Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse

Shihan (David) Ma, M.S. – corresponding author
Corresponding address:
[206C Florida Gym
P.O. Box 118209
Gainesville, FL 32611-8209, USA]
Email: david.ma@ufl.edu
Phone: +1 352 294 1679

Doctoral Candidate
Dept. of Tourism, Hospitality and Event Management
College of Health and Human Performance
University of Florida

Andrei P. Kirilenko, **Ph.D.**
Associate Professor
Dept. of Tourism, Hospitality and Event Management
College of Health and Human Performance
University of Florida
Email: andrei.kirilenko@ufl.edu
Phone: +1 352 294 1648

Svetlana Stepchenkova, Ph.D.
Associate Professor
Dept. of Tourism, Hospitality and Event Management
University of Florida
Email: svetlana.step@ufl.edu
Phone: +1 352 296 1652

Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse

Abstract

This study puts to empirical test a major typology in the tourism literature, mass versus special interest tourism (SIT), as the once-distinctive boundary between the two has become blurry in modern tourism scholarship. We utilize 41,747 geo-located Instagram photos pertaining to the 2017 Great American Solar Eclipse and Big Data analytics to distinguish tourists based on their choice of observational destinations and spatial movement patterns. Two types of tourists are identified: opportunists and hardcore. The motivational profile of those tourists is validated with the external data through hypothesis testing and compared with and contrasted against existing motivation-based tourist typologies. The main conclusion is that large share of tourists involved in what is traditionally understood as SIT activities exhibit behavior and profile characteristic of mass tourists seeking novelty but conscious about risks and comforts. Practical implications regarding the potential of rural and urban destinations for developing SIT tourism are also discussed.

Keywords

Big Data; Instagram photos; Social media; Spatial analysis; Special interest tourism; Astro-tourism (or solar eclipse)

Highlight

- Tests major tourism typology, mass versus SIT tourist, with Big Data analytics.
- 41,747 Instagram photos of 2017 Great American Solar Eclipse are used.
- Two types of tourists, *opportunists* and *hardcore,* are identified and validated.
- Opportunists exhibit behavior and profile characteristic of mass tourists.
- The two segments require different approaches for development of SIT tourism.

Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse

1 INTRODUCTION

Special interest tourism (SIT), interchangeably referred to as niche tourism, has become a noticeable phenomenon in the tourism industry and tourism literature since the 1980s. The concept of SIT is generally defined as travel for a specific interest or motivation with a provision for a customized experience (Douglas, Douglas, & Derrett, 2001; Weiler & Hall, 1992), which traditionally attracts small number of tourists (Robinson & Novelli, 2005; Weiler & Hall, 1992). The rapid growth of SIT is largely due to the heterogeneity of the market products, as well as the increasing demand for more focused activities and interest-based travel experiences (Douglas et al., 2001; Trauer, 2006).

From this point of view, special interest (SI) tourists have a natural desire to shift away from mainstream mass tourism (Robinson & Novelli, 2005) and demand more specialized activities and interest-based tourism experiences (Ali-Knight, 2010). Therefore, SIT has traditionally been regarded as the counterpoint alternative to mass tourism (Douglas et al., 2001), (Robinson & Novelli, 2005). However, there have been arguments that SIT and overall tourism (or mass tourism) are not necessarily mutually exclusive and often overlap (Hall, 1992, 2003; Trauer, 2006). From the demand perspective, SI tourists vary in motivations, sub-segmenting on the continuum from GIT (general interest tourism, namely the mass tourism) to SIT (Brotherton & Himmetoglu, 1997). A “dabbler” SI tourist (an unconvinced mass tourist who has a preliminary interest in SIT) is at the transitional stage from GIT to SIT and may still strongly resemble a mass tourist. From the supply perspective, destinations and tourism and hospitality businesses cater to modern tourists' growing demand for excitement and personalized experiences (Robinson & Novelli, 2005; Wearing, 2002) and multi-motivational choice for destinations (Ryan, 2003). As a result, tourists are now presented with a variety of activities at the destination, including those supposed to be special-interest activities (McKercher & Chan, 2005). Hence, SIT has evolved from a narrow niche market to one that appeals to a more mainstream audience, and what once seemed to be a distinctive boundary between mass tourism and SIT has become blurred (Agarwal, Busby, & Huang, 2018).

So, is SIT merely a kind of a fashionable tourism product or true special-interest tourists still exist? If so, then how can we segment SI tourists from the mass tourists? Segmentation in tourism research has shifted from the traditional conceptual approaches (Cohen, 1983; Plog, 1987), i.e., *a priori* typological construction based on demographic characteristics and tourist socio-psychographic variables, to data-driven approaches with statistical analysis and quantitative measurement in the past two decades (Dolnicar, 2002a). The advantage of such *a posteriori* approaches is to incorporate the complexity of tourist destination consumption behaviors (Težak, Saftić, Težak, & Bošković, 2011) with fact-based data (Dolnicar, 2002a). Multiple studies have used behavioral constructs as group criteria to identify tourist types and market segments (Dolnicar, 2002b; Dolnicar & Fluker, 2003; McKercher & du Cros, 2003; Phillips & Brunt, 2013). However, these studies are commonly based on small-sample survey data. The massive amount of tourist-generated data in social media was not fully leveraged in tourism segmentation until recently (De Cantis, Ferrante, Kahani, & Shoval, 2016; Donaire, Camprubí, & Galí, 2014; Hernández, Kirilenko, & Stepchenkova, 2018; Kirilenko, Stepchenkova, & Hernandez, 2019). The research questions addressing those methods are just a few at present, and the overall body of such research is still small.

The tourism industry has entered the “digital tourist era” with high penetration of social media usage before, after, and most importantly, during travel (Amaro, Duarte, & Henriques, 2016). The potential of social media to discern tourist-related travel behaviors and patterns while they are traveling has been convincingly demonstrated (Hernández et al., 2018; D. Leung, Law, van Hoof, & Buhalis, 2013) via the use of data mining, content analysis, network analysis, and other techniques (Donaire, 2011; Donaire et al., 2014; Stepchenkova & Zhan, 2013; Y.-T. Zheng, Zha, & Chua, 2012) although the samples in some of those studies were rather small. Additionally, social media messages frequently contain fine-resolution data on the geographical location of the travelers, which is beneficial for generating tourist-related movement routes (García-Palomares, Gutiérrez, & Mínguez, 2015). All of these attributes, including availability of mobile devices, high rate of social media participation, massive user-generated content (both textual and visual) stored online, identifiable geographical location, and availability of demographic information from user profiles now converge and make it possible not only to approach a specific research question but also to test standing theories of tourists motivations and behavior (Vecchio, Mele, Ndou, & Secundo, 2018) using big data analytics.

Thus, the main purpose of this study was to put tourism typologies associated with SIT to the empirical test in order to get further insight into the debated relationship between SIT and mass tourism. More specifically, using social media big data, we are set to investigate the question that has been posited by McKercher and Chan (2005): “*How special is special interest tourism?*” For the context of the study, we chose astro-tourism, because it is a “classical” example of SIT (Matos, 2017; Soleimani, Bruwer, Gross, & Lee, 2018; Wen, 2017).

Specifically, we used the 2017 Great American Solar Eclipse event and the photography it generated on Instagram, the largest photo-sharing platform at present (Statista, 2018). Using big data analytics, the data allowed us to investigate the travel behavior of eclipse-chasers and to identify the most popular destinations in eclipse observations, the origins of these visitors, and their spatial movement patterns to gauge the commitment of these tourist to pursue this SIT activity. We further identified two types of tourists based on their choice of destinations and movement behaviors. We validated and profiled these segments by formally stating and testing two hypotheses using the data that was external to the one used to identify the segments. This segmentation is validated by physical distance traveled in addition to the socio-economic characteristics of visitors.

We see the main contributions of this study in two areas. First, the derived and validated segments of eclipse-chasers allow verification of the SIT tourist taxonomy produced in early tourism literature when big data and methods to handle it were not yet available. While the previous studies largely used samples, which were not necessarily representative and often small (Dolnicar, 2002a) or approached the issue qualitatively, our study incorporated practically the whole assembly of data from tourists who reported their eclipse experiences on Instagram. We examined the segmentation results against the backdrop of traditional typology theories in SIT and motivation theories in leisure and tourism studies to show the concordance of our main findings with existing tourist typologies based on psychographic variables. The theoretical division between SIT and mass tourists as discussed by Brotherton and Himmetoglu (1997) and Hall (2003) was put to the empirical test, thus contributing to answering the question by McKercher and Chan (2005): "*How special are special interest tourists?*" with results supporting the view that people with characteristics of mass tourists actively participate in SIT. Second, with reference to its astro-tourism context, the results allow for critical evaluation of marketing recommendations proposed in the literature for development of SIT destinations. We re-examined the topological and environmental conditions qualifying the astro-tourism destinations and offered suggestions for destinations to leverage SIT resources for effective marketing and promotion.

