety performance in safety networks. *Safety Science, 82*, 338–351.
Weiler, B. (2002). Tour guide textbooks: A review of two recent publications. *Tourism Management, 23*(2), 188–190.
Weiler, B., & Walker, K. (2014). Enhancing the visitor experience: Reconceptualising the tour guide's communicative role. *Journal of Hospitality and Tourism Management, 21*(21), 90–99.
Wenchin, T., Mingtsang, H., Liwen, S., & Lin, T. M. Y. (2015). Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management, 46*(April), 99–111.
Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51–65.
Xiang, Z., Schwartz, Z., Gerdes, J. H., Jr., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*(44), 120–130.
Xiang, P., Tian, K., & Huang, Q. (2007). Semantic association identification and knowledge discovery for business applications. *The international conference information computing and automation* (pp. 480–483). .
Xu, Y., & Mcgehee, N. G. (2016). Tour guides under zero-fare mode: Evidence from China. *Current Issues in Tourism, 20*(10), 1–22.
Yang, S. B., Shin, S. H., Joun, Y., & Koo, C. (2016). Exploring the comparative importance of online hotel reviews' heuristic attributes in review helpfulness: A conjoint analysis approach. *Journal of Travel & Tourism Marketing, 34*(7), 1–23.
Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management, 28*(1), 180–182.
Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior, 27*(2), 634–639.
Ye, Q., Li, H. Y., & Wang, Z. S. (2014). The influence of hotel price on perceived service quality and value in e-tourism: An empirical investigation based on online traveler reviews. *Journal of Hospitality & Tourism Research, 38*(1), 23–39.
Yin, C., & Peng, Q. (2010). Sentiment analysis for product features in Chinese reviews based on semantic association. *International conference on artificial intelligence and computational intelligence* (pp. 81–85). .
Yuan, J., & Wu, Z. T. (2016). Research on effect of convergence between creative industry park and tourism based on network text: A case study of shenzhen dafen oil painting village. *Resource Development & Market, 32*(5), 631–636.
Zhang, D. (2014). *Challenges in marketing strategy of online travel booking industry in China: A case study of Ctrip.com and Qunar.com.* School of Business & Engineering.
Zhang, D., & Guo, G. (2014). A comparison of online social networks and real-life social networks: A study of sina microblogging. *Mathematical Problems in Engineering, 2014*(6684), 1–6.
Zhang, F., Song, Y., Cai, W., Liu, S., Liu, S., Pujol, S., ... Feng, D. D. (2016a). Pairwise latent semantic association for similarity computation in medical imaging. *IEEE*

