WILEY | Hindawi

Research Article

Discovering the Graph-Based Flow Patterns of Car Tourists Using License Plate Data: A Case Study in Shenzhen, China

He Bing,[1] Kong Bo [ID],[2] Yin Ling,[1] Wu Qin,[3] Hu Jinxing,[1] Huang Dian,[4] and Ma Zhanwu[1]

[1]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*
[2]*Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610041, China*
[3]*Chengdu University of Information Technology, School of Computer Science, Chengdu 610225, China*
[4]*National Supercomputing Center in Shenzhen, Shenzhen 518055, China*

Correspondence should be addressed to Kong Bo; kongbo827@imde.ac.cn

Identifying flow patterns from massive trajectories of car tourists is considered a promising way to improve the management of tourism traffic. Previous researches have mainly focused on tourist movements at the macro-scale, such as inbound, domestic, and urban tourism using flow maps. Compared with modeling the flow patterns of tourists at the macro-scale, modeling tourist flow at the microscale is more complicated. This paper takes Dapeng Island located in Shenzhen as the study area and uses the car recognition devices to collect traffic flow. Firstly, car tourists are separated from the mixed traffic flow after analyzing the spatial-temporal characteristics of tourists and residents. Next, daily graphs of tourist movements between road segments and tourist attractions are constructed. Finally, a frequent subgraph mining algorithm is used to extract the flow patterns of car tourists. The experimental results show that (1) car tourists have obvious preferences in the selection of trip time and tourist attractions; (2) the intercity tourists tend to take multidestination trips rather than a single destination trip in the same type of attractions; (3) car tourists are inclined to park their cars in an easy-to-access place, even if the attractions visited are changed. The main contribution of this paper is to present a new method for discovering the flow patterns of car tourists hidden in massive amounts of license plate data.

1. Introduction

Due to the flexibility and convenience of road transportation, car-based tourism (travel in owned or rented cars, also named driving tours [1], car tourism [2], and self-driving tours [3]; for simplicity, this paper uses the term car tourism) has been one of the popular forms for leisure and recreation. Recently, car tourism has been growing rapidly in China, and its scale is continuing to expand with the improvement of road infrastructure and the growth of car ownership. A statistical report indicated that by 2015, there were 2.34 billion car tourists in China, accounting for more than 58.5% of the total domestic tourists [4]. It can be foreseen that the percentage of car tourists will increase over time. However, car tourists need to share roads in urban areas or tourist attractions with residents, and they depend on the road network to achieve circulation between the places of origin and multiple tourist attractions. Currently, urban roads are heavily crowded. When a large number of tourist cars enter the road network during peak tourist season, the pressure on road traffic management may be increased. It is worth noting that this phenomenon is severe for coastal tourist attractions.

Coastal islands are one of the favorite tourist destinations. The development of road network on the islands often precedes the development of tourist attractions, and new infrastructures and facilities are being built to handle the increase in tourist traffic. Thus, tourism activities tend to be superimposed on a spatial system and infrastructure network that was not explicitly designed to cater to them and tourism activities can be unevenly distributed [5]. Additionally, some islands are connected to the mainland. The roads entering and leaving these islands have become bottlenecks for tourism transportation, which poses challenges

to the coordination of traffic on and off islands. Moreover, in contrast to commuting transportation, tourism transportation has different characteristics in time and space, and it requires more comfort and convenience. The problems mentioned above show that if tourism transportation is not taken seriously in traffic management, it is likely to increase travel difficulties for residents, and it will affect the travel willingness of tourists and the sustainable development of tourism transportation.

The recording and analysis of trajectories are essential for understanding the movement of tourists and the management of tourist traffic, such as the optimal location and development of transportation facilities and the redistribution of tourists. However, the lack of practical approaches for the collection of relevant data limits the detailed exploration of tourist mobility. The traditional method involves paper-and-pencil or computer interviews, which are expensive and time-consuming. The collected data are also typically limited in terms of personal information such as family composition, age structure, and favorite tourist attractions [6]. Recently, with the development of sensors such as GPS tracker, video recognition device, and RFID, which can capture movement data in real-time and with spatial and temporal details, the trajectory-based data analysis methods have been widely used in transportation research. The analysis results provide real-time and future traffic information for road traffic managers and travelers, as well as technical support for the relief of traffic jams. However, current observations of road traffic are limited to statistical information such as traffic volume, occupancy, and speed. Movement patterns are depicted in a flow graph or reported by visual descriptions rather than exploring flow patterns. Additionally, road traffic has the characteristics of variability and correlation in time and space. Previous researches have demonstrated that sectional traffic flow is interrelated to the distances and locations of monitoring points and the topology of road network. Therefore, it is necessary to consider the structure of road network and the correlation between time and space in the analysis of tourist traffic. This consideration is more useful in explaining the deeper behavior of tourist traffic.

This study is an attempt to investigate the flow patterns of car tourists by applying a frequent subgraph mining algorithm. This algorithm can take into account the correlation of traffic flows captured by video sensors. From the graph and flow perspective, a coastal island is used as an experimental area to explore the dynamic relationship between multiple tourist attractions and key road segments. This paper is organized as follows. The next section reviews related work on movement pattern mining and the methods for analyzing trajectory-based data on tourists and traffic flows. Section 3 introduces the study area (Dapeng Island, Shenzhen, China). Section 4 describes the distribution of the monitoring points in detail. Section 5 introduces the data and methods used in this paper. Section 6 presents the results of flow patterns generated by car tourists. Finally, we finish with a discussion and conclusion.

2. Related Work

In recent years, movement patterns have been analyzed frequently from transportation to tourism, such as the research of movement patterns hidden in taxis [7–11], buses [12, 13], railways [14], tourist movements [15–20], and even in geo-tagged media datasets [21, 22]. In terms of tourism transportation research, the efficient management of tourist traffic requires a sound understanding of car tourists' spatial movement patterns because these patterns provide critical information, e.g., the flow volume and spatial transfer direction, for the planning of new transportation facilities and the redistribution of tourist flow. As is well known, movement is an intrinsic attribute of traffic flow that changes over time with respect to the spatial location of people, goods, and cars. The patterns implied in moving datasets are not repeatedly produced by a single car tourist, but rather by a huge number of cars that appear in the same area. In most cases, the collected moving datasets of traffic entities are relatively large in volume and complex in structure. Therefore, it is necessary to use data mining algorithms and visual analytics techniques to extract useful and relevant information, regularities, and structures from massive movement datasets. The data mining algorithms used in transportation are varied. These algorithms focus on clustering [8], density, and sequential characteristics [9, 10] in time and space. The leisure activities of car tourists are carried out in a road network. The activity sequences can be modeled in a graph that consists of different nodes (for example, parking lots and cultural sites) and edges with direction that are the order of locations visited. For this kind of dataset, graph mining is a widely used method that finds interesting patterns in graph representation data [23]. The detected patterns are typically expressed as graphs, which may be subgraphs of graphical data or more abstract expressions of the trends reflected in data [24]. One form of graph mining is frequent subgraph mining, which is used to identify frequently occurring patterns (subgraphs) across a collection of "small" graphs or in a "large" graph [25]. Various subgraph mining algorithms have been proposed. These algorithms can be further classified based on the search strategies, i.e., either breadth-first or depth-first searches. The depth-first search strategy is more computationally efficient, such as in gSpan (graph-based Substructure pattern mining) [26], MoFa (Molecule Fragment Miner) [27], FFSM (Fast Frequent Subgraph Mining) [28], and Gaston (GrAph/Sequence/Tree extractiON) [29], SPIN (Spanning tree based maximal graph mining) [30]. However, FFSM and Gaston cannot be used for directed graphs without major changes. Only MoFa is suitable for finding directed frequent subgraphs, and for gSpan, only minor changes are necessary [31]. Other related works include significant pattern mining Leap [32], maximal frequent subgraph mining Margin [33], and frequent subgraphs in multigraphs [34].

During the past few years, trajectory-based methods have been used to analyze transportation systems

[10, 11, 13, 14]. In many applications, moving entities are considered moving points whose trajectories (i.e., paths through space and time) can be visualized and analyzed. In transportation, the collected trajectory data can be presented in origin-destination (OD) data with aggregation methods [35]. Such OD data can be visualized with a set of techniques, including flow maps [36, 37] and OD maps [38]. Nevertheless, the study of the spatial dimensions of tourism remains a mostly underexplored area of research, although this research area is expanding due to the advances of new information and communication technologies (ICT). Traditional approaches in tourism can be divided into two categories, which are direct observation techniques (e.g., interviews, trip diaries, and recall diaries) and non-observation techniques (e.g., GPS tracking and video tracking), but the use of nonparticipant observation only is the best technique for privacy reasons [5]. Even with ICT support, this technique has difficulties in data collection, and the large-scale sampling of passenger data is costly. Most published studies related to movement patterns are still descriptive, and they employ small sample sizes that are highly controlled. Moreover, this kind of research is focused on the human movement in tourist intradestination. There have been a few studies that have explored car-related spatial movement patterns in tourist destinations. Even so, the research was aimed at large-scale car tourists' activities [39] or used the questionnaire method, which is prone to biases and errors [40]. Therefore, given the requirements of protecting privacy, increasing data volume, and avoiding investigator biases, it is necessary to conduct flow pattern research based on continuous time-series data acquired from sensors.

