

Multidimensional Analyses of Social Media Related Geographic Information: a Study Concerning the Urban Area of Cagliari (Sardinia, Italy)

Roberta Floris

(Roberta Floris, University of Cagliari DICAAR, roberta.floris@unica.it)

1 ABSTRACT

The widespread diffusion of tourism social media platforms is playing an increasingly important role as information source both for tourists, who gather reliable information supporting destinations' choice and services from peers, and for businesses, who can use the same information for improving their marketing strategies. The use of this type of information could also offer new opportunities for decision-support in tourism planning. By means of improved understanding of the travel motivations and by tailoring tourism service supply, decision making can be facilitated by emphasising the strengths of tourist destinations for past and potential visitors. However, information about tourists' perceptions and opinions is not always properly analysed by planners. User satisfaction depends on factors related to the location and the services quality that the local industry proposes. Moreover, its understanding may offer valuable knowledge in tourism planning at the regional and local levels.

The goal of the study of the paper is to propose an integrated approach to investigate, qualitatively and quantitatively, the relationships between tourists' satisfaction, geographic locations and tourist enterprises in supporting decision-making processes. The methodology implemented in the study includes data collection from Booking and TripAdvisor.com and their integration with authoritative territorial data. Spatial and statistical analysis techniques are applied in order to assess tourists' perceptions on success factors, which may be used as planning support tools. The case study concerns the municipality of Cagliari and demonstrates the value of social media-related data integrated by authoritative information in tourism planning. Finally, the paper proposes a critical discussion on the effectiveness of using the implemented integrated approach in order to address other planning issues. The discussion underlines the potential of the proposed approach in order to address other planning questions as well.

2 INTRODUCTION

This paper focuses on the tourism, analysing the relationships between demand, industry and location, identified as fundamental variables and it aims to understand if the use of the Social Media related Geographic Information (SMGI) may offer knowledge bases for decision making in tourism planning. The investigations are carried on exploring the potential of public volunteered comments, represented by tourists' preferences on destination and tourism supply, for providing useful knowledge about these inclinations in space and time.

For research purposes, a traditional method for collecting information about such preferences, performed via ad-hoc surveys, can be expensive and time consuming. For this reason an alternative approach is presented, by which tourist preferences are discovered by processing and analysing public available social media data. In addition, the investigation aims to demonstrate the opportunities of SMGI as support for design, analysis and decision making in tourism planning, and the consequent value to inform regional and local initiatives.

The adopted methodology deals with several questions related to tourist preferences:

- (1) Which are the most popular destinations?
- (2) Why people chose those destinations?
- (3) What attracts tourists' attention and what do they appreciate or disregard?
- (4) How this knowledge can be use in supporting tourism planning approaches?

This kind of study and the methodology adopted, which couple SMGI and A-GI from open SDI, may provide a novel kind of information, which may be integrated with the traditional knowledge, and successfully used in urban and regional planning as much as in tourist planning, for in both cases they contribute to take into account a multifaceted customers- (or citizens) –oriented view on strategic development issues. In addition, using SMGI may disclose opportunities for further analysis scenarios in urban and regional planning, and may offer useful suggestions for tourism planning strategies. In an integrated planning support framework, SMGI analytics might help to understand tourists' observations,

preferences, interests, feelings, or needs, and possibly affect decision-making dynamics and planning processes with customer oriented strategies.

The method builds on a set of spatial analysis and statistics techniques, useful in describing and visualizing the spatial distribution and detecting patterns and hot-spots. In addition, textual analytics techniques (Campagna et al., 2013; Massa and Campagna, 2014) have been applied in order to discover the knowledge enclosed in the huge amount of qualitative social media comments. The findings could provide important insights into the Sardinia tourism industry, which could be used to develop innovative planning approaches. They could also offer a benchmark for future comparative trend analysis and directions for tourism policy design. After examining the past studies on the travel consumers' online social networks and the most popular web sites (Section 3), the paper focuses on the destination choices and judgments represented in numerical and linguistic terms (Section 4). In Section 5 the research methodology and the results of the case study, concerns the municipality of Cagliari, are summarized and briefly discussed. Section 6 concludes the paper discussing obtained findings and future research perspectives.

3 ESTINATION CHOICE, TOURISTS' PERCEPTION AND CURRENT DEVELOPMENTS IN TOURISM

Tourism is one of the largest industries all over the world and is based on multifaceted activities, which may generate both positive and negative impacts within the social, cultural and environmental domain. According to Smith (1991), as a complex activity, tourism comprises the travel to and around a destination, with the purpose of exploiting particular natural or no-natural attractions, accommodation and specialized and general services. These types of resource have been classified by Jafari (1982) into background tourism elements and facilities services. For this reason, tourism has commonly been recognised both as spatial (Peroni, 2007; Cooper, 2008) and soil consumption (Boccagna, 2010) phenomenon. This fact becomes more evident when tourist activities are not adequately developed and planned. Using programmatic approaches and empathising the sense of place and identity, could help to develop more effective planning strategies for sustainable development based on tourism, in order to reach services quality goals, preserving the natural resources, the cultural heritage and life quality of the host communities (Briassoulis, 1992).

In the last decade, the fast evolution of ICTs enabled users to ubiquitous access a broad range of information services. The degree of interactivity, established by the Web 2.0 paradigm, enhanced the role of the Internet as information source, with a secondary role as opinion source (Grabner et al., 2012). Tourism is one of the sectors where the wide exploitation of ICTs leads to tourist online communities' development. Travel topics are among the most popular in on-line social networks (Buhalis and O'Connor, 2005; Baggio et al., 2008).

In the light of the above premises, this section focuses on three main topics: (1) the current developments in tourism planning and tourism destination, (2) the role and the support of customers' preferences in planning processes and (3) a review of the evolution of the e-Tourism phenomenon.

3.1 Destination choices and travel motivation

Determining the factors that influence tourists' choice for a destination is essential in developing appropriate marketing strategies and planning approaches in tourism field. Most studies of tourists' preferences address destination choices as the key element in the travel decision-making process. According to Dellaert et al. (1998) this element is combined with accommodation or activity choices. On the one hand, the investigation of decision-making processes, mostly conceptual in nature, focuses on the types of decision rules and the decision-making stages that are likely to be adopted by tourists. On the other hand, research in choice factors has been primarily addressed with empirical examinations of critical attributes, used by tourists as criteria for determining their travel alternatives (Crompton, 1979).

Knowledge of consumers' psychology is extremely important in determining the success of a destination (Rodriguez del Bosque and San Martin, 2008). The exploration of psychological concepts such as attitudes, decision-making processes, emotions, experience and satisfaction is necessary for understanding customers' choices and preferences in tourist destination. The destination consists of a well-defined geographical area, such as a country, a region or a city (Hall, 2000) that it can be referred to a product or a brand (Yoon and Uysal, 2005). Many studies on tourists' travel choice distinguish among various approaches to the definition of tourist destination. Van Raaij (1986) defined the travel destination as a product, which is partly given and partly man-made. The given part refers to the natural features of a destination, such as climate, landscapes,

beaches, mountains and historic-culture buildings. The man-made part refers to physical features such as hotels, package tours, transportation facilities, sports and recreation facilities, which can be modified to satisfy customers' preferences and their budget. Ferreira (2011) claims that tourist destinations should be conceived as brands and they should be managed from a strategic point of view.

