International Journal of Hospitality Management 117 (2024) 103633

Contents lists available at ScienceDirect

International Journal of Hospitality Management

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

Intellectual landscape and emerging trends of big data research in hospitality and tourism: A scientometric analysis

Yanting Cai [a,c,1], Gang Li [b,2], Long Wen [a,*,1], Chang Liu [a,1]

[a] School of Economics, University of Nottingham, Ningbo, China
[b] Research Centre for Competitiveness of the Visitor Economy, School of Hospitality and Tourism Management, University of Surrey, UK
[c] School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Keywords:
Big data analytics
Hospitality and Tourism
Scientometric analysis
Intellectual structure
Future directions

ABSTRACT

Big data contain a vast amount of information which is valuable for researchers and decision-makers both in normal and crisis situations. This bibliometric study aims to present the progress, theoretical foundations, and intellectual structure of big data analytics in the hospitality and tourism research domain. Literature records were collected via the Web of Science and screened to maximize relevance. The overall literature dataset included 1184 papers, comprising both review and empirical articles. From this dataset, 47 publications related to the COVID-19 pandemic were identified and formed a sub-dataset to capture the specific research focuses during the crisis. The research themes and their evolutionary paths were identified by keyword clustering and keyword Time Zone analysis. Co-citation analysis was implemented to visualize the intellectual structure. Based on the systematic review, this study proposes future research directions.

1. Introduction

With ongoing technological progress, a vast amount of big data has been generated and stored, bringing revolutionary changes to both the academia and the industries (Li et al., 2018; Li and Law, 2020; Mariani et al., 2018). While traditional data are more likely to contain a representative sample, big data adoption virtually allows capturing characteristics of the entire population under inspection (George et al., 2016; Mariani et al., 2018). Thus, properly organized, analyzed and/or incorporated with other data types, big data can facilitate a decision-making process with more information, adding value for decision-makers (Verhoef et al., 2016). Such advantages prompt the adoption of big data analytics into diverse disciplines, including hospitality and tourism. Big data's valuable properties also make it a dependable tool in crisis situations when data collected via traditional methods such as official statistics remain inaccessible and data collected via small-scale surveys are unable to proxy real tourist behavior from the whole population under a real-life circumstance. Big data sources (such as bookings, reviews, transactions, and movements) can instead provide recovery-oriented policy implications with reliable real-life evidence of behavioral changes in a wide population.

The superiority of big data has led to ample research on big data analytics in the hospitality and tourism context. It is thus important to capture the overall intellectual landscape by reviewing extant relevant literature. Despite that the impact of COVID-19 is fading away, a critical reflection on how big data analytics aided hospitality and tourism research amid a global pandemic may provide useful guidance for future research directions, especially when new exogenous shocks or crises occur. As Yang et al. (2016) emphasized, emerging trends and development patterns in a research area can be identified intuitively by scientometric analysis. This study, by systematically reviewing extant literature and adopting scientometric methods, has four research aims: 1) to shed light on the theoretical foundations of big data analytics research, especially in hospitality research domain; 2) to visualize the main themes and co-citation network of big data analytics in the hospitality and tourism literature; 3) to illustrate how big data analytics aided hospitality and tourism research amid the COVID-19 crisis; and 4) to propose potential future research directions.

* Corresponding author.
E-mail address: long.wen@nottingham.edu.cn (L. Wen).
[1] Address: No. 199 Taikang East Road, Ningbo, China
[2] Address: School of Hospitality and Tourism Management, Austin Pearce Building (AP), University of Surrey, Guildford, Surrey, GU2 7XH

https://doi.org/10.1016/j.ijhm.2023.103633
Received 19 September 2022; Received in revised form 23 October 2023; Accepted 24 November 2023
Available online 14 December 2023

2. Literature review

2.1. Conception and categorization of big data

The past decade has witnessed a rapidly growing body of hospitality and tourism research that uses big data analytics. Although there is little agreement on how large data must be to fall into the scope of "big data", to date, a feature-oriented representation of big data, such as the 5 V definition, is widely accepted by scholars such as Li et al. (2018), and Mariani (2019). The typology of big data employed by hospitality and tourism studies is another core of discussion. Lv et al. (2021) argue that big data used in hospitality and tourism studies can either be structured or unstructured. Structured data are those from professional databases (e.g., Wind and some library databases), government databases and enterprise databases. Li et al. (2018), by reviewing 165 journal articles published up to 2017, conclude that tourism studies have adopted three major big data types including user-generated content (UGC), device data and transaction data. These three big data types identified by Li et al. (2018) mostly fall in the scope of unstructured big data defined by Lv et al. (2021).

2.2. Big data analytics in overall hospitality and tourism research

Emerging trends underlying the adoption of three main big data types namely, UGC, transaction data and device data, and corresponding analytical techniques in hospitality and tourism studies will be presented hereafter. Firstly, compared with other data types, UGC data have been most widely adopted in hospitality and tourism research (Li et al., 2018). Analytical methods associated with UGC are becoming more advanced and sophisticated. As shown in Fig. 1, studies using UGC data have departed from solely mining contents (i.e., Layer One) to more sophisticated analyses (i.e., Layer Two and Layer Three, and the combination of multiple layers), amalgamating values extracted from either textual or visual contents with numerical or geographical information attached to the data, or integrating the UGC with other types of big data or multiple-sized data.

On Layer One, content analysis is often utilized to mine information from unstructured text, photos, or videos. For textual data, topic modeling and sentiment analysis have been intensively adopted. User-generated visual contents have been increasingly used given technological advancements: more researchers such as Deng and Liu (2021) have used image recognition tools to identify visual information from photos. In contrast, although manual narrative analysis of online videos has been found, computer-aided video content analysis in this research domain remains rare.

As indicated by Layer Two in Fig. 1, recent studies tend to use numerical values derived from Layer One such as topic distributions (Park et al., 2021) and sentiment scores (Zhu et al., 2021) from textual data, and even the number of visual contents from online photo data (Li, Kwok, Xie, Liu and Ye, 2021) in different ways. The first type of research (i.e., Layer 2a) integrates mined contents with geographical information to track tourists' movements or generate a sentiment map (Zhang et al., 2021). Another bypass of studies (i.e., Layer 2b) integrates the mined content information with other numerical values such as hotel/-restaurant classes, and hotel ratings. Statistical analyses have been conducted to explore the underlying patterns and identify potential between-group heterogeneities.

On Layer Three, to explore causal relationships and identify determinants, constructed indices from previous layers can be used as inputs for econometric modeling. Customer- and firm-generated big data in numerical forms such as online ratings, number of likes, the timeliness of business responses to consumer reviews (Kwok and Xie, 2016) and structured small-scale data are used together with textual or visual contents on the third layer. Li, Ye, Nicolau and Liu (2021), for instance,

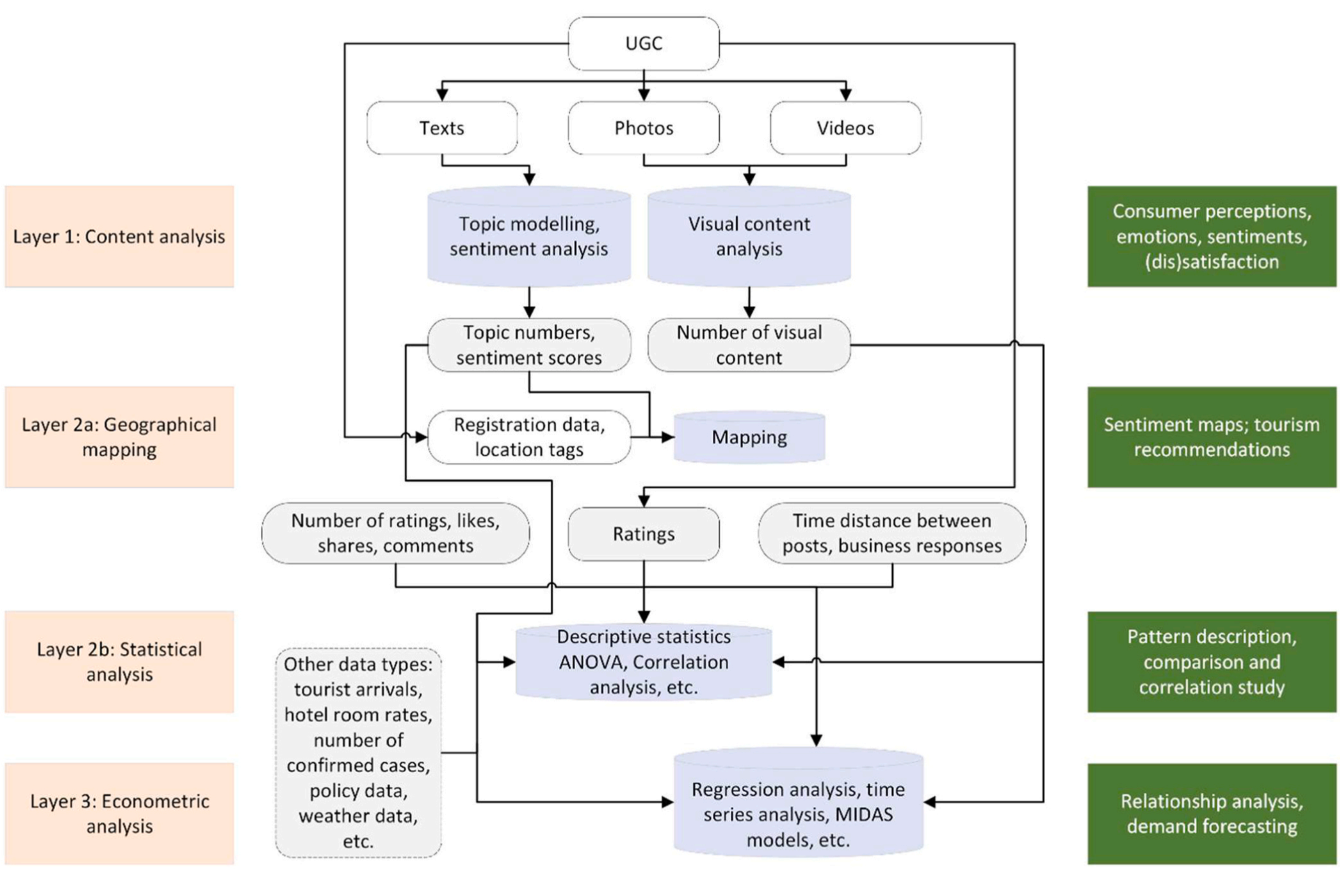

Fig. 1. Categorization of user-generated content data, analytical techniques, and research themes.