郭，Y.,巴恩斯，S. J.,贾，Q.（2017）。从在线评分和评论中挖掘意义：使用潜在dirichlet分配的游客满意度分析。旅游管理，59467-483。
He, C. (2013). Business intelligence analysis algorithm: A research based on semantic association analysis. 《情报杂志》，32（4），134–137。
Hemmatian, F., & Sohrabi, M. K. (2017). A survey on classification techniques for opinion mining and sentiment analysis. 人工智能评论（1），1–51。
Herrera, F., & Martinez, L. (2000). A 2-tuple fuzzy linguistic representation model for computing with words. IEEE模糊系统学报，8（6），746–752。
香，V. C. S.（2008）。领队服务质量对旅行社声誉和顾客口碑的影响。《假期营销杂志》，14（4），305–315。
北卡罗来纳州霍达克、南卡罗来纳州卡森和威尔士摩尔（2013）。正面和负面在线客户评论的影响：品牌实力和品类成熟度重要吗？市场营销杂志，77（6），37–53。
Hu, Y. H., Chen, Y. L., & Chou, H. L. (2017). Opinion mining from online hotel reviews – a text summarization approach. 信息处理与管理，53（2），436–449。
Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, A continuous graph layout algorithm for handy network visualization designed for the gephi software. 《公共科学图书馆一号》，第9（6）条，第1–12条。
詹宁斯，G.（2001）。定性方法与旅游研究&数据分析的定性方法。澳大利亚悉尼：约翰·威利父子公司。
季克强、李俊杰、陈福福（2016）。基于社会网络分析的云南自助旅游空间结构。干旱土地资源与环境杂志，30（6），204–208。
Kim, K. K., Karunanayaka, P., Privitera, M. D., Holland, S. K., & Szaflarski, J. P. (2011). Semantic association investigated with functional MRI and independent component analysis. 癫痫与行为，20（4），613–622。
Kim, D.,和Park, B. J.（2017）。情境在选择属性对酒店选择影响中的调节作用：一项离散选择实验。旅游管理，63439-451。
Kim, K., Park, O., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. 《技术预测与社会变革》，123362-369。
Kuhlmann, M., Hofmann, M. J., & Jacobs, A. M. (2017). If you don't have valence, ask your neighbor: Evaluation of neutral words as a function of a□ective semantic associates. 心理学前沿，8（8），1–7。
库塔斯，M.,和希利亚德，S. A.（1984）。阅读过程中的脑电位反映了词汇透视和语义关联。《自然》，307（5947），161–163。
拉德哈里，R.,和米肖，M.（2015）。eWOM对酒店预订意向、态度、信任和网站感知的影响。《国际酒店管理杂志》，46（3），36–45。
Latapy, M. (2008). Main-memory triangle computations for very large (sparse power-law) graphs. 理论计算机科学，407（1），458–473。
Leclerc, D.,和Martin, J. N.（2004）。导游沟通能力：法国、德国和美国游客的感知。《国际跨文化关系杂志》，28（3–4），181–200。
Lian, T., & Yu, C. (2017). Representation of online image of tourist destination: A content analysis of huangshan. 《亚太旅游研究杂志》，22（2），1–20。
Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. 旅游管理，29（3），458–468。
Liu, Y., & Du, Y. (2017). 网络数据可视化和分析工具：Gephi中文教程。北京：电子工业出版社。
Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. 旅游管理，47140-151。
刘克强，舒克特，M.,罗，R.（2018）。身份寻求过程中的功利主义和知识增长：来自在线评论文本挖掘的证据。旅游管理，66, 38–46。
Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. 旅游管理，68301-323。
Li, M., & Yi, L. (2014). Application and optimization of customer relationship management in tongcheng website. 商业经济，36（11），136–141。
Lui, T. W., Bartosiak, M., Piccoli, G., & Sadhya, V. (2018). Online review response strategy and its e对竞争绩效的影响。旅游管理，67180-190。
Lu, C., & Liu, S. (2016). Cultural tourism O2O business model innovation: A case study of ctrip. 组织中的电子商务杂志，14（2），16–31。
罗X.顾B.杰Z.潘C. W.（2017）。专家博客和消费者对竞争品牌的看法。41,社会科学电子出版物371–395 2。
Mccain, S. L. C., Jang, S. C.,和Hu, C.（2005）。顾客忠诚度服务质量差距分析：赌场酒店实用指南。《国际酒店管理杂志》，24（3），465–472。
Miguéis, V. L., & Nóvoa, H. (2017). Exploring online travel reviews using data analytics: An exploratory study. 服务科学，9（4），315–323。
Min, H. N., Lim, Y. M.,和Magnini, V. P.（2015）。影响顾客对负面在线酒店评论满意度的因素：同理心、释义和速度的影响。康奈尔酒店季刊，56（2），223–231。
Mitchell, J. C. (2010). Social networks in urban situations: Analyses of personal relationships in Central African towns. 《美国社会学杂志》，22（7），19–21。
莫斯伯格，L. L.（1995）。导游及其在包机旅游中的重要性。旅游管理，16（6），437–445。
Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. 旅游管理，32（2），215–224。