Travel motivation is a dynamic concept, that changes from one person to another and from one destination to another (Kuang Hsu et al., 2009). One of the most popular conceptual model for understanding travel motivation is the push and pull model by Crompton (1979). The push motivations are useful in order to explain the desire for travel while the pull motivations explain the actual choice of destination. Moreover, the Crompton model identifies seven socio-psychological (push) motivations (escape, self-exploration, relaxation, prestige, regression, kinship-enhancement, and social interaction) and two cultural (pull) motivations (novelty and education). Uysal and Jurowski (1994) further developed Crompton model and summarized internal (push) and external (pull) motivators to travel. Internal motivators include desire for escape, rest, relaxation, prestige, health and fitness, adventure, and social interaction. External motivators are based on attractiveness of the destination, including tangible resources (i.e. beaches, cultural attractions and recreational activities), and travellers' perceptions and expectations (novelty, benefit expectations, and marketing image). In more recent studies, researchers added shopping as a motivational factor in destination choices (Uysal and Jurowski 1994; Cooper, 2008).

3.2 Tourists' perceptions and preferences

Most studies of tourists' travel choices address tourist destination choice as the key element in travel decision-making. This process is influenced by a number of psychological (internal) and no psychological (external) variables, and consists of a number of different stages that are marked by specific actions (Uysal and Jurowski, 1994). Consumers' judgments basically depend on the strength of their beliefs or expectations about the quality of various features or attributes associated with services. Personal preferences, like motivations, may be both intrinsic, reflecting individual likes and dislikes, and extrinsic or socially conditioned (Crouch et al., 2004). The weight of an attributes is usually related to the relative importance that consumers confer to each attribute. This means that each opinion strictly depends on tourists' direct past experiences with other services of analogous nature. Tourists' decisions are complex and multi-faceted in which the choices for different elements are interconnected and evolve in a decision process over time (Kuang Hsu et al., 2009). In order to meet the target of tourists' expectations, hotels should provide an ample range of quality services, including reception, meals, room service, tennis courts, beach nearby location, swimming pools and gardens, among others.

But how can we define service quality? Service quality can be considered as a composite measure of various attributes. According to Dubè and Renaghan (1999) it consists not only of tangible attributes but also of intangible or subjective attributes such as safety or quietness, which are difficult to measure accurately and which are usually studied by linguistic information (Benitez et al., 2007). Lewis and Booms (1983) define service quality as a measure of how well the service matches customers' expectations. The quality perceived by consumers in a service is a function of the magnitude and direction of the gap between expected service and perceived service (Berry et al., 1990; Benitez et al., 2007). Judgments expressed by numbers are easy to interpret, but linguistic information is more difficult to measure through a mathematical function. Linguistic information characterizes subjective knowledge and is usually ignored by analysts when forming mathematical models that represent real world phenomena. However, attributes measuring service quality are characterized by uncertainty, subjectivity, imprecision and ambiguity (Benitez et al., 2007). When consumers make decisions, they usually employ this subjective knowledge and linguistic information.

Beside tourism marketing and planning, tourists are an important target audience for urban planning: in order to take into account tourists' preferences, planners must deeply study the phenomenon of tourism and attempt to understand and internalize tourists' needs and perceptions (Dickey, 2005). An accurate identification of customers' perception is the first step to maintaining the status of a city image as a popular travel destination.

3.3 Revolution in tourism: the role of Information and communication technologies

In the last decade, the diffusion of the Information and Communication Technologies (ICTs) have revolutionised the travel industry. As part of service sector, tourism has not surprisingly been associated with

developments in new technologies and restored by organizational and structural innovations (Stamboulis and Skayannis, 2003). In the competitive scenario of tourism sector, any location or business aiming to do better than others, should become either a learning industry. Even more, emerging alternative tourism has to engage the element of culture, which improves in importance and has to be constantly transformed.

The e-Tourism advent reflects the digitalisation of all processes and value chains in the tourism, travel, hospitality and catering industries (Buhalis and Jun, 2003). It emerges as a term describing the entire range of applications of the ICTs on tourism and the implications for the tourism value chain (Buhalis and Deimezi, 2004). These changes are particularly evident in the way that tourism organisations communicate with their individual and institutional clients and how they manage their distribution function (Buhalis, 1998). These developments present a major opportunity for innovative tourism organisations and destinations, in order to improve their relative position in the international market (O'Connor, 1999). New sophisticated and demanding travellers require interacting with suppliers so as to satisfy their own specific needs (Buhalis, 2008). The reaction to online inquiries can thus influence customer satisfactions and booking behaviour. As a consequence, response behaviour becomes a crucial factor for the success of tourism enterprises (Pechlaner et al, 2002).

Finally, the Internet enables travellers to access to transparent and easy to compare information on destinations, holiday packages, travel, lodging and leisure services, as well as about their real time prices and availability. consumer generated content, through review portals such as TripAdvisor.com, multimedia sharing such as Panoramio.com, and blogs also create accessible content that increase the level of information available on a global basis (Buhalis and Jun 2003). These facts enhanced the role of the Internet as information source, with a secondary role as opinion source (Grabner et al., 2012). In addition with developments in social media every piece of information, can be commented or rated in some way. Tourists could have ubiquitous access to a broad range of tourist services and also be the producers and broadcasters of information on destination-based social networks. These capabilities have disclosed innovative opportunities for tourists to become themselves users of travel social networks.

4 DEVELOPMENTS IN TOURISM SOCIAL MEDIA

In the age of Digital Information (DI), planning considerations highlight the need to use advanced tools to produce, gather and manage spatial information collected into new digital formats and available to support planning processes. Online Social Networking (SN) sites are the most popular sites on the Internet. The second generation of web-based services is characterized by having consumer-generated contents, which allow people to share information. This type of information, available on forums and reviews, is generated by users/consumers and provides relevant data for travel planning. Hence, after the early introduction of Geographic Information Systems (GIS), we are now facing a knowledge revolution in spatial planning, design and decision-making, for opportunities in knowledge building are unprecedented: not only Spatial data infrastructures (SDIs) give access to a wealth of Authoritative geographic information (A-GI), but also Volunteered GI (VGI) and Social Media related Geographic Information (SMGI) sources are growing exponentially at fast rates: integrating these sources together can substantially improve both our understanding of environmental and social ecosystems, and of people perceptions, insights, and needs. However, while A-GI entails more traditional representations and analytical models, VGI and SMGI for the impetuous modes of production feature Big Data (BD) nature.

In the light of the above premises, this section deals with the opportunities of SMGI as valuable support for analysis, design and decision-making in tourism planning practices at both local and regional level. Moreover, this section presents a brief overview of current available tourism platforms (oriented to data creation, dissemination and collection) and describes opportunities and weakness that still limit their integration into planning practices.