3. Study Area

Dapeng Island, located in the east of Shenzhen (as shown in Figure 1), is an essential node in the "Guangdong-Hong Kong-Macao Greater Bay Area," and it is the only pioneer zone of national tourism reform and innovation in Guangdong province. Dapeng has abundant tourism resources, such as Dapeng Ancient City, National Geological Park, and Folk Village. The "Shenzhen Tourism Statistics Bulletin" showed that a total of 139 million tourists visited this city in 2018. The increase was 5.97 percent each year, of which only one-tenth was group tourists. This indicates that most tourism activities are carried out by individual visitors. Because of topographical constraints, tourism transportation on the island has not been fully developed. Owned and rented car tours are the main modes of visiting the island for individual tourists. The Dapeng Transportation Bureau has analyzed the trend of motorized travel demand on the island and predicted that the total annual traffic flow would be 504000 cars/year. At the peak time of "Golden Week," about 35000 cars/day entered into the island, of which 79% were car tourists.

4. Distribution of Monitoring Points

Urban transportation systems usually employ GPS technology to capture taxi and bus tracks. Different from this kind of public transportation research, this paper aims to analyze the flow patterns of car tourists at multiple attractions. It is difficult to install GPS device on each personal car. Therefore, we chose roadside monitoring devices to collect traffic flow. In addition, urban road network includes expressways, ordinary roads, and community roads. It has a large number of nodes and complex structure. In order to monitor each road segment, many devices will need to be deployed on roads. So, the key road segments and tourist attractions were selected as the locations of monitoring points. Five video devices were deployed at key roads, and two video devices were deployed in parking lots. The labels of monitoring devices are $A1$, $A2$, $A5$, $A7$, $A10$, $P1$, and $P2$ (as shown in Figure 2).

The detected tourist flow at each monitoring point is shown in Table 1.

5. Methods

In this study, a cloud-based database system was established to store traffic data after they were uploaded via 4G communication technologies. The collected data includes license plate numbers, time of passage, and labels for monitoring points. In order to protect tourists' private information, license plate numbers were changed into car IDs and only the registration places of cars were extracted. After data collection period, the license plate numbers will be deleted from the database.

The proposed approach is outlined in Figure 3. This section introduces the detailed steps for the mining of frequent flow patterns of car tourists at the microscale in a tourist intradestination. First, car tourists were separated from mixed traffic flow after analyzing the temporal characteristics of collected data. Second, spatial movement graphs of traffic flow were reconstructed for each day. Each movement graph is a connected and directed graph where vertices are monitoring points and directed edges are tourist flow between two monitoring points. Then, a frequent subgraph mining algorithm (gSpan) was used to detect the flow patterns between road segments and tourist attractions. Next, in order to reduce the number of frequent flow patterns, small overlapping subgraphs were removed from the results. Finally, we analyzed the spatial-temporal characteristics of flow patterns intradestination.

5.1. Preprocessing Source Data. When collecting data, it is inevitable that problem data will be collected. This can be caused by a problem with the device, such as an aging or damaged camera. Additionally, a license plate may be blurred, blocked, or damaged, especially in bad weather, which can affect the efficiency of car recognition. Furthermore, it is challenging to identify some special characters and confusing numbers on license plates. The above issues can lead to data distortion. To ensure the accuracy and reliability of the analyzed results, the collected data need to be processed at first. The rules of processing are as follows:
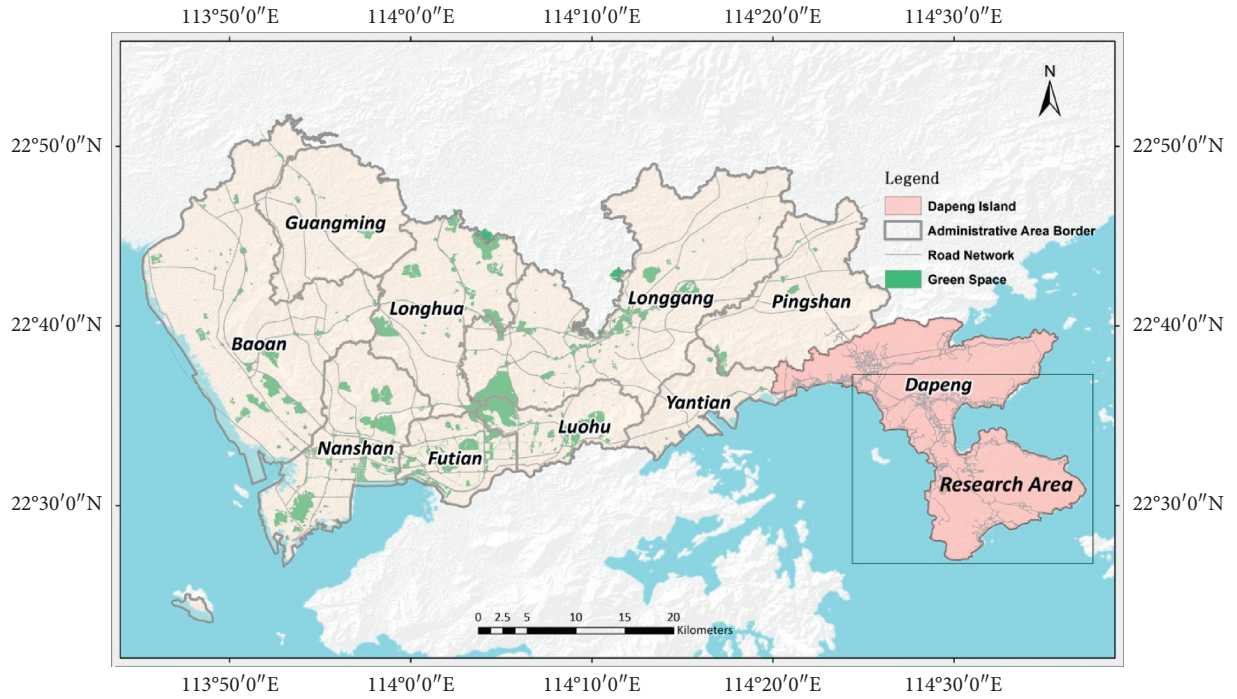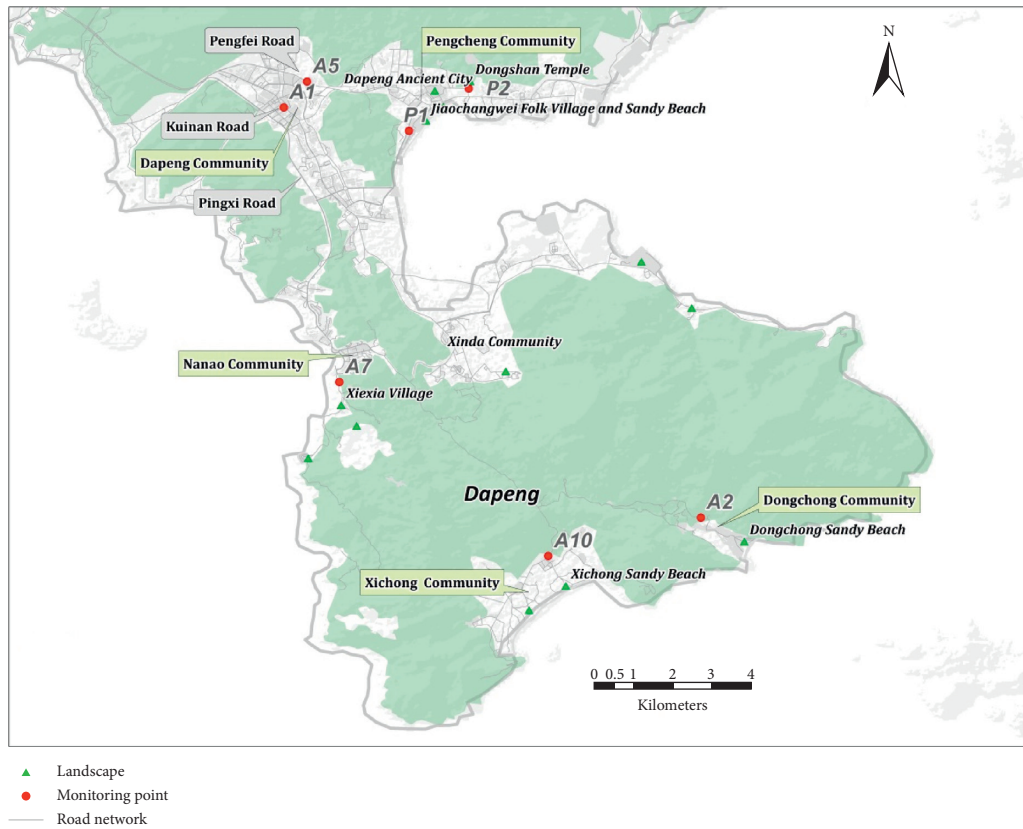
Figure 1: The location of dapeng island, Shenzhen.

Figure 2: Distribution of monitoring points for tourist traffic flow.

