This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

COOC: Visual Exploration of Co-Occurrence Mobility Patterns in Urban Scenarios

Xiangjie Kong, *Senior Member, IEEE*, Menglin Li, Gaoxing Zhao, Huijie Zhang, *Member, IEEE*, and Feng Xia, *Senior Member, IEEE*

***Abstract*—A co-occurrence pattern is an interesting pattern in human mobility, which has essential values in business intelligence, social activities, and urban planning. However, due to the deluge and complexity of mobile big data, as well as the complicated intrinsic features of the co-occurrence pattern, mining and analyzing the co-occurrence pattern are computationally highly expensive. Therefore, in this paper, we propose a framework to mine co-occurrence event data from mobile data and to explore the urban co-occurrence pattern visually. Our framework contains two modules: data modeling, to obtain the co-occurrence event data effectively utilizing frequent itemsets mining algorithm based on traffic GPS records, and visualization, to explore the co-occurrence pattern in urban scenarios from global, regional, statistical, and location perspectives. Our visualization system has been demonstrated using case studies with a real-world data set.**

***Index Terms*—Co-occurrence pattern, human mobility, spatiotemporal data, visualization.**

I. Introduction

THE rapid popularization of wireless communication infrastructure and high-speed advancements in data acquisition technologies lead to the explosive growth in the size and variety of data, including a large number of data which directly track human trajectories such as telecommunication data, and various crowdsourced data that hide residents' tracks, such as smart card data [1], [2]. The analysis of human mobility attracts the research interest of scholars mainly due to its extensive applications in numerous fields, especially in urban computing, such as passenger flow prediction [3], [4], route planning, functional regions mining, and abnormal traffic events detection [5]. Human mobility contains diverse interesting subpatterns, such as black holes, which have the overall inflow greater than the overall outflow, volcanos, which have the overall outflow greater than the overall inflow [6], and lark patterns, which refer to people used to going out early. In this paper, we focus on co-occurrence patterns, which denote people from two regions visit an urban region within a certain time period. It emphasizes regional relations.

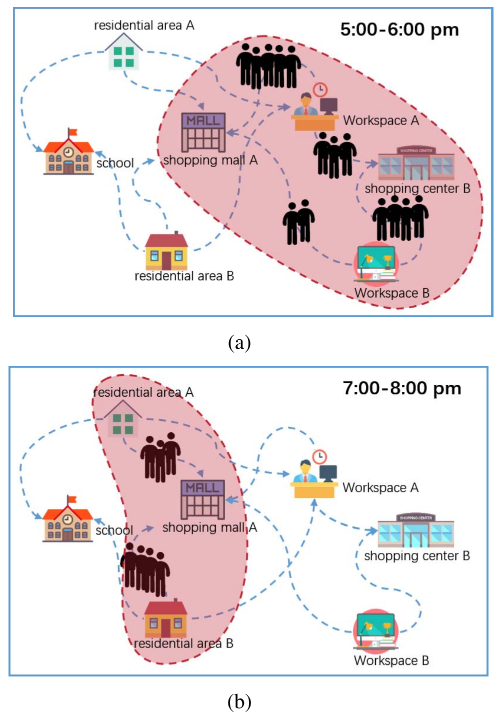


Fig. 1. Examples of co-occurrence patterns in the same urban area at different time periods. (a) Co-occurrence pattern at 5:00–6:00 P.M. (b) Co-occurrence pattern at 7:00–8:00 P.M.

In order to better understand co-occurrence patterns, we present two examples shown in Fig. 1. In Fig. 1(a), five people from workspace A and two people from workspace B visit shopping mall A, respectively, from 5:00 P.M. to 6:00 P.M. Three people from workspace A and four people from workspace B go to shopping mall B, respectively, in the same time period. Therefore, we say that a co-occurrence event between workspace A and workspace B occurs in shopping mall A as well as in shopping mall B. Similarly, from 7 p.m. to 8 p.m. in the evening, three people from residential area A and four people from residential area B visit shopping mall A. A co-occurrence event between residential area A and residential area B occurs in shopping mall A.

