Improving the resident–tourist relationship in urban hotspots

Huy Quan Vu , Birgit Muskat , Gang Li & Rob Law

Published online: 07 Oct 2020.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Check for updates

Improving the resident–tourist relationship in urban hotspots

Huy Quan Vu[a] (iD), Birgit Muskat[b] (iD), Gang Li[c,d] (iD) and Rob Law[e] (iD)

[a]Department of Information Systems and Business Analytics, Deakin University, Melbourne, Australia; [b]ANU Research School of Management, Australian National University, Canberra, Australia; [c]School of Information Technology, Deakin University, Melbourne, Australia; [d]Chinese Academy of Sciences, Xinjiang Technical Institute of Physics and Chemistry, Ürümqi, China; [e]Hospitality and Tourism Research Centre (HTRC), School of Hotel & Tourism Management, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

ABSTRACT

High volumes of tourists often pose a threat to tourism and decrease the quality of life for local residents, particularly in attractive urban tourism places. Yet, to date only a few solution-oriented studies have attempted to alleviate the overtourism problems and to improve the resident-tourist relationship. This study aims to present potential solutions, based on data analytics. Combining venue-referenced social media data with topic modelling from a case study in Paris, this research reveals both similarities and differences in the temporal and spatial activity patterns of tourists and residents. Results offer strategic support to tourism planners on how to manage over-crowded urban tourism hotspots, which consequently facilitate the improvement of the resident–tourist relationship and improve destination attractiveness in the long run. Results further indicate that the exchange of social media-based information for residents and tourists are part of the practice-based solution for better sustainable tourism planning.

1. Introduction

Local residents' life quality has often been negatively impacted by uncontrolled high volumes of tourist streams (Cheung & Li, 2019; Lalicic, 2020; Pearce, 2019; Phi, 2020). The negative social impact that tourism brings to residents lies in the strong economic growth paradigm of tourism planners (Gibson, 2019) and their 'strategic failure' to address the precarious situation of some urban hotspots (Lalicic, 2020). For attractive urban tourist destinations, the spatial density concentrates around popular salient hotspots, leading to a tense resident–tourist relationship. Thus, despite the importance of the positive economic contributions of tourism to residents' income, the unprecedented growth prior to the global COVID-19 pandemic comes at a high cost for residents (World Tourism Organization (UNWTO), 2018).

We suggest that this present time of global economic crisis can be used to reflect on and implement sustainable tourism. Specifically, we posit that tourism planners now strategize how to better manage tourism growth and prepare for implementing the necessary strategic changes to make tourism in general, and urban tourism in particular, more sustainable and similarly, achieve higher destination attractiveness. To help tourism planners to improve their strategic sustainable tourism planning, we adopt a spatial-temporal perspective (Andereck et al., 2005;

McGuire et al., 2010; Namberger et al., 2019). Before the COVID-19 crisis, the vast majority of urban destinations were conflicted with immense competition between residents and tourists. Space and access to local resources has been the key issue that has caused the perception of overtourism (Mody et al., 2019). This problematic tourist-resident relationship is often exacerbated by residents' restricted opportunity to access and use the same infrastructure, increased waiting times, congested areas and transportation; these are aspects that negatively shape the residents' view on the value of tourism (Kim et al., 2020; Pasquinelli & Trunfio, 2020; World Tourism Organization (UNWTO), 2018; World Tourism Organization (UNWTO) & IPSOS, 2019).

This research contributes to the rather under-researched area of sustainable urban tourism (Ashworth & Page, 2011; Kim et al., 2020; Selby, 2012; Shoval, 2018). Prior research confirms that it is necessary to come up with approaches to tackling the manifold and complex issues in sustainable tourism (Font & McCabe, 2017; Maxim, 2016). Solutions are needed to mitigate the negative social impact of tourism (Deery et al., 2012). However, till now, solution-oriented studies still remain scarce (Joppe, 2018) and scholars decry that existing approaches are too abstract, instead of offering concrete and practice-based solutions. Even if solutions exist, they often merely focus on residents and exclude tourists, thereby limiting the capability to solve the problem in a holistic manner (Sharpley, 2014). We argue that finding concrete measures and transforming them into managerial practice, balancing the needs of residents and tourists alike, is an important prerequisite for effectively designing and co-creating urban tourism (Lin et al., 2017). However, to date, only a few practical contributions have come from academic research that offer specific recommendations on how to manage the spatial and temporal aspects of large tourist crowds (Joppe, 2018; Namberger et al., 2019).

This study contributes to knowledge development with a solution-oriented study to the overtourism problem by leveraging an innovative analytics-based approach. The method has potential to generate a positive impact to enhance the resident-tourist relationship caused by hotspot overtourism. Our method can determine both the similarities and differences in tourists' and residents' spatial behaviours at an attractive urban tourism destination, so that appropriate destination management strategies can be developed to address the challenges in overtourism. With a case study of Paris, one of the world's most popular destinations (Samuel, 2019), we demonstrate the effectiveness of our method, and address the challenges of hotspot overtourism from a spatial-temporal perspective. Our method utilises venue-referenced social media data (VR-SMD) (Vu et al., 2020) and a data analytics approach based on topic modelling (Blei et al., 2003) to extract common activity patterns for a comparative analysis of activity patterns. The current research is amongst the first solution-oriented studies in the context of overtourism, which aims to improve the resident–tourist relationship at popular urban destinations.

2. Literature review

2.1. The overtourism phenomenon

The overtourism phenomenon is closely linked to early debates in tourism geography on mass tourism (e.g. Hall, 1970), carrying capacity (e.g. O'Reilly, 1986) or crowding in urban destinations (Van der Borg et al., 1996). Since 2017, the term 'overtourism' has been used to describe the damage caused by the high volumes of tourists and the negative social impact leading to a significant deterioration in the quality of life of local residents (Cheung & Li, 2019; Lalicic, 2020; Novy & Colomb, 2019; Pearce, 2019; Phi, 2020).

The ongoing urgency to find solutions and manage the negative effects of overtourism comes with the increasing volumes of urban tourists (World Tourism Organization (UNWTO), 2018). Particularly, popular urban overtourism hotspots that have continuously been exposed to managing extreme levels of crowding at particular times have become a major problem for tourism, especially as high growth in international urban tourism has now led to a significant decline in

residents' livelihoods. Feeling disadvantaged by tourism, residents increasingly express their anti-tourism sentiments, articulate their stressful experiences and the discomfort and negative transformation that tourism brings to their home environment (Lalicic, 2020; Monterrubio, 2016; Pasquinelli & Trunfio, 2020).

2.2. Hotspot overtourism in urban tourist destinations

Hotspot overtourism is triggered at a particular place and time, where crowding occurs due to exceedingly high tourist numbers and is exacerbated by daily peak times and high season demands (Jacobsen et al., 2019). Jacobsen et al. (2019, p. 54) state that individuals who experience extreme levels of density in hotspots feel that their personal spaces are violated. Therefore, negative affective responses occur including 'worry, feelings of being unsafe in crowds, aversion to noise', leading to negative behaviours and avoidance.

Urban tourism is specifically susceptible to crowded hotspots as their spatial density defines cities. Managing the negative various effects that hotspot overtourism has in urban areas is complex and to date, contributions are rather limited. One related issue within the extant tourism literature appears to be the fact that urban tourism remains an under-researched area (Ashworth & Page, 2011; Selby, 2012; Shoval, 2018). This lack of interest is rather surprising considering its high importance arising from economic impact but also the potential negative social and environmental impact on popular urban destinations. Another issue lies in the absence of research that offers specific ideas to mitigate overtourism problems (Maxim, 2016). This void presumably exists due to the high complexity of generating impactful methodologies for sustainable tourism (Font & McCabe, 2017). Torres-Delgado and Palomeque (2014, p. 135), for example, state that 'unfortunately, most of the systems that have been proposed to date are too theoretical and therefore of limited use in practice'.

Specifically, poor data quality has been identified as one major problem hindering research from contributing concrete solutions to improve negative tourism impacts. Moreover, Torres-Delgado and Palomeque (2014) criticize that various data issues in both data collection and analysis create obstacles to improving sustainable tourism measures. These data-bound issues cause biases in both validity and reliability of results and subsequent interpretations, as often data are insufficient, unavailable, or have issues with levels of aggregation and importantly, are not up to date. In line with these overarching issues to strategically manage all dimensions of sustainable tourism, finding concrete solutions to manage the social dimension often remains unaddressed and measuring and implementing policies to reduce remains a difficult task (Deery et al., 2012).

2.3. The social dimension of hotspot overtourism

The negative social impact that hotspot overtourism creates is the negative resident-tourist relationship, as residents are forced to compete with tourists for space and access to local resources (Mody et al., 2019; Namberger et al., 2019). Studies have shown that lack of access to urban resources and long waiting times have a significant negative influence on perceived crowding (McGuire et al., 2010). This imbalanced relationship generates residents' negative perceptions towards tourism, such as anti-tourism sentiments and it is known that residents tend to exhibit negative attitudes towards tourism during peak seasons (Kim et al., 2020; Williams & Lawson, 2001).