2 RELATED WORKS

2.1 SIT and market segmentation

The traditional tourist typology theories were mainly conceptual taxonomies based on tourists' sociological and psychological attributes. Tourists were distinguished on bipolar dimensions in motivation and behavior such as seeking for novelty – staying in familiarity (Cohen, 1972), quest for pleasure – pursue ultimate meaning (Cohen, 1979), psychocentrics – allocentrics personality (Plog, 1974, 1991), as well as tourists interaction with destinations (1982). Based on these theoretical backgrounds, Brotherton and Himmetoglu (1997) conceptualized SIT as a "special" form of tourism, in which the tourists had a specific, interest-based motivation in their travel decision, differentiated from "general" tourists. A "Tourism Interest Continuum" was posited ranging from General Interest Tourism (GIT) to Mixed Interest Tourism (MIT) and, finally, to SIT. "Dabbler" tourists are at the transitional stage from GIT to MIT. With concerns about risk and unfamiliarity, they only seek "fashionable" or "popular" products; while on the other end of the continuum, "expert" tourists have a specific interest to pursue, which make them more dedicated to SIT activities. This motivational theory has been the fundamental ground for multiple follow-up studies about various SIT forms, such as health and spa tourism (Hall, 2003), wine tourism (Charters & Ali-Knight, 2002) as well as astro-tourism (Fayos-Solà, Alvarez, & Cooper, 2014; Soleimani et al., 2018; Wen, 2017).

Data-driven segmentation approaches derived from marketing research (Balasubramanian, Gupta, Kamakura, & Wedel, 1998) have been widely used for segmenting tourists (Dolnicar, 2002a). The studies have employed advanced statistical procedures and examined tourist types with a wide range of objective and subjective measurements and factors, such as the purpose of travel, travel behaviors, prices, demographic and geographical features, psychographic personalities, etc. (Dey & Sarma, 2006; Díaz-Martín, Iglesias, Vázquez, & Ruiz, 2000; Lehto, O'Leary, & Morrison, 2002; Middleton & Clarke., 2001; Neuts, Romão, Nijkamp, & Shikida, 2016; Pride & Ferrell, 2016). Such typologies are commonly market-driven and extremely popular in identification of (sub-)segments within certain SIT activities, such as cultural tourism (Dolnicar, 2002b; McKercher & du Cros, 2003), ecotourism (Arnegger, Woltering, & Job, 2010; Hvenegaard, 2002), medical tourism (Wongkit & McKercher, 2013), sport tourism (Dolnicar & Fluker, 2003; Phillips & Brunt, 2013), etc. It is noteworthy that behavioral constructs (such as participation, involvement, investment) are frequently used as group criteria in SIT segmentations (Dolnicar, 2002b; Dolnicar & Fluker, 2003; McKercher & du Cros, 2003; Phillips & Brunt, 2013), and tourists' behavioral disparities are frequently associated with sociodemographics (such as income, education experience, and gender) in SI tourist profiles, especially when SI activities are closely pertaining to one's lifestyle, aesthetic and cultural preferences (Marzo-Navarro & Pedraja-Iglesias, 2010; Nella & Christou, 2014; Shani, Wang, Hutchinson, & Lai, 2010)

Segmentation criteria have been greatly expanded with the employment of social media data. Tourist demographic traits on social media profiles and their usage preference have been utilized for market segmentation (Amaro et al., 2016; Mavragani, Nikolaidou, & Theodoraki, 2019). More advanced, all derived attributes such as visitation location, consumption choices, technology orientation can contribute to the segmentation criteria with big data analytics (Vecchio et al., 2018). For instance, Donaire, Camprubí, and Galí (2014) clustered tourists into four groups based on their common preferences shared in their social media photographs. Using GPS tracking data, De Cantis et al. (2016) segmented cruise passengers based on their mobiliaty patterns and soci-dempographic profiles. Hernández, Kirilenko, and Stepchenkova (2018) identified tourist market segments with different attraction choices and travel interests accoriding to tourist online reviews. However, depiste the emerging studies in the field of tourist market segmentation with big data analytics, a more global question of testing and verifying the existing tourism typologies that were obtained through qualitative and small-sample quantitative methods have not yet been approached. This study provides the first take on the issue with regard to one of the most fundamental typologies existing in the tourism literature: mass tourists versus SI tourists. The study's context, that is, astro-tourism, and existing segmentaiton works in this area are discussed in the next section. The recent advances in using photograph data from social media relevant for this study are described in section 2.3.

2.2 Eclipse chasing and observation as astro-tourism

Solar eclipse observation is arguably the most popular astronomy-related activity (Wen, 2017). Activities with astronomic attributes date back centuries and have been embedded in such world-famous heritage attractions as Stonehenge and Woodhenge in the UK, Chichen Itza in Mexico, and Machu Pichu in Peru (Malville, 2008). These locations are historic sites related to the timing of celestial objects and traditional cultural practices (Collison & Poe, 2013), offering archaeoastronomy experiences to the public (Fayos-Solá, Marín, & Jafari, 2014). The modern form of astro-tourism, somehow, is an emerging market in the tourism industry and a less-studied area to which little attention was paid until recently. Existing discussions are still focusing on astro-tourism definitions and the activities it contains, with conceptualizations and corresponding typologies proposed in competing yet adjacent ways, such as terrestrial space tourism (Crouch, 2009; Crouch, Devinney, Louviere, & Islam, 2009), astrotourism (Cater, 2010), celestial ecotourism (Weaver, 2011), or astronomical tourism (Collison & Poe, 2013). But it is generally accepted in most recent studies that astro-tourism is a form of special-interest tourism (SIT), i.e., traveling to destinations for celestial observation, visitation to astronomy-related sites, and participation in astronomical activities (Matos, 2017; Soleimani et al., 2018; Wen, 2017).

More specifically, three types of astro-tourism activities have been summarized. First, travel to destinations with suitable natural conditions for observation and astrophotography of the celestial objects and astronomical phenomena (Cater, 2010; Fayos-Solá et al., 2014; Soleimani et al., 2018). Traditionally, these destinations have a "dark sky" for stargazing or locations with aurora display (Collison & Poe, 2013; Weaver, 2011). In this respect, astro-tourism has been regarded as a "sustainable tourism" that frequently directs travelers to remote locations with clear skies and low levels of light pollution. Discussions of astro-tourism as an existing phenomenon also increase the awareness of the light pollution issue and the urgency of protecting the "starlight" and minimizing light pollution in local environment (Fayos-Solá et al., 2014; Rodrigues, Rodrigues, & Peroff, 2015). Second, astro-tourism increases visitation to scientific infrastructures, such as observatories, science museums, and laboratories, as well as astronomy-related historical sites (Burtnyk, 2000; Fayos-Solá et al., 2014; Robson, 2005; Weaver, 2011). Those activities are more knowledge-driven and are comparably more likely to attract amateur and professional astronomic travelers, highlighting its nature as a form of special-interest tourism (Soleimani et al., 2018). Finally, astro-tourism includes astronomy-related activities and community interactions, such as star parties, which attract tourists with similar interests and hobbies (Wen, 2017).The existing typologis regarding astro-tourists genenrally follow the theoritcal framework of those in SIT. Fayos-Solá et al. (2014) categorized astro-tourists into two types, the general public and the amateur/professional astronomers, and argued that the amateur and astronomic communities played a significant role in cultivating and accelerating the market. Matos (2017) proposed an astro-tourist classification based on travel motivation and involvement, grading the astro-tourists into specific astro-tourist, casual astro-tourist, and serendipitous astro-tourist. Wen (2017) integrated the Serious Leisure Theory (Stebbins, 1982, 1997) and SIT tourist continuum (Brotherton & Himmetoglu, 1997), classifying astro-tourists into dabblers, enthusiasts, fanatics, and specialists according to their travel history and involvement. This classification, however, together with other astro-tourism typologies (Fayos‐Solá et al., 2014; Matos, 2017), was based on presumed theoretical framework and small-sample self-reported measures, and, therefore, lacked evidential behavioral support. The destination preference of different categories of astro-tourists or segment identification based on astro-tourists' destination attributes was also missing.

