Leveraging tourist trajectory data for effective destination planning and management: A new heuristic approach

Weimin Zheng [a], Mengling Li [a], Zhibin Lin [b,*], Yangyu Zhang [a],

[a] *School of Management, Xiamen University, 422 South Siming Road, 361005, Xiamen, China*

[b] *Durham University Business School, Mill Hill Lane, Durham, DH1 3LB, United Kingdom*

Corresponding author: Zhibin Lin

Zhibin Lin, Associate Professor
Durham University Business School, Mill Hill Lane, Durham, DH1 3LB, United Kingdom
E-mail: zhibin.lin@durham.ac.uk

Leveraging tourist trajectory data for effective destination planning and management: A new heuristic approach

Abstract:

Understanding tourist movements provides insights for destination planning, service design and marketing. The key challenge is to develop a tool that can capture the value in the tourist mobility data. This study presents a new heuristic approach that combines adaptive spatial clustering with frequent pattern mining to improve the performance and efficiency of trajectory data analytics. The aim is to fully leverage the semantic information in the tourist data and the duration that tourists stay at an attraction. Anonymous mobile positioning data from 741 tourists to one of China's leading destinations are used to illustrate the application of the new analytical approach. The results reveal a four-level destination spatial structure ranging from core to peripheral areas. The findings provide practical implications for facilitating intra-destination cooperation and optimizing destination resource allocation and service design.

Keywords: Tourist mobility; Frequent pattern mining; Time-space; Core-peripheral; Data mining; Intra-destination cooperation.

1. Introduction

Tourist mobility in a destination is limited to two essential resources, i.e., space and time (Grinberger & Shoval, 2019). When organizing trips, tourists must allocate time for visiting attractions and the transport between different locations (Neutens, et al., 2011). An increase in time spent on one activity (e.g., transport) means less time for another activity (e.g., enjoying a rare exhibition in the museum). The spatio-temporal distribution of tourists reflects the composition of a network of attractions within a destination. The core-periphery model of attractions provides a systematic understanding of the geographical distribution of attractions in the destination (Prideaux, 2002). A primary attraction can attract more tourists than smaller ones and a cluster of attractions can attract more tourists who stay longer than isolated attractions (Lue, et al., 1993). The spatial structure of a destination is constantly shaped and reshaped by tourist movement. Mapping tourist movements provides an understanding of the structure of the destination and the interconnections among destinations (Park, et al., 2020).

With the constraints of space and time, the patterns of tourist movement at a destination are determined by tourists' decision making and the duration that tourists stay at an attraction reflects their preferences. Tourist decision making is influenced by various psychological, social, and cultural factors (Chen, et al., 2016; Dejbakhsh, et al., 2011; McCormack & Schwanen, 2011). The semantic information in tourist mobility data offers a richer understanding of tourists' spatio-temporal experience (Shoval, et al., 2018). Such an understanding is valuable for destination planning, service design, and management.

Much of the tourism mobility research has been devoted to advance the mining and processing of a large amount of spatio-temporal data. Earlier studies have explored various data sources, such as on-site surveys (Yang, et al., 2013), travel diaries (McKercher & Lau, 2008). user-generated photos (Önder, et al., 2016) and officially published secondary data (Yang & Wong, 2012, 2013). In today's digital era, mass data about tourist mobility can now be automatically captured for tourism research thanks to the advances in mobile tracking technologies in recent years (Shoval, et al., 2020; Shoval, et al., 2018). One of the key challenging questions for tourism mobility researchers is the analysis of a large amount of spatio-temporal data (Birenboim & Shoval, 2016). Such an analysis requires cross-disciplinary efforts to develop practical tools that are based on highly advanced and sophisticated algorithms (Grinberger & Shoval, 2019).

Several studies have examined the mining of frequent trajectory patterns (Wang, et al., 2013; Wang, et al., 2016), i.e., the discovery of frequent repetitive sequences from a large number of trajectory data (Zheng, 2015). Recently, Shoval, et al. (2018) used methods that combine spatio-temporal data with tourist emotion data and semantic information to understand tourist experience in time and space. Shoval, et al. (2020) combined mobile tracking data with a traditional survey to understand tourist time-space activities. Despite the progress, two key information points have not been fully leveraged in the tourism research, namely, a) the semantic information of tourist data, and b) the duration that tourists stay at an attraction (Shoval, et al., 2020; Shoval, et al., 2018). As such, existing methods cannot be directly applied in the tourism field. Specifically, given the uneven distribution of tourists, it is difficult for the traditional clustering methods to mine the semantic information of tourism activities from the spatio-temporal frequent trajectory patterns in multidimensional spatial granularity. Recent studies have begun to consider the time factor; however, they are more about the time interval between locations than the duration at the locations, which has rich semantic information that has not been fully leveraged for effective destination planning and management.

The primary goal of this study is to address the above gaps by adopting an advanced data mining and analytical approach that integrates adaptive spatial clustering with frequent pattern mining. This study contributes to the tourist mobility literature by offering a new approach to fully utilize the spatio-temporal information contained in the tourist trajectory data. This new approach helps to improve the performance of spatial clustering and the efficiency of frequent pattern mining by (a) designing an adaptive spatial clustering approach based on DBSCAN for the tourism contexts; (b) proposing BrowseRank to fully capture the spatio-temporal characteristics of tourist movements, especially the time tourists stay at each attraction, and (c) using the Depth-First-Search (DFS) strategy to guarantee better trajectory integrity and higher efficiency. Particularly, our proposed approach significantly advances trajectory data analytics by exploiting both the semantic and the time information contained in the tourist mobility data, which have not yet been adequately explored in the extant literature. Furthermore, this study provides practical implications for destination planning and management to improve tourist experience and enhance destination competitiveness. Specifically, the findings derived from our proposed approach will equip destination planners with greater insights into tourist movement patterns, and help them to optimize the layout of service facilities and the connections between tourist attractions.

2. Literature review

2.1. Theoretical foundations

The theoretical foundations of tourist mobility can be found in time geography, core-peripheral model, cumulative attraction, gravity theory and consumer behavior theories.

2.1.1. *Time geography*

Space and time are two inseparable features of tourist movement or mobility because every activity includes spatial and time dimensions (Grinberger & Shoval, 2019). Time geography provides a valuable theoretical foundation for understanding tourist time-space behaviors (Ellegård & Svedin, 2012). Visiting an attraction or traveling between the hotel and the attraction are considered as spatial behaviors, and when organizing these behaviors one must consider the time to be spent in each attraction and the time for travel between attractions (Neutens, et al., 2011). Hence, time and space are involved in every activity (Neutens, et al., 2011). Tourist spatio-temporal resources are limited by numerous constraints, such as the length of holiday, opening hours of attraction, individual differences in holiday budget, knowledge of the destination, personal preferences, decision-making styles, cultural background, and many others (Long & Nelson, 2013). With spatio-temporal constraints, the pattern of tourist movement is determined by tourists' conscious decision making to maximize the utilities (Chen, et al., 2016; Yang, et al., 2013).

2.1.2. *The core-periphery model*

The core-periphery model is a concept that originated in economic geography (Krugman, 1991; Weaver, et al., 2021), which suggests that in a region, there are certain core areas and several periphery areas. Major social, political and economic activities are concentrated in the core areas, while the periphery areas play a supporting role and are normally dependent on the core. Core areas are generally located in the urban center, and the periphery ones are usually in the suburban and rural areas (Chancellor, et al., 2011). In the tourism context, the core areas are those that have a concentration of popular attractions, and the periphery areas are those receiving very few visitors. Traditionally location and distance are used to demarcate the core and the periphery (Prideaux, 2002). However, rapid improvement in transportation infrastructure may create a "time-space compression" effect (Yin, et al., 2019), which may change a destination's core-periphery structure.

2.1.3. *Cumulative attraction*

The idea of cumulative attraction suggests that a cluster of attractions or destinations can attract more tourists who stay longer than a single attraction site or destination that does not closely connect to a cluster (Lue, et al., 1993). This is because tourists usually prefer to visit several attractions or destinations when having a pleasure trip (Hwang, et al., 2006; Lue, et al., 1993). Visiting attractions or destinations that are closely interconnected saves time, effort, costs, and other resources. Therefore, combining a popular primary attraction with several less popular ones to form a cluster of attractions offers efficiency and a satisfactory experience for the tourists, thus enhancing the competitiveness of the destination as a whole (Lue, et al., 1993).

2.1.4. Gravity theory

Tourism attractions may scatter in various places around the destination (Buhalis, 2000; Edwards, et al., 2008). Some are large, primary attractions, while others are small, secondary ones. According to gravity theory, large, primary attractions have a greater gravitational pull of visitors than secondary ones. A cluster of primary attractions combined with several secondary attractions can enhance the competitiveness of the whole destination. Moreover, a destination that has multi-functionality and attraction diversity is likely to attract a large number of tourists who have various motivations and interests (Edwards, et al., 2008). Similarly, clustering a primary destination with several secondary destinations can enhance the competitiveness of the whole region (Lue, et al., 1993).