Theoretically, this unhealthy competition over access to urban resources is often explained with power and social exchange theory (Andereck et al., 2005; Ap, 1992). Blau (2017) states that power theory can be used to analyse how an imbalance of power in social life leads to conflict. Solutions to restore perceptions of a balance of power amongst stakeholders can take place by effectively managing reciprocity and exchanging resources and benefits (Blau, 2017). Thus, a

balance of power helps by ensuring that the disadvantaged groups neither use force to regain their power – or that they feel that they remain 'left-behind' and suppressed (Blau, 2017).

Following social exchange theory, the reciprocity of the exchanged benefits from tourism needs to be balanced between residents and tourists (Ap, 1992; Andereck et al., 2005). In cases of overtourism, residents feel disadvantaged and perceive an imbalance occurring at their expense. Residents develop negative attitudes towards tourism and use forceful and vocal anti-tourism sentiments possibly because of the lack of adequate exchange. Despite residents' awareness and being in favour of tourism's economic gains (e.g. as a source of income), the limited or restricted availability, access and opportunity to use communal facilities highly influences views on the value of tourism (Andereck et al., 2005; Namberger et al., 2019; World Tourism Organization (UNWTO), 2018). Subsequently, managing power imbalances in the tourism sector has been identified as a key task for tourism planners (Beritelli & Laesser, 2011).

2.4. Data analytics based on VR-SMD

This research adopts a space and temporal perspective on tourism, 'which is co-determined by the information-based and knowledge-based attitudes of visitors' (Romão et al., 2018, p. 70). We posit that knowledge of the accessibility of spatial resources can address a part of the resource-competition problem in overtourism. We concur with recent studies (Joppe, 2018) that knowledge exchange over real-time spatial and temporal data on crowded hotspot overtourism data could help balance the resident-tourist relationship and enable governments to develop long-term evidence-based strategies to enhance tourists' experiences in less crowded spaces with fewer anti-tourism sentiments. In this regard, Laurini (2017) refers to a city's geographic knowledge, showing that knowledge created and shared through data analytics can aid in overcoming this knowledge gap.

The use of data analytics, especially with large-scale data sources, has been introduced in tourism research, such as bank transactions (Sobolevsky et al., 2015), mobile roaming (Raun et al., 2016) and tourist pass data (Scuderi & Nogare, 2018). However, their applications are limited owing to their restrictions to public access. Instead, studies consider alternative data sources such as travel reviews that are widely available on social media platforms (Zhang et al., 2017), Facebook messages (Yoo & Lee, 2015) and geotagged photos (Vu et al., 2016) and tweets (Chua et al., 2016). A major limitation of these data sources is the lack of contextual information on the activity of a tourist at their visited locations.

Recently a new type of social media data called VR-SMD (Vu et al., 2020) was introduced – defined as data on social media platforms that determine users' visits to specific venues. The major advantage of VR-SMD compared with other geo-referenced social media data (Chua et al., 2016; Vu et al., 2016) is its capability to associate users' locations with specific venues rather than raw GPS coordinates. A visit to a venue, such as a restaurant or shopping centre, directly reflects the corresponding activities, the dining or shopping of users, thereby making VR-SMD an ideal resource for analysing and comparing activities between residents and tourists in urban destinations.

3. Methodology

Our approach involves the extraction and analysis of VR-SMD, which are available on various platforms such as TripAdvisor (Zhang et al., 2017), Expedia (Stringam et al., 2010), Yelp (Xiang et al., 2017) and Foursquare (Luo et al., 2019). These platforms provide comprehensive lists of venues in various categories (e.g. restaurants, shops, malls and attractions), which are useful for inferring the associated activities of visitors. The data on these platforms are constantly being generated by users, thus they are timely, up to date and available at large volume, which can address the data quality issues as outlined by Torres-Delgado and Palomeque (2014). VR-SMD in

the form of venue-check-ins on a mobile social media platform, namely, Foursquare, is applied for our case study. This type of data was also proven effective in capturing tourist activities (Luo et al., 2019). Apart from the spatial and activity information embedded in venue check-ins, accurate information on local date and time is also available. Thus, this platform is convenient for temporal analysis and crowd management at hot tourism spots.

A challenge in studying the similarities/differences in the behaviours between residents and tourists is the complexity of the activities involved. Residents and tourists would certainly engage in essential activities, such as dining and travel, because these common activities are essential for daily life. Local residents may also like to visit a few tourism hotspots in urban areas, which are common venues, spaces or squares for public use. Hence, comparing the activity preferences between residents and tourists should account for their entire activity itineraries to capture the similarities/differences in their daily activity patterns, rather than individually assessing venue popularity similar to prior studies (Luo et al., 2019). As such, this research proposes an analytics approach based on topic modelling to the analysis of VR-SMD, which captures concurrent daily activities. The overall framework of our process consists of three major steps: (1) VR-SMD extraction, (2) activity topic modelling of activities and (3) comparative analysis. Due to the high complexity of the extracted data, our methodology adopts multiple steps of data retrieval and analysis. The statistical procedures are needed to reveal overall spatial and temporal patterns of residents and tourists as well as detailed insights about their behaviour at a specific time and at specific tourist venues. These steps will be described in detail in the subsequent sections.

3.1. VR-SMD extraction

The data extraction process is tailored for the specific data type of venue check-ins from Foursquare, which is the data source of this study. First, Foursquare users who posted check-ins at a specific tourism destination of interest are identified. The Foursquare platform does not contain functions to directly support this task. We carry out data collection via the Twitter Application Programming Interface (API), whose system has been integrated with Foursquare. Twitter API provides a streaming function that enables users to retrieve tweets being generated in real time using keywords or geographical locations. A bounding box, as specified by GPS coordinates (i.e. latitude and longitude), can be provided to extract tweets within a location of interest. Tweets generated by venue check-ins can be identified using the keyword 'swarmapp', which is the name of the mobile application that users employ for check-ins. Free access to the Twitter streaming API has a quota limit, in which only a small proportion of tweets are returned. Thus, the streaming function can be deployed for a long time (i.e. weeks or months), thereby providing sufficient data for analysis. Check-in tweets returned by the streaming function are random, and do not cover all possible check-ins made by specific users. Therefore, another data collection step is needed for the extraction of activity history.

The second step extracts all possible check-ins made by the identified users in the location of interest using the Search UserTimeline function. Three search criteria are used, including the user ID of the check-in tweets collected previously, the 'swarmapp' keyword and the GPS coordinate of the bounding box. Consequently, the entire check-in history of each user at the location of interest is extracted. Each venue check-in is accompanied by the following metadata: venue name, venue category, local check-in date and time, and location in the form of GPS coordinate. These metadata are important for sub-sequence analysis. Foursquare has a comprehensive list of venue categories (e.g. restaurants, shops and attractions) that are useful for inferring the corresponding activities. User profiles, including the location of origin, are also extracted, helping to indicate if users are local residents or tourists; i.e. if users state their location of origins as the same city or country as the location of interest, then they are considered local residents. Otherwise, they are likely to be tourists.

3.2. Topic modelling of activities

Topic modelling was originally proposed to discover the hidden semantic structures of topics in textual documents (Papadimitriou et al., 2000). If a document focuses on a certain topic, then words related to such a topic are assumed to appear frequently in the corresponding document. Various techniques for topic modelling have been proposed, such as latent semantic indexing (Papadimitriou et al., 2000), latent semantic analysis (Hofmann, 1999) and latent Dirichlet allocation (LDA) (Blei et al., 2003). Amongst these techniques, LDA is the most advanced technique with its powerful modelling capability and has been applied in various fields, such as political science (Greene & Cross, 2015), marketing science (Jacobs et al., 2016), hospitality (Lim & Lee, 2020) and tourism (Bi et al., 2019; Guo et al., 2017; Mazanec, 2017).

VR-SMD shares similar characteristics with ordinary text documentation, in which the activity history of each user and activities/venues can be regarded as document and words, respectively. However, LDA was mainly applied to textual data in previous studies to discover the discussed topics, opinions and perceptions of people. As such, its potential in modelling people's activities from VR-SMD has not been fully explored. Thus, LDA is adopted as an analytic technique for modelling activity topics reflected in the activity history records of residents and tourists.

