

## **Understanding sustainable tourism behaviors: a comprehensive longitudinal study based on tourist digital footprint.**

This empirical study presents a longitudinal fieldwork (spanning a period of ten years) that investigates how tourists experience the attractions of a tourist destination by leveraging the digital footprint left during their journey. This research contributes to filling the literature gap on real sustainable behaviors in tourism, overcoming the limitations of traditional research methods that fail to address the attitude-behavior gap. Combining multiple sources of big data (including mobile phone network data and user-generated content) and identifying tourists' mobility flows among points of interest in a travel destination, this research offers a new methodological approach to analyzing real behavior patterns.

To achieve this goal, this research extends the application of Social Network Analysis techniques to identify sustainable behavior patterns and aims to understand how tourists' behaviors evolve over time based on the structure of the user-generated content network. In particular, it seeks to identify sustainable behaviors that can reduce overcrowding in specific areas, distribute visitor flows throughout the destination, and maximize the diversification of economic opportunities. Furthermore, the proposed methodology has potential utility for local authorities, enabling them to monitor and track tourists' real behaviors. It also facilitates the identification of emerging patterns favoring a more sustainable approach to tourism. Consequently, local authorities will derive significant benefits from this analytical framework, fostering a more informed decision-making process in the context of sustainable tourism management.

Tourism is a double-edged sword. On the one hand, it produces numerous economic benefits for travel destinations. On the other hand, it can have negative impacts on travel destinations, ranging from reduced safety in overcrowded sites to environmental degradation due to higher levels of pollution and energy consumption from transportation (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 involves evaluating environmental and social dimensions, focusing on reducing negative environmental, social, and economic impacts (Bramwell et al., 2017).

Among various aspects related to sustainable tourism, reducing over-tourism has become a priority for many destinations (Mihalic, 2020). Therefore, Destination Management Organizations (DMOs) are increasing their focus on how tourists use each destination, as well as understanding how to limit the unsustainable number of tourists in a location (Gowreesunkar & Thanh, 2020). Examples of sustainable tourist behaviors that mitigate over-tourism include choosing to visit lesser-known destinations or specific locations within a destination to avoid crowded places and allow popular tourist spots to recover from the damage caused by the crowd (Lee et al., 2013).

UNWTO (2017, 2022) provides guidelines for promoting sustainable tourism. Consequently, DMOs should manage tourist flows to alleviate congestion in crowded areas and promote diversification of activities. These policies have multiple sustainable outcomes. Firstly, they reduce the overall tourist load, mitigating noise and pollution while improving the quality of the tourist experience and the quality of life for local residents. Secondly, they create diversification and new economic opportunities related to tourism, emphasizing the protection of local territories and cultures by decongesting urban centers (Camatti et al., 2020).

In this context, it is imperative to implement analysis techniques that effectively allow an understanding of tourist behaviors in travel destinations that align with the sustainable development of these destinations and counteract the implications of over-tourism. Although recent studies have shown that tourists are broadly interested in incorporating all dimensions of sustainability into their travel experiences (Dolnicar et al., 2017), in reality, they still behave differently (Agag et al., 2020). This inconsistency between intentions and actual behaviors is known as the intention-behavior gap (Ajzen, 1991, 2001). Consequently, many authors have highlighted the difficulty in extracting evidence of real behaviors from intentions or attitudes, analyzed through survey-based or interview-based methods. Beyond the intention-behavior gap, these methods suffer from the "social desirability" of responses, which represents a significant limitation of studies investigating pro-environmental behaviors because respondents usually want to be perceived as socially and environmentally respectful to avoid criticism (Holtgraves, 2004).

Since sustainable tourism is primarily the result of aggregated behavior and the impact is made collectively, big data and its analysis could help the tourism sector analyze tourists' real behaviors related to sustainability while reflecting their longitudinal change over time (Agrawal et al., 2022; Xu et al., 2020). Additionally, the tourism industry represents a fertile context for collecting and analyzing big data from various sources, such as unstructured data, including user-generated content (UGC), and data from social media, or structured data, such as mobile device data, GPS, and Bluetooth data, as well as transaction data from web searches (Li et al., 2018).