2.1.5. Factors influencing the tourist trajectory patterns

Consumer behavior theories provide useful insights into how the patterns of tourist movement are determined by tourists' intentional behaviors within the limit of time and space. According to random utility theory, tourists attempt to maximize the utilities when making travel decisions. They choose to visit and spend time at those attractions that have the greatest utility (Yang, et al., 2013). They use cognitive representations of the destination to aid their decision making (Grinberger & Shoval, 2019). However, tourists are not entirely rational decision-makers because their rationality is bounded, given the limitations of information, time, and cognitive capabilities (Chen, et al., 2016). The cognitive representation depends on various factors such as individual characteristics, familiarity with the destination, pre-travel preparation, and the use of navigation aid (McKercher, et al., 2012).

Tourist cognitive representations are usually biased or distorted, leading to a gap between tourists' original intentions and the resulting actions. Moreover, many tourist behaviors are guided by intuition or emotions rather than rationality. Furthermore, social factors such as power relations within a travel group may influence travel decision making and shape spatio-temporal patterns (McCormack & Schwanen, 2011). Finally, tourist cultural background plays a crucial role in tourist choices of transport mode, activities, attractions, and the type of accommodation, which consequently influence the movement patterns at the destination (Dejbakhsh, et al., 2011).

2.2. Mining and analyzing tourist trajectory patterns

Data captured from GPS-based tracking or social media in their raw form are not useful unless properly processed (Shoval & Ahas, 2016). Researchers have used the Sequence Alignment Method (SAM) to identify patterns of tourist behavior from a large dataset of tourist trajectories and cluster groups of tourists (Delafontaine, et al., 2012; Lee & Joh, 2010; Shoval, et al., 2015). SAM uses dynamic programming algorithms to seek optimal alignments through similarity or distance measures (Wilson, 2008). The method was first developed in biology and later has been applied in geography for analyzing individuals' mobility and identifying the trajectory patterns (Stehle & Peuquet, 2015; Wilson, 1998). In geography, researchers have adopted SAM to identify the various spatio-temporal patterns of tourist movement (Stehle & Peuquet, 2015), for example, similar modes of transportation (Crawford, et al., 2018), movement habits (Dharmowijoyo, et al., 2017; Millward, et al., 2019) and social change (Delmelle, 2016). SAM is generally applied to the tourism field to classify the sequence of tourists' movements and extract the common features of each pattern. For example, Shoval and Isaacson (2007) summarise tourist movement behaviors into three groups based on the GPS track data of 305 tourists; and Shoval, et al. (2015) propose five typical spatial movement patterns based on the similarity of trajectories of 139 tourists.

Another stream of research has focused on mining the frequent trajectory pattern. Cao, et al. (2005) define frequent patterns as the spatial regions around frequent line segments. They first divide original sequences into sequence segments and then adopt a line simplification technique named Douglas-Peucker algorithm to transform the sequence segments into directed line segments. Finally, they group segments considering similar shapes and closeness in space. Wang, et al. (2016) characterize original trajectories with the corners in a road network for mining the frequent trajectory patterns. Several studies have tried to divide the space into grids, characterize trajectories with the sequences of grids, and then mine the frequent trajectory patterns (Tsoukatos & Gunopulos, 2001; Wang, et al., 2013). However, gridding spatial partition may cause strict spatial constraints or the sharp boundary problem, and thus fails to identify meaningful trajectory patterns (Han, et al., 2010).

Recent studies use clustering-based methods to solve this problem. Researchers first cluster all location points in the spatio-temporal trajectory database into regions of interest (ROIs), characterize trajectories with ROIs, then mine frequent trajectory patterns (Zheng, 2015). Kalnis, et al. (2005) consider patterns in the form of moving regions within time intervals from a clustering-based perspective. Giannotti, et al. (2007) propose a frequent sequence mining method, in which the trajectory sequence is first transformed into a sequence of ROIs in the data preprocessing phase. In the modeling phase, individual GPS data is preprocessed into a usable data format and inputted to the mining phase wherein separate strategies are applied to discover frequent trajectory patterns. Kang and Yong (2010) adopt a prefix-projection approach to uncover meaningful spatio-temporal regions. Huang, et al. (2016) first derive the ROIs using DBSCAN clustering algorithm and then identify ordered sequences of these spatial regions using an Apriori-like algorithm.

Both SAM and frequent pattern mining take into account the influence of the sequence of activities to identify behavioral patterns from a large dataset of tourist trajectories. However, there are fundamental differences between the two methods. SAM is based on two types of analysis: clustering the similarities in tourist mobility sequences and detecting the movement patterns from the sequence data (Wilson, 1998). SAM focuses on the commonalities of the whole sequences of tourists to obtain the characteristics of the complete movement routes. It does not quantify the intensity of the associations between attractions. In contrast, frequent pattern mining extracts the high-frequency tourist movement and identifies the connections between attractions and their combination patterns. It contains the patterns among two, three, or four attractions, in a more microscopic perspective. Nevertheless, the weight of the time dimension has rarely been considered in either SAM or the frequent pattern (Shoval & Isaacson, 2007). This is to be addressed in the current study.

3. Methodology

3.1. Data collection

We obtained access to users' mobile positioning data in Xiamen, through one of the largest telecommunication companies in China. Xiamen is a typical littoral city located on the southeast coast of China, which is a well-developed, urban destination (Fig. 1). In 2019, over 100 million tourists visited Xiamen, increasing by 12.5% in 2018 and the total tourism revenue of Xiamen was 165.59 billion yuan. Xiamen has charming and abundant tourism resources, including coastal scenery, historical culture and gastronomy. Especially, the Gulangsu, which is located in the southwest of Xiamen, is on China's list of National Scenic Spots and is also listed as a world heritage site by UNESCO, with its ancient relics and varied architecture. Hence, Xiamen is an appropriate case with important practical value. The city can be divided into two parts: one is the southern island area and Kulangsu, in which most attractions are located; another is the mainland area. The two parts are mainly connected by several sea-crossing bridges and undersea tunnels.

Compared with other data sources, the mobile phone positioning data have advantages of wide coverage, large sample size and non-disturbed collection (Ratti, et al., 2006; Zhao, et al., 2018). Our data was passively collected without the direct participation of tourists, generating massive and continuous tracking data. The data was very detailed, including the specific latitude and longitude of each location and some demographic information, such as age and gender. The data was strictly desensitized and anonymized, which cannot correspond to any individual. According to the registration area and frequent activity area of mobile phone users, the operators can distinguish whether the users are local residents or tourists (Ahas, et al., 2008). Based on this, we screened 1,996 tourists who visited Xiamen during the Spring Festival holiday (7 days) in 2017, and their spatio-temporal trajectories in Xiamen were obtained. The data was collected from mobile phone stations at the 30-minute interval and the individual level. The station densities in space, which are usually higher in urban areas, affect the spatial granularity of the dataset. Therefore, our data has a high resolution of spatio-temporal granularity and rich information. Through data cleaning and filtering, we excluded trajectories with poor quality, such as trajectories with data missing due to signal problems, or trajectories with sparse location points, which could be from passengers who had a short stopover in Xiamen. Finally, 741 trajectories were selected for the study. Each of them contains more than 20 location points. The distribution of trajectory points of all tourists is shown in Fig. 2.