Let $A = \{a_1, a_2, \ldots, a_N\}$ be a collection of N documents, where each document is a record of activities or venues visited by users. Given that the collected venue check-in data can span multiple days, each user may have multiple records of daily activities. Each document may have M words, which varies dependent on the number of check-ins made by the user. An activity record is only considered reliable for determining the daily activities of users if approximately m number of check-ins were made in one day. Otherwise, such a record should be excluded from the modelling process. LDA models each document a_i as a topic mixture vector $\Theta^i = \{\theta_1^i, \theta_2^i, \ldots, \theta_T^i\}$ over T topics, where each element θ_t^i specifies the likelihood of topic t being contained in document a_i. Each topic is characterised by a vector $\Phi^t = \{\phi_1^t, \phi_2^t, \ldots, \phi_V^t\}$, where each element specifies the likelihood of word v belonging to topic t. V denotes the number of words or number of venue categories in the VR-SMD dataset. A word may be associated with different topics with various degrees of memberships. LDA models the distributions of topics Θ in documents, and words in topics Φ follow sparse Dirichlet prior distributions with parameters α and β, respectively (Blei et al., 2003). The Dirichlet distribution assumes that documents cover only a few topics, whereas topics are represented only by a few words. Inputs into LDA for the model learning process is a *word $\times$ document* matrix, where each row corresponds to a daily activity record and each element pertains to the frequency of a word or number of check-ins at corresponding venue categories. After the learning process, word probabilities Φ, which represent topics of activity, and topic probabilities Θ, which represent the topic profile of activity records, are extracted for further analysis.

An issue with LDA is that users should specify the number of topics for model learning, which varies depending on the nature of each dataset. A common approach is to perform an experiment with various numbers of topics and evaluate the models' goodness-of-fit, as measured by the *perplexity* value (Blei et al., 2003). Generally, a low *perplexity* value indicates good fit. However, the model may overfit the data if numerous topics are provided for a dataset of small size. Analysis using the standard elbow method can be employed to identify a suitable topic number based on the experiment result (Ketchen & Shook, 1996).

3.3. Comparative analysis

Once the LDA model completes training on the dataset of activity records with a suitable number of topics, the succeeding step is to analyse the identified topics. The word probabilities Φ is used to examine the most popular activities in each topic. Word cloud, which is a commonly used visualisation technique for text analytics (Trattner et al., 2014), is employed to support the

interpretation of the topic meanings. Thereafter, the topic distributions Θ are computed with respect to the activity records of each user group (i.e. residents and tourists). The distribution of topics between user groups can be compared and validated using standard charting and statistical tests (i.e. t-test). Given the availability of time information, the activities can also be compared according to the temporal dimension, which adds values to the discovered insights. Analysis can be performed at multiple levels by venue categories or individual tourism hotspots.

4. Case study

4.1. Data collection

Our case study focuses on Paris, France, which is a popular urban tourism destination and represents an overtourism destination (Samuel, 2019). The tourism strategy for Paris has been evidently growth-oriented. In 2018, a record high of 24.5 million visitors arrived, of these, 16.5 million visited the inner urban areas (Office du Tourisme et des Congrès (OTCP), 2019). With 40,000 Airbnb accommodation sites, Paris has been offering more Airbnb rentals than other cities (Heo et al., 2019) and the French capital's popularity is associated with visits to popular landmarks (e.g. Tour Eiffel (30%), gastronomy (12%) and culture (e.g. Louvre museum (8%)). Given this growth, negative media reports and images of overtourism have also been increasing (Pasquinelli & Trunfio, 2020; The Guardian, 2019; Washington Post, 2019) and associations of Paris include unfriendly residents (12%) and being considerably busy (9%) (OTCP, 2019).

To collect data, a bounding box, the coordinates of which are specified as [2.023, 2.896, 48.597, 49.107] corresponding to the minimum latitude, maximum latitude, minimum longitude and maximum longitude, respectively, is entered into the streaming function of Twitter API (as described in Section 3.1). The selected bounding box is sufficient to cover the entire geographical area of Paris. The software continued to run for 12 months (i.e. from 2017 to 2018) to collect venue check-in tweets and identify the user IDs of Foursquare users. A second round of data collection was carried out using the Search UserTimeline function to collect users' complete activity history in Paris. User location is used to identify residents and tourists. Users without information on the location of origin are excluded from the dataset. The final dataset contains 52,782 venue check-ins generated by 3,715 users at 11,008 unique venues throughout Paris.

Table 1 presents the statistics of our dataset with respect to resident and tourist groups. Data collection identified fewer residents than tourists, probably since Foursquare is more popular amongst tourists from other countries than for local residents. Nevertheless, the collected dataset is sufficient for our case study owing to the numerous venue check-ins generated. It is noted that the number of venues for two groups do not add up to 11,008, as there are many common venues visited by both residents and tourists.

The majority of the check-ins were generated from 2014 to 2018 (Figure 1a) because the search can extract venue check-ins posted on user timelines before the streaming period in 2017. A peak in tourist numbers was observed for 2017, which is also the deployment period of the streaming function. Residents generally stayed in Paris, thereby resulting in a similar number of residents who made check-ins across years. The plot by month in Figure 2b is consistent with the fact that summer is the peak period for tourists visiting destinations in Europe, including Paris. Table 2 shows the number of tourists and check-ins by country of origin. Official statistics

Table 1. VR-SMD collection based on venue check-ins.

	No. of check-ins	No. of users		No. of venues
		Male	Female	
Residents	26,808	325	106	6,830
Tourists	25,974	1,974	1,206	6,153
Total	52,782		3,715	11,008

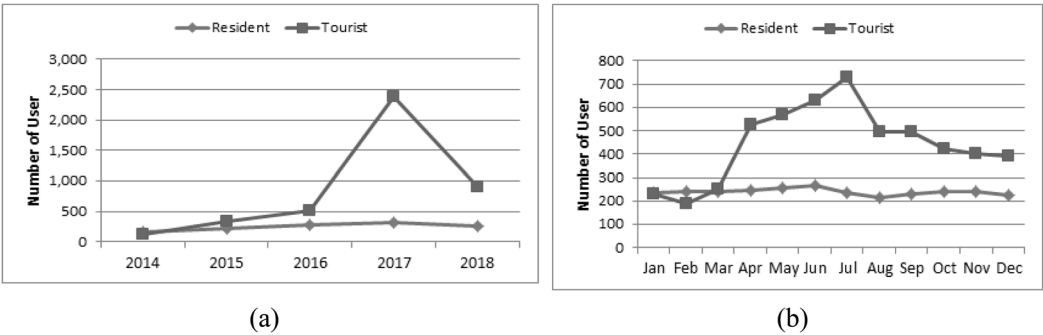(a)   (b)

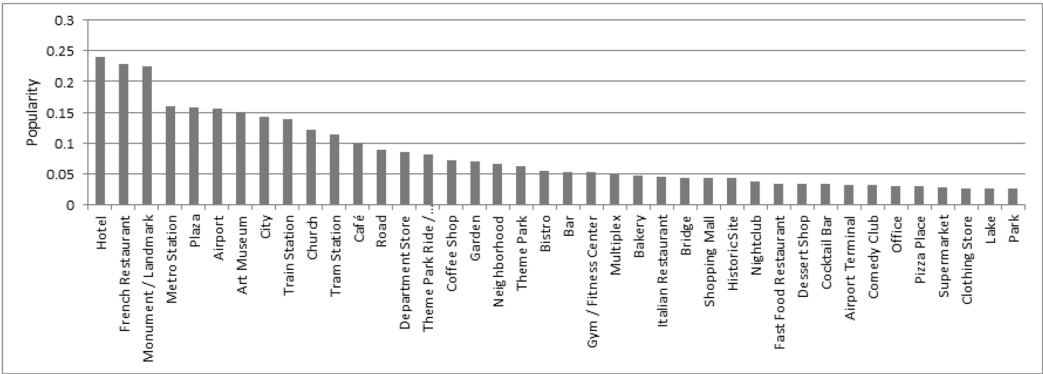**Figure 1.** Number of Foursquare users by year and month.

Figure 2. Leading venue categories.

indicate that many of these countries are amongst the top source markets of inbound tourism for France (DGE, 2016). China was not included amongst the top countries on our list, probably because Foursquare is not a popular check-in platform in this country. However, the collected dataset includes many other countries that could represent the overall activity pattern of tourists in Paris.

4.2. Activity modelling experiments

We modelled the activity patterns using the LDA technique. The activity records are converted into a *word × document* matrix format as outlined in Section 3.2. Each venue category is treated as a word in the vocabulary. Each daily active record is treated as a document. Only activity records with at least four check-ins are considered sufficient to capture user activity and eventually included in the dataset. As check-in records can span multiple days, one user may have more than one activity record. Activity records from 1,400 and 2,665 local residents and tourists were obtained. In topic modelling, words with considerably low frequency are excluded in the modelling process (Blei et al., 2003). Thus, venues that were visited less than five times were omitted. The final word vocabulary for the LDA modelling contained 276 venue categories. Figure 2 presents some of the most frequent venue categories.

Thereafter, an experiment is performed with the aim to determine the suitable number of topics for the final modelling of activities. For our experiment, a 10-fold cross validation approach is adopted, and the dataset is grouped into 10 subsets of equal size (Cawley & Talbot, 2010). Each subset is individually used as validation data, whereas the other subsets are used as training data. Figure 3 shows the average perplexity value of the LDA model on the test subsets

Table 2. Number of tourists by country of origin.

