Journal of Sustainable Tourism

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rsus20

Integrating tourists’ walk and talk: a methodological approach for tracking and analysing tourists’ real behaviours for more sustainable destinations

Ilenia Confente, Valentina Mazzoli, Nicola Camatti & Dario Bertocchi

JOURNAL OF SUSTAINABLE TOURISM
https://doi.org/10.1080/09669582.2024.2322120

Integrating tourists' walk and talk: a methodological approach for tracking and analysing tourists' real behaviours for more sustainable destinations

Ilenia Confente[a], Valentina Mazzoli[a], Nicola Camatti[b] and Dario Bertocchi[c,d]

[a]Department of Management, University of Verona, Italy; [b]Department of Economics of the Ca' Foscari University of Venice, Italy; [c]Department of Languages, Literature, Communication, Education and Society, University of Udine, Italy; [d]School of Tourism and Service Management, Modul University Vienna, Austria

ABSTRACT

This study provides a methodological contribution in the analysis of tourists' sustainable behaviour. Traditional methods encounter challenges in addressing the attitude-behaviour gap, prompting the need for new approaches. Using diverse big data sources, the paper introduces a novel methodology for analysing real behaviour patterns, extending Social Network Analysis techniques. The aim is to identify sustainable behaviours that ease overcrowding, distribute visitor flows, and optimize economic diversification. The methodology is valuable for local authorities in monitoring tourist behaviours, aiding in informed decision-making for a more sustainable approach to tourism.

ARTICLE HISTORY
Received 30 April 2023
Accepted 13 February 2024

KEYWORDS
Sustainable tourism; social network analysis; tourism flows; user-generated content; big data; real behaviour

Introduction

Tourism is a double-edged sword. On one hand, it produces several economic benefits for travel destinations. On the other hand, it can have negative impacts on travel destinations, ranging from reduced security of overcrowded sites to deterioration of nature and the environment due to higher levels of transportation-caused pollution and energy consumption (Agrawal et al., 2022).

Reducing the harmful environmental effects of tourism has become a priority for many travel destinations, as highlighted in the 2030 Agenda for Sustainable Development Goals (SDGs) (UN, 2021). Sustainable tourism focuses on reducing negative environmental, social, and economic impacts of tourism (Bramwell et al., 2017). Destination Management Organisations (DMOs) are increasingly attentive to tourist usage patterns and are working to comprehend and address the issue of overtourism, defined as the overutilization of resources and deterioration of cultural landmarks due to rapid tourism industry expansion (Mihalic, 2020). Overtourism is influenced by various factors, including the rise of short-stay accommodations (e.g. Airbnb), low-cost airline routes, and the predominant economic focus of DMOs over other sustainability dimensions (Butler & Dodds, 2022). Efforts to mitigate overtourism often encounter challenges due to insufficient control over policies at different stakeholder levels (Butler & Dodds, 2022; Gowreesunkar & Vo Thanh, 2020). This paper seeks to equip DMOs and policymakers with a tool to track and manage tourist flows and behaviours, with a specific emphasis on promoting sustainability.

CONTACT Ilenia Confente ilenia.confente@univr.it Via Cantarane, 24, 37129 Verona, Italy
Supplemental data for this article can be accessed online at https://doi.org/10.1080/09669582.2024.2322120.

The UNWTO and UNDP (2017), UNWTO UNWTO, UNWTO, (2022) provides a list of guidelines for DMOs to promote sustainable tourism, through which DMOs should manage tourist flows to alleviate congestion in the most crowded areas and to promote a diversification of activities by decongesting the city centres (Camatti et al., 2020).

The success of sustainable tourism policies depends on tourists' behaviour. Although recent studies have demonstrated that tourists are largely interested in incorporating the dimensions of sustainability into their tourism experiences (Confente & Scarpi, 2021; Dolnicar et al., 2017), they still behave differently in reality (Agag et al., 2020). This inconsistency between intentions and real behaviours is known as the intention–behaviour gap (Ajzen, 1991, 2001). Accordingly, many authors have highlighted the difficulty in deriving real behaviour evidence from intentions or attitudes (e.g. Holmes et al., 2021; Viglia & Acuti, 2022), measured through survey-based or interview-based methods. Traditional survey or interview-based methods face limitations, including social desirability bias, where respondents tend to present themselves as environmentally responsible (Holtgraves, 2004).

Given that sustainable tourism impacts occur collectively, big data emerges as a valuable tool to analyse tourists' real behaviours over time (Agrawal et al., 2022; Xu et al., 2020). The tourism industry provides ample opportunities for collecting and analysing big data from various sources, including unstructured data (Balducci & Marinova, 2018; Eberendu & Madonna University, 2016) and structured data (Li et al., 2018). The combination of structured data (e.g. mobile phone network data embedded in a database) and unstructured data (e.g. text, images, email, personal information, and other data types that are not part of a database; Feldman & Sanger, 2007) allows for overcoming the limitations of the latter. However, research combining multiple big data sources to analyse sustainable patterns remain debated (Xu et al., 2020).

Accordingly, this study provides a methodological contribution in the analysis of tourists' data, proposing this research question (RQ):

> RQ: How can big data approaches, encompassing structured and unstructured data, help tourism organizations better understand tourists' sustainable behaviours, particularly in identifying and analysing effective sustainable behaviours?

The proposed approach combines offline structured data (mobile phone network data) and online unstructured data (UGC) to depict tourists' sustainable behaviours concerning mobility flows in a destination over time. The outlined analytical protocol serves as a guide for policymakers to optimize mobility flows and enhance destination offerings.

To test and validate this methodological approach, the city of Verona, located in northeast Italy, was selected for this study. Like many cultural cities worldwide, such as Barcelona and Venice (Coldwell, 2017), Verona has encountered challenges associated with overtourism. This has prompted the city and its institutions to develop sustainable solutions for managing tourist flows.

Theoretical background

From self-reported to real behaviour data: a necessary shift for tourism sustainability research

Methods for analysing sustainable tourism are now evolving to face the emerging changes in tourists' behaviours and in innovation development levels. In fact, tourism research has mainly relied on methods such as surveys and in-depth interviews (Xu et al., 2020), limiting the use of field experiments, longitudinal studies, neuromarketing techniques and big data analysis (Bramwell et al., 2017; Savelli et al., 2022).

One of the issues presented by methods relying on self-reported measures (e.g. questionnaires or face-to-face interviews) is the apparent intention–behaviour gap who highlight the difficulty in deriving real behaviour evidence from intentions or attitudes (Holmes et al., 2021). It is a long-term issue in academic debate that has become critical for sustainable studies (Viglia & Acuti, 2022).

This is in line with previous studies in different fields that have shown attitude as a predictor of behavioural intentions, rather than real behaviours (Ajzen, 1991, 2001). For instance, Juvan and Dolnicar (2014) reported that an environmentalist group's members, who were assumed to hold strong pro-environmental attitudes, revealed that they did not always behave in an environmentally friendly manner when on vacation. In their research, Passafaro et al. (2015) found similar results.

Beyond the intention–behaviour gap issue, another deficiency in most traditional methods, is the well-known bias related to social desirability. This occurs because individuals aim to create positive representations of themselves and gain recognition from their favourable social image (Juvan & Dolnicar, 2016). Despite various approaches used to control for social desirability bias, such as the social desirability scale (Crowne & Marlowe, 1960), this problem still exists (Hibbert et al., 2013; Juvan & Dolnicar, 2016).

Based on these issues and since sustainable tourism is mainly a result of aggregated behaviour, and its impact is made at a collective level, big data and related analyses may provide more extensive and complete information regarding social and environmental sustainability of a tourism destination. This is done by tracking and mapping the relations between visitors and the natural and social contexts and assessing their impact on the destination. Such an approach also allows capturing information on a much larger scale and reflecting longitudinal change over time (Kitchin, 2013) at a moderate cost. This advantage, together with technology development, has led to the rapid evolution of the approaches utilised for tourism research (Xu et al., 2020).

The term *big data* has been defined as "the enormous volume of both unstructured and structured data generated by technology developments and the exponentially increasing adoption of devices allowing for automation and connection to the internet" (Mariani et al., 2018, p. 3515). The big data approach can be useful for overcoming the limitations of the conventional methodologies used in scientific research, and some researchers have identified it as a new paradigm in the field of investigation (Serrano et al., 2021; Xiang et al., 2017).

The tourism industry offers a rich context for collecting and analysing big data, coming from different sources: UGC data (e.g. social media), device data (e.g. from mobile devices, GPS and Bluetooth), and transaction data (e.g. web search data and online forms; Li et al., 2021). Among them, the most popular source of data for the tourism and hospitality sector is UGC, particularly online reviews (Mariani & Borghi, 2021).

UGC constitutes both structured (e.g. the rating by a review) and unstructured data (e.g. the written text). These data help provide an understanding of customer satisfaction and experience that is useful for tourism services, as well as capture customer sentiment that is valuable for tourist destinations (Mariani et al., 2018).

Among several advantages of big data analysis, recent research has emphasised that through this approach, researchers can collect and analyse large amounts of data with real-time information and no spatial limitations, switching from anecdotal to evidence-based data (Song et al., 2015). Moreover, big data facilitate conducting longitudinal studies due to constant/regular data capture, easy data storage and low costs (Xu et al., 2020).

Most of the big data analyses have been applied to the general tourism field. Despite a few exceptions where these analyses have been adopted to understand sustainable tourism behaviour (for extensive reviews, see Agrawal et al., 2022; Xu et al., 2020), scholars have not yet ascertained their full potential. Hence, such an approach can provide novel opportunities for sustainable tourism research, ranging from better visualisation of tourists' real choices and mobility to the identification of tourist saturation zones and mapping of tourists' frequented attractions and places in specific destinations (Serrano et al., 2021).

