



Selecting the appropriate leading journal in Hospitality and Tourism research: a guide based on the topic-journal fit and the JCR impact factor

Ramón Barrera-Barrera¹

Received: 15 March 2021 / Accepted: 27 January 2022 / Published online: 21 February 2022
© The Author(s) 2022

Abstract

Selecting the appropriate academic journal is a priority issue for researchers in the process of publishing a manuscript. If researchers could quantify the research topic in terms of its fit with the journal requirements before the submission of a paper, then the assessment of journal suitability could be much easier. Basing any decision on journal impact factors alone might obviously result in a mismatch, eventual rejection and a consequent loss of time. Taking the twelve leading Tourism and Hospitality journals as a reference, the main research topics mentioned in the abstracts of 20,381 articles are identified, using the Latent Dirichlet Allocation algorithm and other text-mining techniques running the R programming language. Subsequently, a quantitative measure of the fit of the research topics in each journal is offered according to their frequency of occurrence. The results suggested that the importance of the topic-journal fit with respect to the impact factor depended on the variance of the fits among the journals. Finally, a guide of the most suitable journals for the topics is presented, based on the JCR impact factor and the fit of the topic. Some recommendations are likewise offered on the use of this methodology and its limitations.

Keywords Tourism and Hospitality research · Bibliometric analysis · Text-mining · Topic modelling

Introduction

The scientific production of tourism papers has significantly expanded over recent years. In 2005, the number of papers within this field on the Web of Science Core Collection was around 5000 articles. As of January 7, 2021, the search for the term “tourism” in this database retrieved more than 91,866 documents, of which 63,316 were articles published in over 8800 academic journals. Most of the articles were published in journals within the “Hospitality, Leisure, Sport & Tourism” category (43.22%). The others were published in journals of other categories: “Environmental Studies” (10.1%), “Management” (9.4%),

✉ Ramón Barrera-Barrera
rbarrera@us.es

¹ Management and Marketing Department, University of Seville, Ramón y Cajal, 1, +3441018, Seville, Spain

“Environmental Sciences” (8.6%), “Geography” (6.3%), “Economics” (5.5%), “Sociology” (5.0%), “Business” (4.2%), “Green Sustainable Science Technology” (4.8%), and “Ecology” (2.4%). These results reflect the interdisciplinarity of tourism research (Oviedo-García, 2016).

Given the high number and the variability of academic journals in the field of tourism research, the proper selection of a journal is particularly important to researchers when submitting a paper for two main reasons: (1) the submission and the review process of a manuscript is time consuming; and, (2) the submission of a manuscript is only to one journal at any one time (Knight & Steinbach, 2008). In this way, dissemination of the results of an investigation may be delayed by an inappropriate choice of journal.

Two important factors stand out when selecting a journal for the submission of a manuscript: (1) the journal impact factor; and, (2) the fit between the journal and the topic of the manuscript (van Teijlingen & Hundley, 2002; Knight & Steinbach, 2008; Walters & Markgren, 2019). Currently, the impact factor is often considered as a benchmark for the prestige or reputation of a journal and it is used to rank and to compare journal quality, based on the premise that a higher citation rate will indicate higher quality (Klinger et al., 2005; Kurmis, 2003). Publishing in high-impact journals has a double benefit. At an individual level, researchers who wish to pursue an academic career are expected to publish in these journals and, at an institutional level, the University or the research center is likely to attract greater financial resources and to occupy a prominent position in international rankings.

The concept of the journal impact factor was introduced in 1955 by Dr. Eugene Garfield and its calculation is based on a ratio: the numerator represents the number of citations from the current year that refers to items published over the previous two years; the denominator represents the number of citable materials (source items) published over the same two years (Garfield, 2006). There can be several variants of Garfield’s index, depending on the formulation of the ratio and its numerator and denominator. The best-known indices within the international scientific community are the Journal Citation Reports (JCR), provided by the Web of Science (WOS) database of Clarivate Analytics, and the Scimago Journal Rank (SJR), provided by the Scopus database of Elsevier. The differences between these indices are mainly due to the different scientific databases from which the citations are sourced and the methodologies that are employed for their estimation (Falagas et al., 2008).

According to Knight and Steinbach (2008), the fit between the journal and the manuscript is probably the most important point in selecting a journal. In fact, if a paper is not in line with the journal content, some editors will reject it outright without a peer-review process (desk-rejection). A manuscript presents a good fit if the research topic reflects the research area of the journal and complies with the style and other formal requirements (e.g., article length, coherency of literature review, statistical analyses and practical value for practitioners) (Knight & Steinbach, 2008). However, the question of how to assess whether a research topic fits the requirements of an academic journal is not as straightforward as might appear. In a first approximation, authors should consult the scope and the range of topics noted on the journal website. If the journal defines its lines of research in a generic way, authors may when adopting an interdisciplinary approach have doubts over whether the topic of the manuscript fits the editorial line. Besides, other journals specify the range of topics with more precision, but authors may be unaware of the weight attributed to one topic among the others that the journal publishes.

The objective of this investigation is twofold. A tool will be proposed to quantify the fit of a research topic for a series of Tourism and Hospitality journals according to the

frequency of occurrence of that topic. It will also set out a guide to the most appropriate journals based on the JCR impact factor and the topic-journal fit. To do so, text-mining techniques will be used to perform a bibliometric analysis of the abstracts of all the articles that have been published in the Tourism and Hospitality journals with the highest impact factor (Q1) indexed in the Web of Science within the “Hospitality, Leisure, Sport & Tourism” category. The literature review will be presented in Sect. 2; in Sect. 3, the research methodology will be described; the results will then be discussed in Sect. 4; and the conclusions will finally be presented in Sect. 5.

Literature review

The term bibliometrics was first introduced by Pritchard (1969) as “the application of mathematical and statistical methods to books and other means of communication” (p. 349). Different tools are used to perform a bibliometric analysis—some using very simple mathematical operations and others of greater complexity, which require greater knowledge of statistics and computing. Koseoglu et al. (2016) identified three different traditional bibliometric methods: (1) review techniques, (2) evaluative techniques, and (3) relational techniques.

Review techniques refer to structured literature reviews, systematic literature reviews, and meta-analyses/reviews (Zupic & Cater, 2015). Evaluative techniques can be classified into three groups (Hall, 2011): (1) productivity measures (e.g., number of papers per author, journal, institution or country); (2) impact metrics (e.g., number of citations per paper, author and journal); and (3) hybrid metrics, such as the average number of citations per paper or the h-index. Relational techniques can be divided into four categories: co-citation analysis, co-word analysis, co-authorship analysis, and bibliographic coupling (Zupic & Cater, 2015). Co-citation analysis seeks to determine the similarity between documents, authors, and journals (McCain, 1990). Co-citation analysis occurs when two papers, authors, or journals are jointly cited in the same paper. Co-word analysis or co-occurrence of words quantifies the extent to which two words occur in the same title, abstract, or keyword list (Zupic & Cater, 2015). Co-author analysis examines the networks that authors create when collaborating on scientific articles (Acedo et al., 2006). Finally, bibliographic coupling uses the number of references shared by two documents as a measure of their similarity (Zupic & Cater, 2015).

Most bibliometric studies in the Tourism and Hospitality literature are based on evaluative techniques, and their objective is to offer rankings of authors, articles, journals, and academic institutions using productivity and impact indicators, mainly the number of articles and number of citations (Table 1). Bibliometric analyses based on relational techniques are less common and have focused mainly on co-citation analysis and co-word analysis. The co-citation analyses generate different graphs with VOSviewer© software (Van Eck & Waltman, 2010) that represent the network connections of journals, authors or documents cited in the references. Performed with keywords, the co-word analysis identifies those used with the highest frequency.

