Information Technology & Tourism
https://doi.org/10.1007/s40558-025-00325-3

ORIGINAL RESEARCH

Exploring hot spots at tourism destination by flow-based density clustering method

Yingqi Yuan[1] · Sangwon Park[2] · Yang Xu[3] · Sun-Young Koh[4]

Received: 1 July 2024 / Revised: 24 April 2025 / Accepted: 2 June 2025

Abstract
Advancements in mobile technology allows tourism researchers to access fine-grained location data reflecting the travelers' flow at destinations. While travel flow data contains inclusive information about travelers' mobility—including origins (i.e., where they depart) and destinations (i.e., where they visited)—existing studies grounded in central place theory focus predominantly on destination to identify tourism 'hot spots'. This approach overlooks the dynamic spatial interactions, which provides limited understanding about travel mobility. Thus, this study aims (1) to propose an origin-to-destination flow-based density clustering (OD-FDC) algorithm taking into account the directional travel movement derived from central flow theory, and (2) to demonstrate the usefulness of the method by analyzing over 150,000 car navigation records from Jeju, South Korea. This study clearly delineates how the flow-based method, OD-FDC algorithm, models directional travel movement and identifies tourism hotspots by fully utilizing flow directions, intensities and spatial distributions. Results reveal that the OD-FDC algorithm outperforms point-based analysis method—traditional hotspot analysis (Getis-Ord Gi*) by uncovering dynamic spatial interactions. As a result, this research provides theoretical contributions to the literature on travel mobility and methodological implications in spatial analytic of flow data. The findings provide destination marketers with actionable insights in developing regional planning and marketing.

Keywords Tourism big data · Travel mobility · Travel flow data · Tourism hotspots detection · Origin-to-destination flow-based clustering

Extended author information available on the last page of the article

Published online: 25 June 2025

1 Introduction

The evolution of information technology has reshaped the entire structure of the tourism industry in communication, information searching behaviors, and decision-making processes (Hwang et al. 2006). Mobile technology essentially affects how travelers search for information and change their planned behaviors due to its characteristics, such as portability and context-oriented services (Wang et al. 2012). This feature provides an innovative approach in obtaining fine-grained information about travel mobility, which enables tourism researchers to understand tourists' behaviors in a detailed and accurate manner (Hardy et al. 2017; Wang et al. 2018). The mechanism of mobile technology that equips location awareness systems constantly communicates with the intimate phone towers and/or satellites by transmitting sensors. Sensor data include spatial and temporal information, respectively representing where and when travelers (or mobile users) visited the place within the coverage of a certain phone tower in a sequential manner.

The existing literature on travel flow and/or movement patterns has explored the various formats of data for tourist mobility, such as travel diary, social media, mobile sensor, and Global Positioning System (GPS) data. One of the key tasks in previous studies was to identify tourism hotspots (THs). To do so, researchers have mostly used point-based clustering methods, including K-means and density-based spatial clustering of application with noise (DBSCAN; Wong et al. 2017; Park et al. 2020), as well as hotspot spatial analysis through a GIS tool (e.g., ArcGIS software; Van der Zee et al. 2020). Previous studies have assumed that THs generally refer to a higher concentration of visitations by travelers compared to a random distribution of the visits (or events) (Park et al. 2020). In other words, the detection of THs has focused on the study of point distributions or spatial arrangements of points in a space (Chakravorty 1995) by examining the point patterns (or the density of points visited places within a certain destination).

As opposed to other location data representing hotel, park, and museum locations, mobile sensor data consist of origin from which a traveler departs and destination to which the individual visits. The directions of travel flow during the specific period of time, labeled origin and destination (OD) data, are also included. Detecting spatiotemporal clusters in OD flow data is a critical issue in assisting decision-making support to applications including destination planning and management as well as intelligent recommendation. Importantly, however, the existing studies applying the point-based clustering method have failed to accommodate the benefits of travel flow data comprising the OD information of movements. Recognizing that the locations that travelers visit involve their interests and preferences (Li et al. 2008), the directional patterns of travel flow should be considered to discover the spatial clustering based on interconnected travel activities. Given the key feature of mobility data, THs must be discovered in terms of spatial (i.e., where travelers visit) and temporal (i.e., when travelers visit) dimensions. This argument is consistent to a concept of travel restrictions that embody spatiotemporal constraints, suggesting that traveler behaviors occur with limited spatial and temporal resources (Cooper 1981). That is, limited resources of spatiotemporal constraints form the pattern of information searching and travel behaviors (Kang 2016). Likewise, the graph theory highlights the value of understanding directional *links* that connect two places travelers visited (Park et al. 2021) with weight indicating volume or density. The graph theory facilitates presenting the complex travel mobility using vertices (places visited by travelers) and directed edges (or links; directions of travel flow) based on statistical theory. The links reflect the spatial interactions between destinations formed by travel flow, which is important to define THs.

Therefore, this research aims to propose an innovative method for identifying THs by leveraging comprehensive travel flow information, which encompasses not only the points of departure and destinations but also the volume and direction of travel. The proposed approach integrates two key theories in travel mobility: graph theory, which emphasizes the importance of spatial links through travel flow, and cumulative attraction theory, which focuses on grouping spaces with multiple attractions to enhance their overall appeal. This integration addresses the limitations of existing spatial clustering methods primarily focusing on where travel visited (point-based approach) (e.g., K-means and DBSCAN algorithms). Consequently, the findings of this study provide valuable implications for the literature on tourism big data as well as for destination planning and management.

2 Literature review

2.1 Tourists' spatial behavior

Spatial behavior refers to a sequence of attractions visited by tourists within a geographic space in the tourism context (Xia 2007). Investigating spatial behavior movement provides fruitful insight for managing tourism destinations effectively (Bauder and Freytag 2015). Relevant studies have been conducted from two perspectives: (1) a cognitive approach that examines the objective decision-making process and (2) a behavioral approach that focuses on actual movements. Cognitive perspectives define tourist movements as an objective behavior resulting from a subjective decision-making process (Lloyd 1997; Tussyadiah and Zach 2012), influenced by factors such as perceptions (Tussyadiah and Zach 2012), motivations (Crompton 2025), and preferences (Hardt et al. 2024). On the contrary, research understanding physical movements in a behavioral perspective perceives *spatial behavior* as either travelers' movements among tourist attractions within a single geographic space (Xia 2007). Both approaches provide critical insights to destination management organizations (DMOs) for planning, developing, and marketing destinations efficiently and effectively.

While cognitive approaches dominated early research, behavioral methods have gained attentions in the recent literature due to advancements in geospatial big data. Historically, behavioral studies were limited by data scarcity, but advanced online sharing services and mobile technologies provide abundant geotagged data from social media platforms (e.g., Flickr and Twitter), enabling detailed analysis of tourists' spatial behaviors (Wong et al. 2017). For example, Vu et al. (2018) analyzed geotagged photographs (i.e., Flickr) to capture Australian outbound travelers' behaviors and preferences using sequential rule mining method, thereby assisting DMOs in destination marketing. Similarly, Cheng et al. (2023) using four sequential pattern mining algorithms to Twitter data, uncovering interactional visitors' movement patterns during a mega event, which informed DMOs' destination management strategies.

However, some researchers have argued that social media data containing geotagged information have limitations in describing tourists' movements: (1) data sparsity, which indicates that tourists selectively share experiences on social media; (2) data scarcity, which indicates that limited number of geotagged data are available online; and (3) potential selection bias, given that only few travelers post on social media during trips (Wong et al. 2017; Park et al. 2020). Thanks to the development of tracking devices, tourism researchers of behavioral science facilitate to collect more accurate geographic information and minimize the potential issue of data bias (Raun et al. 2016; Shoval and Isaacson 2007). For example, mobile sensor and GPS data record spatial and temporal information accurately, minimizing possible bias and data loss (Xu et al. 2021). This advancement allows tourism scholars to better understand tourists' spatial behaviors through behavioral approach. Thus, this research examines travel movement patterns using behavioral methods and proposes spatial analytics of travel flows, contributing to studies that investigate these dynamics.

2.2 Tourism hotspot identification from tourists' spatial behaviors

One primary application of behavioral approaches in tourism studies is tourism hotspot identification. These methods enable tourisms scholars to decipher collective location preferences that emerge from tourists' actual spatial behaviors. Existing literature has primarily focused on identifying and characterizing THs through static geographic proximities using geospatial big data. For example, Park et al. (2020) identified THs (popular areas) in three South Korea cities (Jeonju, Gangneung and Chuncheon) by analyzing mobile phone data with DBSCAN clustering algorithm and SPADE algorithm for spatial pattern mining. Similarly, Zheng et al. (2022) revealed frequently visited areas in Xiamen, China, using adaptive spatial clustering (e.g., DBSCAN algorithm) and frequent patterns mining methods (e.g., Depth First Search algorithm) with mobile positioning data. They highlighted clustering multiple destinations based on place-centric theories based upon cumulative attraction theory and attraction compatibility theory, which evaluate how proximity between business influences tourists' spatial behaviors.