For instance, previous studies have revealed the importance of big data approaches for analysing hotels' sustainability and travellers' behaviours (Vu et al., 2015), as well as for understanding the impact of online environmental discourses on tourists' WOM regarding the hospitality sector (Mariani & Borghi, 2021). Other studies have utilized big data analyses to comprehend tourism mobility flows for the purpose of managing transportation networks and urban mobility (e.g. Serna et al., 2017), albeit without a specific emphasis on tracking and understanding tourists' sustainable behaviours and choices.

Along with the benefits of adopting big data for tourism research, big data have a few dark sides as well (Mazanec, 2020). Mazanec (2020, p. 2) advocates the "let the data speak" approach, asserting that while quantitative methodologies can benefit from big data through descriptive, predictive, and prescriptive analytics, decision-making still involves intervention. This relies on causal relations, which are often not supplied by a single source of big data. For sustainable tourism research, big data provide the "full picture" of the phenomenon (in this case, the identification and evolution of sustainable behaviours) but do not explain its causal mechanisms (in understanding the reasons for changes in tourists' behaviours). Additionally, big data analyses can cover a large sample but not necessarily the entire sample; finally, it is not always possible to have access to open data, and it involves managing personal data privacy concerns (Xu et al., 2020).

To overcome these limitations, researchers and destinations should try to implement a combination of data and information and develop subsequent qualitative studies to interpret the results derived from big data.

In the case of UGC, social media and online platforms allow us to understand the relationships between individuals, as well as between people and places. This type of information is critical in managing environmental sustainability, understanding the sociocultural contexts of visitors, and providing better services (Xu et al., 2020). Such information can also be compared and verified through geo-information and visitors' movements.

The analysis of tourist flows can be expanded to incorporate more fine-grained analysis, identifying the movements of specific actors and their relationships with one another (Xu et al., 2020). This can help in providing practical implications for impact management, destination development and transportation planning (Demeter et al., 2023).

Hence, our study offers a methodological contribution in the analysis of tourists' data with a specific focus on sustainable tourist behaviours. The integration of UGC with mobile devices' data (i.e. TELCO data) allows a comparison of tourists' behaviours across time. Such an approach provides more granularity, integrating spatial and temporal data, which cannot be supplied by self-reported data derived from questionnaires.

Social network analysis as a valuable big data analysis technique

Social Network Analysis (SNA) is valuable for processing big data to discover destination dynamics (Baggio, 2008; Provenzano & Baggio, 2020). Recent tourism literature has underscored the aims and contributions of network analysis, especially applied to big data, for the study of tourists' behaviours and destination performances. They are briefly summarised as follows:

- Understanding the flow of tourists. Mapping out the connections between destinations and tourist visits can provide an understanding of how tourists move through a particular area or tourist destination (Kádár & Gede, 2021; Li et al., 2021; Seok et al., 2021; Taczanowska et al., 2014).
- Identifying key attractions. SNA can help detect or confirm the key players in a tourist network, such as popular attractions or hot-spot areas to be visited. This can provide insights into the significant factors for shaping tourists' behaviours (Baggio, 2017; D'Angelo et al., 2023; Van der Zee et al., 2020; Van der Zee & Bertocchi, 2018).

- Analysing online behaviour. SNA can be used to study online behaviour, such as the posted messages, comments or responses on the websites or social media platforms that are most popular among tourists (Kang et al., 2021).
- Predicting future behaviour and identifying current trends. This can be done by analysing patterns of tourists' behaviours over time (Li et al., 2021; Li & Law, 2020; Mukai & Ikeda, 2022; Valeri & Baggio, 2021).

Furthermore, SNA aids in elucidating complex destination systems in a more precise manner. Network science clarifies the properties related to a network's influence and behaviour, along with its overall dynamic behaviour (Baggio et al., 2010). Indeed, it facilitates the examination of a tourism destination from a network perspective, representing linkages with places, attractions, and stakeholders. Valeri and Baggio (2021) contemplate the use of SNA to enhance understanding and predict collective behaviour as the sum of individual (visitors') behaviours.

However, limitations arise from the incomplete exploitation of these studies in examining tourists' sustainable behaviours. This deficiency underscores the necessity for the development of robust metrics capable of incorporating sustainability criteria, thus enabling longitudinal comparisons in line with dynamic network analysis (Baggio et al., 2010).

The need for longitudinal comparisons highlights the dynamic nature of tourist behaviours and the evolving landscapes of destinations. Encompassing extended timeframes, longitudinal studies will provide a more comprehensive understanding of the sustainability of tourist interactions tracking changes and patterns over time.

Research design

As stated before, we introduce a novel methodological approach for identifying and evaluating the sustainable behaviours of tourists during their visits to destinations.

Figure 1 illustrates the analytical framework that we propose for achieving this goal. We focus on tourists' real behaviours, represented by the type of the network of activities that they undertake at the intra-destination level. The network of tourists' activities is used as an indication of their real commitment to support sustainable tourism development in the visited areas, regarding the decongestion of the most touristy places and diversification of tourist experiences.

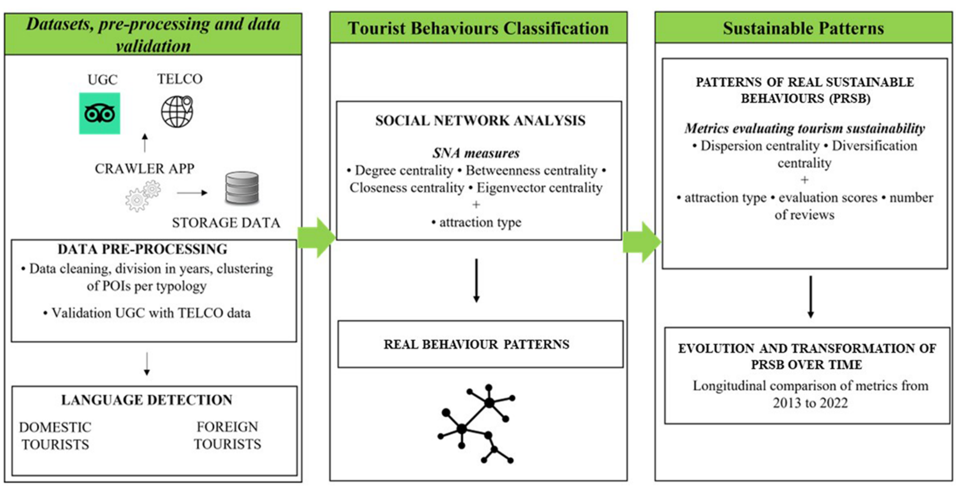

Figure 1. Research design's analytical framework.

To overcome the limitations of collecting such data at the intra-destination level, we rely on the digital footprints left by tourists in the form of UGC as a useful proxy of their patter of visit among tourist attractions within a set time. We use SNA techniques to identify sustainable behaviour patterns, based on the network structure of the collected UGCs. High-frequency TELCO data are further used to consolidate the collected dataset and enhance the results.

The sustainability of actual behaviours is identified based on a series of discriminating criteria applied to the structure of the network, aligning with the UNWTO and UNDP (2017), UNWTO UNWTO, UNWTO, (2022) guidelines for the development of sustainable tourism. Specifically, on one hand, these criteria include the identification of the tourists' behaviours that favour less touristic areas in the destination to limit overcrowding issues. On the other hand, these criteria focus on the types of activities performed by tourists.

Based on these conditions, we evaluate the nature of the network of visited attractions and measure the importance of the patterns of detected sustainable real behaviours, adapting the main metrics of SNA, including degree centrality, betweenness centrality, closeness centrality and Eigenvector centrality, combined with other relevant information extracted from UGC, such as attraction type, number of visits, and reviews. These measures are explained in detail later in the study.

This form of graph theory analysis is valuable in determining the significance of any given node in a network. It has the capacity to eliminate noise in relationships, uncovering segments of the network with distinct characteristics. The centrality indicators developed by SNA can address biases and gaps in social science studies, particularly in relation to the tourism industry and visitors' behaviours.

In this study, our methodology allows us to analyse real behaviour patterns of both domestic and foreign tourists. It also longitudinally captures the evolution and transformation of sustainable real behaviours over time. We apply this analysis framework to the case study of the city of Verona, Italy, using UGC posted on the TripAdvisor platform from 2013 to 2022. The analyses are performed using special libraries suitable for the Python environment, particularly the NetworkX package (Hagberg et al., 2008).

Case study

Verona is a good example of an overcrowded city that currently struggles to decongest its highly visited areas and redirect tourist flows by promoting alternative solutions. Verona attracted 1,228,362 tourists in 2022 (Sistema Statistico Regionale, 2023), and the city represents the fourth Italian destination for the number of foreign tourists over the total number of tourists.

Due to this flourishing interest in Verona, the city provides hospitality and tourism-related services, with growing opportunities and economic performance. However, it also leads to over-tourism in high seasonality months, which causes some problems for both residents and tourists, such as traffic congestion, limited parking availability, pollution and attraction queues (Mazzara, 2023; Scarpi et al., 2022).

Accordingly, among the goals of the most recent Strategic Plan of the DMO (Piano Strategico Destinazione Verona, 2017), one is related to reducing traffic and tourist congestion at the city centre. Another goal is to differentiate the destination offers to allow multiple tours and experiences for tourists and residents. Moreover, because this city mostly depends on jobs and revenues generated by tourism and this industry has a substantial environmental footprint, Verona represents an appropriate context of analysis of socially and environmentally sustainable economic growth patterns, as recommended by SDGs 2030 (UN, 2021).