In recent years, the bibliometric analysis of literature has improved with the use of various text-mining techniques (Yang et al., 2018). Specifically, text-mining techniques in most bibliometric studies are applied to abstracts in a way that can identify research topics through word frequency analysis, relationships between words, and topic modelling. For example, Jiang et al. (2016) analyzed 1726 articles related to hydropower

Table 1 Bibliometric studies in tourism research

Authors	Number of articles and journals	Period of analysis	Bibliometric method	Results
Palmer et al. (2005)	1790 articles in 12 tourism journals	1998–2002	Analysis based on productivity measures	Methodological approaches used in tourism research
Benckendorff and Zehrer (2013)	2486 articles in <i>Annals of Tourism Research</i> , <i>Journal of Travel Research and Tourism Management</i>	1996–2010	Analysis based on impact metrics and co-citation analysis of authors and documents	Rankings of articles, authors, journals, and books with most cites
García-Lillo et al. (2016)	847 articles in <i>International Journal of Hospitality Management</i>	2008–2014	Analysis based on impact metrics and co-citation analysis of documents	Ranking of research papers with the most cites in <i>IJHM</i> and the academic journals where those papers were published
Durán-Sánchez et al. (2017)	238 articles on wine tourism in 122 journals	Until 2014	Analysis based on productivity measures	Rankings of most productive authors and journals
Leung et al. (2017)	406 articles on social media research in 16 business and tourism journals	2007–2016	Co-citation analysis of documents and co-word analysis of keywords	Evolution of research topics in social media
Strandberg et al. (2018)	292 articles in <i>Tourism and Hospitality Research</i>	2000–2014	Analysis based on productivity measures and impact metrics	Rankings of articles and authors with the most cites, most productive authors, and most important research themes
Shen et al. (2018)	4654 articles in 32 tourism journals	2002–2011	Analysis based on productivity measures and co-citation analysis of countries	Rankings of most productive countries and regions, most popular topics of research and types of tourism by country
Okumus et al. (2018)	462 articles on food and gastronomy in hospitality and tourism in 16 tourism journals	1976–2016	Analysis based on productivity measures	Rankings of most productive journals, authors, institutions and countries, and main research topics and methodologies

Table 1 (continued)

Authors	Number of articles and journals	Period of analysis	Bibliometric method	Results
Martínez-López et al. (2018)	All articles in European Journal of Marketing	1967–2017	Analysis based on productivity measures and impact metrics Co-citation analysis of journals, authors, and institutions Bibliographic coupling of institutions and countries Co-word analysis of keywords	Rankings of most productive authors, institutions, and countries Ranking of articles with most cites Rankings of articles, authors, universities, and countries with most cites in EJM documents
Martorell et al. (2019)	1996 articles in International Journal of Hospitality Management	1982–2017	Analysis based on impact metrics Co-citation analysis of journals and authors Bibliographic coupling of countries Co-words analysis of keywords	Ranking of articles with most cites Ranking of most frequently cited papers in articles published in the IJHM Rankings of authors, universities, countries, and journals that cite IJHM documents with greater frequency Graphical analyses of bibliographic coupling of countries, co-occurrence of author keywords, co-citation of journals and co-citation of authors
Zach et al. (2019)	318 articles on tourism research on technology and innovation	1982–2018	Analysis based on topic modelling on abstracts with LDA algorithm	Ten topics that represent different research areas
Qian et al. (2019)	2024 papers published in Tourism Management	2001–2017	Analysis based on productivity measures and impact metrics Co-citation analysis of authors Co-word analysis of keywords	Rankings of most productive authors. Rankings of most cited documents and journals by TM papers Graphical analyses of co-authorship and co-occurrence of author keywords
Rodríguez-López, et al. (2020)	740 articles on restaurant research in 102 journals	2000–2018	Analysis based on productivity measures, impact metrics and co-word analysis of keywords	Ranking of most productive authors, and journals. Graphical analyses of main research topics
Loureiro et al. (2020)	56 journal papers and 325 conference proceedings	2010–2020	Analysis based on text-mining techniques and a citation network analysis	Rankings of journal and proceeding papers with most cites Most important research topics

Table 1 (continued)

Authors	Number of articles and journals	Period of analysis	Bibliometric method	Results
Leong et al. (2021)	693 articles published in Tourism Review	2001–2019	<p>Analysis based on productivity measures and impact metrics</p> <p>Co-citation analysis of journals and authors</p> <p>Bibliographic coupling of authors</p> <p>Co-word analysis of keywords</p>	<p>Rankings of most productive authors, institutions, and countries</p> <p>Ranking of authors and articles with most cites</p> <p>Rankings of authors, institutions and countries that cite articles from TR</p> <p>Graphical analyses of bibliographic coupling of authors, co-occurrence of author keywords, co-citation of journals and co-citation of authors</p>

research between 1994 and 2013. They identified the terms with a high frequency of occurrence, as well as the main research topics, and their development. Gómez-Barroso et al. (2017) collected all papers published between 1976 and 2016 in the journal *Telecommunications Policy*. Text-mining techniques were used to identify the key topics, the dominant combinations of concepts, and the main areas of research. Ding et al. (2018) analyzed 1600 research articles on building energy savings between 1973 and 2016. The results identified the research trends at different stages using various timestamps. Yang et al. (2018) collected 1000 abstracts from Google Scholar, 249 full texts from ScienceDirect, and six articles published in 2016 or 2017 on solar forecasting. The results showed the most important and the most frequent abbreviations, as well as key innovations for recent advances in solar forecasting. Demeter et al. (2019) analyzed 566 papers on inventory research in the *International Journal of Production Economics* between 1994 and 2016, identifying the main topics, terms, research areas and their evolution over time. More recently, Correia et al. (2020) performed a bibliometric analysis of the last 20 years of research on virtual reality and augmented reality in tourism contexts. The full texts of 56 journal papers and 325 conference proceedings were analyzed using a topic modeling approach with the statistical package R. The most important latent topics that emerged from the journal papers and the conference proceedings were shown separately.

The use of text-mining techniques also plays an important role in analyzing the text of online consumer reviews, an important type of electronic word-of-mouth (eWOM) recommendations (Litvin et al., 2008). These comments on products and services, whether positive or negative, are used by other consumers to reduce any risks associated with purchase decisions and have great influence on consumer shopping behavior (Mauri & Minazzi, 2013; Sparks & Browning, 2011; Zhang et al., 2010). Particularly in the tourism industry, online consumer reviews are a very important source of information, due to the unique characteristics of hospitality services (e.g., intangibility, heterogeneity, and inseparability), especially when the consumer has no previous experience (Litvin et al., 2008). In this context, previous studies have analyzed the influence of online reviews on firm performance (e.g., sales and prices of hotel rooms) (Ogüt & Tas, 2012; Ye et al., 2009), the determinants of online reviews credibility (Munzel, 2016; Xie et al., 2011) or how firms manage online responses to consumer reviews (Mauri & Minazzi, 2013; Xie et al., 2014).

One of the most common techniques used to analyze the text of online reviews is topic modeling, especially the Latent Dirichlet Allocation (LDA) algorithm. This methodology is used to discover the underlying topics of massive data volumes of unstructured text (Guo et al., 2017). In the hospitality and the tourism literature, several researchers have employed LDA to analyze online reviews. For example, Guo et al. (2017) collected 266,544 online reviews for 25,670 hotels located in 16 countries with the objective of identifying the most important factors influencing consumer satisfaction. The LDA algorithm uncovered 19 dimensions such as location of hotel, breakfast, hotel staff, car parking, check in and check out, etc. The authors also demonstrated that the relative importance of the dimensions differed according to the gender and the age of the reviewers. Bi et al. (2019) analyzed 24,276 online reviews from two five-star hotels from TripAdvisor with the LDA algorithm. The results showed that the most important attributes of hotel performance were value, location, room, cleanliness, service, food/drink, check in/out, facility and Wifi/Internet. Song et al. (2021) collected 20,476 online reviews from 18,387 TripAdvisor users to understand how visitors experience public spaces on the Las Vegas Strip. The results revealed 30 distinct topics and the most important ones were exploring different hotels, night scenes, people watching, and walking long distances.

Other researchers used a text-mining method called Structural Topic Model (STM) (Roberts et al., 2014), a variant of the LDA model. For example, Hu et al. (2019) analyzed 27,864 hotel reviews in New York City from Tripadvisor with this algorithm to identify the antecedents of hotel customers' dissatisfaction. The most negative topics in the reviews were severe service failure, dirtiness, booking and cancellation, noise, etc. The research also revealed how customer complaints varied across different hotel grades. Finally, Korfiatis et al. (2019) also applied STM to 557,208 Tripadvisor online reviews from airline passengers. The analysis was performed with R statistical software and the results suggested 20 topics relating to customer satisfaction, such as business class, value for money, low cost, delays, staff, check-in and airport experiences, etc.

Research methodology

Text mining is an emerging sub-discipline of data mining, which aims to extract useful information and knowledge from unstructured and semi-structured text (Miner et al., 2012). The idea behind text mining is the transformation of text into numerical data (Silge & Robinson, 2017). There is a wide range of text-mining analyses—some relatively simple and others based on more complex algorithms, as previously mentioned—to perform various interrelated tasks: information retrieval, document clustering, information extraction, natural language processing, sentiment analysis, and social network analysis (Kwartler, 2017).