Nelson (1959) discussed the cumulative attraction theory in the context of retail, and Hunt and Crompton (2008) developed the concept into a local level of tourism destinations. Cumulative attraction theory posits that clustering similar attractions creates advantages like diversity and economies of scale, while attraction compatibility theory highlights how complementary attractions boost customer inflow (Nelson 1959; Crompton and Gitelson 1979). Both theories imply that clustering destinations as a group of multiple attractions increase the place's popularity. Although different scholars have stated different perspectives on understanding the interactions between destinations, they mutually agree that (1) people prefer visiting multiple attractions at a destination; (2) clustering multiple attractions as a group enhance the popularity of the destination; and (3) understanding attraction or destination interaction provides practical insights on planning and developing tourism destination and helpful for making destination marketing decisions. However, these theories focus on static geographic proximities, overlooking dynamic spatial interactions between places.

Addressing this gap, many behavioral studies have linked the spatial behavior of individuals traveling between different places as sequential events (Pettersson and Zillinger 2011). Applying these human movements to the theory from a tourism perspective provides fruitful insights for establishing tourism policies or designing tourist destinations (Hall 2005). Studies related to human movement are highly rooted in graph theory, which is a mathematical structure described as the set of vertices and edges (a link connecting a pair of nodes; Hayes 2000). Graph theory is developed in network science applied to other disciplines (e.g., transportation and geography) to quantify the pattern and structure of human mobility (Dale and Fortin 2010). Likewise, it has been considered to understand travel movement behaviors, given that tourists visit multiple destinations within their travel itinerary rather than a single destination. In other words, multiple locations and pathways exist in tourists' movement behavior, forming a complex network structure (Sainaghi and Baggio 2017). Relevant studies suggest complexity, interaction, and diversity of travel mobility (Strogatz 2001). Baggio (2017) measured the dynamic structure of tourism destinations and mapped a turning point to explore the evolution of mobility behavior patterns. Park et al. (2021) applied graph theory to identify the similarity of complex travel mobility network patterns. Recent studies applying network science approaches analysis demonstrated the disparities in travel behaviors with regard to temporal dimensions such as length of stays and when travel activity takes place (Huang et al. 2024; Park and Zhong 2022) and spatial dimensions such as different types of destinations (Park et al., 2022), deepening our understanding of the complexity in tourists' travel mobility patterns associated to dynamic spatial interactions between tourism attractions or places visited.

Consequently, this study advocates for the integration of flow-centric theories like graph theory with place-centric theories such as cumulative attraction theory and attraction compatibility theory, proposing an innovative approach to tourism spatial analytics. Specifically, it highlights the importance of spatial links through travel flows (i.e., graph theory) and the grouping of spaces that include multiple attractions (i.e., cumulative attraction theory and attraction compatibility theory) to more effectively identify well-represented THs at a destination. The details of the spatial data analysis in this context are discussed in the next section.

2.3 Tourism hotspot analysis

Relevant studies have proposed several methods to identify popular attractions and destinations, which can be classified into: (1) spatial statistics-based methods, such as Getis-Ord Gi* statistics (Van der Zee et al. 2020), (2) clustering-based algorithms in machine learning, including hierarchical clustering (e.g., Nearest neighbor hierarchical clustering algorithm; Su et al. 2020), partitioning clustering (e.g., K-means clustering algorithm; Salas-Olmedo et al. 2018), and density-based clustering methods (e.g., OPTICS and DBSCAN algorithms; Rodríguez-Echeverría et al. 2020), and (3) unsupervised deep learning algorithm, such as Self-Organizing Maps (Provenzano and Giambrone 2023). These studies not only detected tourist hotspots but also analyzed their spatio-temporal characteristics, with the majority utilizing geotagged data or mobile sensor data.

When reviewing the relevant literature, most studies have employed point-based analysis when detecting THs, treating each origin or destination as a unit but neglecting overall structure (e.g., direction) and the features of flow data (e.g., density). This suggests that existing clustering methods often rely on partial information when identifying hotspots, primarily focusing on visited places (or events). To address these limitations, this paper advocates for a theoretical shift from central place theory to central flow theory (Taylor et al. 2010). Central place theory explains how certain tourist spots within a destination gain popularity due to their proximity to populated areas, aligning with the law of distance decay. Conversely, central flow theory shifts the focus from static points to dynamic flows, highlighting that some places attract more tourists due to spatial interactions.

Based on the graph theory (Yang et al. 2010), considering the spatial interactions (or links) between the OD of travel flow is essential to comprehensively understand travel mobility and theoretically suggest spatial clustering (or THs). THs can be defined as places connected by a high density of travel flow, encompassing both spatial and temporal dimensions (Schneider et al. 2013). In this vein, Zhou and Chen (2023) used community detection algorithm in network analysis to cluster attractions in Hong Kong for effective marketing strategies, which primarily considers volume attributes of OD flow but overlooks the spatial distributions of these flows. Therefore, this study proposes an innovative flow-based density clustering method grounded in central flow theory, which leverages raw OD flow information, such as direction, intensity and spatial distribution of OD flow, to holistically identify THs.

3 Methodology

3.1 Data collection

This research relies on a car navigation dataset collected in Jeju Island, South Korea (hereafter Jeju). This dataset is provided by TMAP, one of the largest telecommunication companies in South Korea that offers navigation services to mobile application users. Travelers are defined as individuals whose residential address is outside Jeju and used TMAP mobile application for navigation services while traveling in Jeju.

Jeju is located in the south part of South Korea (see Fig. 1). Well-known for its nature-based tourism destinations, it is one of the most popular vacation spots in South Korea. When visiting Jeju, tourists can only choose from three travel modes: buses, taxis, and rental cars. However, given the poor accessibility and limited level of service of public transportation, most tourists usually choose to travel by rental cars in Jeju (Kim et al. 2021). Therefore, this research uses a car navigation dataset to study major travel patterns in Jeju.

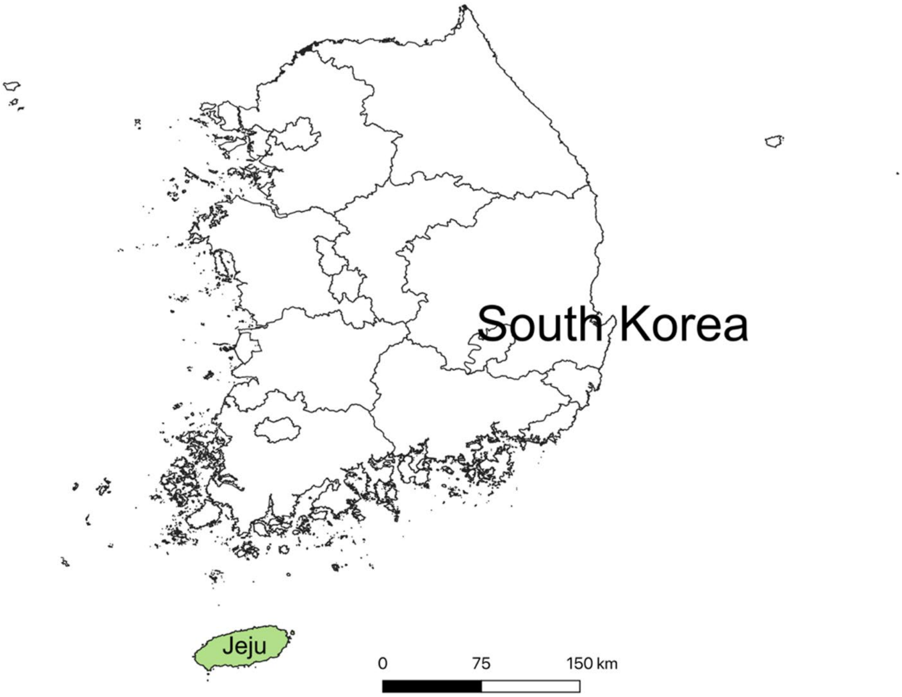

Fig. 1 Location of Jeju in South Korea.
(Source: http://www.gisdeveloper.co.kr/?p=2332)

Table 1 Example of road trip records in the navigation dataset

Index	Travel Date	O_{yi}	O_{xi}	D_{yi}	D_{xi}	NRT
0	2019-06-01	33.***	126.***	33.***	126.***	1
1	2019-06-01	33.***	126.***	33.***	126.***	50
2	2019-06-01	33.***	126.***	33.***	126.***	29
…						
154, 681	2019-06-30	33.***	126.***	33.***	126.***	1

(Note: $(O_{xi},\ O_{yi})$ and $(D_{xi},\ D_{yi})$ refer to the coordinates (xi=longitude, yi=latitude) of the initial and final locations of OD flow; NRT refers to number of road trips.)