revealed the relationship between the photo content information (such as smiley faces) and online review lengths for hotels. Depending on research questions, while in some cases data mined from the UGC form independent variables to explain variables proxied by structured data such as tourist arrivals (e.g., Li, Hu and Li, 2020) and hotel performance (e.g., Xie et al., 2017), in other cases, the dependent variable – such as customer's emotional responses (e.g., Li, Meng and Pan, 2020) – can come from the UGC data.

Secondly, regarding operation/transaction data, web search data have been adopted in demand forecasting practices for tourism (Wen et al., 2021) and hotel industries (Wu et al., 2022). Mobile app visiting data (Wang et al., 2018), payment and credit card transaction data (Carvalho et al., 2022), and flight search and pick data (Gallego and Font, 2021) are other examples of operation data being used in recent hospitality and tourism studies. Analytical methods such as sensitivity analysis (Wang et al., 2018) and econometric modeling (Carvalho et al., 2022) have been utilized.

Thirdly, recent years have witnessed expansion in sources of device data such as bikeshare data (Buning and Lulla, 2021) as well as enrichment in research topics using device data such as informing tourists of potential geologic hazards on tourist routes (He et al., 2023) and revealing values of costal tourism (Kubo et al., 2020). While tourist tracking studies continuously use mobile roaming data from mobile network operators, more diverse follow-up analyses have been conducted in recent studies such as between-group comparisons of visits (Altin et al., 2022) by different tourists. Due to limited accessibility, sensor-based device data remains the least-adopted type of big data.

To address research questions from more comprehensive perspectives, the mix of multi-source and multi-type big data appeared in recent studies. Pramana et al. (2022), for example, used statistical methods to analyze multi-type big data such as mobility data, search index and online reviews, holistically to showcase the impact of the COVID-19 pandemic on the Indonesian tourism industry. Triangulation from different methods and data sources can enhance the reliability and credibility of a finding by alleviating the limitations of a single research design (Turner et al., 2017). The adoption of a comprehensive mixed-method approach is also emerging (e.g., Meek et al., 2021). Recent scholars started to triangulate the findings from secondary big data analytics by separate experimental studies carried out via online surveys, or vice versa (Pachucki et al., 2022).

2.3. Big data analytics and theoretical fundaments in hospitality-specific research

Table 1 presents data types, research themes, theoretical foundations, and analytical techniques mentioned in an exclusive sample of hospitality studies. The selected studies are grouped based on the primary data type adopted. Overall, UGC data are used most often among all types of big data in hospitality studies. Among studies investigating customer satisfaction, which is one of the dominant research themes in hospitality UGC analytics, the two-factor theory is mostly used to incorporate satisfier and dissatisfier factors (Berezina et al., 2016). To study the impact of multisensory stimuli on customer satisfaction, researchers have applied sensory cognitive theory, attribution theory, and perceptual symbol systems theory (e.g., Lee et al., 2019). Particular attention from hospitality field has been paid on the reliability and perceived helpfulness of online reviews (e.g., Kwok and Xie, 2016; Xu et al., 2020, Hu, 2020), due to their impacts on consumers' evaluation of hotels and thus consumers' purchase decisions. In this vein, psychological theories such as signal theory, the loss aversion property of prospect theory, uncertainty reduction theory, source credibility theory, and expectation-confirmation theory are often used to unravel consumer psychologies (Zhao et al., 2019; Lai et al., 2021; Kwok and Xie, 2016; Xu et al., 2020; Hu, 2020). Furthermore, in modeling hotel performance using user-generated information, Xie et al. (2017) focus on consumer-manager interactions, and reciprocity theory is thus used. Kim

Table 1
Summary of theoretical fundaments for big data research in hospitality.

Main data type	Research theme	Example paper	Theoretical fundament (s)	Analytical technique (s)
UGC				
Textual and numerical content				
• (Determinants of) customer experience/(dis) satisfaction/ sentiment		Berezina et al. (2016)	Herzberg's two-factor theory	Text analysis
		Lee et al. (2019)	Sensory cognitive theory; attribution theory; perceptual symbol systems theory	Text analysis, sentiment analysis, regression analysis
		Zhao et al. (2019)	Signal theory	Regression analysis
		Lai et al. (2021)	Loss aversion property of prospect theory	Sentiment analysis, threshold regression
• Review helpfulness/ review reliability/ credibility		Kwok and Xie (2016)	Uncertainty reduction theory	Regression analysis
		Xu et al. (2020)	Risk aversion theory; source credibility theory	Regression analysis
		Hu (2020)	Expectation-Confirmation theory	Topic modeling
• Determinants of hotel performances/ ratings/prices		Xie et al. (2017)	Reciprocity theory	Panel data regression with fixed effects
		Kim and Park (2017)	Regulatory focus theory	Hierarchical multiple regressions
• Consumer revisit behavior		Park et al. (2020)	Theory of planned behavior	Sentiment analysis
• Employee's perspectives		Stamolampros et al. (2019)	Herzberg's two-factor theory	Topic modeling
Visual content				
• Review helpfulness		Li, Kwok et al. (2021)	Media richness theory	Object detection model, YOLOv3, regression analysis
		An et al. (2020)	Negativity bias theory	Manifested photo content, regression analysis
• Heterogeneity underlying photo content		Ren et al. (2021)	Deep learning theory	Deep learning
Transaction/ Operation data				
• Demand forecasting		Yang et al. (2022)	Customer information processing theory; demand theory; utility theory; consumption behavior theory	Mixed data sampling (MIDAS) models
• Determinants of hotel performances/ ratings/prices		Sánchez-Lozano et al. (2021)	Walrasian theory of value	Regression analysis with random effects
• Revenue management		Saito et al. (2019)	Random utility theory	Discrete choice model
• Consumer online/ purchase behavior		Cheng and Anderson (2021)	Consumer purchase journey theory	Latent-class, two-stage Logistic regression model
		Lee and Kim (2020)	Gravity theory	Machine learning:

(continued on next page)

Table 1 (*continued*)

Main data type	Research theme	Example paper	Theoretical fundament (s)	Analytical technique (s)
UGC Textual and numerical content				
				decision tree method
Device data				
• Consumer overnight stays behavior		Nyns and Schmitz (2022)	Clayton Christensen's disruptive innovation theory	Statistical analysis; Regression analysis

and Park (2017) employ regulatory focus theory to postulate why customers evaluate online review ratings before making a purchase decision, while Park et al. (2020) use the theory of planned behavior to study guests' revisit behavior. Stamolampros et al. (2019) build on the two-factor theory to investigate employees' perspectives. One of the recent research interests in hospitality studies adopting UGC is the detection and extraction of specific topics, such as the implementation of sustainability practices (D'Acunto et al., 2023), corporate social responsibility (D'Acunto et al., 2020), privacy issues (D'Acunto et al., 2021), service robots (Orea-Giner et al., 2022), and mechanical AI (Mariani and Borghi, 2021) evaluated by hotel guests. Customers' concerns regarding these specific themes extracted from their UGC, in many studies, were then used to explain consumer satisfaction or hotel performance, often measured by positive WOM (D'Acunto et al., 2023) or online ratings (Orea-Giner et al., 2022).

Although user-generated visual content is less commonly used in hospitality literature than other types of UGC, its great potential in enhancing consumers' perceived review helpfulness has attracted scholars' attention according to media richness theory (Li, Kwok et al., 2021). An et al. (2020) further explain that guests who stayed at two-star hotels are more likely to share photos in negative reviews than those who stayed at higher class hotels, using negativity bias theory (Ito et al., 1998).

As transaction/operation data such as bookings and web visit data directly reflect consumption behavior or demand, relevant theories such as demand theory, random utility theory, consumption behavior theory, and Walrasian theory of value have been adopted in hospitality research (e.g., Yang et al., 2022; Sánchez-Lozano et al., 2021; Saito et al., 2019). Consumer purchase journey theory has also been utilized to explore consumers' online web visiting behavior through transaction data analysis (Cheng and Anderson, 2021). A smaller group of hospitality researchers have adopted device data, such as mobile tracking data, to capture consumers' overnight stay behavior. Nyns and Schmitz (2022) suggest using Clayton Christensen's disruptive innovation theory in the adoption of Web 2.0 Internet technology in the accommodation sector.