Country	No. of tourist	No. of check-ins
United States	618	5,183
Turkey	345	1,827
United Kingdom	243	2,552
Mexico	238	1,664
Belgium	210	1,774
Saudi Arabia	193	1,311
Kuwait	184	1,336
The Netherlands	109	832
Malaysia	105	790
Italy	98	839
Brazil	94	740
Germany	80	570
Japan	78	602
Spain	54	674
Thailand	49	458
Others (94 countries)	574	4,822
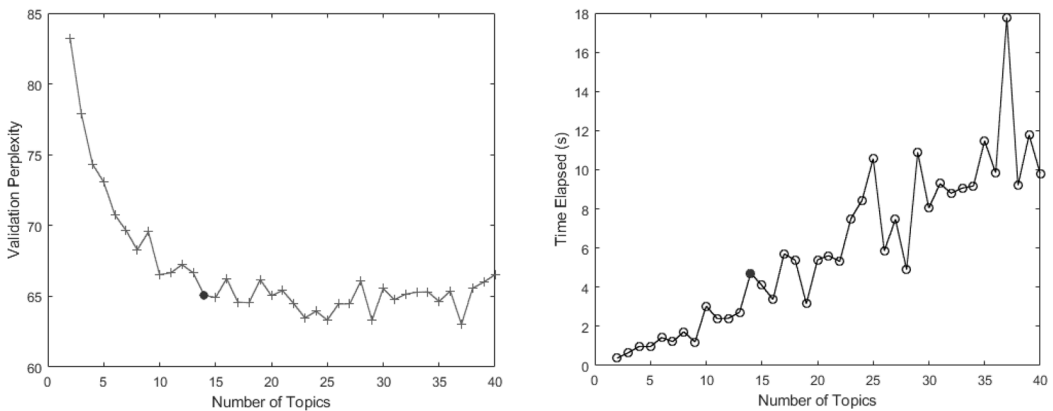
Figure 3. Model performance with varied number of topics.

with various numbers of topics. Generally, LDA tends to achieve a better fit for the data as more topics are used for the modelling. However, the performance does not substantially improve and tends to flutter when numerous topics are used, whereas additional time is needed to train the model (Figure 3b). The Elbow method (Ketchen & Shook, 1996) is applied and the final number of topics selected for the LDA model is 14.

We acknowledge that initialising the topic probabilities in LDA for training is a random process, which may produce slight differences in performance on the same dataset for the same number of topics across runs. Therefore, multiple LDA models are trained with the same number of topics (14) and the best model is selected for further analysis. Apart from perplexity, another evaluation metric is used, called topic concentration, to evaluate the quality of the identified topics. Figure 4 shows the performance of the 100 LDA models sorted according to topic concentration value. The model at the index position is selected as the final model, which has the best (lowest) value for topic concentration and is amongst the models with the lowest perplexity values.

4.3. Activity analysis of the topics

The probability distribution of the topics in documents and words in topics are extracted to examine the activity themes. The venue categories are visualised using the word cloud technique (see Figure 5). The topics are displayed in decreasing order of popularity in the collected
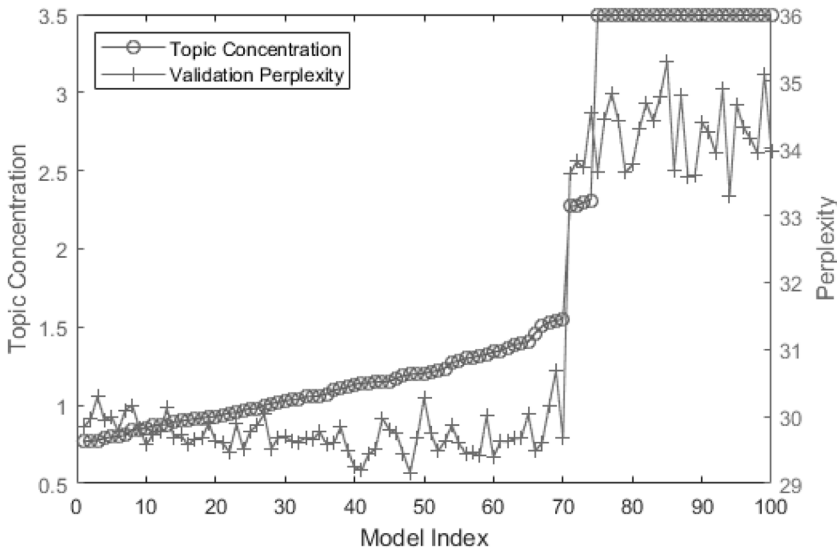
Figure 4. Performance of the different models with 14 topics.

datasets. The text size is determined by the probability value of the venue categories in the corresponding topics. For example, the most popular venue categories in Topic 1 include Monument/Landmark, Church and Art Museum. Meanwhile, Hotel, French Restaurant and Department Store are the most popular venue categories in Topic 2. Therefore, different venue categories could represent a similar type of activity. The venue categories are grouped into 10 super categories as defined by Foursquare for ease of activity interpretations (https://developer.foursquare.com/docs/resources/categories).

The probability distributions of activity types by topic are visualised using a heat map (see Figure 6). Some topics have dominant activity types, such as T2 (Food and Shop & Service), T5 (Food) and T7 (Art & Entertainment). Some topics have a combination of activity types, such as T1 (Art & Entertainment, Food, Outdoor & Recreation and Shop & Service) and T4 (Food, Nightlife Spot, Professional & Other Places and Travel & Transport). Note that some topics may have similar dominant activity types, such as Art & Entertainment in T7 and T12; however, they are treated as separate topics because the visited venues belong to different categories. For example, Art & Entertainment in T7 is about Multiplex, Toy/Game Store and Gym/Fitness Centre, whereas Art & Entertainment in T12 is about Theme parks. Thus, Figures 5 and 6 should be simultaneously examined when interpreting activity themes in the activity records. It is worth noting that the LDA models each topic as a distribution of word frequency. The same word can appear in different topics, but their probabilities are different among the topics. Besides, the themes of each topic are often derived by concurrent venue types. For example, venue types belonging to food categories (e.g. French Restaurant) would appear frequently in multiple topics as dining is one of the essential daily activities, but the other concurrent activities (e.g. shopping, entertainment, outdoor & recreation) are varied.

Besides, we would like to highlight that the meaning of 'topic' in this study is different from prior studies employing LDA (Guo et al., 2017; Lim & Lee, 2020; Luo et al., 2020). Prior works usually apply LDA to textual data (e.g. review, comments, posts), where a topic refers to a common theme being discussed. Thus, topics were labelled according to the identified themes for the ease of interpretation. In this paper, we applied LDA to a new type of data (VR-SMD), where each topic represents a common type of daily itinerary with a mixture of activities belonging to multiple themes. Some topics have similar dominant activity types but with different proportions. The use of a single label (name) is not sufficient to express topic meaning. As such, we use the

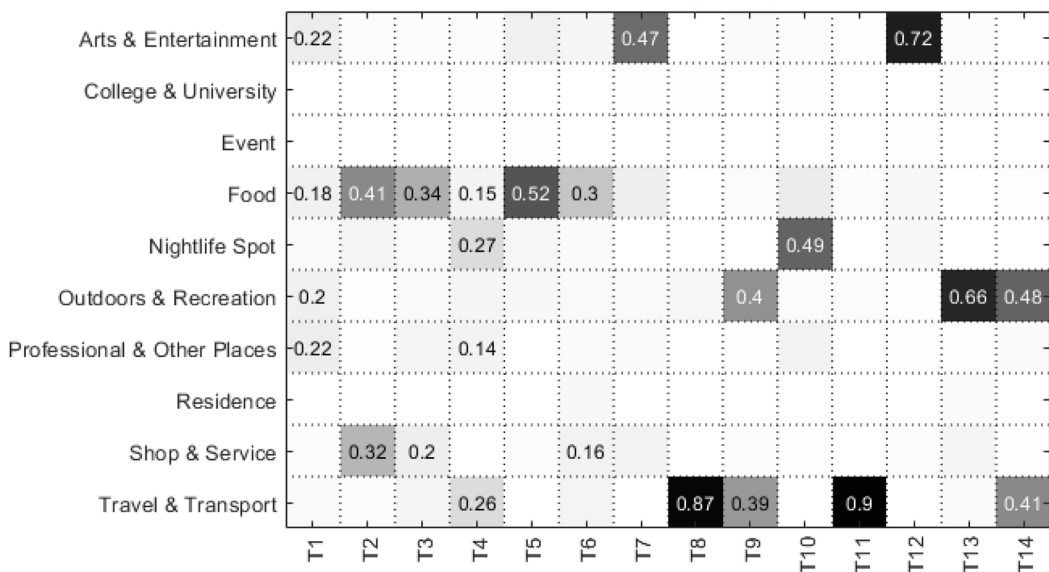**Figure 5.** Word clouds of venues in the topics.

Figure 6. Dominant activity types by topic.

Table 3. Activity records with dominant topics.