3.2 Data description

This anonymized dataset tracks 154,681 road trips of tourists via TMAP mobile application during driving trips in Jeju during the period of June 1–30, 2019. Table 1 shows an example of a road trip record in the navigation dataset. Each record tracks the index of each road trip; travel date (the day when tourists had road trips in Jeju); the location (longitude: xi and latitude: yj) of start points (origins: $\boldsymbol{O}$) and end points (destinations: D; at 1×1 km grid cell level); and the number of road trips (NRT), representing the number of tourists traveling on the same date from the same origins to the same destinations. Notably, this dataset captures the actual point-to-point (origin-to-destination) travel behaviors by using rental cars without intermedi-

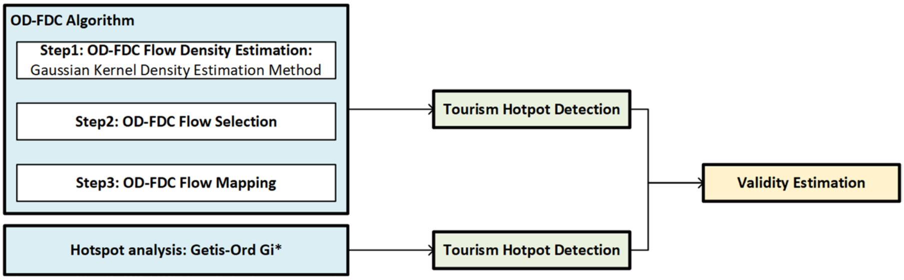

Fig. 2 Flow chart of data analysis in this research

ate stops, enabling precise analysis of OD flow dynamics across Jeju. To improve data validity, the actual movement of a TMAP application user who searches for a destination and then travels at least 100 m for more than 1 min by car is estimated.

3.3 Data analysis

This research aims to apply the origin-to-destination flow-based density clustering (OD-FDC) algorithm, which comprises three steps – OD flow density estimation, OD flow selection, and OD flow mapping. These steps are used to characterize multiscale OD flow patterns, taking into account both the attributes of OD flows (i.e., travel frequency) and their spatial attributes (e.g., spatial locations), thereby enabling to identify THs. Its effectiveness of TH detection is evaluated by comparing it to a traditional hotspot analysis approach (Getis-Ord Gi*). A series of data analytics implemented in this research are summarized in Fig. 2. The OD-FDC algorithm and hotspot analysis will be explained in detail the subsequent section.

3.3.1 OD-FDC algorithm

Unlike flow-clustering algorithms, which treat each OD flow as a unit and group similar OD flows into different clusters (Tang et al. 2021; Zhang et al. 2023), the OD-FDC algorithm is a data-driven clustering method used to discover and visualize the multiscale OD flow pattern by setting varying parameters, bandwidth (smoothing parameter) h, and neighborhood radius R. The OD-FDC algorithm consists of three steps (Zhu et al. 2019):

(1) OD flow density estimation: This step estimates density values (controlled by bandwidth h) of the OD flow within its OD flow neighborhood (controlled by neighborhood radius R).

(2) OD flow selection: This involves choosing the subset of OD flow with local maximal density (controlled by neighborhood radius R) on the basis of the result of OD flow density estimation; and

(3) OD flow mapping: The final step visualizes OD flow with local maximal density on a map.

Each of these steps is detailed in the following section, illustrated in Fig. 3.

Step1: OD flow density estimation.

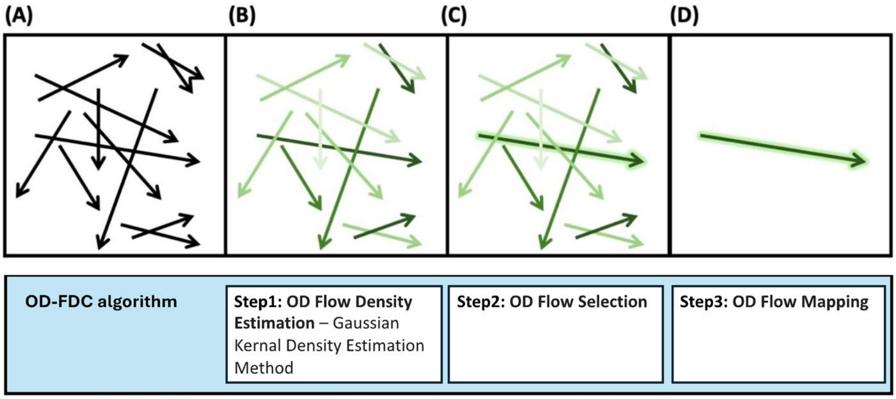

Fig. 3 Illustration of OD-FDC algorithm. (**A**) raw OD flow data (directional lines from origins to destinations with travel frequency (NRT)); (**B**) Step1: OD flow density estimation; (**C**) Step2: OD flow selection; (**D**) Step3: OD flow mapping

KDE is an effective tool for exploring data patterns (Chen 2017) and visualizing event clusters across a study area (Cai et al. 2013). Thus, this research estimates the density of road trips by applying Gaussian KDE (GKDE) method based on the following definitions.

OD flow *OD flow is one of geographic trajectory data with OD locations (coordinates). A record of road trip in the dataset can be represented as a directional OD flow, which can be expressed as follows.*

$$T_i = (O_{xi}, O_{yi}, D_{xi}, D_{yi}) \tag{1}$$

where i denotes the ith record; and (O_{xi}, O_{yi})(x=longitude, y=latitude) and (D_{xi}, D_{yi}) denote the locations of OD flow, respectively.

OD flow neighborhood *The neighborhood of OD flow* T_i(T_i*'s* R*-size neighborhood) is a set of OD flows whose distance to OD flow* T_i *is less than a given radius of the OD flow neighborhood (neighborhood radius)* R*, which can be expressed as follows*

$$N(T_i, d) = \{T_j \in T | dist(T_i, T_j) < R\} \tag{2}$$

where $N(T_i, d)$ is the neighborhood of OD flow T_i; OD flow T_j is one of the OD flows in the neighborhood of OD flow T_i; and $dist(T_i, T_j)$ is the distance between T_i and T_j.

In this case, Euclidean distance is applied to calculate the distance of any pair of OD flows. The Euclidean distance between a pair of OD flows (T_i and T_j) is defined as

$$dist\,(T_i,\,T_j) = \sqrt{\left(O_{xi} - O_{xj}\right)^2 + \left(O_{yi} - O_{yj}\right)^2 + \left(D_{xi} - D_{xj}\right)^2 + \left(D_{yi} - D_{yj}\right)^2} = \sqrt{{d_O}^2 + {d_D}^2} \quad (3)$$

where d_O is the Euclidean distance between two origins, and d_D is the Euclidean distance between two destinations.

Based on the aforementioned definition, the GKDE for OD flow density estimation in this research is (Zhu et al. 2019):

$$\widehat{f}\,(x) = \frac{1}{nh}\sum\nolimits_{i=1}^{n} K\left(\frac{dist\,(T - T_i)}{h}\right) Y\,(T_i) \quad (4)$$

where $T_1,\,T_2,\ldots,\,T_n$ denote the OD flows; $dist\,(T,\,T_i)$ is the distance between OD flow T_i and other T within the neighborhood of OD flow T_i; h is the bandwidth (smoothing bandwidth); n is the number of OD flows; $K\left(\frac{dist(T-T_i)}{h}\right)$ is the Gaussian kernel density function; and $Y(T_i)$ is the tourist volume of OD flow T_i(NRT or travel frequency), representing the weight in the OD flow density estimation.

Smoothing parameter h in KDE has a significant influence on the result of density estimation. An extremely small bandwidth h may cause undersmoothed estimation, whereas an extremely large bandwidth h may cause oversmoothed estimation. Thus, an optimal bandwidth should be identified to avoid such problems. Silverman's rule of thumb (SROT) is the most widely used and well-known algorithm to generate an optimal bandwidth (Harpole et al. 2014). Among various methods for optimal bandwidth calculation, SROT is the first choice to deal with many cases due to its robustness in obtaining a reasonable result regardless of observed data distribution (normal or nonnormal distribution; Bashtannyk and Hyndman 2001). SROT is also a data-driven method used to avoid arbitrary and subjective bandwidth choices. Thus, it has been widely applied in spatial data analysis (Nakaya and Yano 2010). SROT suggests that the optimal bandwidth h is (Silverman 1986)

$$h = \left(\frac{4\sigma^{\,5}}{3n}\right)^{\frac{1}{5}} \quad (5)$$

where σ is the standard deviation of the observed data, and n is the number of observed data.

In this case, the standard deviation σ, which quantifies the amount of variation or dispersion in the dataset (OD flows) to select the optimal bandwidth h, is defined as

$$\sigma = \sqrt{\frac{\sum\nolimits_{i=1}^{n} (d_i)^2}{n}} \quad (6)$$

where d_i is the distance between OD flow T_i and the mean center, $(\overline{O}_x,\,\overline{O}_y,\,\overline{D}_x,\,\overline{D}_y)$ of the dataset (OD flows).

Step2: OD flow selection.

In this part, the main idea of OD flow selection is to generalize a subset of OD flows, with local maximal density as the most significant and representative OD flow, revealing major travel patterns in the dataset. In other words, an OD flow will be selected only if it has the local maximal density within its OD flow neighborhood. The neighborhood radius R is set as an input parameter to determine the number of selected OD flows according to two primary criteria for OD flow selection: (1) only one OD flow with local maximal density is selected from a single OD flow neighborhood, and (2) any two selective OD flows should keep at least a minimum distance (neighborhood radius R) from each other. Given that the neighborhood radius R should cooperate with the smoothing parameter h, the neighborhood radius R is suggested to be equal to or greater than the smoothing parameter h(Zhu et al. 2019).