This study further develops an intellectual map of big data analytics in hospitality and tourism research, given the significant progress made in this area. This study extends previous scientometric works in this research direction from three perspectives. Firstly, this study broadens the coverage of both the time span and keywords used in the literature search. Secondly, while more search keywords help increase the coverage of the literature pool, they may also distort the relevancy of search outcomes. According to Mariani and Baggio (2021), certain big data studies in hospitality and tourism were conceptual in nature (Buhalis and Sinarta, 2019). To better address our research question, only empirical studies that applied big data analytics in hospitality and tourism contexts and relevant review articles were included. Thirdly, a COVID-19-related literature data subsample was investigated to identify the emerging topics during the crisis and provide guidance for future crisis management studies using big data analytics in the hospitality and tourism domain.

3. Data and Methods

Literature record data were retrieved from Web of Science, a well-recognized online literature database (Thelwall, 2008), by entering key terms that are relevant to big data analytics and hospitality and tourism, respectively. According to big data types proposed by Li et al. (2018), more terms related to big data can be included. As did Li and Law (2020), "social media" was also employed as one of the search key terms. Similarly, Lv, Shi and Gursoy (2021) included a specific data type (i.e., volume data) to supplement the generic big data term. We thus included more keywords related to big data types. In line with Mariani and Baggio (2021), "big data analytics" was also included to essentially help restrict the search results to the analytics of big data. Thus, in the spectrum of big-data-related topic key terms, we input "big data" OR "big data analy* " OR "social media" OR "user-generated content* " OR "device data" OR "transaction data" OR "search engine" OR "search query". Meanwhile, regarding key terms related to hospitality and tourism, we recruited "tourism" OR "tourist* " OR "hospitality" OR "hotel* ". The asterisk denotes variable spellings. The two spectrums of keywords were connected by the logic word "AND" to retrieve the intersect of big data analytics and the hospitality and tourism research domain. Web of Science provides diverse literature indexes from which we focused on Science Citation Index Expanded, Social Science Citation Index, Arts & Humanities Citation Index, and Emerging Source Citation Index. Then, we extracted articles and review articles as two document types, and the publication language was restricted to English. Data searching was conducted in February 2022 using Web of Science. Publications from irrelevant research disciplines, such as urology and nephrology, were excluded from the dataset. This initially yielded 3224 literature records. Articles not in English, not full-length, or retracted were also removed. After the initial round of filtering, a total of 3034 publications remained for further screening.

Each literature record was then manually screened to exclude irrelevant publication items. Only articles which genuinely adopted big data and corresponding analytics in tourism or hospitality contexts and relevant review articles passed the final screening condition. After the exclusion, the overall literature record dataset (named as "overall dataset" in the later analysis) included 1184 publications. Moreover, a total of 47 publications which paid attention to the COVID-19 pandemic were extracted and formed a second literature record, referred to as the "COVID-19-related dataset" in later analysis. The retrieved literature record dataset included a series of publication attributes namely, title of paper, year of publication, author(s), institution(s), journal of publication keywords, and abstracts.

This study employed CiteSpace, one of the most commonly used scientometric analytical tools (Chen, 2006), to visualize the evolutionary process of research themes in this area via Time Zone analysis. In addition, this study utilized VOSviewer to conduct co-authorship analysis (as presented by Figs. S1, S2 and S3 in the supplementary file) and co-citation analysis, as well as to visualize the clustering results of keywords.

4. Results and Findings

4.1. Descriptive statistics

Fig. 2 shows the distribution of publication years for the 1184 publications and 25,268 citations in the overall dataset. A total of 47 publications with 221 citations were included in the COVID-19-related dataset. Distribution details are included in Tables S1 and S2 in the supplementary file for both datasets, respectively. Fig. 2 illustrates an exponential growth pattern in the number of publications on big data analytics in hospitality and tourism over time. The first study in the overall dataset was published in 2008. Prior to the end of 2016, only 119 studies were identified, whereas more than half of the total articles were published from 2020 onwards, indicating a significant increase in

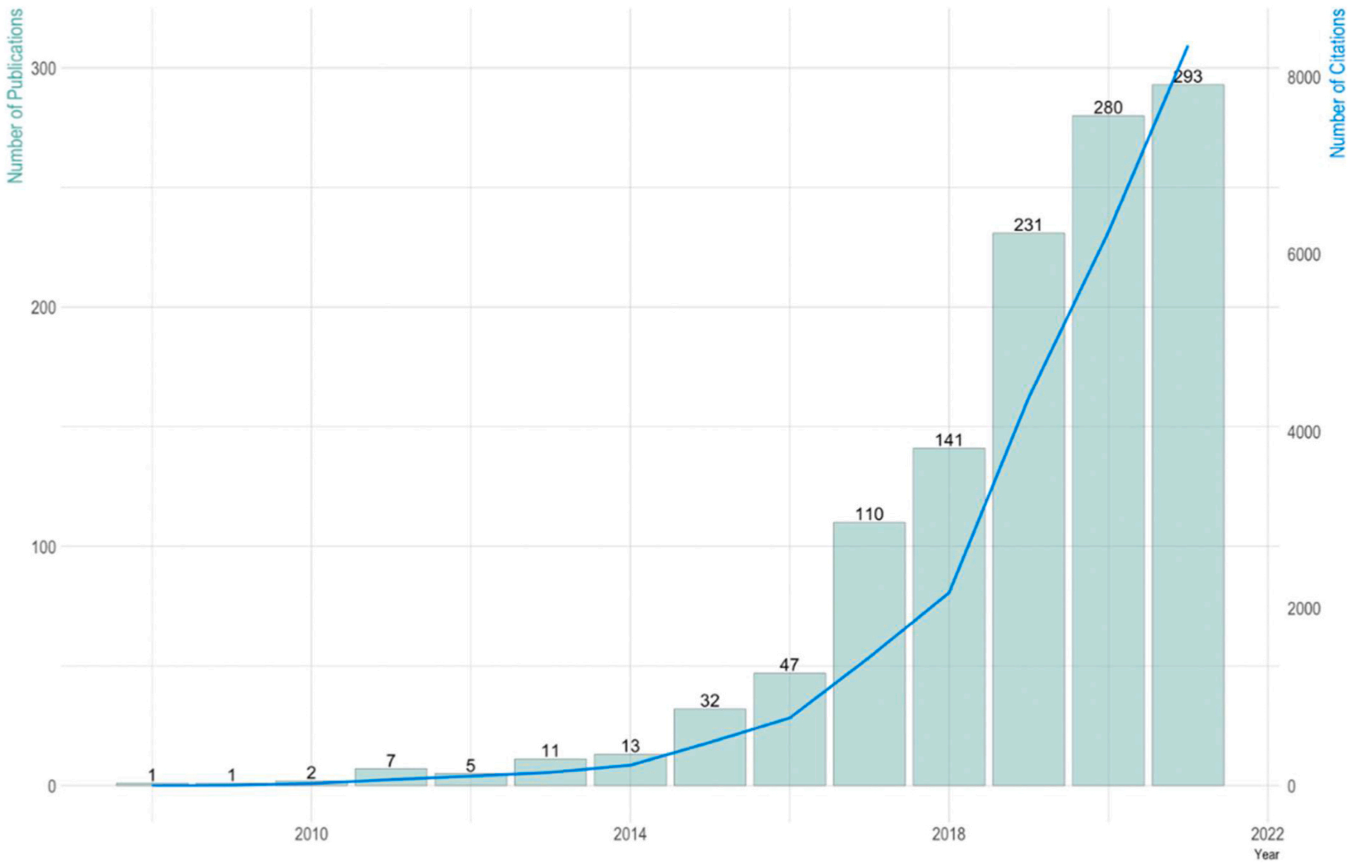

Fig. 2. Distribution of publications and citations in the overall dataset from 2008 to 2021.

research interest. The number of citations for the overall dataset followed a similar trend to the number of publications.

4.2. Co-citation network analysis

In the co-citation network analysis (Van Eck and Waltman, 2010), the value that scholars place on each cited object is shown based on bibliographic coupling, which occurs when two studies cite a common third publication in their references. Figs. 3 and 4 display the networks and the strength of the connections for co-cited journals and documents, respectively. Due to space constraints, the co-citation analysis amongst authors is presented in Fig. S4 in the supplementary file. Each pair of journals/publications co-cited is connected by curved lines.

For a clear visualization of the co-cited network of journals, we selected the top 161 most cited journals with a minimum citation frequency of 50. They were grouped into four clusters as shown in Fig. 3. *Tourism Management* is the most influential journal in this research domain. The blue cluster incorporates journals that focus on tourism management, planning, and vacation marketing. Journals that are related to information technology also appear in the blue cluster. *Journal of Travel Research*, *Annals of Tourism Research*, *Current Issues in Tourism*, and *Tourism Management Perspectives* are the most frequently co-cited sources in the blue cluster. Journals in the green cluster mostly aim to develop certain technological systems for tourism management issues with respect to two subthemes: 1) geography, landscape, and urban planning; and 2) information and computer science. Journals in the yellow cluster mainly focus on economics applications, econometric modeling, and statistical analysis of big data sources. The most influential journals in this cluster include *Tourism Economics*, *Journal of Econometrics*, *Econometrica*, *International Journal of Forecasting* and *Journal of Business and Economic Statistics*.

As psychographic information of individuals can be mined from big data, journals in the red cluster have focused on utilizing this property to provide managerial implications. The most influential journals included in the red cluster are mainly hospitality journals, namely *International Journal of Hospitality Management*, *International Journal of Contemporary Hospitality Management*, and *Journal of Hospitality and Tourism Research*; as well as generic business, marketing, and psychology journals, namely *Journal of Business Research*, *Journal of Marketing*, *Journal of Interactive Marketing*, *Computers in Human Behavior* and *Psychology and Marketing*. Overall, Fig. 3 suggests that this research area has attracted broad attention from not only the hospitality and tourism field, but also general marketing, business, economics and econometrics, urban planning, geography, and information and technology fields.