Activity record	Venue name	Category	Activity type
T1 – 0.948	Starbucks	Coffee Shop	Food
(Resident)	Square de Cluny	Garden	Outdoors & Recreation
	La Dame de Pic	French Restaurant	Food
	Square Paul Painlevé	Park	Outdoors & Recreation
	Cathedral of Notre Dame de Paris	Church	Professional & Other Places
	Shakespeare & Company	Bookstore	Shop & Service
	Paris City Hall	City Hall	Professional & Other Places
	BHV Marais	Department Store	Shop & Service
T1 – 0.942	Boutique Lindt	Chocolate Shop	Shop & Service
(Tourist —United Kingdom)	Garnier Opera	Opera House	Arts & Entertainment
	Shakespeare & Company	Bookstore	Shop & Service
	Les Tours de Notre-Dame	French Restaurant	Food
	Panthéon	Monument/Landmark	Professional & Other Places
	Sénat/Palais du Luxembourg	Capitol Building	Professional & Other Places
	Luxembourg Garden	Garden	Outdoors & Recreation
	Statue of Liberty of Paris	Outdoor Sculpture	Arts & Entertainment

index to refer to the topic rather than a single topic name. The interpretation of these topics can be based on the dominant activity types (Figure 6) and highly frequent words in the word cloud (Figure 5).

Table 3 shows an example of activity records that are strongly associated with specific topics. The first activity record is derived from local residents, who visited various venues, such as coffee shops, gardens, restaurants, parks and churches. The second activity record relates to a tourist who visited a chocolate shop, opera house, bookstore and restaurant, amongst other venues. Both activity records have a strong association with T1 as the visited venues belong to Food, Outdoor & Recreation, Shop & Service and Professional & Other Places. If the venues are examined separately (e.g. Vu et al., 2019), minimal similarity is found between the activities of the residents and tourists. However, LDA was able to identify similar preferences for activity types by considering the concurrence of the venue categories in the modelling process. Thus, the topics identified by the LDA model can be used to examine the similarities/differences between residents and tourists, which will be presented in the following section.

4.4. Comparative analysis of activities

Topic probabilities are computed on the basis of the activity records of each user group and plotted in Figure 7. Overall, our data show that tourists and residents tend to have different preferences for daily activities. T6, T7 and T8 are more popular amongst the residents than tourists. The records in these topics are consistent with the normal daily activities of residents, such as staying at home and going to bakeries (T6), visiting multiplex centres and game stores (T7), visiting fitness centres, traveling on trams and waiting at metro stations (T8). In contrast, T3, T4, T11 and T12 are more popular amongst tourists than residents. The venues in these topics are consistent with common tourism activities, such as being at airports (T4 and T11), staying at hotels (T4), attractions (T3), and visiting theme parks (T12). The t-test in Table 4 verifies the statistical significance of the differences. However, results also show that some residents and tourists appear to have similar preferences for daily activities following these topics. For instance, both tourists and residents are equally likely to visit venues belonging to food and shopping venues (T2 and T5). No significant difference was found between these two groups on these topics.

Resident's perception towards tourism often develops overtime through multiple encounters with tourists at common venues/places. These interactions do not need to be extreme to construct a negative perception. Minimal but frequent encounters over time at overcrowded venues is sufficient to induce the negative perception. Thus, identifying popular topics of activities is beneficial to identify such potential engagement in daily activities and travel itineraries between
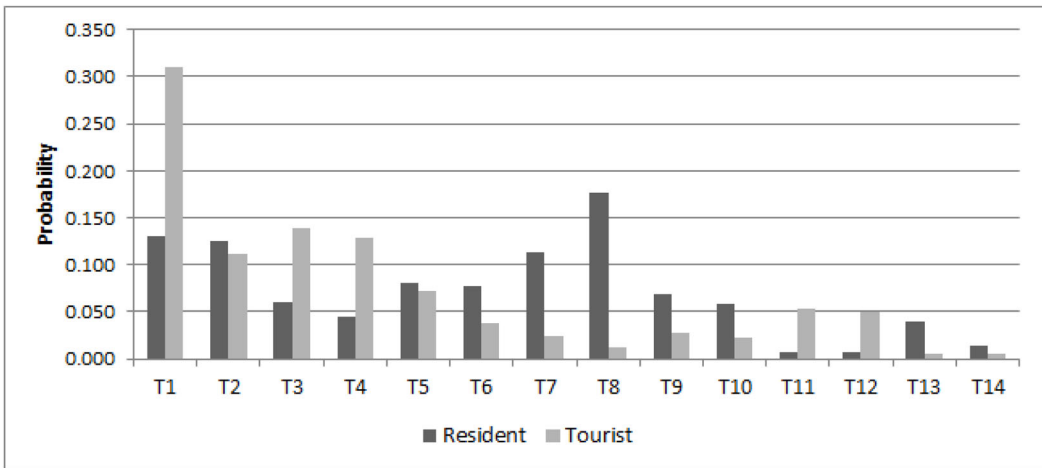
Figure 7. Topic popularity.

Table 4. T-test of the topic popularity (residents vs. tourists).

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14
T-statistics	376.73	3.68	129.08	159.89	2.08	71.61	314.10	686.05	77.40	74.36	109.58	88.56	108.33	13.31
p-Value*	0.00	0.06	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

*Significant at $p \leq 0.05$.

residents and tourists. Travel itineraries following T1 to T5 may need revision to identify and avoid potential clashing points with the residents. Such an analysis can be performed at the venue category level or at individual venues.

Figure 8 illustrates the popularity of sited venues in Paris with respect to the two groups. Results show that residents and tourists are likely to clash at popular venues, such as the Eiffel Tower, The Louvre, Paris Nord Railway, Arcde Triomphe, Cathedral of Notre Dame, Avenue des Champs-Elysees and Galeries Lafayette Haussmann. Moreover, this clash has been covered by recent media reports. Reports and imagery of Paris' two major tourist attractions, namely, the Eiffel Tower and the Louvre, have been controversially discussed and depicted as congested and crowded places with anti-tourism sentiments (e.g. The Guardian, 2019; Washington Post, 2019). A potential suggestion can be providing tourists with alternative venues with minimal local residents during particular times. Alternatively, tourism planners could develop long-term strategies and aim to diversify interest beyond these hotspots and attract tourists to visit alternative attractions at peak times. To mitigate the overtourism situation on a short-term basis, another approach is to recommend that tourists visit hotspots when only a few local residents are present, which can be identified by the temporal analysis of activities. This concept will be discussed in the following section.

4.5. Temporal analysis of activities

We analyse user activities from the temporal aspect. Figure 9 shows the popularity of the topics according to days of the week, which is computed on the basis of the dates of check-ins. Our analysis also reveals that popularity of the topics amongst the residents differs across the days of the week (Figure 9a), yet no remarkable difference was observed amongst tourists in terms of days of the week (Figure 9b). Tourists could consider organising their itineraries for T1 (e.g.
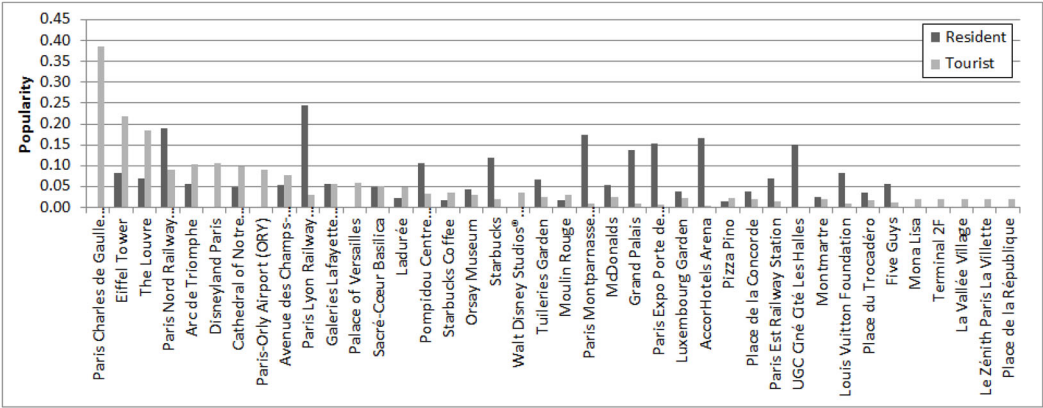
Figure 8. Hotspots of the residents and tourists.
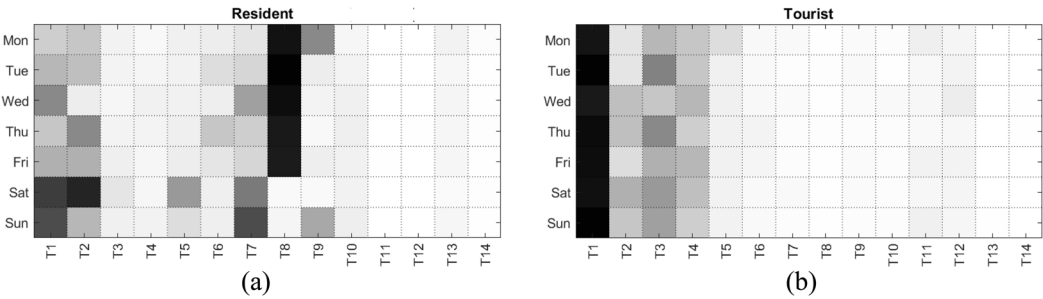

Figure 9. Topic popularity by day of the week.

dining, outdoor & recreation, art & entertainment) during weekdays to avoid clashing with local residents who are engaging in similar types of activities during weekends. For travel itineraries including dining and shopping activities following T2 and T5, tourists can select any day of the week except Saturday to avoid overcrowded conditions at related venues. Ideally, Wednesday is suitable for travel itineraries following T2, as residents tend to go to work in the middle of the week than engage in other leisure activities. Minimal recommendations can be made for T3 and T4 because tourists and residents are participating in these activities on all days of the week. Other topics were identified as less popular or bear less overlap between groups; thus, little recommendation can be made for tourists in these. Nevertheless, the heat map captured interesting facts about resident behaviour. For example, the residents are likely to visit Art & Entertainment venues during weekends (T7) or Travel & Transportation venues during weekdays (T8), probably for work or study.