Step3: OD flow mapping.

After selecting the most significant and representative OD flows, all these OD flows representing driving travel mobility patterns are visualized on a map. This part comprises three steps: (1) arrows indicate the directions of OD flow point from the origins to the destinations, and (2) OD flows with larger density values are labeled in darker color and laid on top of OD flows with lower density values.

3.3.2 Hotspot analysis (Getis-Ord Gi*)

To validate the suitability and superiority of OD-FDC algorithm for tourism hotspot identification, this research compares it with traditional hotspot analysis using Getis-Ord Gi* statistic. The OD-FDC algorithm identifies THs where OD flows with higher density values connect together. In contrast, hotspot analysis (Getis-Ord Gi*) divides the trip destinations into hot spots (cluster of high values of travel frequency) and cold spots (cluster of low values of travel frequency) based on z-scores (standard deviations) and p-values (probability; ESRI 2019). For Getis-Ord Gi* statistic, a statistically significant positive z-score (with small p-value) implies the intense clustering of high visitation (hotspots), while a statistically significant negative z-score suggests the clustering of the low visitation (cold spots). A z-score near zero indicates no spatial clustering. The uniqueness of OD-FDC algorithm, a flow-based analysis method, lies in its ability to process raw OD flow data (e.g., direction, intensity and spatial distribution of OD flow). Conversely, Getis-Ord Gi* statistic, as a point-based analysis method, can only process point data (e.g., aggregated visitation at destinations). Accordingly, it can be stated that OD-FDC advances traditional hotspot analysis in tourism hotspot detection by incorporating spatial interactions (how tourists travel between places) and accounting for geographical proximities (how high-visitation places cluster geographically).

4 Results

4.1 Descriptive statistics

Figure 4 initially presents the histograms of tourism demand in the dataset from an aspatial perspective (Note: "aspatial" is relative to "spatial").

The number of road trips (i.e., tourism demand) at the OD and the flow level are shown on the x-axis, and the frequency associated to certain trip numbers is shown on the y-axis. Three histograms are in a right-skewed (positive) distribution, revealing a spatial heterogeneity disproportion. This can be characterized as the 80/20 principle (Koch 2011), which indicates that a vast majority of points/flows (e.g., 80%) show a low travel frequency, whereas a minority of points/flows (e.g., 20%) present high travel flow in geographic space. Table 2 shows the fundamental statistics about the number of road trips at point and flow levels.

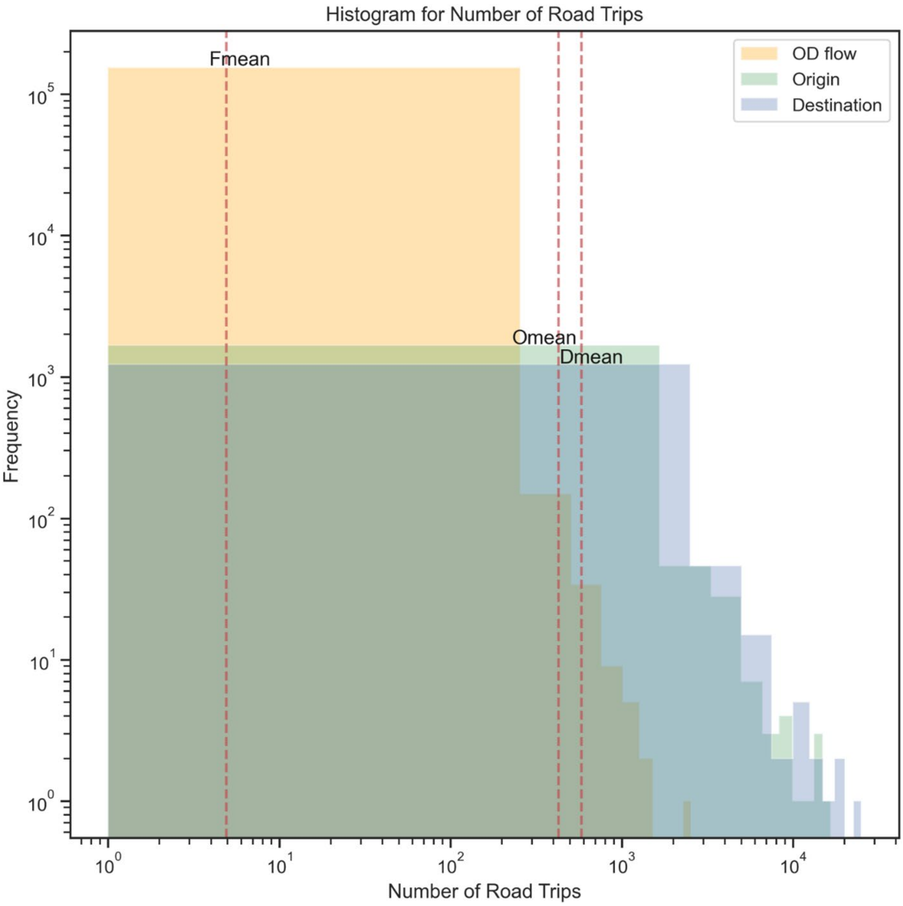

Fig. 4 Histogram for trip numbers in Jeju
(Note: Fmean refers to average number of trips at the flow level; Omean and Dmean refer to average number of trips at the origin and destination levels, respectively.)

About 428 road trips on average started the trips from the same origin, and approximately 583 trips on average visited the same destinations. Interestingly, upon checking the flow level, only five road trips present the same travel flow, which implies identical OD by road trip.

This result suggests that tourist mobility patterns at the flow level are more complex to characterize in terms of their geographic features compared to those at the point level. Greater spatial heterogeneity exists at the flow level than at the point level, as reflected in the shorter head (number of trips below average) and longer tail (number of trips above average) distributions. At the flow level, 82% of the data are in the head, whereas, at the point level, 80% are in the head, indicating a slightly higher spatial heterogeneity at the flow level.

To verify that travel frequency follows a heavy-tailed distribution (e.g., power law, log-normal, or exponential distributions), likelihood ratio tests are conducted to compare a power law distribution against a log-normal distribution (see Table 2). The negative log-likelihood ratio (R) indicates that the log-normal distribution best describes the travel demand at point and flow levels ($p<0.001$). This finding implies a scaling pattern where far more locations and flows are low in popularity, whereas only a few are highly popular.

4.2 Spatial clustering for flow-based density clustering

The OD-FDC algorithm effectively and efficiently extracts inherent scaling patterns from massive spatial mobility data by (1) estimating the density of individual travel flows, (2) selecting a representative OD flow, and (3) visualizing a major OD flow. The following part is organized to describe the three steps in the OD-FDC algorithm.

4.2.1 OD flow density estimation

The density of 154,682 travel flows is calculated as an initial step for the clustering method. According to SROT, this research derives the optimal bandwidth h of road

Table 2 Statistics about number of road trips at point and flow level

Statistics	Point level: Origin	Point level: Destination	Flow level: OD Flow
Mean	428.44	582.89	4.90
Number of points / flows	1,770	1,301	154,682
Number of points / flows in Head	1,416	1,041	127,230
Number of points / flows in Tail	354	260	27,452
Percentage of points / flows in Head	0.80	0.80	0.82
Percentage of points / flows in Tail	0.20	0.20	0.18
Loglikelihood Ratio R	-123961.421796***	-34643.152976***	-3392137.00***

(Note: Mean refers to the average number of trips at different aggregated levels: point levels (i.e., origins or destinations) and flow levels; the significance of the differences in fit between power law and lognormal distributions assessed by likelihood ratio tests; *** $p<0.001$, ** $p<0.01$, * $p<0.05$.)

trips to be 2.156 km. The optimal bandwidth h is set as a fixed value (2.156 km) and the neighborhood radius R is set in the range of 2 h(4.312 km) to 7 h(15.092 km) to show the multiscale OD flow patterns and identify the optimal value of R in this specific context. Table 3 presents the results of OD flow density according to different neighborhood radii (*R*).

As the value of *R* increases from 2 h to 7 h, the median and standard deviation of the density values show decreasing patterns. This can be explained by Tobler's law (the first law of geography): "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). As expected, when *R* increases, reflecting the larger spatial coverage to be defined as a flow neighborhood, the number of inclusive links dramatically increases. Thus, for each OD flow, as *R* increases, more OD flows are considered as its nearby OD flows. Each OD flow density is similar to those within its flow neighborhood. Furthermore, the statistical estimation in Table 3 demonstrates that OD flow densities across different neighborhood radii tend to follow heavy-tailed distributions (i.e., log-normal distribution). This suggests a greater number of low-density flows compared to high-density ones, pointing to spatial heterogeneity, a characteristic of scaling laws. To reveal and visualize the main OD flow patterns in Jeju from the massive OD data, the next step is to select representative OD flows.

