GEO-REFERENCED SENTIMENT ANALYSIS FOR TOURISTS' POINTS OF INTEREST: THE CASE OF MATERA EUROPEAN CAPITAL OF CULTURE

Luigi Celardo, Michelangelo Misuraca, Maria Spano

Abstract. Known as the 'city of Sassi', Matera underwent a renewal process involving all of the Basilicata regions in the last few years. The city's Tourism resources were almost unknown at a national and international level, although the Sassi were included in the UNESCO World Heritage List since 1993. The European Capital of Culture 2019 nomination triggered an intense regeneration, opening the city to global tourism and revealing a high resilience. Tourists' experiences and opinions have been valuable resources for designing tourism activities and creating a new symbolic identity for the city, especially in the Web 2.0 era. Here we propose to compute the reviews' polarity scores and use them with other characteristics (e.g., price, offered services and type of tourist facilities) to build spatial clusters according to the logic of Local Spatial Association Indicators (LISA). The geo-referenced semantic orientation of reviews concerning a particular activity or attraction represents a useful quantitative feature for further analyses and the production of territorial statistics. The proposal can be extended to other cases to monitor the change of sentiment towards specific areas of interest and plan possible intervention policies.

1. Introduction

Cyberspace has opened new opportunities to improve communication and develop customer acquisition and retention approaches. A crucial aspect is the phenomenon of online interpersonal influence (Senecal and Nantel, 2004), which can be referred to as *electronic word-of-mouth* (eWOM), based on the previous Westbrook (1987) definition of word-of-mouth (WOM). The eWOM is an informal communication of consumers about the characteristics and the mode of consumption of particular goods and services when they need to make a purchase decision. Among the different sectors, such opinions became increasingly significant in hospitality and tourism, where the intangible nature of products makes it difficult for an assessment and interpersonal confidence in products is extremely important.

The reputation generated by digital platforms is used by diverse actors in the tourism industry. A good eWOM about facilities and attractions helps potential visitors find information and reinforces images and opinions, as it contributes to creating the symbolic image of tourist destinations. On the other hand, eWOM takes on a strategic role in destination management and operators begin to integrate traditional contextual analysis with tourist review data. In this framework, monitoring tourist sentiment is crucial for analysing the impact of tourism growth in specific areas, considering external and internal factors that influence the image. Here we propose to compute polarity scores of reviews about touristic points of interest (attractions, restaurants, accommodation) on social media platforms such as TripAdvisor, to have both a spatial representation of the spread of sentiment and to construct spatial clusters, according to the logic of *Local Spatial Association Indicators* (LISA), to identify hot or cold spots where positive and negative sentiments are more concentrated (Anselin, 1995). Some authors have already used LISA to classify tweets into different topics and group them spatially to show the relationship between content and place characteristics (Hu *et al.*, 2020) or to find the distribution of topics associated with urban features (Zhong *et al.*, 2018). The proposal effectiveness is shown through a case study about the city of Matera, which recently experienced a regeneration process involving, more generally, the entire Basilicata region. The city's tourism resources were almost unknown, both nationally and internationally. Although in 1993, the Sassi became the first site in southern Italy to be included in the UNESCO World Heritage List as a 'cultural landscape', the real change came with its nomination as the European Capital of Culture for 2019. The nomination triggered an intense development, opening the city to global tourism and demonstrating a high resilience (Ivona *et al*., 2019). The Internet played a key role in reinterpreting the local urban reality, where tourists want to have a unique immersive experience in the Lucanian community rather than being mere onlookers. Our strategy is developed within the framework of the so-called *Ambient Geographic Information* (AGI) system of Stefanidis *et al.* (2013), and it can be extended to other cases to monitor the evolution of sentiment towards specific areas of interest and plan possible intervention policies. The geo-referenced semantic orientation of tourists' opinions represents a useful quantitative feature for further analysis and the production of spatial statistics.