The temporal analysis of activity is performed at the individual venue level for the exact time of day owing to the availability of local time information in the check-in data. Figure 10 displays the time of visit patterns at popular hotspots for the two groups. Correlation and *p*-value are also indicated in the figure titles. The analysis aims to identify the peak time of each user group and propose recommendations to avoid overcrowding at these venues at a particular time. Tourism managers may recommend that tourists visit the Eiffel Tower early in the day to avoid overcrowded conditions in the afternoon from 17:00, when many local residents also visit this tourist spot (Figure 10a). Tourists can avoid traveling to the Paris Nord Railway Station during the rush hours (18:00–19:00) of the local residents (Figure 10c). Tourists who want to visit Avenue de Champs-Elysees can avoid noon (12:00–13:00) and night (20:00) hours, when many local residents also visit this venue. The peak hours that tourists can visit The Louvre, Cathedral of Notre Dame and Galleries Lafayette Haussmann (Figure 10b, d, and f) do not directly overlap
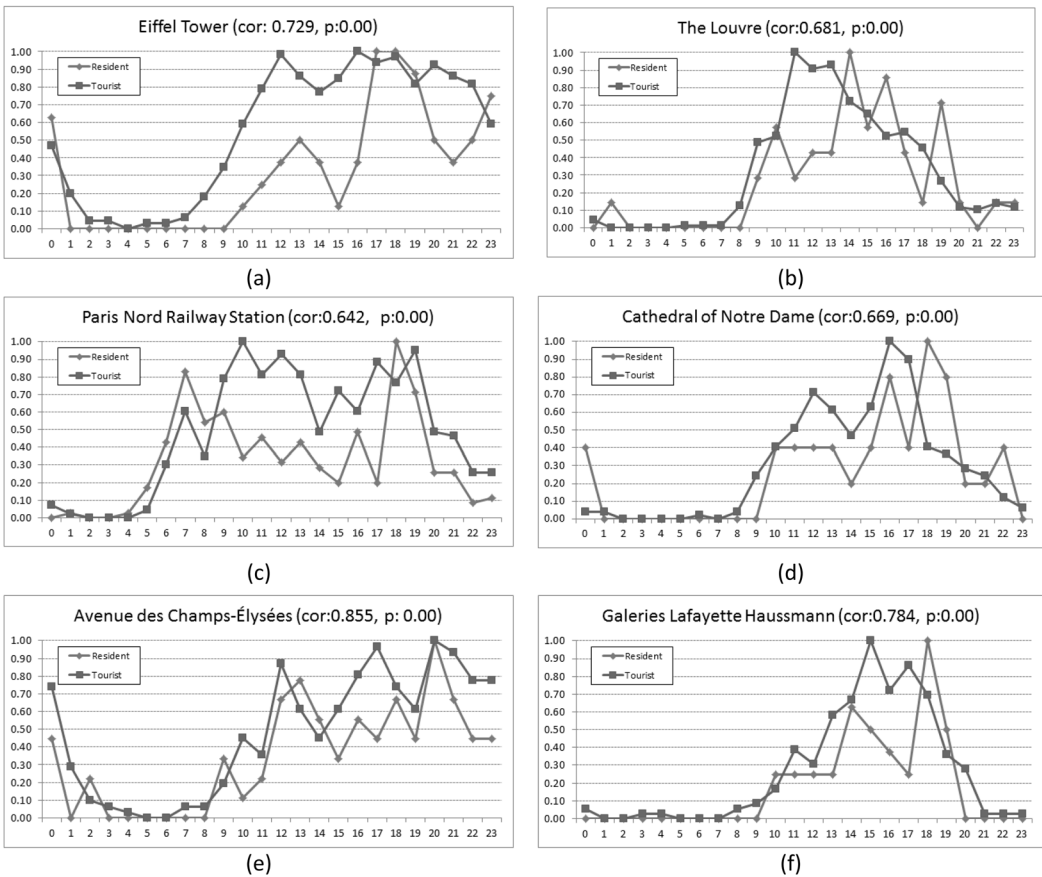
Figure 10. Venue popularity by time of day.

with the peak visiting hours of local residents but are relatively close. The reason for this finding is that local residents tend to visit these venues in the afternoon or evening. Thus, tourists may shift their visiting hours towards early in the morning or noon to minimise overlapping with the visiting hours of the local residents.

4.6. Discussion

The present study demonstrates the capability of our analytics-based approach to identify activities of residents and tourists that coincide at the same time, at a particular place and in this way, generate hotspot overtourism. Our contribution to the sustainable tourism literature is twofold:

First, with our case study of Paris, we add to the urban tourism literature (Ashworth & Page, 2011; Selby, 2012; Shoval, 2018) and provide new insights on how the social impact of overtourism can be tracked down to a few salient urban overtourism hotpots. Our results reveal the activity patterns, related venue categories and comparisons of these patterns for residents and tourists. Interestingly, our data provide evidence that, in some urban spaces, tourists and residents undertake different activities, but we also detect that few salient overtourism hotspots do exist. In these crowded hotspots, the time and activities of residents' daily life and travel itineraries coincide. Thus, in these congested places, high competition between residents and tourism occurs at particular times, and subsequently leads to perceptions of overtourism, which have

proven to impact overall destination images and destination attractiveness (e.g. Kim et al., 2020), and thus needs to be managed by tourism planners.

Second, this study offers an important methodological contribution and responds to problems identified in previous studies that often provided out-dated, insufficient or even unavailable data that caused issues with data accreditation, validity and reliability of results in prior studies. Although data analytics and social media data have been extensively used in the literature on tourism, our method adapts and extends the recent technological advancement in the field, namely, VR-SMD (Vu et al., 2020) in combination with topic modelling (Bi et al., 2019; Guo et al., 2017) as a technological aid for addressing the overtourism problem. Thus, we extend the sustainable tourism literature with an analytics-based approach, using innovative geo-tagged data. With this novel approach, we respond to calls to better address the complex issues in the sustainable tourism domain complexity (Font & McCabe, 2017). We advance the extant literature with concrete ideas on how tourism practice can be more sustainable (e.g. Lalicic, 2020; Joppe, 2018).

Viewing tourism as a place-based activity (e.g. Romão et al., 2018), we suggest that the exchange of spatial and temporal knowledge of activities and travel patterns can help residents, governments and tourists to better manage competition over urban resources at overtourism hotspots. Moreover, as prior studies often solely focus on data from either tourists or residents (e.g. Sharpley, 2014), we integrate data from both residents and tourists to offer a solution-based approach. Thus, our research approaches the overtourism problem in a holistic manner with short- and long-term strategies. In practice, residents benefit from this knowledge as they obtain up-to-date information on crowded hot spots and when to avoid them. Residents obtain information on access to tourist infrastructure and activities. Residents can be empowered with destination-based spatial and temporal knowledge, which may contribute to better participation and a more balanced social exchange and exchanging the benefits from tourism (e.g. Andereck et al., 2005).

The presented case study has several limitations. Firstly, we focused on the analysis from the spatial and temporal aspects, while resident's attitudes and sentiments about tourists at the hotpots have not been examined. This is because the aim of this study is to detect salient tourism hotspots by the use of a new analytics-based approach. Prior research has already focussed on resident's attitudes (e.g. Andereck et al., 2005; Kim et al., 2020; Namberger et al., 2019). We recommend that future studies could provide further insights by analysing textual comments available on various VR-SMD in future works. Secondly, we acknowledge that this study only provides a short-term solution for residents. Yet, governments can use our findings as a starting point for their development and implementation of long-term destination planning, such as diversifying marketing strategies beyond promoting popular sites and venues (e.g. Eiffel Tower and Notre Dame). Governments can focus on down-seasons for major events, adapt pricing strategies or restrict the number of visitors in the long run for severe cases. Lastly, tourists can use this knowledge to prepare for trips and understand residents and their activity patterns and obtain local knowledge about other avenues of available activities.