Despite the advantages associated with big data, research that combines multiple sources of big data to analyze tourists' real behaviors, especially their sustainable patterns over time, remains a topic of debate (Xu et al., 2020). Consequently, in this study, we propose to investigate this research question (RQ): RQ: How can big data approaches that include both structured and unstructured data help tourism organizations better understand tourists' sustainable behaviors? In particular, how can tourism organizations identify and analyze effective sustainable tourist behaviors through big data analysis?

We achieve this goal through the methodological approach of our study that combines multiple sources of big data (e.g., offline structured data: mobile phone network data; online unstructured data: UGC) to represent tourists' sustainable behaviors regarding mobility flows during a journey. This is done by detecting the chosen destination through tourists' points of interest indicated in UGC (i.e., nodes) and tracing their relationships through Social Network Analysis (SNA). UGC from social media represents a form of unstructured text data from online reviews (Balducci & Marinova, 2018). Unstructured data typically includes text, images, emails, or personal information and other types of data that are not part of a database (Feldman & Sanger, 2007). To overcome the limitations of unstructured data (Mazenec, 2020; Brave, Butters, and Fogarty, 2022), the actual tourist presence and mobility in the destination are detected from mobile phone network data (also known as TELCO). TELCO represents a form of structured data due to its defined format and length, based on date and time, location, and SIM ID, which are subsequently anonymized (Eberendu, 2016).

In this study, we use big data to identify those sustainable behaviors related to tourists' choices towards less touristy places and more diversified experiences (Mihalic, 2020). By proposing an analytical protocol, we help destination organizations interpret real behaviors associated with aggregated tourist behaviors and provide practical guidance for monitoring and developing solutions for sustainable destination development, focusing on the diversification of tourist flows and activities. This study offers a comprehensive list of methodological phases as guidelines for policymakers to monitor and optimize mobility flows, as well as review and diversify the destination's offerings in terms of routes, tourist attractions, and activities.

The case used in this study is that of the city of Verona and two datasets were used, one for structured data (TELCO data) and the other for unstructured data (UGC).

We used SNA techniques to create the network structure of Verona, with its points of interest (i.e. nodes) and their relationships. Subsequently, sustainability criteria were selected to detect and measure real sustainable behavior patterns. According to the guidelines for sustainable behaviors, presented by the UNWTO (2017, 2022), real sustainable behaviors were measured using the metrics of dispersion centrality and diversification centrality. These two measures represent a useful support for describing mobility flows across the destination, while identifying emerging mobility trends. To decongest areas with high tourist density, these trends, once detected, can be promoted through ad hoc campaigns by municipalities. The identification of behaviors 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 the diversification and promotion of cultures and traditions locals. We have described the evolution of real behavior models over time, capturing the existence of any sustainable transformation of tourists' choices. We compared the network structures over the period 2013–2022, considering two levels of analysis. At the first level of analysis we attempted to identify the degree of difference in the structures over the years through correlation matrices. The correlation matrices demonstrated their similarity at 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 changes in behavior as these may result from several factors; one may be related to the COVID-19 pandemic, which has changed the degree of risk and fear perceived by tourists when visiting crowded areas (Czarnecki, Dacko, & Dacko, 2023).

Another factor with a potential influence on tourism behavior may be associated with the revision of a destination's offering, indicating a recent trend towards more sustainable activities. Notably, this change is in line with a strategic objective outlined in the destination's development plan, emphasizing improved sustainability. A key objective of this plan concerns the intentional decongesting of highly frequented areas and attractions.

The results of our study allow the city of Verona to review its offer, communication and promotion of its attractions. For example, our data analysis reveals that most tourists visit two main attractions, namely the Arena amphitheater and Juliet's House, while only a few tourists visit other significant but lesser-known nearby attractions, such as Castelvecchio and the his museum. Conducting ad hoc communication activities, increasing awareness of less popular attractions, as well as those that are not concentrated in the city center, combined with a more personalized ticketing system, can offer tourists package-price tickets that can redirect tourists flows and induce them to visit the less frequented attractions in combination with the most visited ones. This will be able to better spread mobility flows in the city center, adapting the route through Verona.