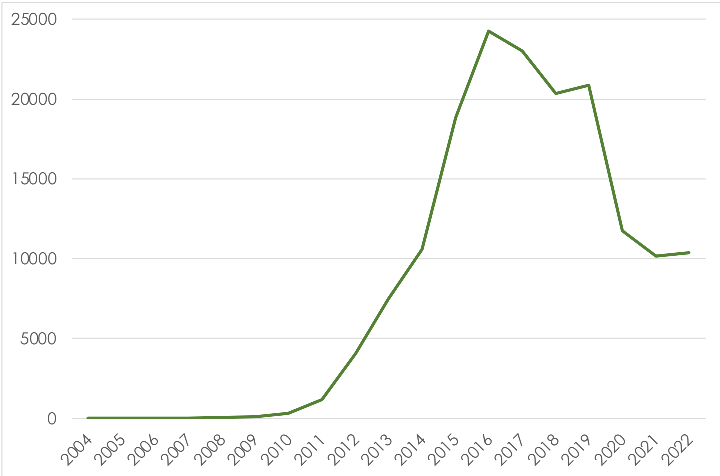
2. Data and methods

The so-called *experiential* tourism, which stems from the "desire to see life as it is lived and to get in touch with the locals" (Lemmi, 2009), is a concept of tourism strongly inherent to the European Capital of Culture event promoted by the European

Union (Guala, 2002). This initiative gives the winning city 365 days to enable its environmental and cultural heritage, involving a broader territorial context than the purely urban area to create virtuous synergies. Matera's strategic project has enabled the Lucanian city to offer alternatives and specific accommodation to around 700,000 temporary citizens in an urban area that, after having been condemned to painful isolation along with the entire regional system of Basilicata since the Second World War, is now undergoing an intense internal regeneration, opening up to competition from glocal tourism and showing a high degree of resilience (Pollice and Urso, 2014). The dossier submitted to the jury for the 2014 European Capital of Culture nomination was entitled *Open Future*, highlighting the possibility of combining science and technology with an exceptionally creative streak to consolidate the city (and Basilicata) position in the creative sector at a European level. The report also mentions the creation of an online cultural platform to provide tourist information at all spatial levels and in real-time to reach and be reached directly by potential visitors (Pollice, 2010) through eWOM. Embracing the same idea, we developed a dataset including the opinions of tourists that visited Matera in the last years. All the restaurant, accommodation, and attraction reviews used in this study have been extracted from the Italian TripAdvisor website, using `MATERA` in the query and restaurants, accommodation and attractions as the main category. Reviews posted between 2004 and 2022 (up to March 23, 2023) have been stored in a local repository together with some metadata, like *address*, *latitude,* and *longitude* (validated with the corresponding Google Maps ID_place), *rating*, # *number of reviews*. We considered the total number of activities and filtered the facilities by latitude and longitude values not geolocated outside the sub-areas defined by the Municipality of Matera[1]. The data extraction procedure belongs to the so-called *web scraping*. Generally, web scraping can be defined as a systematic process of identification and retrieval of content of interest from the Web. A software agent mimics the browsing interaction between the web servers and a human user. Step by step, the agent accesses as many web pages as necessary, analysing the content to extract the data and structure them in the desired form. In our case, two scripts were built using *Python* and *SeleniumLibrary* to extract information on the different POIs and the related on individual reviews. We obtained in this way 806 points of interest with 164,651 reviews.

[1] http://dati.comune.matera.it/dataset/aree-sub-comunali-comune-di-matera

Figure 1 - *Reviews in Italian on Tripadvisor (2004-2022).*



Analysing the evolution over the years is useful to assess the overall trend since the number of reviews published over time is an indicator of the consolidation of eWOM about Matera. Figure 1 shows how the number of reviews tends to increase over the years, reaching a peak in 2014, when Matera officially presented its bid to become the European Capital of Culture. The subsequent years of preparation for the event increased until reaching a peak in 2019. The year Matera became the Capital of Culture was crucial, but it should be remembered that the spread of the Covid-19 pandemic had a decreasing effect on the number of visits. It is worth noting that the number of reviews returns to the years before 2015, indicating that even though the number of visitors decreases, the number of reviews remains at around 10,000.