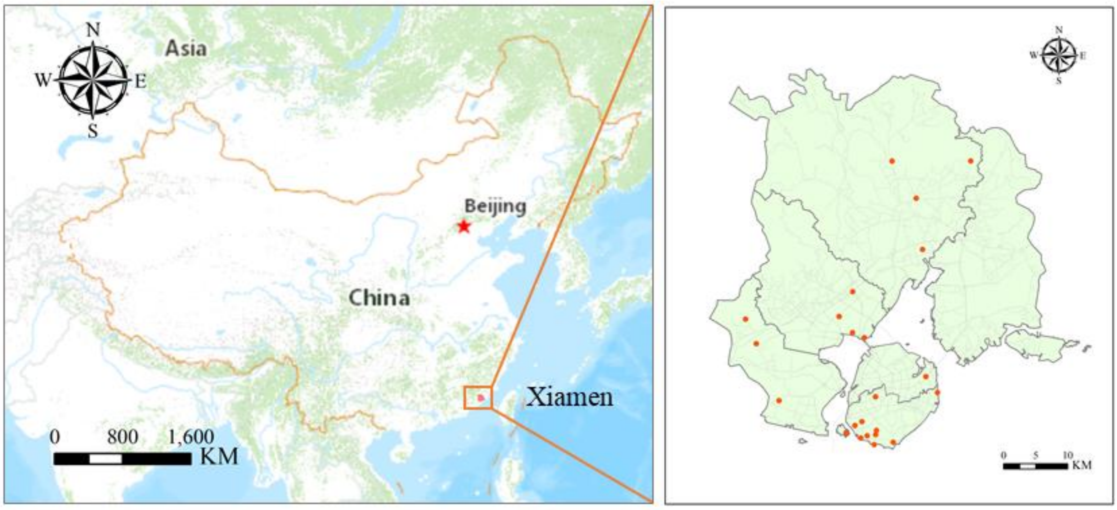

Fig. 1 Xiamen Map

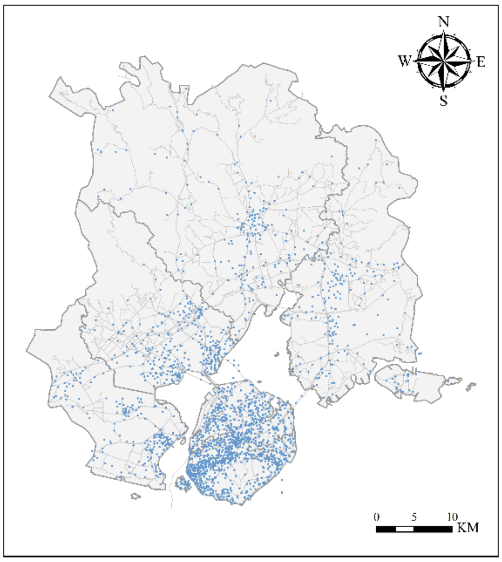

Fig. 2 The distribution of trajectory points of all tourists

3.2. Data analysis

The data analysis consists of two main stages: adaptive spatial clustering based on DBSCAN and frequent pattern mining based on BrowseRank and Markov model. In the first stage, we cluster all location points in the spatio-temporal trajectory database into clusters based on the adaptive DBSCAN. Compared with the traditional spatial clustering methods, the adaptive DBSCAN has the advantages of clustering efficiency, noise processing ability, and shape adaptability. It can mine the semantic information of tourism activities in multidimensional spatial granularity. Moreover, the convex hull algorithm is adopted in this stage to deal with the sharp boundary problem encountered by most spatial clustering algorithms.

In the second stage, we first characterize the trajectories with the sequences of clusters, then mine the frequent trajectory patterns based on BrowseRank. Compared with the existing methods, the BrowseRank can fully take the duration time into consideration when extracting frequent patterns, which better reflect the tourists' attraction preference and characteristic of movements. The detailed description of the two stages is presented in the following subsections.

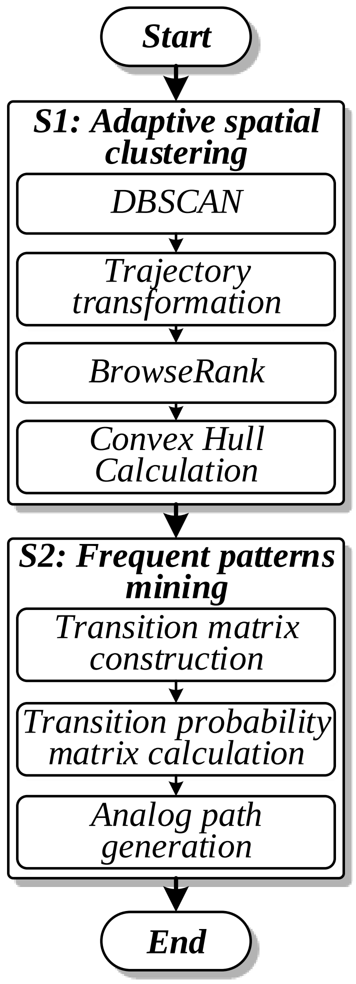

Fig. 3 Methodological framework

3.2.1. *Adaptive spatial clustering*

We design an adaptive spatial clustering algorithm based on DBSCAN by introducing an adaptive parameter (δ). DBSCAN has the advantages of clustering efficiency, noise processing ability, and shape adaptability (Park, et al., 2020). In this algorithm, regions with sufficient density are divided into clusters, and clusters of arbitrary shapes can be found in noisy spatial databases. However, the clustering performance of DBSCAN is often poor in the scene of uneven distribution of location points because selecting the algorithm parameters (*eps* and *minPts*) suitable for all clusters is very difficult. It is common that tourists are unevenly distributed in time and space (Riganti & Nijkamp, 2008), therefore, an adaptive parameter is required. The adaptive parameter (δ) helps us to mitigate the effects of the differences in population density distribution, by judging the number of times DBSCAN used in different areas.

Besides, since two parameters, distance measurement (*eps*) and the minimum size of a cluster (*minPts*), are significantly affect the clustering performance of DBSCAN, Dunn Validity Index (DVI) (Dunn, 1974) is applied to choose the appropriate parameter values. The index is an effective method to evaluate the validity of clustering algorithms, and it quantifies not only the degree of compactness of clusters but also the degree of separation between individual clusters (Fahad, et al., 2014; Liu, et al., 2013). DVI is widely applicable, not limited to a specific destination or the amount of data. The index is determined by the ratio of the shortest distance between any two clusters and the maximum distance within any cluster. The higher the DVI value, the closer the clustering results are in the same cluster, and the different clusters are far away from each other. In other words, the higher the DVI value, the better the clustering result. Once the value of DVI is determined, we can decide the corresponding combination of the two parameters. This way of determining DBSCAN parameters is a data-driven approach, which is an adaptive and self-learning mechanism that can effectively avoid the subjectivity in choosing the right parameters, thus improving the performance of DBSCAN. Therefore, we chose the most suitable parameters (*minPts* and *eps*) for DBSCAN through DVI, to overcome the shortcomings of the DBSCAN and make the results more reasonable. The detailed algorithm process is as follows and we illustrate the process with a schematic figure (Fig. 4).

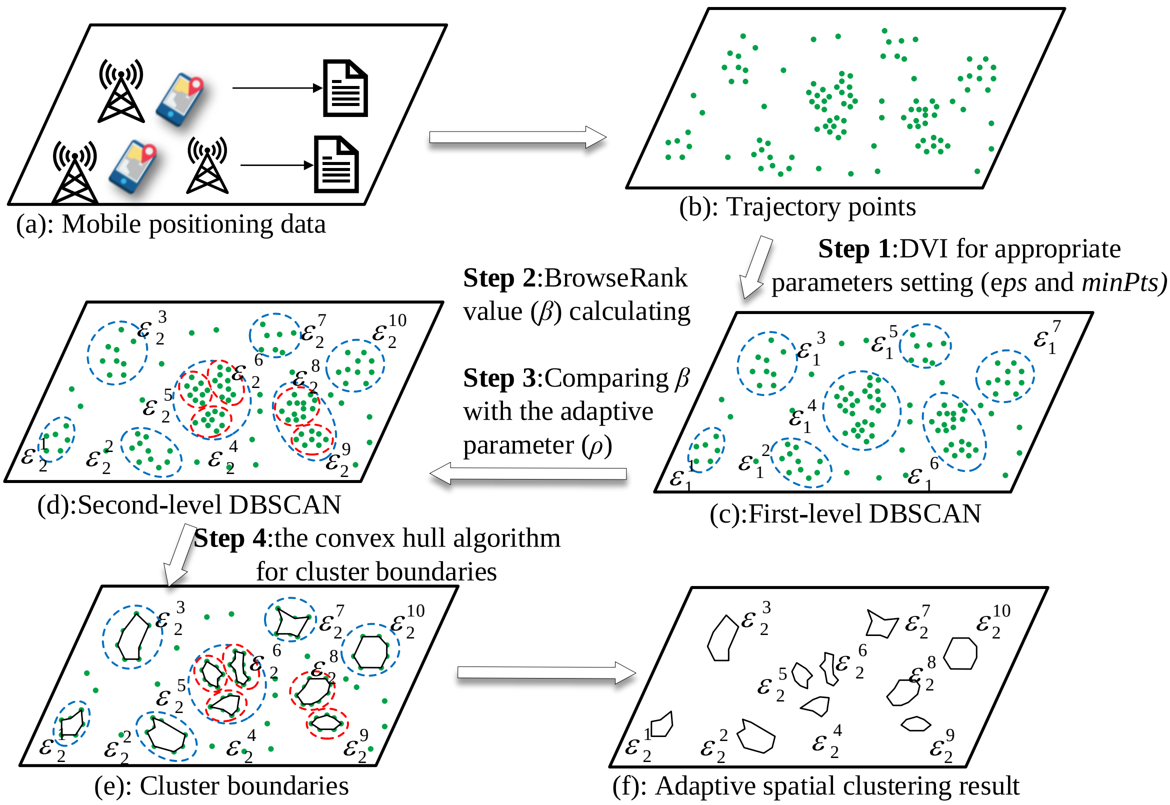

Fig. 4 The process of adaptive spatial clustering

Given the set of points ($\boldsymbol{\rho}$) and adaptive parameter (δ), the output of this algorithm is the set of clusters ($\boldsymbol{\mathcal{E}_n}$) and the BrowseRank value (β) of the clusters contained in $\boldsymbol{\mathcal{E}_n}$ ($\boldsymbol{\beta}(\boldsymbol{\mathcal{E}_n})$). First, the points in $\boldsymbol{\rho}$ based on DBSCAN are clustered, and the set of clusters ($\boldsymbol{\mathcal{E}_n}$) is obtained, where

Θ is set to be the number of clusters contained in $\boldsymbol{\mathcal{E}_n}$. The original trajectories are transformed into characteristic trajectories based on $\boldsymbol{\mathcal{E}_n}$. The DVI is used for the parameter determination of DBSCAN. In Fig. 4(c), the first-level DBSCAN makes 7 clusters.