Datasets, pre-processing and validation

We collected a database of all tourism sites and services listed as "Things to Do" on the TripAdvisor platform using a crawler that we developed. This web app can automatically collect information voluntarily left by visitors on the internet, providing a database structure to the unstructured data available in social networks and reviews posted on the website. The database is structured with 93,879 reviews from 2013 to 2022, provided by visitors to the 692 attractions in Verona. These attractions are grouped into eight macro categories: Cultural Sites (e.g. museums), Landmarks and Sites of Interest (e.g. monuments), Tours and Activities (e.g. guided tours and experiences), Entertainment and Events (e.g. theatres), Natural Sites (e.g. parks), Shopping Facilities (e.g. outlets), Bars and Clubs (e.g. nightlife) and Transport and Services (e.g. train stations and shuttle services). Each record of the database contains this information (related to both the experience and the profile of the reviewer), including the place of the review, date of the review (dd-mm-yyyy), date of the visit (month and year), title of the review, score of the review (from 1 to 5), text of the review, name of the reviewer and home location of the reviewer.

Since reviews reflect consumer experience "*in natura* without any interference from researchers" (Sánchez-Franco & Rey-Moreno, 2022, p. 441), these are unable to completely describe the full experience. Therefore, we decided to find the correlation between the number of visitors and reviews at weekly and monthly levels to confirm and validate the dataset regarding the usage of reviews as a proxy of tourism flows. This was done to overcome the intrinsic limitations of big data, as expressed by Mazanec (2020) and illustrated in advance, and to address the needs regarding data quality and consolidation (as indicated by Brave et al., 2022). Matching two different sources of big data—unstructured UGC data about the tourism experience and structured mobile phone data regarding visitor presence and mobility—allowed us to consolidate the results regarding tourists' real behaviours.

The two databases were compared through a correlation to confirm whether the number of reviews represented the actual number of visitors hosted by the destination under examination.

In this research, we did not affirm that the augmented number of reviews about a single attraction caused the increased number of visitors and tourists in Verona. We simply verified whether the two sets of big data related to each other by showing the same tourist flow dynamic.

To verify this, we used the temporal (monthly) trends of the reviews, divided into domestic and foreign users (based on the home location and language of each review) in 2022, and compared those trends with the number of Italian and foreign visitors (i.e. day visitors and tourists) using TELCO data retrieved from Vodafone Italy. These data reported the presence of people who visited Verona for at least two hours, based on an active SIM, clustered into Italian visitors (Italian SIM cards) and foreigners (SIM cards from other countries connected to the network by roaming). The final value regarding the presence of people in a particular territory (in this case, Verona) was developed by the provider, applying calibration and inference algorithms to represent not only the provider's active sims (or non-Italian sims connected to the Vodafone network) but also an estimation of the total number of users (Cavallo et al., 2022). In this research, day visitors (visiting Verona for one day but without an overnight stay) and tourists were considered (Italians and foreigners visiting Verona during the day and the night and sleeping in the destination). Other users, such as residents and commuters, were excluded from this analysis, assuming their small impact on the total number of reviews about the city where they lived and worked. Figure 2 shows the monthly levels of people's presence in Verona, dividing the users into domestic and international groups. These values represent day visitors and tourists together and comprise the total number of visits to the destination.

We decided to use this data source instead of the official dataset because tourist arrivals based on TELCO data also include day trippers who do not sleep at a destination, which is an insight that is usually missing in official statistics (Bertocchi et al., 2021).

Moreover, the correlation between the monthly numbers of reviews and visitors' presence in Verona indicated a solid value of 0.7624, reaching a higher value (0.94525) when we compared only the foreigners' reviews and presence. This indicated similar seasonality in both visits and reviews, without assuming any causality between the two phenomena.

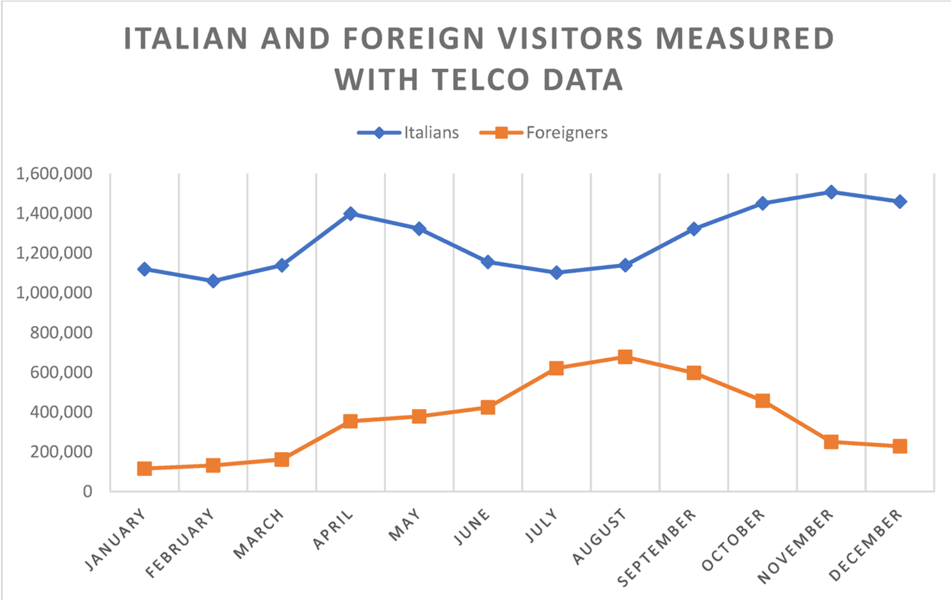

Figure 2. Monthly levels of visitors to Verona, divided into Italian and international groups (2022).

Network structure and patterns of real behaviours

To analyse the social network created by visitors who posted UGC, the collected data were conveyed in a matrix form by combining the user profiles with the tourist attractions reviewed, obtaining a first bimodal network (i.e. crossing visitors and destination attractions), which illustrated the user behaviours in the points of interest in the destination. This matrix was subsequently converted into a single-mode form that depicted only the relations between the attractions. Assuming that a user who reviews two places has visited both, the monomodal matrix lists the attractions in both rows and columns, indicating in each cell the number of visits that can be associated with each user. Hence, the network of visits is ultimately represented by attractions abstracted as nodes, and the tourist flows between the attractions are shown as arcs between the nodes. Not knowing the direction of the flows formally, we depict such a network as a bidirectional graph $G := (V, E)$, with the nodes |V| and edges |E| being the sets of attractions and links between the nodes, respectively. The arcs are weighted, based on the number of users, who have each made a visit between two considered nodes. This weighted one-mode matrix can be converted into an unweighted binary network matrix that presents the value of 1 at a connection between the nodes; otherwise, it presents the value of 0. The former provides a better understanding of the network's structure and density, while the latter provides an overview of the aggregate tourism behaviour that creates the network. For the purposes of analysing and displaying the graph underlying the network, a threshold value can be introduced for the minimum number of visits with which a connection between two nodes must be provided. Our analysis considered scenarios with the minimum number of interactions between two nodes within the range of 5–20, and we focused on an intermediate minimum of 10. The visitors' behaviours, expressed by interactions between two or more attractions, were calculated within a period of two weeks from the review. This was done to restrict the timeframe of the analysis by considering each holiday experience and no other types of behaviours as monthly or seasonal repetitions of the visit.

The fundamental metrics used to analyse the characteristics of the Verona tourists' behaviour network are as follows:

- *Node degree*. It considers the sum of the arcs passing through a node, weighted by the number of visits of each. It also considers the number of times that a general user undertakes a connection between two nodes.
- *Degree centrality*. It is the fraction of nodes to which a node is connected, given the totality of existing nodes in the network. Degree centrality values are normalised and defined as follows:

$$C_D(v_i) = \frac{k_i}{n-1},$$

- where n = |V| denotes the number of nodes in G, and n–1 represents the largest possible degree. The degree centrality reflects the central position of a single tourist attraction in the overall network. The higher the degree centrality, the more connections a tourist attraction has to other tourist attractions, and the more central the tourist attraction is. In real behaviour terms, the degree centrality can explain the number of connections of each individual, depicting the perceived significant points of interest and what visitors consider the DNA of the destination. Each visit is based on nodes with a higher score of degree centrality.
- *Betweenness centrality*. It is the number of times that a node is on the shortest path between two other nodes in the graph. The betweenness centrality of node v is the sum of the fraction of all pairs'
shortest paths that pass through v and are defined as follows:

$$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t \mid v)}{\sigma(s,t)}$$

- where (s,t) signifies the number of shortest (s,t) paths, and (v) denotes the number of those paths passing through some node v other than (s,t). If $s = t$, $(s,t) = 1$, and if $v \in s,t,(v) = 0$. An attraction with a high centrality of betweenness is imperative as it appears to pass through many of the shortest paths in the graph and identifies places that act as bridges between various parts of the graph. In real behaviour terms, the betweenness centrality can measure how an individual connects to others in the destination complex system. The visitor's behaviour, expressed with this value, demonstrates the role of some nodes with higher betweenness centrality as a connection between several places and typologies of attractions.
- *Eigenvector centrality*. It considers the position of the node in the graph and the degree of centrality of the nodes to which it is connected. The *ith* element of vector x, formed by the equation $Ax = \lambda x$, where A is the adjacency matrix of graph G with eigenvalue λ, is the Eigenvector centrality for node i. The Perron-Frobenius theorem states that if λ is the largest eigenvalue of adjacency matrix A, then there exists a singular solution x, whose elements are all positive. A node with high Eigenvector centrality is connected to several other essential nodes, which identify the attractions that have several connections to other important attractions. In real behaviour pattern detection, the Eigenvector measure shows which attraction/typology node is connected to other important nodes in the network. Eigenvector centrality can be used to identify influential points of interest, as well as communities or clusters of attractions visited together that are tightly knit but relatively disconnected from the rest of the network.

