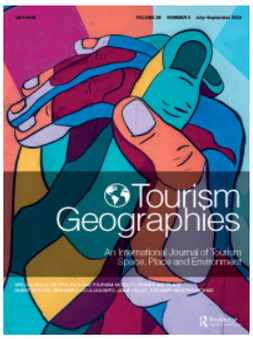

Routledge
Taylor & Francis Group

Tourism Geographies

An International Journal of Tourism Space, Place and Environment

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rtxg20

Tracking tourist mobility in the big data era: insights from data, theory, and future directions

Jinyan Chen, Noam Shoval & Bela Stantic

To cite this article: Jinyan Chen, Noam Shoval & Bela Stantic (19 Apr 2024): Tracking tourist mobility in the big data era: insights from data, theory, and future directions, Tourism Geographies, DOI: 10.1080/14616688.2024.2341249

To link to this article: https://doi.org/10.1080/14616688.2024.2341249

TOURISM GEOGRAPHIES
https://doi.org/10.1080/14616688.2024.2341249

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Tracking tourist mobility in the big data era: insights from data, theory, and future directions

Jinyan Chen[a], Noam Shoval[b] and Bela Stantic[c]

[a]School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong SAR, China; [b]Department of Geography, The Hebrew University of Jerusalem, Jerusalem, Isreal; [c]Information and Communication Technology, Griffith University, Nathan, Australia

ABSTRACT

With the increasing popularity of big data analytics, significant research has been conducted in the field of tourism and hospitality studies, particularly in the form of reviews that examine current works. Even though tourism is closely intertwined with the movements of tourists, a comprehensive review that investigates how big data analytics is utilised to track tourist mobility is still lacking. Hence, the primary objective of this study is to delve into the understanding of tracking tourists' mobility through the utilisation of big data analytics, considering data sources, methodologies, theoretical contributions, limitations, and future research directions. To accomplish this, an extensive literature review was conducted, encompassing publications from tourism and hospitality journals spanning a decade, from 2013 to 2023. This paper thoroughly examines five distinct types of data sources and explains how they enable researchers to monitor and track tourists' mobility for various research objectives. Moreover, the paper contributes to the academic discourse by identifying the gap in the literature where the application of theoretical frameworks has not kept pace with the advancements in data collection and analysis technologies. To bridge this gap, we introduce an innovative conceptual model that aligns the theoretical aspects of tourist studies with the practical application of big data, thereby offering a richer understanding of tourist mobility patterns. In addition, the research identifies and discusses the limitations of current studies, such as concerns regarding data privacy, representativeness, and the dichotomy between qualitative and quantitative data analysis in the field. By establishing a foundational framework and advocating for theoretical development, this paper sets a new standard for the integration of big data into tourism research, paving the way for future scholarly exploration.

ARTICLE HISTORY
Received 5 December 2023
Accepted 5 April 2024

KEYWORDS
Tourist mobility; Big data; Tourist movements; Tourist flow; Social media; GPS

CONTACT Jinyan Chen Jinyan.chen@polyu.edu.hk School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong SAR, China

1. Introduction

Tourist mobility refers to the movement and travel patterns of tourists, which encompasses the decisions and choices in which tourists navigate through different destinations and engage in activities, such as visiting attractions, exploring new cities or countries, and experiencing different cultures. Tracking and analysing tourist mobility patterns can provide valuable insights for understanding tourist behaviours, preferences, and trends, which can be useful for destination planning, marketing, mitigating negative outcomes for the local population and improving overall tourist experiences. In the context of big data analytics, tracking tourist mobility nowadays involves utilising large volumes of data to monitor and analyse tourists' movements derived from sources such as mobile devices, social media, check-ins, transportation data, and other digital footprints. Tourists move between destinations, and these movements create trajectories that sometimes share similarities or vary in terms of itineraries, which contain valuable information that can be analysed for tourism product development, such as route recommendations, helping to improve marketing services or assisting infrastructure and transportation development.

Tourist mobility traditionally relied on small-scale data collected through interviews, surveys, or direct observation. However, the emergence of big data, which is a broad term that refers to the large and complex data that are generated by various sources, and the wide utilisation of big data analytics in tourism and hospitality has garnered significant attention. The utilisation of various data sources and advanced analytics methods has opened up new possibilities for studying tourists' behaviour, particularly concerning their mobility. Big data analytics has revolutionized the study of tourist mobility, providing a new dimension of understanding and insights (Chen et al., 2022b). Its ability to analyse vast amounts of data from various sources, such as social media, mobile devices, and transactions, allows for a comprehensive and accurate representation of tourist movements. This granularity enables researchers to examine specific travel patterns and behaviours, leading to more targeted and personalised destination strategies. Additionally, big data provides spatial and temporal insights into tourist mobility, identifying popular attractions, routes, and optimal resource allocation. The richness of data and metadata, including textual content and images, captures subjective aspects beyond traditional quantitative data, enhancing the understanding of tourist experiences. Furthermore, predictive analytics enables the anticipation of future tourist behaviours. Collectively, these features of big data analytics enrich the study of tourist mobility, empowering destination managers with comprehensive insights for decision-making and strategy development.

Despite the benefits of big data, there are also some limitations and criticisms of using big data analytics in tourism research, and in many cases, the points raised are contradicting. To get a clear view of tracking tourist mobility in the big data era based on the information and debates that research has in tourist mobility literature, this review aims to analyse relevant literature and present the current knowledge of using big data analytics to track tourist movements, identify major challenges, and to suggest future directions. Although there are some relevant review work on tourist mobility (Chen et al. 2021b; Gomezelj, 2016; Mariani & Baggio, 2021), they have not considered the role of big data or how it enriches the mobility study. For example,

Chen et al. (2021b) discussed the role of social media only in studying tourist mobility; Li et al. (2018) explained the big data in tourism research. But they also do not focus on the topic of big data and tourist mobility. Therefore, this work is built based on the concerns raised from current work about big data and tourist mobility, hoping to expand the knowledge of different data sources used, study spatial scale, movements, and density to identify limitations and elaborate on potential future directions. Specifically, this review plans to answer the following questions:

1. What are the characteristics of big data, and how do they contribute to tracking tourist mobility?
2. How can the integration of theoretical frameworks and practical implications enhance the study of tourist mobility patterns using big data analytics?
3. What are the existing limitations and shortcomings in the current studies that rely on big data sources to track tourist mobility?
4. What are the major challenges and future directions in utilising big data to track and analyse tourist mobility patterns effectively?

2. Literature review

2.1. Tourist mobility

Tourist mobility, which encompasses the movement patterns of visitors between and within various destinations, has been an essential component of tourism geography and has significant implications for destination management and marketing (Birenboim et al., 2013; Mashkov & Shoval, 2023; Mckercher et al., 2011). Tourist mobility between different destinations creates trajectories, which are the spatial paths or routes followed during travel. Effective tourist mobility is crucial for accessing attractions, accommodations, and other essential services, enhancing overall travel experiences, and contributing to the economic development of tourism destinations (Baggio & Scaglione, 2018; Cheng et al., 2023).

Tracking and analysing tourist mobility patterns refers to the collection, processing, and interpretation of data related to the movement of tourists. This information can be obtained from various data sources like travel surveys, social media, and transportation systems (Chen et al., 2021b; Cheng, 2024; Shoval et al., 2018b). Analytics of tourist mobility patterns can reveal valuable insights into tourist behaviour, preferences, and travel trends. These insights can be used by tourism industry professionals, governments, and researchers to optimize transportation systems, plan marketing strategies, and develop tourism infrastructure and services to improve overall traveller experiences and economic growth. For example, understanding popular tourist trajectories can inform the development of efficient transportation networks, optimize the placement of tourism infrastructure, and improve overall travel experiences.

2.2. Big data

Big data is a broad term that refers to the large and complex data that are generated by various sources, it is a collection of everything: traditional legacy databases, data

warehouses, social networking, open data, web pages, blogs, logs, data streams generated by remote sensors, Internet-of-Things, genomics, etc. Big data is challenging the typical current state-of-the-art technologies to be handled within a reasonable time frame. There are many challenges including how to capture, curate, store, efficiently search, share, transfer, analyse and visualize the data. However, what we call big data today, might not be big data in 10 or even 5 years because the 'typical' tools and technologies are advancing rapidly.

Big data is often defined with four Vs (*Volume, Velocity, Veracity,* and *Variety*) (Stantic et al., 2014) because it is not just about the vast *Volume* of data but also refers to the speed at which data is being produced and how quickly data must be processed (*Velocity*). In addition, big data has complexity (structured, unstructured, semi-structured) and is in diverse data formats (text, numbers, images, audio, videos, etc), which is defined by *Variety*. While the *Veracity* refers to the trustworthiness of data. The more data is collected and automatically analysed (due to the high volume and velocity), the higher the uncertainty is about the accuracy of data. For example, it is particularly challenging to verify the truthfulness of posts on social media platforms. Detecting fake posts, fake reviews, and even fake friends is an active research area (Nguyen et al., 2019).

In addition to often considered four V's, the literature mentioned other V's that define the big data, for example, *Value* – indicates if the data is worthwhile and has value for the business, *Variability* – indicating the different meanings associated with a given piece of data, and *Volatility* – addressing the problem of how long data is valid and how long should it be stored or is it no longer relevant to the current analysis and decisions (Stantic et al., 2014). Big data is often used in various fields such as tourism, finance, and healthcare to help businesses to gain insight into market trends, and customer behaviour, to improve the efficiency and effectiveness of businesses.

2.3. The connection between tourist mobility and big data

Tourist mobility traditionally relied on small-scale data collected through interviews, surveys, or direct observation. However, the emergence of big data analytics in tourism and hospitality has garnered significant attention. The utilisation of various data sources and advanced analytics methods has opened up new possibilities for studying tourists' behaviour, particularly concerning their mobility (Li et al., 2018; Shoval & Isaacson, 2009). Social media data, for instance, has proven valuable in understanding tourist movements over broader geographic areas, thanks to its rich content comprising locations, time, and textual information (Chen et al., 2021b). Nevertheless, it is important to note that social media is not the sole source of big data for mobility analysis. Li et al. (2018) identified three other types of big data sources: User Generated Content (UGC), device-generated data (e.g. mobile phone data), and transaction data (e.g. data obtained from webpages). However, as Li et al. (2018) review focused on a broader spectrum of big data applications within tourism and hospitality, many studies addressing other areas lack specifics regarding tourist mobility. Consequently, there remains a clear gap in the literature regarding a review investigating tourism mobility and its intersection with big data analytics.

Big data analytics has revolutionized the study of tourist mobility, providing a new dimension of understanding and insights (Chen et al., 2022b). Its ability to analyse vast amounts of data from various sources, such as social media, mobile devices, and transactions, allows for a comprehensive and accurate representation of tourist movements. This granularity enables researchers to examine specific travel patterns and behaviours, leading to more targeted and personalised destination strategies. Additionally, big data provides spatial and temporal insights into tourist mobility, identifying popular attractions, routes, and optimal resource allocation. The richness of data and metadata, including textual content and images, captures subjective aspects beyond traditional quantitative data, enhancing the understanding of tourist experiences. Furthermore, predictive analytics enables the anticipation of future tourist behaviours and integrates multiple data sources, offering a holistic view of mobility (Park et al., 2021). Collectively, these features of big data analytics enrich the study of tourist mobility, empowering destination managers with comprehensive insights for decision-making and strategy development.

Despite the benefits of such data, there are also limitations and criticisms of using big data analytics in tourism research. Some studies pointed out privacy issues (Hardy, 2020), while McKercher et al. (2021) mentioned that sometimes big data analytics could be descriptive and relevant qualitative data would be better to support the studies. However, Chen et al. (2022a) showed that extracting information from text data to identify tourists' behaviour is also possible. We will elaborate on these challenges in this work.