In terms of methodological limitation, data were collected from a single VR-SMD source, namely, Foursquare, representing behaviours of a specific groups of users, such as young people who use the Twitter and Foursquare platforms. Therefore, the findings may not represent and fully capture the activities of tourists and residents across age groups. Tourists from major markets, such as China, were not well represented in the collected dataset as Foursquare is not common in use amongst Chinese users. Therefore, other VRM-SMD platforms that are popular for interested tourist groups may be considered in the future for practical applications. Besides, differences in activities amongst tourists from different countries were not discussed, as our focus did not spotlight providing insights into tourist behaviour. Yet, we suggest that analysing cultural distinctions in behaviour could be an important aspect to consider in future research. Nevertheless, this paper demonstrates that data analytics using VR-SMD has the potential to

support decision-making in sustainable tourism management by comparing the behaviours of general tourists and residents for managerial applications.

5. Conclusions

This study uses Paris as a case study for overtourism and applies an innovative data analytics approach to better understand the potential social impact that tourism brings to residents. The results from social media data generated by tourists and residents show that (1) the spatial-temporal activity and movement patterns of tourists and residents are mostly different. However, our data analysis also reveals (2) a few salient overtourism hotspots where extreme crowding occurs at particular points in time. At these places where the competition over access to urban infrastructure is high, we finally show (3) that urban tourism hotpots vary according to time of the day, season and type of activity.

Our study advances the literature on social impact in the under-researched area of urban tourism. With our analytics-based approach and up-to-date high volumes of venue-referenced social media data, we specifically contribute to the sustainable tourism literature by offering a practice-based solution to mitigate urban hotspot overtourism in order to improve the resident–tourist relationship. Our holistic approach creates value for residents and tourists alike and assists governments for long-term strategic planning. While this study concentrated on contributing to the social dimension of overtourism, we acknowledge that a deep contextual understanding of other urban destinations is required because destinations attract various types of visitors, and their residents might hold different perceptions and values.

Our study also offers practical value and offers some concrete ideas for urban tourism planners to implement sustainable tourism practices. Importantly, we suggest that, at this moment, practitioners ought to reflect on how to better prepare their strategic planning for the time after the COVID-19 crisis. For example, strategic planning initiatives can include: 1) planning how to best get insights from an analysis of activity; 2) preparing for check-in data retrieval at the individual venue level at the exact time of day, owing to the availability of local time data to obtain visit patterns at popular hotspots from both residents and tourists when demand grows again; 3) strategizing for communication practices on how to best share this up-to-date information on crowded hot spots and when to avoid the hotspots for both tourists and residents; 4) developing mobile applications that support social distancing for safe travel during and after a pandemic based on the introduced approach owing to their capability of pinpointing users' locations and urban hotspots.

We acknowledge that perhaps not every tourism practitioner can easily model data in the way we did. Yet, with the increased knowledge on analytics-based methods and learning progress on Big Data analysis in university programs – within tourism and information technology degrees – we recommend hiring recent university graduates or working with student interns or with other specialists in this area. We recommend that tourism planners make an acknowledgement and effort to communicate their intent to empower residents with destination-based spatial and temporal knowledge. This will both help to improve the resident-tourist relationship and overall destination attractiveness.

Future studies should explore the solution-based approach considering three other dimensions (e.g. value alignment, destination context and resident demographics). Specifically, future research should provide a profound understanding of the role of tourism education and its positive impact on the broader community, thereby highlighting the benefits and risks of tourism. This aspect is similarly important to the levels of tourism-savviness and education in residents. In this way, residents may gain a higher appreciation of the negative and positive effects of tourism (World Tourism Organization (UNWTO) & IPSOS, 2019).

Acknowledgement

Disclosure statement

Funding

Notes on contributors

Huy Quan Vu is a senior lecturer at the Business school of Deakin University. His main interests are tourists behaviour analysis, tourism/hospitality data mining, and business intelligence.

Birgit Muskat is a senior lecturer at the Research School of Management, Australian National University. Her main interests are Entrepreneurship, knowledge transfer, and innovation behaviour and experience management.

Gang Li is an associate professor at the school of information technology, Deakin University. His research interests are artificial intelligence, data privacy and tourism/hospitality data mining.

Rob Law is a professor at the school of hotel & tourism management, the Hong Kong Polytechnic University. His research interests are technology applications to tourism and information management.

ORCID

Huy Quan Vu http://orcid.org/0000-0003-1947-2879
Birgit Muskat http://orcid.org/0000-0003-2905-6836
Gang Li http://orcid.org/0000-0003-1583-641X
Rob Law http://orcid.org/0000-0001-7199-3757

References

Andereck, K. L., Valentine, K. M., Knopf, R. C., & Vogt, C. A. (2005). Residents' perceptions of community tourism impacts. *Annals of Tourism Research*, *32*(4), 1056–1076. https://doi.org/10.1016/j.annals.2005.03.001
Ap, J. (1992). Residents' perceptions on tourism impacts. *Annals of Tourism Research*, *19*(4), 665–690. https://doi.org/10.1016/0160-7383(92)90060-3
Ashworth, G., & Page, S. J. (2011). Urban tourism research: Recent progress and current paradoxes. *Tourism Management*, *32*(1), 1–15. https://doi.org/10.1016/j.tourman.2010.02.002
Beritelli, P., & Laesser, C. (2011). Power dimensions and influence reputation in tourist destinations: Empirical evidence from a network of actors and stakeholders. *Tourism Management*, *32*(6), 1299–1309. https://doi.org/10.1016/j.tourman.2010.12.010
Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, *70*, 460–478. https://doi.org/10.1016/j.tourman.2018.09.010
Blau, P. (2017). *Exchange and power in social life*. Routledge.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://doi.org/10.5555/944919.944937
Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*, 2079–2107. https://doi.org/10.5555/1756006.1859921
Cheung, K. S., & Li, L. H. (2019). Understanding visitor–resident relations in overtourism: Developing resilience for sustainable tourism. *Journal of Sustainable Tourism*, *27*(8), 1197–1216. https://doi.org/10.1080/09669582.2019.1606815

Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in Southern Italy. *Tourism Management*, *57*, 295–310. https://doi.org/10.1016/j.tourman.2016.06.013

Deery, M., Jago, L., & Fredline, L. (2012). Rethinking social impacts of tourism research: A new research agenda. *Tourism Management*, *33*(1), 64–73. https://doi.org/10.1016/j.tourman.2011.01.026

DGE. (2016). *82.6 million foreign tourists in France in 2016*. Études Économiques, Direction Générale des Entreprises, Retrieved Octorber 28, 2018, from https://www.entreprises.gouv.fr/files/files/directions_services/etudes-et-statis-tiques/4p-DGE/2017-06-4p-N71-EVE-ENGLISH.pdf

Font, X., & McCabe, S. (2017). Sustainability and marketing in tourism: Its contexts, paradoxes, approaches, challenges and potential. *Journal of Sustainable Tourism*, *25*(7), 869–883. https://doi.org/10.1080/09669582.2017.1301721

Gibson, C. (2019). Critical tourism studies: New directions for volatile times. *Tourism Geographies*, 1–19. https://doi.org/10.1080/14616688.2019.1647453

Greene, D., & Cross, J. P. (2015). *Unveiling the political agenda of the European parliament plenary: A topical analysis* [Paper presentation]. Proceedings of the ACM Web Science Conference, Oxford, UK. (Article No. 2) https://doi.org/10.1145/2786451.2786464

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using Latent Dirichlet Allocation. *Tourism Management*, *59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Hall, P. (1970). A horizon of hotels. *New Society*, *15*, 389–445.

Heo, C. Y., Blal, I., & Choi, M. (2019). What is happening in Paris? Airbnb, hotels, and the Parisian market: A case study. *Tourism Management*, *70*, 78–88. https://doi.org/10.1016/j.tourman.2018.04.003

Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval* (pp. 50–57). https://doi.org/10.1145/312624.312649

Jacobs, B. J. D., Donkers, B., & Fok, D. (2016). Model-based purchase predictions for large assortments. *Marketing Science*, *35*(3), 389–404. https://doi.org/10.1287/mksc.2016.0985

Jacobsen, J. K. S., Iversen, N. M., & Hem, L. E. (2019). Hotspot crowding and over-tourism: Antecedents of destination attractiveness. *Annals of Tourism Research*, *76*, 53–66. https://doi.org/10.1016/j.annals.2019.02.011

Joppe, M. (2018). Tourism policy and governance: Quo vadis? *Tourism Management Perspectives*, *25*, 201–204. https://doi.org/10.1016/j.tmp.2017.11.011

Ketchen, D. J., Jr., & Shook, C. L. (1996). The application of cluster analysis in Strategic Management Research: An analysis and critique. *Strategic Management Journal*, *17*(6), 441–458. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6 < 441::AID-SMJ819 > 3.0.CO;2-G

Kim, G., Duffy, L. N., & Moore, D. (2020). Tourist attractiveness: Measuring residents' perception of tourists. *Journal of Sustainable Tourism*, *28*(6), 898–916. https://doi.org/10.1080/09669582.2019.1708919

Lalicic, L. (2020). Solastalgia: An application in the overtourism context. *Annals of Tourism Research*, *82*, 102766. https://doi.org/10.1016/j.annals.2019.102766

Laurini, R. (2017). *Geographic knowledge infrastructure: Applications for territorial intelligence and smart cities*. Elsevier.