齐明，李，X，朱，E，和史，Y.（2017）。中国五星级酒店感知室内环境质量评估：在线评论分析的应用。《建筑与环境》，111, 1–9。
Racherla, P.,和Friske, W.（2012）。在线消费者评论的感知"有用性"：一项跨三个服务类别的探索性调查。电子商务研究与应用，11（6），548–559。
T.拉多耶维奇、N.斯坦尼西奇和N.斯坦尼奇（2015）。确保正面反馈：影响当代酒店业顾客满意度的因素。旅游管理，51, 13–21。
拉希米，R.,和科扎克，M.（2016）。顾客关系管理对顾客满意度的影响：以一家廉价连锁酒店为例。《旅游与旅游营销杂志》，34（1），1–12。
Raisi, H., Baggio, R., Barratt-Pugh, L., & Willson, G. (2017). Hyperlink network analysis of a tourism destination. 旅游研究杂志，57（2），1–27。
Ren, G., & Hong, T. (2017). Investigating online destination images using a topic-based sentiment analysis approach. 可持续性，9（10），1–18。
Schuckert, M., Liu, X., & Law, R. (2015a). Hospitality and tourism online reviews: Recent trends and future directions. 《旅游与旅游营销杂志》，32（5），608–621。
舒克特，M.,刘，X.,罗，R.（2015b）。按语言组对在线评论进行细分：英语和非英语使用者对酒店的评价如何不同。《国际酒店管理杂志》，48143-149。
《陕北日报》（2018）。2018年全球旅游目的地营销峰会暨世界文化旅游论坛。
Sirgy, M. J.（2010）。走向休闲旅游满意度的生活质量理论。旅游研究杂志，49（2），246–260。
Somabai，P. B., Varma, T.,和Somabai，P. P.（2015）。基于特征的旅游业意见挖掘综述。工程计算机与应用科学杂志，4（3），83–86。
Song, Z., & Wang, Y. (2013). Analysis on the access system and management of tour guides. 《旅游论坛报》，28（7），57–63。
Sotiriadou, P., Brouwers, J., & Le, T. A. (2014). Choosing a qualitative data analysis tool: A comparison of NVivo and leximancer. 《休闲研究年鉴》，17（2），218–234。
Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. (2011). The ubiquity of small-world networks. 大脑连接，1（5），367–375。
蔡志永、王敏婷和曾浩婷（2015）。导游的外表吸引力、幽默感和资历对导游注意力和效率的影响。《旅游与旅游营销杂志》，33（6），1–13。
Wang, W., & Wang, L. (2018). Merger case analysis of tongcheng and yilong, two online travel agents. 商业与经济，（5），93–94+113。
王勇、杨立军、孙春川、江俊杰（2017）。基于文本挖掘的典型旅游网站旅游共享比较研究——以甘肃省为例。资源开发与市场，33（1），100–104。
瓦瑟曼，S.,和福斯特，K.（1994）。社交网络分析。社会网络分析与挖掘百科全书，22（补充1），109–127。
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "小世界"网络。《自然》，393（6684），440–442。
Wehbe, F., Hattab, M. A., & Hamzeh, F. (2016). Exploring associations between resilience and construction safety performance in safety networks. 安全科学，82338–351。
韦勒，B.（2002）。导游教材：最近两次出版的评论。旅游管理，23（2），188–190。
韦勒，B.,沃克，K.（2014）。提升游客体验：重新认识导游的沟通角色。《酒店与旅游管理杂志》，21（21），90–99。
Wenchin, T., Mingtsang, H., Liwen, S., & Lin, T. M. Y. (2015). Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. 《国际酒店管理杂志》，46（4月），99–111。
向志强、杜、Q、马、Y、范伟（2017）。主要在线评论平台的比较分析：社交媒体分析对酒店业和旅游业的影响。旅游管理，58, 51–65。
项，Z.,施瓦茨，Z.,格德斯，J. H., Jr.,和乌萨尔，M.（2015）。关于酒店客人体验和满意度，大数据和文本分析能告诉我们什么？《国际酒店管理杂志》，44（44），120–130。
向平，田坤，黄问（2007）。用于商业应用的语义关联识别和知识发现。信息计算与自动化国际会议（第480–483页）。
徐勇和麦格赫，N. G.（2016）。零票价模式下的导游：来自中国的证据。旅游业的当前问题，20（10），1–22。
Yang, S. B., Shin, S. H., Joun, Y., & Koo, C. (2016). Exploring the comparative importance of online hotel reviews' heuristic attributes in review helpfulness: A conjoint analysis approach. 《旅游与旅游营销杂志》，34（7），1–23。
叶问，罗，R.,顾，B.（2009）。在线用户评论对酒店客房销售的影响。《国际酒店管理杂志》，28（1），180–182。
叶问，罗，R.,顾，B.,陈，W.（2011）。用户生成内容对旅行者行为的影响：一项关于电子口碑对酒店在线预订影响的实证调查。计算机与人类行为，27（2），634–639。
Ye, Q., Li, H. Y., & Wang, Z. S. (2014). The influence of hotel price on perceived service quality and value in e-tourism: An empirical investigation based on online traveler reviews. 《酒店与旅游研究杂志》，38（1），23–39。
尹灿和彭问（2010）。基于语义关联的中文评论产品特征情感分析。人工智能和计算智能国际会议（第81–85页）。
Yuan, J., & Wu, Z. T. (2016). Research on e□ect of convergence between creative industry park and tourism based on network text: A case study of shenzhen dafen oil painting village. 资源开发与市场，32（5），631–636。
张丹（2014）。中国在线旅游预订行业营销战略面临的挑战：携程的案例研究。com和去哪儿。通用域名格式。商业与工程学院。
张德国（2014）。在线社交网络与现实社交网络的比较：新浪微博研究。工程中的数学问题，2014（6684），1–6。
张，F.,宋，Y.,蔡，W.,刘，S.,刘，S.,普约，S。，... 冯博士（2016a）。医学成像中相似性计算的成对潜在语义关联。IEEE