Based on analyzing urban co-occurrence patterns, we can infer interesting temporal and spatial information, such as regions that people from the same functional regions prefer to co-occurring, and regions that people (who visit certain regions frequently) come from [7]. Therefore, shop owners can carry out targeted promotional activities according to the

Manuscript received May 31, 2018; revised September 14, 2018; accepted November 19, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61572106 and Grant 41671379, in part by the Dalian Science and Technology Innovation Fund under Grant 2018J12GX048, and in part by the Fundamental Research Funds for the Central Universities under Grant DUT18JC09. *(Corresponding author: Feng Xia.)*

X. Kong, M. Li, G. Zhao, and F. Xia are with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China (e-mail: xjkong@ieee.org; cookies.s@outlook.com; zhaogaoxing_fh@126.com; f.xia@ieee.org).

H. Zhang is with the School of Information Science and Technology, Northeast Normal University, Changchun 130032, China (e-mail: zhanghj167@nenu.edu.cn).

Digital Object Identifier 10.1109/TCSS.2018.2883582

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

types of groups that go shopping at special time periods, such as young office staffers or seniors. By studying empirical networks of human contact from co-occurrence patterns, an infectious disease model can be built to pave the way for controlling contagious disease transmission based on big data [8]. Benefiting from the characteristic that co-occurrence patterns are more sensitive to crowd assembling, urban co-occurrence pattern-oriented anomaly detection is a quite potential research direction [6]. In other words, such patterns have essential values in business intelligence, social activities [9], [10], and urban planning [11].

Whether people visit a region depends on various factors [12]. For distance, the possibility of choosing a near region is higher than that of a distant region. Weather also has an impact on travel; people tend to stay home if the weather is bad. In terms of occupations, freelancers avoid the stress of commuting time. Personal preferences are unignored, and people tend to shop in shopping malls in their favorite stores. Human mobility brings comprehensive results of interactions between multiple factors including regional functions, interregion distances, and personal preferences [13]. In other words, human mobility is a complex sequential transition regularity with a high degree of time dependence and spatial dependence [14]. Its time dependence manifests itself not only in the changes of human mobility over time but also in multilevel periodicities, weekly or monthly. In addition, trajectory data usually contain a variety of attributes, such as spatial property, temporal property, and even semantic property [15]. The data have strong heterogeneity and embrace sparse trace information. Therefore, the analysis of co-occurrence patterns based on the trajectory data is extremely challenging.

Therefore, we seek the help of visualization techniques to reduce the difficulty of the co-occurrence analysis [16]. Data visualization, as the name suggests, is to present data in an intuitive form. In recent years, with the fiery research of big data, the superiority of visualization in analyzing data has been discovered and widely used. Visualization can merge machine intelligence and human wisdom, to help extract truly useful information from overwhelming data. Hence, in this paper, we propose a visual system COOC to explore co-occurrence patterns from trajectory data.

The major contributions of this paper can be summarized as follows.

1) We propose a co-occurrence event data mining scheme utilizing frequent itemsets mining algorithm based on traffic GPS data.
2) We design a novel visual form to display the global co-occurrence pattern and integrate it with other state-of-the-art visual techniques to explore the urban co-occurrence pattern.
3) We demonstrate that our framework can help users get interesting insights and effective analysis from multiple perspectives based on a real-world data set.

II. Related Work

We provide an overview of related studies in this section. Two most relevant topics, co-occurrence pattern mining and visualization of many-to-many relations, are in focus.