Second, the β of each cluster in $\boldsymbol{\mathcal{E}_n}(\beta(\boldsymbol{\mathcal{E}_n}))$ is calculated using BrowseRank algorithm. BrowseRank is an alternative to PageRank from Google that evaluates the popularity of a page (for details see Liu, et al., 2008). Inspired by the application of BrowseRank in the field of web searching and ranking, we innovatively apply it to the tourism field by considering elements of the tourism scenes. In this study, the β of each cluster is calculated according to the number of visits to this cluster, the number of transitions from other clusters to this cluster, and the duration of time spent in the cluster. The larger the β of the cluster, the more popular the cluster is. Since the mobile positioning data is at 30-min intervals unchangeably, the number of points that appear in the cluster can be used to refer to the duration time of the tourist in the cluster, and the duration time of each cluster is obtained. By adding BrowseRank to the algorithm design, several crucial features of tourist spatio-temporal behaviors are fully considered, including the duration time in each cluster for the temporal dimension, the sequence and direction of tourist movement among clusters for the spatial dimension. Our final clustering results are closely in line with tourists' actual preferences. In contrast, prior tourism research has rarely considered the important value of temporal dimension in spatial clustering.

Third, we compare each β of the cluster with the adaptive parameter (δ). When the cluster's β exceeds the adaptive parameter (δ), it means the cluster is where tourists gather highly. For the purpose of generating greater management insights, these places need to be further refined. Therefore, our adaptive DBSCAN is applied again to achieve finer clusters, and the clustering results and β of each cluster are derived. In Fig. 4(d), since the β of the clusters ε_1^4 and ε_1^6 exceed the adaptive parameter (δ), a second-level DBSCAN makes them to ε_2^4, ε_2^5, ε_2^6 and ε_2^8, ε_2^9 respectively.

After clustering the spatial points, the convex hull algorithm proposed by Graham (1972) is employed to handle the boundary of the clusters (Fig. 4(e)). The term of the convex hull is the smallest polygon that covers all the points in the given set (Barber, et al., 1996), and in this study, a cluster can be regarded as a point set (for the details of the convex hull algorithm, please see Barber, et al., 1996). Finally, 10 clusters with clear geographical boundaries are made (Fig. 4(f)).

3.2.2. Frequent pattern mining

In the stage of adaptive spatial clustering, the set of clusters $\boldsymbol{\mathcal{E}_n}$, the β of clusters in $\boldsymbol{\mathcal{E}_n}$ ($\beta(\boldsymbol{\mathcal{E}_n})$), and the set of characteristic trajectories ($\boldsymbol{\Gamma}$) are obtained. In this stage, the transition matrix is first constructed according to $\boldsymbol{\Gamma}$. If a visit to the cluster ε_n^i is followed by ε_n^j in the θth characteristic trajectory, then we set 0–1 discrete variable x_{ij}^{θ} to 1, and 0 otherwise. Thus, the transition frequency from ε_n^i to ε_n^j can be calculated according to Eq. (1), where $\boldsymbol{\lambda}$ denotes the number of characteristic trajectories. Correspondingly, the transition probability from ε_n^i to ε_n^j can be calculated based on Eq. (2). After that, the transition probability matrix (**PM**) is generated, as shown in Eq. (3).

$$y_{ij} = \sum_{\theta=1}^{\lambda} x_{ij}^{\theta} \tag{1}$$

$$p_{ij} = \frac{y_{ij}}{\sum_{j=1}^{\theta} y_{ij}} \tag{2}$$

$$\mathbf{PM} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1\theta} \\ p_{21} & p_{22} & \cdots & p_{2\theta} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\theta 1} & p_{\theta 2} & \cdots & p_{\theta\theta} \end{bmatrix} \tag{3}$$

We take into account both the duration time of tourist stay and spatial displacement by proposing the process of analog path generation. Specifically, we search the frequent trajectory patterns whose path strength (as shown in Eq. (4)) exceeds a predetermined frequent pattern threshold (ξ). In this step, we take DFS (depth first search) as the search algorithm. When the path strength is higher than the threshold (ξ), a further search is carried out to obtain a longer path; otherwise, the search process stops. That is, the goal of DFS in this study is to find the path with the greatest depth value. Suppose that a path $P = \{\varepsilon_n^1, \varepsilon_n^2, \cdots, \varepsilon_n^{\psi}\}$, then the strength of this path ($\Phi(p)$) can be calculated according to Eq. (4), where β_n^i denotes the β of ε_n^i, and $\mathbf{PM}[i][i+1]$ denotes the transition probability from ε_n^i to ε_n^{i+1}. In this manner, with β, the index reflecting the duration time in clusters, incorporated into the calculation formula of path strength, our approach overcomes the drawbacks of most frequent pattern mining methods that do not take into account the temporal dimension and enhances the practical value of the results.

$$\begin{aligned} \Phi(P) &= \beta_n^L \times \prod_{i=1}^{\psi-1}\left[\beta_n^i \times \mathbf{PM}[i][i+1]\right] \\ &= \beta_n^1 \times \mathbf{PM}[1][2] \cdot\cdots\cdot \times \beta_n^{\psi} \times \mathbf{PM}[\psi-1][\psi] \end{aligned} \tag{4}$$

4. Results

4.1. Sample profile

The profiles of the 741 sample tourists are presented in Table 1. The typical length of stay was between three and four nights. Male tourists far outnumbered female tourists, and the majority were 36–45 years old. There were only two minors, and the reason could be most juveniles do not have their own mobile phones. This is an inherent bias of mobile positioning data. Moreover, the vast majority of tourists came from economically well-developed Eastern China.

Table 1 Sample description

Days of stay	1	2	3	4	5	6	7	Total
Gender								
Male	2	86	126	110	61	84	8	477
Female	2	45	71	73	30	39	4	264
Age								
17 or below	0	0	0	1	1	0	0	2
18–35	1	36	29	39	17	46	7	175
36–45	0	35	68	53	40	28	1	225
46–55	1	27	49	45	14	24	2	162
56 or above	2	33	51	45	19	25	2	177
Length of access								
12 months or below	0	11	17	8	3	15	3	57
13–60 months	1	38	28	38	14	53	4	176
61–120 months	1	16	35	17	14	11	0	94
121–180 months	1	20	41	26	20	24	3	135
181–240 months	1	45	74	89	39	18	2	268
241 months or above	0	1	2	5	1	2	0	11
Origin								
North China	1	21	46	35	27	49	2	181
East China	2	98	143	142	55	64	9	513
South Central China	1	12	8	6	9	10	1	47

4.2. Adaptive spatial clustering

First, the adaptive spatial clustering method based on DBSCAN is applied to cluster all the points. We determined the approximate parameter range for this study, based on a review of the relevant literature (Ester, et al., 1996), and traversed more than 80 parameter combinations within that range. For each combination of parameters, we obtained its clustering result. Each clustering result corresponds to a DVI value. In the first level, the maximum DVI ($\boldsymbol{\mathcal{E}}_1$) is obtained when *minPts* = 4 and *eps* = 0.014. The first-level DBSCAN clustering set ($\boldsymbol{\mathcal{E}}_1$) can be found (Fig. 5(left)). Nine clusters are contained in this set. However, the β of ε_1^1 and ε_1^2 exceed the adaptive parameter δ, which means these two clusters are the areas with a high concentration of tourists. If we do not take a more nuanced division, the whole inland area will be grouped into one cluster (ε_1^1) (Fig. 5(left)), which is opposed to the real situation and not conducive to the targeted management. Therefore, we then further cluster these two clusters in the same way and achieve the second-level DBSCAN clustering set ($\mathcal{E}_2$), which includes 30 clusters (Fig. 5(right)). The β of each cluster in $\mathcal{E}_2$ is less than δ. With $\mathcal{E}_2$ as the output of this stage, we further handle the clusters in $\mathcal{E}_2$ using the convex hull algorithm (Fig. 6).