Detection and measurement of sustainable real behaviour patterns over time

Locations and typologies of tourist activities were selected as discriminating criteria for extracting patterns of sustainable real behaviours. Specifically, our goal was to evaluate the extent to which tourists' behaviours were associated with actual visits to other places as alternatives to major tourist points of interest, contributing to alleviating the costs generated by the high congestion that distinguishes the city of Verona. To achieve this goal, two types of flows were considered: one that was generated in areas of Verona that were usually subject to high tourist pressure (Type A), and the other outside these areas (Type B). We therefore distinguished between tourist attractions (nodes) located in each area, as well as the flows (arcs) that were exclusive to Areas A and B, or both. We also evaluated whether tourists' behaviours were linked to the participation in sustainable tourism activities that contributed to the diversification of tourism products. In line with Verona's strategic objectives, we considered the set of attractions (within the network) related to experiential tourism (Tours and Activities) as a benchmark to evaluate the promotion of sustainable cross-sector diversification in the tourism industry. This approach of identification of sustainability criteria reflects the UNWTO and UNDP (2017), UNWTO UNWTO, UNWTO, (2022) routes for the development of sustainable tourism.

We assessed the importance and characteristics of sustainable real behaviour patterns through an in-depth analysis of the metrics of our network, adapting them to the introduced sustainability criteria. To this end, our methodology introduced the metrics of *dispersion centrality* and *diversification centrality*, as follows:

- *Dispersion centrality*. This was intended as a measure of Eigenvector centrality, which calculates the centrality for a node, based on the centrality and nature of its neighbouring nodes, giving greater importance to the nodes located in Area B. It can be obtained by adding weight w_i of importance to each node v_j in Area B that is greater than that in Area A. In our case, we assumed that $w_i = 1\, if\, v_j \in A$; $w_i = 2\, if\, v_j \in B$.
- *Diversification centrality*. This was intended as a measure of Eigenvector centrality, which calculates the centrality for a node, based on the centrality of its neighbours and the nature of the nodes, giving greater weight to nodes v_j located in Area B and connected to experiential tourism (B-exp) (i.e. v_j belongs to the Tour and Activities group). In our case, we assumed that $w_i = 1\, if\, v_j \in A$; $w_i = 2\, if\, v_j \in B$; $w_i = 3\, if\, v_j \in B\text{-}exp$.

Based on these criteria and metrics, we extracted and evaluated the nodes within our graph, distinguishing them by type, as well as the generated arcs between them in all their combinations according to the dataset. Finally, we also distinguished and extrapolated the attractions of Type B that belonged only to the Tours and Activities group. This distinction was made on the structure of the network of Verona attractions (created until the year 2022) as a whole, regardless of the type of tourist, as well as between domestic and foreign tourists.

We then performed an in-depth analysis of the evolution and transformation of sustainable real behaviour patterns over time, focusing on the structure assumed by the network from year to year, from 2013 to 2022. Specific metrics for the longitudinal comparisons were based on the use of *similarity matrices,* defined as the intersection of the arcs of two graphs (e.g. only that of the years 2017 and 2018), divided by the union of these arcs. Specific details of the nature of these changes were observed through measurements of the annual variations in the numbers and degrees of the nodes, distinguishing between their locations and typologies.

Results on the network structure and patterns of visitors' behaviours in Verona

We built a network of nodes and edges representing the most common visitor behaviour patterns in Verona considering at least ten connections between an attraction and another. Three different behaviours represented by the network were developed: a. overall tourists' behaviours (using all languages of the reviews and nationalities of the reviewers to show a comprehensive set of visitors' behaviours); b. domestic tourists' behaviours (using only the Italian text and Italian reviewers); c. foreign tourists' behaviours (using all other languages and nationalities reviewers; Figure 3).

As the complete network is the sum of all possible behaviours in the destination, the structure has a stronger density, with a high number of connections and nodes, namely 145 nodes equalling 56 Cultural Sites, 45 Tours and Activities, 37 Landmarks and Sites of Interest, 3 Shopping Facilities and one each of Bars and Clubs, Entertainment and Events, Natural Sites, and Transport and Services (see Supplementary Material *A* for the detailed dataset). The recognisable sub-networks (Figure 3(A)) are present in the centre of the network (black circle) and represent the most important attractions of the destination. The figure also shows a higher degree centrality value and a small network outside the nest of the edges, represented by several different tours and experiences, each connected to a particular node with a high value of betweenness (named Way Tours).

Considering only the network of domestic flows (comprising 79 nodes, distributed among 38 Cultural Sites, 28 Landmarks and Sites of Interest, 7 Tours and Activities, 3 Shopping Facilities and one each of Bars and Clubs, Entertainment and Events, and Natural Sites), it can be noticed that it is less dense and has a similar centre sub-network of the entire behaviour, depicting the main attractions of the destination as having been visited by both Italian and foreign tourists. This means that for all Italian clusters, such attractions represent the main motivations for visiting Verona.

Those nodes are also visible in the third network of international visitors' behaviours, which considerably varies from the previous two networks in terms of diversification of the types of attractions. The international visitors' behaviours, represented by the network, consist of 104 nodes, distributed among 37 Cultural Sites, 40 Tours and Activities, and 27 Landmarks and Sites of Interest, depicting a significantly higher number of experiences (tours and activities) compared with the domestic visitors' behaviours. This is also confirmed by a sub-network (in light blue at the bottom right of the network) that shows the same structure that we already found in the entire users' network (see Supplementary Material *B* for a review of the distinction between Italian and international tourists).

Another similarity to the domestic network, but strongly visible in the international network, is the red constellation of tours and activities (experienced together by visitors) that were still not connected to the main attractions and the most important nodes of the destination. Those activities often occurred in rural areas outside the city centre and referred to varying tourism typologies, such as gastronomy, wine tourism and nature-based tourism (Figure 3(A)).

Delving into further details about the differences, we made a list of the top ten attractions within the city centre (Type A) and outside it (Type B), showing degree, betweenness and diversification centrality values using the ranking of the total number of reviews as a filter. These values were divided into three sections in Table 1, which illustrates the structure of the three networks presented earlier. In Figures 3(B,C), we use heat maps as a representation of the density of attractions in Verona and their dispersion centrality.

The following results depict the most reviewed attractions: Arena di Verona, Juliette's House, Piazza delle Erbe and Piazza Bra. All these attractions are located in the city centre and have higher scores in degree/betweenness/diversification centrality for all types of networks, considering only Type A.

Regarding the points of interest outside the centre, Romeo's House plays a key role in determining the degree and diversification centrality values since the reviews indicate lower satisfaction with such points of interest, as expressed in an average score of 3 out of 5 stars in TripAdvisor. As previously observed, Way Tours' point of interest, located outside the city centre, has a higher centrality value (0.12 for the entire network and 0.14 for the foreigners'

