A network perspective of cognitive and geographical proximity of sustainable tourism organizations: evidence from Italy

Silvia Blasi
Department of Business Administration, University of Verona, Verona, Italy

Shira Fano
Fondazione Nord Est, Venezia, Italy

Silvia Rita Sedita
Department of Economics and Management, University of Padova, Padova, Italy, and

Gianluca Toschi
Fondazione Nord Est, Venezia, Italy

Abstract

Purpose – This research aims to contribute to the literature on sustainable hospitality and tourism by applying social network analysis to identify sustainable tourism business networks and untangle the role of cognitive and geographical proximity in their formation.

Design/methodology/approach – Data mining and machine learning techniques were applied to data collected from the websites of tourism companies located in northeastern Italy, namely, the Veneto region. Specifically, the authors used Web scraping to extract relevant information from the internet.

Findings – The results support the existence of geographical clusters of tourist accommodation providers that are linked by strong cognitive proximity based on sustainability principles that are well communicated via their websites. This does not appear to be greenwashing because companies that have agreed on sustainability principles have also implemented concrete actions and tend to signal these actions through a variety of sustainability certifications.

Practical implications – The results may guide tourism managers and policymakers in developing tourism initiatives directed at the creation of fruitful collaborations between similarly oriented organizations and methods to support clusters of sustainable tourism accommodation. Identifying sustainable tourism networks may assist in the identification of potential actors of change, fueling a widespread transition toward sustainability.

Originality/value – In this study, the authors adopted an innovative methodology to detect sustainability-oriented tourism business networks. Additionally, to the best of the authors' knowledge,

this study is one of the first to simultaneously explore the cognitive and geographical connections between tourism businesses.

1. Introduction

Tourism is one of the most important economic sectors in many countries (World Travel and Tourism Council, 2020), and travelers are increasingly considering sustainability as a key criterion in their travel choices (Antonides, 2017; Beaumont, 2011; Croce *et al.*, 2019; Kitthamkesorn and Chen, 2017; Shaw, 2004; Lehto and Lehto, 2019). This emergent group of consumers, who seek more authentic and fulfilling experiences that benefit the body, mind and soul, is often referred to as the "Lifestyles of Health and Sustainability" (LOHAS) market segment. According to Kotler (2011, p. 144), "the market for LOHAS products is growing"; thus, "producers will have to decide more carefully what to produce, how to produce it, how to distribute it, and how to promote it." This search for holistic experiences is prompting tourism operators to develop new business models focused not only on profits but also on social and environmental sustainability (Pan *et al.*, 2018; Więckowski, 2021). However, caution should be taken when considering these behaviors, which may simply be a form of greenwashing to attract conscious consumers. Careful attention should be paid to real sustainability engagement, which is verifiable through certification (Lesar *et al.*, 2020; Sampaio *et al.*, 2012).

The existing literature has extensively explored sustainable tourism from a variety of perspectives (Butler, 1999; de Lange and Dodds, 2017; Ivars-Baidal *et al.*, 2021; Streimikiene *et al.*, 2021; Timur and Getz, 2008). Given that the tourism industry contributes approximately 8% of total carbon dioxide emissions worldwide (Lenzen *et al.*, 2018), most studies have focused on green tourism (Font *et al.*, 2021; Jones, 1987; Lenzen *et al.*, 2018; Line *et al.*, 2018), while others have studied the impact of tourism on the quality of life of residents in tourist destinations (Hassan, 2000; Jeon *et al.*, 2016; Mihalic and Kuščer, 2022; Ramkissoon, 2020).

Few studies have analyzed the tourism sector by adopting a systematic approach in which the unit of analysis is not the individual actor but the overall ecosystem in which actors interact to ensure economic, social and environmental sustainability. However, given the increasing impact of globalization on human mobility worldwide, much is yet to be done to analyze ways to support the tourism industry's transition to sustainability (Elche *et al.*, 2018; Osti and Goffi, 2021). One way to improve our understanding of sustainability in tourism is to simultaneously consider the cognitive and geographical dimensions of tourism businesses (Weidenfeld *et al.*, 2016), thus adopting a sustainable ecosystem approach, as suggested by Boschma's (2005) research on evolutionary economic geography.

This work aims to contribute to the literature on sustainable hospitality and tourism by applying a social network analysis to identify sustainable tourism business networks and untangle the role of cognitive and geographical proximity. Our empirical analysis responds to the following research questions:

RQ1. Do tourism businesses commit to sustainability? If so, to what extent?

RQ2. Can we differentiate between soft (mainly greenwashing) and strong (actions and certifications) sustainability engagement?

RQ3. Are firm networks linked by geographical proximity?

RQ4. Are firm networks linked by cognitive proximity?

The empirical setting for this research was the Veneto region in northeastern Italy, where we analyzed a sample of tourist accommodations using an original research method. Cognitive proximity was determined by applying textual analysis and machine learning techniques to data collected from websites using information retrieval techniques. Specifically, using the bag-of-words approach (Zhang *et al.*, 2010), we extracted text from websites related to a specific set of words (a "query"), then analyzed website content using the term frequency (TF)–inverse document frequency (IDF) text classification method (Jones, 1972). This methodology allowed us to create an index of the presence and degree of discussions about sustainability on each website. Data collected through web scraping enabled the construction of sustainable accommodation provider networks. Then network nodes represented the accommodation provider websites and were connected by links representing the use of common terms related to sustainability.

Geographical proximity was measured by adopting the local labor system (LLS) criterion [1], which identifies clusters of firms with reference to local labor force mobility. This method has previously been used for tourist destination analysis in Italy (Lazzeretti and Capone, 2008; Lazzeretti and Petrillo, 2006). LLSs are aggregations of neighboring municipalities characterized by high demographic density. Each LLS is identified by the name of its most populous municipality, which usually has a greater availability of productive, commercial and administrative resources, and thus is the likely focus of the local labor market.

Based on original data and methods, our findings contribute to the present understanding of the relational structure of tourism businesses in Italy and may be used to inform tourism managers and policymakers on developing sustainable tourism initiatives. The creation of fruitful collaborations among similarly oriented organizations and the maintenance of sustainable tourism ecosystems are ways to guarantee a widespread transition toward sustainability.

This article proceeds as follows. Section 2 presents the literature review, Section 3 illustrates and discusses the empirical findings and Section 4 offers some conclusive remarks.

2. Literature review
Collaborative innovation in tourism and hospitality is of paramount importance (Marasco *et al.*, 2018). The innovation literature has long maintained the crucial role of cognitive proximity for benefiting network relationships. Developed by Nooteboom (1999), the concept of cognitive proximity is commonly defined as similarities in the way in which actors perceive, interpret, understand and evaluate the world (Knoben and Oerlemans, 2006). This concept is particularly important for promoting innovation because it fuels knowledge exchange between organizations, leveraging absorptive capacity (Cohen and Levinthal, 1990) and the recombination of related knowledge bases (Tanner, 2016). Organizations that are connected through cognitive proximity create an environment that affects their strategic behaviors (Ruiz-Ortega *et al.*, 2021). Meyer and Rowan (1977) observed that organizational behavior is strongly influenced by the environment in which organizations operate, generating isomorphic behaviors led by shared cultures, norms and beliefs (Miska *et al.*, 2018). DiMaggio and Powell (1983) argued that isomorphism between organizations is not a result of competition nor a requirement for efficiency; rather, it indicates a search for legitimacy within the environment in which they operate. Isomorphism may also be driven by a common orientation toward corporate social responsibility, where firms may decide to

legitimize their common goals via certification (Gehman and Grimes, 2017), thus limiting the greenwashing effect. Tourist destinations that share a sustainability orientation exhibit cognitive proximity, which can act as a lever for adopting entrepreneurial strategies and actions to pursue sustainability goals (Ruiz-Ortega *et al.*, 2021).

To clarify the role of cognitive proximity in directing tourist destinations toward a more sustainable orientation, it is useful to recall social capital theory (Putnam, 2000). Nahapiet and Ghoshal (1998) define social capital as the resources embedded in a firm's network of relationships that enable the firm to access relevant knowledge and improve its strategic decisions (Baggio and Cooper, 2013). In the context of tourist destinations, the role of social capital is of particular interest because firms draw on complex networks of relationships with a variety of stakeholders within a tourist ecosystem (Córcoles Muñoz *et al.*, 2022). Over the past two decades, this concept has become pervasive in the strategic literature (Adner, 2017). The ecosystem is characterized by the existence of both competition and cooperation between firms. Recently, a growing number of researchers have used the concept of the ecosystem to analyze business and innovation phenomena (Blasi and Sedita, 2020). An interesting approach to analyzing the business ecosystem is social network analysis. Jacobides *et al.* (2018) underline that research on networks and ecosystems can be mutually beneficial. However, while some scholars have applied network analysis to investigate ecosystems, none have analyzed tourism business networks as ecosystems that share norms, values and beliefs as a result of their cognitive proximity.

Collaborative networks and ecosystems have also been analyzed in terms of their spatial dimensions, especially in research on economic geography (Asheim, 1996; Maskell, 2017; Saxenian, 1996; Scott, 1988; Storper, 1997) and by social economists (Becattini *et al.*, 2001; Becattini, 1979, 2000; Brusco, 1989) and business strategic experts (Porter, 1990). The concept of geographical proximity dates back to studies by Marshall (1920, pp. 4–9), who postulated the existence of external economies that "can often be secured by the concentration of many small businesses of a similar character in particular localities." Following Marshall's pioneering ideas, scholars have built a generous corpus of literature on the evolution of clusters and industrial districts, which has been extensively summarized by Lazzeretti *et al.* (2014) and Sedita *et al.* (2021). The link between geographical proximity and sustainability has been addressed more recently by Sedita and Blasi (2021), who explored the role of clusters, local development, green industry dynamics, local production systems and social entrepreneurship from the perspective of innovation and sustainability transition. To maintain sustainability, access to resources that guarantee the development of sustainability-oriented actions is necessary. Notably, these resources may be internal or external to the clusters, and there may be multiple forms of sustainability governance led by individual organizations, collective organisms, institutions and policymakers. Moreover, universities and research institutions may also play a role (Belussi *et al.*, 2022).