The co-citation network can be used to identify the seminal and the most influential literature in this research domain. As shown in Fig. 4, a total of 167 cited documents were grouped into four clusters. In the blue cluster, three publications – namely, Xiang et al. (2015), Xiang et al. (2017) and Guo et al. (2017) – are most influential. These three seminal hospitality-focused articles commonly adopted the textual analytics of online hotel reviews. While Xiang et al. (2015) and Guo et al. (2017) both examined the important relationship between hotel ratings and consumer experience, Xiang et al. (2017) was the first study to comparatively investigate the variations underlying information quality on three online review platforms, namely TripAdvisor, Expedia, and Yelp, using the hotel industry as an example. In the green cluster, most influential works are review articles which focus on the adoption of social media (Xiang and Gretzel, 2010; Zeng and Gerritsen, 2014), electronic word-of-mouth (Litvin et al., 2008), information technology (Buhalis and Law, 2008) and user-generated content (Lu and Stepchenkova, 2015) in hospitality and tourism research. Ye et al. (2011), Sparks and Browning (2011) and Ye et al. (2009) represent the most frequently co-cited works in the red cluster. The objectives of these articles overlapped in the adoption of user-generated online reviews

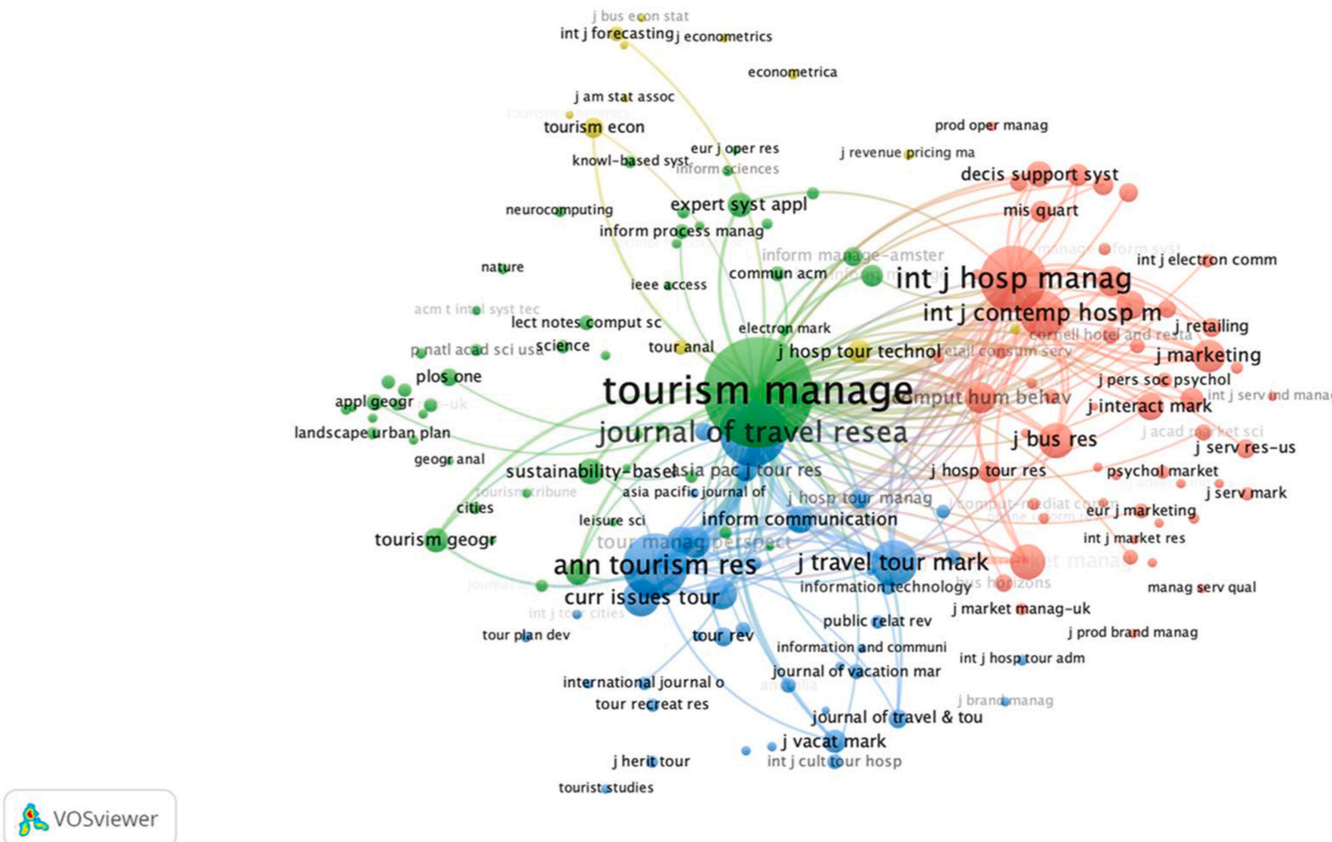

Fig. 3. Network visualization of co-cited journals in the overall dataset.

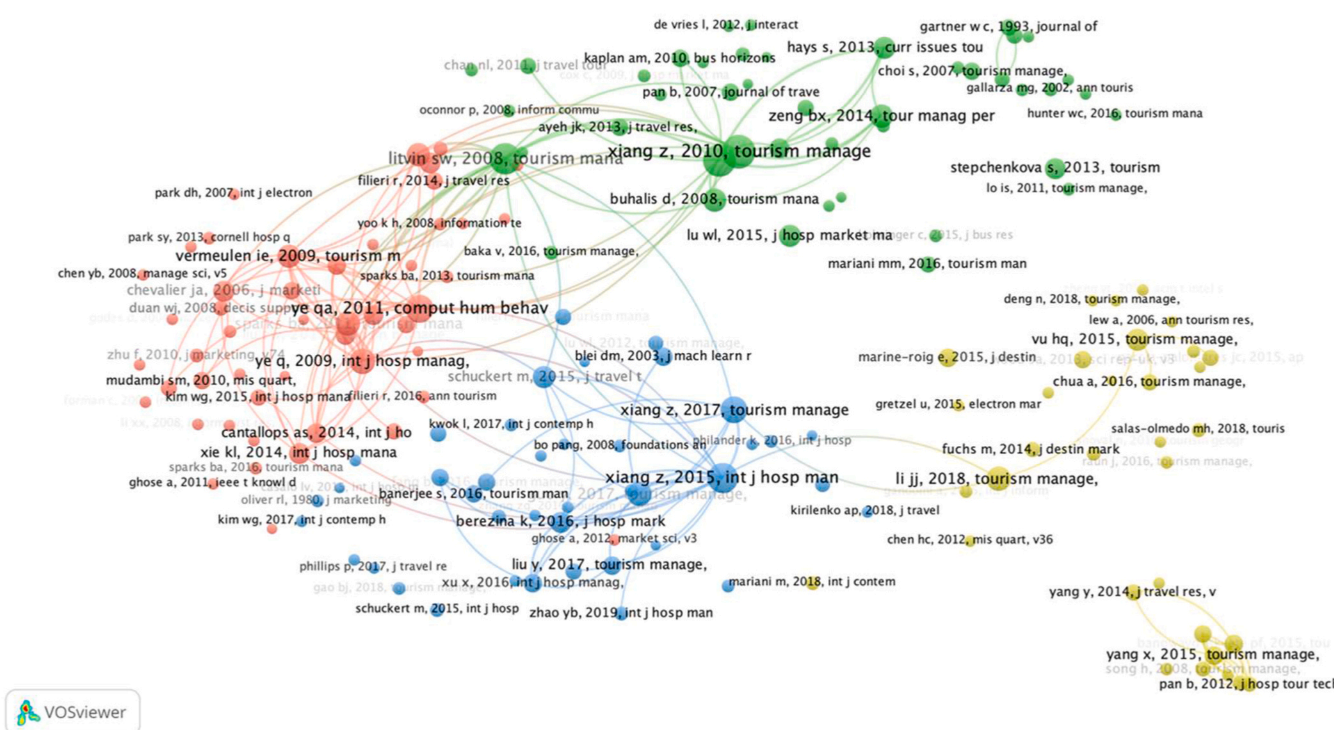

Fig. 4. Network visualization of co-cited references in the overall dataset.

and/or managerial responses to study consumer booking behaviors or purchase intentions. The research foci of frequently co-cited documents located at the upper right corner of the yellow cluster, are mainly tourist flow tracking, tourist movement studies and tourism hot spot identifications that adopt the location data from GIS (García-Palomares et al., 2015), geo-tagged photos (Vu et al., 2015) or other geo-tagged social media data (Chua et al., 2016). Meanwhile, the lower right corner of the yellow cluster exhibits another research branch which mainly adopted big data to aid forecasting performance. In this sub-cluster, Yang et al. (2014), Yang et al. (2015) and Pan et al. (2012) are the three most influential articles.

4.3. Keyword clustering

Fig. 5 represents keyword clustering of the overall dataset by grouping the keywords in publications' titles, abstracts, and keywords into multiple themes according to their co-occurrence. Nodes with shared attributes were assigned to color-coded clusters so as to identify distinct conceptual subcategories and significant gaps (Zopiatis et al., 2021). By merging keywords with the same meaning but different spellings and setting the minimum number of keyword occurrence to 15, a total of 124 keywords were recognized and clustered. As indicated by Fig. 5, four distinct clusters were identified. The clustering details are recorded in Table S3 in the supplementary file.