4.1 Social Media related Geographic Information

Social networks are online communities of people who share common interests and activities (Miguens et al, 2008). They provide a user with a collection of various interaction possibilities, ranging from a simple chat, to multiple video conferences, and from the exchange of plain email messages to the participation in blogs and discussion forums. Consumers have ubiquitous access to a broad range of information services and visit the Internet to look for information and communicate or simply spend time and shop (Buhalis 1998; Chung

and Buhalis, 2008). Increasingly, it is evident that people meet online to express views, share information and often keep online blogs. The development of a shared knowledge base is the driving force of an online community. Social, economic and technological aspects are incorporated into the community (Fernback; 1999; Gleave, 2009). In addition with developments in social media every piece of information, can be commented or rated in some way. Social media employ mobile and web-based technologies to create highly interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content (Kiezman, 2011). Social media also offer different ways for management, sharing and extraction of contents, provoking a degree of uncertainty for the knowledge processing. By contrast, with the traditional geographic information, SMGI regards users' perceptions on the Earth surface related to a specific period of time and requires advanced tools to support real-time monitoring, analysis and decision-making. SMGI platforms can be used both for leisure and for more professional purposes, ideally allowing for the integration and sharing of all the resulting information streams.

This information has been, and still is collected through effective data acquisition techniques such as global positioning systems (GPS), high-resolution remote sensing, location-aware services and surveys, and Internet-based volunteered geographic information (Guo and Mennis, 2009). In addition SMGI may be geocoded in different ways, using either the position of the author (if public), or the location of the post (i.e. recorded through a GPS sensor of a mobile device if available), or through toponyms parsing in the text.

But the central challenge is how to manage a big amount of data and to find efficient methods to extract useful knowledge from spatial data sets of unprecedentedly large size and complexity. To address these challenges, spatial data mining and geographic knowledge discovery has emerged as an active research field, focusing on the development of theory, methodology, and practice for the extraction of useful information and knowledge from complex spatial databases (Andrienko and Andrienko, 1999; Miller and Han, 2009; Guo and Mennis, 2009). Spatial data mining and geographic knowledge discovery is an iterative process that involves multiple steps, including data selection, cleaning, pre-processing, and transformation. These methods are exploratory in nature, more inductive than traditional statistical ones, including clustering, classification, association rule mining, information visualization, and visual analytics (Miller and Han, 2009). Their goal is to integrate and further develop methods in various fields for the analysis of large and complex spatial data.

TripAdvisor.com and Booking.com are among the most popular platforms of the latter kind. They play a significant role in the online tourism market. They can be considered as market-driven social media. While on the one hand, these platforms represent an important marketing channel through which destinations and tourism enterprises can reach and persuade potential visitors (Biassoulis, 2002), on the other hand they assist consumers in posting and sharing their travel-related comments. Travellers opinions and personal experiences based on reconstruction of their trips in turn serve as information to others.

When location also is available, all these type of information, as all information derived from forums, discussion blogs or social network, could be considered as VGI. In recent years the term VGI became popular to indicate the avalanche of information which every second is shared on the web by users acting as sensors (Goodchild, 2007). According to Sui and Goodchild (2011), more recently the convergence of GIS and social media granted by interoperability of geo-web tools is further enriching the possibility of sharing the knowledge not only about the Earth surface but also about all the biological, social and cultural phenomena there happening. In facts, as Campagna et al. argue (2013), VGI may include both geographic information collected by groups of people within crowdsourcing initiatives and geo-tagged multimedia collected for personal purposes by the Internet users and publicly shared through archives in the cloud. Social media information may be geocoded in different ways, using either the position of the author (if public), or the location of the post (i.e. recorded through a gps sensor of a mobile device if available), or through toponyms parsing in the text. VGI has been proven useful in many application contexts such as emergency response, environmental monitoring and spatial planning (Poser and Dransch, 2010).

5 MULTIDIMENSIONAL ANALYSES OF TOURISM SOCIAL MEDIA INFORMATION

The methodological approach builds on a preliminary exploratory analysis of social networks contents of Sardinia searching for the most popular destination, the relationships between service quality and location and the spatial distribution of tourists' preferences at regional and local level. the purpose of this research is to use geographic information data, in particular both Authoritative Geographic Information (A-GI) and

Social Media related Geographic Information (SMGI). Combined A-GI and SMGI data are used to express location, while service quality is investigated through the SMGI support.

The analysis framework is provided for two scale (regional and local levels) and two dimensions (services quality and locations) using both A-GI and SMGI. First of all, analyses at the regional scale is implemented to describe spatial patterns of tourists' preferences and to identify locations of interest; the latter may include clusters of positive or negative preferences, or individual spots of interest. Afterwards, at the local level (i.e. within the single cluster or spot of interest) further analyses were developed aiming at understanding the possible reasons underneath the detected patterns and singularities, with the assumption that they may help in explaining success or failure factors with regards to destinations and services features. At the regional and the local levels, an investigation method is adopted, including descriptive spatial analysis and spatial statistics techniques coupled with explanatory SMGI analyses, encompassing Spatial-Temporal Textual analysis, which can be defined as the textual analysis constrained by space and time boundaries (STTx; Campagna, 2014).

Operationally, the study was carried on according to the following workflow:

- (1) data collection and geocoding; data were extracted by Booking.com and TripAdvisor.com, geocoded and integrated in a geodatabase for analyses;
- (2) regional preferences dynamics analysis; data were analysed for the entire region at the municipal unit of analysis with spatial analysis, spatial statistics, and STTx; in order to detect clusters and hot/cold-spots;
- (3) local preferences dynamics analysis; data are integrated with authoritative information from the regional SDI and other official open data sources, in order to find explanatory hints on the tourist preferences dynamic and to get deeper insights on the relationships among these preferences, the local territorial features and the quality of the industry services in selected destinations;
- (4) geographically weighted regression (GWR), to investigate how the detected patterns spatially changes within a particular place. This cartographic approach may hinder the exploration of spatial non-stationarity by inadequately illustrating the spatial distribution of the sign, magnitude, and significance of the influence of each explanatory variable on the dependent variable.

5.1 Data collection e geocoding

In the first step of the methodology, a database is created extracting data from TripAdvisor.com and Booking.com, in the period between May 2012 and May 2013. Through these applications customers can book, rank and review hotels and restaurants (or Tourism Services, TS). The focus of the portals is to filter contents based on rankings, which derived from other users' ratings. Thus, rankings are split into several categories, such as value/price, rooms, location, cleanliness and sleep quality. Available rating categories, however, are determined by the type of reviewed item. The reviews are enriched by the possibility to add multimedia elements or travel maps of previous trips or to take part in discussion forums. Thanks to the availability of the services location, they can thus be considered SMGI (Campagna, 2014).

The big issue is to manage this big amount of information. Thus, the study required the adoption of a mixed methods approach, in which quantitative information, concerns the score of tourist evaluation criteria, and qualitative information, includes customers textual descriptive review, were collected in a database for analyses. It should be noted that in TripAdvisor.com the rating scale consists of five ordinal values (or stars), ranging from 'terrible' to 'excellent'. A separate mandatory overall rating summarizes the total customer satisfaction. In Booking.com a rating scale consists of numerical integer ordinal values, ranging from 1 to 10 (i.e. the higher the better). Beside quantitative assessment, in both platforms, a text box record allows to qualitative natural language reviews. The title is a concise short text formulation of the assessment, while the comment is a long text field.