A. Co-Occurrence Pattern Mining

Co-occurrence patterns show great value in various fields, including biological symbiosis [17], mobile phone user application mode [18], and computer vision [19]. In this paper, we focus on spatial–temporal co-occurrence pattern in urban human mobility, which represents subsets of event types that occur together in both space and time. The computational expensiveness and the excessive data size cause the great resistance of the co-occurrence pattern analysis. So, a monotonic composite interest measure for discovering mixed-drove spatiotemporal co-occurrence pattern (MDCOP) and a novel MDCOP mining algorithm are proposed to improve the computational efficiency [20], [21]. Pillai *et al.* [22] present a general framework to identify the spatiotemporal co-occurrence patterns for continuously evolving spatiotemporal events that have polygon-like representations. Specifically designed spatiotemporal indexing techniques are utilized to mine the co-occurrence pattern from spatiotemporal data sets [23]. Hong *et al.* [6] propose a two-step black hole detection algorithm based on a well-designed spatiotemporal graph index. Machens *et al.* [8] build an infectious disease model on empirical networks of human contact to bridge the gap between dynamic network data and contact matrices. Akbari *et al.* [24] design a method to extract implicitly contained spatial relationships algorithmically, to deal with different feature types that are with points, lines, and polygon data, to mine the spatiotemporal co-occurrence pattern in space and time, and they apply this method on a real case study for air pollution.

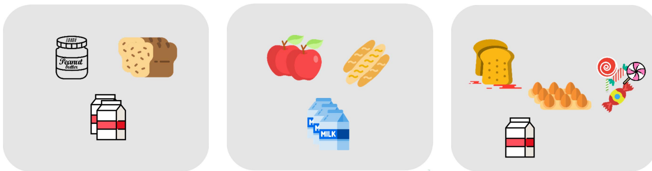
A co-location pattern is highly similar but different from the co-occurrence pattern. It represents subsets of spatial features whose instances are often located at close spatial proximity [25]. A spatial co-location rule issue is different from the traditional association rule because there is no natural notion of transactions in spatial data sets embedded in continuous geographic space. Using the concept of proximity neighborhood, Huang *et al.* [26], [27] provide a transaction-free approach to mine the colocation pattern, and they address the problem of mining the co-location pattern with rare spatial features.

In summary, co-occurrence pattern mining is challenging. Due to its nature of interdisciplinary integration, association relationship mining algorithms, visualization techniques, and many other methods are applied in research on co-occurrence patterns. In this paper, we understand the co-occurrence pattern in human mobility from frequent patterns (FP) and mine co-occurrence event data effectively utilizing classical frequent itemsets mining algorithms.

B. Visualization of Many-to-Many Relations

In the previous examples of the co-occurrence pattern, we can see that a co-occurrence event is the relation of two regions and one region. Many co-occurrence events form co-occurrence patterns, and the relation turns into a complex relation between multiple regions and multiple regions. Researches on how to visualize the relation between multiple entities and multiple entities have arisen, which

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (b) (c)

Fig. 2. Example of checking out at a supermarket. (a) Your basket. (b) Couple's basket. (c) Woman's basket.

provide the technical support for our work. Yang *et al.* [28] design a new visualization called MapTrix to illustrate the dense many-to-many flow and compare its effect with a bundled node-link flow map representation and OD maps. An interactive visual analytics system called TelCoVis is presented to help analysts leverage their domain knowledge to gain insight into the co-occurrence in urban human mobility based on telco data [11]. Biclusters represent two sets of related entities with close relationships, and we regard bicluster as a special kind of co-occurrence pattern. A five-level design framework for bicluster visualizations is proposed to provide a potential solution to ease the process of exploring and identifying coordinated relationships (e.g., four people who visited the same five cities on the same set of days) within some large data sets [29]. Zhao *et al.* [30] propose a visualization technique, named BiDots, which overcomes several limitations by encoding biclusters in a more compact and cluster-driven manner to allow analysts to interactively explore biclusters over multiple domains.

Based on the research of visualizing biclusters and many-to-many relation, we design a novel visual form, circular view, to present the global co-occurrence pattern intuitively and combine it with other state-of-the-art visual techniques to explore the urban co-occurrence pattern from multiple perspectives.

III. Background

There are various facts about the origin of co-occurrence patterns. Some scholars think that the co-occurrence pattern is a variant of the co-location pattern [25]. In this paper, we understand and mine co-occurrence patterns in human mobility from the perspective of FPs.