The overall theme of the red cluster, reflected by the two dominant keywords, is the adoption of social media information in tourism research. Keywords in the upper right corner of red cluster form the first subtheme: tourists' spatial movement patterns/behavior for tourism route design, destination planning, and so on. The spatial analysis/mapping of tourism activities have been specifically constructed in heritage sites, nature-based sites, urban areas, protected areas, or a specific region such as Hong Kong. This subtheme uses data types like device data which mainly include GPS data and UGC data which contain location information such as location-tagged online texts or photos from Flickr or online check-in/registration data. Evaluation of tourists' experiences and destination images is the second subtheme in the red cluster, denoted by the keywords in the lower right corner. Tourists' perceived image of a city, a place, or a specific form of tourism such as sustainable tourism is investigated by this branch of research using data from social media platforms such as Twitter and Instagram. Customer engagement, which is defined as "a customers' personal connection to a brand as manifested in cognitive, affective, and behavioral actions outside the purchase situation" (So et al., 2014, p. 310), has become a core objective of relationship marketing within the hospitality and tourism context. This engagement can be seen in consumers' contributions to a brand or firm through activities such as (electronic) word-of-mouth, recommendations and writing reviews. As such, motivated by the great value of research, another subtheme formed by the keywords in the lower left corner of the red cluster, is customer engagement and communication on social media community, which was developed by analyzing social media text data.

The second-largest keyword cluster, coded green, is focused on hospitality. This cluster provides useful information on hospitality scholars' specific research focus. One of the most significant research topics in this group is the evaluation of customer (dis)satisfaction using text mining, topic modeling, or sentiment analysis via techniques such as machine learning. Service quality, considered a major determinant of customer satisfaction, has also attracted the attention of hospitality scholars (Nilashi et al., 2022). Investigating behavioral intention, including revisit intention, is another research topic that hospitality big

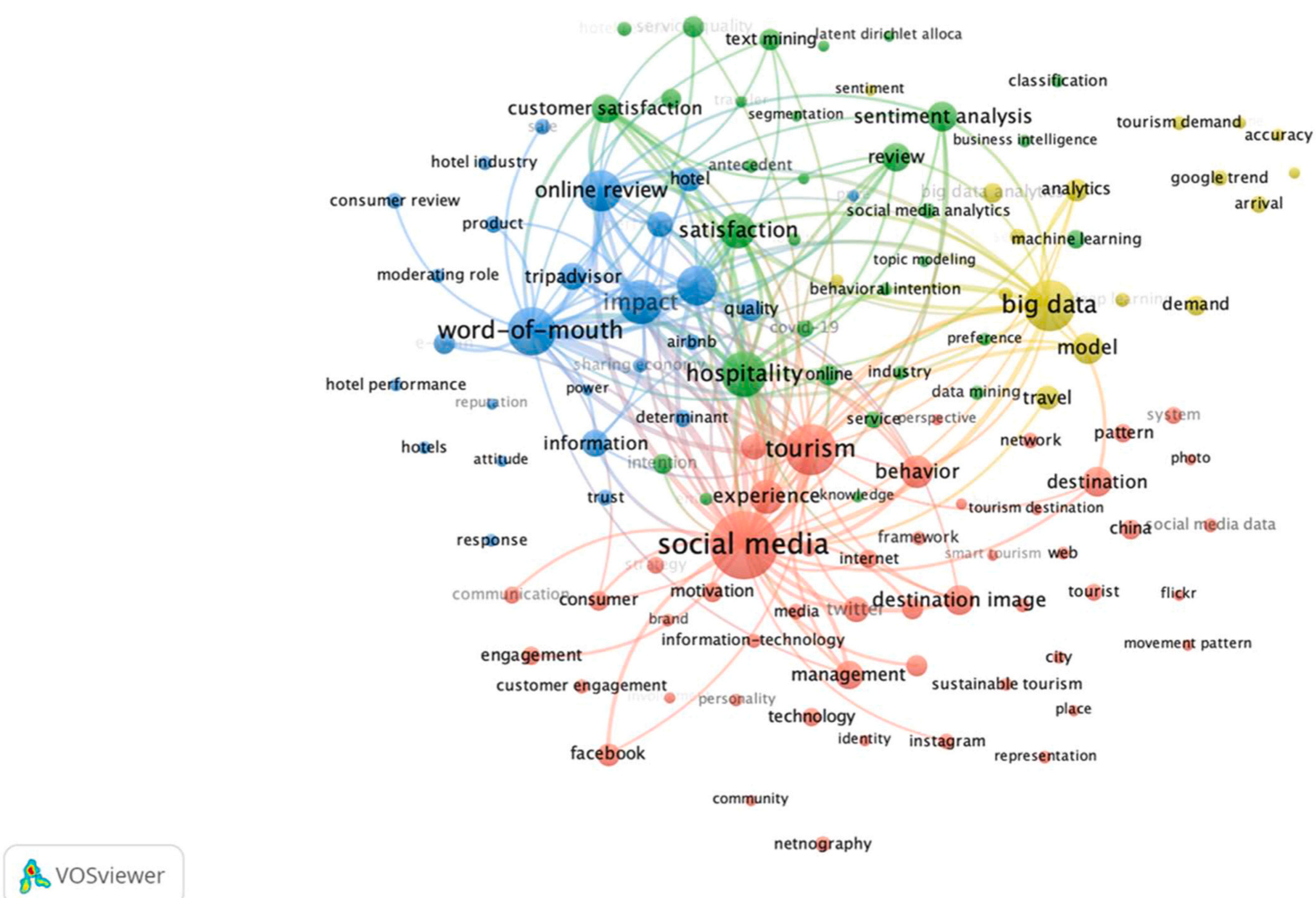

Fig. 5. Co-occurrence of keywords in the overall dataset.

data analytics can contribute to. Market segmentation is another subtheme in the hospitality research domain. For instance, Ahani et al. (2019) developed a segmentation strategy for spa hotels by applying the machine learning method on online review and rating data.

The blue cluster is closely connected to the green one, with hospitality-related keywords such as hotels, hotel performance, and hotel industry indicating the dominant research context in this research branch. Other frequently detected keywords in this cluster include word-of-mouth, online review, impact, performance, quality, sharing economy, response, TripAdvisor, and Airbnb. The exhibited keywords reflect the research themes, which are impact studies and performance analysis through the analysis of online hotel reviews and business responses from platforms in a sharing economy (Cheng et al., 2019), such as TripAdvisor and Airbnb. It is worth noting that not only are customers' online reviews valuable, but managerial responses are also a crucial source of textual data. For example, Sheng et al. (2019) examined the impact of online managerial responses on returning customers' future satisfaction measured by hotel review ratings. Certain studies in this cluster also attempted to explore how users perceived the helpfulness of online reviews or the credibility/trust of an online review regarding the hospitality industry (Lee et al., 2021). The keyword "quality" may reflect research themes regarding either big data quality (Alaoui and Gahi, 2019) or the quality of information on online review platforms (Xiang et al., 2017).

The smallest collection of keywords is colored yellow. The research theme of this cluster includes tourism/hospitality demand forecasting (Huang and Hao, 2021; Sun et al., 2021), and tourism/hospitality demand modeling (Liu et al., 2021). To improve forecasting accuracy, Google trend data, machine learning (Sun et al., 2019) and deep learning respectively appeared to be the big data source and that analytical technique that frequently appeared in this research subtheme. Lowering the minimum number of keyword occurrences to 10 reveals additional subthemes in the yellow cluster, with the most significant being the proposal and validation of personalized tourism attractions or a service recommender system (Meehan et al., 2016).

4.4. Evolution and trend analysis

The Time Zone analysis result, as obtained via CiteSpace, is shown in Fig. 6 to visualize the change of research foci across the pandemic. The Time Zone analysis of the overall dataset is included in Fig. S5 in the supplementary file due to space constraints. Studies published in 2019 from the overall dataset were used to generate keywords for 2019 in Fig. 6 to showcase research themes in normalcy. In 2019, before the pandemic outbreak, general consumer-side keywords such as customer satisfaction, destination image, experience, and behavior (as well as

Fig. 6. Evolution of keywords in the COVID-19-related dataset in contrast to keywords in 2019.

keywords associated with data and method) were recognized from the overall dataset.

The following three years (i.e., 2020–2022) in Fig. 6 exhibit the Time Zone analysis results solely for the COVID-19-related dataset. From the COVID-19-related articles published in 2020, the initial year of the pandemic outbreak, big data and pandemic-related keywords were frequently identified. In addition, as the pandemic outbreak profoundly distorted the global economy, studies have been conducted to capture the current situation/market state of the hospitality and tourism industry amid COVID-19. In this sense, online travel agencies (OTA) were used to retrieve hotel price data, and descriptive analyses of hotel prices were adopted by studies such as Wu et al. (2020) to showcase the fluctuations in hotel room rates amid the outbreak.