Geocoding is performed on the extracted addresses, exactly providing the place of tourism operators, in order to spatially analyse the location of the tourism business patterns in Sardinia. Point locations are found automatically for around 80% of the items and approximately 20% of the geocoding required manual editing.

As working set, a unified database of 992 records is used. The records provide TLS name, category, location, and related quantitative score, and include not only hotels, but also other types of accommodation such as resort, Bed and Breakfast or agritourism.

The analysis results revealed that the spatial distribution of the customers review on the TLS in Sardinia is divided into five main categories: agritourisms (6%), Bed and Breakfast (15.7%), hotels (42%), private accommodations (29%), residences and resorts (7.3%) of total numbers of operators). In addition three provinces have emerged as important tourist destinations in the tourists' perception: Olbia-Tempio (27,8%) Sassari (24%) and Cagliari (20,6%). Other four provinces (Nuoro 8,6%, Oristano 7,3%, Ogliastra 5%; Carbonia-Iglesias 4%) are well represented by tourism businesses; whilst the province of Medio Campidano is only represented by the 3% tourist enterprises. Finally analysis of the significance of tourist appreciation in the coast and in inner areas in Sardinia revealed that 92% of tourism reviews sample concerns TLS (917) in the coastal areas, while only less than 8% of popular tourism businesses are found inland. Nevertheless, Nuoro and Medio Campidano provinces together provide notable inland popular TLS with almost the 13% of the total number of reviewed tourism businesses. This may mean that tourists visit these areas to discover a less popular side of the island, which is characterised by its nature, cultural heritage and traditions. However, in terms of number of tourists, this kind of tourism still does not compete with massive preferences for tourism along the coastal area.

5.2 Regional preferences dynamics analysis

After the preliminary descriptive analyses of the preferences dataset, the second step of the methodology concerns the application of spatial analyses of tourist preferences, to explore spatial patterns of positive judgments at the regional level.

The application of spatial analytical techniques allows the exploration of the spatial dynamics of tourists' perception and their relationships with other territorial variables. For each TLS the database includes a score record, which is the average of six main attributes:

- (1) location, which is related to the geographic position of the structure;
- (2) services, referring to all transport facilities, shopping areas, bars and restaurants;
- (3) price/quality ratio, referring to structure cleanness, staff kindness and all type of comforts offered by the operators;
- (4) staff (kindness);
- (5) room cleanness (cleaning);
- (6) comfort, referring to all facilities and services that hotels provide to their customers.

The attributes location and services explain the territorial features of the destination, while others express the perceived quality of the TLS supply. Thus the data model allows investigating the spatial patterns of preferences on territorial and tourism industry features at the local level across the whole region.

Figure 1 shows the distribution of the TPPI (left). The TPPI shows an overall high spatial concentration in the North-East of Sardinia. The Costa Smeralda district appears as the only area where the global tourism preferences fulfil overall tourists' expectations. Looking at individual municipalities, the analysis shows that Alghero exposes the highest TPPI rate. The other two municipalities with a high TPPI are Cagliari and Olbia.

The purpose of this map is to identify surfaces already affected by the phenomenon, but also potential development areas. It is notable that most of the municipality located into the coastal area attracted the attention of the participants, while the mostly inland area represents a cold spot. Is notable the presence of three major clusters representing the spatial distribution of the TPPI within Sardinia.

The first cluster is located into the Nurra district and includes Sassari, Stintino and Alghero municipalities. The presence of this cluster is probably due to the presence of Alghero, unique municipality which preserved Catalan tradition. The second cluster is placed in Cagliari metropolitan area. This is one of the Sardinia economy point, and due to its strategic location has become one of the most popular tourism destination.

Finally, the last cluster covers a large part of the Gallura district, which represents the ancient tourism area for the presence of the Costa Smeralda, which started to be recognised by the foreign market from the sixties thanks to Prince Aga Khan and his massive investment in this area. The Costa Smeralda has been aimed at an élite type of tourism; this should be the reason why international tourism starts to increase. Is notable the spatial continuity between the old tourism areas and the young one. In addition the map on the right shows the pattern of the negative judgments.

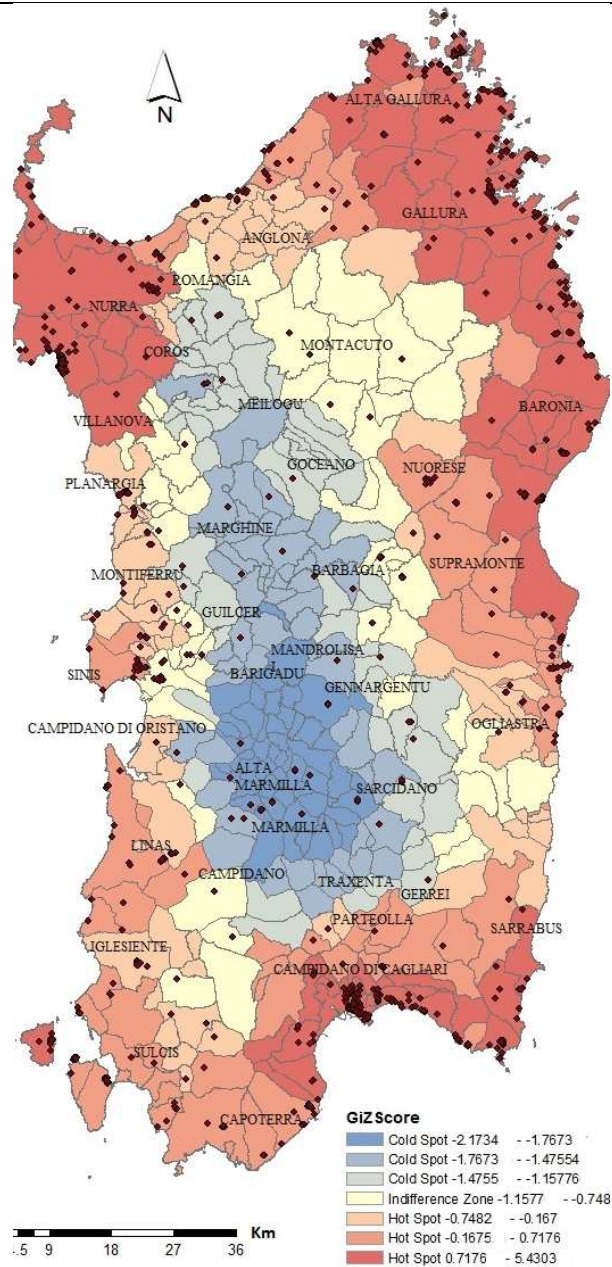


Figure 1 Cluster of positive tourist preferences (TPPI)

5.3 SMGI analytics at the local level:

After the analysis of tourism dynamics at the regional level, identifying clusters and spots of successful destinations through the preference patterns, the methodology adopted for this study has shifted to the local scale for further analyses aiming at finding explanatory answers for the phenomena under observation. The shift from the regional to the local scale is also conducted relying on spatial analysis and spatial statistics techniques on an integrated SMGI/A-GI data database. Thus, the following questions one should answer were:

- (1) Why tourists' interest concentrates in this destination?
- (2) What exactly in the destination does attract the tourists' attention?
- (3) Why tourists chose those destinations?