Recent research on tourism management has paid attention to the geography of collaborative networks for pursuing mutual benefits and tapping into complementary knowledge assets (Jain and Sharma, 2021; Stylidis, 2018). Sustainability goals may also be reached through interconnections between tourism businesses, allowing companies to codesign tourism services and cooperate in the implementation of sustainable tourism solutions (Elche *et al.*, 2018).

Theories on agglomerations and clusters suggest that geographically proximate organizations may exhibit isomorphic behaviors arising from vicarious learning (Manz and Sims, 1981), a type of imitative learning typical of industrial districts (Belussi, 1999). This issue draws attention to absorptive capacity at the cluster level (Giuliani, 2013), reinforced by the actions of specific actors in knowledge networks (Morrison, 2008). In addition, the

related variety approach (Frenken *et al.*, 2007), which has received growing attention in the literature, highlights the need for a certain degree of cognitive proximity in local systems to promote innovation and economic development in the region (Capone and Lazzeretti, 2018) and increase resilience to external shocks (Sedita *et al.*, 2017). In the tourism research field, Lazzeretti *et al.* (2015) illustrated the existence of knowledge spillovers between neighboring destinations, further supporting the idea of destination networks (Lue *et al.*, 1993). Recent studies have also highlighted the utility of social network analysis to map the interconnections between proximate tourist destinations, suggesting an overlap between cognitive and geographical proximity (Casanueva *et al.*, 2016; Lazzeretti and Petrillo, 2006; Prats *et al.*, 2008; Sørensen, 2007).

Colocation facilitates knowledge sharing and collective action planning. Sustainability-oriented organizations that exhibit high cognitive proximity may either be stimulated by geographic proximity to embark on collaborative sustainable initiatives or have the ability to distribute their perspectives among proximate organizations, generating local sustainability-oriented projects.

Our intellectual curiosity led us to explore the relationship between the cognitive and geographical dimensions of proximity and their role in fueling isomorphic sustainability behaviors in tourist destinations.

3. Methods and data

3.1 Methods

We applied data mining and machine learning techniques to data collected from the websites of tourism companies located in the Veneto region using Quantitas Intelligent Business Analyzer (QIBA), a Web crawling and scraping tool.

There are many advantages to analyzing information collected from websites for research purposes. First, corporate websites can showcase a company's products and services. Therefore, it is in the company's best interest to describe its products and, more generally, provide an accurate representation of itself. Moreover, information from websites is relatively inexpensive to obtain (Gök *et al.*, 2015), publicly available and up to date given a company's interest in informing its market about new products, markets and technologies. Collecting data from the Web is also nonintrusive (Arora *et al.*, 2016), which is increasingly relevant in a period when response rates to research questionnaires are in constant decline. The internet can now be considered a source of data instead of or in combination with data collected using traditional tools such as surveys (Barcaroli *et al.*, 2014; Ten Bosch *et al.*, 2018). However, there are some limitations to the use of data obtained from websites. For example, the information in company websites is self-reported. Moreover, it is not standardized, and thus needs accurate processing for appropriate analysis (Arora *et al.*, 2016; Kinne and Resch, 2018). Nevertheless, there is an increasing number of applications to facilitate the use of data collected from the Web for economics and tourism analysis.

Given the rapid rise of the internet, textual data has emerged as a major type of big data in tourism, particularly on websites showing reviews of hotels and restaurants. Text mining and analysis have great potential to inspire innovations for tourism professionals (Li *et al.*, 2019). Indeed, online customer reviews are essential for understanding consumer experiences in the tourism industry. Therefore, researchers and applications use a range of techniques to analyze the content of data from customer review websites. For example, the effect of electronic word of mouth on tourism products in the hospitality industry has been studied by Sparks and Browning, (2011), Mishra *et al.* (2019), Calheiros *et al.* (2017) and Zhang (2019). Li *et al.* (2019) provide a complete review of text corpus-based tourism data mining from online reviews. In contrast, we scraped text from tourism facility websites

rather than from customer reviews to monitor the sustainability orientation of the facilities. We adopted a novel content analysis method, which relies on the TF-IDF weighting scheme, a commonly used tool in information retrieval. Among the existing TF-IDF versions, we implemented that proposed by Paik (2013) because it allowed us to consider documents of different lengths. Given the significant variability in the volume of text on the websites from which we extracted our data (see Figure A1 in Appendix 1), Paik's (2013) document length hypothesis was more relevant to our study.

3.2 Data

Data for this study were collected from the websites of accommodation providers located in the Veneto region. Veneto was chosen because it is one of the most appealing regions in Italy, attracts many tourists and is particularly active in producing guidelines to promote a gradual transition from mass tourism to quality sustainable tourism and responsible hiking (Bonzanigo et al., 2016; Della Lucia and Franch, 2017; Scuttari et al., 2019). The initial data set for this study was the "Elenco delle strutture ricettive turistiche della regione Veneto" ("List of tourist accommodation facilities in the Veneto region"), made available by the region's Open Data Veneto project [2] and updated daily. We downloaded the data set on February 3, 2022, and obtained an initial 9,132 accommodation facilities, classified into 11 categories, as shown in Table 1. We excluded 3,026 facilities for which no website was listed. We excluded an additional 680 cases that had typographic errors or whose websites referred
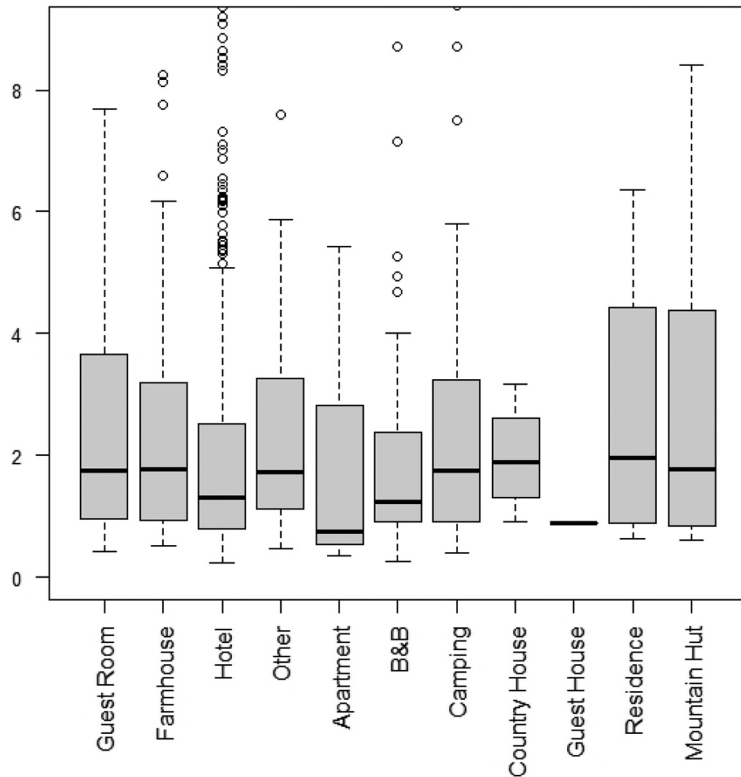

Figure 1.
Boxplots of total TF-IDF by type of structure

to Web aggregators. Of the 5,426 remaining sites processed by QIBA, 1,327 sites no longer existed, lacked information or used technology that blocked the scraper from collecting information. Of all websites, 44.9% contained correct and usable information. As a result, we restricted our analysis to 4,099 accommodation facilities.

Text from the websites was scraped using QIBA and collected in plain text format. After implementing several cleaning and preprocessing steps, we analyzed the Italian versions of the Web pages. Text was transformed into lower case to ensure equivalence among the same strings denoted in different cases.

To analyze the degree and characteristics of sustainability in tourism accommodation facilities, we adopted information retrieval techniques – the extraction of relevant information from the internet – using a query consisting of specific keywords. Assuming that a coherent list of words can effectively represent a given topic, we calculated the weight of each set of text according to the weight of the individual words contained in it.

Given the complexity of the concept of sustainability, we identified and analyzed three separate subtopics:

(1) sustainability in tourism;
(2) sustainability in tourism facility buildings; and
(3) certifications.

For each topic, we identified a basket of words (or "query") that characterized it. To select the terms for each basket, we used a hybrid approach by applying a two-step procedure. In the first step, we used a qualitative lexicon-based approach and manually selected the relevant terms for each basket based on the current literature. In the second step (query expansion), to enrich the initial baskets, we applied a word-embedding methodology by implementing a machine learning algorithm proposed by Mikolov *et al.* (2013). For the estimation of the model, we used the *word2vec* package of the statistical programming language R. With each word in the baskets being the input, we were able to identify similar words as the output, which were then added to the initial baskets.

The fundamental idea of this approach is that terms are first coded as *n*-dimensional vectors, which are then used to extract similar terms as outputs using natural language programming techniques [3]. Moreover, this model allows for the inclusion of unigrams and n-grams. The advantage of this methodology is that it not only enables the retrieval of terms initially missed by the researcher but also allows the researcher to search for terms in the

Type (in Italian)	Type (in English)	N	Website	%
Affittacamere	Guest room	995	425	42.7
Agriturismo	Farmhouse	897	426	47.5
Albergo	Hotel	3,165	2,052	64.8
Altra Ricettività	Other	246	88	35.8
Appartamento	Apartment	877	109	12.4
Bed and Breakfast	Bed and breakfast	2,437	713	29.3
Campeggio	Camping	185	128	69.2
Country House	Country house	44	20	45.5
Foresteria	Guest house	15	2	13.3
Residence	Residence	119	65	54.6
Rifugio	Mountain hut	152	71	46.7
Total		*9,132*	*4,099*	*44.9*

Table 1.
Tourist
accommodation
facilities in the
Veneto region

specific language of the corpus being analyzed. Finally, after applying *word2vec* and enriching the initial baskets, we manually checked the added words and selected only those that were relevant (see Appendix 2 for the basket lists).