TABLE 1: Car tourists detected at monitoring points.

| Label	Location	Detected car tourists
A1	Kuinan Road	Dapeng Community and Xinda Community
A2	Dongchong Road	Dongchong Community
A5	Pengfei Road	Dapeng community and Dapeng Ancient City, Jiaochangwei Folk Village, Dongshan Temple
A7	Fumin Road	Nan'ao Community
A10	Nanxi Road	Xichong Community
P1	Jiaochangwei Park Lots	Jiaochangwei Folk Village
P2	Ancient City Park Lots	Dapeng Ancient City, Dongshan Temple
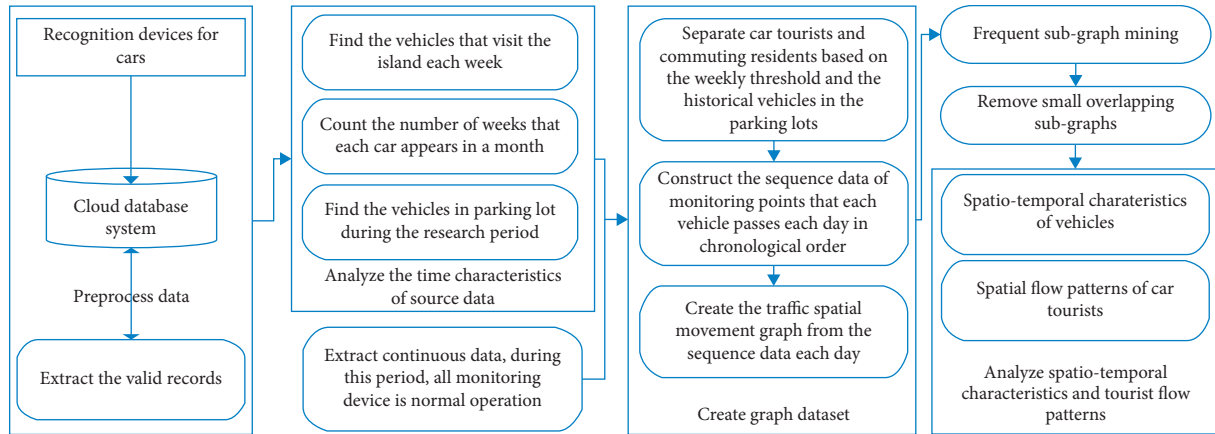
FIGURE 3: The detailed steps for mining frequent patterns in tourist flow.

(1) Deleted irrelevant fields: The primary information such as car license plate number, label of monitoring device, and collection time is retained.

(2) Removed null values and corrected license plate attributions.

(3) Eliminated invalid data, such as special car license plates and duplicate data.

(4) Corrected confusing letters and numbers in car icense plates.

5.2. Identifying Car Tourists.

In this study, tourist flow was divided into three types. One type consisted of commuters on the island, the next consisted of intracity tourists (local weekend tourists in Shenzhen), and the last type consisted of intercity tourists (leisure tourists from outside of Shenzhen). The collected data came from the cameras on the roads and in parking lots. These three types of flows were mixed together in the collected data. It is necessary to separate the different types of traffic flows. The detailed steps are shown in Figure 4.

The data collected from the parking lots were classified into intracity tourists and intercity tourists according to the registration location of car license plates.

As we know, the number of trips made by tourists and local commuters is different. Tourists only visit the island occasionally on weekends or holidays. Local residents on the island might drive more times per week. Therefore, we took a week as a unit and determined if a car had visited the island during that week. If so, this car would be tagged once. Then, we counted the number of weeks a car appeared in each month. If the number of weeks visited exceeded a predefined threshold, this car was considered to be a commuter. Otherwise, this car was considered to be a tourist.

Therefore, for the data collected from roads, we first set a threshold manually based on the statistics of the number of weeks visited in one month to distinguish island commuters and tourists. Next, we categorized visitors into intracity car tourists and intercity car tourists based on where their license plates are registered. Finally, different types of traffic flows were separated and aggregated.

In order to verify the usability and reliability of the proposed method, the visiting characteristics of all cars were analyzed. The result is shown in Figure 5. As can be seen, the proportion of weeks in one month in which the car appears is the highest, reaching 88.86%. The percentage of cars appearing on the island for less than two weeks is 95%. In addition, we counted the percentage of cars in the parking lots relative to the total number of cars. The ratio is 78.6%, which is close to the ratio of 79% counted by Dapeng Transportation Bureau during peak tourist periods. The ratio of the number of weeks visited in one month (88.86%) is greater than the statistical result of Transportation Bureau (79%). We think that the Transportation Bureau only considered the tourists in parking lots. Therefore, the threshold in this study is one week for extracting car tourists.

5.3. Reconstructing the Spatial Movement Graphs.

In order to model the flow patterns of car tourists, labeled direct graphs are used to construct movement relationships between monitoring points. In particular, each vertex of the direct
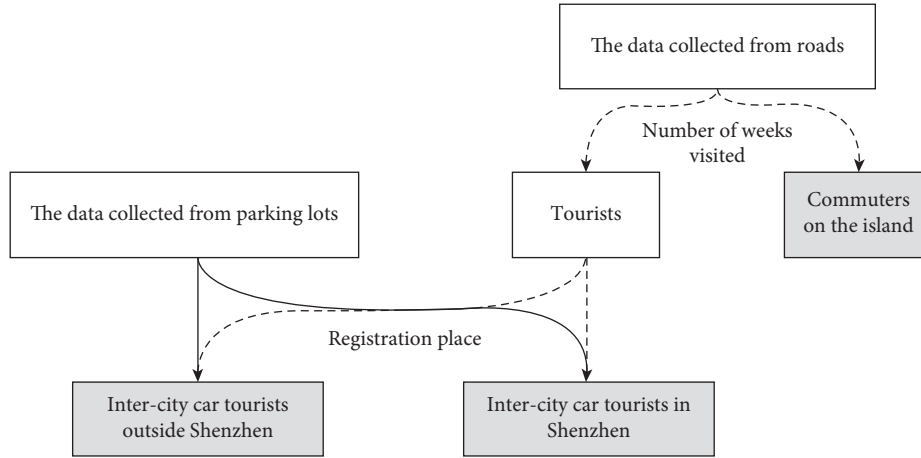
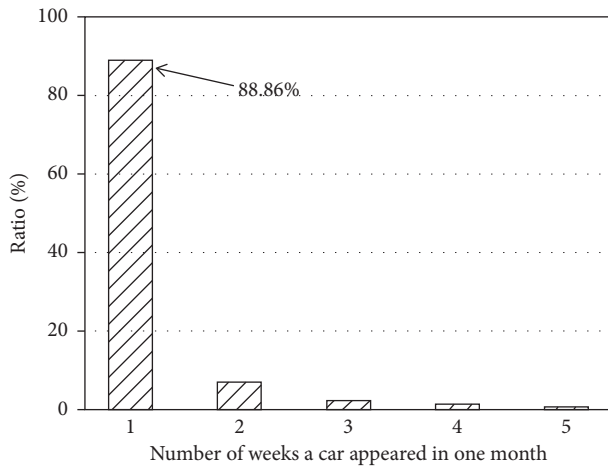Figure 4: The workflow of traffic flow separation.

Figure 5: Number of weeks visited by tourists in one month.

graph corresponds to a monitoring point, and each edge corresponds to a directed connection between two monitoring points passed by car tourists. The related definitions are as follows:

Definition 1. Label Graph. Given a set of vertices $V = \{v_1, v_2, \ldots, v_k\}$, a set of edges connecting two vertex in V, $E = \{e_h = (v_i, v_j)|v_i, v_j \in V\}$, a set of vertex labels $L(V) = \{lb(v_i)|\forall v_i \in V\}$, and a set of edge labels $L(E) = \{lb(e_h)|\forall e_h \in E\}$, e_h is a direct edge that has the start vertex v_i and end vertex v_j, then a label graph G is represented as

$$G = (V, E, L(V), L(E)). \tag{1}$$

In the graph dataset, the label of an edge is represented by a label pair of two monitoring points in the tourist visiting order. A graph consists of edges connecting the monitoring points visited by each tourist in one day. The advantage of using the vertex label pair as an edge label is that it maintains the temporal and spatial order of the two monitoring points that tourists pass through. When mining a labeled graph, the spatial-temporal order of monitoring points in results could be

preserved. Using this representation, the problem of finding frequent flow patterns of car tourists becomes a problem of mining frequent subgraphs in all movement graphs.

5.4. Mining Frequent Subgraphs. Some definitions related to frequent subgraph mining are given below.

Definition 2. Subgraph. A subgraph $g_2 = (V_2, E_2)$ of graph g_1 is a graph in which $V_2 \subseteq V_1$, $E_2 = E_1 \cap (V_2 \times V_1)$

Definition 3. Support of a subgraph g. Given a labeled graph dataset $G_D = \{g_1, g_2, \ldots, g_n\}$, the support or frequency of a subgraph g is the percentage (or number) of graphs in G_D.

Definition 4. Frequent subgraph. A frequent subgraph is a graph whose support is not less than a minimum support threshold. The minimum support threshold represents the minimum number of occurrences of a subgraph. To obtain the frequent patterns, we chose to manually set the value of minimum threshold.

Definition 5. Mini Code. First, a depth-first search is performed on the graph to form a DFS (depth-first search) tree, and then this tree is scanned. The order of the scanned edges constitutes a sequence called the DFS Code. The DFS Codes are sorted in a lexicographic order to find the smallest DFS code that uniquely identifies the graph. This minimum DFS Code is called the mini Code.

