Quality & Quantity
https://doi.org/10.1007/s11135-025-02105-x

Biclustering *sustainable* local tourism systems by the Tabu search optimization algorithm

Wassim Ayadi[1,2] · Joseph Andria[3] · Giacomo di Tollo[4] · Gerarda Fattoruso[5]

Accepted: 18 February 2025

Abstract
Tourism is nowadays fully acknowledged as a leading industry contributing to boost the economic development of a country. This growing recognition has led researchers and policy makers to increasingly focus their attention on all those concerns related to *optimally* detecting, promoting and supporting territorial areas with a high tourist vocation, i.e., *Local Tourism Systems*. In this work, we propose to apply the biclustering data mining technique to detect Local Tourism Systems. By means of a two-dimensional clustering approach, we pursue the objective of obtaining more in-depth and granular information than conventional clustering algorithms. To this end, we formulate the objective as an optimization problem, and we solve it by means of Tabu-search. The obtained results are very promising and outperform those provided by classic clustering approaches.

Keywords Biclustering · Local tourism systems · Tabu-search · Optimization problem

1 Introduction

Tourism has now become one of the main engines of development in the world and one of the major industries influencing both national economies and natural environment. Tourism is a source of economic development for several countries, and its relevant impact on a countries' economic growth has been worldwide acknowledged. Tourism industry involves many economic sectors, such as, among others, transport, financial services, infrastructure, food, health, information and technology. With reference to the economic impact, tourism boost destinations competitiveness by promoting all those kind of investments needed for enriching and improving overall tourist experience. The ongoing increasing importance of

✉ Giacomo di Tollo
g.ditollo@univpm.it

1 SMART Lab, University of Tunis, ISG Tunis, Tunis, Tunisia

2 FSEGT, University of Tunis El Manar, Tunis, Tunisia

3 Department of Economics, Business and Statistics, University of Palermo, 90100 Palermo, Italy

4 Department of Management, Polytechnic University of Marche, Piazza Roma, 22, 60121 Ancona, Italy

5 Department of Economics, Management and Territory, University of Foggia, 71121 Foggia, Italy

Published online: 02 April 2025

the tourism industry along with its multidimensionality and complexity have attracted in recent years great attention from both academics and policymakers (Gormsen 1997). A great concern in public policy has been given to the themes pertaining to local government systems in order to strengthen territorial cooperation and governance (Ababkova and Vasileva 2020). With this regard, particular attention has been paid to the definition, identification and modeling of Local Tourism Systems (LTS) in order to promote tourism development policies and assess governance strategies that support comprehensively systems growth and tourism demand. LTS represent *clusters* of homogeneous or integrated territories with a tourist vocation which are characterized by an integrated offer of cultural and environmental heritage, including also either typical agricultural or local handicraft products, and where, individual or partnership of, tourist oriented firms are observed (Glover 1986).

The interest in *clustering* approaches stems from the acknowledgment of their positive influence on companies' performance, countries' competitiveness and regional development (Rocha 2004). As also stated by Porter (2000), clusters facilitate innovation by contributing to the development of innovative processes, promote and strengthen the networking with other institutions, they are also able to better intercept consumers' demand and interests and boost technology development by inducing the synthesis of knowledge and information needs of all stakeholders.

Some examples of real-experienced clustering strategies in tourism are, just to cite a few examples, the Heritage Tourism Districts (Moebius 2022) (HTSs), the Tourism Improvement Districts (TIDs) (Tourism Economics 2021), the Adventure Tourism Districts (ATDs) (Tennessee Department 2024), the Urban Tourism Districts (UTDs) (Howard 2010) and the Tourism Marketing Districts (TMDs) (Testa and Sipe 2021). They all aim to foster destination development through policy orientations, marketing programs and attracting investments. The public sector, through the implementation of effective laws and policies, plays a crucial role in supporting their growth and promoting their sustainable development. European and national legislative authorities have defined and regulated the role of LTS by recognizing these systems as part of a wider administrative, institutional and political context. In Italy, for example, the Legislative Decree No. 79-2011 has defined the role of LTS by allowing Regions to accomplish the task of recognizing these systems as part of their administrative structure. Identifying clusters is relevant not only for regulatory purposes. Indeed, the objective of clustering a given geographical area with a tourist vocation is a complex multi-criteria decision problem, which involves many factors at different levels (Andria et al. 2021). The study by Andria et al. (2015) is among the recent research focusing on this topic.

Moreover, clusters often arise spontaneously and the task of their creation cannot be addressed by introducing a list of simple constraints and rules. Industry clusters are intrinsically dynamics, with boundaries and activities constantly changing as the whole network in which they operate evolve over time. Clusters generally start from a bottom-up "spontaneity" process rather than a top-down policy approach.

In our line of work, following the work in Andria et al. (2015), we detect the LTS in Sicily by applying the biclustering technique, with the aim of revising and improving the therein obtained clusters. In the literature, several studies have already demonstrated the effectiveness of using biclustering for different aspects related to the tourism sector. For example, Dolnicar (2013) demonstrates that biclustering is a tool to overcome data dimensionality problems in tourism services. Belayadi et al. (2022) presents a scheme based on Apriori biclustering algorithms to extract association rules from LBSN data that can be used for the tourism recommendation system. Finally, Dolnicar et al. (2012) introduce biclustering, to address the problem of high dimensionality in tourism segmentation studies by discussing. the cases in which biclustering should be used rather than parametric or non-parametric clustering techniques.

Clustering has reasonable results in LTS, however it has some limits: for instance, there is no overlap between clusters, and the elements of a cluster must be homogeneous on all characteristics, which is not the case with LTS.

Biclustering is a data mining technique used for discovering patterns in data sets where the rows and columns are organized into coherent clusters. Unlike traditional clustering, which focuses on grouping objects based on similarity, biclustering aims to identify sub-matrices within a data matrix where the elements are similar to each other both across rows and columns.