2.3 Geo-tagging social media photography

The availability of large sets of travel photographs publicly shared through social media have provided an accessible source for tourism researchers. Numerous studies have utilized data extracted from early photography sharing social media: Flickr and Panoramio (Donaire, 2011; Donaire et al., 2014; Kim & Stepchenkova, 2015; Kisilevich, Krstajic, Keim, Andrienko, & Andrienko, 2010; Stepchenkova & Zhan, 2013; Y.-T. Zheng et al., 2012). However, with Panoramio being discontinued following its purchase by Google and Flickr shifting its priorities towards professional photographers, the amateur photographers *en masse* have shifted towards alternative platforms, and researchers followed. Using Instagram data has become a trend in recent publications (Chen, Parkins, & Sherren, 2018; Mukhina, Rakitin, & Visheratin, 2017), reflecting the Instagram's status as the most popular photo sharing platform (Statista, 2018).

Social media photography frequently comes with auxiliary data (metadata). Among these metadata, time and location of the photographs are extremely valuable assets. The geotagged data can be used as a proxy for space attractiveness, helping to identify the major tourist attractions and the intensity of the land use (García-Palomares et al., 2015; Kisilevich et al., 2010; Yuan & Medel, 2016). When temporal information is used in addition to the spatial data, tourist photography can be used to identify tourist movements, visit preference, mobility patterns, and to assess tourist routes. The validity of this approach was proven by De Choudhury et al. (2010) who successfully compared the tourist trajectories identified from Flickr photographs with bus routes. Önder, Koerbitz and Hubmann-Haidvogel (2016) traced the travel pattern of tourists in Austria based on the geographical and textual analysis of over one-million photographs. Leung, Vu, and Rong (2017) used Flickr data to analyze tourist movements and visit patterns in Hong Kong. Straumann, Çöltekin, and Andrienko (2014) analyzed the visitation locations and travel routes of foreign and domestic tourists in Zurich from their posted photographs, and found significant difference. Certain groups of tourists may share similar travel routes and movement patterns, and it was suggested that such tempo-spatial features can be used as tourist segmentation criteria (W. Zheng, Huang, & Li, 2017).

The spatiotemporal visitation pattern has been leveraged to successfully distinguish between locals and visitors in several studies. The researchers made the fine distinction based on the series of user photographical "footprints", that a photographer is classified as a visitor if he/she is publishing photographs taken within the area of interest during a short period ranging from few days to few weeks, then moving to a different place, whereas a local is more likely to be present in the specific area with higher frequency or during an extensive timeframe. The timeframe threshold to cut between local and visitors may vary. Girardin, Dal Fiore, Ratti, and Blat (2008) used a 30-day local photo timeframe to identify visitors with no further justification on the period length provided. Contrasting, Donaire, Camprubí, and Galí (2014) used a 5-day local photo timeframe to differentiate visitors with locals. Similar considerations also allow identification of the visitor's origin, e.g., from the area where the visitor made the most photographs or from the area where the visitor made the photographs for the longest time (Järv, Tenkanen, Toivonen, & Hiippala, 2018). This approach was validated by Heikinheimo et al. (2017) who verified the photography-based origin identification using survey data.

Thus, identification of spatio-temporal travel patterns from online shared photography, as well as classification of the travel tracks into those left by the visitors and locals, are well established in the recent tourism literature. Segmentation based on tourist mobility pattern has been suggested (W. Zheng et al., 2017) and found adequate in studies (De Cantis et al., 2016). The following section details how we applied the outlined methods to the analysis of the spatial presentations and movement patterns of tourists during the 2017 total solar eclipse.

3 DATA

3.1. Study area

The “Great American Eclipse” occurred on August 21, 2017 with the total eclipse phase starting at 17:16 UTC at the US west coast and ending at 18:44 UTC at the east coast. At any given place, the total eclipse phase lasted for 2.7 minutes, and the partial eclipse lasted for about1.5 hours. The totality path was selected as the study area to represent the terrestrial footprint of the eclipse, which was projected into a 110-km wide ribbon crossing 14 US states: Oregon, Idaho, Wyoming, Montana, Nebraska, Iowa, Kansas, Missouri, Illinois, Kentucky, Tennessee, Georgia, and North and South Carolinas (See Appendix Figure A1).

3.2 Data collection

We collected the Instagram posts geotagged within the eclipse totality path using a Picodash hashtag search. Picodash (www.picodash.com) is an Instagram photography archive allowing a keyword and geolocation search. A pilot search was used to identify search terms based on additional hashtags correlated with the top three hashtags pertaining to the event: #eclipse2017, #solareclipse, and #solareclipse2017. The data was cleaned by removing the duplicates and photographs from outside the eclipse path. Then, the photographs posted by the same users on the same day from the same location were aggregated so that each eclipse photographer at a specific location would be represented by a single photograph. Thus, we retained 41,747 geotagged photographs taken by 37,652 unique users. This dataset is further referred to as Dataset 1 (see Figure 1 for the spatial distribution of eclipse photographs). The timeline of the photography (Figure 1 for the temporal distribution) demonstrates a distinct spike at 19:00 UTC, coinciding with the end of the total eclipse at the East coast, indirectly supporting our assumption that the majority of collected photographs were indeed posted by the eclipse watchers.

Figure 1. Spatial and temporal distribution of photographs taken within the totality path. These photographs were used as a proxy for eclipse observational points. Timeline of eclipse related posts published between August 21, 2017 0:00 UTC and August 22, 2017 23:59 UTC.

Next, we collected all photographs posted by 37,652 identified eclipse watchers within a 6-month period centered at the eclipse event with the purpose of identifying the home location of the photographers. Similar to the eclipse photographs, we cleaned and aggregated the dataset, thus reducing it to approximately 3 million photographs (mean = 80 photographs per user). This dataset is further referred to as Dataset 2. To maintain the photographer's privacy, the only fields retained were the user ID, latitude and longitude of the posted photograph, and the timestamp; all other non-empty, including the image itself, were discarded.

4. METHOD

This section describes the methodological aspects of the analyses necessary to segment astro-tourists based on their online photo-sharing behavior. These aspects include identification of popular eclipse observation destination (section 4.1) and identifying travel origins, that is, home locations, of astro-tourists (section 4.2). This section is oriented toward the technically inclined reader and can be passed over without loss of understanding of the main results, discussion and conclusion.

4.1 Identifying popular eclipse observation destinations

To find the spatial distribution of popular eclipse observation locations within the solar eclipse path, we performed a point density analysis (Silverman, 1986) on the Dataset 1. Then, we identified separate popular observational areas with cluster analysis. Specifically, we used the DBSCAN (density-based spatial clustering analysis) algorithm (Ester, Kriegel, Sander, & Xu, 1996) for clustering. Compared to a more commonly used K-means clustering, DBSCAN does not require a pre-set parameter regulating the number of clusters and is insensitive to both cluster shapes and outliers. These features make it a suitable tool for clustering noisy data, which is a typical case for geo-located social media. Following recommendations by REF, we used the following DBSCAN parameters: minPts (minimum number of points in a cluster) = 100, ε (search radius of neighboring points) = 200 km.

4.2 Identifying major travel origins

We identified travelers' origins by combining spatial and temporal approaches as discussed in Section 2. Specifically, traveler's origin (home location) was identified from Dataset 2 as the area from which this traveler (1) makes many photographs while (2) exhibiting prolonged presence. Specifically, we used the following algorithm steps for each user (traveler):

1. Cluster analysis was applied to the geographical locations of the user's photographs to find the spatial clusters of the points from where the photographs were posted, and

2. The centroid of the most populous cluster was assumed to be the provisional user's home location P_1.

The provisional location was then validated as follows:

3. The photographs taken during the 3-month periods before and after the eclipse events were processed as described be steps 1 and 2, thus generating locations P_2 and P_3, accordingly;

4. The distances between locations P_1, P_2, and P_3 were computed; if the distances were found to be lesser than 50 miles (80 km), the provisional user's home location P_1 was confirmed; otherwise, the user was discarded from the analysis.

The 50-mile (80-km) distance was based on the tourist's definition as a person traveling over 50 miles from their place of residence (Smith, 1999; UNWTO, 1994). Users with fewer than 30 photographs posted over the 6-month period were excluded from the analysis to abide by the cluster analysis requirements (9,964 users excluded). The validation process successfully identified the origins of 76.96% of travelers with the mean distance between locations P_1, P_2, and P_3 = 4.2 km (Table 1). The remaining 23% users were mis-identified with over 2,110 km locational error among P_1, P_2, and P_3; there were also considerately fewer photographs posted by these users from the wrongly identified home location. After discarding mis-identified users, the home locations of 21,310 users were estimated.