After constructing the movement graphs of tourist flow, the subgraph mining method was used to explore the frequent patterns. As demonstrated in related works, there are many kinds of subgraph mining algorithms. The AGM (Apriori-based Graph Mining) [41] can discover all frequent subgraphs (both connected and disconnected) in a graph database that satisfy a specific minimum support constraint. This algorithm uses an approach similar to Apriori, and it requires 40 minutes to 8 days to find the result subgraphs in a dataset containing 300 chemical compounds. The algorithm FSG (finding frequently occurring subgraphs in large graph) [42] adopts an adjacent representation of a graph and an edge-
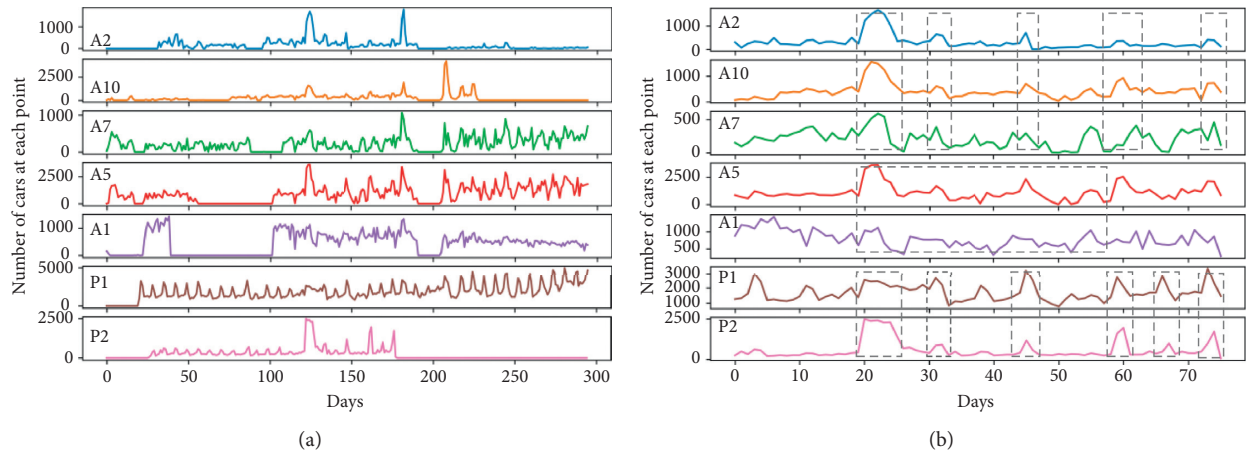
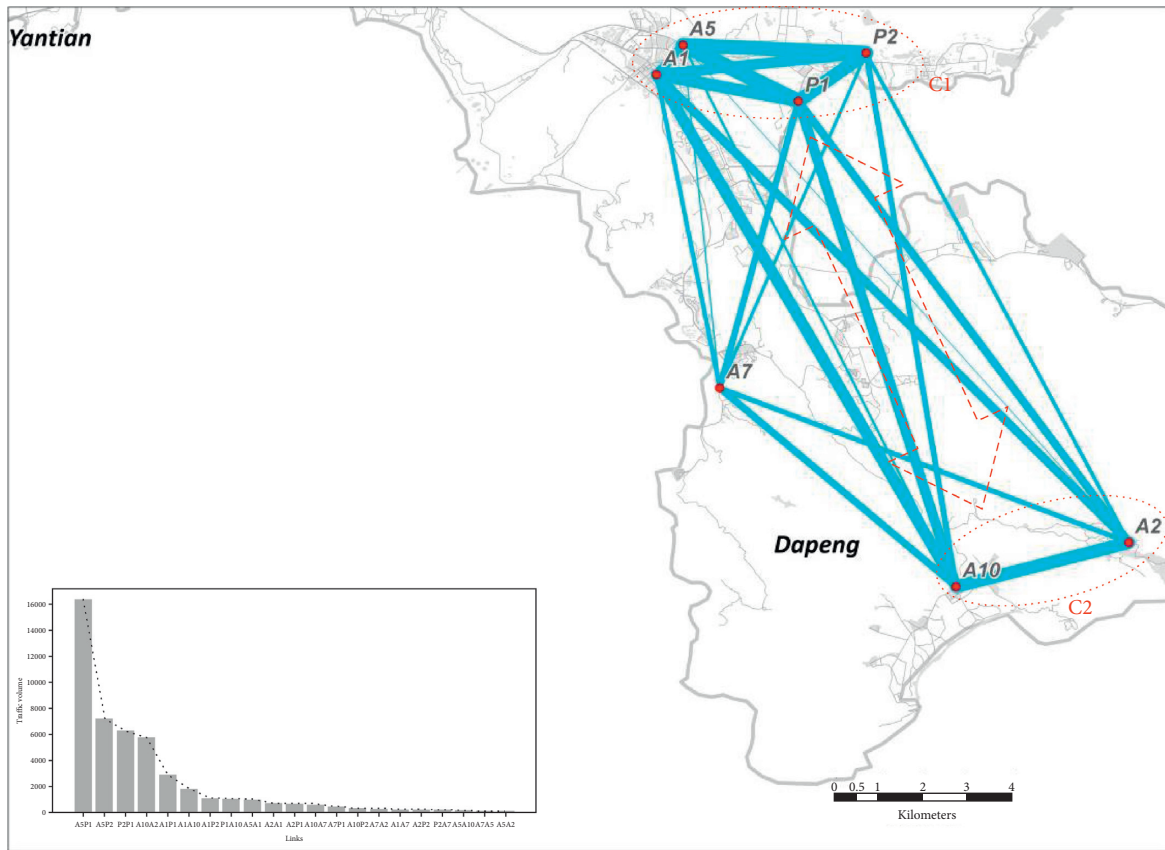Figure 6: Daily traffic flow at each monitoring point.

Figure 7: Flow map of tourism traffic.

Table 2: Types of car tourist visiting the island.

Types	Number of records	Number of graphs
All car tourists	361693	76
Intracity car tourists in Shenzhen	240452	76
Intercity car tourists outside Shenzhen	121241	75

growing strategy to find all of the connected subgraphs that frequently appear in a graph database. The results have shown that FSG can be finished in 600 seconds. gSpan [26] is designed to reduce or avoid the candidate generation and

pruning false positives used in AGM and FSG. gSpan can complete the same task in 10 seconds. Considering the efficiency, in this study, gSpan was used to mine frequent subgraphs in a directed graph dataset and then find the
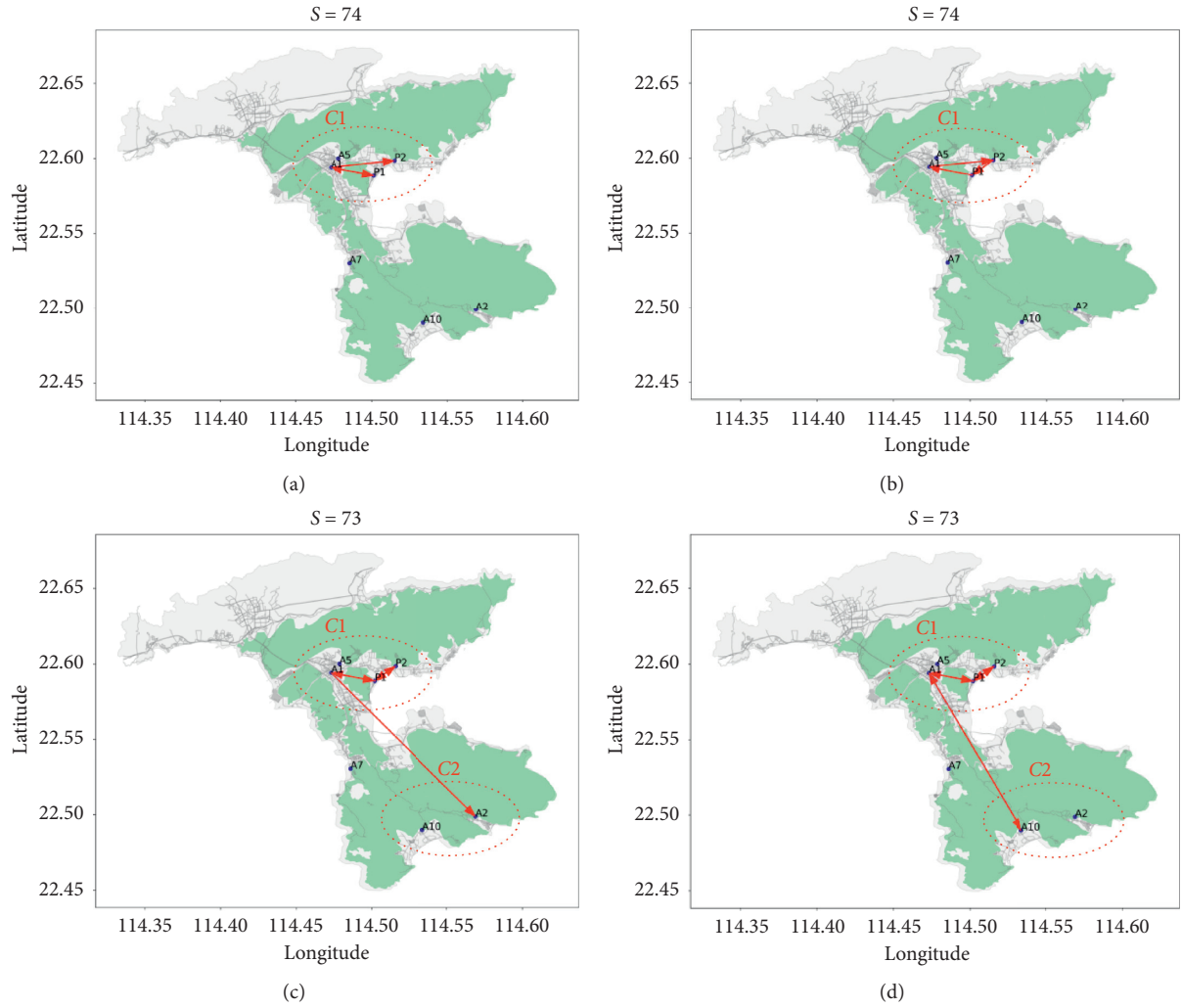
(a)

(b)

(c)

(d)

FIGURE 8: Flow patterns inferred from intracity tourists.