In other words, biclustering is a method for finding dense regions in a data matrix, where the elements in the region have similar (i.e., coherent) values across the rows and columns. Although the application of biclustering techniques can be found in problems such as Local Multiple Sequence Alignment (Ayadi 2018) and recommendation systems (Sun and Zhang 2022), the term was first coined in the microarray gene expression analysis context (Houari 2018; Maâtouk et al. 2021; Maâtouk etal. 2022). In this case, the goal was that of identifying groups of genes that showed similar activity patterns in a subset of experimental conditions. In fact, unlike classical one-way clustering, biclustering can reveal crucial information for the following reasons: i) only a small set of genes may be involved in some cellular process of interests; ii) the cellular process of interest may be active only in a subset of the conditions; iii) a single gene may participate in multiple processes that may, or not, be co-active under all conditions.

With respect to our objective of study, we may rephrase by saying that: i) only a small set of units may be involved in some processes of attractiveness ii) these processes may be active only in a subset of the features (attributes of a tourist destination) iii) a single unit may participate to more than one process that may, or not, be co-active under all the considered attributes.

Biclustering algorithms typically use optimization techniques to find the sub-matrices with the highest level of similarity, and they can be used to discover both static and dynamic patterns in data. Applied to our case study, bi-cluster would improve the quality of LTS clusters by using the combination between rows and columns, finding a subset of columns to classify territories as a subset of features:

This would provide more useful granular insights on the relationship between the typicality and attractiveness of a LTS in terms of an *optimal* policy-making framework that looks at the commonalities and peculiarities across different territorial areas.

The biclustering optimization problem is solved by means of a two-stage Tabu Search metaheuristic algorithm which uses a tabu list to store recently evaluated solutions so to avoid reconsidering them for a given number of iterations.

The rest of the paper is organized as follows: in Sect. 2 the problem formulation is described, in Sect. 3 the biclustering method is presented, while, in Sect. 4 our proposed biclustering algorithm is introduced. The Sects. 5 and 6, respectively, describe and comment on the obtained results and conclude the paper.

2 Problem formulation

A geographical area is made up of territorial units, the municipalities. Such a zone can be represented as a complete non oriented graph whose vertices are the territorial units (Andria et al. 2015). We aim at identifying biclusters of vertices forming tourism systems. The graph is denoted by:

$$G = (V, E)$$

where:

- $V = \{1, ..., n_U\}$ is the set of vertices corresponding to the n_u territorial units;
- $E = \{1, ..., n_u\}$ is the set of edges.

To each vertex $i \in V$ is associated a value β representing the attractiveness of the territorial unit. To each edge $k \in E$ connecting vertices $i - j$ is associated a value e_{ij} corresponding to the inverse distance between i and j.

In our work we aim to identify biclusters of vertices forming tourism systems. To reach this goal, we have to introduce the concept of attractiveness. The attractiveness of a geographical entity can be defined as its ability to improve the tourist's well-being. It refers to all those combined attributes of an area which are positively perceived or that draw in potential tourist visitors. We translate this concept into a measure for the aggregate attractiveness as perceived by tourists. A geographical area can offer a number of features such as hotels, outdoor accommodations and bed and breakfast, restaurants, night clubs, travel agencies, tourist guides, cultural life, transportation infrastructure etc. Each feature is assessed by the quantity of homogeneous elements offered by all operators working in a given territorial unit. The potential of a given unit to attract tourism is a function of these features. Features taken into account in our work are detailed in Table 1 where each bicluster is characterized by a specific number of features.

In Andria et al. (2015), authors solved the tourism cluster problem with respect to both the *Unconstrained* and *Constrained* formulations, where the imposed constraints, which comply with the guidelines provided by the Sicilian Regional Assembly Actuative Decree No. 4-2010 (Regione 2010), are:

$$\sum_{i \in Q} F_{2i} \geq 150,000 \tag{1}$$

$$\sum_{i \in Q} F_{5i} \geq 7,500 \tag{2}$$

Table 1 Overall features statistics. Data refer to the whole region Sicily, over the period 1998-2002

Feature	Mean	Median	SD	Min	Max
Municipalities area (km2)	66.01	37.00	79.82	1.14	551.12
Population	12,914.00	4,713.00	40,537.98	229.00	663,173.00
Cultural heritage goods	0.98	0.00	3.50	0.00	56.00
Transport	41.98	15.00	139.40	0.00	1,906.00
Beds in Hotels	461.15	15.00	1,215.74	0.00	10,603.00
Beds in Hospitals	48.78	0.00	312.29	0.00	4,284.00
Sport, Cultural activity	47.19	17.00	165.14	0.00	2,480.00
Financial intermediaries	25.85	6.00	107.27	0.00	1,610.00
Commercial Businesses	435.88	135.00	1,603.42	5.00	24,664.00
Distance from airport (km)	81.69	75.90	44.45	0.00	203.00

$$\sum_{i \in Q} F_{2i} \le 350 \sum_{i \in Q} F_{9_i} \tag{3}$$

In (1–3), Q is a feasible subset of V, i.e., $Q \subseteq V$, and F_{2i}, F_{5i}, F_{9i} are, respectively, the number of residents, the number of beds places and the number of commercial activities related to the i^{th} territorial unit.

Before introducing our proposed biclustering algorithm, we first briefly give a description of the biclustering problem.

3 The biclustering problem

Unlike clustering which gives a global view of knowledge hidden in the data by creating disjoint clusters, Biclustering gives a local view by creating overlapping biclusters covering generally a portion of the input dataset. This overlap is planned when every bicluster represents a good LTS when the lines represent the municipalities and columns are their features. Thus, the extraction of these coherent subsets represents a very important task in the analysis of LTS.