4. Results and discussion

4.1 Do tourism businesses commit to sustainability? If so, to what extent?

First, we analyzed whether tourism companies in the Veneto region included sustainability topics on their websites. We built a basket of terms linked to sustainability and labeled company websites containing at least one of these words as "websites containing sustainability topics," as shown in Table 2.

Among the 4,099 sites analyzed, 19.9% (815) contained sustainability topics, suggesting a moderate commitment. Campsite and farmhouse accommodation showed the highest commitment to the communication of sustainability issues (40.6% and 34.5%, respectively). Results suggest a positive inclination toward sustainability but a heterogeneity in levels of commitment.

Second, to determine the strength of the sustainability commitment, we adopted the TF-IDF weighting scheme to classify websites according to the three baskets of words previously developed (see Table A1 in Appendix 2). The idea behind term weighting schemes, which are central to information retrieval systems, is that the higher the frequency of a certain term, the greater the emphasis that the company places on it, hence the more important it is to the company. Therefore, we used a TF-IDF indicator to reflect the importance of terms in the corpus, and then ranked company websites according to the strength of their communication about sustainability. The TF-IDF score for a term is higher when the term only appears in a small number of websites, thus helping with the identification of infrequent terms on an individual company's website. When a term appears on almost all websites, its TF-IDF score is lower. This enabled us to rank accommodation providers according to each sustainability subtopic as well as overall sustainability, which we obtained by combining their scores for each subtopic into a final score. This score shows the position of a company in the process toward complete sustainability.

Paik's TF-IDF uses two within-document TF normalizations, one being more effective for short queries and the other performing better on long queries. The final weight was then calculated by considering a weighted combination of these two components. The TF component follows three key hypotheses: the TF hypothesis, where a term's weight depends on its frequency; the advanced TF hypothesis, where the rate of change of a term's weight

Type	Website	Websites containing sustainability topics	%
Guest room	425	53	12.5
Farmhouse	426	147	34.5
Hotel	2,052	405	19.7
Other	88	28	31.8
Apartment	109	8	7.3
Bed and breakfast	713	95	13.3
Camping	128	52	40.6
Country house	20	5	25.0
Guest house	2	1	50.0
Residence	65	9	13.8
Mountain hut	71	12	16.9
Total	*4,099*	*815*	*19.9*

Table 2.
Websites containing sustainability topics

decreases with increased frequency (e.g. the change in term weight caused by increasing TF
from 2 to 3 is higher than that caused by increasing TF from 25 to 26; and the document
length hypothesis, which posits that long documents are more likely to use terms repeatedly;
thus, if two documents have different lengths and the same TF values for term t, the TF of t
should be higher for the shorter document.

In our TF-IDF version, TF included the following two aspects:

(1) relative intra-document, which measures the importance of a term by considering
its frequency relative to the average TF of the document; and

(2) length-regularized TF, which normalizes TF by considering the number of terms
present in the document.

Figure 1 shows a box plot analysis of the distribution of TF-IDF values in terms of the type
of structure. As shown in Figure 1, the structure type with the highest value (median and
third quartile) was residence, followed by country house, mountain hut and farmhouse.

Overall, the analysis responds to the first research question by showing evidence of the
existence of a moderate sustainability orientation in tourist accommodations in the Veneto
region. Camping sites and guesthouses exhibited the highest commitment to sustainability.
However, they were less intense in their communication about sustainability compared with
other facilities (country house, mountain hut and farmhouse). This demonstrates high
heterogeneity in tourist accommodation behaviors and suggests the lack of a cohesive
strategy for sustainability communication.

4.2 Soft and hard dimensions of sustainability engagement

Next, we explored whether companies that announce their compliance with sustainability
principles also implement concrete sustainability actions and officialize their actions by
obtaining sustainability certifications. We obtained estimates from two simple ordinary
least squares (OLS) regression analyses. The first tested the correlations between the TF-
IDF values of Basket 1 (declaration of commitment to sustainability principles) and Basket 2
(implementation of concrete sustainability actions), while the second tested the correlations
between the TF-IDF values of Basket 2 and Basket 3 (sustainability certifications). We chose
to estimate two simple OLS models because, given the exogeneity of regressors, the OLS
estimates were consistent and, as per the Gauss–Markov theorem, they were considered
optimal in the class of linear unbiased estimators. For both regressions, the coefficients were
positive and statistically significant (p-values < 0.001) (see Table A2 in Appendix 3). The
scatter plots shown in Figure 2 provide a graphical representation of the outcomes of the two
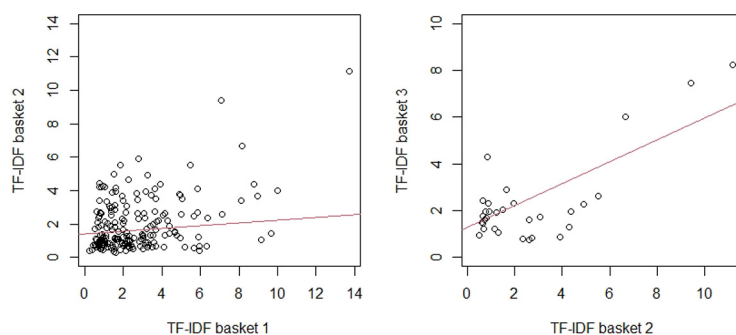

Figure 2.
Correlation between
TF-IDF of Baskets 1
and 2 (*left*) and
Baskets 2 and 3
(*right*)

correlations. Figure 2 shows that when a greater number of companies agree on sustainability principles, this leads to the implementation of more concrete actions, further prompting more companies to officialize their actions by obtaining sustainability certifications.

The correlation between Basket 1 and Basket 2 was positive and statistically significant, although R^2 was lower than that of the correlation between Basket 2 and Basket 3 [4]. These results are consistent with the fact that a generic adhesion to sustainable tourism was detected in Basket 1. Adherence to sustainable tourism values by tourist structures can manifest in various ways. In some cases, an adherence to sustainability values preceded the execution of some actions (collected in Basket 2), while in others it did not. For example, the terms *environmental impact* and *ecological footprint* (in Basket 1) should be accompanied by concrete actions such as *reduced water consumption*, *separate collection* or *recycling* (listed in Basket 2). Terms such as *slow tourism*, *eco-trekking* and *cycle tourism* are not necessarily linked to concrete actions. The strong link (a high R^2) between Baskets 2 and 3 is not surprising – it is necessary to implement concrete sustainable actions to obtain certification (Basket 3), which sometimes requires interventions in the facility's buildings. The terms listed in Basket 2 describe several of these actions. The regression results indicate that companies that obtain sustainable certifications advertise their efforts on their websites by emphatically declaring the sustainability actions undertaken (high TF-IDF). This suggests the existence of a trend toward strong sustainability engagement that goes beyond greenwashing, thus answering our second research question.

4.3 Geographical and cognitive proximity networks
4.3.1 Geographical proximity. We explored the spatial distribution of sustainable facilities by adopting LLS as a unit of analysis across two stages. In the first stage, we mapped the proportion of sustainable tourist accommodation facilities in the total number of facilities in the LLS (see Figure 3). We further classified LLSs according to the tourism intensity index, which measures the relevance of tourism to the local economy. The indicator shown in Figure 4 represents the ratio of the number of workers employed in the tourism and
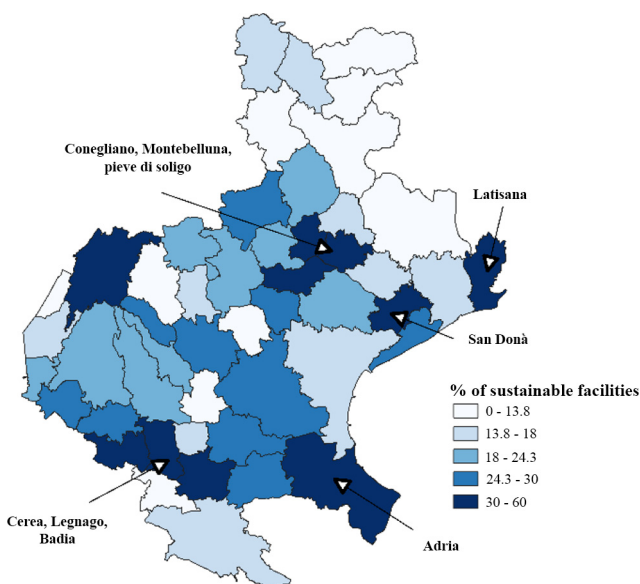

Figure 3.
Share of sustainable facilities by LLS

Tourism Intensity Indix

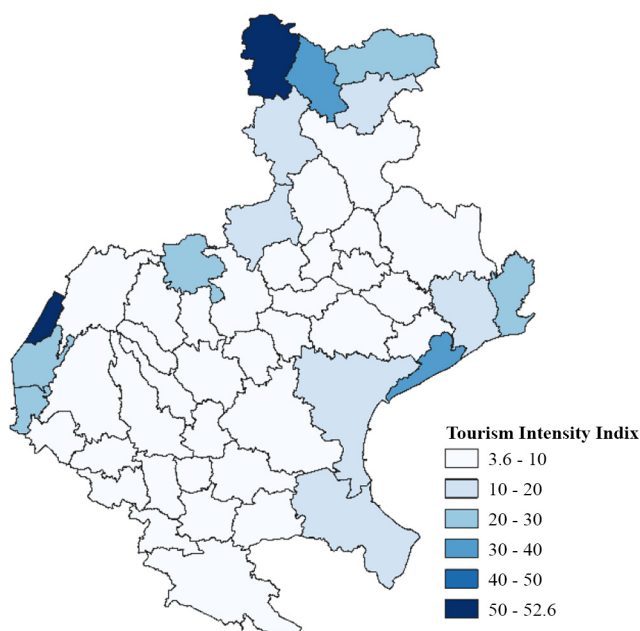| | 3.6 - 10 |
	10 - 20
	20 - 30
	30 - 40
	40 - 50
	50 - 52.6

Figure 4.
Touristic LLS

hospitality industry (ATECO code "I – Attività dei servizi di alloggio e di ristorazione") to the total number of employees in the LLS. We define a "touristic LLS" as one with more than 10% of its total workforce employed in tourism (see Figure 4).