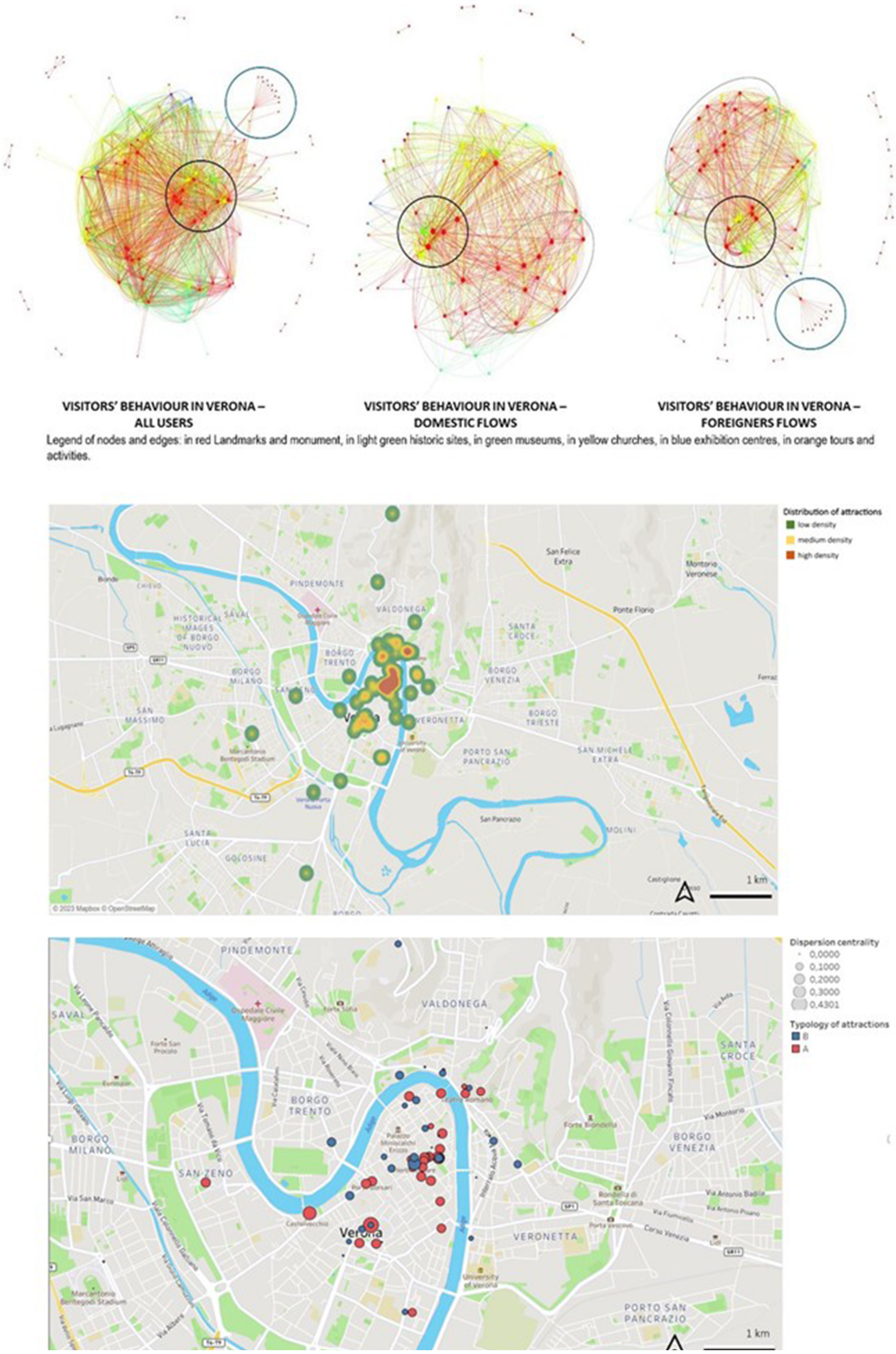

Figure 3. (A). The network between attractions representing the behaviours of all visitors and domestic and foreign tourists in Verona. (B) Heat map of the tourist attractions density in Verona. (C). Heat map of Verona attractions (Type A and Type B) weighted by their dispersion centrality.

Table 1. Social network analysis measure to detect real behaviour patterns between tourist attractions in the city centre and outside it.

All visitors' nodes measures - top ten A and B attractions by number of reviews							
Name	Macro	Type	Degree Centrality	Betweenness Centrality	Diversification Centrality	N. Reviews	Score
Arena di Verona	Cultural sites	A	0,61805556	0,09097653	0,14682523	18542	4.5
Casa di Giulietta	Cultural sites	A	0,59722222	0,04213205	0,14529945	12084	3.5
Piazza delle Erbe	Landmarks and sites of interest	A	0,63888889	0,07030523	0,15012602	9355	4.5
Piazza Bra'	Landmarks and sites of interest	A	0,59027778	0,03361083	0,1479044	5821	4.5
Museo di Castelvecchio	Cultural sites	A	0,49305556	0,01970184	0,1294816	3453	4.5
Ponte Scaligero	Landmarks and sites of interest	A	0,52777778	0,02134497	0,13945143	3202	4.5
Basilica di San Zeno Maggiore	Cultural sites	A	0,50694444	0,0187071	0,13150375	2957	4.5
Torre dei Lamberti	Landmarks and sites of interest	A	0,52777778	0,02067947	0,14032625	2888	4.5
Ponte Pietra	Landmarks and sites of interest	A	0,54166667	0,01224905	0,13878602	2588	4.5
Basilica di Santa Anastasia	Cultural sites	A	0,54861111	0,02255769	0,14364926	2569	4.5
Name	**Macro**	**Type**	**Degree Centrality**	**Betweenness Centrality**	**Diversification Centrality**	**N. Reviews**	**Score**
Ways Tours	Tour and activities	B	0,14583333	0,1207265	0,0511592	1439	5.0
Giardino Giusti	Cultural sites	B	0,29861111	0,00360297	0,12078312	1110	4.5
Verona Porta Nuova Railway Station	Cultural sites	B	0,29166667	0,00012223	0,12863559	520	3.5
Pagus Wine Tours	Tour and activities	B	0,04166667	0,01146076	0,02562847	491	5.0
Centro Commerciale Adigeo	Shopping	B	0,04861111	0	0,02380375	441	4.0
Romeo's House (Casa di Romeo)	Landmarks and sites of interest	B	0,39583333	0,00128701	0,16517514	430	3.0
Free Walking Tour Verona	Tour and activities	B	0,00694444	0	0,00522228	354	4.5
TENUTA SANTA MARIA VALVERDE	Tour and activities	B	0,03472222	0,00097125	0	311	5.0
Explore Verona	Tour and activities	B	0,02083333	0,00029138	0	283	5.0
Colors of Italy	Tour and activities	B	0,02083333	0,01146076	0,01057535	282	5.0

network), which is even higher than those of the main attractions located in the historical area. While this point is important for foreign tourists, Italian visitors do not show this pattern. Overall, only two experiences outside the city centre, namely Tenuta Santa Maria Valverde (Valpolicella Vineyards) and Explore Verona are visited the most by domestic tourists.

Results on the sustainable behaviour patterns in Verona over time

This step of the analysis investigated the evolution and transformation of sustainable behaviour patterns over time by comparing the structural variations of the network of attractions visited each year, from 2013 to 2022. The findings indicate a significant change in the network composition after 2019, particularly during and after the COVID-19 pandemic. This change is confirmed by the Similarity Matrix A (Figure 4), which demonstrates stronger similarities in tourists' behaviours in terms of choosing attractions between 2015–2019 and 2020–2022. Moreover, a

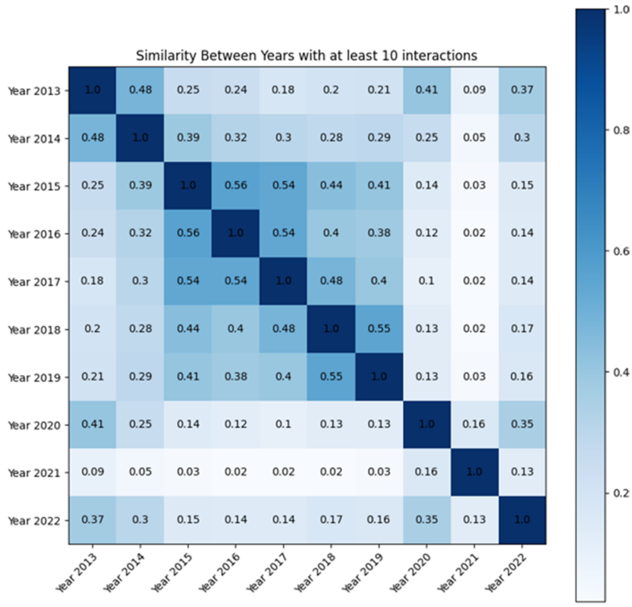

Figure 4. Behaviour similarity over the years.

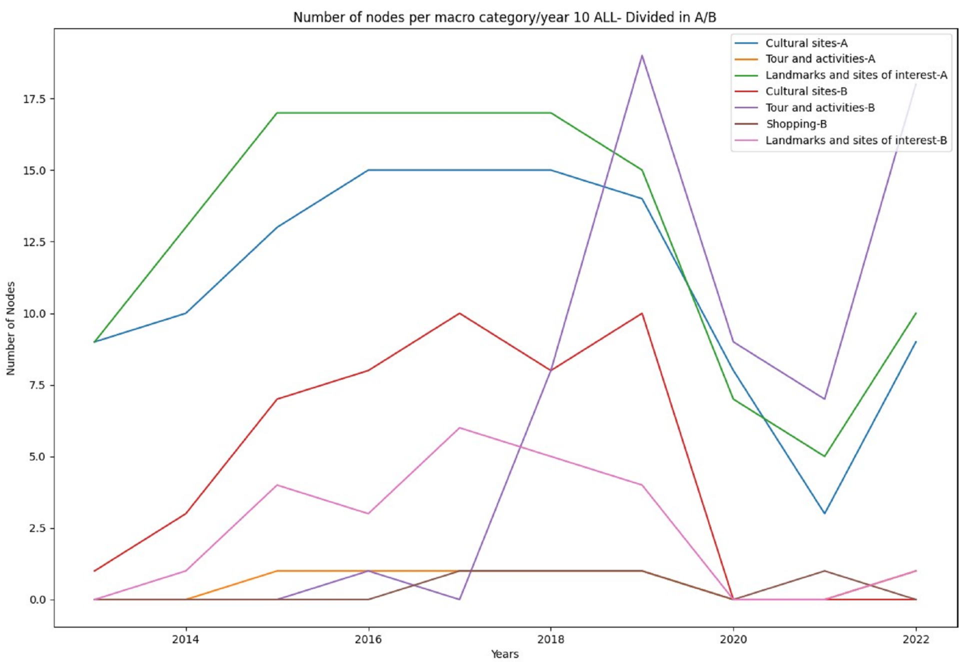

Figure 5. Number of nodes per macro category, divided into Type A and Type B. Trends over the years.

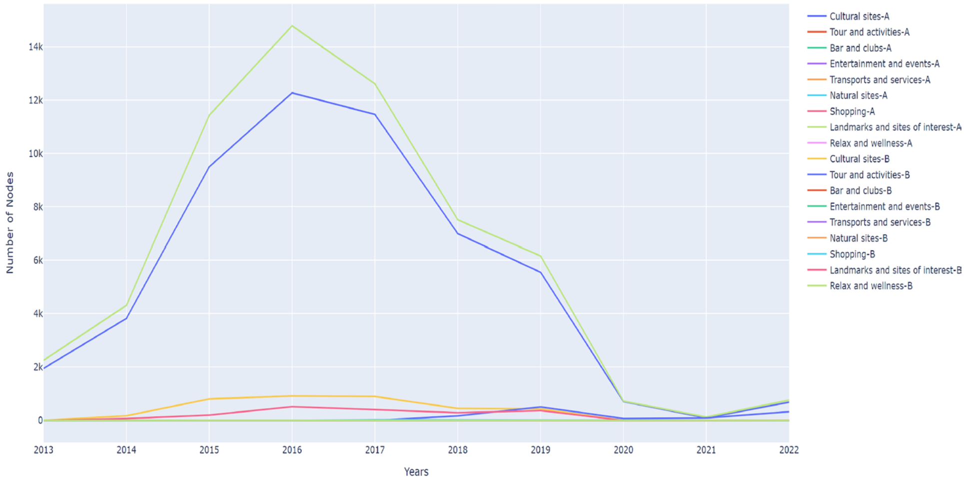

Figure 6. Degree value per number of nodes, per macro category, divided into Type A and Type B. Trends over the years.

fundamental difference is observed between these two sets, with a lack of homogeneity in the comparison of each year after 2019, with respect to the previous years.