4.2.2 Selection of representative flows

To identify major travel flow patterns from data exhibiting a heavy-tailed distribution, this study assesses the OD flow density and selects those with local maximum estimation. The OD-FDC algorithm is run with varying neighborhood radii *R*, ranging from 2 h(4.312 km) to 7 h(15.092 km). The results are shown in Table 4. As the neighborhood radius *R* increases to 7 h, the number of selected OD flows decreases from 17,107 to 4,879, suggesting that these selected OD flows are less likely to overlap. *R* represents the minimum distance between any two selected OD flows. Thus, the OD flows with a local maximum density within their respective neighborhoods will be farther away from each other as the neighborhood radius grows.

4.2.3 OD flow mapping

On the basis of the result of OD flow density estimation, the selective OD flows within the neighborhood radius in the range of 2 h(4.312 km) to 7 h(15.092 km) are plotted on choropleth maps using equal interval classification (details in Appendix 1.). Five equal intervals of 20% in the range of 0–100% are chosen to represent density values: very low (0–20%), low (21–40%), moderate (41–60%), high (61–80%), and very high (81–100%). The colors in the choropleth maps, from light yellow to dark green, denote the density values of the selected OD flows, from very low to very high (top 20% of OD flow density). OD flows with the highest density indicate the highest travel frequency within each neighborhood. Given that only OD flows with the largest density in each neighborhood are selected and drawn on the choropleth maps, those areas where selective OD flows connect together can be regarded as popular tourism destinations (THs).

Table 3 Statistical properties of OD flow density with different neighborhood radius R

Statistics	$R=2h$	$R=3h$	$R=4h$	$R=5h$	$R=6h$	$R=7h$
# OD flow	154,682	154,682	154,682	154,682	154,682	154,682
Mean	6,337,316.51	2,608,603.61	1,228,619.98	716,925.51	460,299.32	317,560.64
Median	6,140,465 (1337295.86)	2356376.5 (1166866.06)	955590.5 (918320.561)	484900.5 (704439.66)	277001.5 (547544.74)	173142.5 (438641.18)
% OD flow in Head	0.56	0.59	0.63	0.66	0.69	0.71
% OD flow in Tail	0.44	0.41	0.37	0.34	0.31	0.29
Loglikelihood Ratio R	-705052.04***	-586053.3***	-503972.53***	-460893.63***	-436262.13***	-420278.79***

(Note: Standard deviations in parentheses; optimal bandwidth h generated by SROT, is 2.156 km; Mean refers to the average density values of OD flows; # is number of; % is percentage of number of; the significance of the differences in fit between power law and lognormal distributions assessed by likelihood ratio tests; *** $p<0.001$, ** $p<0.01$, * $p<0.05$)

The result shown in Appendix 1. indicates that not only the intensive driving travel patterns at global and local geographic scales can be obtained from the figures, but spatial dispersion and concentration of self-driving movement patterns can also be visualized. Specifically, a majority of OD flows in dark green (selective OD flows with densities in the top 40%) are the ones departing from and arriving at coastal areas/beaches, indicating that a large proportion of self-driving tourists prefer traveling from and to coastal regions.

Prior to spatial dispersion and concentration of patterns, two types of TH should be defined according to the level of flow density, namely, even TH (E-TH) and concentrated TH (C-TH). Specifically, E-TH is the area where OD flows in light green connect together (OD flows with density labeled as low and moderate), whereas C-TH is the area where OD flows in dark green connect together (OD flows with density labeled as high and very high). As shown in Figs. 5 and 10 THs are detected by the OD-FDC algorithm, including 3 E-THs and 7 C-THs. Tourists congregate at C-THs but travel to/from E-THs, depicting dispersion travel patterns (in hub-to-spoke or spoke-to-hub style), whereas tourists congregate at C-THs and travel to/from C-THs, revealing concentration travel patterns (in hub-to-hub style; Park and Zhong 2022). Different from the K-means algorithm, THs uncovered from the flow-based density clustering method are positioned in contiguous and disjoint areas, suggesting that the OD-FDC algorithm effectively reveals THs with diffusive tourism activities in the destination.

4.3 Validity estimation of THs

This part attempts to estimate the validity of THs using the proposed method by comparing a typical hotspot analysis (Getis-Ord Gi*). Hotspot analysis presents the cluster results of hot and cold spots based on Getis-Ord Gi* statistic (see Fig. 6). Given that the mobility data distinguish OD information, a separate hotspot analysis is conducted. The results between the OD seem to be similar, but important variations exist in northeast and southwest areas. Although a TH is present in the northeast area (O/D-H-C4) with regard to the origin data (i.e., dots in orange), the destination data (i.e., dots in red) suggest a unique TH at the southwest area (O/D-H-C6) at Jeju. This result indicates the significance of considering flow-level data in detecting THs.

To verify the effectiveness of the OD-FDC algorithm in detecting THs, the results of Getis-Ord Gi* statistic are overlaid on top of the OD flow maps (see Fig. 7). Other results ranging from $R = 2$ h to $R = 7$ h can be found in Appendix 2.

In general, the hotspots identified through flow-based clustering overlap significantly with areas highlighted by hotspot analysis (Getis-Ord Gi*). Both types of THs are consistently located near coastal areas in Jeju. Notably, spatial zones identified as cold spots in the hotspot analysis are also not classified as hotspots using the flow-based density clustering method. This result demonstrates that the analytical results from flow-based clustering are validated by their high similarity to those derived from traditional clustering analysis. For additional validation, tourism-related POI (Points of Interest) data are analyzed within the corresponding areas, including attractions (e.g., parks, beaches, botanical and zoological gardens, museums), transportation facilities (e.g., airports, terminals, ferry ports), and lodging. As shown in Fig. 7,

Table 4 The neighborhood radius R and the corresponding statistics of selective OD flows

Statistics	$R=2h$	$R=3h$	$R=4h$	$R=5h$	$R=6h$	$R=7h$
# OD flow	154,682	154,682	154,682	154,682	154,682	154,682
# Selective OD flows	17,107	13,063	9,289	7,282	5,590	4,879
Mean	8,327,251	4,675,997	3,112,115	2,328,978	1,846,512	1,546,150
Median	8,184,829 (1337295.86)	4,508,223 (1166866.06)	2,937,307 (918320.561)	2,206,732 (704439.66)	1,687,272 (547544.74)	1,407,638 (438641.18)

(Note: optimal bandwidth h is 2.156 km; # is number of; Mean refers to the average density values of OD flows; Median refers to the median density values of OD flows; Standard deviations in parentheses.)

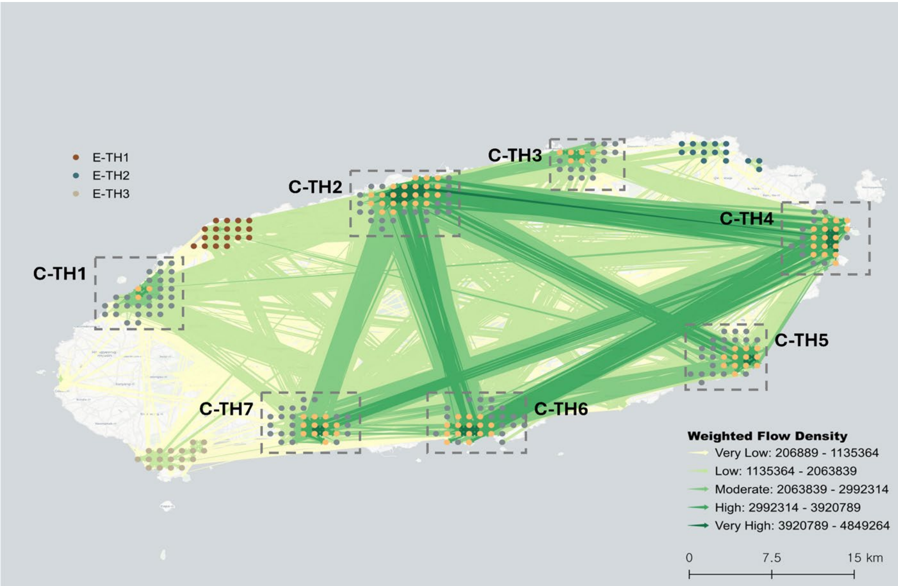

Fig. 5 5,590 selective TOD flows with neighborhood radius $R = 6\ h$(12.936 km)

all ten THs encompass numerous and diverse tourism attractions. Certain hotspot clusters, such as C-TH1, E-TH1, and C-TH4, particularly exhibit a higher concentration of various tourism assets, including natural and cultural facilities. Other THs, like C-TH2, C-TH6, and C-TH7, function as transportation hubs, featuring airports, car rental services, and ferry ports. Additionally, unique THs detected solely by the OD-FDC algorithm, such as C-TH3, E-TH2, and E-TH3, include diverse tourism facilities and services, such as lodging, tourism attractions, and cultural facilities (e.g., traditional markets and local festivals). These findings validate the suitability and superiority of the OD-FDC algorithm in identifying comprehensive THs within a destination.