TABLE 3: The list of flow patterns inferred from intracity tourists.

Figures	Patterns
Figure 8(a)	$(A1 \longrightarrow P2$ and $A1 \Leftrightarrow P1)$
Figure 8(b)	$A1 \longrightarrow P2 \longrightarrow P1 \longrightarrow A1$
Figure 8(c)	$A1 \longrightarrow A2$ and $A1 \Leftrightarrow P1 \Leftrightarrow P2$
Figure 8(d)	$A1 \longrightarrow A10$ and $A1 \Leftrightarrow P1 \Leftrightarrow P2$

frequent subgraphs with maximum length in the results as the flow patterns of car tourists. The algorithmic details of gSpan are available in reference [26]. An extended instruction is given below. The method consisted of two steps: (1) Finding the frequent subgraphs using gSpan: First, the frequencies of edges and nodes of all graphs was calculated. Second, the frequencies were compared with the minimum support threshold and the infrequent edges and nodes were removed. Then, the remaining nodes and edges were reordered according to the frequency. And, the frequency of each edge was calculated again. Finally, the subgraphs of the restored graph were mined according to mini Code and it was determined whether the current DFS encoding is the minimum code or not. If so, current edges were added to the results, and

further attempts were made to add possible edges. If not, the mining process was finished. (2) Finding frequent subgraphs with maximum length: There are a large number of subgraphs in the obtained results, and some subgraphs are partial graphs of the others. Therefore, this kind of subgraph was deleted by comparing the labels of nodes and edges, and the final results were the maximum frequent subgraphs.

6. Results

In this section, we first analyzed the spatial and temporal distribution of tourist flows by statistical methods and maps. Then, we divided the tourist flows into intracity car tourists and intercity car tourists and used a frequent subgraph mining algorithm for pattern recognition. Finally, we summarized the movement patterns of all tourists.

6.1. Spatial-Temporal Characteristics of the Tourist Traffic Flow

6.1.1. Temporal Characteristics. Figure 6(a) depicts 295 days of data before any preprocessing was applied. Due to
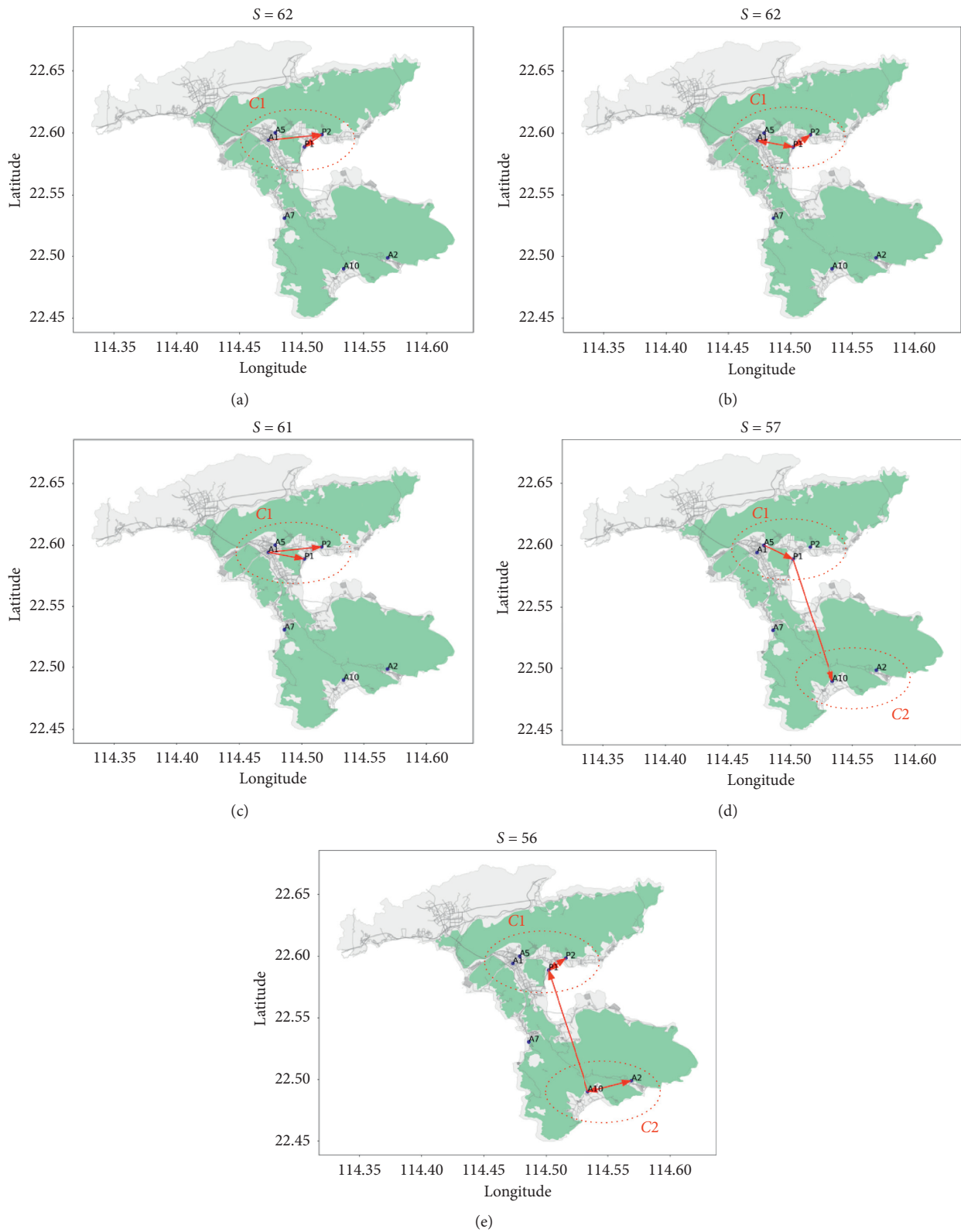
FIGURE 9: Flow patterns inferred from intercity tourists.

Table 4: The list of the flow patterns inferred from intercity tourists.

Figures	Patterns
Figure 9(a)	$A1 \longrightarrow P2 \longrightarrow P1$
Figure 9(b)	$A1 \Leftrightarrow P1 \Leftrightarrow P2$
Figure 9(c)	$A1 \longrightarrow P1$ and $A1 \longrightarrow P2$
Figure 9(d)	$A5 \longrightarrow P1 \longrightarrow A10$
Figure 9(e)	$A10 \longrightarrow P1 \Leftrightarrow P2$ and $A10 \Leftrightarrow A2$

the failure of device communication or power, the constructed movements graphs may be incomplete, which would lead to the loss of frequent subgraphs. Therefore, we selected 76 days of valid data as the dataset for frequent pattern mining. Figure 6(b) shows that (1) the tourist flows at $A2$ and $A10$ near the sandy beaches has similar temporal characteristics. Their peak hours of tourist volume occur both on holidays and weekends, while on weekdays, the flow curves are relatively stable. (2) $P1$ and $P2$ are located in two parking lots. Although there are differences in the tourist volumes, the trends are similar. (3) For the two entrances to the tourist attractions, the average daily volume of tourist traffic at $A1$ is 0.7 times that of $A5$. After sorting the volume of tourist traffic, Pengcheng Community has the most car tourists, followed by two tourist attractions with a sandy beach (i.e., Dongchong and Xichong Community), and the least visited attraction is Nanao Community.

6.1.2. Spatial Characteristics. Figure 7 shows the flow map of the aggregated tourists transferring in multiple monitoring point pairs. The tourist volumes are depicted and sorted in the left bar chart in the figure, and the link thickness represents the traffic volume. As can be seen, the links with the largest traffic volume are $A5\,P1$, $A5\,P2$, $P2\,P1$, $A10\,A2$, $A1\,P1$, and $A1\,P2$ ($A5\,P1$ is a simplified form for $A5 \leftrightarrow P1$, which represents the forth and back tourist flow between $A5$ and $P1$). These links could be divided into two areas, $C1\,(A1, A5, P1, P2)$ and $C2\,(A2, A10)$. Further inspection reveals that the number of tourists transferring between $P1$ and $P2$ in area $C1$ is close to the value between $A2$ and $A10$ in area $C2$, but the number of tourists visiting $P1$ and $P2$ is 3.12 times that of $A2$ and $A10$.