Assume that you are lined up to check out at a supermarket, and there are milk, bread, and butter in your shopping basket, as shown in Fig. 2(a). Then, you notice that the couple in front of you is paying for milk, bread, and cereals, as shown in Fig. 2(b), and the shopping basket of the woman behind you contains milk, bread, sugar, and eggs, as shown in Fig. 2(c). Furthermore, you find that the three baskets all contain milk and bread. Does this mean that if customers buy milk, they are likely to buy bread at the same time? The relationship between goods in shopping baskets, which reflects customers' shopping habit, is exactly the basis to make good marketing strategies for retailers.

In FPs, each entity is an item, and a set of items form an itemset. A frequent itemset is an itemset that appears in the transaction data set frequently. For example, the set of milk and bread forms a frequent itemset, which appears in the above shopping baskets example frequently. Just as the shopping baskets example reveals the relationship between items that appear in the same shopping baskets at the same time, we focus on the relationship between regions which appear in the same region at the same time. (The statement that region A appears in region B means that there is a travel from region A to region B.) The research entity is exactly a region, so the itemset is the set of regions, and the frequent itemset is the set of regions that frequently appear in the same region. So, from the perspective of human mobility, we define a co-occurrence event as follows.

If people from a set of regions R_a visit another set of regions R_b at a certain time period frequently, we say that "regions in R_a co-occurs with each other at R_b."

Then, how do we determine if an event is frequent enough? Support is a primary and important index for FPs, and it forms the basic support-confidence framework together with confidence to measure the interest of the rules to reflect their usefulness and certainty. Support is defined as the frequency of an itemset in the transaction data set. In co-occurrence patterns, we define the support of an event as the number of destinations where the set of regions R_a visit at a certain time period, that is, the number of regions in set R_b.

Then, we need to set a minimum support minsup. Just as only the itemset that satisfies minsup is a frequent itemset, only an event that meets minsup is a co-occurrence event. We assume that R is a frequent itemset. If there is no set larger than R, which has the same support as R in the transaction data set, then R is closed and called a closed frequent itemset. Similarly, if there is no set larger than R, which is frequent in the transaction data set, then R is a maximal frequent itemset. The statement that set A is larger than set B denotes that B is a subset of A and at least one item in A does not belong to B. The difference of closed frequent itemset and maximal frequent itemset is that the support of all subsets of a closed frequent itemset is datum and is the same as that of the closed frequent itemset, while the maximal frequent itemset only guarantees that its all subsets are frequent. The process of mining FP is actually looking for all closed frequent itemsets because closed frequent itemsets store all the information of FP. An important property of frequent itemsets, which is worth paying attention to, is transcendental nature, which refers to that all nonempty subsets of frequent itemsets are frequent. This property is widely used in FP mining.

In the FP analysis based on the shopping baskets example, researchers only focus on bread and milk appearing in shopping baskets and the number of goods being ignored. In Fig. 1, we can see that the quantitative relationship among regions, such as two people and five people, is crucial to the analysis of co-occurrence patterns in human mobility and is unignored. So, we provide a detailed definition of a co-occurrence event as follows.

If a set of people Flow $= \{f_{11}, f_{12}, \ldots, f_{1m}, f_{21}, \ldots, f_{mn}\}$ from a set of regions $R_a = \{a_1, a_2, a_3, \ldots, a_m\}$ visits another set of regions $R_b = \{b_1, b_2, b_3, \ldots, b_n\}$ at a certain time, where $n \geq$ minsup, we can say that "regions in R_a co-occurs with each other at R_b."

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

IV. System Architecture

With a multilevel and multiperspective analysis of co-occurrence patterns as the design goal, we compile a list of analytical tasks.

R1 (Global Exploration): How do co-occurrence events distribute throughout the city? What is the co-occurrence interaction among regions? How do co-occurrence patterns change over time? Such information can help users understand co-occurrence patterns in the global city.