Transactions on Biomedical Engineering, 63(5), 1058–1069.
Zhang, Z., Zhang, Z., & Yang, Y. (2016b). The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior. *Tourism Management, 55*, 15–24.
Zhang, C., & Zhou, Q. (2018). Online Investigation of users' attitudes using automatic question answering. *Online Information Review, 42*(3), 419–435.
Zhao, X., Gao, X., Guo, J. a., & Tao, N. (2009). Research focus analysis based on the frequency of topic words and g-index. *Library & Information Service, 53*(2), 59–61.
Zhao, X., Liang, W., Xiao, G., & Law, R. (2015). The influence of online reviews to online hotel booking intentions. *International Journal of Contemporary Hospitality Management, 27*(6), 1343–1364.



< au id="1"/ > **Zhiping Hou**, Ph.D., is a Professor and Associate Dean in Business School of Guilin University of Technology, China. He received his PhD in information science from Nanjing University, Nanjing, China, in 2014. His research interests include data mining, social media marketing, social network, information management, and tourism electronic commerce.



< au id="2"/ > **Fasheng Cui**, is a Master student in Business School of Guilin University of Technology, China. His research interests include social network analysis and electronic commerce.

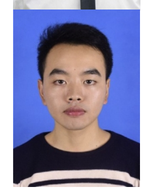
< au id="3"/ > **Yongheng Meng**, Ph.D., is a Professor and Associate Dean in Business School of Guilin University of Technology, China. His current research interests include travel information search and economic development.

< au id="4"/ > **Tonghui Lian**, Ph.D., is Associate Professor in the School of Business Administration at Nanjing University of Finance and Economics, China. His research interests include information management and electronic commerce.

< au id="5"/ > **Caihua Yu**, Ph.D., is Associate Professor in the School of Economics and Management at Nanjing University of Information Science and Technology, China. Her research interests include information management and electronic commerce.

生物医学工程学报，63（5），1058–1069。
张志强，张志强，杨勇（2016b）。专家身份的力量：网站认可的专家评论如何影响旅行者的在线评级行为。旅游管理，55, 15–24。
Zhang, C., & Zhou, Q. (2018). Online Investigation of users' attitudes using automatic question answering. 在线信息评论，42（3），419–435。
Zhao, X., Gao, X., Guo, J. a., & Tao, N. (2009). Research focus analysis based on the frequency of topic words and g-index. 图书馆和信息服务，53（2），59–61。
赵，X.,梁，W.,肖，G.,罗，R.（2015）。在线评论对在线酒店预订意向的影响。《国际当代酒店管理杂志》，27（6），1343–1364。

支平候，博士，中国桂林工学院商学院教授、副院长。2014. ，他在中国南京大学获得了信息科学博士学位。他的研究兴趣包括数据挖掘、社交媒体营销、社交网络、信息管理和旅游电子商务。



法胜翠是中国桂林工学院商学院的硕士研究生。他的研究兴趣包括社交网络分析和电子商务。

雍恒梦，博士，中国桂林工学院商学院教授、副院长。他目前的研究兴趣包括旅游信息搜索和经济发展。

童慧莲，博士，南京大学中国财经学院工商管理系副教授。他的研究兴趣包括信息管理和电子商务。

蔡华雨博士是南京信息工程大学中国经济管理学院的副教授。她的研究兴趣包括信息管理和电子商务。