This raw index analysis shows that only two of the LLSs – Latisana and Adria, both of which were characterized by a high proportion of sustainable accommodation facilities among the total facilities available – could be labeled as touristic. Notably, apart from Latisana, which houses the seaside destination of Bibione, these LLSs are situated away from the classic tourist destinations. This indicates that the more popular a tourist destination is, the lower the sustainability commitment of its tourism facilities. It appears that the lesser known areas were the ones that invested the most in sustainability.

Two clusters of LLSs with a high ratio of sustainable facilities emerged – one located in the southernmost part of the Veneto region (Cerea, Legnago and Badia) and the other in the central part of Veneto, including the LLSs of Conegliano, Montebelluna and Pieve di Soligo. These two clusters appear to share a need to offer alternative experiences to those offered by traditional tourism destinations (e.g. beaches, mountains or art). The first cluster is located in the heart of the Po Valley, while the second offers a tourism experience linked to wine (prosecco) production and hilly landscapes. However, the two clusters differed significantly in terms of the number of facilities: the first cluster had only 28 facilities, while the second had 111.

In the second stage, we mapped the spatial distribution of the top 246 facilities ranked in the sustainable intensity index (third quartile – Top 25%) with regard to their website content. Figure 5 illustrates the number of facilities in each LLS included in the Top 25%, while Figure 6 illustrates the share of the Top 25% facilities in the TF-IDF ranking in the total number of facilities.

The second LLS cluster (Conegliano, Pieve di Soligo and Montebelluna) ranked the highest in terms of having the greatest proportion of facilities strongly addressing sustainability issues
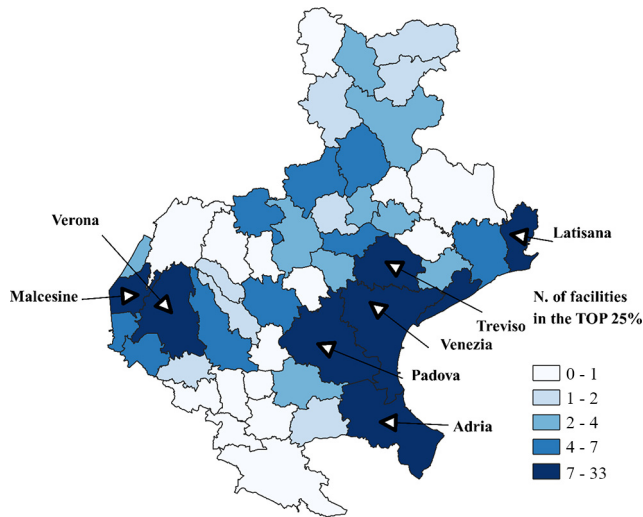
Figure 5.
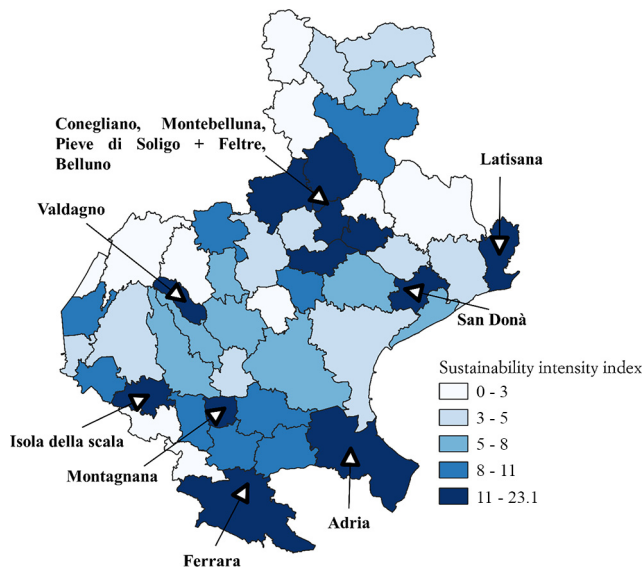Sustainability
intensity index –
number or facilities
by LLS

Figure 6.
Sustainability
intensity index, share
of facilities on 3rd
quartile by LLS

in their websites (see Figure 6). By comparing these results with those derived from simply counting the number of sustainable facilities in each LLS (see Figure 3), we observed that the cluster expands to the LLSs of Feltre (also a touristic LLS) and Belluno. As shown in Figure 3, San Donà di Piave, Latisana and Adria appear to be leaders in sustainable facilities. This analysis shows the emergence of the LLSs of Montagnana, Valdagno, Isola della Scala and Ferrara. It should be emphasized that Montagnana and Ferrara are LLSs with a tiny number of facilities (fewer than ten units), and thus are probably of little significance for our analysis.

Overall, the empirical evidence from the second stage of the analysis confirms that peripheral LLSs, which are probably not the most attractive from a tourist point of view, are characterized by a higher proportion of facilities that place importance on communicating sustainability through their websites. This is an unexpected and interesting result that is worthy of further investigation.

This empirical evidence responds to the third research question, suggesting the need to consider clusters of sustainable tourism as important units of analysis for managers and policymakers willing to enhance the sustainability of their facilities. Over time, economic actors in neighboring LLSs facing the same challenge – the need to offer an alternative to classic tourism offerings – can develop synergic strategies to plan a transition toward sustainable tourism.

4.3.2 Cognitive proximity. Data collected through Web scraping and filtered through the three baskets of sustainability terms (Table A1) enabled the construction of sustainable accommodation facility networks. Each unweighted, undirected network consisted of a set of nodes, each representing a single accommodation facility (or, more precisely, its website) connected by links or edges denoting the use of one of the terms in our baskets. In other words, if two tourism companies used a term from one of our baskets on their websites, we considered them linked. The general properties of the networks are listed in Table 3. Modularity is a measure of the quality of a particular division of a network (Newman and Girvan, 2004), while network density is the ratio of observed edges to the number of possible edges for a given network. The Clauset–Newmann–Moore hierarchical agglomeration algorithm (Clauset *et al.*, 2004) was deployed to discover the community structure in the networks for each basket of words.

From the analysis of Basket 1 (terms that refer generically to sustainability), four clusters of companies emerged (see Figure 7). The first cluster (black nodes) comprised 293 websites characterized by rich sustainability terminology such as *sustainable*, *km0*, *environmental impact*, *biodiversity*, *eco-friendly* and *ecology*. The websites in this cluster had the highest TF-IDF values. The second cluster (green nodes) comprised 203 websites characterized by the strong use of terms such as *cycle tourism* and *environmental sustainability* and low TF-IDF values. The third cluster (red nodes) comprised 163 websites characterized by the use of terms such as *biodiversity* and *ecology* and intermediate TF-IDF values. Finally, the fourth cluster (blue nodes) comprised 47 websites with a less intense focus on sustainability and using generic words such as *ecological*, among others.

Six clusters emerged from the analysis of Basket 2 (terms concerning the sustainability of buildings) (see Figure 8). The first cluster (black nodes) comprised 73 websites focusing on energy efficiency. The general intensity with which they spoke of sustainability was not particularly high. The second cluster (light green nodes) comprised 72 websites that used highly articulate language, including terms such as *emissions*, *green hotels*, *energy efficiency*, *photovoltaic*, *waste sorting*, *environmentally friendly* and *solar panels*. The general strength of terms pertaining to sustainability was highest for this cluster. The third cluster (red nodes) included 61 sites using terms such as *recycling*, *renewable energy*, *waste sorting* and *eco-hotel*

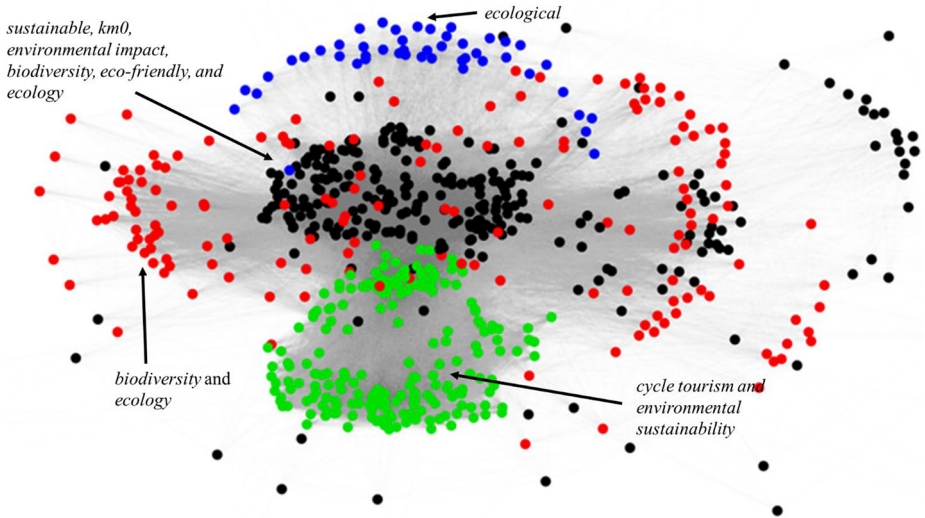| Network properties	Basket 1	Basket 2	Basket 3
Number of nodes	706	299	54
Number of links	70,323	12,329	320
Modularity	0.31	0.25	0.53
Density	0.28	0.28	0.22