R2 (Regional Exploration): What are co-occurrence patterns between a concrete region and all other regions? How do these co-occurrence patterns change over time? What is the time-varying law of the co-occurrence quantity? When the co-occurrence pattern is specific to a certain region, the information above will be quite attractive.

R3 (Location Exploration): Human mobility is highly location-dependent. So, what are locations of regions that are involved in co-occurrence events you are interested in? Regional functions have closed relationship with human mobility. What kinds of regional functions do these regions possess? Furthermore, how do we accurately extract the desired location information from a map with complicated information? These challenges need to be solved in the location exploration.

R4 (Statistical Exploration): Human mobility is highly time-dependent. After obtaining the distribution of co-occurrence patterns, users will be interested in the time-varying law of global co-occurrence patterns and regional co-occurrence patterns from multiple time dimensions, such as months, weeks, days, or even hours. In addition to co-occurrence quantity, traffic flow among regions is also of interest to users. These tasks will be settled through the statistical exploration.

Based on such analytical tasks, we design the COOC system to visualize co-occurrence patterns as a meaningful form. Fig. 3 presents the system architecture, which is composed of two modules: data modeling module and visualization module. In data modeling module, based on the raw data set, we first perform data preprocessing operation, including data cleaning from multiple aspects or organizing data into a database. Then, we execute time division, region division, and transaction construction to construct the data structure of the preprocessed data. Following that we utilize an efficient FP mining algorithm to mine co-occurrence event data from the data with a good data structure. Finally, we integrate the extracted co-occurrence event data and analytical tasks to conduct the visual design. Our visualization system consists of four parts, and each part contains one or more views to support the corresponding analytical tasks.

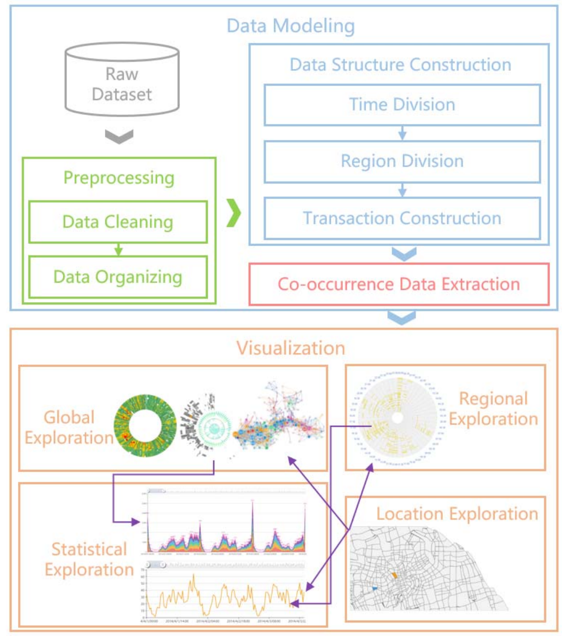


Fig. 3. System architecture of COOC.

V. Data Modeling

In this section, we first describe the data preprocessing and then introduce how to construct the data structure. Finally, we present how to mine co-occurrence event data effectively.

A. Data Preprocessing

Our system is based on taxi GPS data. The collected raw GPS data have a great number of error records due to complex attributes. What is more, for our analytical goal, there are redundant fields. Therefore, adequate data preprocessing operations are necessary. Co-occurrence patterns focus on origins and destinations of human travel instead of actual trajectories. Thus, GPS records without passengers are useless and are filtered out. Then, we extract trajectory data from GPS records according to taxi ID, and OD data can be obtained from trajectory data. Travel time, travel distance, and average speed are the directest indicators of one trip. We calculate the three indicators of OD data. Most urban road planning complies with rectangular rules. We use Manhattan distance to get travel distance based on the longitude and latitude information of origins and destinations. Manhattan distance is also called taxi distance, and its calculation formula is as follows:

$$\begin{aligned}\text{Distance} = d[(O_{\text{lat}}, O_{\text{lng}}), (O_{\text{lat}}, D_{\text{lng}})] \\ + d[(O_{\text{lat}}, D_{\text{lng}}), (D_{\text{lat}}, D_{\text{lng}})] \quad (1)\end{aligned}$$

where O_{lat}, O_{lng}, D_{lat}, and D_{lng} stand for origin latitude, origin longitude, destination latitude, and destination longitude respectively, and function d utilizes the latitude and longitude information to get distances. That is, Manhattan distance is the sum of the horizontal distance and the vertical distance between the origin and the destination. We set thresholds of travel time, travel distance, and average speed to further improve data quality by eliminating OD records whose field values are too high or too low.