The biclustering aims to combine simultaneously the rows and the columns of a matrix to obtain consistent, homogeneous and stable biclusters. A bicluster is a subgroup of lines that exhibit a common behavior under a subgroup of features. Each line or each feature can participate in one or more biclusters. Formally, the data matrix is represented as $M(I, J)$, where I is a set of lines or municipalities and J a set of features, and the cell m_{ij} represents the relative level that the j^{th} feature is observed within the i^{th} municipality. A bicluster $B(G, C)$ associated with a data matrix $M(I, J)$ is a sub-matrix such that $G \subseteq I$ and $C \subseteq J$. The biclustering problem aims to extract biclusters of maximal size that satisfy coherence constraints.

The biclustering problem is a highly combinatorial problem with a search space size $O(2^{|I|+|J|})$ (Freitas etal. 2013).

Moreover, in the general case, the biclustering problem is NP-hard (Cheng and Church 2000), which explains why the majority of existing biclustering methods do not guarantee the optimality of their solutions.

Biclustering algorithms can be grouped in two classes (Freitas etal. 2013): constructive algorithms and metaheuristics. In the first class, we can cite as example the BiMax (Prelic etal. 2006) which starts with the entire data matrix as the initial bicluster. Then, divided iteratively this bicluster into several biclusters satisfying certain features until the verification of termination criterion. This approach is characterized by its swiftness, but it may exclude good biclusters by performing the division before evaluating them. Also, we can cite the Greedy Iterative Search Approaches OPSM (Ben-Dor et al. 2003) and ISA (Bergmann et al. 2003). While, metaheuristics include the local search method called CC (Cheng and Church 2000) and the evolutionary algorithms (EA) SEBI (Divina and Aguilar-Ruiz 2006) and AMoDeBic (Charfaoui et al. 2024). Hybrids are also proposed and we can cite HMOBI (Seridi et al. 2015), EBA (Maâtouk et al. 2019) and BOBEA (Maâtouk etal. 2022).

For more details, we can cite some surveys (Noronha et al. 2022; Castanho etal. 2022; Garcia et al. 2023; Balamurugan and Raja 2022; Castanho etal. 2024).

In the next section, we introduce our proposed biclustering algorithm based on metaheuristic optimization for detecting LTS.

4 The proposed method

To solve the formulation problem we choose to employ metaheuristics. The use of metaheuristics appeared as a way out of combinatorial problems whose search spaces are exploding in size and for NP-hard problems (Garcia et al. 2023). Our problem is among the problems that can be solved by the neighborhood search method. We adopt a Tabu Search (TS) method which is a particular neighborhood search algorithm introduced by Fred Glover in 1986 (Glover 1986). The principle of TS is as follows: at each iteration, the neighborhood (complete or subset of neighborhood) of the current solution is examined and the best solution is selected if it does not belong to the tabu list. In this section, we represent a new biclustering algorithm called Tabu Search Biclustering (TSB) to resolve tourism problem faced by Italian Sicily Region.

The aim of our algorithm is to improve the attractiveness of the intial solution. Indeed, the attractiveness can be defined as its ability to contribute to tourist well-being (see Eq. 6).

The remainder of this section will outline the main components of our experimental approach: Sect. 4.1 will discuss data to be given as input to our procedure; Sect. 4.2 will discuss about the initial solution; Sect. 4.3 will outline the objective function used, and Sect. 4.4 will be devoted to our search algorithm.

4.1 Data input

Our algorithm requires two main inputs for the biclustering process. First, it uses the initial LTS clusters generated by the *TA-clusters* algorithm applied to Sicily's 390 municipalities, as previously defined in Andria et al. (2015). These clusters are essential as starting points for detecting coherent tourism biclusters. The dataset contains several features describing each municipality, such as the number of hotels, restaurants, airports, museums, and the distances between different territorial units. The choice of these features is crucial to reflect the multi-dimensional attractiveness of each municipality, capturing both infrastructure and tourism-related services.

4.2 Initial solution

The initial solution of our proposed algorithm is derived directly from the clusters formed by the TA-clusters algorithm (Andria et al. 2015). These clusters comprise subsets of municipalities characterized by all available features. However, not all features are equally relevant to a tourism-based analysis. Consequently, our method uses a feature reduction step by identifying and eliminating trivial attributes, i.e. those whose values are predominantly zero across all municipalities in a cluster. In addition, a pairwise correlation analysis is performed between all features within each cluster. Highly correlated features, which may represent redundant information, are identified using a Pearson correlation coefficient threshold (greater than 0.9). Instead of mechanically eliminating one feature from each highly correlated pair, we analyze groups of redundant features to determine which provides the most information or is most relevant in the context of the analysis. This approach preserves data quality by retaining the most useful features and avoiding the loss of potentially important information. These steps ensure that only relevant characteristics are retained, enabling the creation of more focused and meaningful biclusters. The resulting biclusters are made up of subsets of municipalities linked by consistent patterns of tourism-related characteristics. This refined solution forms the basis for further optimization using our TSB method described in the following subsections.

4.3 Objective function

Attractiveness geographical zone was proposed by Andria et al. (2015) to measure tourist well-being. The Eq. 6 describes this measure.

Let P_i be the tourist presence (i.e., the quantity of nights spent by tourists in accommodation facilities belonging to territorial unit i) and $p_i = P_i / \sum_{j=1}^{N_U} P_j$ indicate the relative tourist presence of a geographical entity i w.r.t. the whole region taken into account. To measure the potential of a feature f in a territorial unit i we introduce S_f as the value of feature f in a given territorial unit i. The ratio $S_{fi} / \sum_{j=1}^{N_U} S_{fj}$ represents the relative frequency of feature f in the territorial unit i. Let n_F be the number of features considered. The attractiveness a_i of the territorial unit i is measured by

$\beta_i = \sum_{f=1}^{N_F} \frac{S_{fi}}{p_i \sum_{j=1}^{N_U} S_{fj}} \; i = 1, \ldots, n_U.$

To achieve our objective of clustering we also take into consideration a sustainability-oriented criterion by introducing two more conditions, i.e., the *Environmental* sustainability, $s^{(E)}$, and the *Social* (S) sustainability, $s^{(S)}$, as defined by:

$$s_i^{(E)} = \frac{A_i}{p_i \sum_{j=1}^{n_U} A_j}. \tag{4}$$

$$s_i^{(S)} = \frac{P_i^r}{p_i \sum_{j=1}^{n_U} P_j^r}. \tag{5}$$

where A_i is the i-th municipality's surface area (sq.m.) and P_i^r is the i-th municipality's residents amount.