The aim is twofold: on the one hand the study is performed to discover why tourists prefer some destination rather than other at the regional scale (qualitative analysis), while on the other hand to investigate quantitatively why tourists' preferences are located in these areas and what factors contribute to higher TPPI rate. As a case study for the sake of illustrating the methodology steps, the tourist destination of Cagliari have been chosen as the regional analyses demonstrated its highly successful performance. The SMGI

analytics are intended to investigate the success factors within this destination in order to extract useful hints to be used for further planning in the same or other destinations.

Cagliari, the capital of Sardinia is located in the South of the region. It is nationally and internationally well connected thanks to the airport and the marina. Cagliari is the most important airport in the island in terms of traffic and size. In fact, it operates about 50% of Sardinia air traffic and can serve up to 4 million passengers per year (Benedetti et al., 2012). This destination has been recognized as one of the best-selling destination from different tourists' typologies. Thus, the following questions one should answer were Why tourists interest concentrates in Cagliari? and What exactly in the destination does attract the tourists attention? In order to answer these questions, summarising the review by neighbourhoods, the map in Figure 5 shows the spatial clusters of preferences:

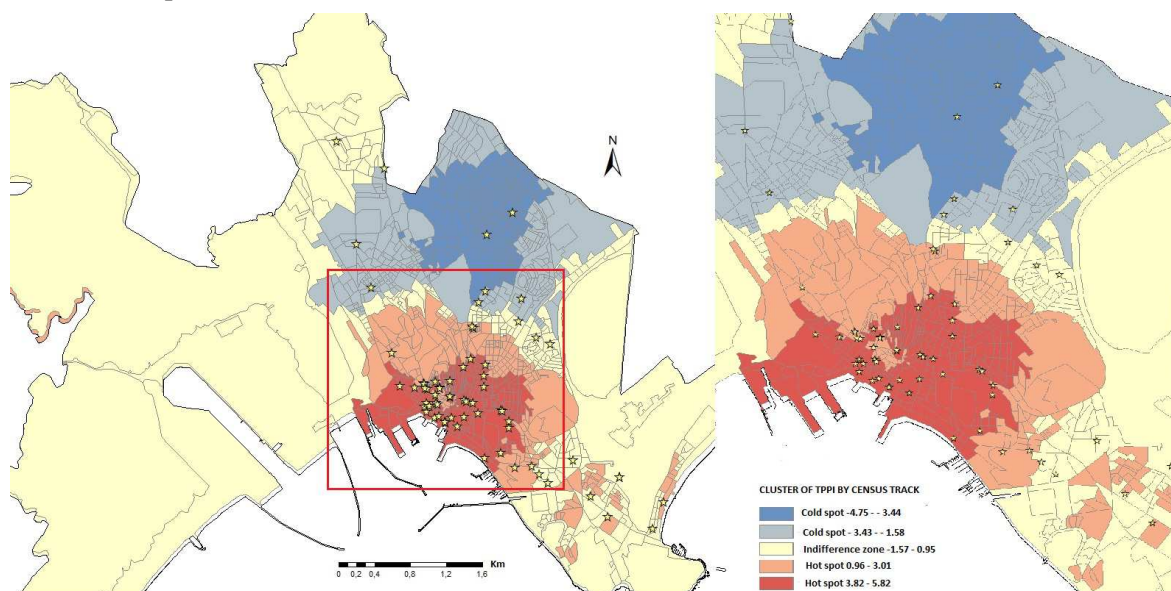


Figure 2 Significant patterns in Cagliari municipality

The location of each TLS could immediately allow detecting where the attention of the tourists who visited Cagliari has been focused on; thus, is possible to answer such question as what areas, places, or artefacts in the city attract the user attention. Spatial clusters of preferences are detected by hot-spot analysis. Firstly, a threshold distance of 1700 meters has been calculated and the spots by census track summarised: the map in Figure 2 shows the spatial distribution of TPPI: red areas, located in city centre, indicate surfaces affected by high concentration of TPPI phenomenon (hot spot), while blue areas represent locations where the phenomenon is less intense (cold spot). Not surprisingly the Cagliari city centre attracted the main attention of the tourists, while the mostly residential Pirri district represents a cold spot.

| Category | WORDS [frequency] |
|-----------------------------------|--|
| Geographic location | location [1010]; town [476] |
| Services | staff [890]; restaurant [643]; room [459]; hotel [469]; pool [230]; food [180] |
| Accessibility | minutes [250]; harbour [237]; proximity [164]; walking [146] |
| Natural and no natural components | city centre [426]; beach [378]; old city [132] |

Table 1: Top 15 words related to Cagliari divided by category

The next step focuses on the reviews content in order to understand not only where but also what tourists think about Cagliari. Hundreds of textual reviews only located into detected spatial pattern have been investigated by STTx analysis. The possibility to analyse tourist preferences may help to investigate the move in general spatial interest pattern in order to detect possible useful hints to be used for further planning in terms of tourism development at local scale. Very interesting results have been obtained applying STTx to local subsets of data obtained by selecting high TPPI values by location. Here the tag cloud (Table 1) clearly shows the focus of the majority of the words in the posts refers to spatial or physical aspects of the city such as centre, city, location and church. The results includes keywords related both to leisure sites such as restaurants, shopping and dinner and services such as stuff and room. Additionally, outcomes indicate a high level of satisfaction with the places accessibility: words as minutes, proximity and walking could be related to the services' spatial location, natural resources or monuments. In this sense the capacity of tourists to

effortlessly move from a place to another and to affect same areas of interest generate a positive tourist destination image. In addition different sectors within the local community could benefit by the presence of tourists to different degrees. Business sectors are more likely to hold favourable views of tourism because of the economic benefits that the industry is perceived to bring.

Nevertheless, local residents could have negative views, especially where their lives are affected by the noise, overcrowding and overuse of facilities. This is not the kind of information we usually find in land use planning documents, but his potential for support design and decision-making may be highly valuable.