Regarding the COVID-19-related studies published in 2021, the research emphasis was on crisis management and tourism recovery, with hospitality and tourism scholars having access to various types of big data associated with the pandemic. Analytics of textual data via natural language processing, machine learning and deep learning have allowed researchers to investigate the impact (Zhang et al., 2021) of the pandemic on public perception, sentiment, (dis)satisfaction towards an industry such as hospitality (Hu et al., 2021) and the aviation industry (Gallego and Font, 2021), a destination (Chen, 2021), or a specific service type such as service robots (Zhang, 2021). Amid a major global health crisis, "risk perception" and "personality" became notable keywords in 2021 in the COVID-19-specific studies. More data types, especially device data including sensor-based smartphone data (Mikhailov and Kashevnik, 2021) and operation data – namely, credit card transaction data (Donaire et al., 2021) and location-tagged user registration data (Zhang et al., 2021) – have been adopted in 2021 by COVID-19-related studies to capture tourists' spatial-temporal movement or emotion distribution patterns amid the pandemic.

Amongst the COVID-19-related articles published in 2022, scholars sought to revive the market by implementing cluster analysis of public opinions (Zha et al., 2022) and exploring the impact of service quality on customer satisfaction (Nilashi et al., 2022). While the research theme of emotion and public sentiment amid negative events remained dominant, research contexts became more diverse in 2022. Given that the tourism industry was once under stringent COVID-related pressure in some destinations, alternative forms of tourism were sought by COVID-19-related studies published in 2022. For instance, Zhang et al. (2022) explored public sentiment towards virtual tourism to evaluate its supplementary effects on on-site tourism. Travel distance has become another concern. Public interests in local tourism were investigated as an alternative to international tourism (Lee and Leung, 2022). Moreover, while the real tourism activity was profoundly impacted by travel restrictions, big data sources such as search query and social media sentiments were available on a daily basis and carried rich information on people's interest and future travel intention. High-frequency big data have thus been employed using mixed data sampling (MIDAS) models to aid demand forecasting in the pandemic era (Wu et al., 2022).

In contrast to pre-pandemic research, COVID-19-related studies have focused more on comparisons across the pandemic to investigate the impacts of a major crisis. Thus, for analytical techniques, statistical analyses such as ANOVA and *t*-tests were adopted in COVID-related research to evaluate the statistical differences in destination images caused by the pandemic (Jia, 2021). Likewise, temporal (Gallego and Font, 2021) and spatial variations (Zhang et al., 2021) of sentiment scores were investigated in COVID-19-related studies to reflect the impact of the pandemic. Importance-performance analysis was implemented in COVID-19-related studies to reveal consumers' perceptions of hotel service attributes throughout the pandemic (Hu et al., 2021). Moreover, in terms of data types adopted, with actual tourism practices being impeded by worldwide travel restrictions, tourist tracking in a specific destination using device data was relatively scarcer in COVID-19-related studies. Finally, COVID-19-related studies were mostly policy-oriented for short-term industrial recovery, while research topics with regard to long-run industry development were largely set aside.

5. Concluding remarks and future research directions

In conclusion, this study discusses the up-to-date types of big data and analytical strategies used in the hospitality and tourism research domain, and summarizes the theoretical fundaments in extant big data analytics in hospitality research. Using co-citation analysis, keyword clustering, and keyword Time Zone analysis, this study presents the intellectual structures and themes of big data analytics in hospitality and tourism research under both normal and crisis circumstances.

In general, the main themes summarized by the clustering results include: 1) tracking tourists' movement patterns, and capturing spatial-temporal sentiment maps; 2) evaluating consumers' psychographic aspects such as sentiments, experience, perceptions, attitudes, or satisfactions towards hospitality and tourism products or services, and identifying the heterogeneity and/or determinants of the aforementioned psychographic aspects; 3) social media marketing themes, including customer engagement and consumer-business online communication; 4) evaluating performance and its determinants; 5) demand modeling and forecasting; 6) developing and empirically validating new methods and algorithms, such as personalized recommender systems for tourism attractions or routes; and 7) issues associated with managerial responses and their perceived helpfulness, trust, and the credibility of online reviews. During the COVID-19 pandemic, research foci have shifted towards tourists' risk perceptions, the impact of exogenous shocks on hotel prices, changes in tourist perceptions/sentiments changes over a crisis period, and alternative tourism forms.

With the analytical techniques of big data becoming more diversified and advanced, certain research themes proposed by Li et al. (2018) as future research directions have already become a reality in the most recent years. These include the utilization of online news data (Park et al., 2021) and photo content data (Li, Ye et al., 2021) and the use of a mixed-methods approach and multi-type data. Regarding the expansion of the research contexts proposed by Li et al. (2018), emergencies and tourism precautions have been extensively investigated in the last two years amid the COVID-19 pandemic. Based on the literature review and scientometric analysis, six future research directions are proposed hereafter.

Firstly, with regards to hospitality-specific studies, consumers' perception/evaluation/concerns towards a specific topic such as sustainability practices, CSR, review helpfulness, and its influence on hotel performance/rating/prices emerged as an important theme. In this sense, consumer psychology theories are commonly adopted in this branch of studies. While extant research mostly collects and relies on the UGC on a specific time slice, future studies are encouraged to monitor the UGC longitudinally on a relatively large time span to explore changes in customer attitudes over time. For instance, whether heterogeneous individuals are transiting to sustainable consumption patterns can be investigated, and this is of great managerial importance for industrial practitioners to facilitate sustainable development. Moreover, hospitality studies using big data analytics are found dominated by micro-level consumer psychographic facets, firm-level hospitality performance studies and demand forecasting practices. This is possibly because the major big data types used by hospitality research are UGC and transaction data which are mostly generated at the individual and firm levels. Whereas from a macro supply-side perspective, the application of big data to facilitate industrial development still calls for further scholarship. For instance, regional inequality of destinations in terms of their hospitality and tourism industries' resilience or vulnerability to future crises and other risk factors is yet to be examined. Likewise, how big data analytics would contribute to the construction of a regional competitiveness index regarding hospitality and tourism industries remains rarely mentioned. Furthermore, the development of a dynamic, longitudinal, and multi-destinational industrial competitiveness evaluation system can be facilitated on a macro scope.

Secondly, as the progress in advancing big data analytical techniques and the processability of big data imposes major obstacles for researchers in this domain (Lv et al., 2021), methodological advancement will continue to provide ample scope for future applications. For instance, with the continuous upgrades of visual recognition technologies, informative online photos and videos contents can be analyzed and used in a more sophisticated manner via statistical analysis, econometrics modeling or geographical visualization. Moreover, from journal co-citation and keyword clustering analysis, it is evident that tourism big data analytics are more frequently associated with geographical studies such as point of interest recommendation and tourists' trajectory tracking. Whereas the adoption of device data analytics in hospitality research is still rare, probably due to the lack of relevant data, or the absence of relevant research topics. In this lens, hospitality researchers may consider the utilization of device data to aid industrial performance or strategic decisions such as hotel location choice, and consumer segmentations.

Thirdly, due to the growing storage capacity which allows larger datasets to be captured, what is considered "big" today may not be big enough in the future. Amongst COVID-related literature dataset, the size of big data adopted varies from 1418 (Davras and Durgun, 2022) for TripAdvisor online review data to over 5 billion for Skyscanner flight search data (Gallego and Font, 2021). Given continuous growth in data storage and processing capacity, the size of big data should be carefully considered to present a more complete representation of the population.

Fourthly, the great value underlying the concept of "real-time" deserves more scholarship (Ariyaluran Habeeb et al., 2019). Essentially, the accumulation of continually moving data from linked devices and sensors has made it critical to prompt anomalies-detection mechanisms and make policy decisions in real time. The real-time and dynamic property is especially important especially in a crisis circumstance where the exogenous environment is fluctuating. Despite that few studies have investigated the dynamics of consumers' satisfaction in a crisis time (Kim et al., 2023), a real-time decision-making solution based on various big data sources is still crucial to aid agile and effective policy responses in both normalcy and crisis situations.

Fifthly, experimental research, especially online experiment-based research, has been gaining increasing popularity given its superiority for causality inference. Nevertheless, laboratory/online experiments encounter limitations as many of them focus on behavioral intentions in a specific scenario rather than real behavioral outcomes in a generic setting. However, in many cases intentions do not necessarily translate into real behavior, and the external validity of online experiments needs to be verified. Big data, as a representation of real tourist behavior, provide a feasible solution to triangulate laboratory/online experiments, verify external validity, and thus generate more robust research findings and more reliable managerial implications. Experimental research triangulated by secondary online big data has thus become an emerging trend but is still rare. Future research may consider the usage of multiple sources and types of big data to complement online/laboratory experimental research.

Sixthly, other than traditional self-reported methods via experiments, reflections of tourists' behavior may also be collected via neurophysiological sensors such as electroencephalography (EEG) (Li et al., 2022). Hence, tourists' attitudes towards a specific attribute can be simultaneously examined via secondary big data and EEG experiments to triangulate the results and thus provide more valid managerial implications.

This research also exhibits some limitations and thus calls for future review works. Firstly, the findings of this study can be updated in future to capture the most recent dynamics in this research domain. Secondly, the coverage of publications included in this study can be expanded further by employing more online library databases or adding more keywords while searching for literature records. It will, however, challenge the accuracy of retrieved literature records, and extra efforts are thus required to carry out more intensive screening. Although the practice of manual screening can significantly reduce irrelevant items in retrieved literature database, it can be time-consuming and may not be a representative of full accuracy. In future scientometric research, more advanced automatic filtering processes can be developed to ensure accuracy and improve filtering efficiency.

Declaration of Competing Interest

None.

Data Availability

Data will be made available on request.

Acknowledgement

The authors would like to acknowledge the National Natural Science Foundation of China (Grant No. 72004106).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ijhm.2023.103633.