By aggregating territorial units in clusters, we want to create aggregates whose attractiveness is bigger than the sum of attractiveness of its components. Finally, the attractiveness of bicluster b can be computed as defined in

$$ATC(b) = \sum_{i \in V_b} \left(\beta_i + s_i^{(E)} + s_i^{(S)} \right) + \sum_{c \in E_b} (\beta_i + \beta_j) e_{ij} \tag{6}$$

where V_b and E_b are respectively the sets of vertices and edges forming bicluster (local system) b, with $i - j$ being endpoints of edge e.

In Andria et al. (2015), an approximate solution of (6) is given, although without taking into account the aforementioned sustainable objectives. In this study, we aim to improve the quality and accuracy of the obtained results by either i) applying sustainability criteria or ii) adopting a biclustering approach.

4.4 The search procedure

First of all, after obtaining the clusters generated by the TA-Clusters algorithm (Andria et al. 2015; Andria and di Tollo 2015), our method first preprocesses them to retain only the most relevant features within each cluster.

Afterwards, we adopt the Tabu Search method to improve the quality of biclusters. To do that, several steps are necessary:

- We build a list called *T* that contains municipalities whose distance between them does not exceed 80 km. A second list called *Mov* that contains two types of movement: the first move consist to add a municipality to the bicluster. The second move consist to replace one municipality with a low attractiveness from the bicluster with another one how has a high attractiveness. All movement have to respect the 80 km constraint.
- After creating *T* and *Mov*, we choose randomly a move from *Mov*. If this movement does not exist in Tabu List *TL*, we will apply it to our current bicluster *Bic* to obtain *Bic′* (a neighbor of *Bic*) after inserting it in the *TL*. If not, we choose another move. The *TL* list is based on the First-In First-Out (FIFO) principle.
- After that, we compute the attractiveness of *Bic′*. If ATC(*Bic′*) > ATC(*Bic*), then we check the cited conditions imposed by the Italian legislator ("*the Sicilian Regional Accusative Decree n.4-2010*" (Regione 2010)) to recognize a geographical entity as being a local tourism system.
- Our method iterates these steps until reaching a maximum number of iterations, or when the current bicluster is no longer improved for *y* iterations.

The flowchart representation of the proposed TSB algorithm is given in Fig. 1. In the proposed two-stage metaheuristic biclustering algorithm, the first stage is to discard all those features for a cluster which account for a very low percentage. The second stage, starting from the initial solution found at the previous stage, uses the Tabu Search algorithm to explore local neighborhoods of temporarily accepted solutions. Tabu Search is used to prevent the biclustering algorithm from re-evaluating solutions (biclusters) already processed and targeted as unpromising.

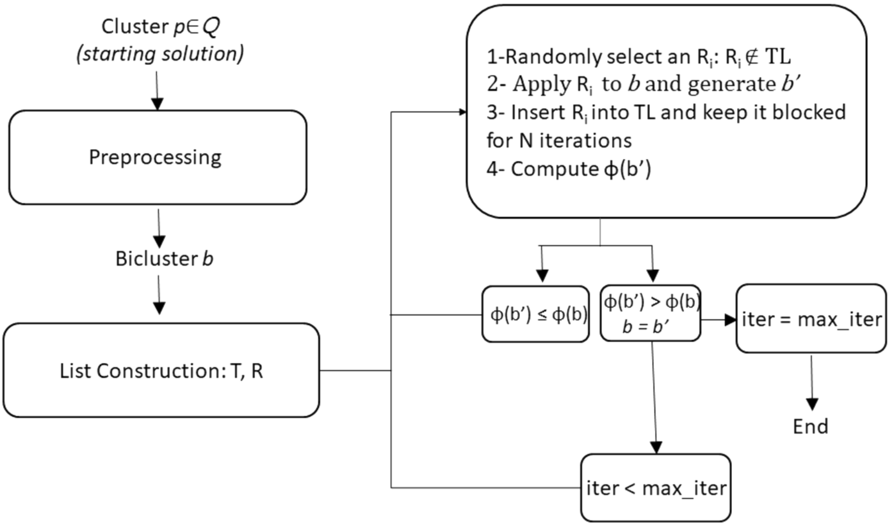


Fig. 1 Flowchart representation of the two-stage metaheuristc Tabu Search Biclustering (TSB) Algorithm. On the left panel, all-relevant feature selection procedure is addressed. On the right one, the Tabu Search algorithm is executed to explore local neighborhoods of previous stage passed solutions

5 Evaluation results

The performance of the proposed algorithm TSB is evaluated in this section, with a focus on comparing our results to those obtained by Andria et al. (2015). The primary metric for comparison is the attractiveness of each bicluster, defined as the capacity of a geographical entity to contribute to tourist well-being. Additionally, we compare the number of territories in both clusters and biclusters, providing a comprehensive analysis of the effectiveness of our approach.

Table 2 illustrates the key findings from this evaluation, highlighting the advantages of biclustering over traditional clustering methodologies. Furthermore, Fig. 2 shows the profile plot of Biclusters reported in Table 2. The result obtained through biclustering is a set of biclusters, each composed of territories within Sicily, for which we have calculated their attractiveness - our objective function. The attractiveness metric reflects how well the territories within each cluster or bicluster meet the predefined criteria. Notably, the attractiveness of all clusters is improved when applying biclustering. This improvement demonstrates that the biclustering algorithm not only enhances existing clusters but also ensures that all regulatory constraints are maintained. Furthermore, we want to stress that the overall bicluster attractiveness, as defined by (6), is also sustainable-compliant, so that the obtained results validate the effectiveness of our proposed TSB algorithm and show managerial insights for a sustainability-oriented policy implementation.