Moreover, the OD-FDC algorithm provides additional insights compared to traditional hotspot analysis, revealing the intensity (indicated by flow density color) and diversity of source markets (characterized by the number of unique inflows and outflows at each destination or origin) for each TH. The number of OD flows and

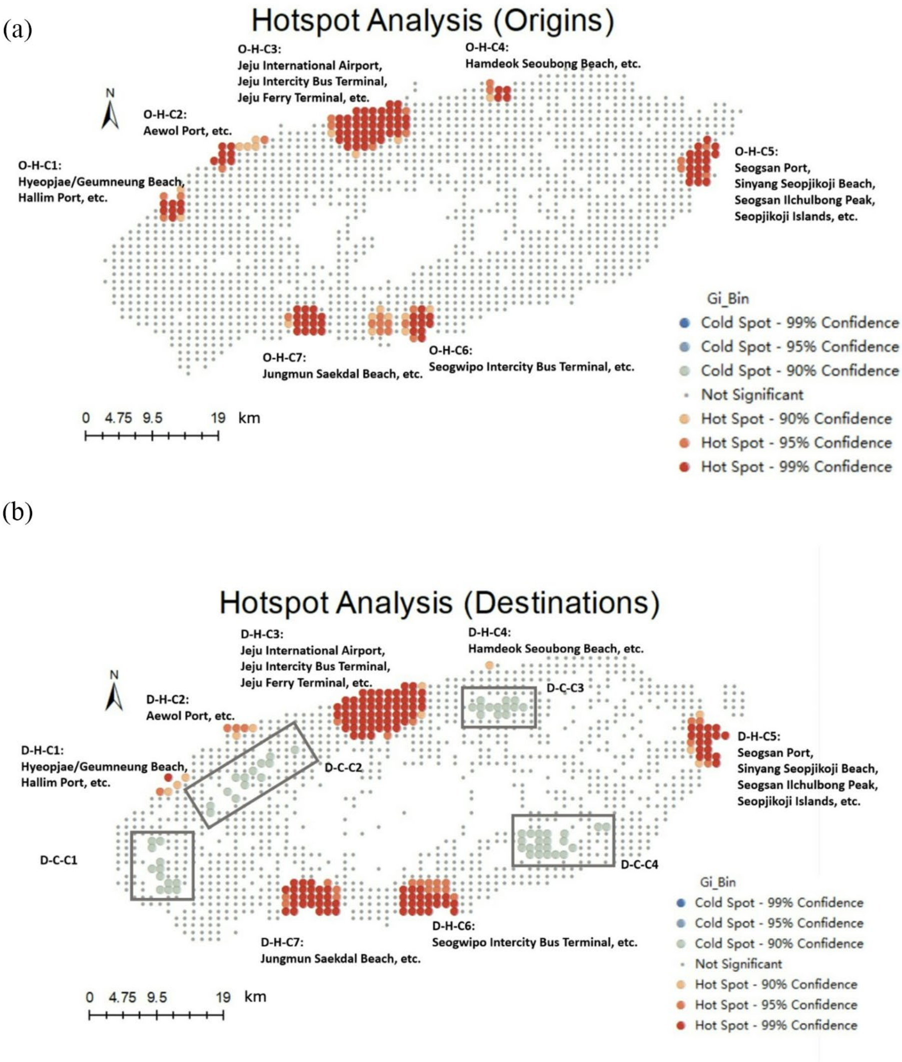

Fig. 6 Hot Spot Analysis (Getis-Ord Gi*) for (**a**) origins and (**b**) destinations in Jeju

their chroma levels among the ten THs reflect the intensity and source diversity of specific destinations (see Fig. 7). This information serves as a data-driven approach to representing destination competitiveness and attractiveness (Chen et al. 2025).

5 Conclusion

The evolution of information and communication technology allows tourism researchers to access massive and fine-grained data, enabling them to understand travel behaviors. It has accelerated the access to not only large but also diverse data

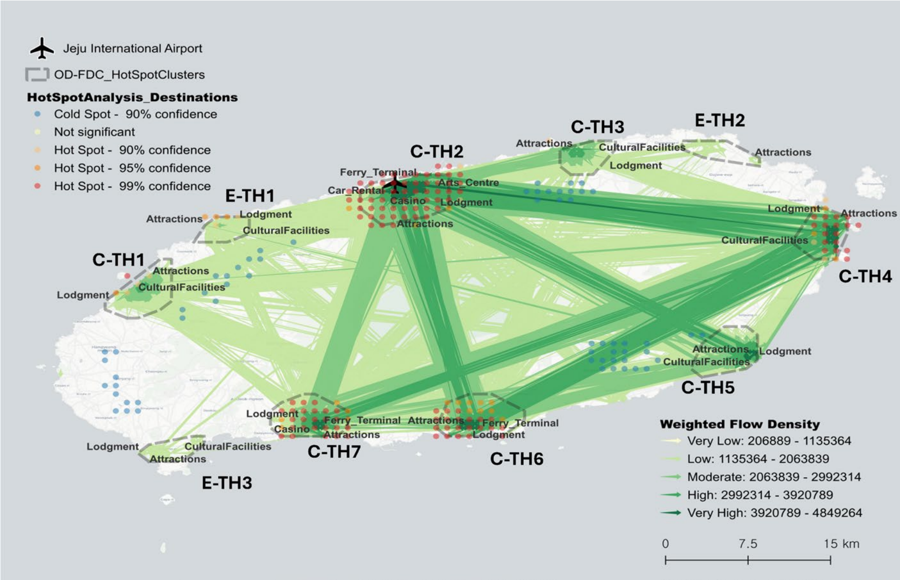

Fig. 7 OD flow map with Origin Tourism Hot-spots, Destination Tourism Cold-spots and Destination Tourism Hot-spots, where neighborhood radius R is 6 h

formats (Li et al. 2018). This opportunity suggests the necessity for tourism researchers to adopt innovative methods in collecting and analyzing data and suggesting important insights. Indeed, this study discusses the car navigation data representing travel mobility, whereby travelers visit various places with a sequential mobility pattern. The data include spatial and temporal information of travel flow in an OD format, which is different from the typical travel demand data that simply shows the number of visitors on a certain place.

This research advances tool development for enhancing spatial analytics of flow data, as well as promotes a methodological shrift from point-based to flow-based analysis. Previous studies exploring travel mobility have attempted to identify THs by applying point-based analysis methods, such as Getis-Ord Gi* statistics (Van der Zee et al. 2020), K-means and DBSCAN approaches (e.g., Al-Saad et al. 2023; Park et al. 2020). As opposed to the travel demand data, the OD data includes inclusive information about the place where people depart and the one where they arrive, referring to directional travel flows. The point-based analysis methods have limitations of managing the flow data because it focuses primarily on a spatial dimension, which provides partial understanding of travel movement patterns. Thus, this paper suggests an innovative flow-based analysis method, an origin-to-destination flow-based density clustering (OD-FDC) method, which accommodates the key characteristics of travel flows considering spatial and temporal dimensions. This method defines the THs that connect the selective flows with high density estimations. The findings of this paper suggest a number of academic and practical implications.

5.1 Theoretical implications

In reshaping the conceptual framework of tourism research, this study challenges the traditional reliance on central place theory (Van der Zee et al. 2020), which has predominantly focused on the ‘point (place)’ as the core of analysis and modeling. Traditionally, this theory has delineated regions based on metrics such as population, emphasizing vertical relationships but it overlooks horizontal or non-hierarchical relations, such as the external connectivity between regions (Taylor et al. 2010). By pivoting from a focus on “place” to “flow”, our research draws on the principles of central flow theory, which posits that dynamic spatial interactions create and define THs.

Moreover, with the growing popularity of big data, the graph theory has increasingly regarded destination structure as a crucial approach for exploring complex systems, emphasizing the edges or links (i.e., travel flows) between nodes (i.e., destinations) (Park et al. 2021; Phillips et al. 2015). Building on the graph theory which consider THs as places connected by significant travel flows, this research proposed the OD-FDC algorithm to identify key travel flows and define THs based on highlighted spatial interactions. As demonstrated in the validity estimation of THs, this study confirmed the inclusion of various tourism-related POIs within the THs identified by the OD-FDC algorithm, aligning with the principles of cumulative attraction theory. Consequently, the OD-FDC algorithm proposed in this research represents an innovative approach by integrating two essential theories in travel mobility.

Furthermore, the findings of OD-FDC algorithm enrich our theoretical understanding of spatial dependency and heterogeneity within tourism geography. Specifically, (1) spatial dependency indicates that nearby locations are more likely to have similar values or characteristics than distant ones, as described by the first law of geography (Tobler’s Law; Miller 2004). The density of each OD flow is determined not only by its own travel frequency but also by the travel frequency of its nearby OD flows. (2) Spatial heterogeneity indicates that differences in locational attributes vary across different spatial units. Important and representative OD flows are visualized on a map at the global (most popular OD flows in Jeju) and local (most popular OD flows in each OD flow neighborhood) scales (Yang et al. 2019). This allows major travel demand patterns across different spatial units to be shown on a map without losing information about the OD of each OD flow. Thus, the structure of self-driving tourism itineraries at different geographic scales within a tourism destination can be visualized and understood well.