6.2. Analyses of the Detected Flow Patterns. The flow map is intuitive, but it suffers from serious visual clutter, and it is difficult to read because of overlapping flows. It can also be seen in Figure 7 that the flow map only shows the traffic volumes between the monitoring points in the tourist destination. But, it could not express the spatial transfer directions of car tourists. Therefore, these facts motivate us to find a new approach to solve these problems. This section describes the use of frequent subgraph mining algorithm to explore the spatial flow patterns with directions in tourist traffic and to obtain the maximum frequent patterns from the daily tourist movement graphs according to a predefined minimum support threshold. By using the identification method of car tourists introduced in Section 5, two types of

data were extracted and used for the subsequent flow pattern mining (listed in Table 2).

6.2.1. Flow Patterns of Intracity Tourists in Shenzhen. The experiments were conducted with the dataset of intracity car tourists in Shenzhen (as shown in Table 2). The minimum support threshold for gSpan was set to 73. The spatial flow patterns are represented in Figure 8 and listed in Table 3. The inferred patterns could be divided into two groups. One group consists of the patterns shown in Figures 8(a) and 8(b), which show the tourist spatial transfer process in area $C1$. The other group contains the patterns shown Figures 8(c) and 8(d), which represent the tourists transferring between area $C1$ and area $C2$. The two groups demonstrate that the intracity car tourists who arrived at $P1$ and $P2$ preferred to choose Kuinan Road at $A1$ instead of Pengfei Road at $A5$. The difference between Figures 8(a) and 8(b) is the existence of the circle tourist flow. The difference between Figures 8(c) and 8(d) is the presence of tourists flow back and forth. The tourist flow from $A1$ to $A2$ have only one direction. One of the reasons may be that this study failed to find a suitable monitoring point on Pingxi Road, resulting in the loss of directionality for this part of tourist flow.

6.2.2. Flow Patterns of Intercity Car Tourists. Figure 9 shows flow patterns of intercity car tourists. The dataset used in this section is composed of intercity tourists (as shown in Table 2). The minimum support threshold for gSpan was set to 56. The result flow patterns are sorted by the support threshold, as listed in Table 4.

From these patterns, the following could be concluded. (1) The most frequent patterns are shown in Figures 9(a)–9(c). The frequencies of the discovered flow patterns are 62, 62, and 61, respectively. These three patterns describe the preference of intercity tourists for area $C1$. This kind of tourists first visited one of attractions near a parking lot $P1$ or $P2$ and then visited another attraction, or just visited one tourist attraction near $P1$ or $P2$. (2) The above three patterns are different from those of the intracity tourists in Shenzhen. There is no circle tour between $P1$, $P2$, and $A1$. The reason for this is that some tourists chose to continue drive to area $C2$. (3) The minimum support threshold of these two patterns, as shown in Figures 9(d) and 9(e), is significantly smaller than that shown in Figures 9(a)–9(e), but it reflects the spatial transfer preferences of tourists from neighboring cities in areas $C1$ and $C2$. In Figure 9(d), the car tourists tended to drive directly from $A5$ to $A10$ after visiting
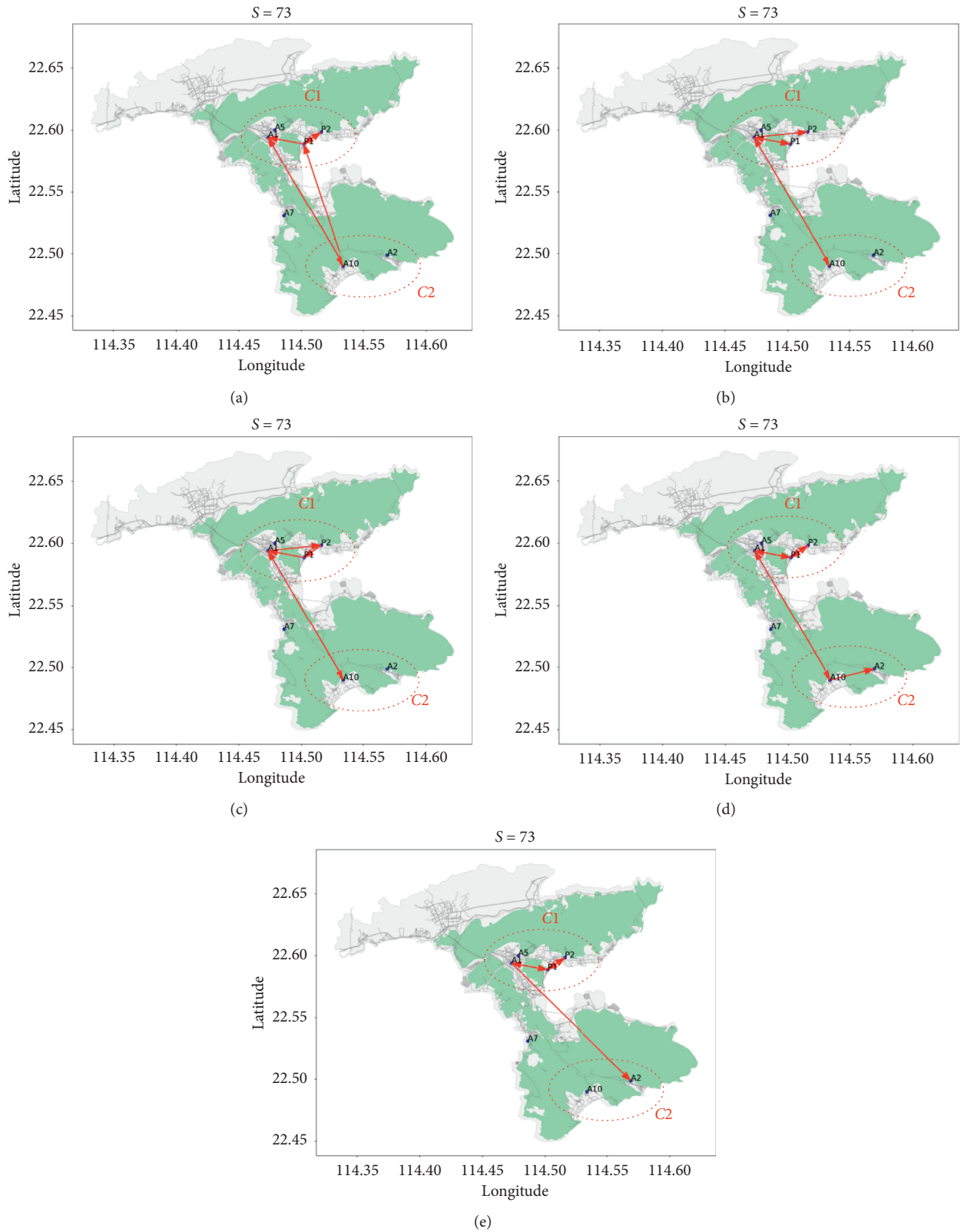
FIGURE 10: Flow patterns of all tourists.

attractions near $P1$. As shown in Figure 9(e), the car tourists that flowed between $A10$ and $A2$ went back to $P1$ and changed car parking lots between $P1$ and $P2$.

However, there is no tourist flow to $A1$ and $A5$. The reason for this may be that the traffic flow of Pingxi Road (the expressway in and out of the peninsula) was not

TABLE 5: The list of flow patterns inferred from all tourists.

Figures	Patterns
Figure 10(a)	$(A1 \longrightarrow A10 \longrightarrow A1)$ and $(A10 \longrightarrow P1 \Leftrightarrow P2)$ and $(P1 \longrightarrow A1)$
Figure 10(b)	$(A1 \Leftrightarrow A10$ and $A1 \longrightarrow P1$ and $A1 \longrightarrow P2)$
Figure 10(c)	$(A1 \Leftrightarrow P10$ and $A1 \longrightarrow P2 \longrightarrow P1 \longrightarrow A1)$
Figure 10(d)	$(A1 \Leftrightarrow P1 \Leftrightarrow P2$ and $A1 \Leftrightarrow A10 \Leftrightarrow A2)$
Figure 10(e)	$(A1 \Leftrightarrow P1 \Leftrightarrow P2$ and $A1 \longrightarrow A2)$

monitored. This part of traffic flow could directly arrive and then leave from $A10$.

6.2.3. Flow Patterns of All Tourists. This section presents an exploration of the spatial flow patterns of all car tourists. The results are shown in Figure 10. The minimum support threshold for gSpan was set to 73, which means that 73 of the 76 graphs contained the discovered flow pattern. The result patterns are listed in Table 5.

All of the patterns shown in Figure 10 have tourist flow between area $C1$ and area $C2$. This is consistent with the trend in the flow map (shown in Figure 7). As the picture shows, $A1$ is the main entrance for tourists to enter and exit Dapeng Island. Some of the car tourists drove to $P1$ or $P2$, and the rest flowed to $A10$ or $A2$. Figures 10(a)–10(c) show the directions of tourist flows between the Dapeng Community, Pengcheng Community, and Xichong Community. Figure 10(d) shows the directions of tourist flows between Dapeng Community, Pengcheng Community, and Dongchong Community. Figure 10(e) represents only the directions of tourist flows between Dapeng Community and Dongchong Community. Furthermore, the directions of tourist flows in area $C1$ and area $C2$ or between area $C1$ and area $C2$ are different. Taking Figures 10(a)–10(c) as examples, although the orders of $P1$ and $P2$ that are accessed from $A1$ have a lack of regularity, they have their own characteristics when considering the accompanying paths to $A10$. These examples indicate that if there is a tourist flow between $A1$ and $A10$, it could be divided into two cases. In one case, some of the tourists returned directly, and in the other case, some of the tourists flowed to $P1$. In the first case, either the tourist flow passing by $A1$ first accessed $P2$ and then visited $P1$, or both $P1$ and $P2$ had tourists at the same time. In the second case, some of the tourists returned from $A10$, visited $P1$ and $P2$, and then left the island.