For instance, in the case of Cluster 12, traditional clustering results in a group of 8 territories with an attractiveness value of 2442. However, when biclustering is applied, the cluster is expanded to include 17 territories, and the attractiveness significantly increases to 6746. This improvement is achieved through the addition of certain territories and the replacement of less optimal ones, which collectively contribute to a higher attractiveness score. Moreover, biclustering leads to an increase in the number of territories within each bicluster compared to traditional clusters, resulting in more comprehensive groupings. This phenomenon is particularly noticeable in smaller clusters, which evolve remarkably in terms of both the number of territories and their attractiveness.

Table 2 Result of objective function and number of territories in TA constrained clusters, as reported in Andria et al. (2015), and biclusters

	Clusters		Biclusters	
ID	Attractiveness	Territories	Attractiveness	Territories
1	4476	15	4,929	17
2	32,4380	63	36,890	65
3	9615	32	10,174	32
4	4476	20	6,573	24
5	16,942	50	17,184	51
6	24,731	48	26,554	49
7	8787	26	11,191	30
8	5992	33	6,988	35
9	120	4	3,613	13
10	0,835	4	3,072	12
11	8374	24	10,550	29
12	2442	8	6,746	17
Total	120	327	144,6	374

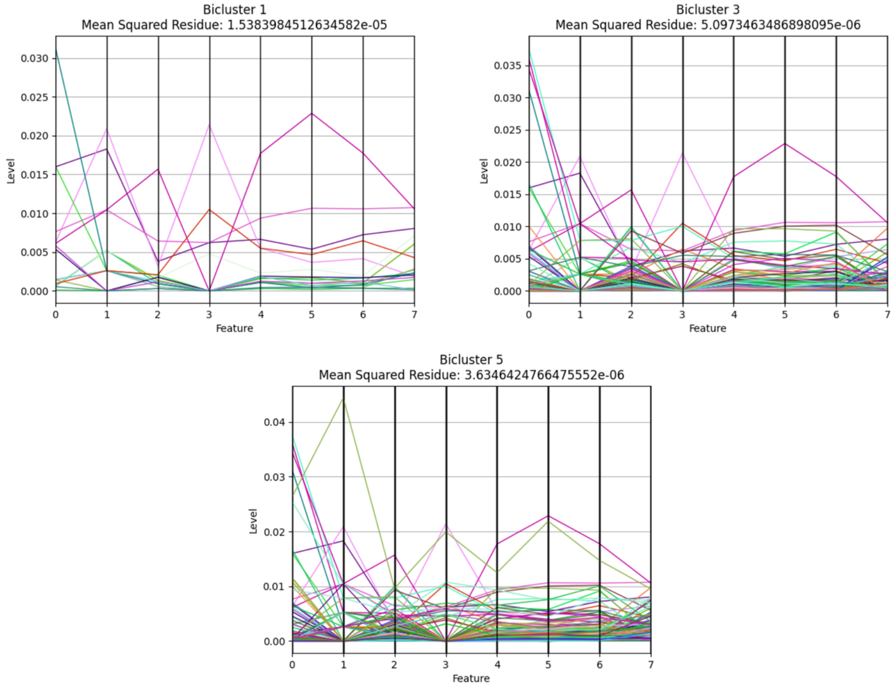
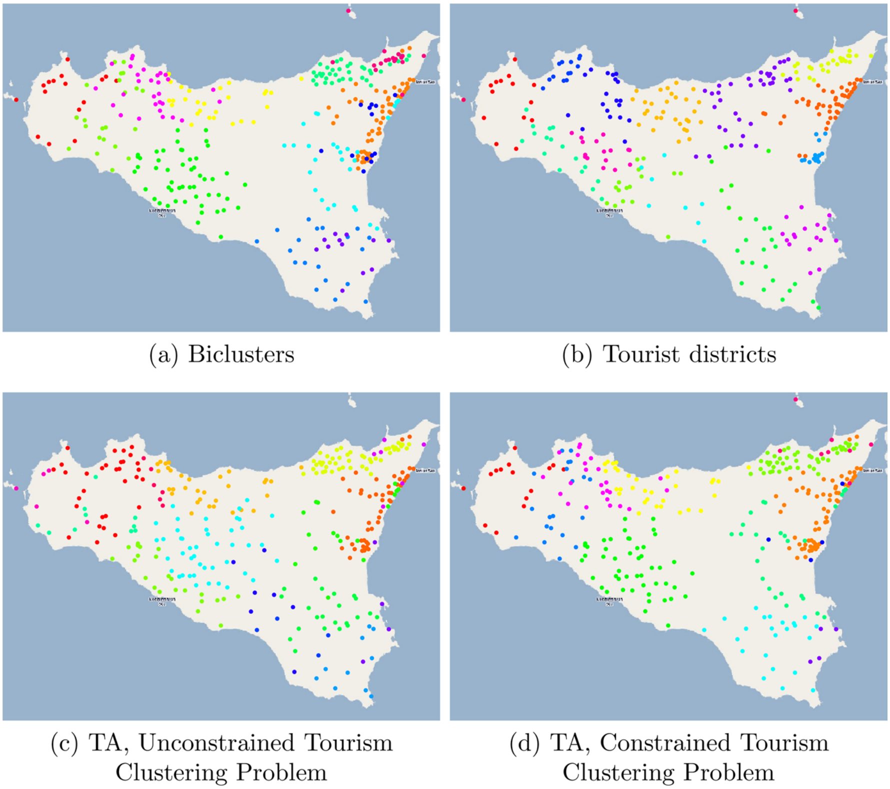