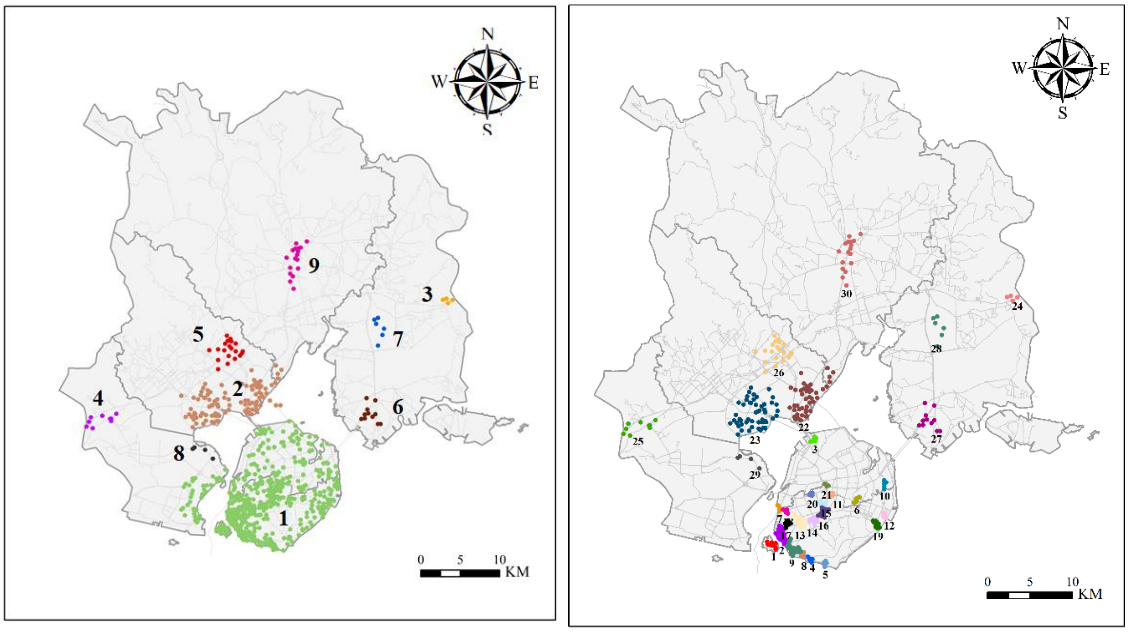

Fig. 5 Two processes of clustering

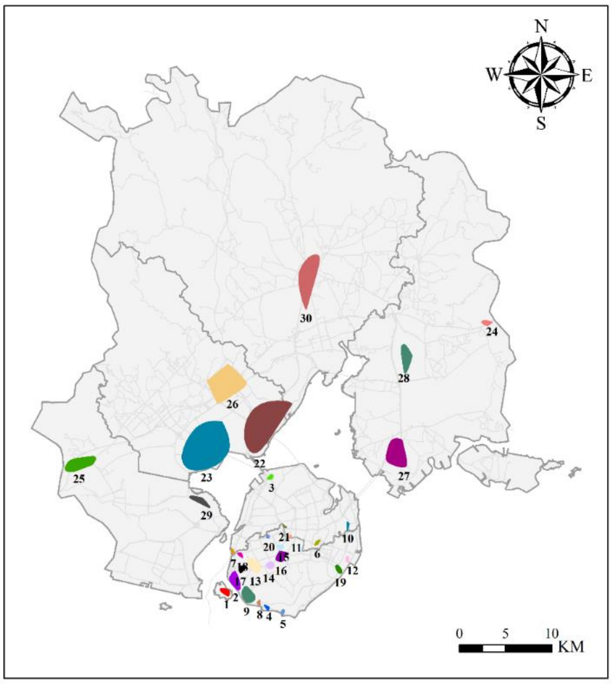

Fig. 6 Final clustering performance with 30 clusters

Table 2 displays the clusters and the primary tourist attractions involved in the corresponding clusters. In the areas outside the island, a cluster may contain a large area because there are few visitors; while on the island, areas with dense tourists can be grouped into finer clusters. Especially in the southwestern of the island, several popular areas are in the close vicinity, such as ε_2^4 (Hulishan Fortress), ε_2^8 (Baicheng Beach), and ε_2^9 (South Putuo Temple). Our method enables the boundaries of these areas to be clearly displayed.

Table 2 Primary attractions in the 30 clusters (partially listed)

Cluster	Primary attraction
ε_2^1	Kulangsu
ε_2^2	Zhongshan Street
ε_2^3	Xiamen Bridge
ε_2^4	Hulishan Fortress
ε_2^5	Zeng Cuo An
…	…

4.3. Frequent pattern mining

According to the results of adaptive spatial clustering, the 741 trajectories can be transformed into characteristic trajectories, and then the transition matrix is constructed. We then mine the frequent patterns according to Section 3.2. The number of frequent patterns depends heavily on the frequent pattern threshold (ξ). For example, when ξ=1E-04, 60 frequent patterns can be mined: $\varepsilon_2^1 \rightarrow \varepsilon_2^8$, $\varepsilon_2^8 \rightarrow \varepsilon_2^1$, $\varepsilon_2^9 \rightarrow \varepsilon_2^1$, …., $\varepsilon_2^{22} \rightarrow \varepsilon_2^{23}$, in which traditionally prevalent routes are included as well as routes that have not been noticed before.

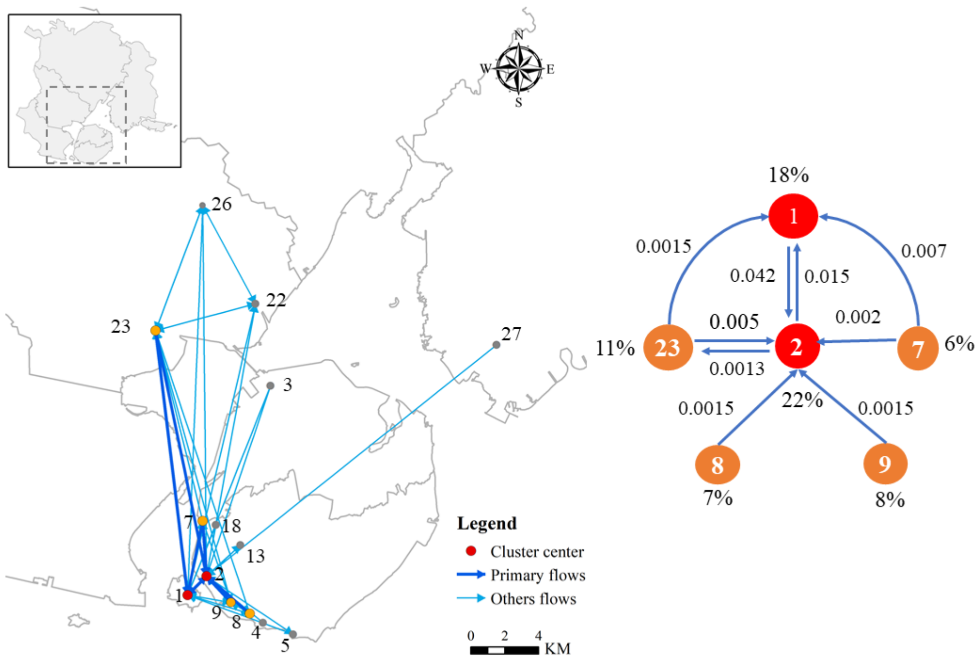

Fig. 7 Frequent patterns with length-2 sequences

Based on the 60 frequent patterns, Fig. 7(left) is drawn to show the directional frequent routes between clusters, in which a clear majority flows concentrate in the southwest area of the island. Among the clusters on the inland, only ε_2^{23} has a tight link with the island. In Fig. 7(right), the number next to the node indicates the proportion of the route passing through the node in all routes, the arrow lines delineate the direction of tourist movement and the number demonstrates the route intensity, reflecting the tourist volume. Furthermore, the main incoming clusters of the flows are ε_2^1 and ε_2^2. ε_2^1 (Kulangsu) is a uniquely appealing attraction as the World Heritage Site, and ε_2^2 (Zhongshan Road) is a well-known commercial block, with tourist pedestrian streets, shopping malls, and a large number of hotels and homestays around. These two areas have a strong appeal to all other areas, regardless of distance.

According to the result of our method, most of the sequences occur in a confined geographic area. Among all the attractions, popular areas of Xiamen are geographically concentrated (i.e., ε_2^1, ε_2^2, and ε_2^9). We can infer that the tourist distribution in Xiamen is greatly uneven. Massive tourists gather in these adjacent attractions, and only a minority of tourists visit other areas, indicating the existence of core-peripheral (Prideaux, 2002) or primary-secondary attractions in Xiamen (Park, et al., 2020).

5. Discussion

This study proposes an innovative method for mining and analyzing tourist mobile big data, generates great insights for understanding tourist movement patterns, and offers practical implications for intra-destination cooperation and optimize destination resource allocation.

5.1. Understanding tourist movement pattern

The study reveals several classic travel itineraries or movement patterns, including "ε_2^1 (Kulangsu) $-\varepsilon_2^2$ (Zhongshan Street)", "ε_2^9 (South Putuo Temple) $-\varepsilon_2^2$ (Zhongshan Street)", and so on. Furthermore, several travel itineraries that are not widely recognized are also found, like "ε_2^1 (Kulangsu) $-\varepsilon_2^{23}$ (Lingling International Circus City)" and "ε_2^2 (Zhongshan Street) $-\varepsilon_2^{23}$ (Lingling International Circus City)", and others. Understanding the tourist movement patterns is of great significance for destination planning, transport network and other service design and development (Lew & McKercher, 2006). In this study, we have generated insights both on the hierarchy of clusters and frequent patterns.