References

Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., Weaven, S., 2019. Market segmentation and travel choice prediction in spa hotels through TripAdvisor's online reviews. Int. J. Hosp. Manag. 80, 52–77.
Alaoui, I.E., Gahi, Y., 2019. The impact of big data quality on sentiment analysis approaches. Procedia Comput. Sci. 160, 803–810.
Altin, L., Ahas, R., Silm, S., Saluveer, E., 2022. Megastar concerts in tourism: A study using mobile phone data. Scand. J. Hosp. Tour. 22 (2), 161–180.
An, Q., Ma, Y., Du, Q., Xiang, Z., Fan, W., 2020. Role of user-generated photos in online hotel reviews: an analytical approach. J. Hosp. Tour. Manag. 45, 633–640.
Ariyaluran Habeeb, R.A., Nasaruddin, F., Gani, A., Targio Hashem, I.A., Ahmed, E., Imran, M., 2019. Real-time big data processing for anomaly detection: a survey. Int. J. Inf. Manag. 45, 289–307.
Berezina, K., Bilgihan, A., Cobanoglu, C., Okumus, F., 2016. Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. J. Hosp. Mark. Manag. 25 (1), 1–24.
Buhalis, D., Law, R., 2008. Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tour. Manag. 29 (4), 609–623.
Buhalis, D., Sinarta, Y., 2019. Real-time co-creation and nowness service: lessons from tourism and hospitality. J. Travel Tour. Mark. 36 (5), 563–582.
Buning, R.J., Lulla, V., 2021. Visitor bikeshare usage: tracking visitor spatiotemporal behavior using big data. J. Sustain. Tour. 29 (4), 711–731.
Carvalho, B.P., Peralta, S., Pereira dos Santos, J., 2022. Regional and sectorial impacts of the Covid-19 crisis: evidence from electronic payments. J. Reg. Sci. 62 (3), 757–798.
Chen, C., 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. J. Am. Soc. Inf. Sci. Technol. 57 (3), 359–377.
Chen, Y., 2021. Public perception of tourism cities before and during the COVID-19 pandemic through the lens of user-generated content. Sustainability 13 (24), 14046.
Cheng, M., Anderson, C.K., 2021. Search engine consumer journeys: exploring and segmenting click-through behaviors. Cornell Hosp. Q. 62 (2), 198–214.
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. Tour. Manag. 71, 366–377.
Chua, A., Servillo, L., Marcheggiani, E., Moere, A.V., 2016. Mapping cilento: using geotagged social media data to characterize tourist flows in southern Italy. Tour. Manag. 57, 295–310.
D'Acunto, D., Tuan, A., Dalli, D., Viglia, G., Okumus, F., 2020. Do consumers care about CSR in their online reviews? An empirical analysis. Int. J. Hosp. Manag. 85, 102342.
D'Acunto, D., Volo, S., Filieri, R., 2021. "Most Americans like their privacy." exploring privacy concerns through US guests' reviews. Int. J. Contemp. Hosp. Manag. 33 (8), 2773–2798.
D'Acunto, D., Filieri, R., Amato, S., 2023. Hotels' environmentally-framed eWOM. The moderating role of environmental culture. Tour. Manag. 98, 104776.
Davras, O., Durgun, S., 2022. Evaluation of precautionary measures taken for Covid-19 in the hospitality industry during pandemic. J. Qual. Assur. Hosp. Tour. 23 (4), 960–982.
Deng, N., Liu, J., 2021. Where did you take those photos? Tourists' preference clustering based on facial and background recognition. J. Destin. Mark. Manag. 21, 100632.
Donaire, J.A., Galí, N., Camprubi, R., 2021. Empty summer: international tourist behavior in Spain during COVID-19. Sustainability 13 (8), 4356.