This paper uses R software to perform the text-mining analysis for three main reasons. Firstly, this software enables rapid processing of massive texts. Secondly, text-mining analyses generally require the use of a programming language that can be adapted to the characteristics of the database and the nature of the research. And thirdly, there are packages and libraries specifically designed with this programming language for processing text databases. Figure 1 presents the research methodology that was used to perform the analyses and follows the usual workflow for text mining. It details the main functions and

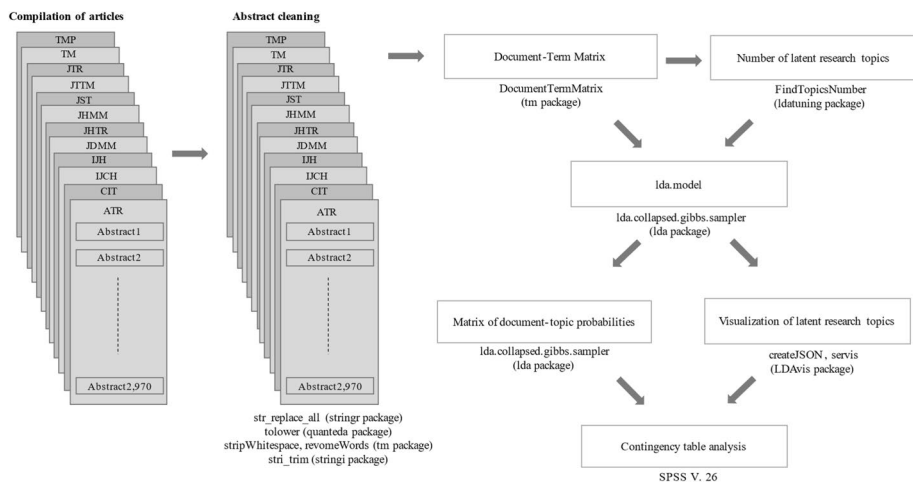


Fig. 1 Text-mining analysis flowchart. *Source:* Adapted from Silge and Robinson (2017)

packages used for the purposes of this study. The statistical software package SPSS V. 26 and the Excel spreadsheet were also used to analyze the data.

Data collection

The first step of the data-collection procedure involved the compilation of the articles from the following academic journals: Annals of Tourism Research (ATR), Current Issues in Tourism (CIT), International Journal of Contemporary Hospitality Management (IJCHM), International Journal of Hospitality Management (IJHM), Journal of Destination Marketing & Management (JDMM), Journal of Hospitality & Tourism Research (JHTR), Journal of Hospitality Marketing & Management (JHMM), Journal of Sustainable Tourism (JST), Journal of Travel & Tourism Marketing (JTTM), Journal of Travel Research (JTR), Tourism Management (TM) and Tourism Management Perspectives (TMP). All the articles under study were extracted from the Scopus database, because it has records of all articles published in these journals since their establishment. In contrast, articles are only indexed in the Web of Science from the year in which the journal was first indexed in this database. The articles were downloaded on 1 October 2020, with a total of 20,381 records, and only those articles with the available abstract were selected. The result was a database with 18,798 articles (Table 2). The abstracts of the articles were compiled as the primary data for text-mining, because they are a better representation of the main contents of the paper than various keywords, and the excessively lengthy time spans and resources needed for full-text analysis ruled out that option (Gómez-Barroso et al., 2017; Jiang et al., 2016).

Abstract cleaning

The process of abstract cleaning involved upper-to-lower case conversion and removal of numbers, punctuation marks, whitespaces, line breaks and stop words (Miner et al., 2012). The process of filtering out the stop words removed the most trivial words such

Table 2 Number of articles in the database

Journal	Number of articles published	Number of articles with abstracts	Mean length of words in abstract	Standard deviation
ATR	2970	2045	116.66	16.59
CIT	1224	1217	168.11	42.66
IJCHM	2161	2091	204.46	93.44
IJHM	2645	2515	137.08	40.11
JDMM	445	428	163.61	42.80
JHTR	1302	1160	136.26	75.42
JHMM	488	487	143.72	30.67
JST	1427	1397	184.93	33.27
JTTM	1409	1356	130.03	35.95
JTR	1911	1800	121.96	36.18
TM	3704	3620	137.34	41.45
TMP	695	682	143.43	30.01
Total	20,381	18,798		

as pronouns, prepositions, and articles (*e.g.*, “the”, “for”, “these”, “in” etc.); and common words in abstracts, the meanings of which were not relevant for the purposes of this study, by comparing each word to a stop word list (*e.g.*, “article”, “results”, “show”, “data”, “research”, “literature”, “sample”, etc.). The main functions used in this phase were: `str_replace_all`, `tolower`, `stripWhitespace`, `revomeWords` and `stri_trim` from various R packages.

LDA topic modelling

Latent Dirichlet Allocation is an unsupervised learning method and one of the most common algorithms for topic modelling (Guo et al., 2017; Köseoglu, 2020). It treats each document as a mixture of topics, and each topic as a mixture of words (Blei et al., 2003). The algorithm estimates two probability matrices: (1) a matrix of term-topic probabilities where each entry is the probability of term i being generated from topic j ; and (2) a matrix of document-topic probabilities where each entry is the probability of terms from document i being generated from topic j . The LDA model was estimated using the collapsed Gibbs sampling method, following the process described by Kwartler (2017).

The number of desired topics have to be entered in the LDA algorithm before the model is run. The function `FindTopicsNumber` in the `ldatuning` package (Nikita, 2020) calculates the value of the metric “CaoJuan2009” for a different number of topics taking the Document-Term Matrix (DTM) as its input (a matrix where each row represents one abstract, each column represents one token and each value contains the number of times that token occurred in that abstract). Small values of this metric are indicative of a possible optimal number. Finally, the `LDavis` package (Sievert & Shirley, 2014) was used to visualize the LDA models that were estimated.

In this study, the LDA algorithm was applied to the entire corpus of abstracts with the purpose of finding the main latent research topics in leading Tourism and Hospitality journals. The latent research topics identified were labeled using the words most likely to be generated by each topic. The matrix of document-topic probabilities was used to create a contingency table showing the number of articles from each journal assigned to the different research topics.

Contingency tables analysis

The contingency tables analysis was used to determine whether there was a statistically significant relationship between the topics and the journals given the categorical nature of the variables (Hair et al., 1999). The Chi-square test of independence contrasts the hypothesis of the independence of the variables against the alternative hypothesis that the distribution of one variable will differ from the various levels of the other (Hair et al., 1999). In turn, the corrected standardized residuals indicate whether the difference between the observed frequency and the expected frequency is statistically significant. At a confidence level of 95%, those residuals greater than 1.96 will indicate that the frequency is greater than expected, if the variables are independent, while residuals less than -1.96 indicate that the frequency is less than expected under the condition of independence. The corrected standardized residuals were taken as a reference, to measure the degree of fit of the research topics in each journal.

Results

Main latent research topics in leading Tourism and Hospitality journals

Once the cleaning process of the abstracts was complete, the LDA algorithm was used to identify the main latent research topics. A document-term matrix was created for use as an input of the FindTopicsNumber function, which calculates the values of the metric “CaoJuan2009” for a different number of topics (Fig. 2). However, the determination of an optimal number of topics is not easy as the metric “CaoJuan2009” can indicate different solutions. Furthermore, choosing a small number generates very generic topics, while too high a number can produce overly specific topics. The LDA algorithm was performed on 18, 24, and 28 topics. Following the recommendations of Sievert and Shirley (2014), the solution of 18 topics was chosen for easy interpretation and for clearly distinguishing between each one without overlaps. Finally, the 18 topics were plotted with the LDavis package for their visualization.