5.1.1. *Understanding tourist movement structure*

The tourist movement structure of Xiamen can be classified into four levels from the core to the peripheral areas: center zone, sub-center zone, sub-edge zone, and edge zone (Fig. 8). The closer to the core, the more tourists. In this study, the tourist movement structure was divided in a way that combined the results of spatial clustering with the results of frequent patterns. This approach is different from the traditional way of determining the hierarchy by focusing only on the number of tourists to attractions. Through the two-level spatial clustering, 30 clusters were derived to represent the distribution of tourists. Then, in the intensity of 1E-03, we counted the proportion of frequent patterns contained in each cluster separately to determine the level of traffic, and divide the tourist movement structure accordingly.

First, the center zone is composed of ε_2^1 and ε_2^2. Nearly one-fifth of the frequent patterns pass the two clusters (ε_2^1, 18%; ε_2^2, 22%), which are far more than other clusters. The two clusters have connections with lots of attractions and tourists diffuse to the surrounding areas and other more distant places from there. They can serve as a tourism distribution center, providing tourists with information consultation and selling tourism products. Except for these two clusters, the proportion of other areas in the frequent patterns is below 10%, so we use 5% as the boundary to divide the second and third hierarchy.

Second, the sub-center zone contains other clusters on primary flows (see the flows in Fig. 7). On one hand, these areas maintain close connections with the center zone; on the other hand, they also radiate tourists to other areas and are important starting and ending areas. In this zone, tourist service facilities can be deployed to meet the needs of tourists for staying, resting and shopping.

Third, the sub-edge zone includes clusters on other flows (also in Fig. 7). Taking in tourists from the upper two grades, this zone also plays an important role in the regional development of tourism.

Finally, the edge zone consists of the rest clusters, where the frequent patterns do not pass through, meaning few tourists choose to visit these areas. Tourism facilities need to be improved and marketing needs to be strengthened in this zone.

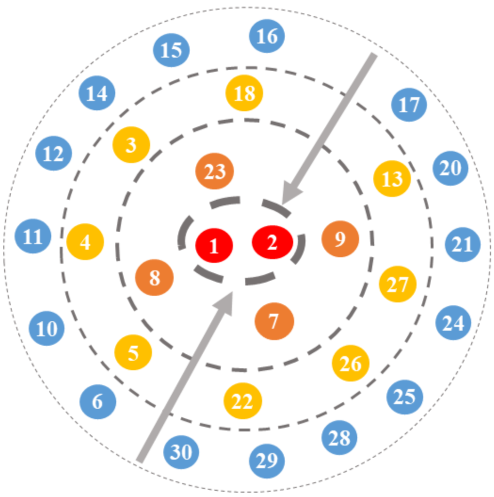

Fig. 8 Four hierarchies of tourist movement structure

5.1.2. *Understanding the frequent pattern hierarchy*

Our study not only enables an understanding of the hierarchical structure for each cluster but also the hierarchy of frequent patterns in different intensity thresholds. When ξ=1E-03, there are 15 frequent patterns, which must be the most active and crowded routes for tourists (as shown in Fig. 9). The excessive concentration of tourist flows is not conducive to the enhancement of the tourist experience. Therefore, tourism managers can disperse the tourist flows to frequent patterns of the next intensity level, such as the 60 frequent patterns when ξ=1E-04. We notice that the official tourism website of Xiamen (www.visitxiamen.com) only provides the attraction list and does not have any route suggestions. According to the results of our study, the official tourism website can show the routes with different levels of crowding based on the intensity thresholds to facilitate individualized choices. For example, when ξ=1E-03, there is one frequent pattern start from ε_2^6 (i.e., ε_2^6 to ε_2^2); when ξ=1E-04, two frequent patterns are added (i.e., ε_2^6 to ε_2^7, ε_2^6 to ε_2^{21}). To avoid overcrowding, tourists could consider choosing the strategy of going to ε_2^7, or ε_2^{21}, and then ε_2^2 at the end. With the support of our study, tourists will have a variety of options.

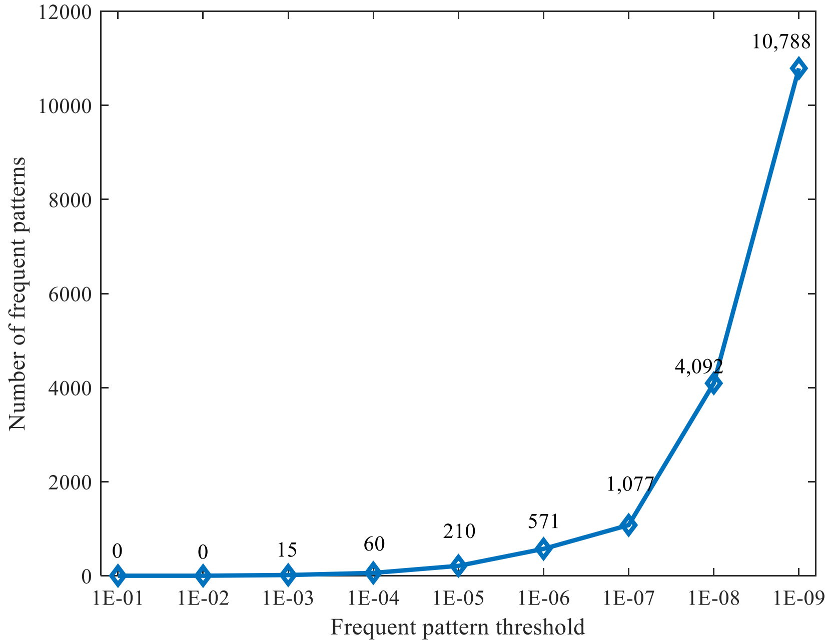

Fig. 9 The relationship between ξ and the number of frequent patterns

5.2. Facilitating intra-destination cooperation

Traditionally, individual attractions determine their partners based on the characteristics of tourism resources, the distance between attractions, or the power of administration, which are not the most effective approach. Frequent pattern mining of tourist trajectories, which provides an internal driving force of cooperation through mining the relationship among tourist attractions. The destination marketing department can design appropriate travel packages through collaboration between attractions to form a cluster, which provides greater efficiency and more satisfying experiences for the tourists (Wu & Carson, 2008). On the one hand, our findings reveal fresh opportunities for linking attractions for cooperation to form a new cluster. For example, Kulangsu (ε_2^1) and Hulishan Fortress (ε_2^4) are often regarded as a combination with good cooperation potential not only because of their close distance and complementary resources, but also because they both have high popularity among visitors to Xiamen (Kulangsu is a National 5A Scenery Sites, while Hulishan Fortress is a 4A). However, the relationship between these two attractions is not supported by the results of frequent pattern mining in Section 4.3. On the contrary, Kulangsu (ε_2^1) and Lingling International Circus City (ε_2^{23}), which are often ignored in the past, have a strong potential for cooperation because "Kulangsu – Lingling International Circus City" and "Lingling International Circus City – Kulangsu" are both frequent patterns. The cooperation is beneficial as it provides convenience to the tourists and enhances the attractiveness of the whole entity (Hwang, et al., 2006; Li, et al., 2008; Lue, et al., 1993; Oppermann, 1995; Yang, et al., 2013).

On the other hand, due to the complexity and dynamics of the destination, there was a lack of clear understanding of the connections between multiple attractions in the past (Caldeira & Kastenholz, 2018; Shao, et al., 2017). The frequent pattern mining has significant advantages over methods like social network analysis, which also investigates attraction links. The former is not limited to explore the relationship between two attractions and it can cover multi-attraction patterns. Taking the frequent pattern ε_2^1 to ε_2^2 as an example, based on our results, we can extend it to three-attraction patterns like ε_2^1 - ε_2^2 - ε_2^7, ε_2^1 - ε_2^2 - ε_2^{21}, or four-attraction patterns ε_2^1 - ε_2^2 - ε_2^7 - ε_2^6, ε_2^1 - ε_2^2 - ε_2^{21} - ε_2^{20}. By clarifying the complex relationships between these attractions, the attraction managers can know which attractions their tourists mainly come from and proactively seek to establish better partnerships with these preceding attractions. Besides, we also observe that some attractions have the advantage of synergistic development. For example, ε_2^{22}, ε_2^{23} and ε_2^{26} are attractions that have both geographic proximity and frequent patterns with each other. By examining their resource attributes, we believe the three attractions are holistically well suited for collaborating on the popular parent-child tours or study tours, since ε_2^{22} (Jimei School Village) helps children to feel the spirit of Tan Kah Kee (a famous, locally born, overseas Chinese philanthropist who funded the establishment of the school village and Xiamen University); ε_2^{23} (Lingling International Circus City) has high quality circus shows that appeal to children; and ε_2^{26} (The Ancient Courtyard Folk Culture Garden) demonstrates the cultural essence of Southern Fujian with rich regional features.