Table 1. Validation of identification process of the user's home location. The error shows the distance between photographer's home location identified from different samples of photographs (pre: pre-eclipse, post: post- eclipse) and all photographs (overall). Notice that the error for

successfully identified home locations is three orders of magnitude lesser than the error for failed identification. Total number of users: 37,652.

Identification	User statistics			Locational error (km)			
	N	%	Valid %	overall to pre	overall to post	pre to post	Mean error
Success	21,310	56.59%	76.96%	5.09	2.50	7.52	4.20
Failure	6378	16.94%	23.04%	2050.68	1209.63	3086.71	2110.83
No data	9964	26.47%	-	-	-	-	-

5 RESULTS

This section reports on the segmentation results of astro-tourists using data and methods described in the previous section. First, we identify observational destinations of eclipse watchers with geo-tagged photographs and group these destinations into clusters reflective of their geographic locations (e.g., rural versus metropolitan). Section 5.2 establishes the visitors' origins (permanent location) by examining their photograph sharing online behavior during six-month period around the eclipse event. Having known these two factors, the movement patterns of astro-tourists are identified, and visitors are differentiated on the distance they traveled (section 5.3). In section 5.4, two main tourist segments are profiled, and their socio-demographic profile is further validated.

5.1 Solar eclipse observational destinations

Spatial distribution of the eclipse observational destinations (see density distribution insert in Figure 2) is consistent with the major urbanized areas located on the west (Portland, OR) and east (the belt spreading from Nashville to Charleston, SC) US coasts. Secondary observational hotspots that do not coincide with any major populated place have also been identified (e.g., the one located close to the I-15 highway near the Idaho and Wyoming border). The observational points are grouped into 16 distinct clusters and multiple outlying locations characterized by just a few observations (Figure 2; Table 2). The largest cluster is located in the densely populated east coast area, stretching from Nashville to Charlotte, NC, but also including major natural areas such as the Smoky Mountain National Park. Other clusters are more compact; they include two metropolitan areas (Portland, OR and St. Louis, MO), three suburbs, and seven locations in rural and natural areas, such as the thinly populated area roughly centered at the border of Wyoming and Idaho, covering Caribou-Targhee National Forest and (partial) Yellowstone (Figure 2; Table 2). Most of these secondary clusters are situated in small cities and rural areas with plain or deserted landscapes.

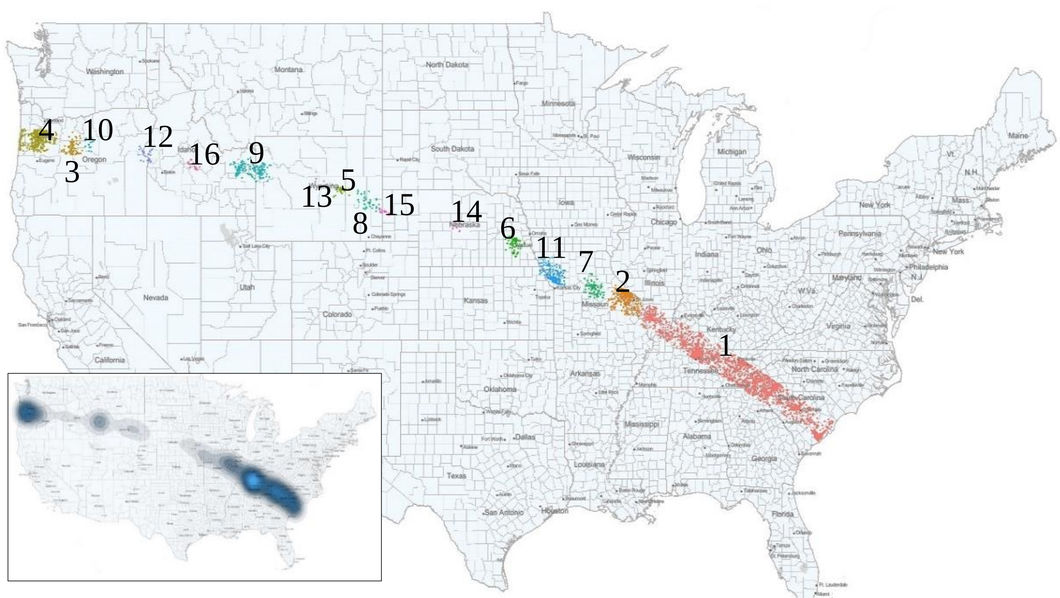

Figure 2. Density distribution (figure insert) and clusters of solar eclipse observational points. For cluster information see Table 2.

Table 2. Area types of solar eclipse observational clusters. For the map see Figure 2.

Cluster ID	Number of points	Location	Area type
1	22,511	East coast	Mixed*
4	4756	Portland	Metropolitan
9	2422	Caribou-Targhee National Forest	Natural
2	2177	St. Louis	Metropolitan
3+10	1831	Oregon	Rural
11	1165	Kansas City north suburb	Suburban
5+13	1022	Wyoming	Rural
7	908	Columbia north suburb	Suburban
8+15	857	Wyoming	Rural
6	523	Omaha-Lincoln suburbs	Suburban
12	329	Idaho-Oregon border	Rural
14	205	Nebraska National Forest	Natural
16	211	Salmon-Challis National Forest	Natural
Outliers	2840		

*Large belt stretching from Nashville to the East coast through populated and natural areas

5.2 Visitor origins

Among the 21,310 users whose home locations have been successfully identified (Table 3), 7,439 (35%) are recognized as locals, that is, those users whose home location was within the eclipse totality path. The mean distance from their home locations to the eclipse observational point was 22.4 km. The remaining 13,871 (65%) users were classified as tourists, with the majority of them (93.85%) being domestic (vs. 6.15% international). The origins of domestic tourists are widely dispersed throughout the country (Figure 3). A significant number of long-distance tourists (15%) traveled from four major metropolitan areas: Los Angeles, Chicago, Atlanta, and New York (Table 4).

Table 3. Identified visitor origins.

Category	Number	Percentage	Mean travel distance (km)
Locals	7,439	34.91%	22.4
International tourists	853	4.00%	6244.8
Domestic tourists	13,018	61.09%	703.9
Total	**21,310**	**100%**	**1044.6**

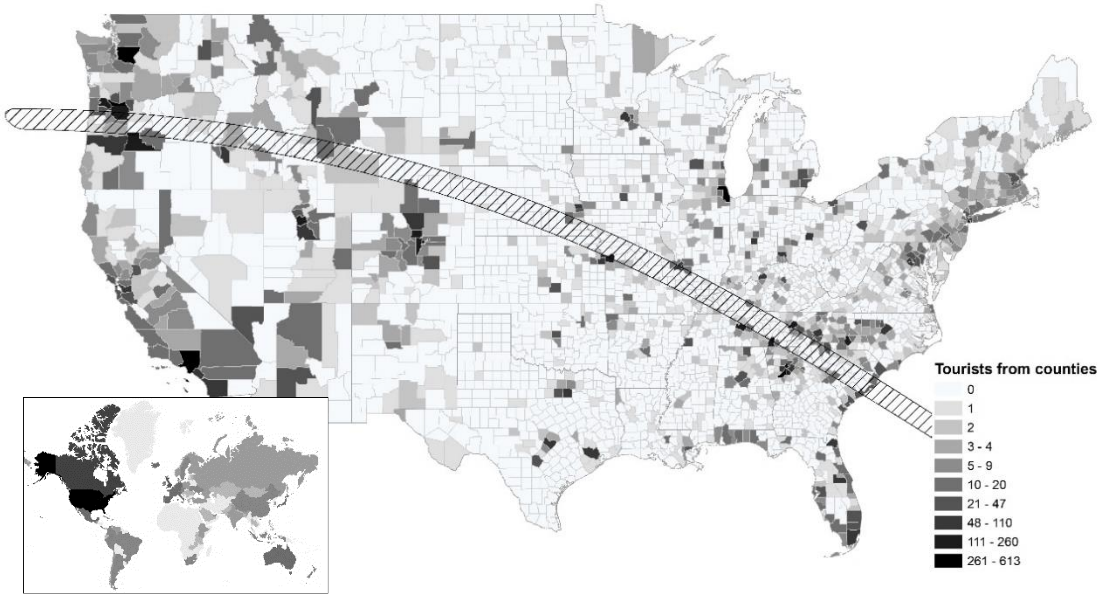

Figure 3. International and domestic tourist origins

Table 4. Top domestic origin counties.