7. Conclusions and Discussions

7.1. Conclusions. In order to facilitate the management of tourist traffic flow, the car tourists in Dapeng Island were taken as a research case. The experimental analysis used real data captured by video devices in the research area. Due to the lack of suitable device installation locations, the captured picture from the video device had a certain distance from the road. The catch rate of the tourists' cars was low. However, the detailed time-series data in one day could be collected. Compared with the manual survey, the collected data was improved in terms of reliability and richness. A day was chosen as the time unit for frequent pattern mining. After selecting the available data, we divided the data of cars into intracity and intercity tourists. Next, the license plate data were transformed into movement graphs according to the visited location sequence between multiple monitoring points. The intricate flow patterns of the car tourists were discovered by gSpan algorithm, which had the best performance in terms of the quality of the results and the execution time and which had already proven to be efficient for frequent subgraph mining. The conclusions are as follows:

(1) The car tourists had obvious preferences in the selection of trip time and tourist attractions (shown in Figures 6 and 7). In terms of time, the curves of tourist flow at each monitoring point were similar. There were a large number of car tourists at various attractions on holidays, but the volume of car tourists was relatively lower on weekdays. The same types of attractions had the similar trends in tourist flow, such as the two attractions with a sandy beach (Dongchong Community and Xichong Community) and the ancient city and cultural attractions (Dapeng Ancient City and Dongshan Temple). In terms of space, the attractions that are close to the entrance of a scenic area and rich in tourist resources were more popular with tourists. However, due to the terrain barrier in the scenic area, the traffic conditions affected the movements of tourists between multiple attractions.

(2) Different types of car tourists had similar spatial choices in scenic area (shown in Figures 8–10). For example, different types of tourists had flow patterns that described the movements in one area (as shown in Figures 8(a), 8(b) and 9(a)–9(c)) and the movements between different areas (as shown in Figures 8(c), 8(d), 9(d), and 9(e)). The intercity tourists and intracity tourists had different choices in scenic spots. The intercity tourists would take multidestination trips instead of single destination trips in the same type of attractions. As can be seen in Figures 8 and 9, there are the two sandy beach attractions in area $C2$, the intracity tourists tend to visit one of them, while the intercity tourists would visit both attractions. Specifically, to save time and money, intercity tourists would visit multiple attractions in one trip instead of making multiple trips. Additionally, another difference was that there was no circlet in area $C1$ for intercity tourists and no traffic flow between $A2$ and $A10$ for intracity tourists.

(3) Although the patterns depicted on the map look complicated and messy, after the patterns are converted to rules, they become clear. In pattern maps, only Figures 8(b), 9(a), and 9(d) show a clear unidirectionality. The rest is complex and is difficult to compare. As we can see from Figure 8(a), the intracity tourists passing point $A1$ could be divided into two groups, one group flowed to parking lot $P1$, and then left the scenic area from $P1$. The other group flowed to parking lot $P2$. Two groups have simultaneity in tourist routes. So, we could convert the flow patterns to rules and use "and" to illustrate the simultaneity of different routes in the same pattern (as shown in Tables 3–5). By this way, all patterns can be applied to traffic control system for regional tourism.

(4) Large primary attractions are more attractive than smaller secondary attractions. Looking from the arrow on the pattern maps, tourists always visit area $C1$ first and then select area $C2$. The main reason is that area $C1$ has abundant tourism resources and diverse tourism activities. For example, there are cultural attractions (e.g., Dapeng Ancient City and Dongshan Temple), sandy beach entertainments, and large parking lots for tourists in area $C1$. These factors are also often considered in the evaluation of the importance of attractions in tourism network.

(5) The tourists tended to park their cars in an easy-to-access place, even if the visited attractions are changed, as shown in Figures 8(a)–8(d), 9(b), and 9(e). Again, here we take area $C1$ as an example. The distance between the Dapeng Ancient City and Dongshan Temple is about 1 km. On the pattern maps, we can see that there are two-way arrows pointing to the two parking lots $P1$ and $P2$. This indicates that the car tourists tended to park their cars to the nearest parking lot, so that they could pick them up when the tour destination is changed.

7.2. Discussions. Tourist flow is the key to traffic management in tourism destinations, and it affects the development of tourism on an island and the experience of tourists. The recent development of transport technologies has shown that traffic flow data will be increasingly collected and it will be available for data analysis. Therefore, advanced data analytics should be used to interpret and depict the complex movements of car tourists. The proposed approach in this study was intended to find (i) the statistical summaries of the spatial-temporal characteristics of car tourists in the research area, helping to discover patterns from the mass car license plate data; (ii) the flow patterns of intercity and intracity tourists, helping to illustrate the different preferences of the two types of car tourists; and (iii) the most frequent patterns of all tourists, helping to identify the law of tourist movement and make efficient policy for the management of tourist traffic flow. The presented approach enriched the analytical methodology of tourist

traffic flow and suggested a shift from the conventional and complicated paper or computer interview-based method to a dynamic flow graph-based method. Furthermore, we have shown how transportation data provides hard-to-obtain insights and quantitative results for tourists.

In order to illustrate the law of spatial movement of tourists, related studies have proposed a variety of macro flow patterns [43–45]. For example, in 2008, McKercher [46] proposed 11 prominent route styles in urban destinations. The macropatterns retain only the main components and simplify the details and are often used in tourism management to guide destination development. Compared with modeling the flow patterns of tourists at the macrolevel, modeling tourist flow at the microscale is more complicated. Lew and McKercher [17] noted that it is a challenge to balance model effectiveness and usability. The reason is that simple patterns may not provide enough details for use and complex patterns may be difficult to interpret and apply. In this study, we used the video devices installed at key nodes of road network to collect tourist traffic flows and used the frequent subgraph mining algorithm to discover flow patterns at the microscale. The extracted patterns can be converted into rules and applied to the traffic control system for the management of regional tourists. The difficulty of finding and applying patterns at microscale could be overcome by this way.

During the peak period of tourism, the number of car tourists in the scenic area increases sharply. It is easy to result in road congestion and uneven distribution of tourists between attractions. With the use of traffic flow data, the daily, monthly, and seasonal characteristics of tourists can be analyzed, the future tourist flow can be predicted, and the flow patterns can be obtained by data mining methods. Thus, the traffic management department can effectively control the tourist traffic flow, and the tourism department can develop attractive tourism products to achieve a spatial balance of the distribution of tourists, reduce traffic congestion and harmful gas emissions and weaken the impact of tourism environment and human body.

Inside a scenic area, attractions form a complex network of tourist flow due to the frequent spatial interaction. Each attraction is both the origin and destination of car tourists. Due to the differences in attractiveness, degree of development, and convenience of transportation, tourists have shown special preferences when choosing attractions and trip routes. Accordingly, effective identification of these preferences will be beneficial to the development of tourist market and will also helpful for tour planners to understand how tourists see the spatial connection of multiple attractions. However, traditional manual surveys are time-consuming and laborious, and the amount of data obtained is small. It is difficult to reveal tourist preferences by this way. Frequent subgraph mining algorithms provide a desirable method for identifying the spatial preferences of tourists. This kind of method could perform well under the support of a large amount of movement data between attractions.

This study has several limitations. Firstly, due to lack of power supply facilities, it was unable to collect the traffic data for each day. When applying the model to the actual control of tourist traffic flow, it is necessary to further co-operate with the traffic management department to obtain comprehensive traffic flow data. Secondly, this study focused on the mining of flow patterns, the influencing factors behind the identified patterns were not further analyzed. As mentioned by Lew and McKercher [17], factors related to tourists and destinations can affect how tourists move or travel in a destination. These factors include family composition, income, and valid information obtained before travelling. It is difficult to obtain these factors by relying only on the traffic flow data collected in the traffic monitoring system. Therefore, in future work, field investigation will be necessary.

Data Availability

As the data also form part of an ongoing study, the raw data needed to reproduce these findings cannot be shared at this time.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] D. W. Eby and L. J. Molnar, "Importance of scenic byways in route choice: a survey of driving tourists in the United States," *Transportation Research, Part A (Policy and Practice)*, vol. 36, no. 2, pp. 95–106, 2003.

[2] B. Wang, Z. Yang, F. Han, and H. Shi, "Car tourism in xinjiang: the mediation effect of perceived value and tourist satisfaction on the relationship between destination image and loyalty," *Sustainability*, vol. 9, no. 1, pp. 1–16, 2016.

[3] X. M. Zhang and Q. Y. Pen, "A study on tourism information and self-driving tour traffic," in *Proceedings of the First International Conference on Transportation Engineering*, Southwest Jiaotong University, Chengdu, China, July 2007.

[4] China National Tourism Administration, *The Yearbook of China Tourism Statistics 2016*, China Travel & Tourism Press, Beijing, China, 2017.

[5] M. A. Mohamad Toha and H. N. Ismail, "A heritage tourism and tourist flow pattern: a perspective on traditional versus modern technologies in tracking the tourists," *International Journal of Built Environment and Sustainability*, vol. 2, no. 2, pp. 85–92, 2015.