3. Methodology

To get insights into the state of tracking tourist mobility in the big data era we conducted a systematic literature review of research articles found in Scopus and Web of Science databases. We have selected Scopus and Web of Science as sources of articles because these databases are widely used for academic research in various fields, including hospitality and tourism. They are known for their extensive coverage of peer-reviewed literature and provide a large number of academic publications. Also, these databases offer a systematic and comprehensive way to retrieve relevant research articles for the study.

3.1. Literature selection

This paper reviews relevant literature and uses qualitative and quantitative methods to analyse papers to answer the proposed questions. For the systematic literature review, we followed widely accepted methods in the hospitality and tourism domain (Gomezelj, 2016; Mariani & Baggio, 2021). Since this review specifically targets understanding tourist mobility, we only limited the subject area to social sciences and searched only for journal articles. We looked for papers published in the English language since 2013. Also, because the review plans to focus on tourism research, we only selected journals recognised by *Scimago Journal & Country Rank* in the subject of *Tourism, Leisure and Hospitality Management*. Since this paper focuses on using big data to understand tourists' mobility, relevant keywords were developed to search

the literature. Given that the recent relevant review was conducted on social media data (Chen et al., 2021b), this paper will further expand the knowledge of different data sources used, identify limitations, and elaborate on potential future directions to explore information in-depth by addressing the research questions.

The process of selecting relevant papers involved three steps. In the first step, after initial reading of relevant papers, keywords were identified to be searched on Scopus and Web of Science. The search keywords were: 'TITLE-ABS KEY (*tourist* OR *tourism* OR *visitor* OR *traveller*)' combined with mobility-related words AND 'TITLE-ABS-KEY (*pattern* OR *visitation* OR *mobility* OR *footprint* OR *flow* OR *network* OR *movement* OR *trajectory*' and the condition related to social media platforms' AND TITLE-ABS KEY '*big data*' OR '*social media*' OR *bluetooth* OR '*mobile phone*' OR '*location tracking*').

The search on Scopus and Web Science returned 4,482 papers. The second step involved filtering by reviewing the titles and abstracts to discard out-of-scope papers (e.g. non-English or non-relevant journals), which led to 112 papers that addressed the focus of this review. Considering the specificity of the reviewed topic, due to the cross-disciplinary research, the relevance, and its high impact factor, we decided to also take into consideration articles from the Journal of Information Technology and Tourism despite being outside of the accepted subject of Tourism, Leisure and Hospitality Management. The search of this journal returned 35 papers, and after the second filtering step, it resulted in 9 relevant papers. A total of 121 papers are considered for further analysis. This focus was defined as research explicitly related to big data analytics-derived travel patterns, flows or visitation insights. Despite there are relevant conference publications such as Wang et al. (2022) to maintain consistency and ensure that only publications from high-impact journals are taken into consideration, conference publications are not taken into consideration in analysis.

Clearly, using big data to analyse tourist mobility is an emerging topic, which is evident as the number of publications per year has been steadily increasing, particularly in the last six years, with more than 83% of papers being published since 2018. Figure 1 also shows the top ten journals with over 90% of papers published. The top three journals that published on this topic are Tourism Management (26), followed by Tourism Geographies (21), and Annals of Tourism Research (13). The trends shown in Figure 1 also clearly indicate the topic's importance with increasing popular trends. Therefore, a comprehensive review of how big data has been utilised in analysing tourist mobility is needed. Additionally, identifying future research directions in this area will be crucial for researchers to explore new avenues and contribute to advancing knowledge in the field. These findings emphasise the significance and relevance of studying the use of big data analytics in tracking tourist mobility and the potential for further research in this domain.

3.2. Quantitative and qualitative analysis methods

We relied on the use of qualitative methods to analyse papers and answer the proposed questions. This analysis is performed inductively, meaning that the themes or concepts are not preconceived but rather emerge from the analysis of literature. Based on the details provided in the papers after careful manual analyses we utilized clustering methods to group and count different aspects including: 'broad grouping of

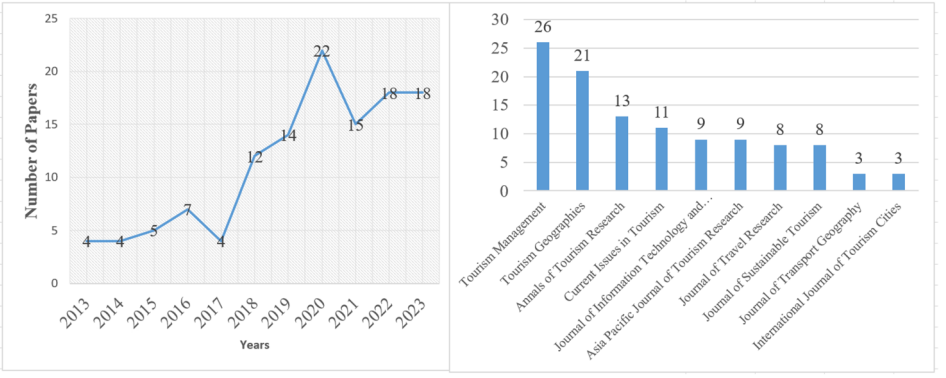

Figure 1. Trend of the publications over the years.

data sources', 'type and specific sources of data', and 'theories applied'. The same method was applied to identify and group 'goals and research aims'.

4. Results and analysis

Following the proposed methodology in this section, we provide results obtained to answer the research questions. Specifically, we present findings about the data source used, by identifying and grouping them into three levels: brood groups, types, and specific sources. We also provide details of the intensity of utilisation regarding all three levels. Furthermore, we perform a comparison of the big data Source used in analysing tourist mobility regarding the different aspects such as 'access and collection', 'spatial-temporal resolution', 'cost-effectiveness', 'data size', and 'completeness of the trajectory'. We also present findings related to the 'goals and research aims' and identify and cluster on 'main research aims' and 'study goals'. Finally, we present findings related to the 'theory contributions'.

4.1. Data source used in analysing tourist mobility by big data

Data has always been the fundamental element in big data analysis, and specific characteristics in big data analytics present challenges and opportunities to utilise the data. After a deeper analysis of the identified paper, this review will list big data sources and explain in detail the advantages, limitations, and challenges of adopting different data sources.

According to the literature, big data analytics in tourism and hospitality studies can be clustered into three main data source groups: user-generated data (for example, by social media), data generated from devices (such as mobile phones), and transaction data (such as online booking data) (Li et al., 2018). However, unlike the previous studies that focus on general research topics using big data in tourism and hospitality, this work specifically focuses on studying tourist mobility, so the data types are quite unique. Tourist mobility is one of the most important aspects of tourist activities. Studying mobility requires data that provides information about geographic locations or in the form of coordinates (e.g. latitude-longitude) or specific location

names (e.g. city names, attraction names) to identify tourists' movements. Additionally, if the sequences of the movements are needed in the research, location data must be paired with temporal information to model tourists' movement patterns or identify trajectories.

To understand different big data sources available for analysing tourist mobility, Table 1 is created to summarise the current data features from the identified literature. For instance, many studies in this review rely on User-Generated Content (UGC) from social media platforms such as Twitter and Weibo. Researchers often highlight the accuracy of social media data in terms of locations and time of posts, making it suitable for analysing tourist movements (Chen et al., 2021a). Bluetooth devices also receive significant attention, as they provide location and time series data, accounting for approximately 30% of data sources that considered studies relied on. This review also identified some new data sources, such as a 'GIS-based library', representing data

Table 1. Big data sources in studying Tourism Mobility.

Broad Grouping	Type	Specific source	Reference
User Generated Content	Social media	Twitter	(Barros et al., 2020; Bigné et al., 2019; Chen et al., 2023; Cheng et al., 2023; Encalada-Abarca et al., 2023; Frenzel et al., 2022; Höpken et al., 2020; Kim & Hyun, 2021; Kim et al., 2019; Park et al., 2019)
		Instagram	(Barros et al., 2020; Falk & Hagsten, 2021; Liu, Wang, Weber, et al., 2022; Mor et al., 2023; Ning et al., 2023; Zhou & Chen, 2023)
		Flickr	(Barros et al., 2020; Encalada-Abarca et al., 2023; Giglio et al., 2019; Höpken et al., 2020; Kim et al., 2019; Leung et al., 2017; Ma et al., 2020; Xu et al., 2022; Zhang et al., 2020)
		Weibo	(Chen et al., 2021a, 2022a; Xue & Zhang, 2020; Zhao et al., 2023; Zhou & Chen, 2023)
		TripAdvisor	(Nolasco-Cirugeda et al., 2022; Van der Zee & Bertocchi, 2018; Xiang et al., 2017)
		Google Map Reviews	(Höpken et al., 2019; Owuor et al., 2023)
		Wikipedia	(Owuor et al., 2023)
Data generated from devices	Mobile phones	mobile phones data	(Baggio & Scaglione, 2018; Chu & Chou, 2021; Hardy et al., 2017, 2022; Park et al., 2020, 2021, 2023; Raun et al., 2020; Shi et al., 2023; Spangenberg, 2014; Xu, Li, Belyi, et al., 2021)
	GPS devices	GPS-device	(Grinberger et al., 2014; Hardy et al., 2020, 2022; Hardy & Aryal, 2020; Mckercher et al., 2012; Shoval & Ahas, 2016)
		bike-share data	(Buning & Lulla, 2020)
GIS-based library	Image Library	ImageNet	(Payntar et al., 2021)
	GIS-trajectory library	LiangBulu.com	(Liu, Wang, Yang, et al., 2022)
Search Engine Data	Text and Time series	Google Trend	(Li, Li, et al., 2022)
	Text	Baidu Index	(Liu et al., 2019)
Sensor Data	Biological Data	Heart Beats, Blood Pressure	(Shoval et al., 2018a)
	Weather Data	Raining, Temperature	(McKercher et al., 2015)

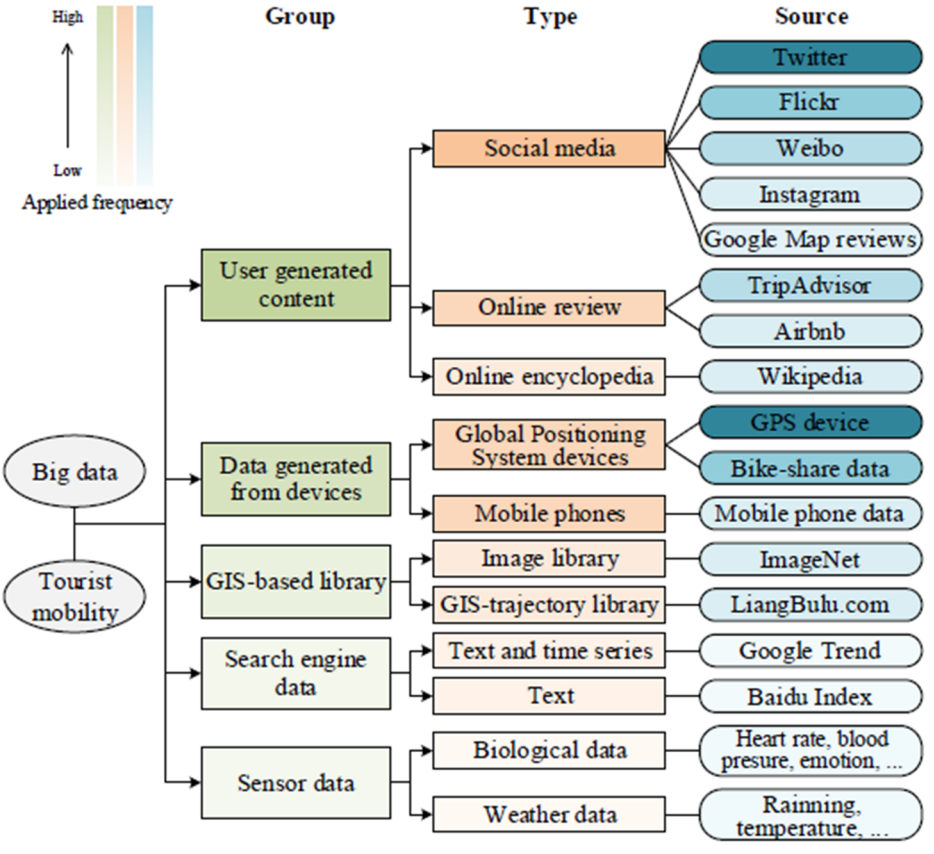

Figure 2. Data used for big data analytics and tourist mobility.

obtained from websites that provide locations and time to model visitors' trajectory and data from search engines (such as Google Analytics and Baidu search), which are text and time-based data.