County	Number of tourists	Population (2010 census)	Visit rate (per 10,000)	Nearest distance to totality (km)
Multnomah (OR)	613	735,334	8.34	20.3
Los Angeles (CA)	486	9,818,605	0.49	1010.5
Cook (IL)	429	5,194,675	0.83	394.4
Fulton (GA)	412	920,577	4.48	61.6
King (WA)	383	1,931,249	1.98	211.0
New York (NY)	357	1,585,873	2.25	971.7
St. Louis (MO)	221	319,294	6.92	0.0
Jackson (MO)	208	674,158	3.09	0.0
Salt Lake (UT)	204	1,029,655	1.98	259.5
Mecklenburg (NC)	202	919,628	2.20	58.1
Knox (TN)	166	432,226	3.84	0.0
District of Columbia (DC)	163	601,767	2.71	630.7
Jefferson County (CO)	149	534,543	2.79	229.5
DeKalb County (GA)	146	691,897	2.11	85.3
Williamson County (TN)	142	183,182	7.75	0.0
Kings County (NY)	140	2,504,700	0.56	962.2

5.3 Tourist movement patterns

The edges connecting the nodes formed by travel origins and destinations make a network representing the tourists’ travel flows during the 2017 solar eclipse event (Figure 4). In agreement with the findings from the previous sections, most of the travel flows were generated between the major metropolitan areas and the hotspots in the totality path. The strongest connections follow the “shortest travel time” path, meaning that the majority of the tourists were traveling from their home locations to the nearest eclipse observational points. It seems reasonable to suggest that such travels are influenced by the road system, with the preferable travel pattern following the main highways and, possibly, time available and travel costs. A secondary factor influencing travel seems to be urbanization rate and availability of natural areas at the observational site. For example, the main travel route from Chicago followed the I-57 interstate highway to a natural area located approximately 450 km south, while an alternative route of a similar length to St. Louis was significantly less traveled. Similarly, more tourists from Charlotte, NC traveled along the I-77 to Columbia, SC than along the I-85 to Greenville, SC, even though the latter destination was closer. Similar travel patterns were observed among tourists originating from Seattle, Salt Lake, Denver, and the Bay Area.

Figure 4. Domestic travel network showing 2017 solar eclipse travel patterns.

Assuming that following the shortest travel time path to the observational pattern is the universal movement pattern among all tourists, the frequency distribution of travel distances from the same origin should approximate a truncated normal distribution starting from the shortest-distance as “distance decay” (Greer & Wall, 1979; Oppermann, 1995). Our analysis, however, showed a multi-modal travel distance distribution (Figure 5A&B). While travel distance for the majority of travelers indeed followed the shortest-distance pattern with travel distance distribution close to truncated normal, there is also a segment of eclipse watchers traveling great distances, presumably to arrive at the best observational locations. For example, two secondary peaks (3000 km and 4000 km) on the travel distance distribution for New York tourists (Figure 5A) correspond to the rural area adjacent to Idaho and Wyoming border and to small towns south of Portland, OR, respectively. Evidently, there is a significant segment of New Yorkers making trans-continental travel to rural locations with lower air humidity, lesser cloudiness, clearer sky, and sparser population in the west part of the country (Figure 4). As compared to the shortest-distance travel, this secondary travel pattern could be named the "best-location" travel. Similar travel patterns were observed among tourists from almost all origin counties, regardless of their distance to the totality path (see the illustration for Washington, DC in Figure 5B).

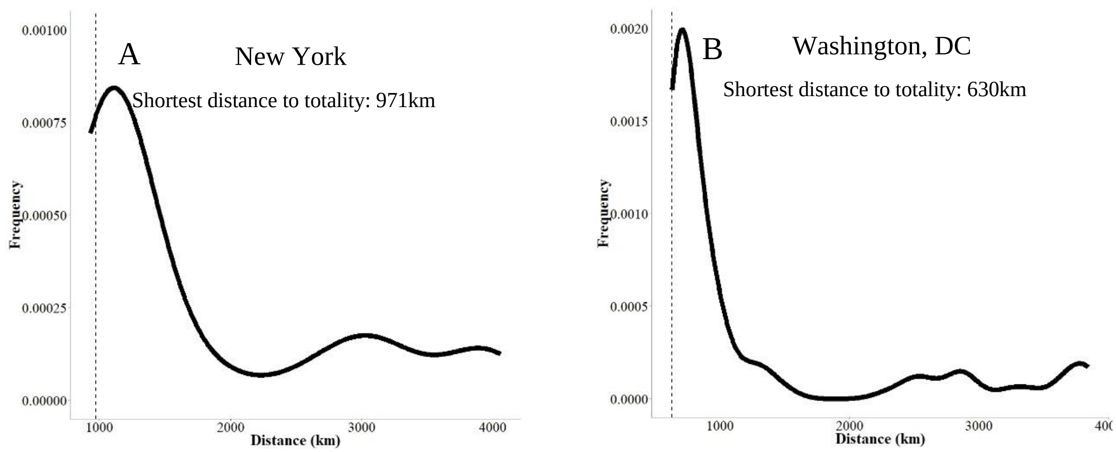

Figure 5. Solar eclipse travel distance distribution for New York (A) and Washington, DC (B) travelers.

5.4 Tourist types: description and validation

The identified movement patterns point toward two distinct types of astro-tourists. The Type I tourists follow the "shortest-distance" principle, preferring geographically proximate, cost-effective observational points. This behavior is generally consistent with low-involvement recreational travel (Fesenmaier & Johnson, 1989). The Type II tourists follow the "best-location" principle, preferring the most advantageous observational points, regardless of the distance to cover, which, typically, result in a significantly longer travel distance. The categorization result of these two types of tourists in the top origin counties and their statistical features are given in

Appendix Table 1A. Whether these segments just an artefact of the method based on limited data points, or whether they reflect the true behavioral pattern of special-interest tourists and can be meaningfully described in demographic terms was further examined We stated and tested three hypotheses:

H_1: Visitation rate to the totality path is negatively correlated with the shortest travel distance for Type I travelers and not correlated for Type II travelers;

H_2: Visitation rate to the totality path is not correlated with education level for Type I travelers and positively correlated for Type II travelers;

H_3: Visitation rate to the totality path is not correlated with income for Type I travelers and positively correlated for Type II travelers.

H_1 posits that usually physical distance has a negative impact on tourists' travel intentions. This is consistent with common sense travel decision-making (Nicolau & Más, 2006) since physical distance fundamentally influences travel cost (Cesario, 1976). A select group of highly motivated tourists, however, is little affected by travel distance. The correlational analysis of visitation rate and the nearest distance to totality path indeed indicates that Type I tourists are indeed distance-sensitive: $r = -0.572$ ($p < 0.001$); Type II tourists, on the other hand, are unaffected by the distance, implying that the distance is not a constraining condition for Type II tourists: $r = -0.071$ ($p = 0.572$).

Hypotheses H_2 and H_3 evaluate the role of education (measured as the percentage of the population with a bachelor's degree or higher in the county) and income (median household income in the county, USD) in the proportion of Type I and Type II tourists for any particular place of origin. We expect that the points of origin with higher educational level and higher mean income of population will have a relatively higher rate of Type II tourists, while the rate of

Type I tourists would be relatively unaffected by education and income. We found that the Type II visitation rate is indeed positively correlated with the educational level ($r = 0.388$, $p = 0.001$) and income ($r = 0.254$, $p = 0.039$). Contrastingly, the Type I tourist visitation rate is not correlated with either educational level ($r = 0.03$, $p = 0.81$) or income ($r = -0.027$, $p = 0.829$). Thus, hypotheses H_1 – H_3 were supported by the analysis.

6 DISCUSSION

Our data analysis strongly suggests the existence of two distinct types of tourists who deviate in their spatial behavioral preferences that are viewed as indicative of interest in astro-tourism. Type I tourists generally follow the "shortest-distance" pattern in their destination choices, presumably selecting convenient observation locations, while Type II tourists are less affected by travel distance. These two types of tourists are also different in their educational and financial characteristics, with the Type II tourists being more prevalent in communities of higher educational levels and financial affluence, whereas Type I tourist characteristics appear to be not directly related to educational and income status.

6.1 Theoretical implications

From the theoretical perspective, these two identified tourist types with their respective disparities in involvement, cost-sensitivity, and travel intentions are likely to be well interpreted by the continuum of SIT tourists as proposed by Brotherton and Himmetoglu (1997) and Serious Leisure Theory by Stebbins (1982, 1997). We suggest that Type I tourist behaviors approximate those expected of ***opportunity tourists*** or ***opportunists***, seeking "fashionable" or "popular" experiences. A similar pattern of behavior was termed by Brotherton and Himmetoglu (1997), Stebbins (1997) and MacKellar (2009) as “dabblers”. While expressing a general interest in astronomical events, quite possibly under media influence (Trauer, 2006), opportunists are less-engaged and less-experienced in such special-interest travel. Given a choice, they are likely to select the most cost-effective and time-efficient locations in which they can combine several types of activities. Their choice is consistent with easy access to adequate facilities as they are “sensitive to risks” and tend to choose destinations “familiar to their previous experiences” (Brotherton & Himmetoglu, 1997). In many aspects, the travel behaviors of opportunists are similar to traditional mass tourists or the “general public” according to the Fayos-Solá et al. (2014) classification, yet with the motivation of “novelty” (trying a new interest and activity) (Brotherton & Himmetoglu, 1997).