Table 3.
Network indicators

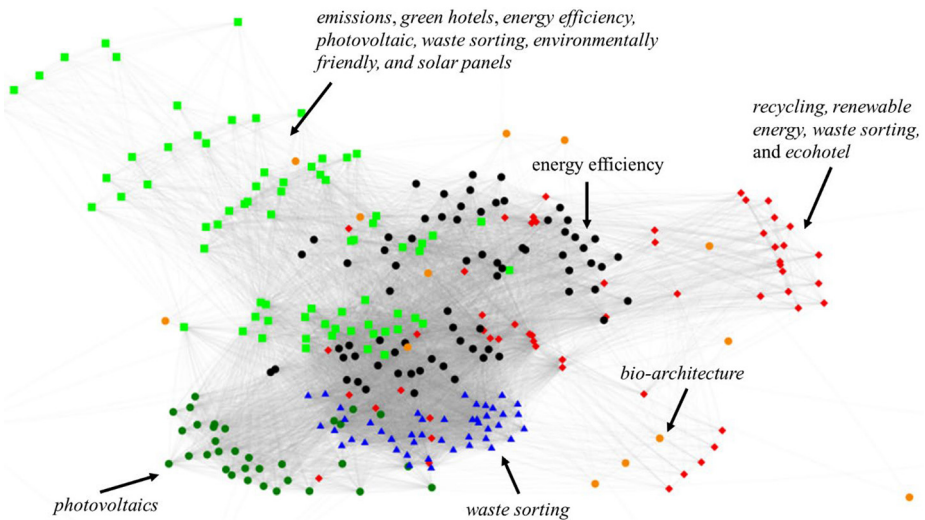**Figure 7.**
Topic clusters,
Basket 1

Figure 8.
Topic clusters,
Basket 2

and carrying a high TF-IDF value. The fourth cluster (blue nodes) comprised 46 websites, with the most frequent term being *waste sorting* and their focus on sustainability being the lowest. The fifth cluster (34 websites, dark green nodes) mainly spoke about *photovoltaics*, with the level of intensity being medium-low. The sixth and final cluster (dark orange nodes) was the least linked to the others. It comprised 13 websites in which the term *bio-architecture* was used. The level of attention to sustainability was medium-high.

From the analysis of Basket 3, which accounted for sustainability certifications, six different clusters emerged (see Figure 9). The first cluster (black nodes) comprised 18 companies that used the terms *Ecolabel* and *Legambiente Turismo* on their websites, while

fsc

Casaclima

ecolabel and *Lega
Ambiente Turismo*
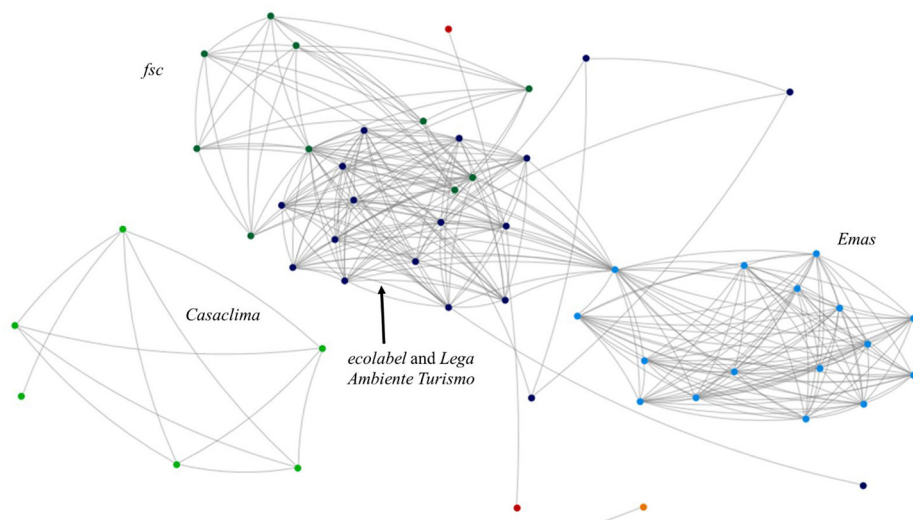
Emas

the second cluster (light blue nodes) comprised 16 companies using the term *EMAS* (Eco-Management and Audit Scheme). The third cluster (dark green nodes) comprised 10 companies characterized by the use of the word *FSC* (Forest Stewardship Council), whereas the fourth cluster (light green nodes) comprised six companies that used the word *CasaClima*. The final two clusters, each consisting of two companies, used the terms *Beluga* and *ESG* (environmental, social and governance). The companies in the fourth cluster (*CasaClima*) were characterized by the highest TF-IDF value, indicating the highest intensity of discussion on the topic of sustainability on their websites.

These results, which respond to our fourth research question, support the existence of networks of tourist accommodation facilities that are linked by strong cognitive proximity.

5. Conclusions and implications
Recently, several initiatives aimed at contributing to the creation of clusters have been undertaken throughout Italy, especially in the field of sustainable tourism. These initiatives focus on the development of circular economy practices and the creation of synergistic relationships with local communities. To boost the resilience and competitiveness of the tourism sector post-COVID-19, it has become imperative to animate the sector economically, socially and culturally, stimulating the entire supply chain. However, knowledge about the relational structure of sustainable tourism networks is scant. This research aimed to apply data mining and social network analysis to explore the cognitive and geographical proximity of tourist destinations in the Veneto region, highlighting the existence of networks of destinations characterized by sustainability-driven isomorphism (Masocha and Fatoki, 2018). Our research indicates that tourism operators in peripheral areas are often more committed to sustainable behaviors and collaborate with other tourist destinations to create a sustainable tourism ecosystem.

5.1 Theoretical implications
Our study of tourism networks is one of the first, to the best of the authors' knowledge, to explore not only the significant interconnections existing between the structural components of

tourism businesses but also the links that allow tourism companies to cooperate in the implementation of sustainable tourism solutions. Recognizing the relational structure of tourist destinations is crucial for developing collaborative innovation networks (Marasco *et al.*, 2018; Monte *et al.*, 2011) because cognitive and geographical proximity can facilitate the flow of knowledge across organizational boundaries from an open innovation perspective (Chesbrough, 2003). From a theoretical point of view, this study makes a threefold contribution. First, it highlights that spatial and relational dynamics are crucial to the functioning of the tourism industry. Sustainable tourism trajectories should be investigated through the lens of a new theoretical framework that combines cognitive and geographical proximity of organizations. Second, social network analysis of data mined from tourism company websites can reveal crucial information about the sustainable orientation of tourism destinations, allowing for the accurate mapping of sustainable tourism ecosystems. Third, we confirmed that sustainable tourism is strongly related to environmental sustainability by revealing its interconnections with the social and governance dimensions.

5.2 Practical implications

Our findings may inform managers and policymakers about the importance of cognitive and spatial proximity in sustaining the innovative trajectories of tourist organizations. Managers may be willing to formalize networks with similar organizations located nearby to create a robust and sustainable tourism offering targeting the national or international LOHAS market segment (Kotler, 2011; Pan *et al.*, 2018; Więckowski, 2021).

At the destination level, sustainable business model innovation may be supported not only by tacit knowledge transfer between organizations through staff mobilization and direct observation but also by trade associations and local research institutions acting as knowledge brokers through initiatives such as seminars, meetings and conventions (Shaw and Williams, 2009) .

Finally, current communications about sustainability appear to be departing from the greenwashing that was prevalent at the beginning of the millennium, creating space for a more genuine sustainability orientation that entails specific environmental and social practices. As a result, managers should be increasingly willing to aim for sustainability certifications (Lesar *et al.*, 2020; Sampaio *et al.*, 2012). A higher commitment toward sustainability among tourist accommodation providers is crucial and should be properly represented throughout their communication channels (e.g. websites and social media), as clearly stated by Blasi *et al.* (2021).

Peripheral tourist destinations appear more likely to adopt sustainability practices to better position themselves against standard touristic offerings. Policymakers should acknowledge this trend and initiate territorial sustainability certifications for tourist destinations to mitigate tourist flow inequalities and avoid the phenomenon of overtourism (Mihalic and Kuščer, 2022). Signaling the sustainability of tourist offerings at the local level may attract tourists away from the most crowded destinations, promoting a more distributed system of tourist attractions. For instance, Venice, a fragile historic city, has attracted an unsustainable flow of tourists (Bertocchi and Visentin, 2019). To mitigate the dark side of cruise tourism in Venice, the Council of Ministers in Italy has mandated that from August 1, 2021, large ships can longer sail in front of San Marco or through the Giudecca Canal. However, policymakers should not limit themselves to responding to these issues but also promote alternative tourism routes directed at the LOHAS market segment.

5.3 Limitations and further research

A limitation of our research is its limited geographical scope. Future studies on sustainable tourism could compare other Italian regions or conduct a cross-country analysis. In terms of

content analysis, we acknowledge the limitations of selecting only specific baskets of words to capture the sustainability of tourist accommodations. Nevertheless, sophisticated mathematical methods and algorithms were applied to limit subjectivity issues. Further analysis could be oriented toward evaluating specific collaborative initiatives in different tourist destination clusters. Additionally, qualitative analyses based on primary data collection could be conducted to better understand whether cognitive and geographical proximity affect the probability of embarking on joint actions .

Notes

1. The methodology of LLS identification was developed by Sforzi (1997).

2. https://dati.veneto.it/opendata/elenco_strutture_ricettive_del_veneto

3. See Mikolov *et al.* (2013) for details on the methodology.

4. R^2 spans from 0 to 1; therefore, values higher than 0.5 indicate high correlation, while values lower than 0.5 indicate low correlation (Hair *et al.*, 2006).

References

Adner, R. (2017), "Ecosystem as structure", *Journal of Management*, Vol. 43 No. 1, pp. 39-58, doi: 10.1177/0149206316678451.

Antonides, G. (2017), "Sustainable consumer behaviour: a collection of empirical studies", *Sustainability*, Vol. 9 No. 10, p. 1686, doi: 10.3390/su9101686.

Arora, S.K., Li, Y., Youtie, J. and Shapira, P. (2016), "Using the wayback machine to mine websites in the social sciences: a methodological resource", *Journal of the Association for Information Science and Technology*, Vol. 67 No. 8, pp. 1904-1915, doi: 10.1002/asi.23503.

Asheim, B.T. (1996), "Industrial districts as 'learning regions': a condition for prosperity", *European Planning Studies*, Vol. 4 No. 4, pp. 379-400, doi: 10.1080/09654319608720354.