Figure 3 shows the latent research topics identified (left-hand side) and the most representative terms of the topic ‘social media’ (right-hand side). The topics are represented with circles and the inter-topic distances are computed using multidimensional scaling, by using first and second Principal Component Analysis (PCA). The area of the circles indicates the prevalence of the topics that are sorted in descending order. The most useful terms for interpreting a topic appear when that topic is selected. The overall frequency

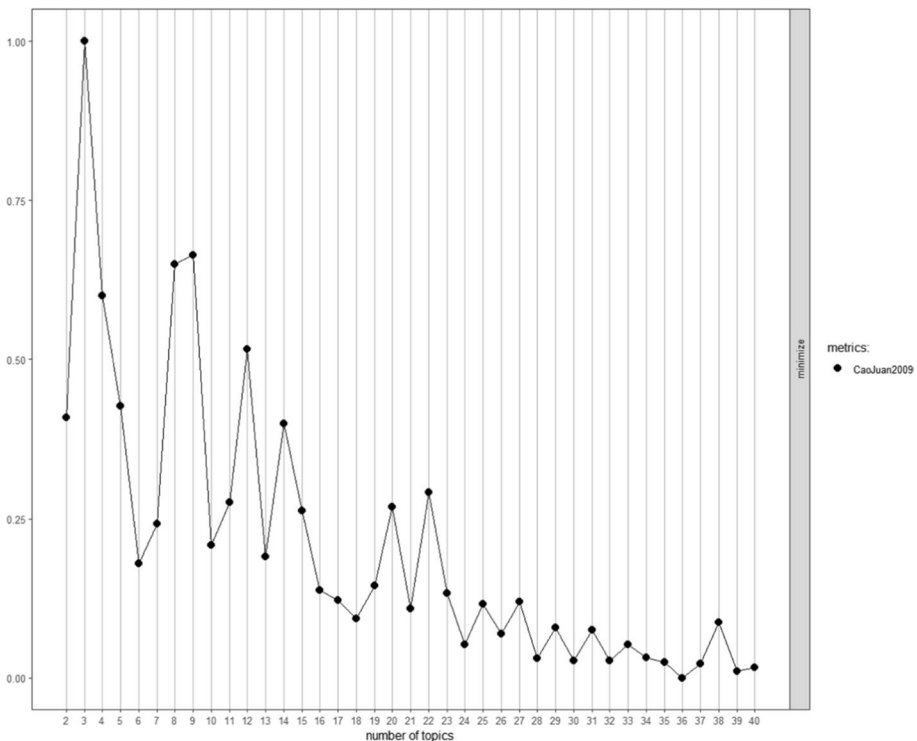


Fig. 2 Selecting the optimal number of topics

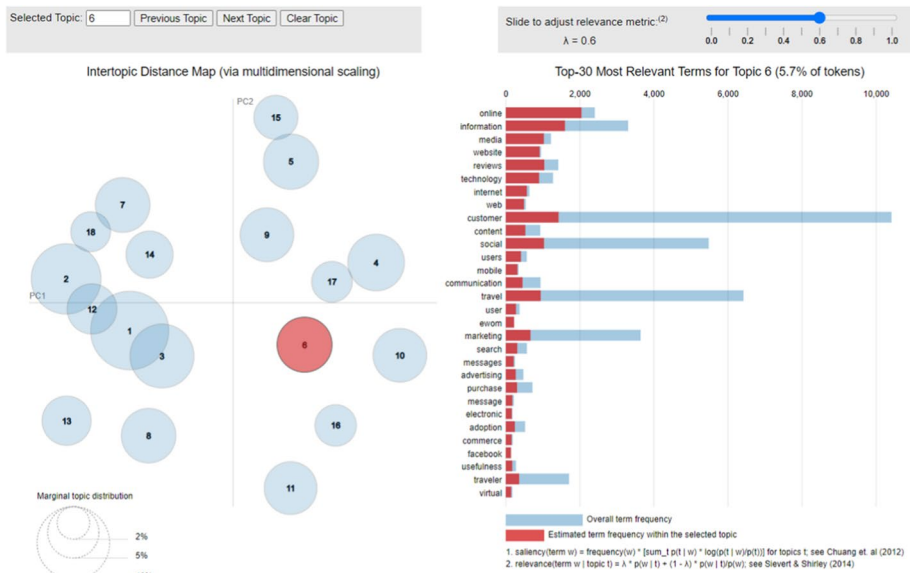


Fig. 3 Main latent research topics in leading Tourism and Hospitality journals

(grey bars) and the estimated frequency of each term (red bars) within the selected topic define the most representative terms of each topic. However, Sievert and Shirley (2014) suggested that if the meaning of each topic is based on the most probable terms in that topic (red bars), then the problem will be that common terms in the abstracts will appear at the top for multiple topics, making it hard to differentiate the meanings of those topics. These authors therefore proposed a measure called “relevance” for ranking the terms, suggesting $\lambda=0.6$ as the optimal value for topic relevance. The latent research topics were labeled with the words that were most likely to be linked with each topic (Table 3).

The LDA algorithm also estimated a document-topic probabilities matrix with 18,798 rows and 18 columns. This matrix was used to assign each article to the latent research topic with the highest probability of belonging. Figure 4 shows the probability distributions of the articles from each journal on the different topics. As an example, most of the articles from the ATR were assigned to four topics: sustainable tourism, tourism economic, cultural tourism, and tourist experience. The total number of articles from each journal assigned to the eighteen research topics appear in Table 4.

Measuring the topic-journal fit

The degree of association between the research topic and the journal cannot solely be based on its observed frequency, since a topic may have a higher frequency in one journal than in another, due to the simple fact that the first might publish a greater number of articles than the second. Therefore, the corrected standardized residuals from the contingency table analysis were used to measure the topic-journal fit. The contingency table analysis reported a statistically significant chi-square value (χ^2 : 12,324.187; p : 0.000). This result suggested a relationship of dependency between the research topics and the journals. The analysis of

Table 3 Main latent research topics and most relevant terms

N°	Latent research topic	Most relevant terms:
1	Sustainable tourism	Tourism, sustainable, community, stakeholder, sustainability, social, network, policy, local, governance, collaboration, destination, rural, planning, actors, entrepreneurs, power, economic, theory, collaborative, poverty, political, global, entrepreneurship, area, organizations, agenda, change, principles, goals
2	Tourism economic	Tourism, crisis, economic, government, region, policy, regional, planning, public, tourist, political, local, destination, island, rural, disaster, area, international, world, economy, national, crises, south, market, state, crime, border, foreign, resident, disasters
3	Destination image	Destination, tourist, image, visitor, travel, segment, motivation, choice, shopping, trip, market, traveler, visit, behavior, risk, cluster, attribute, attraction, marketing, distance, segmentation, tourism, clusters, visiting, perception, film, international, push, competitiveness, stay
4	Service quality	Service, customer, quality, hotel, satisfaction, attribute, guest, luxury, measurement, validity, reliability, expectations, performance, criteria, manager, Importance-Performance Analysis (IPA), delight, perception, instrument, hospitality, items, failure, experience, staff, delivery, provider, providers, cleanliness, dissatisfaction, loyalty
5	Hotel performance	Firm, performance, hotel, innovation, financial, efficiency, company, strategic, competitive, manager, restaurant, capital, hospitality, market, corporate, franchising, small, productivity, profitability, stock, chains, franchise, lodging, size, outsourcing, capabilities, risk, cost, ownership, advantage
6	Social media	Online, information, media, website, reviews, technology, internet, web, customer, content, social, users, mobile, communication, travel, user, electronic Word of Mouth (eWOM), marketing, search, messages, advertising, purchase, message, electronic, adoption, commerce, Facebook, usefulness, traveler, virtual
7	Tourism demand	Demand, tourism, forecasting, economic, casino, income, arrivals, gaming, expenditure, forecasts, time, gambling, run, accuracy, series, rate, panel, estimates, forecast, estimated, tourist, index, expenditures, seasonal, estimate, elasticity, output, seasonality, estimation, econometric
8	Tourist experience	Experience, cruise, tourist, volunteer, tourism, motivation, travel, adventure, pilgrimage, dark, home, religious, life, memory, memorable, backpackers, social, personal, Muslim, volunteers, spiritual, space, constraints, people, holiday, emotions, leisure, ship, phenomenon, experiential
9	Institutional education	Hospitality, student, training, education, manager, career, women, gender, skills, employment, workers, female, university, competencies, men, club, educational, courses, staff, teaching, educators, male, school, graduates, employee, workforce, curriculum, employers, working, harassment

Table 3 (continued)