On the other hand, Type II tourists tend to approach the other end of the SIT spectrum in which the behaviors approximate those of “fanatics”, “specialists”, “hobbyists”, “enthusiasts”, or “experts” as proposed in multiple typologies (Brotherton & Himmetoglu, 1997; MacKellar, 2009; Stebbins, 1997; Wen, 2017). In our study, we named the Type II tourists as ***determined special-interest tourists***, or ***hardcore*** because they have been demonstrated to be more involved and more invested in such astronomical activity tourism and will travel significantly longer distances to specific destinations. The hardcore tourists could be amateurs, professional astronomers, and scientists; however, their intention to attend astronomical activities is universally and highly self-determined. They would spare no effort and easily sacrifice time and convenience to travel to a premium observational point. The motivation and behaviors of the hardcore special-interest tourists can be interpreted using the Serious Leisure Theory that states optimal observational experience is “sufficiently rewarding despite the costs”, in the “acquisition and expression of skills and knowledge” and requires “high investment” (Bartram, 2001; Brotherton & Himmetoglu, 1997; Stebbins, 1982).

While the two segments seemingly fit the previously identified SIT tourist typology, the findings of this study do not fully support the view that SIT tourism is a niche tourism and further led us to the viewpoint presented by Agarwal, Busby, and Huang (2018) in which SIT has evolved to attract a significantly larger and mainstream audience. The apparent existence of a large number of Eclipse images on Instagram (0.57 million) together with a massive number of people willing to travel to observe a major astronomical event seems to contradict the traditional description of SIT as the opposite of mass tourism. Instead, SIT should possibly be counted as a subdivision of mass tourism. The mean visitation rate of opportunists (1.8 per 10,000 population) is estimated to be an order of magnitude higher than that of the hardcore tourists (0.16 per 10,000 population), indicating that an astro-tourism event is not just a peripheral activity of astronomy geeks but a fashionable trend with a remarkable market basis. Therefore, we argue that instead of recognizing ***opportunists*** as entry-level SI tourists, transiting from GIT to SIT (Brotherton & Himmetoglu, 1997), it seems more fitting to identify them as mass tourists who are attracted by major SIT events when participation costs are nominal. The large-scale empirical evidence unearthed by this study does not contradict the view that opportunists resemble the modern mass tourists with multiple motivations and desire for personalized experiences. In contrast, ***determined special-interest tourists (hardcore)*** are well-fit to the traditional concept of SIT as specific interest-based tourists with a smaller number of tourists (as compared to mass tourists). Although there have been conceptualized motivation models for SIT indicating that SI tourists are also multi-motivational (Ryan, 2003) and special interest is not the solo motivation pursued throughout the entire travel experience (Trauer, 2006), further revisit of the traditional conception of SIT is necessary in this regard.

6.2 Marketing implications

Observations of celestial phenomena such as solar eclipse, “supermoon” (full moon coinciding with perigee), “blood moon” (lunar eclipse), and meteor showers are of interest not only to astronomical society members but also to the general public. Festivals and local carnivals at observational destinations are often organized for larger audiences. As findings from this study show, the target market for astro-tourism and other SIT are not limited to those with a special interest in such activities but can also be extended to the general public. Thus, special activities and celebrations around celestial phenomena are likely to reach broad market audiences at substantial distances from observational places if attractions and products are purposely developed. From the perspective of demand, the large opportunist tourist base has high similarity to mass tourists seeking “novelty” yet with concerns for risks. Therefore, similar to many tourism products, “novelty” is a factor and a selling point in promoting SIT to potential opportunists.

It has been argued that not all locations are qualified to be astro-tourism destinations as astro-tourism is nature-based and requires clear night skies and low levels of light pollution (Fayos‐Solá et al., 2014). Thus, the argument states that astro-tourism has a better chance to develop in rural areas and could be economically and environmentally beneficial to these regions (Fayos‐Solá et al., 2014; Rodrigues et al., 2015). Our findings from the 2017 Great American Solar Eclipse observation, while not invalidating those and similar recommendations, pointed in another direction. This large-scale study of actual tourist behavior shows that the most visited observational destinations are in populated and/or urbanized areas (Table 2), and the tourist flows to these locations consists mainly of opportunists who are drawn to the infrastructurally-developed areas with seemingly greater comforts and lower risks. Thus, this study significantly expands the scope of locations that actual tourists view as suitable and, consequently, have a more positive outlook on the marketing potential of the "less than ideal" destinations for astro-tourism.

Topologically, eclipse viewing travel destinations form a line, and the observational locations of the opportunist astro-tourists generally follow this line. For the hardcore eclipse-chasers, however, the majority of activities occur at a finite number of locational clusters (Appendix Figure A2). The activities are concentrated around only three major hotspots: (a) a rural area south to Portland; (b) a rural area close to Idaho and Wyoming border, covering Caribou-Targhee National Forest and (partially) Yellowstone National Park; and (c) a low population density area located between Nashville, TN and Evansville, IN. All three locations are thinly populated and have a flat landscape such as seen on Figure 6. Such choices of landform are consistent with Wen's (2017) findings that plains and deserts are the preferred destinations for highly involved astro-tourists. The network in Figure 4 illustrating distances traveled by eclipse-chasers implies that the three hotspots are the preferred destinations of hardcore astro-tourists. These locations can potentially be developed into astro-tourism destinations with minimum facilities. Even without eclipse opportunities, these are likely to be suitable for other astronomic observations owing to their premium positions.

Figure 6. A photograph of a “true” eclipse observational site near the Idaho – Wyoming border. The eclipse-chaser was driving from San Francisco, a 15-hour drive. Courtesy of Irina Delusina, UC Davis.

6.3 Limitations and future research

For a balanced appreciation of the study findings, we would like to emphasize two points. Studies that use social media data are often criticized for their limited generalizability due to the fact that various social media platforms are designed for specific target audiences (Oteros-Rozas, Martin-Lopez, Fagerholm, Bieling, & Plieninger, 2016). The social media platform used in this study, Instagram, has a very wide reach (60% of mobile users are accessing Instagram – Statista, 2018), outperforming Flickr in representativeness (Tenkanen et al., 2017) , which positively affects generalization of the Instagram-based studies.

While the dataset size was quite large in this study, it is worth noting that a portion of data points was dropped due to unsuccessful origin location identification. We do not think that this data reduction affects the identified typology due to it relatively small size (17%). However, the very fact that there are non-identified locations requires further investigation into the nature of those tourists. The unsuccessful identifications are likely to happen in two scenarios: (1) users are in constant mobility with no stable residence locations being identified or (2) users are frequently present in two or more locations separated by sufficiently large distances. These tourists might represent a new segment of SI tourist (frequent traveler and long-distance commuter); therefore, devising a way to capture the digital footprint and spatial travel trajectory of these tourists in an extended time window is an item on the SIT tourism research agenda.

In conclusion, the most significant contribution of this study is its successful integration of the tourist typology theoretical framework with big data analytics. The typology theories and the place that SIT occupies in the tourist domain were put to an empirical test with social media data. The results are in agreement with the previous argument (Agarwal et al., 2018) in which SIT nowadays has a significantly larger mainstream audience (the *opportunists*) or alternatively, that the concept of mass tourism has expanded its frontier to absorb some tourist activities traditionally classified as belonging to the SIT domain. True SI tourists (the *hardcore*) still exist, yet their "market share" seems to be not as large as one may have assumed prior to the study. Further research is in order to verify this conclusion with other types of SIT (such as wine tourism or dark tourism) using social media as the data source and the approach to big data analysis implemented in this paper. In order to capitalize on the trend of the "traditional SIT" activities receiving more interest from the mass tourists, destinations can leverage their SIT resources as an effective marketing and promotion tool to approach wider audiences. From a theoretical angle, we urge tourist researchers to revisit the concept of SIT since the answer to the question that McKercher and Chan (2005) raised more than a decade ago: "How special is special interest tourism?" seems to be "the special interest tourism is not so special, after all".