Baggio, R. and Cooper, C. (2011), "Knowledge transfer in a tourism destination: the effects of a network structure", in Scott, N. and Laws, E. (Eds), *Advances in Service Network Analysis*, Routledge, Oxon, pp. 183-198.

Barcaroli, G., Nurra, A., Scarnò, M. and Summa, D. (2014), "Use of web scraping and text mining techniques in the istat survey on information and communication technology in enterprises", *Proceedings of Quality Conference*, pp. 33-38.

Beaumont, N. (2011), "The third criterion of ecotourism: are ecotourists more concerned about sustainability than other tourists?", *Journal of Ecotourism*, Vol. 10 No. 2, pp. 135-148, doi: 10.1080/14724049.2011.555554.

Becattini, G. (1979), "Dal settore industriale al distretto industriale. Alcune considerazioni sull'unità di indagine dell'economia industriale", il Mulino, Bologna.

Becattini, G. (2000), *Il distretto industriale: un nuovo modo di interpretare il cambiamento economico*, Rosenberg and Sellier, Turin.

Becattini, G., Bellandi, M., Dei Ottati, G. and Sforzi, F. (2001), *Il caleidoscopio dello sviluppo locale: trasformazioni economiche nell'Italia contemporanea*, Rosenberg and Sellier, Turin, Vol. 2.

Belussi, F. (1999), "Policies for the development of knowledge-intensive local production systems", *Cambridge Journal of Economics*, Vol. 23 No. 6, pp. 729-747, doi: 10.1093/cje/23.6.729.

Belussi, F., Sabbadin, E. and Sedita, S.R. (2022), "Make or buy? Global clusters in sustainability transitions", in Tollin, N.V.J., Masi, D. and Brambini, A. (Eds), *Handbook on Global Production Systems: exploring the Role of the SDGs*, Edward Elgar, Cheltenham.

Bertocchi, D. and Visentin, F. (2019), "'The overwhelmed city': physical and social over-capacities of global tourism in venice", *Sustainability*, Vol. 11 No. 24, p. 6937, doi: 10.3390/su11246937.

Blasi, S. and Sedita, S.R. (2020), "The diffusion of a policy innovation in the energy sector: evidence from the collective switching case in Europe", *Industry and Innovation*, Vol. 27 No. 6, pp. 680-704, doi: 10.1080/13662716.2019.1616535.

Blasi, S., Crisafulli, B. and Sedita, S.R. (2021), "Selling circularity: understanding the relationship between circularity promotion and the performance of manufacturing SMEs in Italy", *Journal of Cleaner Production*, Vol. 303, p. 127035, doi: 10.1016/j.jclepro.2021.127035.

Bonzanigo, L., Giupponi, C. and Balbi, S. (2016), "Sustainable tourism planning and climate change adaptation in the alps: a case study of winter tourism in mountain communities in the dolomites", *Journal of Sustainable Tourism*, Vol. 24 No. 4, pp. 637-652, doi: 10.1080/09669582.2015.1122013.

Brusco, S. (1989), *Piccole Imprese e Distretti Industriali: una Raccolta di Saggi*, Rosenberg and Sellier, Turin.

Butler, R.W. (1999), "Sustainable tourism: a state-of-the-art review", *Tourism Geographies*, Vol. 1 No. 1, pp. 7-25, doi: 10.1080/14616689908721291.

Calheiros, A.C., Moro, S. and Rita, P. (2017), "Sentiment classification of consumer-generated online reviews using topic modeling", *Journal of Hospitality Marketing and Management*, Vol. 26 No. 7, pp. 675-693, doi: 10.1080/19368623.2017.1310075.

Capone, F. and Lazzeretti, L. (2018), "The different roles of proximity in multiple informal network relationships: evidence from the cluster of high technology applied to cultural goods in tuscany", *Industry and Innovation*, Vol. 25 No. 9, pp. 897-917, doi: 10.1080/13662716.2018.1442713.

Casanueva, C., Gallego, Á. and García-Sánchez, M.-R. (2016), "Social network analysis in tourism", *Current Issues in Tourism*, Vol. 19 No. 12, pp. 1190-1209, doi: 10.1080/13683500.2014.990422.

Chesbrough, H.W. (2003), *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Harvard Business Press, Boston.

Clauset, A., Newman, M.E.J. and Moore, C. (2004), "Finding community structure in very large networks", *Physical Review E – Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, Vol. 70 No. 6, p. 6, doi: 10.1103/PHYSREVE.70.066111.

Cohen, W.M. and Levinthal, D.A. (1990), "Absorptive capacity: a new perspective on learning and innovation", *Administrative Science Quarterly*, Vol. 35 No. 1, p. 128, doi: 10.2307/2393553.

Córcoles Muñoz, M.M., Parra Requena, G., García Villaverde, P.M. and Ruiz Ortega, M.J. (2022), "Relational antecedents of sustainability orientation in hospitality and tourism firms: the mediating role of absorptive capacity", *Journal of Sustainable Tourism*, pp. 1-23, doi: 10.1080/09669582.2021.2023166.

Croce, A.I., Musolino, G., Rindone, C. and Vitetta, A. (2019), "Sustainable mobility and energy resources: a quantitative assessment of transport services with electrical vehicles", *Renewable and Sustainable Energy Reviews*, Vol. 113, p. 109236, doi: 10.1016/j.rser.2019.06.043.

de Lange, D. and Dodds, R. (2017), "Increasing sustainable tourism through social entrepreneurship", *International Journal of Contemporary Hospitality Management*, Vol. 29 No. 7, pp. 1977-2002, doi: 10.1108/IJCHM-02-2016-0096.

Della Lucia, M. and Franch, M. (2017), "The effects of local context on world heritage site management: the dolomites natural world heritage site, Italy", *Journal of Sustainable Tourism*, Vol. 25 No. 12, pp. 1756-1775, doi: 10.1080/09669582.2017.1316727.

DiMaggio, P.J. and Powell, W.W. (1983), "The iron cage revisited: institutional isomorphism and collective rationality in organizational fields", *American Sociological Review*, Vol. 48 No. 2, pp. 147-160.

Elche, D., García-Villaverde, P.M. and Martínez-Pérez, Á. (2018), "Inter-organizational relationships with core and peripheral partners in heritage tourism clusters", *International Journal of Contemporary Hospitality Management*, Vol. 30 No. 6, pp. 2438-2457, doi: 10.1108/IJCHM-11-2016-0611.

Font, X., Torres-Delgado, A., Crabolu, G., Palomo Martinez, J., Kantenbacher, J. and Miller, G. (2021), "The impact of sustainable tourism indicators on destination competitiveness: the european tourism indicator system", *Journal of Sustainable Tourism*, pp. 1-24, doi: 10.1080/09669582.2021.1910281.

Frenken, K., Van Oort, F. and Verburg, T. (2007), "Related variety, unrelated variety and regional economic growth", *Regional Studies*, Vol. 41 No. 5, pp. 685-697, doi: 10.1080/00343400601120296.

Gehman, J. and Grimes, M. (2017), "Hidden badge of honor: how contextual distinctiveness affects category promotion among certified B corporations", *Academy of Management Journal*, Vol. 60 No. 6, pp. 2294-2320.

Giuliani, E. (2013), "Network dynamics in regional clusters: evidence from Chile", *Research Policy*, Vol. 42 No. 8, pp. 1406-1419, doi: 10.1016/j.respol.2013.04.002.

Gök, A., Waterworth, A. and Shapira, P. (2015), "Use of web mining in studying innovation", *Scientometrics*, Vol. 102 No. 1, pp. 653-671, doi: 10.1007/s11192-014-1434-0.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2006), "Multivariate data analysis 6th edition", *Pearson Prentice Hall. New Jersey. Humans: Critique and Reformulation, Journal of Abnormal Psychology*, Vol. 87, pp. 49-74.

Hassan, S.S. (2000), "Determinants of market competitiveness in an environmentally sustainable tourism industry", *Journal of Travel Research*, Vol. 38 No. 3, pp. 239-245, doi: 10.1177/004728750003800305.

Ivars-Baidal, J.A., Vera-Rebollo, J.F., Perles-Ribes, J., Femenia-Serra, F. and Celdrán-Bernabeu, M.A. (2021), "Sustainable tourism indicators: what's new within the smart city/destination approach?", *Journal of Sustainable Tourism*, pp. 1-24, doi: 10.1080/09669582.2021.1876075.

Jacobides, M.G., Cennamo, C. and Gawer, A. (2018), "Towards a theory of ecosystems", *Strategic Management Journal*, Vol. 39 No. 8, pp. 2255-2276, doi: 10.1002/smj.2904.

Jain, S. and Sharma, P. (2021), "Understanding customer perception on sustainable tourism and related practices using cluster analysis: a study in indian context", *Paradigm*, Vol. 25 No. 2, pp. 122-140, doi: 10.1177/09718907211046155.

Jeon, M.M., Kang, M. and Desmarais, E. (2016), "Residents' perceived quality of life in a cultural-heritage tourism destination", *Applied Research in Quality of Life*, Vol. 11 No. 1, pp. 105-123, doi: 10.1007/s11482-014-9357-8.

Jones, A. (1987), "Green tourism", *Tourism Management*, Vol. 8 No. 4, pp. 354-356, doi: 10.1016/0261-5177(87)90095-1.

Jones, K.S. (1972), "A statistical interpretation of term specificity and its application in retrieval", *Journal of Documentation*.

Kinne, J. and Resch, B. (2018), "Generating big spatial data on firm innovation activity from text-mined firm websites", *GI_Forum*, Vol. 1, pp. 82-89.

Kitthamkesorn, S. and Chen, A. (2017), "Alternate weibit-based model for assessing green transport systems with combined mode and route travel choices", *Transportation Research Part B: Methodological*, Vol. 103, pp. 291-310, doi: 10.1016/j.trb.2017.04.011.