N°	Latent research topic	Most relevant terms:
10	Job satisfaction	Employee, job, organizational, leadership, service, commitment, behavior, satisfaction, turnover, front-line, emotional, stress, workplace, engagement, recovery, hotel, hospitality, performance, psychological, customer, organization, justice, conflict, supervisors, creativity, burnout, citizenship, supervisor, leader, empowerment
11	Resident's attitude	Intention, resident, attitude, green, behavioral, perceived, behavior, satisfaction, attachment, environmental, perception, revisit, theory, loyalty, emotions, pro, emotional, place, subjective, norms, Theory of Planned Behavior (TPB), norm, social, experiential, environmentally, community, quality, cognitive, tourist, mouth
12	Ecotourism	Park, ecotourism, visitor, conservation, wildlife, natural, area, environmental, national, recreation, beach, nature, site, farm, agritourism, whale, marine, ecological, attraction, species, ecotourist, watching, local, coastal, sustainable, benefit, animals, animal, trail, resources
13	Cultural tourism	Heritage, cultural, authenticity, site, culture, tourist, ethnic, indigenous, identity, city, place, tourism, world, history, creative, museum, attraction, representations, landscape, historic, gaze, language, aboriginal, art, museums, UNESCO, World Heritage Site (WHS), visitor, space, discourse
14	Medical tourism	Travel, tour, medical, leisure, market, convention, operators, package, age, tours, vacation, spa, traveler, agents, segment, older, meeting, agencies, outbound, trips, Visiting Friends and Relatives (VFR), senior, health, product, marketing, family, wellness, timeshare, household, meetings
15	Revenue management	Hotel, price, room, Airbnb, accommodation, pricing, revenue, rate, guest, peer, occupancy, sharing, location, hosts, booking, Revenue Management (RM), property, listings, discount, lodging, properties, home, pay, star, bed, market, competition, economy, Revenue Per Available Room (RevPAR), rating
16	Corporate Social Responsibility (CSR)	Brand, Corporate Social Responsibility (CSR), festival, loyalty, customer, co, creation, equity, branding, sport, corporate, mega, trust, games, Customer Relationship Management (CRM), marketing, responsibility, olympic, exhibition, organizers, social, love, assistant, exhibitor, engagement, expo, image, hosting, personality, reputation
17	Restaurant industry	Food, restaurant, wine, dining, customer, menu, culinary, foods, tipping, chefs, safety, foodservice, healthy, patrons, casual, eating, service, diners, menus, health, beverage, fast, cuisine, meal, servers, winery, wineries, waiting, fine, server
18	Climate change	Climate, change, air, recovery, water, transport, carbon, airline, emissions, ski, airlines, airport, weather, adaptation, airports, flight, passengers, environmental, passenger, waste, winter, cost, traffic, aviation, gas, transportation, emission, snow, pollution, greenhouse

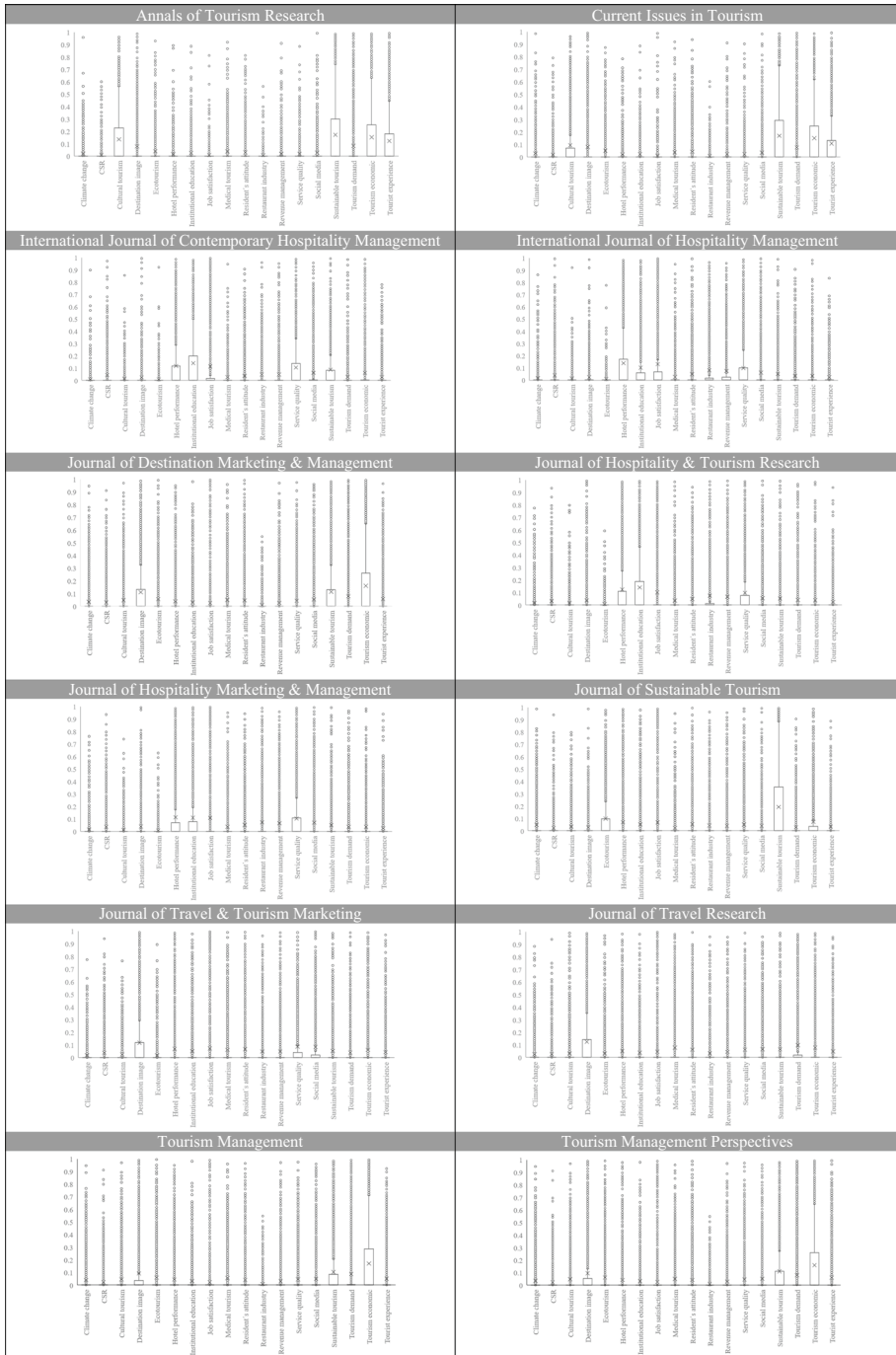


Fig. 4 Probability distributions of the articles on the latent research topics

Table 4 Contingency table of the latent research topics and the journals

Latent research topic	ATR	CIT	IICHM	IJHM	JDMM	JHTR	JHMM	JST	JTTM	JTR	TM	TMP	Total articles
Climate change	24	40	10	37	4	17	5	113	12	38	127	13	440
CSR	12	24	79	87	32	30	34	13	56	35	64	12	478
Cultural tourism	327	87	12	10	19	12	5	56	28	58	151	40	805
Destination image	170	107	29	30	111	60	33	34	329	343	360	76	1682
Ecotourism	61	69	9	7	7	4	3	278	16	72	234	50	810
Hotel performance	19	26	267	401	3	138	27	25	24	27	150	19	1126
Institutional education	31	16	339	293	1	258	17	7	5	16	88	20	1091
Job satisfaction	9	14	278	378	1	76	62	11	19	12	83	26	969
Medical tourism	67	28	40	40	5	46	16	12	117	183	185	20	759
Resident's attitude	54	67	74	111	42	54	37	83	115	115	136	28	916
Restaurant industry	4	6	114	235	1	76	29	6	18	10	21	7	527
Revenue management	30	32	91	188	1	58	20	5	34	44	84	21	608
Service quality	24	30	221	251	8	122	71	9	113	70	153	26	1098
Social media	39	54	141	160	31	47	77	11	160	125	185	39	1069
Sustainable tourism	366	233	177	102	73	56	20	536	58	116	387	102	2226
Tourism demand	180	91	46	72	13	47	11	19	53	246	328	41	1147
Tourism economic	355	167	127	75	46	34	6	129	120	184	700	74	2017
Tourist experience	273	126	37	38	30	25	14	50	79	106	184	68	1030
Total articles	2045	1217	2091	2515	428	1160	487	1397	1356	1800	3620	682	18,798

the standardized residuals pointed to the nature and degree of this statistical relation. The higher the value of a residual, the better the research topic fit with the journal. On the contrary, negative values of the residuals were indicative of a bad fit.

As shown in Table 5, papers related to sustainable tourism, ecotourism and climate change had high adjustments in JST. Similarly, papers on hotel performance, restaurant industry, job satisfaction or Hotel, Restaurant and Institutional Education (HRIs) also had a good fit in IJCHM, IJHM, and JHTR. Papers on destination image also had a good fit in JTTM, JTR, and JDMM. The excellent adjustment of the cultural tourism on ATR also stood out. Furthermore, the high residual variance values indicated that the research topics simultaneously presented a good fit in these journals and a bad fit in the rest.

From another perspective, the variance of the residuals in each journal is an indicator of their degree of specialization. In this sense, the journals of this study can be divided into two groups. On the one hand, the publications with the greatest variability of fit and greatest specialization were: IJHM, JST, ATR and IJCHM (in that order). On the other hand, the publications with less variance of fit and a multidisciplinary approach were: TMP, JDMM, CIT, JHMM, TM, JTTM, JTR and JHTR (in that order).

