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.

**Keywords** Social network analysis, Sustainable tourism, Web data mining, Text analysis, Machine learning

**Paper type** Research paper

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; Kithamkesorn 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
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

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

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

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

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 oficialize 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)](image-url)
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
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.
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

<table>
<thead>
<tr>
<th>Network properties</th>
<th>Basket 1</th>
<th>Basket 2</th>
<th>Basket 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>706</td>
<td>299</td>
<td>54</td>
</tr>
<tr>
<td>Number of links</td>
<td>70,223</td>
<td>12,329</td>
<td>320</td>
</tr>
<tr>
<td>Modularity</td>
<td>0.31</td>
<td>0.25</td>
<td>0.53</td>
</tr>
<tr>
<td>Density</td>
<td>0.28</td>
<td>0.28</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 3. Network indicators
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
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 over-tourism (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


IJCHM


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


Figure A1. Boxplot of the distribution of words in the tourism facilities websites.
## Appendix 2

### Basket 1 – Terms related to sustainability in tourism

<table>
<thead>
<tr>
<th>Italian Term</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>armonia con l’ambiente</td>
<td>environmental harmony</td>
</tr>
<tr>
<td>biodiversità*</td>
<td>biodiversity*</td>
</tr>
<tr>
<td>cicloturismo*</td>
<td>cycle tourism*</td>
</tr>
<tr>
<td>conservazione della natura</td>
<td>nature conservation</td>
</tr>
<tr>
<td>Eco Trekking</td>
<td>Eco Trekking</td>
</tr>
<tr>
<td>eco-camping*</td>
<td>eco-camping*</td>
</tr>
<tr>
<td>eco-compatibile*</td>
<td>eco-compatibility</td>
</tr>
<tr>
<td>eco-compatibilità</td>
<td>eco-compatibility</td>
</tr>
<tr>
<td>eco-friendly</td>
<td>eco-friendly</td>
</tr>
<tr>
<td>ecologia</td>
<td>ecology</td>
</tr>
<tr>
<td>ecologica</td>
<td>ecological</td>
</tr>
<tr>
<td>ecologico</td>
<td>ecological</td>
</tr>
<tr>
<td>ecomuseo*</td>
<td>Eco-museum</td>
</tr>
<tr>
<td>ecosistema*</td>
<td>ecosystem*</td>
</tr>
<tr>
<td>ecoturismo</td>
<td>ecotourism</td>
</tr>
<tr>
<td>ecovegan</td>
<td>Eco-vegan</td>
</tr>
<tr>
<td>eco-vegan</td>
<td>eco-vegan</td>
</tr>
<tr>
<td>equilibrio naturale</td>
<td>natural balance</td>
</tr>
<tr>
<td>greenfriendly*</td>
<td>Green-friendly*</td>
</tr>
<tr>
<td>impatto ambientale</td>
<td>environmental impact</td>
</tr>
<tr>
<td>impronta ecologica</td>
<td>ecological footprint</td>
</tr>
<tr>
<td>ingrediente biologici</td>
<td>organic ingredients</td>
</tr>
<tr>
<td>km0</td>
<td>km 0</td>
</tr>
<tr>
<td>mobilità sostenibile</td>
<td>sustainable mobility</td>
</tr>
<tr>
<td>plasticless*</td>
<td>plasticless*</td>
</tr>
<tr>
<td>riduzione degli sprechi*</td>
<td>waste reduction*</td>
</tr>
<tr>
<td>rispetto dell’ambiente</td>
<td>respect for the environment</td>
</tr>
<tr>
<td>salvaguardia ambientale sostenibile</td>
<td>environmental protection</td>
</tr>
<tr>
<td>sostenibilità ambientale</td>
<td>sustainable</td>
</tr>
<tr>
<td>sostenibilità sociale</td>
<td>environmental sustainability</td>
</tr>
<tr>
<td>turismo consapevole</td>
<td>social sustainability</td>
</tr>
<tr>
<td>turismo ecologico</td>
<td>ecological tourism</td>
</tr>
<tr>
<td>turismo ecosostenibile</td>
<td>eco-sustainable tourism</td>
</tr>
<tr>
<td>turismo eco-sostenibile</td>
<td>eco-sustainable tourism</td>
</tr>
<tr>
<td>turismo etico*</td>
<td>ethical tourism*</td>
</tr>
<tr>
<td>turismo green</td>
<td>green tourism</td>
</tr>
<tr>
<td>turismo lenso</td>
<td>slow tourism*</td>
</tr>
<tr>
<td>turismo responsabile</td>
<td>responsible tourism</td>
</tr>
<tr>
<td>turismo rispettoso*</td>
<td>respectful tourism</td>
</tr>
<tr>
<td>turismo sostenibile</td>
<td>sustainable tourism</td>
</tr>
<tr>
<td>turismo verde</td>
<td>green tourism</td>
</tr>
</tbody>
</table>