In addition, we also show fine-grain classification of data sources used by dividing them into specific types to understand the variety of Big data sources utilized in recent studies. Figure 2 takes into account the number of studies and the types of data involved. Different colours represent the groups of data sources (green), followed by the data types (orange) and specific data sources (blue). The colour saturation used in the figure indicates the extent to which these data sources have been applied in studies, providing a clearer understanding of their popularity and relevance (a lighter saturation indicates a lower level of popularity and the opposite).

4.2. Comparing the big data source used in analysing tourist mobility

In general, User Generated Content is the most popular data source, with nearly 70% of papers relying on this data source. The most popular social media platforms were Twitter and Flicker, and this finding is consistent with the previous study by Chen et al. (2021b). With different types of available data sources, it is up to the researcher to choose the most suitable data that meets the study's specific needs. In Figure 3,

five typical data characteristics were identified (*Flexible access and collection, High spatial-temporal resolution, Cost-effective, Data sizes - richness and completion of the trajectory*) and rated based on their effectiveness in analysing mobility. These characteristics were identified based on clustering the characteristics of data types provided in the assessed literature. The rationale behind using these characteristics is that they allow for a comprehensive and accurate analysis of tourist mobility. Flexible access and collection refers to the ease with which data can be obtained and collected. High spatial-temporal resolution ensures that detailed information about tourist movements is captured. Cost-effectiveness is important as it reduces the time, energy, and budget spent on data collection. Data size richness stands for the energy, time and budget spent on data collection which ensures that there is sufficient data to analyse and identify trends and patterns. Completion of the trajectory is essential as it allows for a comprehensive understanding of tourist spatial and temporal activities. Therefore, these data characteristics are crucial in enabling accurate and effective analysis of tourist mobility using big data analytics.

The scale is from 1 to 3, with 1 being the least effective and 3 being the most effective. The three-point scale is considered to have sufficient granularity for the purpose of indicating the ranking of the specific data source regarding five identified characteristics. Researchers can follow the guidelines for different research directions and purposes to make better decisions when adopting different data types and align their decisions based on data characteristics performance and intended objectives.

Flexible access and collection stand for different methods of collecting data. In this category, social media can be considered the best option when compared with other sources of data because social media data collection is flexible and can be based on different keywords and different geo-location scales. Also, data can be collected near real-time, and adjusting parameters for data collection is relatively easy. Also, Social media data can be collected through an *Application Programming Interface* (API) or purposely built web crawlers with the flexibility of parameters for data collection. Collecting social media data is more efficient and cost-effective than surveys, which require surveying people at different locations.

High spatial-temporal resolution means the data has the exact location and time. Data generated from devices and GIS-based libraries has the best accuracy. For example, mobile phone data can record the point of interest at a specific time, and data can also show the exact point where tourists visit defined by the exact longitude and latitude of the location. Data can be associated with geo-coordinates (latitude and longitude) and time, which provide information on the location where the data was posted. The granularity of space can be very fine, such as a specific attraction; however, it is possible to aggregate data to higher levels, like a city, region, country or even globally. The posting time can also be aggregated to the minute, hour, day, month, season, or year to analyse travel patterns.

Cost-effectiveness stands for the energy, time, and budget spent on data collection. As mentioned earlier, UGC data can be collected through API or programming language. Moreover, in general, UGC data does not need to recruit volunteers to participate, and it can be flexible with dates and keywords for multiple rounds of data collection.

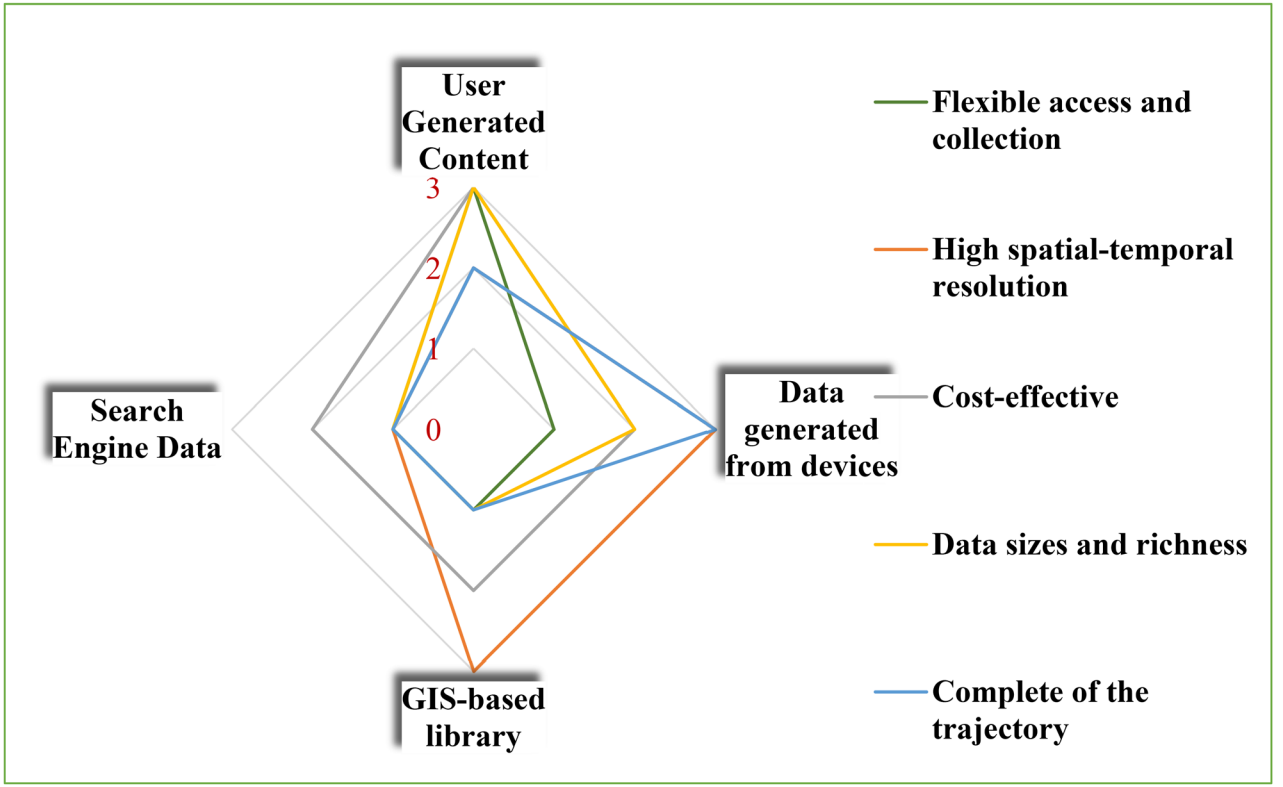

Figure 3. Ranking of five data characteristics for each identified data type.

Data size and richness are the amount of data and different content that the study can collect. Social media data not only comes in high volumes but also provides, in addition to the text, emoticons, images, and videos, as well as metadata (user info), which allows the exploration of visitor interests and levels of satisfaction towards diverse attributes and destinations.

4.2.1. Completeness of the trajectory

Having a complete tourist trajectory is essential for studying tourist mobility as it provides a comprehensive understanding of tourist spatial and temporal activities, which allows the accuracy of analysis. According to Figure 3, data generated from the device, particularly bluetooth and mobile phones, are the most effective regarding the completeness of the trajectory as it is the only form of identified data sources that has the highest value of 3.

In summary, social media as a form of user-generated data source has been well accepted in tourism studies, especially in modelling the travel patterns of visitors (Henderson, 2006; Leung et al., 2013; Önder & Marchiori, 2017; Wang et al., 2011). However, social media data also has its concerns in analysing tourist mobility. Chen et al. (2021) mentioned that utilising social media can be challenging to recognise tourists and locals. Chen et al. (2022) also stated that social media might face uncompleted trips if tourists do not post at all locations they visited. To overcome these concerns, different actions can be taken. Specifically, regarding the lack of explicit geographical coordinates, while it is true that not all social media posts have geo-location attached, it has been shown that in real projects, about 15% of Twitter posts have exact geo-locations reflecting the exact location at the time of posting (Becken et al., 2017).

Furthermore, when geo-location does not need to be fine-grained to the exact location, the metadata about the location the user provided as the place of residence can be used to identify the origin and distinguish locals from tourists. On the other hand, mobile phones provide similar objectives for modelling trajectory, such as the locations and time. However, mobile phone data provides more comprehensive trip details, as researchers could identify when tourists connected with the local signals and when they left. Also, from the background of the data, mobile phone data already provides the users' information. Therefore, it is more accurate to distinguish tourists' or locals' trajectories. However, unlike social media data that is publicly available, accessing mobile phone data is very difficult and costly. For example (Park et al., 2021; Raun et al., 2016) explained how they have collaborated with mobile phone network providers to access the highly-private data. Without such collaboration, access to mobile phone data is not possible and only on rare occasions, and few studies relied on mobile phone data. In addition, the biggest concern when using mobile phone data is how to access the data and the privacy. To work on the mobile phone data, literature always mentioned that they have collaborated with the provider (Park & Zhong, 2022; Raun et al., 2016), which presents difficulty in accessing data, leading to limited work on this data source.

Is there a data type that could reflect a full trip as well as identity tourists? Data generated from devices could be the solution. As proposed by Hardy et al. (2020;

Raun et al., 2020), they provided a device to the tourists and modelled their trajectory at a state level or studied the trajectory of wine tourists (Lewis et al., 2021). Also, using a bluetooth device is popular in assessing tourist mobility in the Park (Birenboim et al., 2013; Shoval et al., 2018a) or cruise (Hardy et al., 2022; Sciortino et al., 2022; Shoval et al., 2020). From the literature, it is also evident that bluetooth device-generated data works better in the clear boundary destination, preferably smaller areas, so it is practical to send and collect the bluetooth devices. Once the destination level becomes bigger, such as the national level or global level, it is challenging to use such methods.

Different data sources provide different valuable insights into tourist mobility, allowing researchers to understand movement patterns, destination choices, and trends. However, it is important to acknowledge that these sources may have limitations. For instance, in the case of using GIS-based libraries, it may be limited to spatial data available within the library and may not capture real-time movements or individual-level data. Similarly, using search engine data may suffer from user limitations and biases, as it relies on voluntary searches and may not capture the entire spectrum of tourist mobility.

4.3. Goals and research aims applied big data analytics in analysing tourist mobility

Tourism is often perceived as a distinct activity that diverges from everyday life. Consequently, the study of tourism encompasses a wide range of perspectives and intertwines various disciplines such as geography, sociology, and economics (Zheng et al., 2017). To gain a comprehensive understanding, it is imperative to examine the research goals and aims in this field. Table 2 presents the primary research aims that have been identified by extracting and analysing information from the literature. This systematic overview facilitates a structured analysis of the main objectives explored in the literature, thereby contributing to the advancement of relevant theories and broadening our insights into the areas of tourist mobility and big data studies.