To better understand this transformation, we focused on the changes in the number and diversity of nodes, distinguishing the types of attractions and the distinction between Types A and B. Figures 5 and 6 illustrate an inversion of behaviour patterns that reaches its peak during and after the pandemic period, with substantial growth in experiential tourism. Attractions under "Tours and Activities" (Type B) grow while other tourist attractions (Type A) decrease, especially for foreign tourists. In other words, over the last few years, tourists have demonstrated a higher tendency to experience the destination in a different way, appreciating experiences outside the most frequent attractions. This behaviour is in line with those sustainable behaviours that enhance experiencing local culture, such as discovering Valpolicella Vineyards, wine tasting, joining Italian risotto recipes and pasta-cooking classes.

The shift towards experiential tourism outside tourist-dense areas can be further analysed in terms of the types of users. Specifically, we can distinguish the changes in the behaviours of domestic and foreign tourists over time by employing similarity matrices that compare Italians' and foreigners' visiting behaviours (nodes A and B, Figures 7 and 8). These matrices confirm the substantial structural changes in the network that occurred after 2019 and highlight the differences in visiting behaviours between Italian and foreign users for both types of nodes. Furthermore, the evaluation of the variations in the numbers of nodes over the years, distinguished by the type of attraction favoured by each group of Italian and foreign users (see Supplementary Material *B*), underscores the way experiential tourism involving Type B attractions has increased after 2019 for both types of users—but more for foreign tourists. For instance, foreign tourists appreciate experiences such as cooking classes and bike tours of the destination, while domestic tourists prefer wine-tasting experiences.

Discussion and implications

The main objective of this study is to provide a methodological contribution to the field of tourism, with a specific focus on sustainable tourism. In pursuit of this goal, our study presents an insightful analysis of big data, utilizing the SNA approach in the context of overtourism. Since overtourism involves tourism practices that are not environmentally or economically

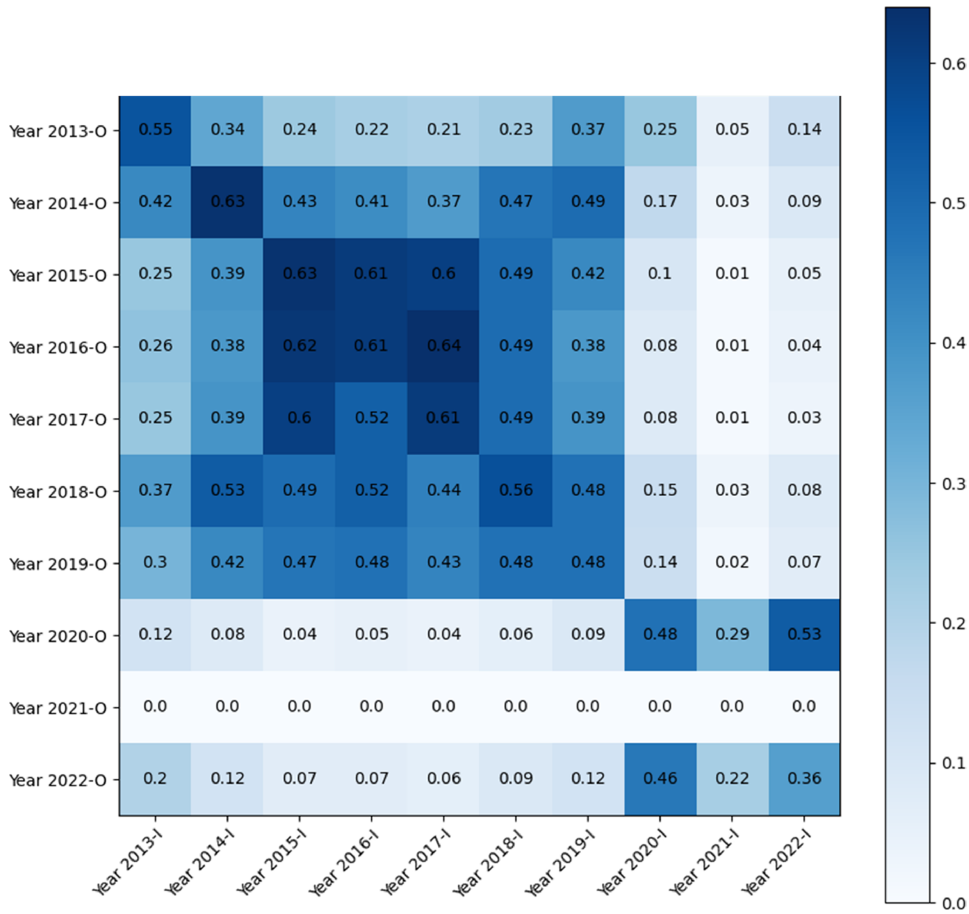

Figure 7. Similarity between Italians and foreigners within attractions in the city centre over the years.

sustainable (Mihalic & Kuščer, 2021), this paper contributes to the field of tourist data applied for sustainable tourism by tracking and monitoring tourist flows. Through the proposed framework, our research has successfully identified and quantified actual sustainable behaviours exhibited by tourists in the selected travel destination, namely Verona, Italy.

Our contribution responds to the need for research methodologies that overcome limitations in studies reliant on self-reported data, particularly within sustainable tourism. Our approach aims to mitigate the intention–behaviour gap challenges (Viglia & Acuti, 2022).

Unlike previous methods (Agrawal et al., 2022; Xu et al., 2020), our study advocates for a dynamic approach to identify and track tourists' real behaviours. It introduces two lenses—spatial and temporal—to understand sustainable tourism. The spatial lens detects chosen tourist attractions, their relations, and how tourists move across them, revealing behavioural patterns. The temporal lens enables longitudinal comparisons of real behaviour patterns over time, facilitating predictions and monitoring the effects of managerial decisions or events (Li et al., 2021; Li & Law, 2020; Valeri & Baggio, 2021).

Applying SNA to trace sustainable real behaviours represents a novel tentative approach extending previous tourism literature (e.g. Hernández et al., 2018; Seok et al., 2021; Van der Zee et al., 2020).

Moreover, our method scientifically validates the use of UGC to describe tourists' real behaviours, overcoming eventual criticisms, in combination with TELCO data (Ma & Kirilenko, 2021).

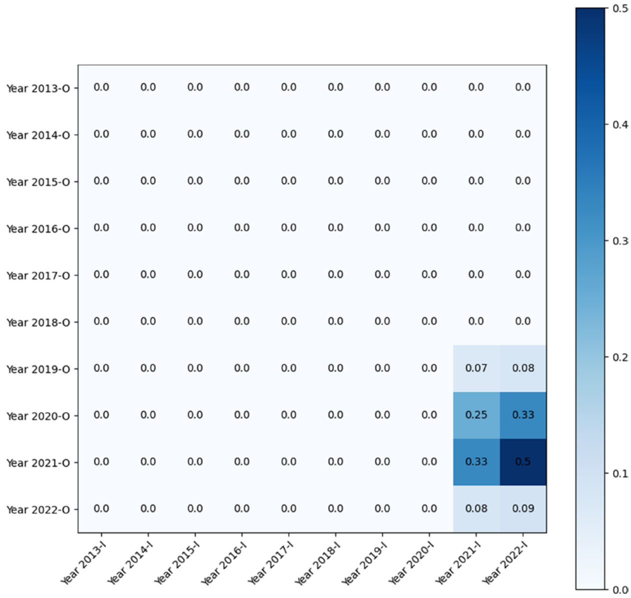

Figure 8. Similarity between Italians and foreigners within attractions outside the city centre.

From a managerial standpoint, our study provides a methodological approach for destination managers to control and interpret sustainable real behaviours. Offering a framework for monitoring tourists' flows and developing solutions for sustainable destination management, it addresses the lack of control over overtourism policies (Butler & Dodds, 2022). The method facilitates optimizing mobility flows and revising destination offerings. It allows capturing the evolution of tours and attractions visited over time, providing insights into the impacts of public policies on tourists' behaviours and choices.

Such a method can be generalised and adopted by scholars and local authorities interested in tracking and monitoring tourists' behaviours. Following the research design's analytical framework, as suggested in Figure 1, the guidelines on checklist phases (outlined below) can be useful for DMOs, researchers and private organisations that are interested in monitoring and evaluating tourists' sustainable real behaviours.

The first phase requires collecting and storing multiple sources of big data related to a destination. In our case, we collected data for the city of Verona and entered them in two datasets, one for structured data (TELCO data) and the other for unstructured data (UGC).

The second phase involves the analysis of the collected data. We employed SNA techniques to create the network structure of Verona, with its points of interest (i.e. nodes) and their relations. Subsequently, sustainability criteria were selected to detect and measure sustainable real behaviour patterns. According to the guidelines for sustainable behaviours, as presented by the UNWTO and UNDP (2017), UNWTO UNWTO, UNWTO, (2022), sustainable real behaviours were measured using the metrics of dispersion centrality and diversification centrality. As these measures allow the understanding of flow distribution and visited attractions, they offer a useful tool to control the effectiveness of overtourism mitigation policies (Butler & Dodds, 2022). To decongest high tourist areas, these trends, once detected, can be promoted through ad hoc campaigns by the municipalities. The identification of behaviours that reflect a growing interest in new diversified activities and attractions, such as experiential tourism or the discovery of local products, can be explained by the creation of new business opportunities that support diversification and the promotion of local cultures and traditions.