To acquire the sentiment scores associated with POIs, we employed an original customised lexicon of Italian terms to perform a lexicon-based sentiment analysis of the reviews. Most resources in the sentiment research area, like lexicons, labelled supplies and NLP tools, are mainly available in English. The lack of linguistic resources is critical in most studies, producing a so-called *lexical gap* (Chiavetta *et al.*, 2016). Thus, we built an Italian lexicon by merging the resources developed in the Opener project (Russo *et al.*, 2016) and other selected studies (e.g., Bolasco and Della Ratta, 2004). The resulting set contains 12,400 polarised lemmas with a value of +1 if positives and -1 if negatives. The reviews were lightly pre-processed. Non-alphabetic characters and symbols - such as numbers or emoticons - were removed to include only content-bearing terms. Polarity scores were calculated using a sentence-level logic (Balbi *et al.*, 2018). Given a review r_i ($i = 1, \ldots, n$), its a_i sentences $\{s_{i1}, \ldots, s_{ik}, \ldots, s_{ia_i}\}$ are identified by considering as separators strong punctuation marks like full stops, question marks and exclamation marks. The k-th sentence s_{ik} is a sequence of its p_k terms $\{t_{ik1}, ..., t_{ikj}, ..., t_{ikp_k}\}$. Each term t_{ikj} in the k-th sentence of the i-th review is compared with the terms in the lexicon, assigning a -1 to negative terms and a +1 to positive terms, respectively. Terms not listed in the lexicon are scored with a null value. The polarity of each term is then weighted considering negators (e.g., *mai*, *nessuno, nessuna*), amplifiers and de-amplifiers (e.g., *poco*, *molto*, *pochissimo*), adversative and contrasting terms (e.g., *ma*, *tuttavia*). This weighting scheme allows for emphasising or dampening the negativity or positivity of each polarised term, leading to a more effective measure of semantic orientation (Vechtomova, 2017). The polarity score of each sentence $\mathrm{PS}_{s_{ij}}$ is obtained as the sum of weighted term scores $\mathrm{PS}_{w_{ikj}}$ on the sentence length:

$$\mathrm{PS}_{s_{ij}} = \frac{\sum_{k=1}^{p_j} \mathrm{PS}^*_{t_{ijk}}}{\sqrt{p_j}} \tag{1}$$

Since we are interested in obtaining a polarity score at a review level, we calculated an overall score PS_{r_i} for each text by an average of sentence polarities:

$$\mathrm{PS}_{d_i} = \frac{\sum_{j=1}^{q_i} \mathrm{PS}_{s_{ij}}}{\#(q_i^-) + \#(q_i^+) + \#(q_i^0)} \tag{2}$$

where q_i^-, q_i^+ and q_i^0 are the numbers of sentences in r_i with a negative, positive, or neutral polarity, respectively.

In the last step, we used a global autocorrelation indicator called Global Moran's I to relate the sentiment score to the spatial location of the points of interest. The basic idea is to assess the absence of spatial randomness in the distribution of a variable. We relate the global value of sentiment to its importance in neighbouring areas to summarise the degree of spatial similarity between neighbouring observations across the study area (Getis *et al.*, 1992)

$$I = \frac{n}{\Sigma_i \Sigma_j w_{ij}} \frac{\Sigma_i \Sigma_j w_{ij} z_i z_j}{\Sigma_i z_i^2} \tag{3}$$

where z_i, is the deviation of sentiment score for each i observation from its means, w_{ij} is the spatial weight for the pair of observations i and j, and n is the number of observations. A positive value of Global Moran's I represent a situation where similar values are found in an adjacent area.

Nevertheless, we are interested in the global spatial representation of sentiment and in identifying regions, i.e. the sub-areas defined by the municipality of Matera, which influence the average positive sentiment at the local level due to their spatial proximity or remoteness. For this reason, local spatial autocorrelation could reflect sentiment spread between scores in neighbouring locations, representing scenarios where scores in one place influence other sites or where scores in one location causally affect other areas. Local measures of spatial autocorrelation focus on the relationships between each observation and its environment rather than providing a single summary of these relationships across the map. In this sense, they are not summary statistics but scores that allow us to learn more about the spatial structure of our data. A Local Moran's I_i identifies cases where the value of an observation and the average of its surroundings are more similar (H-H or L-L) or dissimilar (H-L, L-H) than we would expect randomly. The index is applied to every observation. In this way, there are as many statistics as original observations. The formal representation of the statistic can be written as follows:

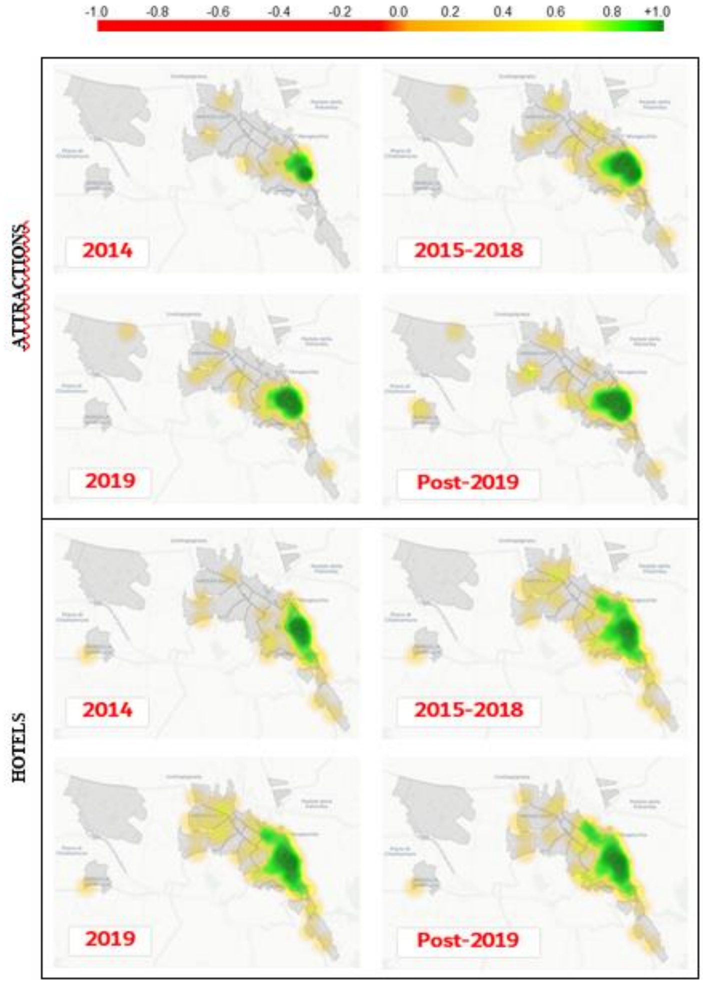
$$I_i = \frac{n(y_i - \bar{y})}{\Sigma_i (y_i - \bar{y})^2} \sum_j w_{ij} (y_j - \bar{y}) \tag{4}$$

where y_i is the value of sentiment score for each *i* observation in the distribution of data, w_{ij} is the spatial weight for the pair of observations *i* and *j*, and *n* is the number of observations.

3. Empirical results

Given the scale of the investment, the question of long-term legacy is essential for all ECOCs (European Capitals of Culture). While all cities that were ECOC in 1995-2004 set long-term objectives for their year of glory, only half set up funds or bodies to pursue them. Therefore, monitoring the widespread opinion is a good way to avoid overlooking the issue of project sustainability over time when planning the event (Corinto and Nicosia, 2016). A first graphical output of the analytical strategy proposed here allows for displaying attractions, hotels, and restaurants' sentiment orientation by means of a geographical heatmap, where the red colour represents a negative sentiment score, whereas the green colour represents a positive score. In Figure 1, the different periods are compared in order to appreciate the temporal evolution of the tourists' opinions about the POIs. In the year of the nomination (2014), a wide green area can be seen around the Sassi districts and the historical city centre for all the POIs categories. In the other analysed periods, the positiveness gradually expanded to reach new areas, even in some districts of the new city centre, such as San Pardo, San Giacomo and Villa Longo, particularly for hotels and restaurants. Thus, an effective spread of positive sentiment aligns with the main objectives for promoting the city of Matera included in the dossier submitted for the Cultural Capital bid.

Figure 2 - *Sentiment Heatmap of attractions, hotels and restaurants (2004-2022).*



Before performing the spatial analysis, we calculated the Global Moran's I from the sentiment scores, obtaining a 0.568 value. To give additional substance to the hypothesis of non-randomness in POIs distribution, by means of a Monte Carlo simulation with 999 permutations, we carried out a statistical test and obtained a $<<0.001$ p-value. Figure 3 depicts the four clusters of POIs in the suburbs, with each area associated with one of the results typically generated by the LISA model (high-

high in red, high-low in orange, low-high in light blue and low-low in dark blue). The output can be read again with respect to an overall improvement in tourists' perception from the nomination as the Capital of Culture to the present day.