Mobility is a fundamental aspect of tourism studies, and analysing mobility patterns has allowed researchers to address a myriad of research objectives within this context, as detailed in Table 2. This comprehensive understanding of visitor mobility not only enhances our grasp of travel behaviour but also facilitates the achievement of the stated research objectives. As evidenced by the analysis in Table 2, the spatial and temporal behaviour analysis of tourists is one of the most prevalent research goals. Given the nature of the topic, researchers are employing various data and analytical methods to study the spatial and temporal patterns of tourists. This invaluable information can guide the establishment of connections between destinations, assist in modifying transport infrastructure, and promote collaboration among tourism organisations (Asero et al., 2016), as destinations are often planned and managed based on different administrative boundaries and geographical factors. This is crucial as destinations are often planned and managed based on distinct administrative boundaries and geographical factors (Paulino et al., 2021).

The advent of big data has ushered in novel opportunities, supplementing traditional data sources and methodologies for scrutinizing travel patterns. Big data facilitates the gathering and analysis of voluminous data from a plethora of sources,

Table 2. Research aims of big data analytics in tourist mobility studies.

Main Research Aims	Achieved study goals	Example papers
Tourist spatial and temporal behaviour analysis	Examining how tourists make travel decisions, their preferences, and their behavioural patterns during their journeys. By leveraging big data analytics, researchers and organisations can gain insights into tourist behaviour on a large scale, enabling a deeper understanding of travel dynamics.	(Chen et al., 2022a; Grinberger et al., 2014; Shoval, 2006; Zhou et al., 2021)
Data science and tourist analysis techniques	Using tourist mobility to develop the algorithms for modelling the tourist's trajectory and pattern mining.	(Chen et al., 2021a; Park et al., 2021; Shoval et al., 2015; Xu, Li, Belyi, et al., 2021)
Destination branding and marketing	Understanding the impacts of tourism on residents and stakeholders and the role of destination branding in attracting international tourists	(Milano et al., 2023; Zheng et al., 2022)
Destination image and satisfaction	The impact of destination image and satisfaction on tourist behavioural intentions	(Chen et al., 2022a; Park & Song, 2021)
Event impacts on tourist movement:	Analysing movements during specific events - Understanding the impacts of mega events on tourism	(Cheng et al., 2023; Jin & Cheng, 2020)
Adapting to changing circumstances	Examining tourism during the pandemic - Understanding transformations in tourist landscapes, such as during the pandemic	(Li, Tao, et al., 2022; Lu et al., 2021; Xu, Li, Xue, et al., 2021)
Tourism impacts and social perspectives:	Study tourist mobility's influence on society, such as on residents and in the crisis.	(Meneghello, 2023; Milano et al., 2023)
Cultural heritage and special interests tourism	Study the extent of visitors to world heritage sites or special interest tourist groups movements, such as wine tourists.	(Falk & Hagsten, 2021; Lewis et al., 2021; Zhao et al., 2023)
Predicting tourist demands	Literature that works on predicting tourist arrivals and tourist demands.	(Bokelmann & Lessmann, 2019; Du et al., 2020; Lee et al., 2021; Zheng et al., 2017)
Factors that influence tourist mobility	Investigating the factors that influence tourist mobility, such as weather conditions, different travel profiles, urban or rural destinations	(Encalada-Abarca et al., 2022; McKercher et al., 2015; Shoval et al., 2018a; Vu et al., 2020; Zhao et al., 2018)

thereby aiding in the identification of concealed patterns, correlations, and trends that may not be discernible with conventional data analysis techniques. Consequently, a significant amount of research in this area is also aimed at 'Data science and tourist analysis techniques'. Leveraging advanced algorithms and machine learning methodologies, researchers can unearth insights and forecast visitor behaviour, preferences, and future travel patterns. Such predictive analysis can empower destination managers to devise targeted marketing campaigns, optimize resource allocation, and formulate sustainable tourism strategies.

Increasingly, research also looks for factors that may impact tourist mobility, such as weather conditions, cultural background and tourist interests. Furthermore, tourist mobility can also be used to analyse destination image and assess tourist satisfaction. In conclusion, understanding visitor dispersal and mobility is vital for effective destination management. Analyzing visitor movements at different scales and utilising big data analytics offer valuable insights into travel patterns through which researchers can obtain high-resolution spatial and temporal features of tourism activities to

facilitate decision-making for destination managers and enable them to make appropriate adjustments to transport infrastructure and foster cooperation between tourism organisations. The spatial interaction between destinations and the spatial distribution of visitors are important aspects to consider when planning and managing destinations. Embracing new data sources and technologies strengthens our understanding of visitor mobility, enhancing the overall understanding and management of the tourism industry.

4.4. Theory contributions in tracking tourist mobility with big data analytics

There has been a notable gap in theoretical contributions in the realm of big data analytics and its application in the study of tourism and hospitality (Li et al., 2018; Mariani & Baggio, 2021). Previous research has often focused on discovering patterns and testing hypotheses using data, yet lacking a strong foundation in theoretical frameworks. To address this gap, Mariani and Baggio (2021) emphasise the need for future studies to clearly articulate how big data analytics can be guided by theories and generate research questions based on theoretical frameworks. For big data analytics in tourist mobility, Chen et al. (2021b) mentioned in their review that there is a lack of theory contributions to serve the study goals as only 10% of papers in the review were found to apply a theory.

However, when the data source is beyond just social media, more theory input has been identified in the review. Unlike reviews that have broader concepts about the big data study in tourism and hospitality, tourist mobility focuses on tourist movements in different locations and timelines. Therefore, some theory that explains tourist movement patterns from a geographical point of view has been a popular area for identifying theory contributions. Figure 4 shows the identified theories from different fields in tourist mobility studies utilising big data analytics. These fields are identified upon analysis of reviewed papers and clustering on the field of research, which was directly indicated in the paper or implied by the paper content. We also identified typical representatives for each field. To provide a better understanding of the prevalence of individual fields and theories we use colour saturation indicating the frequency of usage of specific theories in assessed studies. For example, a lighter saturation associate with a lower number of works relied on a specific theory and the opposite.

One popular theory from Geography is time geography (Grinberger & Shoval, 2019; Shoval & Isaacson, 2009), which explains that tourist activity movements must include spatial and time dimensions in every movement. With different constraints in spatial and temporal resources, tourists will choose patterns or activities to maximise the utilities, which provides the theoretical foundation for tourist time-space behaviours (Park & Zhong, 2022; Zheng et al., 2022). Taking a human geography perspective when examining visitor movements necessitates an understanding of movements at different scales: global, international, regional, and local. Different geographical scales have been employed to analyse and model how visitors travel between destinations. The spatial-temporal relationship between destinations creates visitor flows, particularly evident in multi-destination trips and the resulting networks. These flows have been recognised as a positive phenomenon, as they enable spatial interaction between

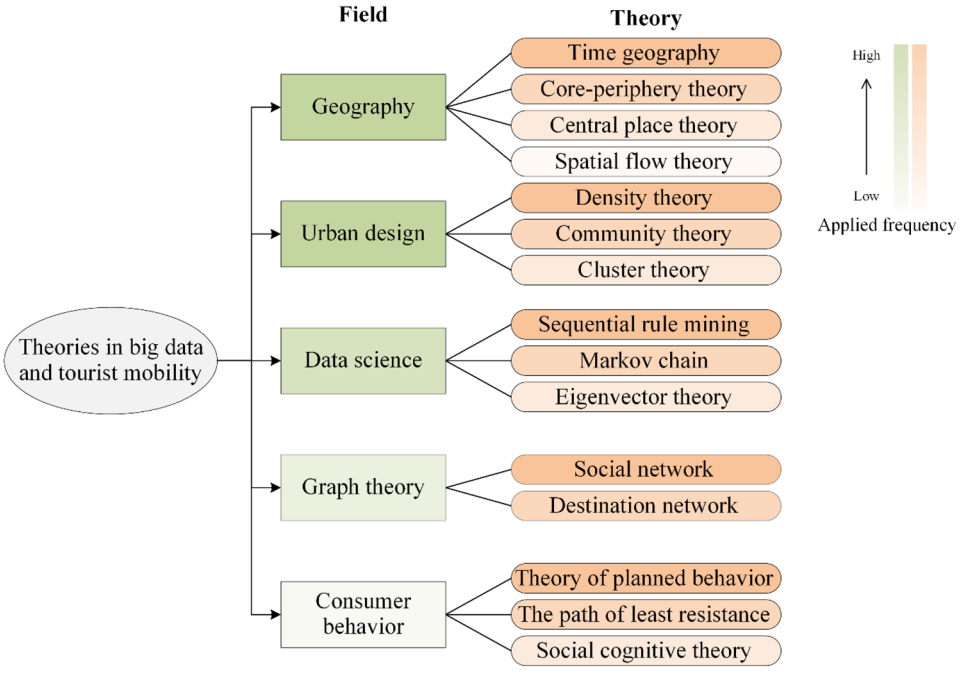

Figure 4. Theories applied in big data and tourist mobility studies.

destinations, providing distinct benefits such as helping city organisations identify suitable partner cities for promotion in specific markets (Grinberger & Shoval, 2019).

Another theory that explains tourist mobility and has been applied along big data is core-periphery from economic geography, which suggests that in different regions, there are always core areas, which are surrounded by social, political and economic activities and periphery, which is generally in suburban and rural areas (Fennell, 1996). In tourism, core destinations represent those that have major tourist resources. In theory, tourist mobility follows the core-periphery theory that most tourists visit the top attractions and then the surrounding places (Chen et al., 2022b). However, with the development of public transportation, region integration (such as the Greater Bay area in China) and destination promotion to rural areas, the core-periphery structure could meet a change (McKercher, 2021). Also, the movements of tourists from one place to another have been considered as a flow, and they normally have a clear starting and ending point, such as (Chen et al. (2021a) modelled the flow of tourists travelling from China to Australia and other countries. Li et al. (2022) discovered that tourist flows to destinations influence and even shape the overall population movement pattern, particularly in a destination that relies on tourism revenue.

Another group of theory contributions is from the distribution of tourists in the destinations, explaining how mobility can be analysed through the density of the crowds. For example, Falk and Hagsten (2021) mentioned that they found that tourists' perceptions heavily impact the concept of visitor density. While the entry numbers measure the traditional concept of tourist density, the number of social media posts of particular places could show better density from the tourist's point of view. The density of the tourists and their movements can also be further analysed for their

spatial and temporal patterns. Using the distribution of tourists in the destinations, Park et al. (2023) proposed that tourist density changes depending on different types of visitors (e.g. same-day visitors) and the intra-city tourist desperation can be important for urban tourism design. Other studies relied on the density theory to group the tourists based on the density to identify the group of attractions and the similarity of movements based on their interests (Giglio et al., 2020; Zhao et al., 2023; Zhou & Chen, 2023).

The theories from data science also play an important role in understanding tourist mobility, particularly those that focus on analysing tourist trajectories, such as the sequences of the paths, the similarity of the itineraries, and movement and flow patterns. For example, 'Sequential rule mining' refers to identifying correlated patterns between the data values in sequential order, which is popular in analysing tourist trajectories (Cheng et al., 2023; Shoval et al., 2015). The 'Markov Chain model', which explains a sequence of events, is a popular model for analysing tourist movement patterns (Liu, Wang, Yang, et al., 2022; Payntar et al., 2021).