Lastly, tourists' behaviours over time are compared. In this regard, we described the evolution of real behaviour patterns over time, capturing the existence of any sustainable transformation of tourists' choices. We compared the network structures over a decade, considering two levels of analysis. At the first level of analysis, we tried to identify the degree of difference in the structures across the years through correlation matrices. The correlation matrices demonstrated their similarity in the beginning of the period, while in more recent years (2021–2022), the matrices have varied. Policymakers should conduct an in-depth analysis of such shifts in behaviours since these can be derived from different factors; one may be related to the COVID-19 pandemic, which changed tourists' perceived degrees of risk and fear when visiting crowded areas (Czarnecki et al., 2023).

Our study's findings enable the city of Verona to revise its offerings and the communication and promotion of its attractions. For instance, our data analysis reveals that most of the tourists visit two main attractions, namely the Arena amphitheatre and Juliet's House, while only a few tourists visit other significant yet less known nearby attractions, such as Castelvecchio and its museum. Conducting ad hoc communication activities, increasing the awareness of less popular attractions, as well as those that are not concentrated in the city centre, combined with a more tailored ticketing system, can offer tourists with bundle-pricing tickets that can re-direct tourists' flows and lead them to visit less popular attractions in combination with the more visited ones. This can better spread the mobility flows in the city centre, adjusting the journey across Verona.

The increasing interest in Type B attractions since 2019 suggests that further experiences can be recommended to promote the local culture and traditions. This is intended to preserve the identity of the destination by focusing on both the environmental and the social sustainability of tourism (e.g. cooking class that features traditional recipes, taught by residents of Verona).

Limitations and venues for future research

While the content of this empirical research is drawn from big data on tourism practices, our work is limited by the analysis of the data obtained for a single destination. Further research could reinforce the validity of our model by applying it to multiple destinations and choosing different sustainability criteria (e.g. decongest certain areas, promote the visibility of areas), according to each destination's strategic objectives related to sustainability and other drivers of overtourism (e.g. duration of the visit, period of the arrival). The reason is that redistribution of flows is only one measure to mitigate overtourism and not the only effective one (Mihalic, 2020). Moreover, as we considered only two segments of tourists, further research could adopt SNA to compare sustainable behaviours considering various segmentation criteria, such as socio-demographic variables (e.g. gender, age) or geographic variables (e.g. country of origin) (Hernández et al., 2018).

Finally, despite the strengths of our adopted methodology, this approach may raise some concerns. First, big data and related analyses can describe a phenomenon but cannot explain the reason for it; in other words, they focus on connectivity rather than on causality (Xu et al., 2020). Mixed-method studies could further enrich our suggested approach *via* integration with qualitative methods to better explain the causal mechanisms behind the shifts in tourists' behaviours towards more or less sustainable choices and behaviours.

Another issue concerns the ethical code of conduct related to the use of big data (Boyd, 2010; Nunan & Di Domenico, 2013). Although our methodological protocol met the code of conduct requirements obtained from suppliers under the terms of a confidentiality agreement, more research is needed for a better understanding of how to use big data in the quest for ethical data management practices. This would require the tourism and hospitality industry to opt for updated, relevant and centralised ethical standards for data privacy and security (Yallop & Seraphin, 2020).

Acknowledgments

We extend our deepest gratitude to Editors and Guest Editors and the team of reviewers for their invaluable reviews and timely support, which greatly enhanced the refinement and quality of this work. Lastly, but certainly, not the least, we would like to thank Dr. Silvia Blasi for her support in crafting the final version of this manuscript.

Disclosure statement

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

Funding

This study was carried out within the PNRR research activities of the consortium iNEST (Interconnected North-Est Innovation Ecosystem) funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.5 – D.D. 1058 23/06/2022, ECS_00000043). This manuscript reflects only the Authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

ORCID

Ilenia Confente http://orcid.org/0000-0003-2970-0206
Valentina Mazzoli http://orcid.org/0000-0002-8344-0356
Nicola Camatti http://orcid.org/0000-0002-1899-2728
Dario Bertocchi http://orcid.org/0000-0002-8973-4558

References

Agag, G., Brown, A., Hassanein, A., & Shaalan, A. (2020). Decoding travellers' willingness to pay more for green travel products: Closing the intention–behaviour gap. *Journal of Sustainable Tourism*, *28*(10), 1551–1575. https://doi.org/10.1080/09669582.2020.1745215

Agrawal, R., Wankhede, V. A., Kumar, A., Luthra, S., & Huisingh, D. (2022). Big data analytics and sustainable tourism: A comprehensive review and network-based analysis for potential future research. *International Journal of Information Management Data Insights*, *2*(2), 100122. https://doi.org/10.1016/j.jjimei.2022.100122

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, *52*(1), 27–58. https://doi.org/10.1146/annurev.psych.52.1.27

Baggio, R. (2008). Symptoms of complexity in a tourism system. *Tourism Analysis*, *13*(1), 1–20. https://doi.org/10.3727/108354208784548797

Baggio, R. (2017). Network science and tourism–the state of the art. *Tourism Review*, *72*(1), 120–131. https://doi.org/10.1108/TR-01-2017-0008

Baggio, R., Scott, N., & Cooper, C. (2010). Network science: A review focused on tourism. *Annals of Tourism Research*, *37*(3), 802–827. https://doi.org/10.1016/j.annals.2010.02.008

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, *46*(4), 557–590. https://doi.org/10.1007/s11747-018-0581-x

Bertocchi, D., Camatti, N., & van der Borg, J. (2021). Tourism Peaks on the Three Peaks. Using big data to monitor where, when and how many visitors impact the Dolomites UNESCO World Heritage Site. *Rivista geografica italiana.*, (2021/3).

Boyd, D. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. In *A networked self* (pp. 47–66). Routledge.

Bramwell, B., Higham, J., Lane, B., & Miller, G. (2017). Twenty-five years of sustainable tourism and the Journal of Sustainable Tourism: Looking back and moving forward. *Journal of Sustainable Tourism*, *25*(1), 1–9. https://doi.org/10.1080/09669582.2017.1251689

Brave, S. A., Butters, R. A., & Fogarty, M. (2022). The perils of working with big data, and a SMALL checklist you can use to recognize them. *Business Horizons*, *65*(4), 481–492. https://doi.org/10.1016/j.bushor.2021.06.004

Butler, R. W., & Dodds, R. (2022). Overcoming overtourism: A review of failure. *Tourism Review*, *77*(1), 35–53. https://doi.org/10.1108/TR-04-2021-0215

Camatti, N., Bertocchi, D., Carić, H., & van der Borg, J. (2020). A digital response system to mitigate overtourism. The case of Dubrovnik. *Journal of Travel & Tourism Marketing*, *37*(8-9), 887–901. https://doi.org/10.1080/10548408.2020.1828230

Cavallo, L., Cerasti, E., Di Torrice, M., Righi, A., Santoro, M. T., Tuoto, T., Valentino, L., Di Sorte, D., Rossi, M., Zaramella, A., Bertocchi, D., Mantegari, G., & Zamengo, B. (2022). Exploring mobile network data for tourism statistics: The collaboration between Istat and Vodafone Business Italia. *Rivista Di Statistica Ufficiale*, *3*, 43–76.

Coldwell, W. (2017). First Venice and Barcelona: Now Anti-tourism Marches Spread across Europe. *The Guardian*, August 10. https://static1.squarespace.com/ static/53109b11e4b05040160f0a8f/t/59bfcc8b90badeccf