Fig. 2 Profile plot of some of the Biclusters reported in Table 2 (refer to Appendix A for the remaining ones). In the figure, colors represent the municipalities and lines trend reveal the underlying homogeneity among them across their characteristics

The resulting biclusters' profiles are reported in Fig. 2 where, for the sake of space, only some plots are shown. The remaining figures are reported in the Appendix A. We can observe that the curves representing the municipalities follow the same trend revealing an underlying homogeneity among these territories in terms of the characteristics assessed. This similarity indicates that these municipalities share comparable dynamics in terms of tourist attractiveness, suggesting a coherent distribution of the resources and infrastructures analyzed. Such a trend may stem from homogeneous public policies, similar geographical contexts or aligned local strategies aimed at strengthening the tourism offering. This observation is particularly relevant for decision-makers, as it suggests the possibility of adopting common policies and coordinated interventions to optimize the tourism potential of these territories. Consequently, these curves testify not only to the validity of the biclustering approach used, but also to its usefulness in identifying patterns that can be exploited in a perspective of integrated and sustainable tourism governance.

Fig. 3 represents biclusters visually, and depicts the geographic distribution of the municipalities grouped by our algorithm compared to the alternative approaches described in Sec. 2. The figure highlights how the biclustering method effectively partitions the region into distinct groups, each with its own set of characteristics that contribute to the overall attractiveness metric. The distinct colors in the figure represent different biclusters, showing how the algorithm has successfully identified and grouped municipalities that share common attributes, leading to an overall increase in the attractiveness metric. This visualization not only provides a clear representation of the results



(a) Biclusters
(b) Tourist districts
(c) TA, Unconstrained Tourism Clustering Problem
(d) TA, Constrained Tourism Clustering Problem

Fig. 3 Comparison between Biclusters (**a**), Tourist districts (**b**), TA Unconstrained clusters (**c**) and TA Constrained Clusters (**d**)

but also serves as a validation of the biclustering approach, showcasing its superiority in forming well-balanced and optimized clusters compared to traditional methods.

In order to assess the performances of the obtained biclusters, we are computing the number of tourists that have visited these biclusters during the years 2003-2007 (Andria et al. 2015), so this assessment is a true out-of-sample analysis. The Fig. 4 shows the overall numbers of tourist arrivals in different method identifying LTS in Sicily between 2003 and 2007. By reference to TSB algorithm over 15 million tourists visited Sicily in 2007, there was an increase of approximately 3,067,788 in the numbers going between TSB and AT algorithm. whereas there was an increase of nearly 2,270,818 tourists visiting Sicily. The number of tourists going there was similar between 2003 and 2005 but after that there was a greater increase in tourists going in 2006 and 2007. This figure takes on an increasing appearance from one year to the next; in 2007 the curves reach their highest values.

We can remark that there are tourist districts that fulfill all constraints imposed by the regulations, but whose activity related to tourism is poor since their tourist presence is negligible and close to zero and we can notice that clusters having the smallest number of visitors are always the ones belonging to the Tourist Districts. The number of tourists containing our bicluster is greater than the number existing in all clusters. This does not happen when defining tourist districts by means of TA (Constrained

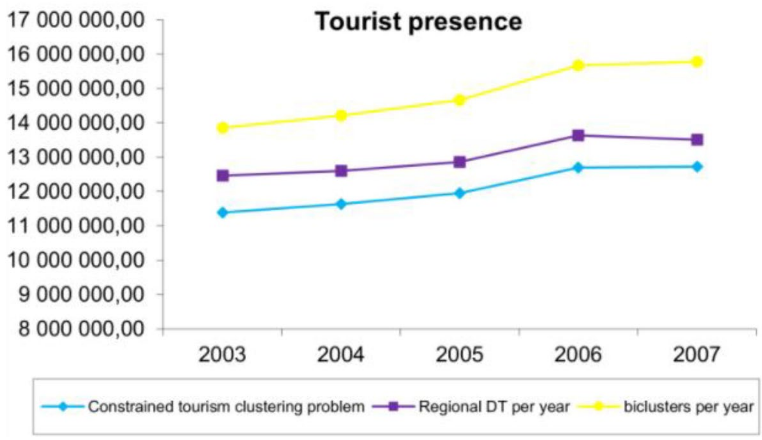


Fig. 4 Total amount throughout the clusters of tourists' presence per year and approach

Tourism Clustering Problems): all clusters have a satisfactory number of visitors over time. and since the tourist districts defined by the TSB algorithm are improvement of clusters,we can remark that all the biclusters have a better number of tourists compared to the clusters. These experiments present our algorithm as an ability to classify the territories of Sicily in a good biclusters. As a last analysis, we want to compare existing LTS with our found bicluster based approach with regards to the multi-destination trip aspect: Hwang and Fesenmaier (2003) reports that, when taking into account drive tourism, the distance of 320 miles (roundtrip) can be identified as a transition point between a single destination and a multidestination travel: this means that policy-makers needs to consider this information in order to create clusters in which the end-user can not exceed the trip of 320 miles as a road trip inside the cluster. This can be formalized by solving the Traveling Salesman Problem (referred to as *TSP*), consisting of finding the shortest possible route for a driver to visit, given a starting point, a number of municipalities and ending in a given ending point (Jünger etal. 1995).

In our case, we want the driver to visit all municipalities in the LTS and to end the cycle in the starting municipality. We have solved the TSP over all bicluster found, and results are shown in Table 3, that shows the length of the best Hamiltonian path in both LTS devised by the policy makers and the biclusters found by our approach. We can remark that our biclusters fulfill the 320 km constraints in 8 out of 12 found biclusters, whilst the LTS devised by public authorities produce only 3 out of 12 LTS that are feasible with respect to this constraint. This shows us that a computational approach may be useful to better model this kind of decision-making exercise.