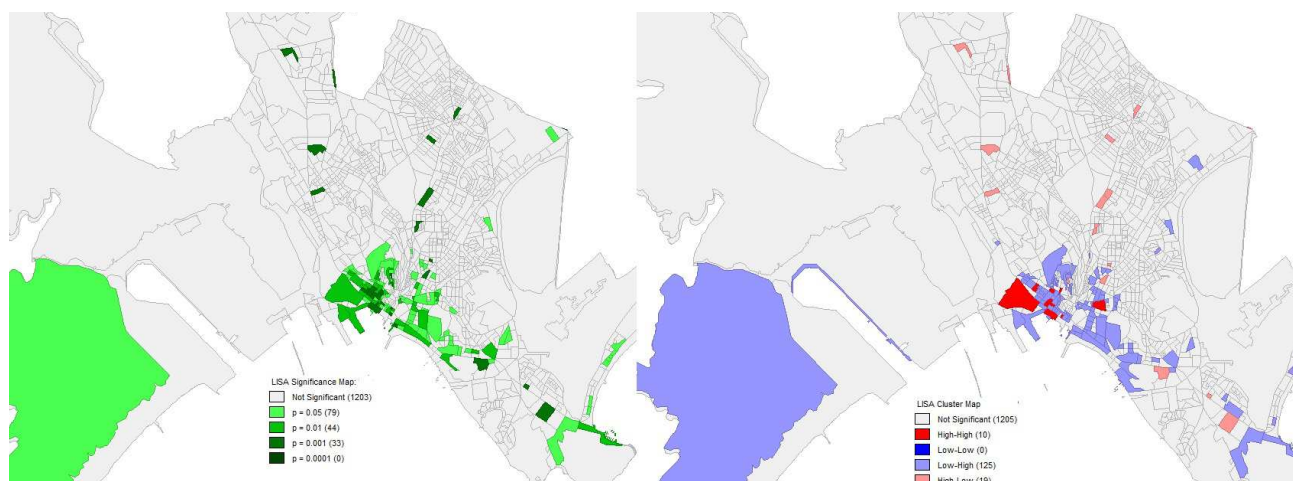
5.4 Geographically weighten regression

Lastly, the study was also supported by the integration of SMGI with other A-GI describing topography, transport infrastructures, cultural heritage sites, and socio-economic feature. The spatial relationships and the explanatory factors behind observed spatial patterns were modeled using GWR (Fotheringham et al., 2003, p 9). The aim of the GWR is twofold: the analysis is performed to investigate quantitatively why visitors' preferences were located in Alghero rather than in other destinations, and to discover what factors contributed to the Cagliari high TPPI rate. The model is applied to a sample of 150 TLS spatially distributed over 100 of 1359 census tracts. The dependent variable is the score of the tourists' preferences (TPPI), normalized – e.g., fraction of local comments that are favorable.

For each census tract, a measure of the set of independent variables was calculated. Preliminary results and elaborations of statistical tests suggested excluding some explanatory variables from the model, because they are not statistically significant. The results of the statistical tests for measuring redundancy suggested including the following candidates variables normalised by total area of census tract:

- (1) number of historical buildings;
- (2) number of restaurants and facilities;
- (3) hectares of natural protected areas;
- (4) distance from the airport;
- (5) proximity to the historical city center;
- (6) distance from the most popular beach.

The assumption was that if the value of the normalised TPPI is similar to the values that it takes in the closest spatial units, the variable is characterized by spatial autocorrelation. This issue can be addressed by adding a spatially-lagged dependent variable to the set of covariates (Anselin 1988; 2003).

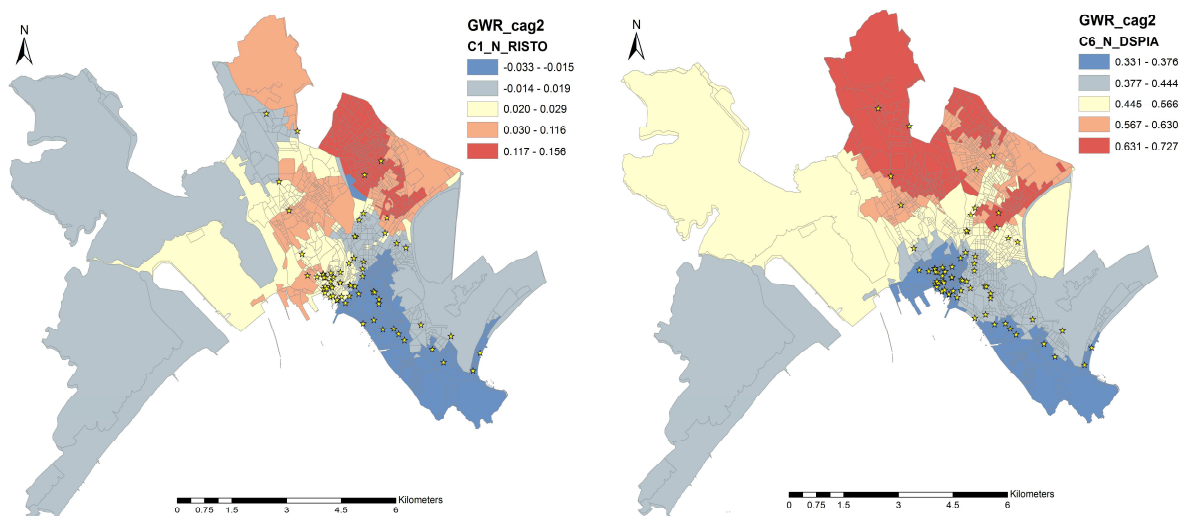


| Variable | Coefficient | Std. Error | z-value | Probability |
|-------------------|-------------|------------|-----------|-------------|
| W_normalised_TPPI | 0.0662904 | 0.0306595 | 2.162143 | 0.0306071 |
| Constant | 0.0030834 | 0.010238 | 0.301150 | 0.7632998 |
| N_restaurants | -0.039035 | 0.0299971 | -1.301316 | 0.1931503 |
| Proxy_city_centre | 0.4747516 | 0.0585060 | 8.114574 | 0.0000000 |
| N_hist_buildings | -0.0273241 | 0.0313221 | -0.872359 | 0.3830122 |
| H_natural_areas | 0.0027261 | 0.0096194 | 0.283397 | 0.0776872 |
| Distance_aeroport | 0.7660687 | 0.0339702 | 22.55116 | 0.0000000 |
| Distance_beach | 0.5470476 | 0.0382759 | 14.29221 | 0.0000000 |

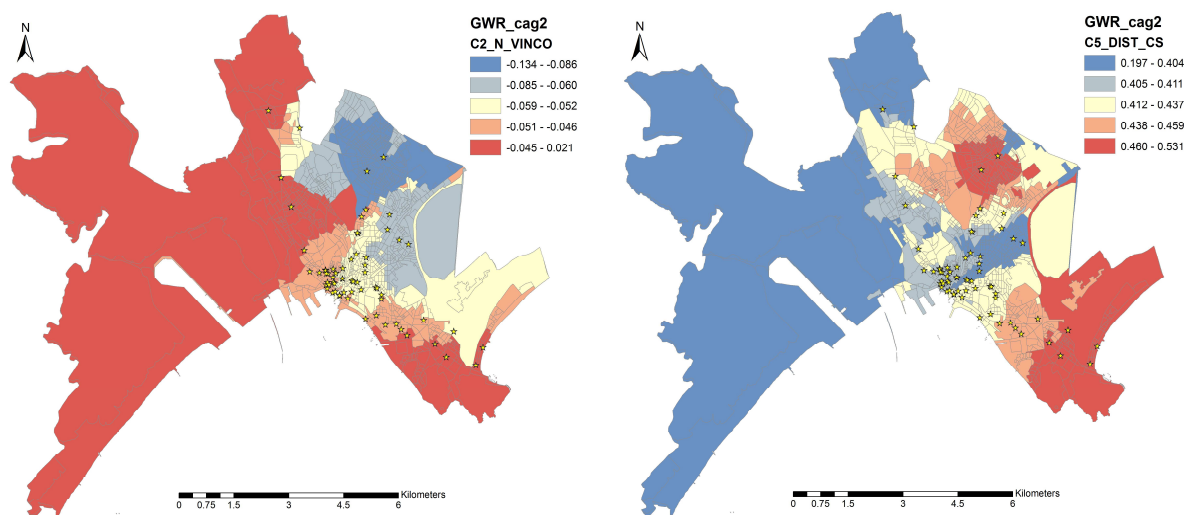
Table 2: Results of the GWR model: influence of each explanatory variable on dependent variable (tourist's preference)