As in the previous output, we observed for the LISA model a general spread of positiveness from the Sassi area to other Matera suburbs, such as the historic city centre, San Pardo and San Giacomo districts. The non-significant clusters for all the tourist attractions tend to diminish over time, showing high-high clusters (in red) from the Sassi districts to more peripheral neighbourhoods such as Agnara and Serra Rifusa. It is interesting to note how for attractions, the different areas of the city flowed across time in the high-high cluster (in red), giving evidence of the success of Matera promotion and of the triggering effect of the nomination as Capital of Culture. As concerns accommodations, the COVID-19 pandemic influenced the spatial distribution of POIs, highlighting how not all the structures were able to maintain a high standard in the tourists' view.

Figure 3 - *Hot-spots and cold-spots of attractions, hotels and restaurants (2004-2022). (segue)*

ATTRACTIONS

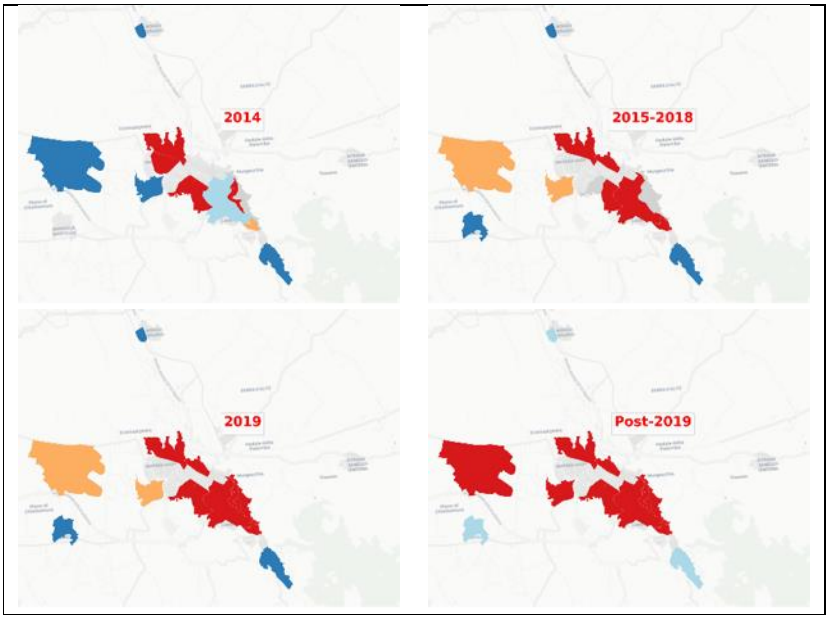
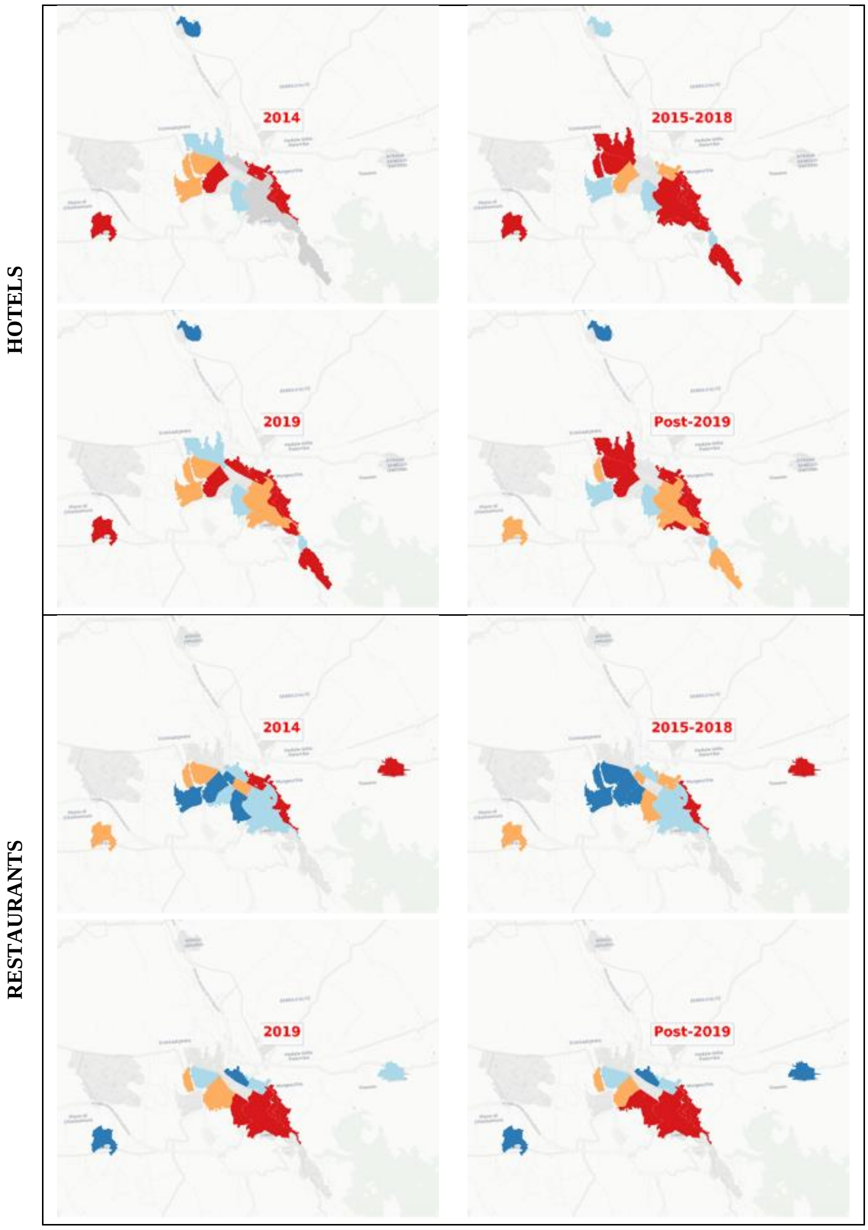