Knoben, J. and Oerlemans, L.A.G. (2006), "Proximity and inter-organizational collaboration: a literature review", *International Journal of Management Reviews*, Vol. 8 No. 2, pp. 71-89, doi: 10.1111/j.1468-2370.2006.00121.x.

Kotler, P. (2011), "Reinventing marketing to manage the environmental imperative", *Journal of Marketing*, Vol. 75 No. 4, pp. 132-135, doi: 10.1509/jmkg.75.4.132.

Lazzeretti, L. and Capone, F. (2008), "Mapping and analysing local tourism systems in Italy, 1991–2001", *Tourism Geographies*, Vol. 10 No. 2, pp. 214-232, doi: 10.1080/14616680802000055.

Lazzeretti, L. and Petrillo, C.S. (2006), *Tourism Local Systems and Networking*, Routledge, London.

Lazzeretti, L., Sartori, A. and Innocenti, N. (2015), "Museums and social media: the case of the museum of natural history of florence", *International Review on Public and Nonprofit Marketing*, Vol. 12 No. 3, pp. 267-283, doi: 10.1007/s12208-015-0136-5.

Lazzeretti, L., Sedita, S.R. and Caloffi, A. (2014), "Founders and disseminators of cluster research", *Journal of Economic Geography*, Vol. 14 No. 1, pp. 21-43, doi: 10.1093/jeg/lbs053.

Lehto, X.Y. and Lehto, M.R. (2019), "Vacation as a public health resource: toward a wellness-centered tourism design approach", *Journal of Hospitality and Tourism Research*, Vol. 43 No. 7, pp. 935-960, doi: 10.1177/1096348019849684.

Lenzen, M., Sun, Y.-Y., Faturay, F., Ting, Y.-P., Geschke, A. and Malik, A. (2018), "The carbon footprint of global tourism", *Nature Climate Change*, Vol. 8 No. 6, pp. 522-528, doi: 10.1038/s41558-018-0141-x.

Lesar, L., Weaver, D. and Gardiner, S. (2020), "Beyond certification: an empirically expanded quality control tool 'multiverse' for sustainable tourism", *Journal of Sustainable Tourism*, Vol. 28 No. 10, pp. 1625-1645, doi: 10.1080/09669582.2020.1745218.

Li, Q., Li, S., Zhang, S., Hu, J. and Hu, J. (2019), "A review of text corpus-based tourism big data mining", *Applied Sciences*, Vol. 9 No. 16, p. 3300, doi: 10.3390/app9163300.

Line, N.D., Hanks, L. and Miao, L. (2018), "Image matters: incentivizing green tourism behavior", *Journal of Travel Research*, Vol. 57 No. 3, pp. 296-309, doi: 10.1177/0047287517697848.

Lue, C.-C., Crompton, J.L. and Fesenmaier, D.R. (1993), "Conceptualization of multi-destination pleasure trips", *Annals of Tourism Research*, Vol. 20 No. 2, pp. 289-301, doi: 10.1016/0160-7383(93)90056-9.

Manz, C.C. and Sims, H.P. (1981), "Vicarious learning: the influence of modeling on organizational behavior", *The Academy of Management Review*, Vol. 6 No. 1, pp. 105-113, doi: 10.5465/amr.1981.4288021.

Marasco, A., De Martino, M., Magnotti, F. and Morvillo, A. (2018), "Collaborative innovation in tourism and hospitality: a systematic review of the literature", *International Journal of Contemporary Hospitality Management*, Vol. 30 No. 6, pp. 2364-2395, doi: 10.1108/IJCHM-01-2018-0043.

Marshall, A. (1920), *Principles of Economics 8th Edition*, Macmillan, London.

Maskell, P. (2017), "Towards a knowledge-based theory of the geographical cluster", *Economy: Critical Essays in Human Geography*, pp. 377-399, doi: 10.4324/9781351159203-18.

Masocha, R. and Fatoki, O. (2018), "The impact of coercive pressures on sustainability practices of small businesses in South Africa", *Sustainability*, Vol. 10 No. 9, p. 3032, doi: 10.3390/su10093032.

Meyer, J.W. and Rowan, B. (1977), "Institutionalized organizations: formal structure as myth and ceremony", *American Journal of Sociology*, Vol. 83 No. 2, pp. 340-363, doi: 10.1086/226550.

Mihalic, T. and Kuščer, K. (2022), "Can overtourism be managed? Destination management factors affecting residents' irritation and quality of life", *Tourism Review*, Vol. 77 No. 1, pp. 16-34, doi: 10.1108/TR-04-2020-0186.

Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013), "Efficient estimation of word representations in vector space", *1st International Conference on Learning Representations, ICLR 2013 – Workshop Track Proceedings*, doi: 10.48550/arxiv.1301.3781.

Mishra, R.K., Urolagin, S. and Jothi J, A.A. (2019), "A sentiment analysis-based hotel recommendation using TF-IDF approach", *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pp. 811-815, doi: 10.1109/ICCIKE47802.2019.9004385.

Miska, C., Szőcs, I. and Schiffinger, M. (2018), "Culture's effects on corporate sustainability practices: a multi-domain and multi-level view", *Journal of World Business*, Vol. 53 No. 2, pp. 263-279, doi: 10.1016/j.jwb.2017.12.001.

Monte, A.D., D'Esposito, M.R., Giordano, G. and Vitale, M.P. (2011), "Analysis of collaborative patterns in innovative networks", *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, Berlin, Heidelberg, pp. 77-84, doi: 10.1007/978-3-642-11363-5_9.

Morrison, A. (2008), "Gatekeepers of knowledge within industrial districts: who they are, how they interact", *Regional Studies*, Vol. 42 No. 6, pp. 817-835, doi: 10.1080/00343400701654178.

Nahapiet, J. and Ghoshal, S. (1998), "Social capital, intellectual capital, and the organizational advantage", *The Academy of Management Review*, Vol. 23 No. 2, pp. 242-266, doi: 10.5465/amr.1998.533225.

Newman, M.E.J. and Girvan, M. (2004), "Finding and evaluating community structure in networks", *Physical Review E*, Vol. 69 No. 2, p. 26113, doi: 10.1103/PhysRevE.69.026113.

Nooteboom, B. (1999), "Innovation, learning and industrial organisation", *Cambridge Journal of Economics*, Vol. 23 No. 2, pp. 127-150, doi: 10.1093/cje/23.2.127.

Osti, L. and Goffi, G. (2021), "Lifestyle of health and sustainability: the hospitality sector's response to a new market segment", *Journal of Hospitality and Tourism Management*, Vol. 46, pp. 360-363, doi: 10.1016/j.jhtm.2021.01.008.

Paik, J.H. (2013), "A novel TF-IDF weighting scheme for effective ranking", *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, doi: 10.1145/2484028.

Pan, S.-Y., Gao, M., Kim, H., Shah, K.J., Pei, S.-L. and Chiang, P.-C. (2018), "Advances and challenges in sustainable tourism toward a green economy", *Science of The Total Environment*, Vol. 635, pp. 452-469, doi: 10.1016/j.scitotenv.2018.04.134.

Porter, M.E. (1990), *The Competitive Advantage of Nations*, The Free Press, New York, NY.

Prats, L., Guia, J. and Molina, F.-X. (2008), "How tourism destinations evolve: the notion of tourism local innovation system", *Tourism and Hospitality Research*, Vol. 8 No. 3, pp. 178-191, doi: 10.1057/thr.2008.24.

Putnam, R.D. (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon and Schuster, New York.

Ramkissoon, H. (2020), "Perceived social impacts of tourism and quality-of-life: a new conceptual model", *Journal of Sustainable Tourism*, pp. 1-17, doi: 10.1080/09669582.2020.1858091.

Ruiz-Ortega, M.J., Parra-Requena, G. and García-Villaverde, P.M. (2021), "From entrepreneurial orientation to sustainability orientation: the role of cognitive proximity in companies in tourist destinations", *Tourism Management*, Vol. 84, p. 104265, doi: 10.1016/j.tourman.2020.104265.

Sampaio, A.R., Thomas, R. and Font, X. (2012), "Small business management and environmental engagement", *Journal of Sustainable Tourism*, Vol. 20 No. 2, pp. 179-193, doi: 10.1080/09669582.2011.602193.

Saxenian, A. (1996), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128, with a New Preface by the Author*, Harvard University Press.

Scott, A.J. (1988), *New Industrial Spaces: Flexible Production Organization and Regional Development in North America and Western Europe (Issue 3)*, Pion Limited.

Scuttari, A., Orsi, F. and Bassani, R. (2019), "Assessing the tourism-traffic paradox in Mountain destinations. A stated preference survey on the dolomites' passes (Italy)", *Journal of Sustainable Tourism*, Vol. 27 No. 2, pp. 241-257, doi: 10.1080/09669582.2018.1428336.

Sedita, S.R. and Blasi, S. (2021), *Rethinking Clusters: Place-Based Value Creation in Sustainability Transitions*, Springer.

Sedita, S.R., De Noni, I. and Pilotti, L. (2017), "Out of the crisis: an empirical investigation of place-specific determinants of economic resilience", *European Planning Studies*, Vol. 25 No. 2, pp. 155-180, doi: 10.1080/09654313.2016.1261804.

Sedita, S.R., Hoffmann, V.E., Guarnieri, P. and Toso Carraro, E. (2021), "Prosecco has another story to tell: the coexistence of multiple knowledge networks in the same value chain", *International Journal of Wine Business Research*, Vol. 33 No. 4, pp. 502-522, doi: 10.1108/IJWBR-06-2020-0024.

Sforzi, F. (Ed.) (1997), *I sistemi locali del lavoro 1991*, Roma, Istat.

Shaw, J. (2004), "Travel choices? The Central Scotland transport corridor studies", *Scottish Geographical Journal*, Vol. 120 No. 4, pp. 289-309, doi: 10.1080/00369220418737213.

Shaw, G. and Williams, A. (2009), "Knowledge transfer and management in tourism organisations: an emerging research agenda", *Tourism Management*, Vol. 30 No. 3, pp. 325-335.