The presence of spatial dependency on positive preference of census tract, which is the value of the normalised TPPI, is detected through the Moran’s test. The result of the local Moran’s Index is quite meaningful for the second order of queen contiguity in respect of results obtained using a 2500 meters spatial autocorrelation distance: adjusted R-squared is less than 40 percent, the coefficient of the dependent variable (p-value b) is less than 5 percent (0.00000010) and the value of Moran’s Index is 0.024. The low p-value indicates that local spatial autocorrelation in dependent variable is much more than one would occur randomly. The output of a spatial lag model of autocorrelation is shown in Table 2. Thus, an analysis of spatial dependency and its significance has been developed through the LISA and results highlight homogeneous areas where spatial dependency is stronger and statistically significant. Thus, the regression with 6 independent variables is estimated using the tourists’ preferences weight matrix

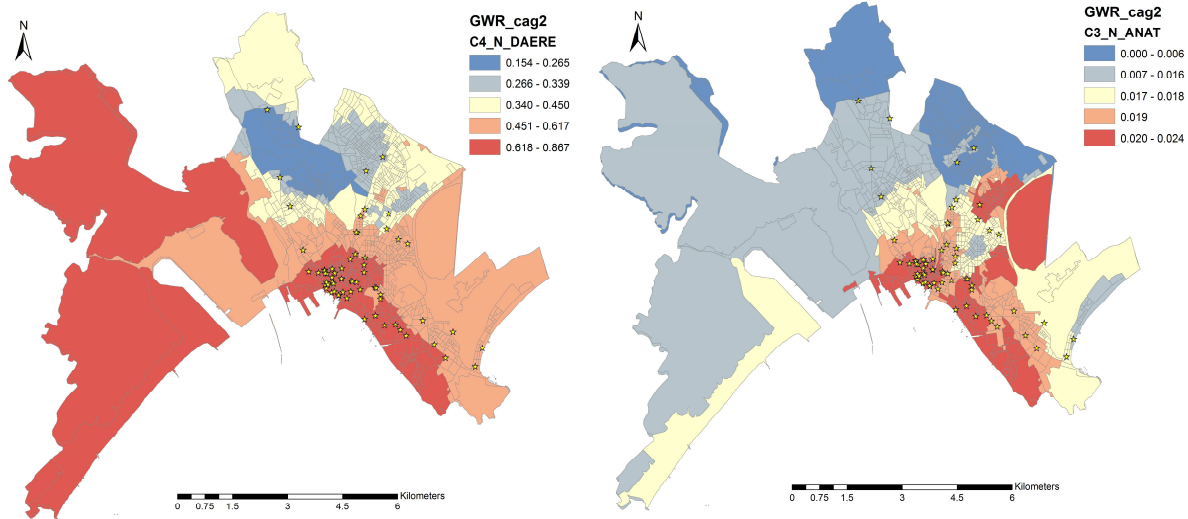
The results concerning the goodness of fit of the regression shown in Table x are significant: R-squared is very high, 0.856636, which indicates that variables in the model explain more than 85% of the variance of the positive tourists’ preferences. The coefficients created to display the spatial patterns located in Cagliari show the relationship between the dependent variable and every explanatory variable (Table 2).



4.a GWR feature class output with rendered residuals of the variables number of restaurant (left) and distance from the beach (right)



4.b GWR feature class output with rendered residuals of the variables number of historical buildings (left) and proximity to the city centre (right)



4.c GWR feature class output with rendered residuals of the variables distance to the airport (left) and hectares of natural protected areas (right)

Figure 4 GWR feature class output with rendered residuals of each variable

The outcomes of the regression model are quite significant for the description of the spatial distribution of tourists' preferences. The coefficients of the variables proximity to the city centre, distance to the airport and proximity to the beach, which are related to location, are almost always significant (p value less than 5 %) and show positive sign. The variables number of historical building and number of restaurants are not significant, for the p value is greater than 10 percent and the negative sign, while the variable hectares of natural protected areas, shows a significant coefficient (0.0027261) and a positive sign.

Overall, these findings suggest that the spatial interest of the participant is quantitatively influenced by the chosen explanatory variables. The selected variables give a more or less significant contribution to tourists' preference explanation through a coefficient. The inclusion of only intrinsic features in the function, allows saying that the values of the coefficients, related to the area, reflect positive effects of geographic position and facilities supply to the spatial patterns of tourist preferences.

6 CONCLUSION

The insights obtained through this novel and integrated approach offer interesting challenges towards the development of more specific analysis, concerning people spatial and thematic perception of places. Spatial analyses and techniques are provided to demonstrate how SMGI may be directly used and integrated with traditional authoritative spatial data layers in GIS environment.

Firstly, the results of analysis carried out by coupling SMGI and A-GI from open SDI show the potential in terms of provision of a novel kind of information which may add value to traditional planning knowledge bases so informing decision-making by community multifaceted. As a matter of facts, SMGI may disclose opportunities for further analysis scenarios in urban and regional planning, and may offer useful suggestions for sustainable development based on tourism strategies. In an integrated planning support framework, SMGI analytics might help to understand tourists' observations, preferences, interests, feelings, or needs, and possibly affect decision-making dynamics and urban and regional planning processes with customer oriented strategies. Moreover, the SMGI potentialities to generate useful knowledge for urban and regional planning, might foster citizens', or tourists in this specific case, dialogue about places and events giving the opportunity of being heard so further facilitating the integration of experiential and multifaceted information and professional knowledge. Thus, the knowledge of SMGI if competently addressed, might be used to support analysis, design and decision-making in tourism planning at differ scales, fostering public participation in processes about the current and future development of destinations.

Secondly, the study demonstrates the opportunities of SMGI as support for analysis in tourism planning. In this case both tourist preferences on destinations and tourism industry services were analysed from the spatial perspective through the review judgments collected by the social media platforms TripAdvisor and Booking.com. Results show which are the most popular destinations or areas and what tourists appreciate or

disregard in Sardinia and then in Cagliari municipality. A set of spatial analysis and statistics techniques were used at different geographic scales (regional, local) to describe and visualise the spatial distribution of tourists' preferences and to detect patterns and hot-spots. The findings provide insights on the Sardinia tourism dynamics which should not be available through other data sources traditionally used in spatial or tourism planning.

Finally, considering users' preferences knowledge in supporting the tourism planning processes could represent a significant implication for future research works in the field of social sciences and tourism management. Indeed, the analysis of the Sardinian case study emphasises the importance of the stakeholders (users or tourists) within the inclusive processes. For instance, their behaviours can reinforce or discourage the existing power relations. As a result, the question concerns how tourists and the political and planning processes are linked. In other words, what the social implications of the tourists' behaviours over the planning or political processes are. In this field the thesis provides different implications for theories concerning the theme of participation and in relation to further studies in other research areas. Nevertheless, other considerations in users' perception, considering for example local communities perceptions, already demonstrate challenging and stimulating research opportunities which may eventually bring innovation to tourism planning, design and decision-making.