Figure 3 - *Hot-spots and cold-spots of attractions, hotels and restaurants (2004-2022). (continua)*



4. Conclusion and final remarks

Following Matera's example, many cities can benefit from the positive effects of enhancing social and cultural assets. The process of urban regeneration is a valuable case study to illustrate the success and importance of coopetition between actors in a territory (Nguyen *et al.*, 2022).

The integrated analysis of polarity scores and geographical data can help explore tourists' opinions and perceptions, providing all stakeholders with a powerful tool to assess and monitor the attractiveness of a given territory and develop actions or policies to maintain or improve its image. Thanks to sentiment analysis, tourism operators and public administrations can study the orientation of reviews published by tourists after a trip (considering all facets of this experience) or after visiting a tourist attraction. The possibility of geo-referencing tourist sentiment increases the informative value of this type of data. The ability to manage these data flows also provides the opportunity to integrate traditional contextual analysis with data-driven strategies by incorporating the mutable opinions of tourists into planning. In this sense, creating interactive maps where the territory can be explored at different levels is possible instead of using a static representation of sentiment. Since the monitoring process included in the Matera OPEN FUTURE dossier requires real-time feedback on the impact of the city's renewal and regeneration processes, it could be useful not only to display static photographs of the various LISA clusters in the area but also to consider the possibility of identifying dynamic clusters. In this sense, other LISA techniques, such as its spatio-temporal version, could be tested. Closely related to the comparative static analysis of LISA statistics, the dynamic performance could provide additional insights into the shape and direction of cold and hot spots of sentiment in a given time interval (Rey, 2019). Sentiment could also be linked to other intrinsic characteristics of POIs and their geographical location. For example, the restaurant's price or the rating associated with the services, cleanliness, or type of cuisine influences tourists' opinions.

Finally, a positive trend in sentiment indicates an optimal reception of the services offered in the area. In promoting these processes, the European Commission is interested in observing and assessing civic engagement when selecting European Capitals of Culture (ECoCs) applications and implementing effective evaluation and long-term monitoring mechanisms. Synergies of this kind are useful to verify that the participation of all local energies remains constant to support the public administration in maintaining the real involvement of the local population in the long term and not only in the short period (Demartini *et al.*, 2020).

References

ANSELIN L., 1995. Local Indicators of Spatial Association - LISA, *Geographical Analysis*, Vol. 27, No. 2, pp. 93–115.