B. Data Structure Construction

We perform data structure construction for better co-occurrence event data mining. In this paper, the definition of co-occurrence events is based on time periods and regions. Therefore, we need to convert temporal and spatial information

TABLE I
STATISTICS OF TRAVEL TIME

Duration (minute) \ Date	5th, April	6th, April	7th, April	10th, April
$[0, 10)$	37.4%	35.19%	34.95%	38.16%
$[0, 20)$	68.33%	66.24%	65.76%	67.63%
$[0, 30)$	80.72%	78.92%	78.78%	79.57%
$[0, 40)$	87.23%	85.56%	85.53%	85.96%
$[0, 50)$	90.81%	89.45%	89.49%	89.67%
$[0, 60)$	92.85%	91.88%	91.88%	91.97%

of OD data into time periods and regions, which are exactly the time division and region division.

1) Time Division: In order to analyze co-occurrence patterns under the same granularity of time division, we adopt regular time division instead of irregular time division, and the division formula is as follows:

$$\text{Time}_i = [i\theta, (i+1)\theta), \quad i = 0, 1 \ldots n-1 \tag{2}$$

where i is the number of a time interval and there are n time intervals. θ is the length of a time interval. For selecting an appropriate time division granularity to analyze co-occurrence patterns, we perform statistics on travel time and display the result in Table I. For taxi travel, the percentage of travel time increases rapidly within 30 min, and then the growth rate slows down significantly. Moreover, about 80% travel time is within 30 min and we set θ as 30 min for the fine-grained analysis to convert the temporal information in OD data into time intervals and to quantify time points with the number of time intervals i.

2) Region Division: Taxis travel along road networks in the city. Thus, we refer to a road network-based region division method [31] and incorporate it into our data modeling module instead of a rough grid division. Urban roads, which are densely distributed, have multiple levels. Therefore, we select the most important roads in the road network, including highways, primary roads, and secondary roads, and map them into a binary image. Dilation operation is performed on the image to eliminate small gaps to avoid too small regions in the final division. Then, we thin the width of roads as a pixel and set the pixels that belong to a region as the same number. Finally, we delete the pixels representing roads. In this way, we divide the entire city into 542 regions, and the spatial information in OD data is also converted to the number of each region.

3) Transaction Construction: After time division and region division, we can obtain time intervals and the number of regions in OD data. The core of co-occurrence patterns in human mobility are regional relations. According to FP mining algorithms, we need to extract transaction data sets, that is, using OD data to build transactions. In the same time period, we aggregate the OD data according to destinations to get the set of regions going to the same destination. A transaction indicates that people from a group of regions arrive at one destination within a certain time interval. Fig. 4 presents an example of transaction construction, in which the trip from the 1st region to the 5th region and the trip from the 3rd region to the 5th region construct a transaction whose destination is the 5th region and origin set contains the 1st region and the 3rd region.

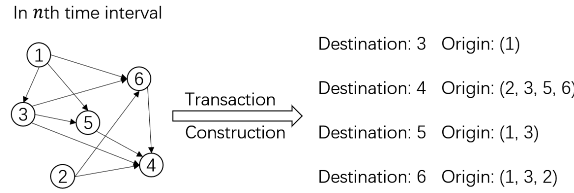


Fig. 4. Example of transaction construction.