7 REFERENCES

- ANDRIENKO Gennady L., and Natalia V. Andrienko. Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, Vol. 13, Issue 4, pp. 355-374, 1999.
- ANSELIN L.: *Spatial econometrics: methods and models*, Vol. 4, 1988.
- BAGGIO R., Costa C., Miguens J.: Social media and tourism destinations: Tripadvisor case study. In *Advances in tourism research*, Vol. 26, Issue 28, 2008.
- BENITEZ J., Martin J., Roman C.: Using fuzzy number for measuring quality of service in the hotel industry. In *Tourism management*. Vol. 26, pp. 544-555, 2007.
- BERRY Michael V., and Jonathan P. Keating.: A rule for quantizing chaos?. *Journal of Physics A: Mathematical and General*, Vol. 23, Issue 21, 1990.
- BOCCAGNA P.: Il turismo è un fenomeno spaziale. *Tema. Journal of Land Use, Mobility and Environment*, Vol. 3, Issue 3, 2010.
- BRIASSOULIS H.: Sustainable tourism and the question of the commons. In *Annals of Tourism Research*. Vol. 29, Issue 4, pp. 1065-1085, 2002
- BUHALIS, D.: Strategic use of information technologies in the tourism industry. *Tourism management*, Vol. 19, Issue 5, pp. 409-421, 1998.
- BUHALIS, D.: Marketing the competitive destination of the future. *Tourism management*, Vol. 21, Issue 1, pp. 97-116, 2000.
- BUHALIS, D.: *eTourism: Information technology for strategic tourism management*, 2003.
- BUHALIS, D., & Jun, S. H. (2003). *E-tourism. Information technology for strategic tourism management*.
- BUHALIS, D., Deimezi, O.: E-tourism developments in Greece: Information communication technologies adoption for the strategic management of the Greek tourism industry. *Tourism and Hospitality Research*, Vol. 5, Issue 2, pp. 103-130, 2004.
- BUHALIS, D., O'Connor, P.: Information communication technology revolutionizing tourism. *Tourism recreation research*, Vol. 30, pp. 7-16, 2005.
- CAMPAGNA Michele. et al.: Place I care! Crowdsourcing planning information. In AESOP-ACSP Joint Congress. Dublin, 2013.
- COOPER C., Fletcher J., Fyall A., Gilbert D., Wahill S.: *Tourism: Principles and Practice*. Harlow, 2008.
- CHUNG J. Y., Buhalis, D.: Web 2.0: A study of online travel community. *Information and communication technologies in tourism*, pp. 70-81, 2008.
- DELLAERT B., Dick F. Ettema L., Christer L.: Multi-faceted tourist travel decisions: a constraint-based conceptual framework to describe tourists sequential choices of travel components. *Tourism Management*. Vol. 19, Issue 4, pp. 313-320, 1998.
- DUBÉ L., Renaghan, L. M.: Building Customer Loyalty—Guests' Perspectives on the Lodging Industry's Functional Best Practices (Part I). *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 40, Issue 5, pp. 78-88, 1998.
- FERNBACK J.: There Is a There There. Notes Toward a Definition of Cybercommunity. In Jones S. (editor), *Doing Internet Research. Critical Issues and Methods for Examining the Net*, Sage, Thousand Oaks, pp. 203-220, 1999.
- FOTHERINGHAM S.A, Brunson C., Charlton M.: *Geographically weighted regression: the analysis of spatially varying relationships*, 2003.
- GLEAVE Eric, et al. A conceptual and operational definition of social role in online community. *System Sciences*, 2009.
- GOODCHILD M. Citizen as Voluntary sensors: spatial data infrastructure in the World of Web 2.0. In *International Journal of Spatial Data Infrastructures*. Vol. 2, pp. 24-32, 2007.
- GRÄBNER D.: Classification of customer reviews based on sentiment analysis. *Information and Communication Technologies in Tourism*, pp. 460-470. Springer, Vienna, 2012.
- HALL C.M.: *Tourism Planning*. Harlow, 2000.
- KIETZMANN J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons*, Vol. 54, Issue 3, pp. 241-251, 2011.
- KUANG HZU H., Tsai Y., Wu H.: The preference analysis for tourist choice of destination: A case study of Taiwan. *Tourism Management*, Vol. 30, Issue 2, pp. 288-297, 2009:
- JAFARI J., Singh, T. V., Kaur, J., Singh, D. P.: The tourism market basket of goods and services: the components and nature of tourism. *Studies in tourism, wildlife, parks, conservation*, pp. 1-12, 1982.

- LEWIS R., Booms B. H.: The marketing aspects of service quality. Emerging perspectives on services marketing, pp. 99-107, Upah Chicago, 1983.
- MASSA P., Campagna, M.: Social Media Geographic Information: recent findings and opportunities for smart spatial planning. Tema. Journal of Land Use, Mobility and Environment, 2014.
- MENNIS Jeremy, and Diansheng Guo.: Spatial data mining and geographic knowledge discover. An introduction. Computers, Environment and Urban Systems Vol. 33, Issue 6, pp. 403-408, 2009.
- MILLER H. J.: The Data Avalanche is here. Shouldn't we be digging?. Journal of Regional Science, Vol. 50, Issue 1, pp. 181-201, 2010.
- O'CONNOR P.: Electronic information distribution in tourism and hospitality, 1999.
- PECHLANER H., Rienzner H., Matzler, K., Osti L. Response attitudes and behavior of hotel industry to electronic info requests, pp. 177-186, Vienna, 2002.
- PERONI G., Formisano M., Matiddi M., Sfodera F.: Fondamenti di marketing turistico territoriale. Milano, 2008
- POSER K., Dransch, D.: Volunteered Geographic Information for disaster management with application to rapid flood damage estimation. Geomatica, Vol. 64, Issue 1, pp. 89-98, 2010.
- RODRIGUEZ DEL BOSQUE I., San Martin H.: Tourist satisfaction. A cognitive affective model. In Annals of tourism research., Vol. 35, Issue 2, pp. 551-573, 2008.
- SMITH V. L.: Anthropology and tourism. Annals of Tourism Research, Vol. 18, Issue 1, pp.12-25, 1991.
- STAMBOULIS Yeoryios, and Pantoleon Skayannis. "Innovation strategies and technology for experience-based tourism." Tourism management Vol. 24, Issue1, pp. 35-43, 2003.
- SUI D., Goodchild M.: The convergence of GIS and social media: challenges for GIScience. In International Journal of Geographical Information Science. Vol. 25, pp. 1737-1748, 2011.
- UYSAL M., Jurowski, C.: Testing the push and pull factors. Annals of Tourism Research, Vol. 21, Issue 4, pp. 844-846, 1994.
- VAN RAAIJ W.: Consumer research on tourism mental and behavioural constructs. Annals of Tourism Research, Vol. 13, Issue 1, pp. 1-9, 1986.
- YOON Y., Uysal, M.: An examination of the effects of motivation and satisfaction on destination loyalty: a structural model. Tourism management, Vol. 26, Issue 1, pp. 45-56, 2005.