5.3. Optimizing destination tourism management

The study results can be used to optimize destination resource allocation. Considering the duration tourists spend in every attraction, our method can accurately discern every cluster and optimize its boundary, helping obtain more comprehensive tourist distribution information. The results generated from our method can be used to improve and optimize the layout of service facilities in tourism destinations. The difficulties and inconveniences in tourist movement are evident if we compare the existing public transportation system with tourist frequent patterns. Fig. 10 (left) displays the bus route map of Xiamen, together with the 30 clusters obtained from this study. We further obtain the following information of each frequent pattern through the Amap, when ξ=1E-04: the bus ride time, the transfer time and the walking distance. We then evaluate the convenience of public transportation for each frequent pattern. We found several patterns are not much convenient and chose four patterns for further examination (Fig. 10 (right)).

Among the four patterns, two patterns (ε_2^{2} to ε_2^{22}, ε_2^{2} to ε_2^{27}) need to transfer midway, and some of the walking distances of the other two patterns (ε_2^{2} to ε_2^{26}, ε_2^{22} to ε_2^{26}) are too long, for instance, the distance from the starting point to the bus stop. Obviously, as the core area of the destination, public transportation from ε_2^{2} to attractions outside the island is not convenient. This will hinder tourists from continuing to visit attractions outside the island. Accordingly, public transportation between the island and the mainland needs to be greatly optimized. Moreover, the important areas ε_2^{22} and ε_2^{26} are not effectively connected, thus the planning of tourist routes outside the island also needs to be improved.

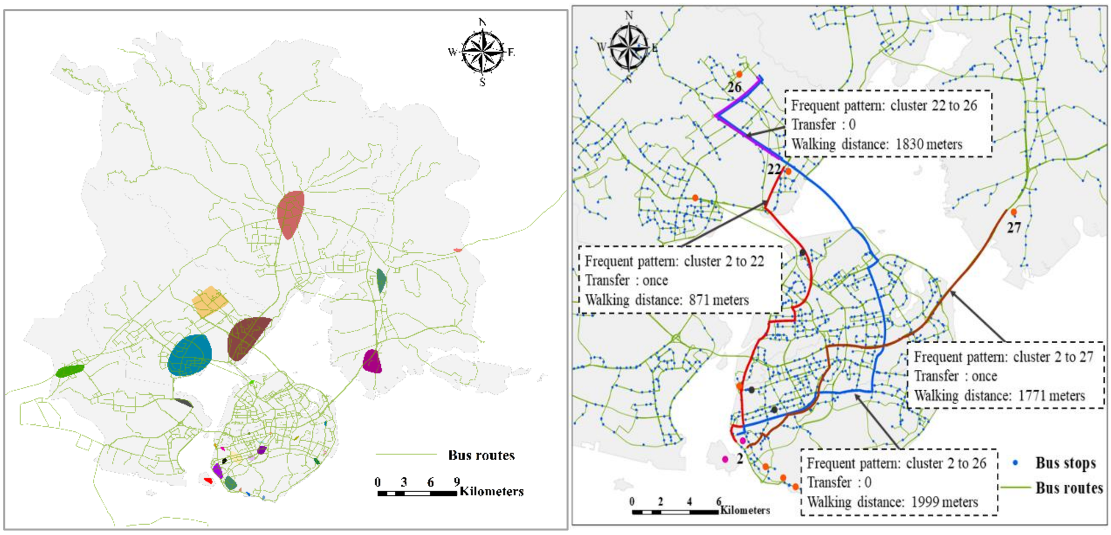

Fig. 10 Bus routes with 30 clusters (left) and four patterns with less convenient public transportation (right)

Our results also uncover the uneven tourism development in the destination. Four attractions outsides the island are included in the frequent patterns, but as far as we know, there are 16 national 3A and above attractions outside the island (China has five levels of rating for national tourist attractions, i.e, from 1A to 5A). This indicates that most of the tourist resources outside the island have not been effectively utilized. The destination planners should shift the development focus away from the mature attractions to these emerging attractions outside the island.

6. Conclusions

In today's digital era, mass spatio-temporal trajectory data can be recorded accurately and retrieved conveniently, which presents great opportunities for tourism research (Shoval, et al., 2020; Shoval, et al., 2018). This study attempts to advance the ways of leveraging such data to generate insights into tourist mobility. A two-stage heuristic approach involving adaptive DBSCAN, BrowseRank algorithm, convex hull algorithm, and DFS strategy is proposed to mine the frequent pattern of tourist trajectories. We illustrate the application of our proposed approach using mobile positioning data captured from real tourists and demonstrate the practical applications of the study findings.

This study advances the data analytics used for tourism mobility research (East, et al., 2017; McKercher, et al., 2015). Compared with the traditional two-stage algorithm (clustering first and then frequent pattern mining), our original contributions are: a) designing an adaptive spatial clustering approach based on DBSCAN to improve the performance of spatial clustering by introducing an adaptive parameter (δ). Our clustering method is designed to address the inadequacy of DBSCAN and adapted to the characteristics of tourism scenarios, where tourists are unevenly distributed in time and space; b) applying BrowseRank value (β) into two stages to fully capture the crucial spatio-temporal characteristics of tourist movements, including the sequence and direction of tourist movement among clusters for the spatial dimension, and the duration time in each cluster for the temporal dimension, which has rarely been considered in previous studies; and c) using the DFS strategy to improve performance and efficiency.

This study offers several practical implications for improving destination planning and management. Combining the results of frequent pattern mining in this study with tourist behavior characteristics, a more thorough understanding of tourist movement patterns in the destination can be obtained. The study results can be used for effective collaboration among tourist attractions to enhance their competitiveness as a collective entity. Moreover, the study results can be used to optimize the layout of service facilities, offering destination managers a proactive management ability. These implications can greatly improve our understanding of the clusters of well-integrated tourism areas and their basic components, and the interconnectedness of elements inside the clusters and between the clusters, and can be regarded as a great addition to the traditional study of tourist mobility. Especially, this study complements the traditional "core-periphery model" through the four hierarchies of tourist movement structure and provides insights into the multi-attraction pattern, which is prevalent in urban tourism today (Edwards, et al., 2008; Hunt & Crompton, 2008).

This study is mainly based on the mobile positioning data of tourists during the Spring Festival. The data is not much large-scale and could be updated due to fast development in the studied destination. Future studies may test the effectiveness of the method by using more diverse and multi-source data. As tourist preferences and movement patterns are influenced by their social demographics, future clustering and frequent pattern mining studies may consider social-demographic variables in the modeling approach. Moreover, with larger-scale data at finer granularities, it will be meaningful to compare the behaviors of tourists at different periods of the day, such as morning, afternoon, and evening, and have a more in-depth insight into the movement patterns of tourists during a day. Given the significant influence of transport infrastructure on tourist mobility (Shoval, et al., 2020; Yang, et al., 2019), future studies may explore the impact of metros on the frequent pattern of tourists. Two metro lines in Xiamen were opened in December 2017 and December 2019 respectively. Both lines extend from the island to the mainland, greatly facilitating tourist access to attractions outside the island. Whether there will be new attractions entering the market, and what changes these new attractions will bring to the destination's spatial structure deserve further research.

References

Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management, 29*, 469-486.

Barber, C. B., Dobkin, D. P., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS), 22*, 469-483.

Birenboim, A., & Shoval, N. (2016). Mobility research in the age of the smartphone. *Annals of the American Association of Geographers, 106*, 283-291.

Buhalis, D. (2000). Marketing the competitive destination of the future. *Tourism Management, 21*, 97-116.

Caldeira, A. M., & Kastenholz, E. (2018). Tourists' spatial behaviour in urban destinations: The effect of prior destination experience. *Journal of Vacation Marketing, 24*, 247-260.

Cao, H., Mamoulis, N., & Cheung, D. W. (2005). Mining frequent spatio-temporal sequential patterns. In *Fifth IEEE international conference on data mining*. Houston, TX, USA.

Chancellor, C., Yu, C. P. S., & Cole, S. T. (2011). Exploring quality of life perceptions in rural midwestern (USA) communities: An application of the core–periphery concept in a tourism development context. *International Journal of Tourism Research, 13*, 496-507.

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies, 68*, 285-299.

Crawford, F., Watling, D. P., & Connors, R. D. (2018). Identifying road user classes based on repeated trip behaviour using Bluetooth data. *Transportation Research Part A-Policy and Practice, 113*, 55-74.

Dejbakhsh, S., Arrowsmith, C., & Jackson, M. (2011). Cultural influence on spatial behaviour. *Tourism Geographies, 13*, 91-111.

Delafontaine, M., Versichele, M., Neutens, T., & Van de Weghe, N. (2012). Analysing spatiotemporal sequences in Bluetooth tracking data. *Applied Geography, 34*, 659-668.

Delmelle, E. C. (2016). Mapping the DNA of Urban Neighborhoods: Clustering Longitudinal Sequences of Neighborhood Socioeconomic Change. *Annals of the American Association of Geographers, 106*, 36-56.