Finally, taking as a reference the values of the residuals (the topic-journal fit) and the JCR impact factor (2019), a guide to the most appropriate journals for the main research topics in Tourism and Hospitality is shown in Fig. 5.

Discussion of results and conclusions

The use of text mining analysis in the bibliometric field is gaining ground due to the new possibilities it offers. In this paper, this methodology has been used to quantify the fit of the main research topics found in the abstracts of the twelve leading journals in Tourism and Hospitality. This quantitative measure of the fit of the topic manuscript and the journal impact factor was designed so that authors may accurately assess journal suitability before the submission of a paper. The results that are discussed below and the main conclusions are then presented.

Table 5 Standardized residuals analysis between latent research topics and the journals

Latent research topic	ATR	CIT	IJCHM	IJHM	JDMM	JHTR	JHMM	JST	JTTM	JTR	TM	TMP	Variance
Climate change	-3.7	2.3	-6.0	-3.1	-1.9	-2.0	-1.9	14.8	-3.7	-0.7	5.2	-0.8	30.45
CSR	-6.0	-1.3	3.8	3.1	6.6	0.1	6.3	-4.0	3.9	-1.7	-3.3	-1.3	17.02
Cultural tourism	27.7	5.1	-8.9	-10.3	0.2	-5.6	-3.6	-0.5	-4.2	-2.3	-0.4	2.1	95.60
Destination image	-1.1	-0.2	-12.8	-14.6	12.5	-4.6	-1.7	-8.9	20.4	15.8	2.4	2.1	119.00
Ecotourism	-3.1	2.4	-9.3	-10.7	-2.8	-6.9	-4.1	29.8	-5.9	-0.7	7.1	4.0	116.18
Hotel performance	-10.2	-3.9	13.9	22.6	-4.7	8.8	-0.4	-6.9	-6.8	-8.4	-5.2	-3.6	103.71
Institutional education	-8.8	-6.9	21.6	13.5	-5.0	24.7	-2.2	-8.8	-8.9	-9.4	-9.7	-3.3	160.10
Job satisfaction	-10.2	-6.5	17.9	24.1	-4.7	2.2	7.7	-7.7	-6.5	-9.1	-8.7	-1.6	126.37
Medical tourism	-1.9	-3.2	-5.2	-6.7	-3.1	-0.1	-0.9	-6.3	8.9	13.9	3.6	-1.5	38.45
Resident's attitude	-3.0	1.1	-3.0	-1.2	4.8	-0.4	2.8	1.9	6.4	3.1	-3.5	-0.9	11.97
Restaurant industry	-7.6	-5.0	7.8	21.3	-3.3	8.0	4.3	-5.6	-3.4	-6.1	-9.0	-2.9	78.25
Revenue management	-4.8	-1.2	3.1	12.9	-3.5	3.5	1.1	-6.3	-1.6	-2.0	-3.5	-0.2	25.86
Service quality	-9.5	-5.2	9.8	9.5	-3.5	7.0	8.3	-8.6	4.1	-3.7	-4.6	-2.3	51.44
Social media	-7.8	-1.9	2.2	1.6	1.4	-2.5	9.8	-8.2	10.1	2.4	-1.7	0.0	31.97
Sustainable tourism	9.0	8.2	-5.1	-13.0	3.4	-7.6	-5.4	31.9	-8.9	-7.5	-2.4	2.6	145.67
Tourism demand	5.4	2.1	-7.9	-7.3	-2.7	-3.0	-3.6	-7.7	-3.5	14.1	8.3	-0.1	46.74
Tourism economic	10.3	3.5	-7.3	-13.5	0.0	-8.9	-6.9	-1.9	-2.3	-0.7	18.6	0.1	75.24
Tourist experience	16.6	7.7	-7.9	-9.4	1.4	-5.1	-2.6	-3.2	0.6	0.8	-1.2	5.2	50.86
Variance	109.21	21.99	103.84	174.13	21.60	64.57	25.00	160.60	59.81	59.95	50.05	6.07	

High residuals values are indicated in green and low residuals values appear in red $p < 0.05$ (corrected standardized residuals in absolute value greater than 1.96)
 $p < 0.01$ (corrected standardized residuals in absolute value greater than 2.58)

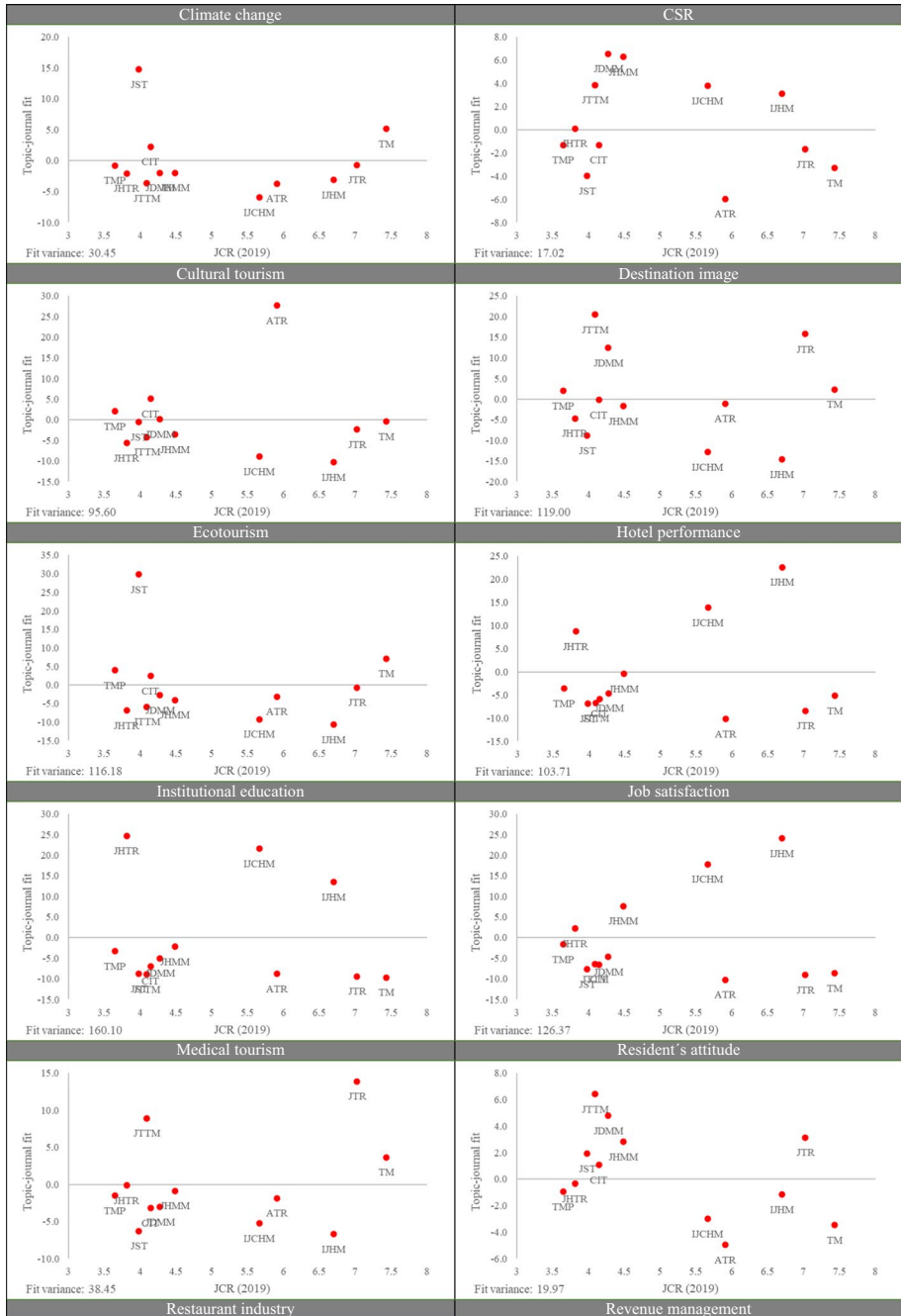


Fig. 5 Journal selection based on the topic-journal fit and the JCR impact factor