REFERENCES

Agarwal, S., Busby, G., & Huang, R. (2018). *Special Interest Tourism: Concepts, Contexts and Cases*. https://doi.org/10.1016/s0261-5177(02)00020-1

Ali-Knight, J. (2010). *The Role of Niche Tourism Products in Destination Development*. Edinburgh Napier University.

Amaro, S., Duarte, P., & Henriques, C. (2016). Travelers' use of social media: A clustering approach. *Annals of Tourism Research*, *59*, 1–15. https://doi.org/10.1016/j.annals.2016.03.007

Arnegger, J., Woltering, M., & Job, H. (2010). Toward a product-based typology for nature-based tourism: A conceptual framework. *Journal of Sustainable Tourism*, *18*(7), 915–928. https://doi.org/10.1080/09669582.2010.485680

Balasubramanian, S., Gupta, S., Kamakura, W., & Wedel, M. (1998). Modeling large data sets in marketing. *Statistica Neerlandica*, *52*(3), 303–323. https://doi.org/10.1111/1467-9574.00086

Bartram, S. A. (2001). Serious leisure careers among whitewater kayakers: A feminist perspective. *World Leisure Journal*, *43*(2), 4–11. https://doi.org/10.1080/04419057.2001.9674225

Brotherton, B., & Himmetoglu, B. (1997). Beyond Destinations—Special Interest Tourism. *Anatolia: An International Journal of Tourism and Hospitality Research*, *8*(3), 11–30. https://doi.org/10.1080/13032917.1997.9687118

Burtnyk, K. (2000). Impact of Observatory Visitor Centres on the Public's Understanding of Astronomy. *Publications of the Astronomical Society of Australia*, *17*(3), 275–281.

Cater, C. I. (2010). Steps to Space; opportunities for astrotourism. *Tourism Management*, *31*(6), 838–845. https://doi.org/10.1016/j.tourman.2009.09.001

Cesario, F. J. (1976). Value of Time in Recreation Benefit Studies. *Land Economics*, *52*(1), 32. https://doi.org/10.2307/3144984

Charters, S., & Ali-Knight, J. (2002). Who is the wine tourist? *Tourism Management*, *23*(3), 311–319. https://doi.org/10.1016/S0261-5177(01)00079-6

Chen, Y., Parkins, J. R., & Sherren, K. (2018). Using geo-tagged Instagram posts to reveal landscape values around current and proposed hydroelectric dams and their reservoirs. *Landscape and Urban Planning*, *170*(July 2017), 283–292. https://doi.org/10.1016/j.landurbplan.2017.07.004

Cohen, E. (1972). Toward a sociology of international tourism,. *Social Research*, *39*(1), 164–182.

Cohen, E. (1979). A Phenomenology of Tourist Experiences. *Sociology*, *13*(2), 179–201. https://doi.org/10.1177/003803857901300203

Cohen, E. (1983). Social Psychology of Tourist Behavior. *Annals of Tourism Research*, *10*(3), 582–583.

Collison, F. M., & Poe, K. (2013). “Astronomical Tourism”: The Astronomy and Dark Sky Program at Bryce Canyon National Park. *Tourism Management Perspectives*, *7*, 1–15. https://doi.org/10.1016/j.tmp.2013.01.002

Crouch, G. I. (2009). The Market for Space Tourism: Early Indications. *Journal of Travel Research*, *40*(2), 213–219. https://doi.org/10.1177/004728750104000212

Crouch, G. I., Devinney, T. M., Louviere, J. J., & Islam, T. (2009). Modelling consumer choice behaviour in space tourism. *Tourism Management*, *30*(3), 441–454. https://doi.org/10.1016/j.tourman.2008.07.003

De Cantis, S., Ferrante, M., Kahani, A., & Shoval, N. (2016). Cruise passengers’ behavior at the destination: Investigation using GPS technology. *Tourism Management*, *52*, 133–150. https://doi.org/10.1016/j.tourman.2015.06.018

De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., & Yu, C. (2010). Automatic construction of travel itineraries using social breadcrumbs. *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia - HT ’10*, 35. https://doi.org/10.1145/1810617.1810626

Dey, B., & Sarma, M. (2006). Tourist Typologies and Segmentation Variables with Regard to Ecotourists. *Tourism Management*, *VIII*, 31–39.

Díaz-Martín, A. M., Iglesias, V., Vázquez, R., & Ruiz, A. V. (2000). The use of quality expectations to segment a service market. *Journal of Services Marketing*, *14*(2), 132–146. https://doi.org/10.1108/08876040010320957

Dolnicar, S. (2002a). A review of data-driven market segmentation in tourism. *Journal of Travel and Tourism Marketing*, *12*(1), 1–22. https://doi.org/10.1300/J073v12n01_01

Dolnicar, S. (2002b). Activity-Based Market Sub-Segmentation of Cultural Tourists. *2002 CAUTHE Conference*, *9*(2), 94–105.

Dolnicar, S., & Fluker, M. (2003). Behavioural market segments among surf tourists: Investigating past destination choice. *Journal of Sport and Tourism*, *8*(3), 186–196. https://doi.org/10.1080/14775080310001690503

Donaire, J. A. (2011). Barcelona Tourism Image Within the Flickr Community. *Cuadernos de Turismo*, *27*, 1061–1062.

Donaire, J. A., Camprubí, R., & Galí, N. (2014). Tourist clusters from Flickr travel photography. *Tourism Management Perspectives*, *11*, 26–33. https://doi.org/10.1016/j.tmp.2014.02.003

Douglas, N., Douglas, N., & Derrett, R. (2001). *Special interest tourism: Context and cases*. New York: John Wiley & Sons Australia.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proceedings*, *96*(34), 226–231. https://doi.org/10.1016/B978-044452701-1.00067-3

Fayos-Solà, E., Alvarez, M. D., & Cooper, C. (2014). Tourism as an Instrument for Development: A Theoretical and Practical Study. In E. Fayos-Solà, M. D. Alvarez, & C. Cooper (Eds.), *Tourism as an Instrument for Development: A Theoretical and Practical Study* (p. iii). https://doi.org/10.1108/s2042-144320140000006000

Fayos-Solá, E., Marín, C., & Jafari, J. (2014). Astrotourism: No Requiem for Meaningful Travel. *PASOS. Revista de Turismo y Patrimonio Cultural*, *12*(4), 663–671. https://doi.org/10.25145/j.pasos.2014.12.048

Fesenmaier, D. R., & Johnson, B. (1989). Involvement-based segmentation. Implications for travel marketing in Texas. *Tourism Management*, *10*(4), 293–300. https://doi.org/10.1016/0261-5177(89)90007-1

García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, *63*, 408–417. https://doi.org/10.1016/j.apgeog.2015.08.002

Girardin, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: A case study. *Journal of Location Based Services*, *2*(1), 41–56. https://doi.org/10.1080/17489720802261138

Greer, T., & Wall, G. (1979). Recreational hinterlands: A theoretical and empirical analysis. *Publication Series, Department of Geography, University of Waterloo*, *14*, 227–246.

Hall, M. (1992). Adventure, sport and health tourism. In B. Weiler & C. M. Hall (Eds.), *Special interest tourism* (pp. 141–158). London: Belhaven Press.

Hall, M. (2003). Spa and health tourism. In S. Hudson (Ed.), *Sport and adventure tourism* (pp. 273–292). https://doi.org/10.5363/tits.5.3_80

Heikinheimo, V., Minin, E. Di, Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey. *ISPRS International Journal of Geo-Information*, *6*(3), 85. https://doi.org/10.3390/ijgi6030085

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

Hvenegaard, G. T. (2002). Using tourist typologies for ecotourism research. *Journal of Ecotourism*, *1*(1), 7–18. https://doi.org/10.1080/14724040208668109

Järv, O., Tenkanen, H., Toivonen, T., & Hiippala, T. (2018). Revealing the origins of social media users: A critical evaluation of measuring techniques. *AAG 2018*. New Orleans.

Kim, H., & Stepchenkova, S. (2015). Effect of tourist photographs on attitudes towards destination: Manifest and latent content. *Tourism Management*, *49*, 29–41. https://doi.org/10.1016/j.tourman.2015.02.004

Kirilenko, A. P., Stepchenkova, S. O., & Hernandez, J. M. (2019). Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tourism Management*, *72*, 400–410. https://doi.org/10.1016/j.tourman.2019.01.001

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., & Andrienko, G. (2010). Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. *Proceedings of the International Conference on Information Visualisation*, 289–296. https://doi.org/10.1109/IV.2010.94

Lehto, X. Y., O'Leary, J. T., & Morrison, A. M. (2002). Do psychographics influence vacation destination choices? A comparison of British travellers to North America, Asia and Oceania. *Journal of Vacation Marketing*, *8*(2), 109–125. https://doi.org/10.1177/135676670200800202

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

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

MacKellar, J. (2009). Dabblers, fans and fanatics: Exploring behavioural segmentation at a special-interest event. *Journal of Vacation Marketing*, *15*(1), 5–24. https://doi.org/10.1177/1356766708098168

Malville, J. M. (2008). *A guide to prehistoric astronomy in the Southwest*. Boulder, CO: Johnson Printing Company.