[6] K. Siła-Nowicka and A. S. Fotheringham, "Calibrating spatial interaction models from gps tracking data: an example of retail behaviour," *Computers, Environment and Urban Systems*, vol. 74, pp. 136–150, 2019.

[7] F. Mao, M. Ji, and T. Liu, "Mining spatiotemporal patterns of urban dwellers from taxi trajectory data," *Frontiers of Earth Science*, vol. 10, no. 2, pp. 205–221, 2016.

[8] L. J. Zheng, D. Xia, X. Zhao, and W. L. Liu, "Mining trip attractive areas using large-scale taxi trajectory data," in *Proceedings of the 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications*, pp. 1217–1222, ISPA/IUCC, Guangzhou, China, December 2017.

[9] I. Rami and M. O. Shafiq, "Detecting taxi movements using random swap clustering and sequential pattern mining," *Journal of Big Data*, vol. 6, no. 39, pp. 1–26, 2019.

[10] G. Kautsar and S. Akbar, "Trajectory pattern mining using sequential pattern mining and k-means for predicting future location," *Journal of Physics: Conference Series*, vol. 801, no. 1, 2017.

[11] D. Guo, X. Zhu, H. Jin, P. Gao, and C. Andris, "Discovering spatial patterns in origin-destination mobility data," *Transactions in GIS*, vol. 16, no. 3, pp. 411–429, 2012.

[12] J. D. Mazimpaka and S. Timpf, "A visual and computational analysis approach for exploring significant locations and time periods along a bus route," in *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, pp. 43–48, San Francisco, CA, USA, October 2016.

[13] H. Faroqi, M. Mesbah, and J. Kim, "Spatial-temporal similarity correlation between public transit passengers using smart card data," *Journal of Advanced Transportation*, vol. 2017, Article ID 1318945, 14 pages, 2017.

[14] Z. Wang, Y. Hu, P. Zhu, Y. Qin, and L. Jia, "Ring aggregation pattern of metro passenger trips: a study using smart card data," *Physica A: Statistical Mechanics and Its Applications*, vol. 491, pp. 471–479, 2018.

[15] X. Zhao, X. Lu, Y. Liu, J. Lin, and J. An, "Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: a case study of Xi'an, China," *Tourism Management*, vol. 69, pp. 368–383, 2018.

[16] M. Versichele, L. De Groote, M. Claeys Bouuaert, T. Neutens, I. Moerman, and N. Van De Weghe, "Pattern mining in tourist attraction visits through association rule learning on bluetooth tracking data: a case study of ghent, Belgium," *Tourism Management*, vol. 44, pp. 67–81, 2014.

[17] A. Lew and B. Mckercher, "Modeling tourist movements," *Annals of Tourism Research*, vol. 33, no. 2, pp. 403–423, 2006.

[18] A. Birenboim, S. Anton-Clavé, A. P. Russo, and N. Shoval, "Temporal activity patterns of theme park visitors," *Tourism Geographies*, vol. 15, no. 4, pp. 601–619, 2013.

[19] H. J. Yun and M. H. Park, "Time-space movement of festival visitors in rural areas using a smart phone application," *Asia Pacific Journal of Tourism Research*, vol. 20, no. 11, pp. 1246–1265, 2015.

[20] S. Kang, "Associations between space-time constraints and spatial patterns of travels," *Annals of Tourism Research*, vol. 61, pp. 127–141, 2016.

[21] C. Comito, D. Falcone, and D. Talia, "Mining human mobility patterns from social geo-tagged data," *Pervasive and Mobile Computing*, vol. 33, pp. 91–107, 2016.

[22] Y. T. Zheng, Y. Li, Z. J. Zha, and T. S. Chua, "Mining travel patterns from GPS-tagged photos," in *Lecture Notes in Computer Science*, K. T. Lee, W. H. Tsai, H. Y. M. Liao, T. Chen, J. W. Hsieh, and C. C. Tseng, Eds., Springer, Berlin, Germany, 2011.

[23] K. M. Borgwardt, H. P. Kriegel, and P. Wackersreuther, "Pattern mining in frequent dynamic subgraphs," in *Proceedings of the 6th IEEE International Conference on Data*

Mining (ICDM 2006), Hong Kong, China. IEEE, December 2006.

[24] D. J. Cook and L. B. Holder, *Mining Graph Data*, Wiley Press, Hoboken, NJ, USA, 2015.

[25] C. Jiang, F. Coenen, and M. Zito, "Frequent sub-graph mining on edge weighted graphs," in *Lecture Notes in Computer Science*, T. Bach Pedersen, M. K. Mohania, and A. M. Tjoa, Eds., Springer, Berlin, Germany, pp. 77–88, 2010.

[26] X. Yan and J. Han, "gSpan: graph-based substructure pattern mining," in *Proceedings of the 2002 IEEE International Conference on Data Mining, ICDM'02*, pp. 721–724, IEEE Computer Society, Maebashi City, Japan, December 2002.

[27] C. Borgelt and M. R. Berthold, "Mining molecular fragments: finding relevant substructures of molecules," in *Proceedings of the 2002 IEEE International Conference on Data Mining, ICDM'02*, pp. 51–58, IEEE Computer Society, Maebashi City, Japan, December 2002.

[28] J. Huan, W. Wang, and J. Prins, "Efficient mining of frequent subgraphs in the presence of isomorphism," in *Proceedings of the Third IEEE International Conference on Data Mining, ICDM'03*, pp. 549–552, IEEE Computer Society, Melbourne, FL, USA, November 2003.

[29] S. Nijssen, "A quickstart in frequent structure mining can make a difference," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'04*, pp. 647–652, ACM, Seattle, WA, USA, August 2004.

[30] J. Huan, W. Wang, J. Prins, and J. Yang, "SPIN: mining maximal frequent subgraphs from graph databases," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, USA, ACM, August 2004.

[31] M. Wörlein, T. Meinl, I. Fischer, and M. &Philippsen, "A quantitative comparison of the subgraph miners MoFa, gSpan, FFSM, and gaston," in *Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases PKDD'05*, Springer, Porto, Portugal, pp. 392–403, 2005.

[32] X. Yan, H. Cheng, J. Han, and P. &Yu, "Mining significant graph patterns by leap search," in *Proceedings of the 2008 ACM SIGMOD*, pp. 433–444, Vancouver Canada, June 2008.

[33] L. T. Thomas, S. R. Valluri, and K. Karlapalem, "Margin: maximal frequent subgraph mining," in *Proceedings of the Sixth International Conference on Data Mining (ICDM'06)*, Hong Kong, China, December 2006.

[34] V. Ingalalli, D. Ienco, and P. Poncelet, "Mining frequent subgraphs in multigraphs," *Information Sciences*, vol. 451-452, pp. 50–66, 2018.

[35] L. Min, Z. C. Wang, J. Liang, and X. R. Yuan, "OD-wheel: visual design to explore OD patterns of a central region," in *Proceedings of the Visualization Symposium 2015*, Hangzhou, China, April 2015.

[36] B. Jenny, D. M. Stephen, I. Muehlenhaus, and B. E. Marston, "Design principles for origin-destination flowmaps," *Cartography and Geographic Information Science*, vol. 45, no. 1, pp. 1–15, 2016.

[37] S. Kim, S. Jeong, I. Woo, Y. Jang, R. Maciejewski, and D. S. Ebert, "Data flow analysis and visualization for spatiotemporal statistical data without trajectory information," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 3, pp. 1287–1300, 2018.

[38] J. Wood, J. Dykes, and A. Slingsby, "Visualisation of origins, destinations and flows with OD maps," *The Cartographic Journal*, vol. 47, no. 2, pp. 117–129, 2010.

[39] M. Zillinger, "Tourist routes: a time-geographical approach on German car-tourists in Sweden," *Tourism Geographies*, vol. 9, no. 1, pp. 64–83, 2007.

[40] S. Page and J. Connell, "Exploring the spatial patterns of car-based tourist travel in loch lomond and trossachs national park, Scotland," *Tourism Management*, vol. 29, no. 3, pp. 561–580, 2008.

[41] A. Inokuchi, T. Washio, and H. &Motoda, "An apriori-based algorithm for mining frequent substructures from graph data," in *Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, PKDD'00*, pp. 13–23, Lyon, France, September 2000.

[42] M. Kuramochi and G. Karypis, "Frequent subgraph discovery," in *Proceedings of the 2001 IEEE International Conference on Data Mining, ICDM'01*, pp. 313–320, IEEE Computer Society, San Jose, CA, USA, August 2001.

[43] M. Oppermann, "A model of travel itineraries," *Journal of Travel Research*, vol. 33, no. 4, pp. 57–61, 1995.

[44] N. Shoval and A. Raveh, "Categorization of tourist attractions and the modeling of tourist cities: based on the co-plot method of multivariate analysis," *Tourism Management*, vol. 25, no. 6, pp. 741–750, 2004.

[45] G. Lau and B. McKercher, "Understanding tourist movement patterns in a destination: a gis approach," *Tourism and Hospitality Research*, vol. 9, no. 1, pp. 39–49, 2007.

[46] B. McKercher and G. Lau, "Movement patterns of tourists within a destination," *Tourism Geographies*, vol. 10, no. 3, pp. 355–374, 2008.