Graph theory is another well-adapted aspect that researchers applied to analyse tourist mobility. Initially, a graph is constructed by sets of nodes and links, so a graph can be seen as a network linked by vertices and edges (Smith & Sarabi, 2022). By studying the characteristics of the network, such as the number of connections a node has, or the position of the node, the importance of the nodes and edges can be shown. In the tourist mobility study, researchers used social network analysis to identify the crucial locations or patterns of the itinerary. For example, Chen et al. (2022b) built a destination network using different cities in Australia as nodes and the number of tourists travelling among the cities as edges. They were able to identify the most attractive destinations and popular routes. Park et al. (2021) applied graph theory to build tourist travel trajectories by considering spatial and temporal dimensions.

Tourists are also a type of customers that consume tourist products, such as purchasing tickets to attractions and experiencing different services. To understand tourist travel behaviour, research also adopted theory from customer behaviours. From a mobility study point of view, the literature shows that most studies focus on factors that influence tourists' travel patterns when applying 'customer behaviour theory'. For example, using the concept of 'the path of least resistance', Türk et al. (2021) built a spatial hedonic price model with geoscience techniques to understand the changes in mobility patterns and how they are reflected in the accommodation (Airbnb in their study). They found that tourists prefer to live in areas that are close to the urban core for easier access to prominent tourist attractions and public transport. Also, the 'theory of planned behaviour' explains how mobility has been impacted by the pandemic and how the perceived risks have influenced different age groups' mobility (Yu et al., 2023).

5. Discussion

Based on the analysis of the literature, we generated the research process to guide future studies in this area, shown in Figure 5. Specifically, we list and elaborate on specific challenges researchers face when utilising big data to track tourist mobility.

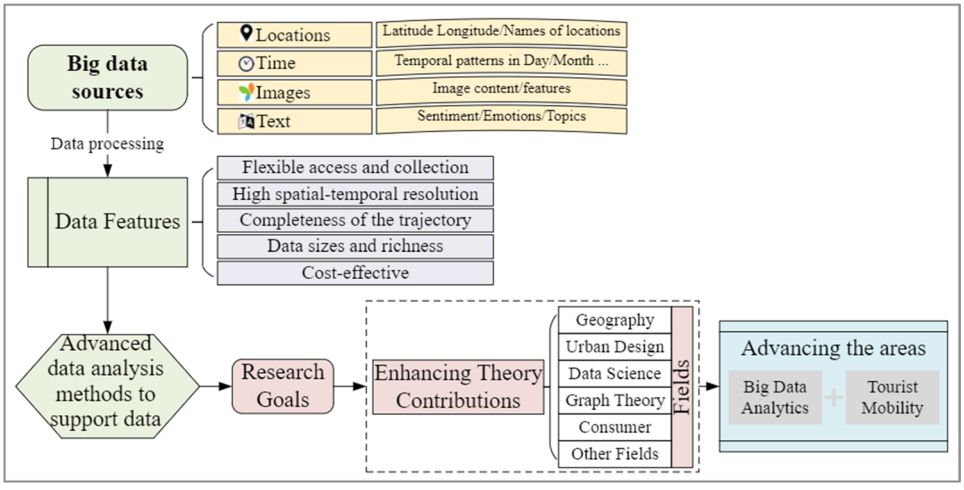

Figure 5. Research process in big data analytics and tourist mobility studies.

We found that future contributions can be enhanced in two directions: Data with advanced data analysis methods and research goals for developing theory contributions. The challenges for studies in tracking tourist mobility using big data analytics are mainly about the methodology to handle data to extract the information and also serving the research goals with the supporting theory contributions. We conclude the discussion by suggesting future directions in studying tourist mobility and big data analytics.

5.1. How to enhance theory contributions

The emergence of big data analytics in tourism and hospitality research has sparked significant interest due to its potential to provide deeper insights into tourists' behaviour. While big data analytics offers opportunities to explore tourist mobility at multiple levels, it has been noted in the literature that there is a lack of theoretical contributions in this field. Studies often focus on descriptive analysis, overlooking the importance of theoretical frameworks that could guide the interpretation of data and uncover underlying explanations and implications, which restrict the depth and breadth of research in understanding tourist mobility.

While big data analytics is considered to be a data-driven approach it does not necessarily mean it is atheoretical. As Mazanec (2020) points out that there is a hidden layer of theorizing that often goes unrecognised in big data analytics. He argues that despite the surface appearance of being atheoretical, big data research in tourism actually involves a significant amount of concealed theorizing. This hidden theorizing shapes the research process, from the selection of data sources to the interpretation of patterns and the derivation of insights. For example, when researchers choose a particular machine learning model or statistical technique, they are implicitly drawing on underlying theories related to probability, decision-making, and even behaviour. The atheoretical critique posited by Mazanec is particularly relevant in the context of tourist mobility, as it raises questions about the validity and applicability of findings

derived from big data analytics. For example, Figure 4 shows that many people concentrate on the theory of the pattern recognition, using the theory that is hidden in the analytics method to explain tourist mobility and movement patterns. Therefore theory contributions could also seek directions from data science theories, such as *Sequential Rule Mining* and *Markov Chain Model*, which could be applied to analyse tourist trajectories and predict future tourist movements. Similarly, graph theory could be used to construct a network of tourist destinations, identifying key nodes (popular destinations) and edges (common travel routes). This could help identify crucial locations or patterns of the itinerary, informing tourism network analysis. Theories from customer behaviour could also be applied to understand tourist mobility. For example, the concept of 'the path of least resistance' could be used to understand tourist accommodation choices and mobility patterns. Furthermore, the Theory of Planned Behaviour could provide insights into how perceived risks influence travel patterns during a pandemic. Integrating theoretical contributions into the study of tourist mobility using big data analytics is critical for several reasons. Firstly, it provides a solid foundation for analysing the data by contextualising the patterns and trends within established theories. Theoretical frameworks offer insights into the motivations, preferences, and behaviours of tourists, enabling a more informed interpretation of the data. Secondly, theoretical contributions enhance the practical implications of the research. Figure 5 shows some current directions to enhance the theoretical contributions to help researchers move beyond descriptive analysis. Regarding the theoretical frameworks, the volume, variety, and velocity of big data pose challenges to existing theoretical frameworks. Traditional theories mostly are not able to fully capture the nuances and patterns in big data. Therefore, new theoretical frameworks are required that are specifically designed for big data analytics and are flexible enough to accommodate the vast amount of data and the diverse ways in which data can be analysed.

In future studies, researchers could consider the data characters from different sources for the study goals. Also, the research goals are supported by the research fields for enhancing the theory contributions. The integration of time geography, core-periphery theory, and density theory could lead to the development of a comprehensive model to predict tourist mobility patterns. This model could provide a more nuanced understanding of how spatial and temporal constraints influence tourist activities and choices. Furthermore, the evolution of public transportation and regional integration could significantly impact the core-periphery structure in tourism, which presents an opportunity to re-evaluate and update the core-periphery theory in the context of modern tourism. Also, advances in artificial intelligence (AI) and Generative AI (Hsu et al., 2024) when utilized in big data analytics can lead to new insights and discoveries that were not previously possible with traditional theoretical frameworks. For example, they can identify patterns and correlations in data that are beyond human perception and expectations. These insights can then be used to develop new theories or to refine existing ones.

5.2. Improving the big data quality in assessing tourist mobility

Despite the advantages of relying on big data to track tourist mobility, existing studies raised a variety of limitations. These limitations are mostly associated with the type

of big data used and the expressiveness of such data. When relying on social media as a data source, concerns were raised about the lack of explicit geographical coordinates and, therefore, that the social media platforms may not accurately reflect the exact location of tourists. In addition, it was mentioned that social media platforms may not be representative of the entire population of tourists, which may limit the ability to provide accurate insights into tourist behaviour.

Furthermore, other limitations are mentioned, such as that the study only covers one destination or does not consider other factors such as transportation, weather conditions, culture, and other cultural factors that may affect tourism behaviour. These types of limitations are related to specific studies and can be overcome by extending locations and including additional data sources required to address specific concerns. For example, to consider transportation factors, it is sufficient to incorporate transportation data in the analysis.

One of the challenges in analysing big data in tourism is the difficulty in distinguishing data from tourists who travel individually versus those who travel as part of a group. Group tourists often follow a predetermined itinerary, which restricts their individual travel choices and can skew the overall analysis if a significant portion of the data is derived from group tourists. The potential solution to this issue would be to acquire detailed knowledge about the location of interest and the typical ways it is visited. This is because, in many cases, the travel patterns of group tourists and individual tourists can vary significantly based on the location and the nature of the visit. Therefore, understanding these nuances can help in more accurately interpreting the data and drawing meaningful conclusions.

Additionally, collected Big data may not be representative of all users, causing the inability to accurately detect different types of tourism activities. The time and cost of collecting and analysing large amounts of data can be a significant barrier for researchers. However, the benefits of using big data analytics to track tourist mobility are significant. By utilising advanced technologies such as machine learning, artificial intelligence, and natural language processing, researchers can gain insights into tourist movements in real-time and make more accurate predictions. Integrating big data analytics with other research methods and theories can help bridge the gap between big data research and other research methods and provide a more comprehensive understanding of tourist mobility. We summarised the points mentioned above and give the recommendations to minimize the limitations as well as highlight challenges:

- **Selection:** The first and most important step is to identify and select relevant data sources. This can be a difficult task because many factors could influence the quality of the data. Finding appropriate data sources, data formats, and relevant features to support the analysis can be challenging, particularly if the data sources are not widely available or if the data formats are not well-defined. As shown in the literature, tracking tourist mobility needs geo-location data (e.g. latitude, longitude and/or name of locations), timeline and other dimensions data (such as text, image, and emotion sensors).
- **Collection:** The availability and quality of data sources are important considerations in tracking tourist mobility. Obtaining high-quality data from diverse sources can be challenging and, in many cases, requires specialised

programming skills, especially when dealing with complex and unstructured data such as social media posts or GPS data.

- **Cleaning:** Cleaning and transforming the data into a format that can be used for analysis also requires specialised skills and expertise. Ensuring that the data is accurate, complete, and up-to-date can be a significant challenge. If the data obtained from different sources or over time is not consistent can lead to inconsistencies in the findings. Therefore, significant attention needs to be directed to cleaning and transformation.
- **Representation:** The results may be influenced by factors such as language barriers, cultural differences, and time of the year. Some concerns are also expressed that social media may not be representative of the entire population; however, this concern can be overcome by relying on diverse social media platforms or combining big data (social media + experiment) (Adamiak & Szyda, 2022; McKercher et al., 2015; Owuor et al., 2023). Alternatively, researchers can also use multiple social media platforms, which attract different demographics of users and ethnicities. For example, the vast majority of *Twitter* (now '*X*') users are 25–34 years old (38.5%), followed by the 35–49 range (20.7%) even 50+ years old users are well represented (17.1%) (Statista, 2021b). TikTok's audience is on the younger side, with 71% of users being between the ages of 18 and 34 years (Statista, 2023a). Similarly, users in their twenties and thirties accounted for almost two-thirds of *Reddit* active users (Statista, 2021a). However, it appears that *YouTube* has a wide demography, with active users by age all across the board. The age range is also dependent, and it is aligned with the covered topic (Statista, 2023b).

5.3. Benefits, challenges, and contradictions of using technology to study mobility

The analysis identified both benefits and challenges of using technology to study mobility, which is not a surprise considering that the technology indeed enables greater precision in tracking and analysing tourist movements. However, it's crucial to also consider the ethical implications of increased monitoring and potential surveillance. Researchers should ensure transparency about data collection methods and to obtain informed consent from participants whenever possible. Alternatively, they should implement privacy-preserving techniques to anonymize results. Furthermore, it's essential to explore the potential of using technology to enhance tourist experiences rather than solely for monitoring purposes. For instance, technology can be used to provide personalised recommendations, improve accessibility, and enhance safety.