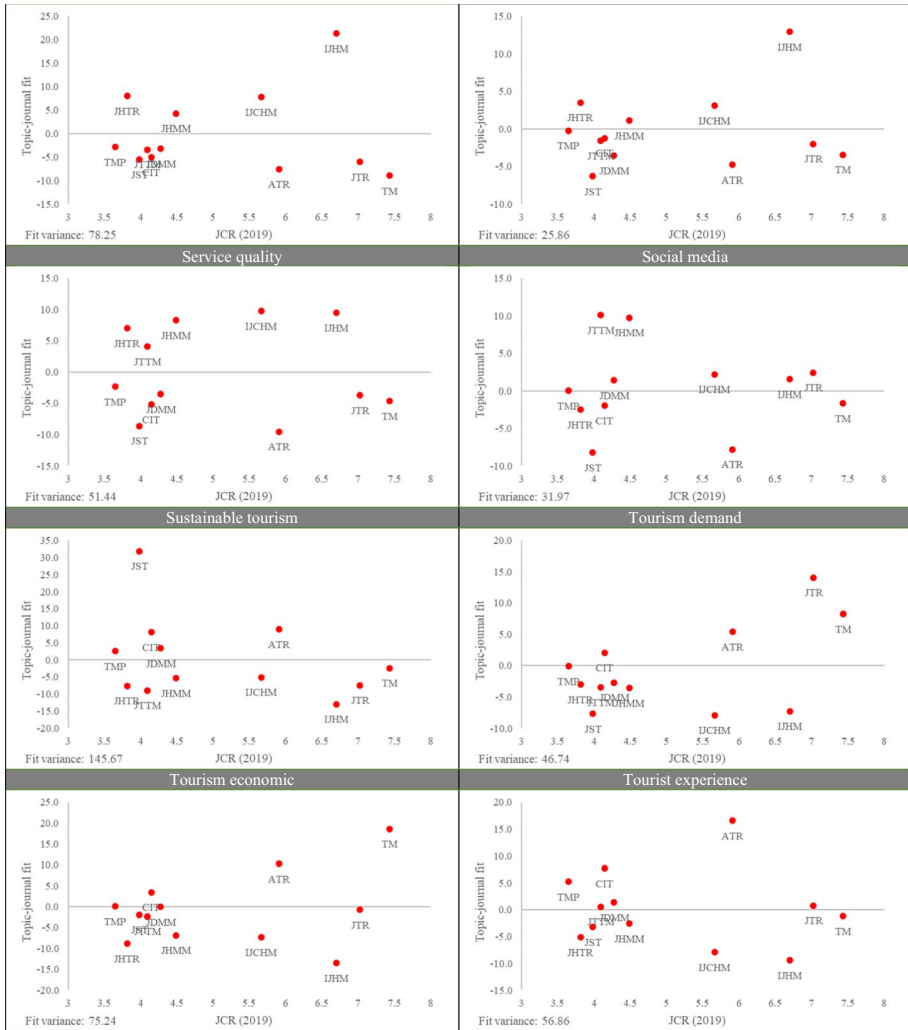


Fig. 5 (continued)

Firstly, the fit is a measure that depends on the research topic of each article and the referenced journals to which to submit the paper. The fit of the research topics will probably differ between other academic journals. For example, the fit for the topic “Tourism demand” in Tourism Economics and “Medical tourism” in Globalization and Health are likely to be higher than those reported in this study. In contrast, the JCR impact factor is an indicator that depends exclusively on the publication to which it refers.

Secondly, the fit of a research topic becomes relevant when it is known for different journals and its importance with respect to the impact factor depends on the variance of the fits from one journal to another. In this way, the fit is more relevant than the impact factor, as the variance of the topic-journal fit increases. On the contrary, the impact factor is more important when the variance of the topic-journal fit decreases. In these cases, it may be advisable to choose journals with a higher impact factor and a lower fit. For example,

although papers related to medical tourism have a greater fit in JTR, they could be sent to TM, which has a higher impact factor.

Thirdly, the focus of the journals also influences the fit of the topics. It was observed that the variance of the topic fit increased in unidisciplinary publications and decreased in multidisciplinary publications. In this sense, it is recommended specialized journals be chosen, if the topic presents a close fit. On the contrary, those publications should be omitted if the topic presents a very low adjustment. The choice of multidisciplinary journals may be more conditioned by the impact factor, as the topics are not closely fitted.

As with all research, this study has also some limitations. First, the use of the impact factor has been questioned, as it may be manipulated by intentionally increasing the citations that a journal might receive (Kurmis, 2003; Vanclay, 2012). Second, authors may consider other factors when choosing a journal such as its visibility, the speed with which a journal will respond and the time taken between submission and publication, philosophical and ethical issues, the journal language, and the monetary cost for the publication, among others (Knight & Steinbach, 2008; van Teijlingen & Hundley, 2002). Thirdly, the temporal stability of the results cannot be guaranteed because the journals incorporate new content in each issue. In addition, the published research topics could change in various scenarios such as a new editorial line, new topics such as Covid, etc. In this sense, Mazanec (2017) demonstrated that the content of a journal's abstracts can vary significantly over time. The results may therefore be a good representation of the time period of the study. However, these analyses should be repeated in the future, to detect new research trends. Moreover, the journal impact factor can change over time. However, a simple update of the metric should be sufficient to update the information in Fig. 5.

Our own method also presents some difficulties and limitations. Text-mining analyses can be more complex than numerical data analysis due to the nature of language. This is a general limitation of the use of this methodology and the results may sometimes not be totally satisfactory. In this sense, the text-cleaning process must be a priority step before performing any text-mining analysis, because this process can significantly influence the quality of the results. Generally, the elimination of those words that have no meaning in themselves (prepositions, conjunctions, articles, adverbs, pronouns, etc.) is recommendable. It may also be necessary to eliminate words with a high frequency of occurrence whose meaning contributes nothing to the purposes of the study (e.g., article, results, objectives, authors, show, obtained, etc.). Sometimes, there are terms within the documents that can distort the results. In this study, all words referring to property rights or copyright (e.g., Elsevier, science, Ltd., all rights reserved, etc.) also had to be removed. Furthermore, the results obtained can show terms that need to be removed in the text-cleaning phase, which means that the entire analysis must be repeated. The text-cleaning phase can be thorny given the diversity of words that exist in any language, but it guarantees results that contribute the fulfillment of the objectives of the study.

Finally, it has been demonstrated that it is possible to quantify the topic-journal fit; a concept of a qualitative nature that is difficult to measure. Researchers from other knowledge areas may therefore also use the same method that has been followed in this study.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acedo, F. J., Barroso, C., Casanueva, C., & Galan, J. L. (2006). Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies*, 43(5), 957–983. <https://doi.org/10.1111/j.1467-6486.2006.00625.x>
- Benckendorff, P., & Zehrer, A. (2013). A network analysis of tourism. *Annals of Tourism Research*, 43, 121–149. <https://doi.org/10.1016/j.annals.2013.04.005>
- Bi, J., Liu, Y., Fan, Z., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, 70(2), 460–478. <https://doi.org/10.1016/j.tourman.2018.09.010>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Correia, S., Guerreiro, J., & Ali, F. (2020). 20 years of research on virtual reality and augmented reality in tourism context: A text-mining approach. *Tourism Management*, 77, 104028. <https://doi.org/10.1016/j.tourman.2019.104028>
- Demeter, K., Szász, L., & Ko, A. (2019). A text mining-based overview of inventory research in the ISIR special issues 1994–2016. *International Journal of Production Economics*, 209, 134–146. <https://doi.org/10.1016/j.ijpe.2018.06.006>
- Ding, Z., Li, Z., & Fan, C. (2018). Automation in construction building energy savings: Analysis of research trends based on text mining. *Automation in Construction*, 96, 398–410. <https://doi.org/10.1016/j.autcon.2018.10.008>
- Durán-Sánchez, A., Del Río-Rama, M., & Alvarez-García, J. (2017). Bibliometric analysis of publications on wine tourism in the databases Scopus and WoS. *European Research on Management and Business Economics*, 23(1), 8–15. <https://doi.org/10.1016/j.iedeen.2016.02.001>
- Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *The FASEB Journal*, 22, 2623–2628. <https://doi.org/10.1096/fj.08-107938>
- García-Lillo, F., Ubeda-García, M., & Marco-Lajara, B. (2016). The intellectual structure of research in hospitality management: A literature review using bibliometric methods of the journal *International Journal of Hospitality Management*. *International Journal of Hospitality Management*, 52, 121–130. <https://doi.org/10.1016/j.ijhm.2015.10.007>
- Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA*, 295(1), 90–93. <https://doi.org/10.1001/jama.295.1.90>
- Gómez-Barroso, J. L., Feijoo, C., Quiles-Casas, M., & Bohlin, E. (2017). The evolution of the telecommunications policy agenda: Forty years of articles in Telecommunications Policy. *Telecommunications Policy*, 41, 853–877. <https://doi.org/10.1016/j.telpol.2016.11.005>
- 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. <https://doi.org/10.1016/j.tourman.2016.09.009>
- Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1999). *Análisis multivariante* (5th ed.). Prentice Hall.
- Hall, C. M. (2011). Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism. *Tourism Management*, 32, 16–22.
- Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*, 72, 417–426. <https://doi.org/10.1016/j.tourman.2019.01.002>
- Jiang, H., Qiang, M., & Lin, P. (2016). A topic modeling based bibliometric exploration of hydropower research. *Renewable and Sustainable Energy Reviews*, 57, 226–237. <https://doi.org/10.1016/j.rser.2015.12.194>