Moreover, this research provides a foundational step towards redefining how tourism researchers perceive and analyze spatial data. The robustness of the OD-FDC algorithm in capturing and visualizing complex travel patterns without extensive empirical parameter tuning highlights the potential for developing more adaptive, real-time analytical tools in tourism research (see Appendix 1).

For additional validity estimation of the findings, several stakeholders associated to Jeju Tourism Organization are invited to confirm the validity and reliability of the THs derived from the flow-based density clustering method. As a result, this research provides important academic and methodological implications for tourism researchers. Increasing accessibility to tourism big data requires current researchers to develop tourism-oriented methods and identify valuable insights from massive and complex data. On the basis of the graph theory, this research suggests tourism researchers a useful method for analyzing OD data reflecting travel flows by addressing the limitations of conventional machine learning methods.

In this sense, the findings of this research provide detailed steps of how to calculate flow density and identify important movements, representing the TH big data. This study also demonstrates the similarity (or validity) of the method compared to the results of a conventional hotspot analysis developed by a Getis-Ord Gi* approach. The results of this research present the excellence of the proposed method by not only showing high similarity of the results to the typical method but also suggesting additional THs that the conventional method cannot identify.

In addition to demonstrating the excellence of the proposed model, this study checks the content validity by inviting relevant stakeholders. By taking advantage of the authors' resources, a number of stakeholders in government and private sectors in tourism industry at Jeju are invited to show the results of the data analysis. The thematic estimation is confirmed to check the accuracy and suitability of the analytical results from the OD-FDC.

5.2 Practical implications

This paper also suggests diverse practical implications. At present, DMOs have developed technological infrastructure to collect massive and real-time data (e.g., tourism big data platforms). However, they are facing challenges in adopting machine learning technology to analyze big data and identify actionable insights. This research provides a foundational approach to model travel behaviors accurately and potentially generate helpful insights into travel movement patterns and dynamic THs for DMOs. First, this information should be helpful for them to develop destination planning and marketing strategies. For instance, the research findings can inform DMOs in optimizing city tour bus routes, which often suffer from low passenger numbers due to inefficient route planning. By applying the proposed OD-FDC algorithm, destination managers can redesign bus routes to connect more attractive locations aligned with actual tourist flow information (e.g., flow intensities and directions), potentially increasing ridership and improving tourist satisfaction. This approach enables more strategic placement of bus stops to match tourist movements between identified THs.

Second, the proposed method can serve as a foundation to develop a destination recommendation system. Unlike traditional systems that suggest isolated attractions for travelers to visit, the results of flow-based clustering method would recommend a TH based on tourists' current location and the popularity of routes. Although it is not a fully personalized recommendation system, it provides a ranked list of destinations based on aggregated hotspot analysis, offering travelers data-driven itineraries aligned with popular movement patterns. Furthermore, the method's applicability extends beyond tourism to airline industry. Airline managers are required to identify THs not only based on how many people visit a destination but also on the density and viability of the route connecting the OD (Spasojevic and Lohmann 2022). This method can aid airlines in evaluating route density and viability, supporting decisions to maintain or discontinue specific routes based on OD flow dynamics, thereby optimizing operational efficiency and customer satisfaction.

5.3 Limitation and future study

The limitations of this study are threefold, presenting opportunities for future research. First, due to data constraints, this research analyzed only one-month data. While focusing on a single month allowed for a more concentrated exploration of THs, it may not capture the seasonal variations in tourists' spatial behavior. Additionally, it may not be sufficient to ensure the continued validity of the algorithm. Future studies could collect longitudinal data spanning multiple seasons to investigate temporal variations in THs across different seasons and evaluate the stability and applicability of the algorithm more thoroughly. Second, this research relies on a single dataset containing OD flow information—namely, flow directions, intensities and spatial distributions—which may limit the understanding of why certain THs attract more visitors than others. Future studies may integrate additional multi-source datasets, such as street network and POIs, to enhance the comprehensiveness of hotspot identification. Third, the navigation dataset used is confined to the nature-oriented city of Jeju, which may limit the algorithm's generalizability. Future research could improve the algorithm's generality by analyzing flow data from other culture-oriented cities to validate the algorithm's usability and applicability across more diverse tourism destinations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40558-025-00325-3.

Declarations

Conflict of interest The authors report that there are no conflict of interests to declare.

References

Al-Saad SA, Jawarneh RN, Aloudat AS (2023) Spatiotemporal cluster analysis of reputable tourist accommodation in greater Amman municipality, Jordan. J Hospitality Tourism Technol 14(4):579–597. https://doi.org/10.1108/JHTT-03-2021-0071

Baggio R (2017) Network science and tourism–the state of the Art. Tourism Rev 72(1):120–131. https://doi.org/10.1108/TR-01-2017-0008

Bashtannyk DM, Hyndman RJ (2001) Bandwidth selection for kernel conditional density Estimation. Comput Stat Data Anal 36(3):279–298. https://doi.org/10.1016/S0167-9473(00)00046-3

Bauder M, Freytag T (2015) Visitor mobility in the City and the effects of travel Preparation. Tourism Geographies 17(5):682–700. https://doi.org/10.1080/14616688.2015.1053971

Cai X, Wu Z, Cheng J (2013) Using kernel density Estimation to assess the Spatial pattern of road density and its impact on landscape fragmentation. Int J Geogr Inf Sci 27(2):222–230. https://doi.org/10.1080/13658816.2012.663918

Chakravorty S (1995) Identifying crime clusters: the Spatial principles. Middle States Geogr 28:53–58

Cheng M, Jin X, Wang Y, Wang X, Chen J (2023) A sequential pattern mining approach to tourist movement: the case of a mega event. J Travel Res 62(6):1237–1256. https://doi.org/10.1177/00472875221126433

Chen J, Wu J, Wang D, Stantic B (2025) Beyond static rankings: A tourist experience-driven approach to measure destination competitiveness. Tour Manag 106:105022

Chen YC (2017) A tutorial on kernel density Estimation and recent advances. Biostatistics Epidemiol 1(1):161–187. https://doi.org/10.1080/24709360.2017.1396742

Cooper CP (1981) Spatial and Temporal patterns of tourist behaviour. Reg Stud 15(5):359–371. https://doi.org/10.1080/09595238100185351

Crompton J (2025) Reflections on the six motives that drive tourists' pleasure vacation behavior. J Travel Res 64(1):3–34. https://doi.org/10.1177/00472875241281520

Crompton JL, Gitelson RJ (1979) The theory of cumulative attraction and compatibility: a case study of two major commercial leisure enterprises. Bayl Bus Stud 10(1):7–16

Dale MRT, Fortin MJ (2010) From graphs to Spatial graphs. Annu Rev Ecol Evol Syst 41:21–38. https://doi.org/10.1146/annurev-ecolsys-102209-144718

ESRI (2019) *How hot spot analysis (Getis-Ord Gi*) works. ArcGIS Pro Tool Reference*. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm#. Access 26 June 2024

Hall CM (2005) Reconsidering the geography of tourism and contemporary mobility. Geographical Res 43(2):125–139. https://doi.org/10.1111/j.1745-5871.2005.00308.x

Hardt D, Glückstad FK (2024) A social media analysis of travel preferences and attitudes, before and during Covid-19. Tourism Manage (1982) 100:104821. https://doi.org/10.1016/j.tourman.2023.104821

Hardy A, Hyslop S, Booth K, Robards B, Aryal J, Gretzel U, Eccleston R (2017) Tracking tourists' travel with smartphone-based GPS technology: a methodological discussion. Inform Technol Tourism 17:255–274. https://doi.org/10.1007/s40558-017-0086-3

Harpole JK, Woods CM, Rodebaugh TL, Levinson CA, Lenze EJ (2014) How bandwidth selection algorithms impact exploratory data analysis using kernel density Estimation. Psychol Methods 19(3):428. https://doi.org/10.1037/a0036850

Hayes B (2000) Computing science: graph theory in practice: part II. Am Sci 88(2):104–109

Huang L, Li M, Zheng W, Gao S (2024) Exploring the evolutionary patterns of urban tourist mobility within a day: A three-step analysis framework. Int J Tourism Res 26(4). https://doi.org/10.1002/jtr.2680

Hunt MA, Crompton JL (2008) Investigating attraction compatibility in an East Texas City. Int J Tourism Res 10(3):237–246. https://doi.org/10.1002/jtr.652

Hwang Y, Gretzel U, Xiang Z, Fesenmaier DR (2006) Information search for travel decisions. Destination Recommendation Systems: Behav Found Appl 42(4):357–371. https://doi.org/10.1079/9780851990231.0003

Kang S (2016) Associations between space–time constraints and Spatial patterns of travels. Annals Tourism Res 61:127–141. https://doi.org/10.1016/j.annals.2016.09.010