We also identified contradictions in the assessed literature. Unfortunately, despite the benefits of using Big data analysis, some potential contradictions or challenges could arise in implementing these directions. Some of these contradictions can be seen in the following areas:

- **Data sources and analysis methods:** While the use of diverse data sources and advanced technologies like machine learning and artificial intelligence can

provide valuable insights, it may also raise challenges related to data privacy, security, and ethical considerations. For instance, how to ensure the confidentiality of individual tourist data while still enabling meaningful analysis? The question that we need to ask is, how to ensure that the use of these technologies does not violate privacy laws or regulations.

- **Balancing precision and surveillance:** On the one hand, the use of big data analytics by harnessing some data sources such as mobile phones can provide greater precision in understanding tourist mobility patterns and trends. However, this increased precision may also lead to greater monitoring and tracking of tourist activities, potentially raising concerns related to surveillance and privacy. How to strike the balance between these two aspects, ensuring that the benefits of data analytics are maximized while minimizing the risks and negative consequences needs to be taken into consideration.
- **Theoretical vs. descriptive analysis:** While theoretical contributions are crucial for advancing knowledge in the field, as mentioned earlier, it has been noted a lack of theoretical contributions in the study of tourist mobility using big data analytics. On the other hand, studies often focus on descriptive analysis, which may not provide a deep understanding of the underlying explanations and implications. How to balance these two approaches and ensure that theoretical contributions are given adequate attention and priority is an open question.
- **Addressing limitations and biases:** While big data analytics offers opportunities to explore tourist mobility at multiple levels, it is important to acknowledge the limitations and potential biases in the data. For instance, data may be skewed towards certain demographics or locations, or may not capture the full complexity of tourist behaviour and experiences. How to address these limitations and biases and ensure that the insights gained from Big data analytics are representative and unbiased is another challenge.

5.4. Future directions in studying tourist mobility and big data analytics

As the field of tourist mobility continues to evolve, so does the need for innovative research directions that can provide valuable insights into tourist behaviour and destination management. One such approach that holds great promise is using big data analytics to study tourist mobility. The analysis of large-scale datasets, encompassing various sources such as social media and mobile phone data, offers a rich opportunity to understand tourist movements, preferences, and behaviours.

Building upon the existing literature, this paper aims to identify future directions and research goals in studying tourist mobility and big data analytics. By exploring these avenues, researchers can expand their understanding of tourist behaviour, inform destination management strategies, and contribute to the development of the tourism industry. Table 3 shows the potential research directions and specific actions that help to enhance the study of tourist mobility through big data analytics into four major directions with adapting new technologies; Analysing the factors that impact tourist mobility; personalization and customization, and extending to other focuses, such as Sustainability. For example, by analysing data on a

tourist's past travel destinations, activities, and reviews, a travel company could provide personalised recommendations for future trips. Personalization and customization are key aspects of big data applications in tourist mobility. As the tourism industry continues to evolve, there is an increasing demand for personalised experiences and services that cater to the unique needs and preferences of each tourist. Big data technologies play a crucial role in enabling this level of personalization and customization.

6. Conclusions

This review aims to provide a comprehensive literature review focusing on the diverse types of big data employed in tourist mobility studies, which, to the best of our knowledge, could be the first attempt to present a detailed study about tourist mobility and big data. Recognizing that different types of big data contain distinct information addresses specific tourist mobility-related issues, which require unique analytical data types and techniques. Understanding the dispersal of visitors and the connections among destinations is crucial for effective destination management. We systematically analysed each data type based on four primary perspectives: *Data characteristics, Research focus, Theory contributions,* and *Challenges*.

The review identified the gap without theoretical contributions to analysing visitors' travel patterns. Big data analytics is considered a mostly data-driven approach in fields such as the empirical sciences, and methodological contributions are as valuable as theoretical contributions (Mazanec, 2020). In line with this observation, most identified papers were data-driven to model tourist spatial and temporal behaviour. The theory contributions focus on five fields, led by Geography fields and Data sciences. Future work can focus on analysing the complex structure of tourist attraction communities to better understand how to direct tourists toward other areas rather than only marketing primary attractions. This will allow for a more in-depth analysis of the tourism industry and its impact on local communities, including using more advanced technologies such as machine learning and artificial intelligence to analyse large datasets and identify patterns and insights that may not be immediately apparent through traditional survey methods.

Big data certainly allows us to see the expand of the study scale, both on the scale of the location (such as local, national, and even worldwide travel flow) and offers unprecedented detail and scope in understanding various phenomena. Meanwhile, the granularity of big data enables researchers to examine specific travel patterns and behaviours, leading to more targeted and personalised destination strategies. However, despite the advances and the potential that big data holds, it appears that we have indeed not fully tapped into its capabilities to address the 'big' problems facing the tourism industry, such as understanding mobility in the context of climate change and promoting sustainable practices. There is a clear gap in the literature regarding a comprehensive review that investigates tourism mobility and its intersection with big data analytics.

Also, researchers could explore the impact of social media on tourist mobility, such as analysing the types of content shared and the impact of social media advertising on travel behaviour. Future studies could also investigate the relationship between

Table 3. Future directions and specific directions.

Future directions		Specific Topics
Adapting new emerging technologies	1.	Explore the use of advances in Artificial Intelligence, such as Large Language models and Generative AI, to process and extract information from a large amount of spatio-temporal data and high-dimensional data.
	2.	The use of augmented reality and virtual reality technologies can enhance the accuracy and efficiency of the tracking process. Moreover, integrating machine learning algorithms can help improve the accuracy and reliability of the predictions, thereby enhancing the effectiveness of joint marketing efforts.
	3.	To investigate the impact of the digital age on tourism and how it affects how tourists engage with each other and the physical environment, transforming how tourists interact with the physical world and how it is changing the way we think about tourism and its impact on communities.
Analyse the factors that impact tourist mobility	1.	Combine spatial and temporal analysis with content analysis to explore the reason or factors that shape particular travel patterns.
	2.	Explore the linkage between tourist flow and the structure of transport networks.
	3.	Analyse tourist mobility patterns in different seasons and locations.
	4.	Explore the influence of transport on tourist movement and how tourists make decisions.
	5.	Consider the social demographics of tourists in clustering and frequent pattern mining studies. This may help researchers better understand the underlying factors that influence the behaviour of tourists.
Personalization and customization	1.	To personalise and customise tourist routines by analysing individual preferences and behaviours to meet tourists' specific needs and interests, enhancing overall satisfaction and engagement.
	2.	Real-time experience based on their current location, the time of day, and their past activities, which could include recommendations for nearby attractions, restaurants, or events that the tourist might be interested in.
Focus on Sustainability	1.	Explore how tourist mobility and big data analytics can contribute to more sustainable tourism practices. Such involves studying how tourists' movement patterns to minimize environmental impact.

tourist mobility and environmental factors such as climate change or air pollution that could impact tourist mobility. Additionally, researchers could explore the feasibility of using machine learning algorithms to predict tourist behaviour based on large datasets and identify areas for further research and development. Furthermore, consider using advanced tracking technologies such as satellite imagery or augmented reality to enhance the accuracy and reliability of GPS data.

In conclusion, integrating big data analytics in studying tourist mobility presents a promising avenue for research, offering valuable insights into travel behaviour and patterns. Future research should aim to bridge this gap by explicitly articulating how big data analytics can be guided by theoretical frameworks. Furthermore, the advent of new technologies, such as GenrativeAI, augmented reality, and machine learning, provide opportunities for more sophisticated data analysis, potentially transforming our understanding of tourist mobility. Despite the challenges, the potential benefits of big data in enhancing tourist mobility are immense, from informing the creation of linkages between destinations to aiding in the development of sustainable tourism strategies. As we move forward, researchers and practitioners must harness the potential of big data and adapt to the evolving technological landscape.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Jinyan Chen is a Research Assistant Professor in the School of Hotel and Tourism Management at the Hong Kong Polytechnic University. She is also an adjunct fellow with the Griffith Institute for Tourism, Australia. Her main research interests include understanding tourist behaviour, modelling tourist travel patterns, sentiment analysis, network analysis, and relying on social media and big data analytics.

Noam Shoval is the Professor in the Department of Geography and the Institute for Urban and Regional Studies at the Hebrew University of Jerusalem. He is also Director of the European Forum and the Director of the Center for Urban Innovation at the Hebrew University of Jerusalem. His main research interests are urban geography and planning, urban tourism and the implementation of advanced tracking technologies in various areas of spatial research such as tourism and urban studies and medicine.

Bela Stantic is Professor in Computer Science and founder and Director of 'Big Data and Smart Analytics' Lab at Griffith University. Professor Stantic is internationally recognised in field of data analytics and efficient management of complex data structures, such as found in big data. He was invited to give many Keynotes and invited talks at highly ranked International conferences and prestigious institutions as well as he has been doing the journal editorial duties and reviews of many top ranked journals.

References

Adamiak, C., & Szyda, B. (2022). Combining conventional statistics and big data to map global tourism destinations before COVID-19. *Journal of Travel Research*, *61*(8), 1848–1871. https://doi.org/10.1177/00472875211051418

Asero, V., Gozzo, S., & Tomaselli, V. (2016). Building tourism networks through tourist mobility. *Journal of Travel Research*, *55*(6), 751–763. https://doi.org/10.1177/0047287515569777

Baggio, R., & Scaglione, M. (2018). Strategic visitor flows and destination management organization. *Information Technology & Tourism*, *18*(1-4), 29–42. https://doi.org/10.1007/s40558-017-0096-1

Barros, C., Moya-Gómez, B., & Gutiérrez, J. (2020). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. *Current Issues in Tourism*, *23*(10), 1291–1310. https://doi.org/10.1080/13683500.2019.1619674

Becken, S., Stantic, B., Chen, J., Alaei, A. R., & Connolly, R. M. (2017). Monitoring the environment and human sentiment on the Great Barrier Reef: Assessing the potential of collective sensing. *Journal of Environmental Management*, *203*(Pt 1), 87–97. https://doi.org/10.1016/j.jenvman.2017.07.007

Bigné, E., Oltra, E., & Andreu, L. (2019). Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities. *Tourism Management*, *71*, 490–503. https://doi.org/10.1016/j.tourman.2018.10.013

Birenboim, A., Anton-Clavé, S., Russo, A. P., & Shoval, N. (2013). Temporal activity patterns of theme park visitors. *Tourism Geographies*, *15*(4), 601–619. https://doi.org/10.1080/14616688.2012.762540

Bokelmann, B., & Lessmann, S. (2019). Spurious patterns in Google Trends Data – An analysis of the effects on tourism demand forecasting in Germany. *Tourism Management*, *75*, 1–12. https://doi.org/10.1016/j.tourman.2019.04.015

Buning, R. J., & Lulla, V. (2020). Visitor bikeshare usage: Tracking visitor spatiotemporal behavior using big data. *Journal of Sustainable Tourism*, *29*(4), 711–731. https://doi.org/10.1080/09669582.2020.1825456

Chen, J., Becken, S., & Stantic, B. (2021a). Using Weibo to track global mobility of Chinese visitors. *Annals of Tourism Research*, *89*, 103078. https://doi.org/10.1016/j.annals.2020.103078

Chen, J., Becken, S., & Stantic, B. (2021b). Harnessing social media to understand tourist mobility: The role of information technology and big data. *Tourism Review*, *77*(4), 1219–1233. https://doi.org/10.1108/TR-02-2021-0090