Gallego, I., Font, X., 2021. Changes in air passenger demand as a result of the COVID-19 crisis: Using Big Data to inform tourism policy. J. Sustain. Tour. 29 (9), 1470–1489.
García-Palomares, J.C., Gutiérrez, J., Mínguez, C., 2015. Identification of tourist hot spots based on social networks: a comparative analysis of European metropolises using photo-sharing services and GIS. Appl. Geogr. 63, 408–417.
George, G., Osinga, E.C., Lavie, D., Scott, B.A., 2016. Big data and data science methods for management research. Acad. Manag. J. 59 (5), 1493–1507.
Guo, Y., Barnes, S.J., Jia, Q., 2017. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. Tour. Manag. 59, 467–483.
He, F., Liu, H., Liu, C., Bao, G., 2023. Analysis of radar technology identification model for potential geologic hazard based on convolutional neural network and Harris Hawks optimization algorithm. Soft Comput. 27 (6), 3493–3507.
Hu, F., 2020. What makes a hotel review helpful? An information requirement perspective. J. Hosp. Mark. Manag. 29 (5), 571–591.
Hu, F., Teichert, T., Deng, S., Liu, Y., Zhou, G., 2021. Dealing with pandemics: An investigation of the effects of COVID-19 on customers' evaluations of hospitality services. Tour. Manag. 85, 104320.
Huang, B., Hao, H., 2021. A novel two-step procedure for tourism demand forecasting. Curr. Issues Tour. 24 (9), 1199–1210.
Ito, T.A., Larsen, J.T., Smith, N.K., Cacioppo, J.T., 1998. Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. J. Personal. Soc. Psychol. 75, 887–900.
Jia, S. (Sixue, 2021. Analyzing restaurant customers' evolution of dining patterns and satisfaction during COVID-19 for sustainable business insights. Sustainability 13 (9), 4981.
Kim, J.M., Liu, J., Park, K.K., 2023. The dynamics in asymmetric effects of multi-attributes on customer satisfaction: evidence from COVID-19. Int. J. Contemp. Hosp. Manag.
Kim, W.G., Park, S.A., 2017. Social media review rating versus traditional customer satisfaction: which one has more incremental predictive power in explaining hotel performance? Int. J. Contemp. Hosp. Manag. 29 (2), 784–802.
Kubo, T., Uryu, S., Yamano, H., Tsuge, T., Yamakita, T., Shirayama, Y., 2020. Mobile phone network data reveal nationwide economic value of coastal tourism under climate change. Tour. Manag. 77, 104010.
Kwok, L., Xie, K.L., 2016. Factors contributing to the helpfulness of online hotel reviews: does manager response play a role? Int. J. Contemp. Hosp. Manag. 28 (10), 2156–2177.
Lai, X., Wang, F., Wang, X., 2021. Asymmetric relationship between customer sentiment and online hotel ratings: the moderating effects of review characteristics. Int. J. Contemp. Hosp. Manag. 33 (6), 2137–2156.
Lee, H.Y., Leung, K.Y.K., 2022. Island ferry travel during COVID-19: charting the recovery of local tourism in Hong Kong. Curr. Issues Tour. 25 (1), 76–93.
Lee, M., Lee, S., (Ally), Koh, Y., 2019. Multisensory experience for enhancing hotel guest experience: empirical evidence from big data analytics. Int. J. Contemp. Hosp. Manag. 31 (11), 4313–4337.
Lee, M., Kwon, W., Back, K.J., 2021. Artificial intelligence for hospitality big data analytics: developing a prediction model of restaurant review helpfulness for customer decision-making. Int. J. Contemp. Hosp. Manag. 33 (6), 2117–2136.
Lee, Y., Kim, D.-Y., 2020. The decision tree for longer-stay hotel guest: the relationship between hotel booking determinants and geographical distance. Int. J. Contemp. Hosp. Manag. 33 (6), 2264–2282.
Li, C., Kwok, L., Xie, K.L., Liu, J., Ye, Q., 2021. Let photos speak: the effect of user-generated visual content on hotel review helpfulness. J. Hosp. Tour. Res. 10963480211019112.
Li, C., Ye, Q., Nicolau, J.L., Liu, X., 2021. Smiley guests post long reviews! Int. J. Hosp. Manag. 96, 102963.
Li, H., Hu, M., Li, G., 2020. Forecasting tourism demand with multisource big data. Ann. Tour. Res. 83, 102912.
Li, H., Meng, F., Pan, B., 2020. How does review disconfirmation influence customer online review behavior? A mixed-method investigation. Int. J. Contemp. Hosp. Manag. 32 (11), 3685–3703.
Li, J., Xu, L., Tang, L., Wang, S., Li, L., 2018. Big data in tourism research: a literature review. Tour. Manag. 68, 301–323.
Li, S., Lyu, T., Chen, M., Zhang, P., 2022. The prospects of using EEG in tourism and hospitality research. J. Hosp. Tour. Res. 46 (1), 189–211.
Li, X., Law, R., 2020. Network analysis of big data research in tourism. Tour. Manag. Perspect. 33, 100608.
Litvin, S.W., Goldsmith, R.E., Pan, B., 2008. Electronic word-of-mouth in hospitality and tourism management. Tour. Manag. 29 (3), 458–468.
Liu, H., Liu, W., Wang, Y., 2021. A study on the influencing factors of tourism demand from mainland China to Hong Kong. J. Hosp. Tour. Res. 45 (1), 171–191.
Lu, W., Stepchenkova, S., 2015. User-generated content as a research mode in tourism and hospitality applications: Topics. Methods, Softw. *J. Hosp. Mark. Manag.* 24 (2), 119–154.
Lv, H., Shi, S., Gursoy, D., 2021. A look back and a leap forward: A review and synthesis of big data and artificial intelligence literature in hospitality and tourism. J. Hosp. Mark. Manag. 31 (2), 1–31.
Mariani, M., 2019. Big data and analytics in tourism and hospitality: A perspective article. Tour. Rev. 75 (1), 299–303.
Mariani, M., Baggio, R., 2021. Big data and analytics in hospitality and tourism: A systematic literature review. Int. J. Contemp. Hosp. Manag. 34 (1), 231–278.
Mariani, M., Borghi, M., 2021. Customers' evaluation of mechanical artificial intelligence in hospitality services: a study using online reviews analytics. Int. J. Contemp. Hosp. Manag. 33 (11), 3956–3976.
Mariani, M., Baggio, R., Fuchs, M., Höepken, W., 2018. Business intelligence and big data in hospitality and tourism: a systematic literature review. Int. J. Contemp. Hosp. Manag. 30 (12), 3514–3554.
Meehan, K., Lunney, T., Curran, K., McCaughey, A., 2016. Aggregating social media data with temporal and environmental context for recommendation in a mobile tour guide system. J. Hosp. Tour. Technol. 7 (3), 281–299.
Meek, S., Wilk, V., Lambert, C., 2021. A big data exploration of the informational and normative influences on the helpfulness of online restaurant reviews. J. Bus. Res. 125, 354–367.
Mikhailov, S., Kashevnik, A., 2021. Car tourist trajectory prediction based on bidirectional LSTM neural network. *Electronics* 10 (12), 1390.
Nilashi, M., Abumalloh, R.A., Minaei-Bidgoli, B., Abdu Zogaan, W., Alhargan, A., Mohd, S., Syed Azhar, S.N.F., Asadi, S., Samad, S., 2022. Revealing travellers' satisfaction during COVID-19 outbreak: Moderating role of service quality. J. Retail. Consum. Serv. 64, 102783.
Nyns, S., Schmitz, S., 2022. Using mobile data to evaluate unobserved tourist overnight stays. Tour. Manag. 89, 104453.
Orea-Giner, A., Fuentes-Moraleda, L., Villacé-Molinero, T., Muñoz-Mazón, A., Calero-Sanz, J., 2022. Does the implementation of robots in hotels influence the overall tripadvisor rating? A text mining analysis from the industry 5.0 approach. Tour. Manag. 93, 104586.
Pachucki, C., Grohs, R., Scholl-Grissemann, U., 2022. No story without a storyteller: the impact of the storyteller as a narrative element in online destination marketing. J. Travel Res. 61 (8), 1703–1718.
Pan, B., Chenguang, Wu, D., Song, H., 2012. Forecasting hotel room demand using search engine data. J. Hosp. Tour. Technol. 3 (3), 196–210.
Park, E., Kang, J., Choi, D., Han, J., 2020. Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. Curr. Issues Tour. 23 (5), 605–611.
Park, E., Park, J., Hu, M., 2021. Tourism demand forecasting with online news data mining. Ann. Tour. Res. 90, 103273.
Pramana, S., Paramartha, D.Y., Ermawan, G.Y., Deli, N.F., Srimulyani, W., 2022. Impact of COVID-19 pandemic on tourism in Indonesia. Curr. Issues Tour. 25 (15), 2422–2442.
Ren, M., Vu, H.Q., Li, G., Law, R., 2021. Large-scale comparative analyses of hotel photo content posted by managers and customers to review platforms based on deep learning: Implications for hospitality marketers. J. Hosp. Mark. Manag. 30 (1), 96–119.
Saito, T., Takahashi, A., Koide, N., Ichifuji, Y., 2019. Application of online booking data to hotel revenue management. Int. J. Inf. Manag. 46, 37–53.
Sánchez-Lozano, G., Pereira, L.N., Chávez-Miranda, E., 2021. Big data hedonic pricing: econometric insights into room rates' determinants by hotel category. Tour. Manag. 85, 104308.
Sheng, J., Amankwah-Amoah, J., Wang, X., Khan, Z., 2019. Managerial responses to online reviews: a text analytics approach. Br. J. Manag. 30 (2), 315–327.
So, K.K.F., King, C., Sparks, B., 2014. Customer engagement with tourism brands: Scale development and validation. J. Hosp. Tour. Res. 38 (3), 304–329.
Sparks, B.A., Browning, V., 2011. The impact of online reviews on hotel booking intentions and perception of trust. Tour. Manag. 32 (6), 1310–1323.
Stamolampros, P., Korfiatis, N., Chalvatzis, K., Buhalis, D., 2019. Job satisfaction and employee turnover determinants in high contact services: Insights from employees' online reviews. Tour. Manag. 75, 130–147.
Sun, S., Wei, Y., Tsui, K.-L., Wang, S., 2019. Forecasting tourist arrivals with machine learning and internet search index. Tour. Manag. 70, 1–10.
Sun, S., Li, Y., Guo, J., Wang, S., 2021. Tourism demand forecasting: an ensemble deep learning approach. Tour. Econ. 13548166211025160.
Thelwall, M., 2008. Bibliometrics to webometrics. J. Inf. Sci. 34 (4), 605–621.
Turner, S.F., Cardinal, L.B., Burton, R.M., 2017. Research design for mixed methods: a triangulation-based framework and roadmap. Organ. Res. Methods 20 (2), 243–267.
Van Eck, N., Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84 (2), 523–538.
Verhoef, P., Kooge, E., Walk, N., 2016. Creating value with big data analytics: making smarter marketing decisions. Routledge,, London.
Vu, H.Q., Li, G., Law, R., Ye, B.H., 2015. Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. Tour. Manag. 46, 222–232.
Wang, Y., Correia, G.H., de, A., van Arem, B., Timmermans, H.J.P., (Harry), 2018. Understanding travellers' preferences for different types of trip destination based on mobile internet usage data. Transp. Res. Part C: Emerg. Technol. 90, 247–259.
Wen, L., Liu, C., Song, H., Liu, H., 2021. Forecasting tourism demand with an improved mixed data sampling model. J. Travel Res. 60 (2), 336–353.
Wu, E.H.C., Hu, J., Chen, R., 2022. Monitoring and forecasting COVID-19 impacts on hotel occupancy rates with daily visitor arrivals and search queries. Curr. Issues Tour. 25 (3), 490–507.
Wu, F., Zhang, Q., Law, R., Zheng, T., 2020. Fluctuations in Hong Kong hotel industry room rates under the 2019 Novel Coronavirus (COVID-19) outbreak: Evidence from big data on OTA channels. Sustainability 12 (18), 7709.
Xiang, Z., Gretzel, U., 2010. Role of social media in online travel information search. Tour. Manag. 31 (2), 179–188.
Xiang, Z., Schwartz, Z., Gerdes, J.H., Uysal, M., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? Int. J. Hosp. Manag. 44, 120–130.
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. Tour. Manag. 58, 51–65.
Xie, K., Kwok, L., Wang, W., 2017. Monetizing managerial responses on TripAdvisor: performance implications across hotel classes. Cornell Hosp. Q. 58 (3), 240–252.

Xu, Y., Zhang, Z., Law, R., Zhang, Z., 2020. Effects of online reviews and managerial responses from a review manipulation perspective. Curr. Issues Tour. 23 (17), 2207–2222.

Yang, S., Han, R., Wolfram, D., Zhao, Y., 2016. Visualizing the intellectual structure of information science (2006–2015): Introducing author keyword coupling analysis. J. Informetr. 10 (1), 132–150.

Yang, X., Pan, B., Evans, J.A., Lv, B., 2015. Forecasting Chinese tourist volume with search engine data. Tour. Manag. 46, 386–397.

Yang, Y., Pan, B., Song, H., 2014. Predicting hotel demand using destination marketing organization's web traffic data. J. Travel Res. 53 (4), 433–447.

Yang, Y., Fan, Y., Jiang, L., Liu, X., 2022. Search query and tourism forecasting during the pandemic: When and where can digital footprints be helpful as predictors? Ann. Tour. Res. 93, 103365.

Ye, Q., Law, R., Gu, B., 2009. The impact of online user reviews on hotel room sales. Int. J. Hosp. Manag. 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. Comput. Hum. Behav. 27 (2), 634–639.

Zeng, B., Gerritsen, R., 2014. What do we know about social media in tourism? A review. Tour. Manag. Perspect. 10, 27–36.

Zha, J., Tan, T., Ma, S., He, L., Filimonau, V., 2022. Exploring tourist opinion expression on COVID-19 and policy response to the pandemic's occurrence through a content analysis of an online petition platform. Curr. Issues Tour. 25 (2), 261–286.

Zhang, S.-N., Li, Y.-Q., Ruan, W.-Q., Liu, C.-H., 2022. Would you enjoy virtual travel? The characteristics and causes of virtual tourists' sentiment under the influence of the COVID-19 pandemic. Tour. Manag. 88, 104429.

Zhang, Y., 2021. A big-data analysis of public perceptions of service robots amid COVID-19. Adv. Hosp. Tour. Res. (AHTR) 9 (1), 234–242.

Zhang, Y., Lingyi, M., Peixue, L., Lu, Y., Zhang, J., 2021. COVID-19's impact on tourism: Will compensatory travel intention appear? Asia Pac. J. Tour. Res. 26 (7), 732–747.

Zhao, Y., Xu, X., Wang, M., 2019. Predicting overall customer satisfaction: big data evidence from hotel online textual reviews. Int. J. Hosp. Manag. 76, 111–121.

Zhu, J., Cheng, M., Li, Z., 2021. Peer-to-peer accommodation experience and guest actual recommendations: a novel mixed-method approach. Tour. Manag. Perspect. 38, 100816.

Zopiatis, A., Pericleous, K., Theofanous, Y., 2021. COVID-19 and hospitality and tourism research: an integrative review. J. Hosp. Tour. Manag. 48, 275–279.