6 Conclusion

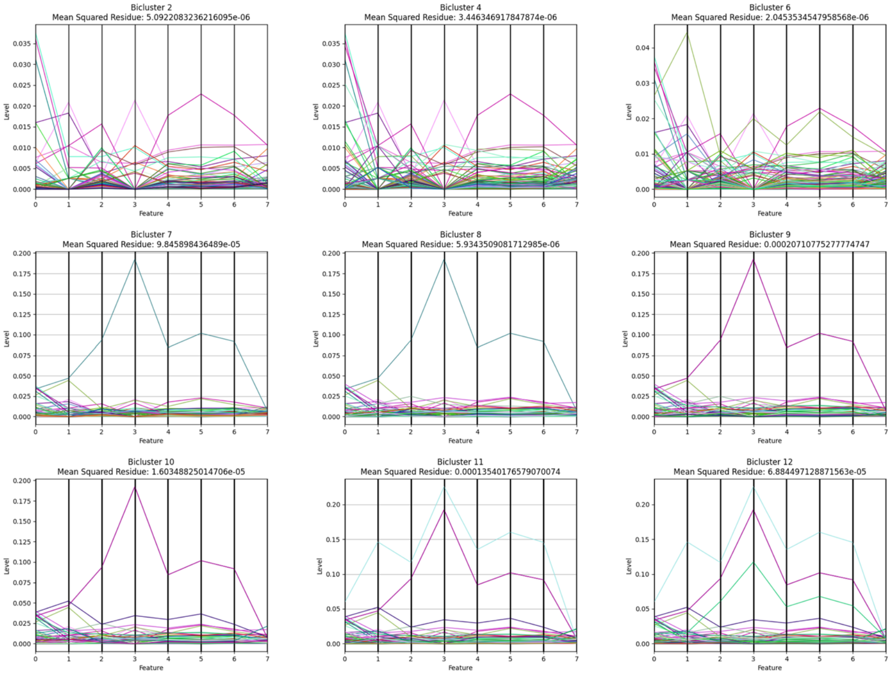
In this paper, we have used biclustering to identify local tourism system. We have emphasized the role of biclustering to ameliorate the LTS clusters that respect the guidelines provided by the regulation. Biclustering problem has been formulated as an optimization

Table 3 Number of cities and length (in km) of the cycle connecting all cities in the LTS, in Italian Local Tourism Districts and in our bi-clustering approach

LTS			Bicluster		
ID	# Cities	Length	ID	# Cities	Length
0	17	364.8	0	89	6473.4
1	59	3071.2	1	14	212.5
2	31	823.1	2	43	1976.6
3	24	552.5	3	30	308.1
4	49	1291.3	4	26	303.3
5	44	862.4	5	16	285
6	27	1952.3	6	7	155.1
7	24	607.3	7	19	427
8	12	205.1	8	17	311.2
9	11	163.9	9	8	177.4
10	28	149.8	10	13	1913.7
11	16	238.4	11	14	218.1

problem and solved by mains of the Tabu search metaheuristics. According to our experimental study which is materialized by comparing our solution with clusters and existing regional tourism districts. Results has shown that our biclusters has a number of visitors higher than the one obtained by clustering. This results highlights the capacity of biclustering to improve the LTS rather than clustering. Our approach represents a valid decision support tool. From a managerial standpoint, it facilitates the implementation of more informed strategic and policy decisions aimed at sustainable development while optimizing the use of available resources. Furthermore, it can also represent a valid tool for monitoring the effectiveness of the implemented interventions and the evaluation of the policies carried out by decision makers. Furthermore, we also point out the exportability of our approach to other contexts and application areas. Our future research will focus on introducing a new biclustering approach able to defining a local tourist system according to a type of tourism in each bicluster.

Appendix A



Author Contributions The authors contributed equally to this work.

Funding Open access funding provided by Università Politecnica delle Marche within the CRUI-CARE Agreement.

Data Availability Data from literature and 3rd party.

Code availability Not applicable

Declarations

Conflict of interest The authors have no competing interest to declare that are relevant to the content of this article.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Yes.

References

Ababkova, M., Vasileva, O.: System approach to evaluation of tourism cluster companies' environment and performance, XIII International Scientific and Practical Conference "State and Prospects for the Development of Agribusiness – INTERAGROMASH 2020" 175 (2020)

Andria, J., di Tollo, G.: Clustering local tourism systems by threshold acceptance. In: Mora, A.M., Squillero, G. (eds.) Applications of Evolutionary Computation, pp. 629–640. Springer International Publishing, Cham (2015)

Andria, J., di Tollo, G., Pesenti, R.: Detection of local tourism systems by threshold accepting, Computational Managagement. Science **12**, 559–575 (2015)

Andria, J., di Tollo, G., Pesenti, R.: Fuzzy multi-criteria decision-making: An entropy-based approach to assess tourism sustainability. Tour. Econ. **27**(1), 168–186 (2021). https://doi.org/10.1177/1354816619885207

Ayadi, W.: Local multiple sequence alignment with biclustering, International Conference on Health Informatics 231–238 (2018)

Balamurugan, R., Raja, S.P.: MSR-based algorithms for biclustering of microarray gene expression data. Curr. Sci. **123**(4), 530–541 (2022)

Belayadi, Y., Khababa, A., Attia, A., Maza, S.: An effective method based on bi-clustering and association rules for user activity analysis in location-based social network. Journal homepage **27**(6), 855–864 (2022). URL:http://iieta. org/journals/isi

Ben-Dor, A., Chor, B., Karp, R., Yakhini, Z.: Discovering local structure in gene expression data: The order-preserving submatrix problem. J. Comput. Mol. Cell Biol. **10**(3–4), 373–84 (2003)

Bergmann, S., Ihmels, J., Barkai, N.: Iterative signature algorithm for the analysis of large-scale gene expression data. Phys. Rev. E **67**(3), 031902 (2003)