Sørensen, F. (2007), "The geographies of social networks and innovation in tourism", *Tourism Geographies*, Vol. 9 No. 1, pp. 22-48, doi: 10.1080/14616680601092857.

Sparks, B.A. and Browning, V. (2011), "The impact of online reviews on hotel booking intentions and perception of trust", *Tourism Management*, Vol. 32 No. 6, pp. 1310-1323, doi: 10.1016/j. tourman.2010.12.011.

Storper, M. (1997), *The Regional World: territorial Development in a Global Economy*, Guilford press.

Streimikiene, D., Svagzdiene, B., Jasinskas, E. and Simanavicius, A. (2021), "Sustainable tourism development and competitiveness: the systematic literature review", *Sustainable Development*, Vol. 29 No. 1, pp. 259-271, doi: 10.1002/sd.2133.

Stylidis, D. (2018), "Residents' place image: a cluster analysis and its links to place attachment and support for tourism", *Journal of Sustainable Tourism*, Vol. 26 No. 6, pp. 1007-1026, doi: 10.1080/ 09669582.2018.1435668.

Tanner, A.N. (2016), "The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases", *Journal of Economic Geography*, Vol. 16 No. 3, pp. 611-635, doi: 10.1093/jeg/lbv011.

Ten Bosch, O., Windmeijer, D., van Delden, A. and van den Heuvel, G. (2018), "Web scraping meets survey design: combining forces", *Big Data Meets Survey Science Conference*, Barcelona.

Timur, S. and Getz, D. (2008), "A network perspective on managing stakeholders for sustainable urban tourism", *International Journal of Contemporary Hospitality Management*, Vol. 20 No. 4, pp. 445-461, doi: 10.1108/09596110810873543.

Weidenfeld, A., Björk, P. and Williams, A.M. (2016), "Cognitive and cultural proximity between service managers and customers in cross-border regions: knowledge transfer implications", *Scandinavian Journal of Hospitality and Tourism*, Vol. 16 No. sup1, pp. 66-86, doi: 10.1080/ 15022250.2016.1244587.

Więckowski, M. (2021), "Will the consequences of covid-19 trigger a redefining of the role of transport in the development of sustainable tourism?", *Sustainability*, Vol. 13 No. 4, p. 1887, doi: 10.3390/ su13041887.

World Travel and Tourism Council (2020), "Economic Impact of Travel and Tourism Report", available at: https://wttc.org/About/About-Us

Zhang, J. (2019), "What's yours is mine: exploring customer voice on airbnb using text-mining approaches", *Journal of Consumer Marketing*, Vol. 36 No. 5, pp. 655-665, doi: 10.1108/JCM-02- 2018-2581.

Zhang, Y., Jin, R. and Zhou, Z.-H. (2010), "Understanding bag-of-words model: a statistical framework", *International Journal of Machine Learning and Cybernetics*, Vol. 1 Nos 1/4, pp. 43-52, doi: 10.1007/s13042-010-0001-0.

Appendix 1
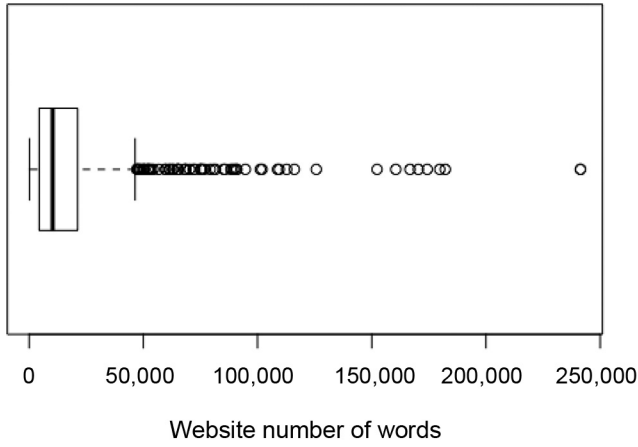
Website number of words

Appendix 2

Basket 1 – Terms related to sustainability in tourism

armonia con l'ambiente	environmental harmony
biodiversità*	biodiversity*
cicloturismo*	cycle tourism*
conservazione della natura	nature conservation
Eco Trekking	Eco Trekking
eco-camping*	eco-camping*
ecocompatibile*	eco-friendly*
eco-compatibilità	eco-compatibility
ecofriendly	ecofriendly
eco-friendly	eco-friendly
ecologia	ecology
ecologica	ecological
ecologico	ecological
ecomuseo*	Eco-museum *
ecosistema*	ecosystem*
ecoturismo	ecotourism
ecovegan	Eco-vegan
eco-vegan	eco-vegan
equilibrio naturale	natural balance
greenfriendly*	Green-friendly *
impatto ambientale	environmental impact
impronta ecologica	ecological footprint
ingredienti biologici	organic ingredients
km0	km0
km 0	km 0
mobilità sostenibile	sustainable mobility
plasticless*	plasticless*
riduzione degli sprechi*	waste reduction*
rispetto dell'ambiente	respect for the environment
salvaguardia ambientale	environmental protection
sostenibile	sustainable
sostenibilità ambientale	environmental sustainability
sostenibilità sociale	social sustainability
turismo consapevole	conscious tourism
turismo ecologico	ecological tourism
turismo ecosostenibile	eco-sustainable tourism
turismo eco-sostenibile	eco-sustainable tourism
turismo etico*	ethical tourism*
turismo green	green tourism
turismo lento*	slow tourism*
turismo responsabile	responsible tourism
turismo rispettoso*	respectful tourism*
turismo sostenibile	sustainable tourism
turismo verde	green tourism

Basket 2 – Terms related to sustainability in tourism facilities buildings

alloggi a basso impatto	low impact housing
architettura bioecologica	bioecological architecture
architettura sostenibile	sustainable architecture
bioarchitettura	bioarchitecture

Table A1.
Baskets of words used in the Web scraping process (English translation in the gray-shaded column)

(*continued*)

bio-architettura	bio-architecture
co-generazione	co-generation
compensazione co2	CO_2 compensation
compostaggio*	composting*
consumo dell' acqua	water consumption
eco alloggi	eco accommodations
Eco-bnb	ecobnb
eco-building	eco-building
ecocompatibile*	eco-friendly*
eco-costruzione	eco-construction
eco-cottage	eco-cottage
Eco-hotel	ecohotel
eco-hotel	eco-hotel
efficienza energetica	energy efficiency
emissioni zero	zero emissions
emissioni*	emissions*
energie rinnovabili*	renewable energies*
eolico*	wind power*
fotovoltaico	photovoltaic
geotermia	geothermal
geotermiche*	geothermal*
green building	green building
hotel green	green hotel
hotel sostenibile	sustainable hotel
impianto eolico*	wind power plant*
impianto geotermico	geothermal system
materiali rigenerabili	regenerable materials
Nearly zero energy building	Nearly zero energy building
Nzeb	Nzeb
pannelli fotovoltaici	photovoltaic panels
pannelli solari	solar panels
raccolta differenziata	waste sorting
riciclo	recycling
riduzione dei consumi	reduction of consumption
riduzione dei rifiuti	waste reduction
riduzione delle emissioni*	emission reduction*
rigenerazione urbana*	urban regeneration*
risparmio dell' acqua	water saving
sistemi passivi per la climatizzazione	passive systems for air conditioning
zero emissioni	zero emissions
Basket 3 – Certifications	
AENOR	AENOR
AENOR Medio Ambiente	AENOR Medium Environment
Audubon International	Audubon International
B Corporation	B Corporation
Bcorp	BCorp
Beluga	Beluga
BIO Hotels	BIO Hotels
Biosphere Responsible Tourism	Biosphere Responsible Tourism
Blue Angel	Blue Angel
Blue Flag	Blue Flag
Carbon Neutral Certification	Carbon Neutral Certification
Carbon Reduction Label	Carbon Reduction Label

(*continued*)

Table A1.

*Carta Europea del Turismo Sostenibile**	European Charter for Sustainable Tourism *
*casaclima**	climate house*
certificazione di turismo sostenibile	sustainable tourism certification
*CETS**	CETS*
*climahotel**	climahotel*
Climatop	Climatop
*demeter**	demeter*
EarthCheck	EarthCheck
Earthsure	Earthsure
ECO certification	ECO certification
*ecolabel**	ecolabel*
*EcoWorldHotel**	EcoWorldHotel *
*emas**	emas*
*ESG**	ESG*
EU Ecolabel	EU Ecolabel
European Ecolabel	European Ecolabel
Fair for Life	Fair for Life
*fsc**	fsc*
Global Sustainable Tourism Council	Global Sustainable Tourism Council
Green Globe	Green Globe
Green Key	Green Key
Green Seal	Green Seal
Green Tick	Green Tick
Green tourism	Green tourism
Green Tourism Business Scheme	Green Tourism Business Scheme
GSTC	GSTC
GSTC C-HTO	GSTC C-HTO
GTBS	GTBS
Ibex Fairstaty	Ibex Fairstaty
International Eco Certification Program	International Eco Certification Program
legambiente turismo	legambiente tourism
Nature's Best Ecotourism	Nature's Best Ecotourism
Sustainable Tourism Education Program	Sustainable Tourism Education Program
Travelife	Travelife
*treedom**	treedom*
Viabono	Viabono

Table A1.   **Note:** *Terms added after word embedding analysis performed with *Word2Vec*

Appendix 3

Independent variables	Dependent variable Basket 2 (1)	Basket 3 (2)
Basket 1	0.082** (0.037)	
Basket 2		0.467*** (0.040)
Constant	1.390*** (0.146)	1.293*** (0.083)
Observations	283	106
R^2	0.017	0.573
Adjusted R^2	0.014	0.569
Residual Error Std.	1.420 (df = 281)	0.671 (df = 104)
F-statistic	4.954** (df = 1; 281)	139.349*** (df = 1; 104)

Notes: **$p < 0.05$; ***$p < 0.01$

Table A2.
OLS models,
correlation between
baskets

Corresponding author
Silvia Rita Sedita can be contacted at: silvia.sedita@unipd.it