Dharmowijoyo, D. B. E., Susilo, Y. O., & Karlstrom, A. (2017). Analysing the complexity of day-to-day individual activity-travel patterns using a multidimensional sequence alignment model: A case study in the Bandung Metropolitan Area, Indonesia. *Journal of Transport Geography, 64*, 1-12.

Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics, 4*, 95-104.

East, D., Osborne, P., Kemp, S., & Woodfine, T. (2017). Combining GPS & survey data improves understanding of visitor behaviour. *Tourism Management, 61*, 307-320.

Edwards, D., Griffin, T., & Hayllar, B. (2008). Urban tourism research: Developing an agenda.

Annals of Tourism Research, 35, 1032-1052.

Ellegård, K., & Svedin, U. (2012). Torsten Hägerstrand's time-geography as the cradle of the activity approach in transport geography. *Journal of Transport Geography, 23*, 17-25.

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., Foufou, S., & Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE Transactions on Emerging Topics in Computing, 2*, 267-279.

Giannotti, F., Nanni, M., Pinelli, F., & Pedreschi, D. (2007). Trajectory pattern mining. In *The 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Graham, R. L. (1972). An efficient algorith for determining the convex hull of a finite planar set. *Information Processing Letters, 1*, 132-133.

Grinberger, A. Y., & Shoval, N. (2019). Spatiotemporal contingencies in tourists' intradiurnal mobility patterns. *Journal of Travel Research, 58*, 512-530.

Han, J., Li, Z., & Tang, L. A. (2010). Mining moving object, trajectory and traffic data. In H. Kitagawa, Y. Ishikawa & Q. Li (Eds.), *Database Systems for Advanced Applications* (pp. 137-155). London: Routledge.

Huang, Q., Li, Z., Jing, L., & Chang, C. (2016). Mining frequent trajectory patterns from online footprints. In *The 7th ACM SIGSPATIAL international workshop on GeoStreaming*.

Hwang, Y.-H., Gretzel, U., & Fesenmaier, D. R. (2006). Multicity trip patterns: Tourists to the United States. *Annals of Tourism Research, 33*, 1057-1078.

Kalnis, P., Mamoulis, N., & Bakiras, S. (2005). On discovering moving clusters in spatio-temporal data. In C. Bauzer Medeiros, M. J. Egenhofer & E. Bertino (Eds.), *Advances in Spatial and Temporal Database* (Vol. 3633). Berlin: Springer.

Kang, J.-Y., & Yong, H.-S. (2010). Mining spatio-temporal patterns in trajectory data. *Journal of Information Processing Systems, 6*, 521-536.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy, 99*, 483-499.

Lee, H.-J., & Joh, C.-H. (2010). Tourism Behaviour in Seoul: An Analysis of Tourism Activity Sequence using Multidimensional Sequence Alignments. *Tourism Geographies, 12*, 487-504.

Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of Tourism Research, 33*, 403-423.

Li, X., Meng, F., & Uysal, M. (2008). Spatial pattern of tourist flows among the Asia-Pacific countries: An examination over a decade. *Asia Pacific Journal of Tourism Research, 13*, 229-243.

Liu, Y., Gao, B., Liu, T.-Y., Zhang, Y., Ma, Z., He, S., & Li, H. (2008). BrowseRank: Letting web users vote for page importance. In *the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Vol. , pp. 451-458).

Liu, Y. C., Li, Z. M., Xiong, H., Gao, X. D., Wu, J. J., & Wu, S. (2013). Understanding and enhancement of internal clustering validation measures. *IEEE Transactions on*

Cybernetics, 43, 982-994.

Long, J. A., & Nelson, T. A. (2013). A review of quantitative methods for movement data. *International Journal of Geographical Information Science, 27*, 292-318.

Lue, C.-C., Crompton, J. L., & Fesenmaier, D. R. (1993). Conceptualization of multi-destination pleasure trips. *Annals of Tourism Research, 20*, 289-301.

McCormack, D. P., & Schwanen, T. (2011). Guest editorial: The space-times of decision making. *Environment and Planning A: Economy and Space, 43*, 2801-2818.

McKercher, B., & Lau, G. (2008). Movement patterns of tourists within a destination. *Tourism Geographies, 10*, 355-374.

McKercher, B., Shoval, N., Ng, E., & Birenboim, A. (2012). First and repeat visitor behaviour: GPS tracking and GIS analysis in Hong Kong. *Tourism Geographies, 14*, 147-161.

McKercher, B., Shoval, N., Park, E., & Kahani, A. (2015). The [Limited] impact of weather on tourist behavior in an urban destination. *Journal of Travel Research, 54*, 442-455.

Millward, H., Hafezi, M. H., & Daisy, N. S. (2019). Activity travel of population segments grouped by daily time-use: GPS tracking in Halifax, Canada. *Travel Behaviour and Society, 16*, 161-170.

Neutens, T., Schwanen, T., & Witlox, F. (2011). The prism of everyday life: Towards a new research agenda for time geography. *Transport Reviews, 31*, 25-47.

Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2016). Tracing tourists by their digital footprints: The case of Austria. *Journal of Travel Research, 55*, 566-573.

Oppermann, M. (1995). A model of travel itineraries. *Journal of Travel Research, 33*, 57-61.

Park, S., Xu, Y., Jiang, L., Chen, Z., & Huang, S. (2020). Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data. *Annals of Tourism Research, 84*, 102973.

Prideaux, B. (2002). Building visitor attractions in peripheral areas-Can uniqueness overcome isolation to produce viability? *International Journal of Tourism Research, 4*, 379-389.

Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Enviroment Planning B: Planning Design, 33*, 727-748.

Riganti, P., & Nijkamp, P. (2008). Congestion in popular tourist areas: a multi-attribute experimental choice analysis of willingness-to-wait in Amsterdam. *Tourism Economics, 14*, 25-44.

Shao, H., Zhang, Y., & Li, W. (2017). Extraction and analysis of city's tourism districts based on social media data. *Computers Environment and Urban Systems, 65*, 66-78.

Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies, 18*, 587-606.

Shoval, N., & Isaacson, M. (2007). Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of American Geographers, 97*, 282-297.

Shoval, N., Kahani, A., De Cantis, S., & Ferrante, M. (2020). Impact of incentives on tourist activity in space-time. *Annals of Tourism Research, 80*, 102846.

Shoval, N., McKercher, B., Birenboim, A., & Ng, E. (2015). The application of a sequence alignment method to the creation of typologies of tourist activity in time and space. *Environment and Planning B: Planning and Design, 42*, 76-94.

Shoval, N., Schvimer, Y., & Tamir, M. (2018). Real-time measurement of tourists' objective and subjective emotions in time and space. *Journal of Travel Research, 57*, 3-16.

Stehle, S., & Peuquet, D. J. (2015). Analyzing spatio-temporal patterns and their evolution via sequence alignment. *Spatial Cognition & Computation, 15*, 68-85.

Tsoukatos, I., & Gunopulos, D. (2001). Efficient mining of spatiotemporal patterns. In C. S. Jensen, M. Schneider, B. Seeger & V. J. Tsotras (Eds.), *Advances in spatial and temporal databases* (Vol. 2121). Berlin: Springer.

Wang, L., Hu, K., Ku, T., & Yan, X. (2013). Mining frequent trajectory pattern based on vague space partition. *Knowledge-Based Systems, 50*, 100-111.

Wang, T., Zhang, D., Zhou, X., Qi, X., Ni, H., Wang, H., & Zhou, G. (2016). Mining personal frequent routes via road corner detection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 46*, 445-458.

Weaver, D., Moyle, B. D., & McLennan, C.-l. (2021). A core/periphery perspective on mega-event sustainability: Dystopic and utopic scenarios. *Tourism Management, 86*, 104340.

Wilson, C. (2008). Activity patterns in space and time: Calculating representative Hagerstrand trajectories. *Transportation, 35*, 485-499.

Wilson, W. C. (1998). Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A, 30*, 1017-1038.

Wu, C. L., & Carson, D. (2008). Spatial and temporal tourist dispersal analysis in multiple destination travel. *Journal of Travel Research, 46*, 311-317.

Yang, Y., Fik, T., & Zhang, J. (2013). Modeling sequential tourist flows: where is the next destination? *Annals of Tourism Research, 43*, 297-320.

Yang, Y., Li, D., & Li, X. (2019). Public transport connectivity and intercity tourist flows. *Journal of Travel Research, 58*, 25-41.

Yang, Y., & Wong, K. K. (2012). A spatial econometric approach to model spillover effects in tourism flows. *Journal of Travel Research, 51*, 768-778.

Yang, Y., & Wong, K. K. (2013). Spatial distribution of tourist flows to China's cities. *Tourism Geographies, 15*, 338-363.

Yin, P., Lin, Z., & Prideaux, B. (2019). The impact of high-speed railway on tourism spatial structures between two adjoining metropolitan cities in China: Beijing and Tianjin. *Journal of Transport Geography, 80*, 102495.

Zhao, X., Lu, X., Liu, Y., Lin, J., & An, J. (2018). Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi'an, China. *Tourism Management, 69*, 368-383.

Zheng, Y. (2015). Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology, 6*, 1-41.