Castanho, E., Aidos, H., Madeira, S.: Biclustering fmri time series: a comparative study. BMC Bioinf. **23**, 192 (2022)

Castanho, E., Aidos, H., Madeira, S.: Biclustering data analysis: a comprehensive survey, Briefings in Bioinformatics, Volume 25, Issue 4 (2024)

Charfaoui, Y., Houari, F., Boufera, A.: Amodebic: An adaptive multi-objective differential evolution biclustering algorithm of microarray data using a biclustering binary mutation operator. Expert Syst. Appl. **238**, 121863 (2024)

Cheng, Y., Church, G.M.: Biclustering of expression data. In Proceedings of the eighth international conference on intelligent systems for molecular biology **8**, 93–103 (2000)

Divina, F., Aguilar-Ruiz, J.: Biclustering of expression data with evolutionary computation. IEEE Trans. Knowl. Data Eng. **18**(5), 590–602 (2006)

Dolnicar, S., Kaiser, S., Lazarevski, K., Leisch, F.: Biclustering: overcoming data dimensionality problems in market segmentation. J. Travel Res. **51**(1), 41–49 (2012)

Dolnicar, S.: Tourism market segmentation: a step by step guide, in: Handbook of tourism economics: Analysis, new applications and case studies, World Scientific, pp. 87–104 (2013)

Freitas, A., Ayadi, W., Elloumi, M., Oliveira, L., Hao, J.: Survey on biclustering of gene expression data, Biological Knowledge Discovery Handbook: Preprocessing, Mining and Postprocessing of Biological Data. Chapter 24, Wiley (2013)

Garcia, A., Jacques, J., Sobanski, V., Dhaenens, C.: Metaheuristic biclustering algorithms: From state-of-the-art to future opportunities. ACM Comput. Surv. **56**, 1–38 (2023)

Glover, F.: Future paths for integer programming and links to artificial intelligence. Comput. Oper. Res. **13**, 533–549 (1986)

Gormsen, E.: The impact of tourism on coastal areas. Geo-Journal **42**(1), 39–54 (1997)

Houari, A., Ayadi, W., Ben Yahia, S.: A new FCA-based method for identifying biclusters in gene expression data. Int. J. Mach. Learn. Cybern. **9**, 1879–1893 (2018)

Howard, R.W.: Urban tourism districts: a taxonomy and a study of a new proposed type. Tour. Hosp. Plan. Dev. **7**(4), 415–428 (2010). https://doi.org/10.1080/1479053X.2010.520468

Hwang, Y.-H., Fesenmaier, D.: Multidestination pleasure travel patterns: empirical evidence from the American travel survey. J. Travel. Res. **42**, 166–171 (2003). https://doi.org/10.1177/0047287503253936

Jünger, M., Reinelt, G., Rinaldi, G.: Chapter 4 the traveling salesman problem, in: Network Models, Vol. 7 of Handbooks in Operations Research and Management Science, Elsevier, pp. 225–330. (1995) https://doi.org/10.1016/S0927-0507(05)80121-5. https://www.sciencedirect.com/science/article/pii/S0927050705801215

Maâtouk, O., Ayadi, W., Bouziri, H., Duval, B.: Evolutionary biclustering algorithms: an experimental study on microarray data. Soft. Comput. **23**(17), 7671–7697 (2019)

Maâtouk, O., Ayadi, W., Bouziri, H., Duval, B.: Evolutionary local search algorithm for the biclustering of gene expression data based on biological knowledge. Appl. Soft Comput. **104**, 107177 (2021)

Maâtouk, O., Ayari, E., Bouziri, H., Ayadi, W.: Bobea: a bi-objective biclustering evolutionary algorithm for genome-wide association analysis, GECCO Companion 344–347 (2022)

Moebius, C.: Equitable & sustainable heritage districts: Lessons learned from little havana, https://planning.maryland.gov/Documents/Offsite/Webinar-Slides/2022-0818-Dr-CMoebius-Slides.pdf (2022)

Noronha, D., Henriques, R., Madeira, S., Zárate, L.: Impact of metrics on biclustering solution and quality: A review. Pattern Recogn. **127**, 108612 (2022)

Porter, M.: Location, competition and economic development: local clusters in a global economy. Econ. Dev. Q. **14**(1), 7–20 (2000)

Prelic, A., Bleuler, S., Zimmermann, P., Wille, B.P.G.W., Zitzler, A.E.: A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics **22**(9), 1122–1129 (2006)

Regione, S.: Criteri e modalità per il riconoscimento dei distretti turistici, Gazzetta Ufficiale della Regione Siciliana (2010)

Rocha, H.: Entrepreneurship and development: The role of clusters. Small Bus. Econ. **23**(5), 363–400 (2004). (https://EconPapers.repec.org/RePEc:kap:sbusec:v:23:y:2004:i:5:p:363-400)

Seridi, K., Jourdan, L., Talbi, E.-G.: Biclustering of expression data with evolutionary computation. Appl. Soft Comput. **33**, 239–249 (2015)

Sun, J., Zhang, Y.: Recommendation system with biclustering. Big Data Min. Anal. **5**(4), 282–293 (2022)

Tennessee Department of Tourist Development, Adventure tourism, https://industry.tnvacation.com/industryresourcestourism-initiatives/adventure-tourism (2024)

Testa, M.R., Sipe, L.J.: Economic storytelling, tourism improvement districts and university partnerships. J. Econ. Manag. Trade **27**(11), 1–8 (2021). https://doi.org/10.9734/jemt/2021/v27i1130373. (https://journaljemt.com/index.php/JEMT/article/view/971)

Tourism Economics - An Oxford Economics Company, The economic case for tourism improvement districts, https://s3.amazonaws.com/tourism-economics/craft/TourismEconomics_TheEconomicCaseforTIDs.pdf (2021)