BALBI S., MISURACA M., SCEPI G. 2018. Combining different evaluation systems on social media for measuring user satisfaction, *Information Processing & Management*, Vol. 54, No. 4, pp. 674–685.

BOLASCO S., DELLA RATTA F. 2004. Experiments on semantic categorisation of texts: analysis of positive and negative dimension. In PURNELLE G, FAIRON C. and DISTER A. (Eds.) *Le poid des mots. Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles,* UCL Presses Universitaires de Louvain, pp. 202–210.

CHIAVETTA F., LO BOSCO G., and PILATO G. 2016. A Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language. In MAJCHRZAK T.A, TRAVERSO P., MONFORT V. and KREMPELS K.-H. (Eds.), In *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, Scitepress, pp.159–170.

CORINTO G. L., NICOSIA E. 2016. The European Capital of Culture. Will Matera 2019 be a successful example? *In development, art(s) and culture UNEEC forum*, Vol. 8, Lucian Blaga University of Sibiu Press, pp. 19–35.

DEMARTINI P., MARCHEGIANI L., MARCHIORI M. 2020. Culture Invites Participation. An Inquiry on Matera as European Capital of Culture 2019. In PIBER M. (Ed.) *Management, Participation and Entrepreneurship in the Cultural and Creative Sector*. Berlin: Springer, pp. 161–186.

GETIS A., ORD J. K. 1992. The Analysis of Spatial Association by Use of Distance Statistics, *Geographical Analysis*, Vol. 24, No. 3, pp. 189–206.

GUALA C. 2002. Towards A typology of Mega-Event, *Bollettino della Società Geografica Italiana*, Vol. 7, No. 4, pp. 743–755.

HU T., SHE B., DUAN L., YUE H., CLUNIS J. 2020. A Systematic Spatial and Temporal Sentiment Analysis on Geo-Tweets, *IEEE Access,* Vol. 8, No. 1, pp. 8658–8667.

IVONA A., RINELLA A., RINELLA F. 2019. Glocal Tourism and Resilient Cities: The Case of Matera "European Capital of Culture 2019, *Sustainability*, Vol. 11, No. 15, pp. 4118–4130.

LEMMI E. 2009. *Dallo «spazio consumato» ai luoghi ritrovati. In Verso una Geografia del Turismo Sostenibile*. Milan: Franco Angeli.

NGUYEN T. Q. T., JOHNSON P., YOUNG T. 2022. Networking, coopetition and sustainability of tourism destinations, *Journal of Hospitality and Tourism Management*, Vol. 50, pp. 400–411.

POLLICE F. 2014. *Territori del turismo. Una Lettura Geografica delle Politiche del Turismo.* Milan: Franco Angeli.

POLLICE F., URSO G. 2014. Turismo vs. globaritarismo. In TURCO A. (Ed.), *Filiere etiche del turismo. In Territori della Vacanza tra Valori, Politiche e Mercati*, Unicopli, pp. 75–96.

REY S., 2019. Spatial Dynamics and Space-Time Data Analysis. In: FISHER M., and NIJKAMP P. (Eds.), *Handbook of Regional Science*, Berlin: Springer, pp. 1–18.

SENECAL S., NANTEL J. 2004. The influence of online product recommendations on consumers' online choices, *Journal of Retailing*, Vol. 80, No. 2, pp. 159–169.

STEFANIDIS A., CROOKS A., RADZIKOWSKI J. 2013. Harvesting ambient geospatial information from social media feeds, *GeoJournal*, Vol. 78, No. 2, pp. 319–338.

VECHTOMOVA O. 2017. Disambiguating context-dependent polarity of words: An information retrieval approach, *Information Processing & Management*, Vol. 53, pp. 1062–1079.

WESTBROOK R. A. 1987. Product/consumption-based affective responses and postpurchase processes, *Journal of Marketing Research*, Vol. 24, No. 3, pp. 258–270.

ZHONG C., ZENG S., TU W., YOSHIDA M. 2018. Profiling the Spatial Structure of London: From Individual Tweets to Aggregated Functional Zones, *International Journal of Geo-Information*, Vol.7, No.10, pp. 386–400.

Luigi CELARDO, University of Naples Federico II, luigi.celardo@unina.it
Michelangelo MISURACA, University of Calabria, michelangelo.misuraca@unical.it
Maria SPANO, University of Naples Federico II, maria.spano@unina.it