7cc1564/1505741995422/First±Venice±and±Barcelona_±n ow±anti-tourism±marches±spread±across±Europe±_±Tr avel±_±The±Guardian.pdf
Confente, I., & Scarpi, D. (2021). Achieving environmentally responsible behavior for tourists and residents: A norm activation theory perspective. *Journal of Travel Research*, *60*(6), 1196–1212. https://doi.org/10.1177/0047287520938875
Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. https://doi.org/10.1037/h0047358
Czarnecki, A., Dacko, A., & Dacko, M. (2023). Changes in mobility patterns and the switching roles of second homes as a result of the first wave of COVID-19. *Journal of Sustainable Tourism*, *31*(1), 149–167. https://doi.org/10.1080/09669582.2021.2006201
D'Angelo, N., Abbruzzo, A., Ferrante, M., Adelfio, G., & Chiodi, M. (2023). GPS data on tourists: A spatial analysis on road networks. *AStA Advances in Statistical Analysis*, 1–23. https://doi.org/10.1007/s10182-023-00484-w
Demeter, C., Fechner, D., & Dolnicar, S. (2023). Progress in field experimentation for environmentally sustainable tourism – A knowledge map and research agenda. *Tourism Management*, *94*, 104633. https://doi.org/10.1016/j.tourman.2022.104633
Dolnicar, S., Knezevic Cvelbar, L., & Grün, B. (2017). Do pro-environmental appeals trigger pro-environmental behavior in hotel guests? *Journal of Travel Research*, *56*(8), 988–997. https://doi.org/10.1177/0047287516678089
Eberendu, A. C, Madonna University. (2016). Unstructured Data: An overview of the data of Big Data. *International Journal of Computer Trends and Technology*, *38*(1), 46–50. https://doi.org/10.14445/22312803/IJCTT-V38P109
Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.
Gowreesunkar, V. G., & Vo Thanh, T. (2020). Between overtourism and under-tourism: Impacts, implications, and probable solutions. *Overtourism: Causes, implications and solutions*., 45–68.
Hagberg, A., Swart, P., & Chult, D. S. (2008). *Exploring network structure, dynamics, and function using NetworkX (No. LA-UR-08-05495; LA-UR-08-5495)*. Los Alamos National Lab (LANL).
Hernández, J. M., Kirilenko, A. P., & Stepchenkova, S. (2018). Network approach to tourist segmentation via user generated content. *Annals of Tourism Research*, *73*, 35–47. https://doi.org/10.1016/j.annals.2018.09.002
Hibbert, J. F., Dickinson, J. E., Gössling, S., & Curtin, S. (2013). Identity and Tourism Mobility: An Exploration of the Attitude-Behavior Gap. *Journal of Sustainable Tourism*, *21*(7), 999–1016. https://doi.org/10.1080/09669582.2013.826232
Holmes, M. R., Dodds, R., & Frochot, I. (2021). At home or abroad, does our behavior change? Examining how everyday behavior influences sustainable travel behavior and tourist clusters. *Journal of Travel Research*, *60*(1), 102–116. https://doi.org/10.1177/0047287519894070
Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality & Social Psychology Bulletin*, *30*(2), 161–172. https://doi.org/10.1177/0146167203259930
Juvan, E., & Dolnicar, S. (2014). The attitude-behaviour gap in sustainable tourism. *Annals of Tourism Research*, *48*, 76–95. https://doi.org/10.1016/j.annals.2014.05.012
Juvan, E., & Dolnicar, S. (2016). Measuring environmentally sustainable tourist behaviour. *Annals of Tourism Research*, *59*, 30–44. https://doi.org/10.1016/j.annals.2016.03.006
Kádár, B., & Gede, M. (2021). Tourism flows in large-scale destination systems. *Annals of Tourism Research*, *87*, 103113. https://doi.org/10.1016/j.annals.2020.103113
Kang, S., Kim, W. G., & Park, D. (2021). Understanding tourist information search behaviour: The power and insight of social network analysis. *Current Issues in Tourism*, *24*(3), 403–423. https://doi.org/10.1080/13683500.2020.1771290
Kitchin, R. (2013). Big data and human geography opportunities, challenges and risks. *Dialogues in Human Geography*, *3*(3), 262–267. https://doi.org/10.1177/2043820613513388
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, X., & Law, R. (2020). Network analysis of big data research in tourism. *Tourism Management Perspectives*, *33*, 100608. https://doi.org/10.1016/j.tmp.2019.100608
Li, Y., Xie, J., Gao, X., & Law, A. (2021). A Method of selecting potential development regions based on GPS and social network models–from the perspective of tourist behavior. *Asia Pacific Journal of Tourism Research*, *26*(2), 183–199. https://doi.org/10.1080/10941665.2018.1515092
Ma, S., & Kirilenko, A. (2021). How reliable is social media data? Validation of TripAdvisor tourism visitations using independent data sources. In *Information and Communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 eTourism Conference*.
Mariani, M., & Borghi, M. (2021). Customers' evaluation of mechanical artificial intelligence in hospitality services: A study using online reviews analytics. *International Journal of Contemporary Hospitality Management*, *33*(11), 3956–3976. https://doi.org/10.1108/IJCHM-06-2020-0622
Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: A systematic literature review. *International Journal of Contemporary Hospitality Management*, *30*(12), 3514–3554. https://doi.org/10.1108/IJCHM-07-2017-0461
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

Mazzara, L. (2023). Il boom dei turisti divide la città: «Ospiti insoddisfatti tra parcheggi impossibili e confusione, *L'Arena*. https://www.larena.it/territorio-veronese/citta/turisti-a-verona-regolamentare-gli-ingressi-divide-1.9731048
Mihalic, T. (2020). Conceptualising overtourism: A sustainability approach. *Annals of Tourism Research*, *84*, 103025. https://doi.org/10.1016/j.annals.2020.103025
Mihalic, T., & Kuščer, K. (2021). Can overtourism be managed? Destination management factors affecting residents' irritation and quality of life. *Tourism Review*, *77*(1), 16–34. https://doi.org/10.1108/TR-04-2020-0186
Mukai, T., & Ikeda, Y. (2022). Optimizing travel routes using temporal networks constructed from global positioning system data in kyoto tourism. *Frontiers in Physics*, *10*, 1001983. https://doi.org/10.3389/fphy.2022.1001983
Nunan, D., & Di Domenico, M. (2013). Market research and the ethics of big data. *International Journal of Market Research*, *55*(4), 505–520. https://doi.org/10.2501/IJMR-2013-015
Passafaro, P., Cini, F., Boi, L., D'Angelo, M., Heering, M. S., Luchetti, L., Mancini, A., Martemucci, V., Pacella, G., Patrizi, F., Sassu, F., & Triolo, M. (2015). The "sustainable tourist": Values, attitudes, and personality traits. *Tourism and Hospitality Research*, *15*(4), 225–239. https://doi.org/10.1177/1467358415576086
Piano Strategico Destinazione Verona. (2017). Curated by Ugolini, M.M. https://musei.regione.veneto.it/c/document_library/get_file?uuid=dcacccbf-7dc8-41e8-920e-79905a039988&groupId=10813
Provenzano, D., & Baggio, R. (2020). A complex network analysis of inbound tourism in Sicily. *International Journal of Tourism Research*, *22*(4), 391–402. https://doi.org/10.1002/jtr.2343
Sánchez-Franco, M. J., & Rey-Moreno, M. (2022). Do travelers' reviews depend on the destination? An analysis in coastal and urban peer-to-peer lodgings. *Psychology & Marketing*, *39*(2), 441–459. https://doi.org/10.1002/mar.21608
Savelli, E., Gregory-Smith, D., Murmura, F., & Pencarelli, T. (2022). How to communicate typical–local foods to improve food tourism attractiveness. *Psychology & Marketing*, *39*(7), 1350–1369. https://doi.org/10.1002/mar.21668
Scarpi, D., Confente, I., & Russo, I. (2022). The impact of tourism on residents' intention to stay. A qualitative comparative analysis. *Annals of Tourism Research*, *97*, 103472. https://doi.org/10.1016/j.annals.2022.103472
Serna, A., Gerrikagoitia, J. K., Bernabé, U., & Ruiz, T. (2017). Sustainability analysis on Urban Mobility based on Social Media content. *Transportation Research Procedia*, *24*, 1–8. https://doi.org/10.1016/j.trpro.2017.05.059
Serrano, L., Ariza-Montes, A., Nader, M., Sianes, A., & Law, R. (2021). Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings: A text mining approach. *Journal of Sustainable Tourism*, *29*(7), 1134–1152. https://doi.org/10.1080/09669582.2020.1838529
Seok, H., Barnett, G. A., & Nam, Y. (2021). A social network analysis of international tourism flow. *Quality & Quantity*, *55*(2), 419–439. https://doi.org/10.1007/s11135-020-01011-8
Sistema Statistico Regionale. (2023). Tourists flows in the Veneto. https://statistica.regione.veneto.it/banche_dati_economia_turismo.jsp
Song, Z., Xing, L., & Chathoth, P. K. (2015). The effects of festival impacts on support intentions based on residents' ratings of festival performance and satisfaction: A new integrative approach. *Journal of Sustainable Tourism*, *23*(2), 316–337. https://doi.org/10.1080/09669582.2014.957209
Taczanowska, K., González, L. M., Garcia-Massó, X., Muhar, A., Brandenburg, C., & Toca-Herrera, J. L. (2014). Evaluating the structure and use of hiking trails in recreational areas using a mixed GPS tracking and graph theory approach. *Applied Geography*, *55*, 184–192. https://doi.org/10.1016/j.apgeog.2014.09.011
United Nations (UN). (2021). *Agenda 21, UNCED 1992. Sustainable development goals knowledge platform*. https://sustainabledevelopment.un.org/outcomedocuments/agenda21
UNWTO & UNDP. (2017). Tourism and the sustainable development goals – Journey to 2030. UNWTO.
UNWTO. (2022). *A Year in Review*. https://www.unwto.org/unwto-2022-a-year-review
Valeri, M., & Baggio, R. (2021). Social network analysis: Organizational implications in tourism management. *International Journal of Organizational Analysis*, *29*(2), 342–353. https://doi.org/10.1108/IJOA-12-2019-1971
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
Van der Zee, E., Bertocchi, D., & Vanneste, D. (2020). Distribution of tourists within urban heritage destinations: A hot spot/cold spot analysis of TripAdvisor data as support for destination management. *Current Issues in Tourism*, *23*(2), 175–196. https://doi.org/10.1080/13683500.2018.1491955
Viglia, G., & Acuti, D. (2022). How to overcome the intention–behavior gap in sustainable tourism: Tourism agenda 2030 perspective article. *Tourism Review*, *78*(2), 321–325. (ahead-of-print). https://doi.org/10.1108/TR-07-2022-0326
Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, *46*, 222–232. https://doi.org/10.1016/j.tourman.2014.07.003
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, F., Nash, N., & Whitmarsh, L. (2020). Big data or small data? A methodological review of sustainable tourism. *Journal of Sustainable Tourism*, *28*(2), 144–163. https://doi.org/10.1080/09669582.2019.1631318
Yallop, A., & Seraphin, H. (2020). Big data and analytics in tourism and hospitality: Opportunities and risks. *Journal of Tourism Futures*, *6*(3), 257–262. https://doi.org/10.1108/JTF-10-2019-0108