Chen, J., Becken, S., & Stantic, B. (2022a). Assessing destination satisfaction by social media: An innovative approach using importance-performance analysis. *Annals of Tourism Research*, *93*, 103371. https://doi.org/10.1016/j.annals.2022.103371

Chen, J., Becken, S., & Stantic, B. (2022b). Harnessing social media to understand tourist travel patterns in muti-destinations. *Annals of Tourism Research Empirical Insights*, *3*(2), 100079. https://doi.org/10.1016/j.annale.2022.100079

Chen, Y., Liu, Y., Wu, L., & Li, X. (. (2023). How does mobile social media sharing benefit travel experiences? *Journal of Travel Research*, *62*(4), 841–858. https://doi.org/10.1177/00472875221098936

Cheng, M. (2024). Social media and tourism geographies: Mapping future research agenda. *Tourism Geographies*, 1–10. https://doi.org/10.1080/14616688.2024.2304782

Cheng, M., Jin, X., Wang, Y., Wang, X., & Chen, J. (2023). a sequential pattern mining approach to tourist movement: The case of a Mega Event. *Journal of Travel Research*, *62*(6), 1237–1256. https://doi.org/10.1177/00472875221126433

Chu, C. P., & Chou, Y. H. (2021). Using cellular data to analyze the tourists' trajectories for tourism destination attributes: A case study in Hualien, Taiwan. *Journal of Transport Geography*, *96*, 103178. https://doi.org/10.1016/j.jtrangeo.2021.103178

Du, X., Pei, Y., Duivesteijn, W., & Pechenizkiy, M. (2020). Exceptional spatio-temporal behavior mining through Bayesian non-parametric modeling. *Data Mining and Knowledge Discovery*, *34*(5), 1267–1290. https://doi.org/10.1007/s10618-020-00674-z

Encalada-Abarca, L., Ferreira, C. C., & Rocha, J. (2022). Measuring tourism intensification in urban destinations: An approach based on fractal analysis. *Journal of Travel Research*, *61*(2), 394–413. https://doi.org/10.1177/0047287520987627

Encalada-Abarca, L., Ferreira, C. C., & Rocha, J. (2023). Revisiting city tourism in the longer run: An exploratory analysis based on LBSN data. *Current Issues in Tourism*, *27*(4), 584–599. https://doi.org/10.1080/13683500.2023.2182669

Falk, M. T., & Hagsten, E. (2021). Visitor flows to world heritage sites in the era of Instagram. *Journal of Sustainable Tourism*, *29*(10), 1547–1564. https://doi.org/10.1080/09669582.2020.1858305

Fennell, D. A. (1996). A tourist space-time budget in the Shetland Islands. *Annals of Tourism Research*, *23*(4), 811–829. https://doi.org/10.1016/0160-7383(96)00008-4

Frenzel, F., Giddy, J., & Frisch, T. (2022). Digital technology, tourism and geographies of inequality. *Tourism Geographies*, *24*(6-7), 923–933. https://doi.org/10.1080/14616688.2022.2142843

Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2019). Using social media to identify tourism attractiveness in six Italian cities. *Tourism Management*, *72*, 306–312. https://doi.org/10.1016/j.tourman.2018.12.007

Giglio, S., Bertacchini, F., Bilotta, E., & Pantano, P. (2020). Machine learning and points of interest: Typical tourist Italian cities. *Current Issues in Tourism*, *23*(13), 1646–1658. https://doi.org/10.1080/13683500.2019.1637827

Gomezelj, D. G. (2016). A systematic review of research on innovation in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, *28*(3), 516–558. https://doi.org/10.1108/IJCHM-10-2014-0510

Grinberger, A. Y., & Shoval, N. (2019). Spatiotemporal contingencies in tourists' intradiurnal mobility patterns. *Journal of Travel Research*, *58*(3), 512–530. https://doi.org/10.1177/0047287518757372

Grinberger, A. Y., Shoval, N., & McKercher, B. (2014). Typologies of tourists' time-space consumption: A new approach using GPS data and GIS tools. *Tourism Geographies*, *16*(1), 105–123. https://doi.org/10.1080/14616688.2013.869249

Hardy, A. (2020). Tracking via geotagged social media data. In *Tracking tourists movement and mobility*. Goodfellow Publishers Limited. https://doi.org/10.23912/9781911635383-4575

Hardy, A., & Aryal, J. (2020). Using innovations to understand tourist mobility in national parks. *Journal of Sustainable Tourism*, *28*(2), 263–283. https://doi.org/10.1080/09669582.2019.1670186

Hardy, A., Birenboim, A., & Wells, M. (2020). Using geoinformatics to assess tourist dispersal at the state level. *Annals of Tourism Research*, *82*, 102903. https://doi.org/10.1016/j.annals.2020.102903

Hardy, A., Hyslop, S., Booth, K., Robards, B., Aryal, J., Gretzel, U., & Eccleston, R. (2017). Tracking tourists' travel with smartphone-based GPS technology: A methodological discussion. *Information Technology & Tourism*, *17*(3), 255–274. https://doi.org/10.1007/s40558-017-0086-3

Hardy, A., Vorobjovas-Pinta, O., Wells, M., Grimmer, L., & Grimmer, M. (2022). Measuring cruise passenger dispersal through technology. *Annals of Tourism Research*, *93*(93), 103319. https://doi.org/10.1016/j.annals.2021.103319

Henderson, J. C. (2006). Destination development: Singapore and Dubai compared. *Journal of Travel & Tourism Marketing*, *20*(3-4), 33–45. https://doi.org/10.1300/J073v20n03

Höpken, W., Eberle, T., Fuchs, M., & Lexhagen, M. (2019). Google Trends data for analysing tourists' online search behaviour and improving demand forecasting: The case of Åre, Sweden. *Information Technology & Tourism*, *21*(1), 45–62. https://doi.org/10.1007/s40558-018-0129-4

Höpken, W., Müller, M., Fuchs, M., & Lexhagen, M. (2020). Flickr data for analysing tourists' spatial behaviour and movement patterns: A comparison of clustering techniques. *Journal of Hospitality and Tourism Technology*, *11*(1), 69–82. https://doi.org/10.1108/JHTT-08-2017-0059

Hsu, C. H. C., Tan, G., & Stantic, B. (2024). A fine-tuned tourism-specific generative AI concept. *Annals of Tourism Research*, *104*, 103723. https://doi.org/10.1016/j.annals.2023.103723

Jin, X., & Cheng, M. (2020). Communicating mega events on Twitter: Implications for destination marketing. *Journal of Travel & Tourism Marketing*, *37*(6), 739–755. https://doi.org/10.1080/10548408.2020.1812466

Kim, J. M., & Hyun, S. (2021). Differences in online reviews caused by distribution channels. *Tourism Management*, *83*, 104230. https://doi.org/10.1016/j.tourman.2020.104230

Kim, Y., Kim, C. K., Lee, D. K., Lee, H. W., & Andrada, R. I. T. (2019). Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management*, *72*, 249–256. https://doi.org/10.1016/j.tourman.2018.12.005

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. *International Journal of Contemporary Hospitality Management*, *33*(6), 2117–2136. https://doi.org/10.1108/IJCHM-06-2020-0587

Leung, D., Law, R., van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing*, *30*(1–2), 3–22. https://doi.org/10.1080/10548408.2013.750919

Leung, R., Vu, H. Q., & Rong, J. (2017). Understanding tourists' photo sharing and visit pattern at non-first tier attractions via geotagged photos. *Information Technology & Tourism*, *17*(1), 55–74. https://doi.org/10.1007/s40558-017-0078-3

Lewis, G. K., Hardy, A., Wells, M. P., & Kerslake, F. L. (2021). Using mobile technology to track wine tourists. *Annals of Tourism Research Empirical Insights*, *2*(2), 100022. https://doi.org/10.1016/j.annale.2021.100022

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009

Li, L., Tao, Z., & Lu, L. (2022). Understanding differences in rural tourism recovery: A critical study from the mobility perspective. *Current Issues in Tourism*, *26*(15), 2452–2466. https://doi.org/10.1080/13683500.2022.2088337

Li, Y., Li, Y., Li, J., Ma, S., & Gao, P. (2022). Tourism demand forecasting from the perspective of mobility: A brand-new predictive variable generated from intercity population mobility big data. *Asia Pacific Journal of Tourism Research*, *27*(5), 526–546. https://doi.org/10.1080/10941665.2022.2091941

Liu, P., Zhang, H., Zhang, J., Sun, Y., & Qiu, M. (2019). Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: From online to arrival. *Tourism Management*, *73*, 105–114. https://doi.org/10.1016/j.tourman.2019.01.021

Liu, W., Wang, B., Yang, Y., Mou, N., Zheng, Y., Zhang, L., & Yang, T. (2022). Cluster analysis of microscopic spatio-temporal patterns of tourists' movement behaviors in mountainous scenic areas using open GPS-trajectory data. *Tourism Management*, *93*, 104614. https://doi.org/10.1016/j.tourman.2022.104614

Liu, Z., Wang, A., Weber, K., Chan, E. H. W., & Shi, W. (2022). Categorisation of cultural tourism attractions by tourist preference using location-based social network data: The case of Central, Hong Kong. *Tourism Management*, *90*, 104488. https://doi.org/10.1016/j.tourman.2022.104488

Lu, J., Xiao, X., Xu, Z., Wang, C., Zhang, M., & Zhou, Y. (2021). The potential of virtual tourism in the recovery of tourism industry during the COVID-19 pandemic. *Current Issues in Tourism*, *25*(3), 441–457. https://doi.org/10.1080/13683500.2021.1959526

Ma, S., Kirilenko, A. P., & Stepchenkova, S. (2020). Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse. *Tourism Management*, *77*, 104021. https://doi.org/10.1016/j.tourman.2019.104021

Mariani, M., & Baggio, R. (2021). Big data and analytics in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management*, *34*(1), 231–278. https://doi.org/10.1108/IJCHM-03-2021-0301

Mashkov, R., & Shoval, N. (2023). Using high-resolution GPS data to create a tourism Intensity-Density Index. *Tourism Geographies*, *25*(6), 1657–1678. https://doi.org/10.1080/14616688.2023.2276910

Mazanec, J. A. (2020). Hidden theorizing in big data analytics: With a reference to tourism design research. *Annals of Tourism Research*, *83*, 102931. https://doi.org/10.1016/j.annals.2020.102931

McKercher, B. (2021). The periphery as a tourism market? *Tourism Recreation Research*, *49*(2), 302–311. https://doi.org/10.1080/02508281.2021.2011592

McKercher, B., Filep, S., & Moyle, B. (2021). Movement in tourism: Time to re-integrate the tourist? *Annals of Tourism Research*, *91*, 103199. https://doi.org/10.1016/j.annals.2021.103199

Mckercher, B., Shoval, N., & Ng, E. (2012). First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong. *Tourism Geographies*, 14(1):147–161. https://doi.org/10.1080/14616688.2011.598542

Mckercher, B., Shoval, N., Mckercher, B., Ng, E., & Birenboim, A. (2011). Hotel location and tourist activity in cities. *Annals of Tourism Research*, *38*(4), 1594–1612. https://doi.org/10.1016/j.annals.2011.02.007

McKercher, B., Shoval, N., Park, E., & Kahani, A. (2015). The [limited] impact of weather on tourist behavior in an urban destination. *Journal of Travel Research*, *54*(4), 442–455. https://doi.org/10.1177/0047287514522880

Meneghello, S. (2023). Mapping tourist landscapes in pandemic times: A dwelling-in-motion perspective. *Tourism Geographies*, *25*(7), 1730–1745. https://doi.org/10.1080/14616688.2023.2172604

Milano, C., González-Reverté, F., & Benet Mòdico, A. (2023). The social construction of touristification. Residents' perspectives on mobilities and moorings. *Tourism Geographies*, *25*(4), 1273–1291. https://doi.org/10.1080/14616688.2022.2150785