Lim, J., & Lee, H. C. (2020). Comparisons of service quality perceptions between full-service carriers and low-cost carriers in airline travel. *Current Issues in Tourism*, *23*(10), 1261–1276. https://doi.org/10.1080/13683500.2019.1604638

Lin, Z., Chen, Y., & Filieri, R. (2017). Resident-tourist value co-creation: The role of residents' perceived tourism impacts and life satisfaction. *Tourism Management*, *61*, 436–442. https://doi.org/10.1016/j.tourman.2017.02.013

Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2019). Tourist behavior analysis in gaming destinations based on venue check-in data. *Journal of Travel & Tourism Marketing*, *36*(1), 107–118. https://doi.org/10.1080/10548408.2018.1494088

Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2020). Topic modelling for theme park online reviews: Analysis of Disneyland. *Journal of Travel & Tourism Marketing*, *37*(2), 272–285. https://doi.org/10.1080/10548408.2020.1740138

Maxim, C. (2016). Sustainable tourism implementation in urban areas: a case study of London. *Journal of Sustainable Tourism*, *24*(7), 971–989. https://doi.org/10.1080/09669582.2015.1115511

Mazanec, J. A. (2017). Determining long-term change in tourism research language with text-mining methods. *Tourism Analysis*, *22*(1), 75–83. https://doi.org/10.3727/108354217X14828625279771

McGuire, K. A., Kimes, S. E., Lynn, M., Pullman, M. E., & Lloyd, R. C. (2010). A framework for evaluating the customer wait experience. *Journal of Service Management*, *21*(3), 269–290. https://doi.org/10.1108/09564231011050760

Mody, M., Suess, C., & Dogru, T. (2019). Not in my backyard? Is the anti-Airbnb discourse truly warranted? *Annals of Tourism Research*, *74*(C), 198–203. https://doi.org/10.1016/j.annals.2018.05.004

Monterrubio. (2016). The impact of spring break behaviour: An integrated threat theory analysis of residents' prejudice. *Tourism Management*, *54*, 418–427. https://doi.org/10.1016/j.tourman.2015.12.004

Namberger, P., Jackisch, S., Schmude, J., & Karl, M. (2019). Overcrowding, overtourism and local level disturbance: How much can Munich handle? *Tourism Planning & Development*, 16(4), 452–472. https://doi.org/10.1080/21568316.2019.1595706

Novy, J., & Colomb, C. (2019). Urban tourism as a source of contention and social mobilisations: A critical review. *Tourism Planning & Development*, 16 (4), 358–375. https://doi.org/10.1080/21568316.2019.1577293

Office du Tourisme et des Congrès (OTCP). (2019). *Tourism in Paris - Key Figures 2018*. Retrieved October 5, 2019, from https://pro.parisinfo.com/etudes-et-chiffres/key-figures-english-version/annual-key-figures

O'Reilly, A. M. (1986). Tourism carrying capacity: Concept and issues. *Tourism Management*, 7(4), 254–258. https://doi.org/10.1016/0261-5177(86)90035-X

Papadimitriou, C. H., Raghavan, P., Tamaki, H., & Vempala, S. (2000). Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences*, 61(2), 217–235. https://doi.org/10.1006/jcss.2000.1711

Pasquinelli, C., & Trunfio, M. (2020). Overtouristified cities: An online news media narrative analysis. *Journal of Sustainable Tourism*, 28(11), 1805–1824. https://doi.org/10.1080/09669582.2020.1760871

Pearce, P. L. (2019). Are tourists interesting? In P. L. Pearce (Ed.), *Tourist behaviour. The essential companion* (pp. 1–19). Edward Elgar.

Phi, G. T. (2020). Framing overtourism: A critical news media analysis. *Current Issues in Tourism*, 23(17), 2093–2097. https://doi.org/10.1080/13683500.2019.1618249

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

Romão, J., Kourtit, K., Neuts, B., & Nijkamp, P. (2018). The smart city as a common place for tourists and residents: A structural analysis of the determinants of urban attractiveness. *Cities*, 78, 67–75. https://doi.org/10.1016/j.cities.2017.11.007

Samuel, H. (2019). *Paris set to ban tourist buses from city centre due to overtourism*, Traveller.com.au. Retrieved October 4, 2019, from https://www.traveller.com.au/paris-set-to-ban-tourist-buses-from-city-centre-due-to-over-tourism-h1fw8c

Scuderi, R., & Nogare, C. D. (2018). Mapping tourist consumption behaviour from destination card data: What do sequences of activities reveal? *International Journal of Tourism Research*, 20 (5), 554–565. https://doi.org/10.1002/jtr.2205

Selby, M. (2012). Geographies of tourism and the city. In J. Wilson (Ed.), *The Routledge handbook of tourism geographies* (pp. 232–239). Routledge.

Sharpley, R. (2014). Host perceptions of tourism: A review of the research. *Tourism Management*, 42, 37–49. https://doi.org/10.1016/j.tourman.2013.10.007

Shoval, N. (2018). Urban planning and tourism in European cities. *Tourism Geographies*, 20(3), 371–376. https://doi.org/10.1080/14616688.2018.1457078

Sobolevsky, S., Bijic, I., Belyi, A., Sitko, I., Hawelka, B., Arias, J. M., & Ratti, C. (2015). *Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity*. In IEEE International Congress on Big Data, pp. 600–607.

Stringam, B. B., Gerdes, J., & Vanleeuwen, D. M. (2010). Assessing the importance and relationships of ratings on user-generated traveler reviews. *Journal of Quality Assurance in Hospitality & Tourism*, 11(2), 73–92. https://doi.org/10.1080/1528008X.2010.482000

The Guardian. (2019). *Mona Lisa fans decry brief encounter with their idol in Paris. Louvre visitors complain of being moved on in seconds after spending hours in queue*. Retrieved September 23, 2019, from https://www.theguardian.com/world/2019/aug/13/mona-lisa-fans-face-scandalous-queues-after-portrait-relocated

Torres-Delgado, A., Palomeque, F.L. (2014). Measuring sustainable tourism at the municipal level. *Annals of Tourism Research*, 49, 122–137, https://doi.org/10.1016/j.annals.2014.09.003.

Trattner, C., Helic, D., & Strohmaier, M. (2014). Tag clouds. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of social network analysis and mining*. Springer.

Van der Borg, J., Costa, P., & Gotti, G. (1996). Tourism in European heritage cities. *Annals of Tourism Research*, 23(2), 306–321. https://doi.org/10.1016/0160-7383(95)00065-8

Vu, H. Q., Leung, R., Rong, J., & Miao, Y. (2016). Exploring park visitors' activities in Hong Kong using geotagged photos. In *Proceedings of 23rd Information and Communication Technologies in Tourism*, pp. 183–196. https://doi.org/10.1007/978-3-319-28231-2_14

Vu, H. Q., Li, G., & Law, R. (2019). Discovering implicit activity preferences in travel itineraries by topic modeling. *Tourism Management*, 75, 435–446. https://doi.org/10.1016/j.tourman.2019.06.011

Vu, H. Q., Li, G., & Law, R. (2020). Cross-country analysis of tourist activities based on venue-referenced social media data. *Journal of Travel Research*, 59(1), 90–106. https://doi.org/10.1177/0047287518820194

Washington Post. (2019). *These are the cities overtourism could threaten next*. Retrieved October 5, 2019, from https://www.washingtonpost.com/travel/2019/06/18/these-are-cities-overtourism-could-threaten-next/

Williams, J., & Lawson, R. (2001). Community issues and resident opinions of tourism. *Annals of Tourism Research*, 28(2), 269–290. https://doi.org/10.1016/S0160-7383(00)00030-X

World Tourism Organization (UNWTO). (2018). 'Overtourism'? – *Understanding and managing urban tourism growth beyond perceptions, executive summary*. World Tourism Organization.

World Tourism Organization (UNWTO) and IPSOS. (2019). *Global survey on the perception of residents towards city tourism: Impact and measures*. UNWTO. https://doi.org/10.18111/9789284421107.

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

Yoo, K. H., & Lee, W. (2015). Use of Facebook in the US heritage accommodations sector: An exploratory study. *Journal of Heritage Tourism*, *10*(2), 191–201. https://doi.org/10.1080/1743873X.2014.985228

Zhang, H-y., Ji, P., Wang, J-q., & Chen, X-h. (2017). A novel decision support model for satisfactory restaurants utilizing social information: A case study of TripAdvisor.com. *Tourism Management*, *59*, 281–297. https://doi.org/10.1016/j.tourman.2016.08.010