C. Co-Occurrence Data Extracting

There are two types of classical frequent itemset mining algorithms. One is the *a priori* algorithm, and the other is the FP-growth algorithm [32]. The former uses an iterative method of layer-by-layer search, to continuously construct candidate sets and to filter candidate sets, to mine frequent itemsets. The mining process requires many scans of transaction data sets. When the data size is large, for instance, GPS data, the algorithm is quite inefficient. The FP-growth algorithm compresses transaction data sets through FP-tree data structure to obtain frequent itemsets with only two scans of transaction data sets. It is much more efficient than *a priori* algorithm. This is why we choose this algorithm to mine co-occurrence event data.

Han *et al.* [33] first proposed the FP-tree structure and proposed an efficient FP-growth algorithm based on such a structure. FP-tree is a special kind of prefix tree, which consists of frequent item header tables and item prefix trees. The FP-tree construction process is to sort transaction data items in transaction data table according to support and to insert the data items of each transaction in descending order into a tree with null as the root node, and at the same time, to record the support of the current node at each node.

The FP-growth algorithm continuously compresses construction and projection of FP-tree, that is, to compress the database that provides frequent itemsets into an FP-tree and to reserve itemsets' association information. The algorithm constructs a conditional projection database and projection FP-tree for each frequent item. This process is repeated for each new-constructed FP-tree until the constructed new FP-tree is empty or contains only one path. When the constructed FP-tree is empty, its prefix is the FP. When the constructed FP-tree has only one path, FPs can be obtained by enumerating all possible combinations and connecting them with the prefix of this tree. Items in header table of FP-tree need to be sorted in the descending order. First, items cannot share the prefix without sorting. Then, sorting in the ascending order will cause items appearing frequently to be in branches of trees and items cannot share more prefixes. Otherwise, sorting in the descending order will make frequent items to appear in upper layers of trees. They can be shared by more items as prefixes. The detailed construction processes of FP-tree and FP-growth can be found in [33].