Mor, M., Dalyot, S., & Ram, Y. (2023). Who is a tourist? Classifying international urban tourists using machine learning. *Tourism Management*, *95*, 104689. https://doi.org/10.1016/j.tourman.2022.104689

Nguyen, T. T., Phan, T. C., Nguyen, Q. V. H., Aberer, K., & Stantic, B. (2019). Maximal fusion of facts on the web with credibility guarantee. *Information Fusion*, *48*, 55–66. https://doi.org/10.1016/j.inffus.2018.07.009

Ning, D., Yujie, Q., XiaoBin, C., & Jing, Q. (2023). Seeing is visiting: Discerning tourists' behavior from landmarks in ordinary photos. *Current Issues in Tourism*, *26*(15), 2494–2512. https://doi.org/10.1080/13683500.2022.2089547

Nolasco-Cirugeda, A., García-Mayor, C., Lupu, C., & Bernabeu-Bautista, A. (2022). Scoping out urban areas of tourist interest though geolocated social media data: Bucharest as a case

study. *Information Technology & Tourism*, *24*(3), 361–387. https://doi.org/10.1007/s40558-022-00235-8

Önder, I., & Marchiori, E. (2017). A comparison of pre-visit beliefs and projected visual images of destinations. *Tourism Management Perspectives*, *21*, 42–53. https://doi.org/10.1016/j.tmp.2016.11.003

Owuor, I., Hochmair, H. H., & Paulus, G. (2023). Use of social media data, online reviews and wikipedia page views to measure visitation patterns of outdoor attractions. *Journal of Outdoor Recreation and Tourism*, *44*, 100681. https://doi.org/10.1016/j.jort.2023.100681

Park, D., Kim, W. G., & Choi, S. (2019). Application of social media analytics in tourism crisis communication. *Current Issues in Tourism*, *22*(15), 1810–1824. https://doi.org/10.1080/13683500.2018.1504900

Park, J., & Song, H. (2021). Variance of destination region image according to multi-dimensional proximity: A case of the Greater Bay Area. *Journal of Destination Marketing & Management*, *20*, 100600. https://doi.org/10.1016/j.jdmm.2021.100600

Park, S., & Zhong, R. R. (2022). Pattern recognition of travel mobility in a city destination: Application of network motif analytics. *Journal of Travel Research*, *61*(5), 1201–1216. https://doi.org/10.1177/00472875211024739

Park, S., Xu, Y., Jiang, L., Chen, Z., & Huang, S. (2020). Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data. *Annals of Tourism Research*, *84*, 102973. https://doi.org/10.1016/j.annals.2020.102973

Park, S., Yuan, Y., & Choe, Y. (2021). Application of graph theory to mining the similarity of travel trajectories. *Tourism Management*, *87*, 104391. https://doi.org/10.1016/j.tourman.2021.104391

Park, S., Zu, J., Xu, Y., Zhang, F., Liu, Y., & Li, J. (2023). Analyzing travel mobility patterns in city destinations: Implications for destination design. *Tourism Management*, *96*, 104718. https://doi.org/10.1016/j.tourman.2022.104718

Paulino, I., Lozano, S., & Prats, L. (2021). Identifying tourism destinations from tourists' travel patterns. *Journal of Destination Marketing & Management*, *19*, 100508. https://doi.org/10.1016/j.jdmm.2020.100508

Payntar, N. D., Hsiao, W. L., Covey, R. A., & Grauman, K. (2021). Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru. *Tourism Management*, *82*, 104165. https://doi.org/10.1016/j.tourman.2020.104165

Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management*, *57*, 202–212. https://doi.org/10.1016/j.tourman.2016.06.006

Raun, J., Shoval, N., & Tiru, M. (2020). Gateways for intra-national tourism flows: Measured using two types of tracking technologies. *International Journal of Tourism Cities*, *6*(2), 261–278. https://doi.org/10.1108/IJTC-08-2019-0123

Sciortino, C., Ferrante, M., De Cantis, S., & Gyimóthy, S. (2022). Tracking cruise passengers' consumption: An analysis of the relationships between onshore mobility and expenditure. *Annals of Tourism Research Empirical Insights*, *3*(2), 100059. https://doi.org/10.1016/j.annale.2022.100059

Shi, C., Zhai, Y., & Li, D. (2023). Urban tourists' spatial distribution and subgroup identification in a metropolis – The examination applying mobile signaling data and latent profile analysis. *Information Technology & Tourism*, *25*(3), 453–476. https://doi.org/10.1007/s40558-023-00255-y

Shoval, N. (2006). The geography of hotels in cities: An empirical validation of a forgotten model. *Tourism Geographies*, *8*(1), 56–75. https://doi.org/10.1080/14616680500392499

Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies*, *18*(5), 587–606. https://doi.org/10.1080/14616688.2016.1214977

Shoval, N., Kahani, A., De Cantis, S., & Ferrante, M. (2020). Impact of incentives on tourist activity in space-time. *Annals of Tourism Research*, *80*, 102846. https://doi.org/10.1016/j.annals.2019.102846

Shoval, N., & Isaacson, M. (2009). *Tourist mobility and advanced tracking technologies*. Routledge.

Shoval, N., McKercher, B., Birenboim, A., & Ng, E. (2015). The application of a sequence alignment method to the creation of typologies of tourist activity in time and space. *Environment and Planning B: Planning and Design, 42*(1), 76–94.

Shoval, N., Schvimer, Y., & Tamir, M. (2018a). Real-time measurement of tourists' objective and subjective emotions in time and space. *Journal of Travel Research*, *57*(1), 3–16. https://doi.org/10.1177/0047287517691155

Shoval, N., Schvimer, Y., & Tamir, M. (2018b). Tracking technologies and urban analysis: Adding the emotional dimension. *Cities*, *72*(1240), 34–42. https://doi.org/10.1016/j.cities.2017.08.005

Smith, M., & Sarabi, Y. (2022). How does the behaviour of the core differ from the periphery? – An international trade network analysis. *Social Networks*, *70*, 1–15. https://doi.org/10.1016/j.socnet.2021.11.001

Spangenberg, T. (2014). Development of a mobile toolkit to support research on human mobility behavior using GPS trajectories. *Information Technology & Tourism*, *14*(4), 317–346. https://doi.org/10.1007/s40558-014-0005-9

Stantic, B., Pokorny, J., & Pokorný, J. (2014). Opportunities in big data management and processing. *Frontiers in Artificial Intelligence and Applications*, 15–26. https://doi.org/10.3233/978-1-61499-458-9-15

Statista. (2021a). *Distribution of Reddit app users in the United States as of March*. https://www.statista.com/statistics/1125159/reddit-us-app-users-age/

Statista. (2021b). *Distribution of Twitter users worldwide*. https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/

Statista. (2023a). *Distribution of TikTok users worldwide*. https://www.statista.com/statistics/1299771/tiktok-global-user-age-distribution/

Statista. (2023b). *Distribution of YouTube users worldwide*. https://www-statista-com.ezproxy.lb.polyu.edu.hk/statistics/1287137/youtube-global-users-age-gender-distribution/

Türk, U., Östh, J., Kourtit, K., & Nijkamp, P. (2021). The path of least resistance explaining tourist mobility patterns in destination areas using Airbnb data. *Journal of Transport Geography*, *94*, 103130. https://doi.org/10.1016/j.jtrangeo.2021.103130

Van der Zee, E., & Bertocchi, D. (2018). Finding patterns in urban tourist behaviour: A social network analysis approach based on TripAdvisor reviews. *Information Technology & Tourism*, *20*(1-4), 153–180. https://doi.org/10.1007/s40558-018-0128-5

Vu, H. Q., Li, G., & Law, R. (2020). Discovering highly profitable travel patterns by high-utility pattern mining. *Tourism Management*, *77*, 104008. https://doi.org/10.1016/j.tourman.2019.104008

Wang, J., Mo, H., Wang, F., & Jin, F. (2011). Exploring the network structure and nodal centrality of China's air transport network: A complex network approach. *Journal of Transport Geography*, *19*(4), 712–721. https://doi.org/10.1016/j.jtrangeo.2010.08.012

Wang, Z., Koroll, L., Höpken, W., & Fuchs, M. (2022). Analysis of instagram users' movement pattern by cluster analysis and association rule mining. In Information and communication technologies in tourism 2022: Proceedings of the ENTER 2022 eTourism Conference, January 11–14 (pp. 97–109). Springer International Publishing. https://doi.org/10.1007/978-3-030-94751-4_10

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Xu, T., Chen, R., Chen, W., Zheng, L., & Zhang, Y. (2022). Comparing the spatiotemporal behavior patterns of local, domestic and overseas tourists in Beijing based on multi-source social media big data. *Asia Pacific Journal of Tourism Research*, *27*(7), 692–711. https://doi.org/10.1080/10941665.2022.2119419

Xu, Y., Li, J., Belyi, A., & Park, S. (2021). Characterizing destination networks through mobility traces of international tourists—A case study using a nationwide mobile positioning dataset. *Tourism Management*, *82*, 104195. https://doi.org/10.1016/j.tourman.2020.104195

Xu, Y., Li, J., Xue, J., Park, S., & Li, Q. (2021). Tourism geography through the lens of time use: A computational framework using fine-grained mobile phone data. *Annals of the American Association of Geographers*, *111*(5), 1420–1444. https://doi.org/10.1080/24694452.2020.1812372

Xue, L., & Zhang, Y. (2020). The effect of distance on tourist behavior: A study based on social media data. *Annals of Tourism Research*, *82*, 102916. https://doi.org/10.1016/j.annals.2020.102916

Yu, L., Zhao, P., Tang, J., & Pang, L. (2023). Changes in tourist mobility after COVID-19 outbreaks. *Annals of Tourism Research*, *98*, 103522. https://doi.org/10.1016/j.annals.2022.103522

Zhang, K., Chen, Y., & Lin, Z. (2020). Mapping destination images and behavioral patterns from user-generated photos: A computer vision approach. *Asia Pacific Journal of Tourism Research*, *25*(11), 1199–1214. https://doi.org/10.1080/10941665.2020.1838586

Zhao, P., Ma, Z., Chen, J., Law, R., Zhang, Y., & Liu, Y. (2023). Depicting urban multi-scale tourist activity spaces using digital footprints for smart destinations. *Current Issues in Tourism*, *26*(18), 2965–2980. https://doi.org/10.1080/13683500.2022.2104696

Zhao, X., Lu, X., Liu, Y., Lin, J., & An, J. (2018). Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi'an, China. *Tourism Management*, *69*, 368–383. https://doi.org/10.1016/j.tourman.2018.06.026

Zheng, W., Huang, X., & Li, Y. (2017). Understanding the tourist mobility using GPS: Where is the next place? *Tourism Management*, *59*, 267–280. https://doi.org/10.1016/j.tourman.2016.08.009

Zheng, W., Li, M., Lin, Z., & Zhang, Y. (2022). Leveraging tourist trajectory data for effective destination planning and management: A new heuristic approach. *Tourism Management*, *89*, 104437. https://doi.org/10.1016/j.tourman.2021.104437

Zhou, J., Yang, S., Xiao, C., & Chen, F. (2021). Examination of community sentiment dynamics due to COVID-19 pandemic: A case study from a State in Australia. *SN Computer Science*, *2*(3), 1–11. https://doi.org/10.1007/s42979-021-00596-7

Zhou, X., & Chen, Z. (2023). Destination attraction clustering: Segmenting tourist movement patterns with geotagged information. *Tourism Geographies*, *25*(2-3), 797–819. https://doi.org/10.1080/14616688.2021.2006769