Marzo-Navarro, M., & Pedraja-Iglesias, M. (2010). Are there different profiles of wine tourists? An initial approach. *International Journal of Wine Business Research*, *22*(4), 349–361. https://doi.org/10.1108/17511061011092401

Matos, A. L. (2017). *Terrestrial Astrotourism – Motivation and Satisfaction of Travelling to Watch the Night Sky*. Aalborg University.

Mavragani, E., Nikolaidou, P., & Theodoraki, E. (2019). Traveler segmentation through Social Media for intercultural marketing purposes: The case of Halkidiki. *Journal of Tourism, Heritage & Services Marketing*, *5*(1), 15–23.

McKercher, B., & Chan, A. (2005). How special is special interest tourism? *Journal of Travel Research*, *44*(1), 21–31. https://doi.org/10.1177/0047287505276588

McKercher, B., & du Cros, H. (2003). Testing a cultural tourism typology. *International Journal of Tourism Research*, *5*(1), 45–58. https://doi.org/10.1002/jtr.417

Middleton, V., & Clarke., J. (2001). *Marketing in travel and tourism*. Jordan Hill.

Mukhina, K. D., Rakitin, S. V., & Visheratin, A. A. (2017). Detection of tourists attraction points using Instagram profiles. *Procedia Computer Science*, *108*(June), 2378–2382. https://doi.org/10.1016/j.procs.2017.05.131

Nella, A., & Christou, E. (2014). Segmenting Wine Tourists on the Basis of Involvement with Wine. *Journal of Travel and Tourism Marketing*, *31*(7), 783–798. https://doi.org/10.1080/10548408.2014.889639

Neuts, B., Romão, J., Nijkamp, P., & Shikida, A. (2016). Market segmentation and their potential economic impacts in an ecotourism destination: An applied modelling study on Hokkaido, Japan. *Tourism Economics*, *22*(4), 793–808. https://doi.org/10.1177/1354816616654252

Nicolau, J. L., & Más, F. J. (2006). The influence of distance and prices on the choice of tourist destinations: The moderating role of motivations. *Tourism Management*, *27*(5), 982–996. https://doi.org/10.1016/j.tourman.2005.09.009

Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2016). Tracing Tourists by Their Digital Footprints. *Journal of Travel Research*, *55*(5), 566–573. https://doi.org/10.1177/0047287514563985

Oppermann, M. (1995). A Model of Travel Itineraries. *Journal of Travel Research*, *33*(4), 57–61. https://doi.org/10.1177/004728759503300409

Oteros-Rozas, E., Martin-Lopez, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2016). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators*. https://doi.org/10.1016/j.ecolind.2017.02.009

Pearce, P. (1982). *The Social Psychology of Tourist Behavior. International Series in Experimental Psychology (vol. 3)*. Oxford: Pergamon.

Phillips, J., & Brunt, P. (2013). Tourist differentiation: Developing a typology for the winter sports market. *Turizam: Međunarodni Znanstveno-Stručni Časopis*, *61*(3), 219–243.

Plog, S. (1974). Why Destination Areas Rise and Fall in Popularity. *Cornell Hotel and Restaurant Administration Quarterly*, *14*(4), 55–58. https://doi.org/10.1177/001088047401400409

Plog, S. (1987). Understanding psychographics in tourism research. *Travel, Tourism, and Hospitality Research. A Handbook for Managers and Researchers*, pp. 203–213.

Plog, S. (1991). *Leisure travel: Making it a growth market...again*. New York: John Wiley and Sons.

Pride, W. M., & Ferrell, O. C. (2016). Foundations of Marketing. In *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2418717

Robinson, M., & Novelli, M. (2005). Niche tourism: An introduction. In M. Novelli (Ed.), *Niche Tourism: Contemporary issues, trends and cases* (pp. 1–14). https://doi.org/10.4324/9780080492926

Robson, I. (2005). The role of the observatories. *The ESO/ESA/IAU Conference: Communicating Astronomy with the Public*, 60–70. Edinburgh.

Rodrigues, A. L. O., Rodrigues, A., & Peroff, D. M. (2015). The Sky and Sustainable Tourism Development: A Case Study of a Dark Sky Reserve Implementation in Alqueva. *International Journal of Tourism Research*, *17*(3), 292–302. https://doi.org/10.1002/jtr.1987

Ryan, C. (2003). *Recreational Tourism: Demand and Impacts*. Channel View Publications.

Shani, A., Wang, Y., Hutchinson, J., & Lai, F. (2010). Applying expenditure-based segmentation on special-interest tourists: The case of golf travelers. *Journal of Travel Research*, *49*(3), 337–350. https://doi.org/10.1177/0047287509346852

Silverman, B. (1986). *Density estimation for statistics and data analysis*. New York: Chapman and Hall.

Smith, S. (1999). How Far is Far Enough? Operationalizing the Concept of '"Usual Environment"' in Tourism Definitions. *Tourism Analysis*, *4*, 137–143.

Soleimani, S., Bruwer, J., Gross, M. J., & Lee, R. (2018). Astro-tourism conceptualisation as special-interest tourism (SIT) field: A phenomonological approach. *Current Issues in Tourism*, *0*(0), 1–16. https://doi.org/10.1080/13683500.2018.1444021

Statista. (2018). Most popular mobile social networking apps in the United States as of October 2018, by reach. Retrieved February 18, 2019, from https://www.statista.com/statistics/579334/most-popular-us-social-networking-apps-ranked-by-reach/

Stebbins, R. A. (1982). Serious Leisure: A Conceptual Statement. *Sociological Perspectives*, *25*(2), 251–272. https://doi.org/10.2307/1388726

Stebbins, R. A. (1997). Casual leisure: A conceptual statement. *Leisure Studies*, *16*(1), 17–25. https://doi.org/10.1080/026143697375485

Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management*, *36*, 590–601. https://doi.org/10.1016/j.tourman.2012.08.006

Straumann, R. K., Çöltekin, A., & Andrienko, G. (2014). Towards (Re)Constructing Narratives from Georeferenced Photographs through Visual Analytics. *The Cartographic Journal*, *51*(2), 152–165. https://doi.org/10.1179/1743277414Y.0000000079

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports*, *7*(1), 17615. https://doi.org/10.1038/s41598-017-18007-4

Težak, A., Saftić, D., Težak, A., & Bošković, D. (2011). Selection of segmentation criteria: Case of wellness tourism. *30th International Conference on Organizational Science Development-Future Organization*. Portorož, Slovenia.

Trauer, B. (2006). Conceptualizing special interest tourism—Frameworks for analysis. *Tourism Management*, *27*(2), 183–200. https://doi.org/10.1016/j.tourman.2004.10.004

UNWTO. (1994). *Recommendations on tourism statistics*. New York.

Vecchio, P. Del, Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from Social Big Data: Implications for Smart Tourism Destinations. *Information Processing and Management*, *54*(5), 847–860. https://doi.org/10.1016/j.ipm.2017.10.006

Wearing, S. (2002). Re-centring the self in volunteer tourism. In *The tourist as a metaphor of the social world* (pp. 237–262). https://doi.org/10.1079/9780851996066.0237

Weaver, D. (2011). Celestial ecotourism: New horizons in nature-based tourism. *Journal of Ecotourism*, *10*(1), 38–45. https://doi.org/10.1080/14724040903576116

Weiler, B., & Hall, C. (1992). *Special interest tourism*. Belhaven Press.

Wen, J. (2017). *Astronomy tourism: Exploring an emerging market: Group culture, individual experience, and industry future*. James Cook University.

Wongkit, M., & McKercher, B. (2013). Toward a typology of medical tourists: A case study of Thailand. *Tourism Management*, *38*, 4–12. https://doi.org/10.1016/j.tourman.2013.02.003

Yuan, Y., & Medel, M. (2016). Characterizing international travel behavior from geotagged photos: A case study of Flickr. *PLoS ONE*, *11*(5). https://doi.org/10.1371/journal.pone.0154885

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

Zheng, Y.-T., Zha, Z.-J., & Chua, T.-S. (2012). Mining Travel Patterns from Geotagged Photos. *ACM Transactions on Intelligent Systems and Technology*, *3*(3), 1–18. https://doi.org/10.1145/2168752.2168770