Kim EJ, Kim Y, Jang S, Kim DK (2021) Tourists' preference on the combination of travel modes under Mobility-as-a-Service environment. Transp Res Part A: Policy Pract 150:236–255. https://doi.org/10.1016/j.tra.2021.06.016

Koch R (2011) *The 80/20 Principle: The Secret of Achieving More with Less: Updated 20th anniversary edition of the productivity and business classic*. Hachette UK

Li J, Xu L, Tang L, Wang S, Li L (2018) Big data in tourism research: A literature review. Tour Manag 68:301–323. https://doi.org/10.1016/j.tourman.2018.03.009

Li XR, Cheng CK, Kim H, Petrick JF (2008) A systematic comparison of first-time and repeat visitors via a two-phase online survey. Tour Manag 29(2):278–293. https://doi.org/10.1016/j.tourman.2007.03.010

Lloyd R (1997) Spatial cognition: geographic environments, vol 39. Springer Science & Business Media

Miller HJ (2004) Tobler's first law and Spatial analysis. Ann Assoc Am Geogr 94(2):284–289. https://doi.org/10.1111/j.1467-8306.2004.09402005.x

Nakaya T, Yano K (2010) Visualising crime clusters in a space-time cube: an exploratory data-analysis approach using space-time kernel density Estimation and scan statistics. Trans GIS 14(3):223–239. https://doi.org/10.1111/j.1467-9671.2010.01194.x

Nelson RL (1959) The selection of retail locations. F. W. Dodge Corporation

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 Tourism Res 84:102973. https://doi.org/10.1016/j.annals.2020.102973

Park S, Yuan Y, Choe Y (2021) Application of graph theory to mining the similarity of travel trajectories. Tour Manag 87:104391. https://doi.org/10.1016/j.tourman.2021.104391

Park S, Zhong RR (2022) Pattern recognition of travel mobility in a City destination: application of network motif analytics. J Travel Res 61(5):1201–1216. https://doi.org/10.1177/00472875211024739

Pettersson R, Zillinger M (2011) Time and space in event behaviour: tracking visitors by GPS. Tourism Geographies 13(1):1–20. https://doi.org/10.1080/14616688.2010.529932

Phillips JD, Schwanghart W, Heckmann T (2015) Graph theory in the geosciences. Earth Sci Rev 143:147–160. https://doi.org/10.1016/j.earscirev.2015.02.002

Provenzano D, Giambrone R (2023) CLUSTERING OF TOURISM PATTERNS WITH SELF-ORGANIZING MAPS: THE CASE OF SICILY. Tourism Anal 28(4):625–641. https://doi.org/10.3727/108354223X16773711119152

Raun J, Ahas R, Tiru M (2016) Measuring tourism destinations using mobile tracking data. Tour Manag 57:202–212. https://doi.org/10.1016/j.tourman.2016.06.006

Rodríguez-Echeverría J, Semanjski I, Van Gheluwe C, Ochoa D, IJben H, Gautama S (2020) Density-Based Spatial clustering and ordering points approach for characterizations of tourist behaviour. ISPRS Int J Geo-Information 9(11):686. https://doi.org/10.3390/ijgi9110686

Sainaghi R, Baggio R (2017) Complexity traits and dynamics of tourism destinations. Tour Manag 63:368–382. https://doi.org/10.1016/j.tourman.2017.07.004

Salas-Olmedo MH, Moya-Gómez B, García-Palomares JC, Gutiérrez J (2018) Tourists' digital footprint in cities: comparing big data sources. Tourism Manage (1982) 66:13–25. https://doi.org/10.1016/j.tourman.2017.11.001

Schneider CM, Belik V, Couronné T, Smoreda Z, González MC (2013) Unravelling daily human mobility motifs. J Royal Soc Interface 10(84):20130246. https://doi.org/10.1098/rsif.2013.0246

Shoval N, Isaacson M (2007) Tracking tourists in the digital age. Annals Tourism Res 34(1):141–159. https://doi.org/10.1016/j.annals.2006.07.007

Silverman BW (1986) Density Estimation for statistics and data analysis. London; Chapman and Hall

Spasojevic B, Lohmann G (2022) Air route development-Lessons from Australia. J Air Transp Manage 104:102274. https://doi.org/10.1016/j.jairtraman.2022.102274

Strogatz SH (2001) Exploring complex networks. Nature 410(6825):268–276. https://doi.org/10.1038/35065725

Su X, Spierings B, Hooimeijer P, Scheider S (2020) Where day trippers and tourists go: comparing the spatio-temporal distribution of Mainland Chinese visitors in Hong Kong using Weibo data. Asia Pac J Tourism Res 25(5):505–523. https://doi.org/10.1080/10941665.2020.1741409

Tang J, Bi W, Liu F, Zhang W (2021) Exploring urban travel patterns using density-based clustering with multi-attributes from large-scaled vehicle trajectories. Phys A 561:125301. https://doi.org/10.1016/j.physa.2020.125301

Taylor PJ, Hoyler M, Verbruggen R (2010) External urban relational process: introducing central flow theory to complement central place theory. Urban Stud 47(13):2803–2818. https://doi.org/10.1177/0042098010377367

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234–240. https://doi.org/10.2307/143141

Tussyadiah IP, Zach FJ (2012) The role of geo-based technology in place experiences. Annals Tourism Res 39(2):780–800. https://doi.org/10.1016/j.annals.2011.10.003

Van der Zee E, Bertocchi D, Vanneste D (2020) Distribution of tourists within urban heritage destinations: a hot spot/cold spot analysis of tripadvisor data as support for destination management. Curr Issues Tourism 23(2):175–196. https://doi.org/10.1080/13683500.2018.1491955

Vu HQ, Li G, Law R, Zhang Y (2018) Travel diaries analysis by sequential rule mining. J Travel Res 57(3):399–413. https://doi.org/10.1177/0047287517692446

Wang D, Park S, Fesenmaier DR (2012) The role of smartphones in mediating the touristic experience. J Travel Res 51(4):371–387. https://doi.org/10.1177/0047287511426341

Wang Z, He SY, Leung Y (2018) Applying mobile phone data to travel behaviour research: A literature review. Travel Behav Soc 11:141–155. https://doi.org/10.1016/j.tbs.2017.02.005

Wong, E., Law, R., Li, G., Schegg, R., & Stangl, B. (2017). Reviewing Geotagging Research in Tourism. In Information and Communication Technologies in Tourism 2017 (pp. 43–58). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-51168-9_4

Wu X, Huang Z, Peng X, Chen Y, Liu Y (2018) Building a Spatially-Embedded network of tourism hotspots from geotagged social media data. IEEE Access 6:21945–21955. https://doi.org/10.1109/ACCESS.2018.2828032

Xia W, Mei B, Yuanjiang J (2007) Tourism planning for the scenic speedway based on landscape evaluation: A case study on funing scenic speedway. Econ Geogr 27(1):161–166. https://doi.org/10.3969/j.issn.1000-8462.2007.01.036

Xu Y, Li J, Belyi A, Park S (2021) Characterizing destination networks through mobility traces of international tourists—A case study using a nationwide mobile positioning dataset. Tour Manag 82:104195. https://doi.org/10.1016/j.tourman.2020.104195

Yang S, Yang X, Zhang C, Spyrou E (2010) Using social network theory for modeling human mobility. IEEE Network 24(5):6–13. https://doi.org/10.1109/MNET.2010.5578912

Yang X, Fang Z, Xu Y, Yin L, Li J, Lu S (2019) Spatial heterogeneity in spatial interaction of human movements—Insights from large-scale mobile positioning data J Transp Geogr, *78*, 29–40. https://doi.org/10.1016/j.jtrangeo.2019.05.010

Zhang H, Zhou X, Ye X, Tang G, Wang H, Jiang S (2023) Strength-weighted flow cluster method considering Spatiotemporal contiguity to reveal interregional association patterns. GIScience Remote Sens 60(1). https://doi.org/10.1080/15481603.2023.2252923

Zheng W, Li M, Lin Z, Zhang Y (2022) Leveraging tourist trajectory data for effective destination planning and management: A new heuristic approach. Tour Manag 89:104437. https://doi.org/10.1016/j.tourman.2021.104437

Zhu X, Guo D, Koylu C, Chen C (2019) Density-based multi-scale flow mapping and generalization. Comput Environ Urban Syst 77:101359. https://doi.org/10.1016/j.compenvurbsys.2019.101359

Authors and Affiliations

Yingqi Yuan[1] · Sangwon Park[2] · Yang Xu[3] · Sun-Young Koh[4]

✉ Sangwon Park
sangwon.park@khu.ac.kr

Yingqi Yuan
yingqi.yuan@khu.ac.kr

Yang Xu
yang.ls.xu@polyu.edu.hk

Sun-Young Koh
sunyko@ijto.or.kr

1 Department of Global Hospitality & Tourism Management, Kyung Hee University, Seoul, Republic of Korea

2 Smart Tourism Education Platform, College of Hotel and Tourism Management, Kyung Hee University, Seoul, Republic of Korea

3 Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

4 Jeju Research Institute, Jeju, Republic of Korea