### Basket 2 – Terms related to sustainability in tourism facilities buildings

<table>
<thead>
<tr>
<th>Italian Term</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>alloggi a basso impatto</td>
<td>low impact housing</td>
</tr>
<tr>
<td>architettura bioecologica</td>
<td>bioecological architecture</td>
</tr>
<tr>
<td>architettura sostenibile</td>
<td>sustainable architecture</td>
</tr>
<tr>
<td>bioarchitettura</td>
<td>bioarchitecture</td>
</tr>
</tbody>
</table>

---

(continued)
bio-architettura  
co-generazione  
compensazione co2  
compostaggio*  
consumo dell’acqua  
eco alloggi  
Eco-bnb  
eco-building  
ecocompatibile*  
eco-costruzione  
eco-cottage  
Eco-hotel  
eco-hotel  
efficienza energetica  
emissioni zero  
emissioni*  
energie rinnovabili*  
eolico*  
fotovoltaico  
geotermia  
geotermiche*  
green building  
hotel green  
hotel sostenibile  
impianto eolico*  
impianto geotermico  
materiali rigenerabili  
Nearly zero energy building  
Nzeb  
pannelli fotovoltaici  
pannelli solari  
raccolta differenziata  
ricido  
r riduzione dei consumi  
r riduzione dei rifiuti  
r riduzione delle emissioni*  
rigenerazione urbana*  
risparmio dell’acqua  
sistemi passivi per la climatizzazione  
zero emissioni  

Basket 3 – Certifications  
AENOR  
AENOR Medio Ambiente  
Audubon International  
B Corporation  
Bcorp  
Beluga  
BIO Hotels  
Biosphere Responsible Tourism  
Blue Angel  
Blue Flag  
Carbon Neutral Certification  
Carbon Reduction Label  

bio-architecture  
co-generation  
CO2 compensation  
composting*  
water consumption  
eco accommodations  
ecobnb  
eco-building  
eco-friendly*  
eco-construction  
eco-cottage  
ecohotel  
eco-hotel  
energy efficiency  
zero emissions  
emissions*  
renewable energies*  
wind power*  
photovoltaic  
geothermal  
geothermal*  
green building  
green hotel  
sustainable hotel  
wind power plant*  
geothermal system  
regenerative materials  
Nearly zero energy building  
Nzeb  
photovoltaic panels  
solar panels  
waste sorting  
recycling  
reduction of consumption  
waste reduction  
emission reduction*  
urban regeneration*  
water saving  
passive systems for air conditioning  
zero emissions  

| Table A1. Sustainable tourism organizations | AENOR  
AENOR Medium Environment  
Audubon International  
B Corporation  
BCorp  
Beluga  
BIO Hotels  
Biosphere Responsible Tourism  
Blue Angel  
Blue Flag  
Carbon Neutral Certification  
Carbon Reduction Label  
(continued) |
| European Charter for Sustainable Tourism | climate house |
| sustainable tourism certification | CETS |
| climahotel | Climatop |
| demeter | EarthCheck |
| Earthsure | ECO certification |
| ecolabel | EcoWorldHotel |
| ESG | Emas |
| EU Ecolabel | Ecolabel |
| European Ecolabel | EuroEcolabel |
| Fair for Life | Fair for Life |
| fsc | 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
Table A2. OLS models, correlation between baskets

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Basket 2 (1)</th>
<th>Basket 3 (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basket 1</td>
<td>0.082** (0.037)</td>
<td></td>
</tr>
<tr>
<td>Basket 2</td>
<td>0.467*** (0.040)</td>
<td>0.146 (0.146)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.390*** (0.146)</td>
<td>1.293*** (0.083)</td>
</tr>
<tr>
<td>Observations</td>
<td>283</td>
<td>106</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.017</td>
<td>0.573</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.014</td>
<td>0.569</td>
</tr>
<tr>
<td>Residual Error Std.</td>
<td>1.420 (df = 281)</td>
<td>0.671 (df = 104)</td>
</tr>
<tr>
<td>$F$-statistic</td>
<td>4.954** (df = 1; 281)</td>
<td>139.349*** (df = 1; 104)</td>
</tr>
</tbody>
</table>

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

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