- Klinger, J. K., Scanlon, D., & Pressley, M. (2005). How to publish in scholarly journals. *Educational Researcher*, 34(8), 14–20.
- Knight, L. V., & Steinbach, T. A. (2008). Selecting an appropriate publication outlet: A comprehensive model of journal selection criteria for researchers in a broad range of academic disciplines. *International Journal of Doctoral Studies*, 3, 59–79.
- Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications*, 116, 472–486.
- Köseoglu, M. A. (2020). Identifying the intellectual structure of fields: Introduction of the MAK approach. *Scientometrics*, 125, 2169–2197. <https://doi.org/10.1007/s11192-020-03719-8>
- Koseoglu, M. A., Rahimi, R., Okumus, F., & Liu, J. (2016). Bibliometric studies in tourism. *Annals of Tourism Research*, 61, 180–198. <https://doi.org/10.1016/j.annals.2016.10.006>
- Kurmis, A. P. (2003). Understanding the limitations of the journal impact factor. *The Journal of Bone and Joint Surgery*, 85(12), 2449–2454. <https://doi.org/10.2106/00004623-200312000-00028>
- Kwartler, T. (2017). *Text mining in practice with R*. John Wiley & Sons Ltd.
- Leong, L., Hew, T., Tan, G., Ooi, K., & Lee, V. (2021). Tourism research progress-A bibliometric analysis of tourism review publications. *Tourism Review*, 76(1), 1–26.
- Leung, X. Y., Sun, J., & Bai, B. (2017). Bibliometrics of social media research: A co-citation and co-word analysis. *International Journal of Hospitality Management*, 66, 35–45. <https://doi.org/10.1016/j.ijhm.2017.06.012>
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468. <https://doi.org/10.1016/j.tourman.2007.05.011>
- Loureiro, S. M. C., Guerreiro, J., & Ali, F. (2020). 20 Years of research on virtual reality and augmented reality in tourism context: A text-mining approach. *Tourism Management*, 77, 1–21. <https://doi.org/10.1016/j.tourman.2019.104028>
- Martínez-López, F. J., Merigó, J. M., Valenzuela-Fernández, L., & Nicolás, C. (2018). Fifty years of the European Journal of Marketing: A bibliometric analysis. *European Journal of Marketing*, 52, 439–468. <https://doi.org/10.1108/EJM-11-2017-0853>
- Martorell, O., Socias, A., Otero, L., & Mulet-Forteza, C. (2019). Thirty-fifth anniversary of the International Journal of Hospitality Management: A bibliometric overview. *International Journal of Hospitality Management*, 78, 89–101. <https://doi.org/10.1016/j.ijhm.2018.10.013>
- Mauri, A. G., & Minazzi, R. (2013). Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management*, 34, 99–107. <https://doi.org/10.1016/j.ijhm.2013.02.012>
- Mazanec, J. A. (2017). Determining long-term change in tourism research language with text-mining methods. *Tourism Analysis*, 22(1), 75–83. <https://doi.org/10.3727/108354217X14828625279771>
- McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of American Society for Information Science*, 41, 433–443. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:63.0.CO:2-Q](https://doi.org/10.1002/(SICI)1097-4571(199009)41:63.0.CO:2-Q)
- Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., & Nisbet, B. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.
- Munzel, A. (2016). Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus. *Journal of Retailing and Consumer Services*, 32, 96–108. <https://doi.org/10.1016/j.jretconser.2016.06.002>
- Nikita, M. (2020). Tuning of the latent Dirichlet allocation models parameters. <https://cran.r-project.org/web/packages/ldatuning/ldatuning.pdf>.
- Öğüt, H., & Taş, B. K. (2012). The influence of internet customer reviews on the online sales and prices in hotel industry. *The Service Industries Journal*, 32(2), 197–214.
- Okumus, B., Ali, M., & Ma, F. (2018). Food and gastronomy research in tourism and hospitality: A bibliometric analysis. *International Journal of Hospitality Management*, 73(January), 64–74. <https://doi.org/10.1016/j.ijhm.2018.01.020>
- Oviedo-García, M. A. (2016). Tourism research quality: Reviewing and assessing interdisciplinarity. *Tourism Management*, 52, 586–592. <https://doi.org/10.1016/j.tourman.2015.08.011>
- Palmer, A. L., Sesé, A., & Montano, J. J. (2005). Tourism and statistics: Bibliometric study 1998–2002. *Annals of Tourism Research*, 32(1), 167–178. <https://doi.org/10.1016/j.annals.2004.06.003>
- Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, 25, 348–349.
- Qian, J., Law, R., Wei, J., & Wu, Y. (2019). Trends in global tourism studies: A content analysis of the publications in tourism management. *Journal of Quality Assurance in Hospitality & Tourism*, 20(6), 753–768. <https://doi.org/10.1080/1528008X.2019.1658149>

- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., et al. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Rodríguez-López, M. E., Alcántara-Pilar, J. M., Barrio-García, S. D., & Muñoz-Leiva, F. (2020). A review of restaurant research in the last two decades: A bibliometric analysis. *International Journal of Hospitality Management*, 87, 1–10. <https://doi.org/10.1016/j.ijhm.2019.102387>
- Shen, Y., Morrison, A. M., Wu, B., Park, J., Li, C., & Li, M. (2018). Where in the world? A geographic analysis of a decade of research in tourism, hospitality, and leisure journals. *Journal of Hospitality & Tourism Research*, 42(2), 171–200. <https://doi.org/10.1177/1096348014563394>
- Sievert, C., & Shirley, K. E. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63–70).
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly.
- Song, Y., Wang, R., Fernandez, J., & Li, D. (2021). Investigating sense of place of the Las Vegas Strip using online reviews and machine learning approaches. *Landscape and Urban Planning*, 205, 103956. <https://doi.org/10.1016/j.landurbplan.2020.103956>
- Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32(6), 1310–1323. <https://doi.org/10.1016/j.tourman.2010.12.011>
- Strandberg, C., Nath, A., Hemmatdar, H., & Jahwash, M. (2018). Tourism research in the new millennium: A bibliometric review of literature in Tourism and Hospitality Research. *Tourism and Hospitality Research*, 18(3), 269–285. <https://doi.org/10.1177/1467358416642010>
- Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.
- Van Teijlingen, E., & Hundley, V. (2002). Getting your paper to the right journal: A case study of an academic paper. *Journal of Advanced Nursing*, 37(6), 506–511. <https://doi.org/10.1046/j.1365-2648.2002.02135.x>
- Vanclay, J. K. (2012). Impact factor: Outdated artefact or stepping-stone to journal certification? *Scientometrics*, 92, 211–238. <https://doi.org/10.1007/s11192-011-0561-0>
- Walters, W. H., & Markgren, S. (2019). Do faculty journal selections correspond to objective indicators of citation impact? Results for 20 academic departments at Manhattan College. *Scientometrics*, 118, 321–337. <https://doi.org/10.1007/s11192-018-2972-7>
- Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43, 1–12. <https://doi.org/10.1016/j.ijhm.2014.07.007>
- XieMiao, H. J. L., Kuo, P. J., & Lee, B. Y. (2011). Consumers' responses to ambivalent online hotel reviews: The role of perceived source credibility and pre-decisional disposition. *International Journal of Hospitality Management*, 30(1), 178–183.
- Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T. C., & Coimbra, C. F. M. (2018). History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Solar Energy*, 168, 60–101. <https://doi.org/10.1016/j.solener.2017.11.023>
- 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. <https://doi.org/10.1016/j.ijhm.2008.06.011>
- Zach, F. J., Tech, V., & Krizaj, D. (2019). Topic modelling tourism literature on innovation and technology. *e-Review of Tourism Research (eRTR)*, 17(3), 317–333.
- Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4), 694–700. <https://doi.org/10.1016/j.ijhm.2010.02.002>
- Zupic, I., & Cater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429–472. <https://doi.org/10.1177/1094428114562629>