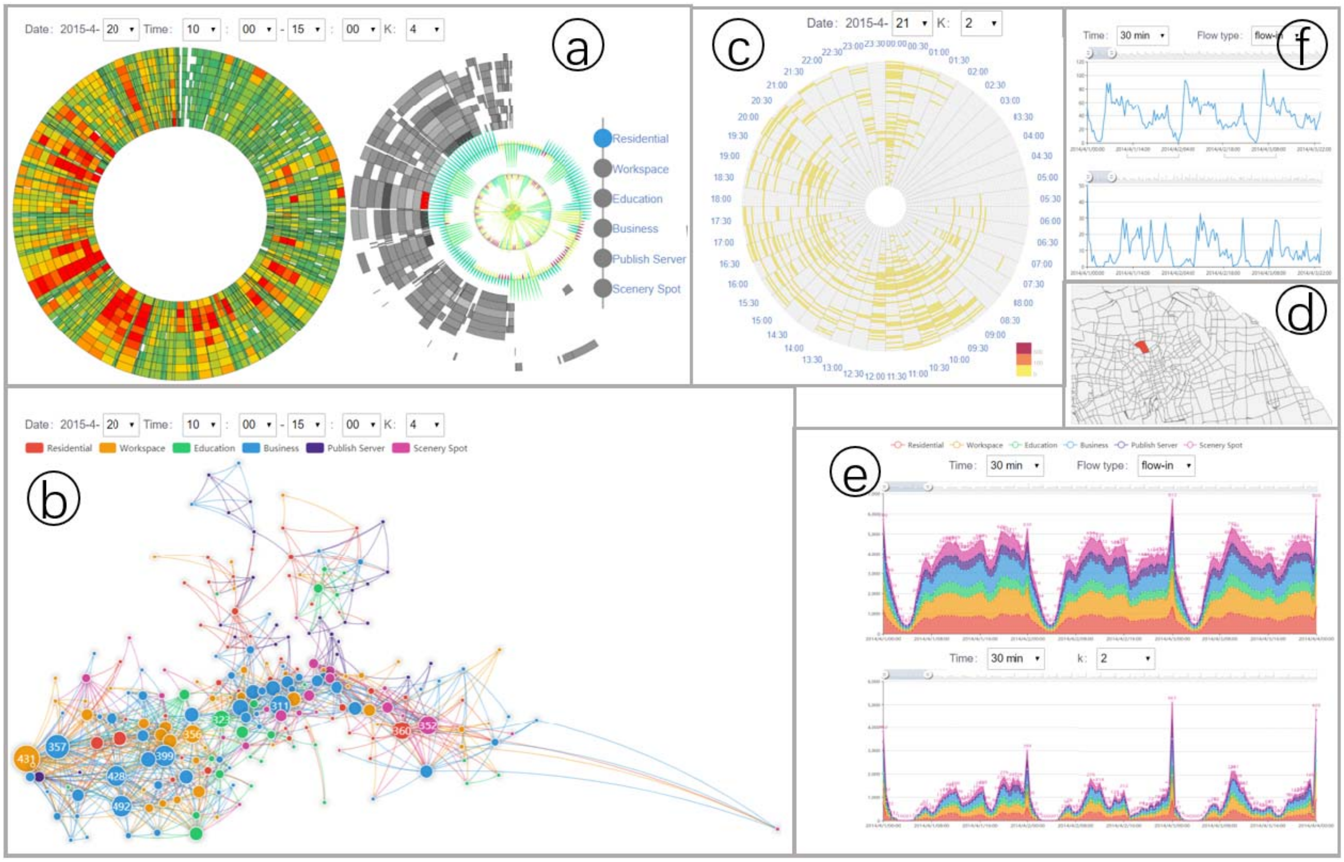


Fig. 5. Interface of our visual design. (a) Circular view shows the regional distribution and temporal distribution of co-occurrence patterns in the global scope. (b) Topology view presents the spatial distribution of co-occurrence regions and interactions between regions. (c) Cyclic heatmap view presents the co-occurrence patterns of specific regions. (d) Regional boundary map supports the location exploration. (e) Stacked view shows the statistical information of a certain type of functional regions. (f) Line view refines the statistical information to a specific region.

TABLE II

EXAMPLES OF CO-OCCURRENCE EVENT DATA

TimeId	Destination	Origin	Support	k
1	(214,348,357,370,380)	(535)	5	1
5	(253,318,319,422)	(273,280,323)	4	3
9	(60,102,132,210)	(18,22)	4	2
21	(141,148,202,222,249)	(165,178)	5	2

Table II presents the examples of extracted co-occurrence event data. Field TimeId is the serial number of time intervals, which is from 1 to 48. Field Destination and field Origin are the destination set and the origin set, respectively. Field Support is the number of elements in the destination set, and field k is the number of elements in the origin set. Take line 1 shown in Table II for example. It implies that in 1st time interval, that is, from 00:00 A.M. to 00:30 A.M., people from the 535th region visit 214th, 348th, 357th, 370th, and 380th regions. Moreover, we obtain the interregional flow data, as displayed in Table III, which contains the flow from an origin to a destination. In Section VI, we describe how to form a co-occurrence pattern visually exploring design based on co-occurrence event data and interregional flow data.

TABLE III

EXAMPLES OF INTERREGIONAL FLOW DATA

TimeId	Destination	Origin	Flow
0	1	82	1
5	11	13	7

VI. VISUAL DESIGN

This section introduces a set of visualization techniques that assist users in exploring co-occurrence patterns from global regional scope and its time-varying statistics.

A. Global Explorer

Before performing detailed exploration, users tend to grasp what the overall situation is (R1). Therefore, in the global explorer part, we are committed to providing a global overview of co-occurrence patterns from the perspective of regional distribution and interregional interaction. This part consists of two views, topology view and circular view.

Topology view mainly presents the spatial distribution of co-occurrence regions and interactions among regions. We regard a region as a vertex, a co-occurrence event as an edge to build a global co-occurrence network, which is exactly the topology view, as shown in Fig. 5(b). We use

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

the degree of vertexes to denote their size so users can intuitively grasp regions with the greatest co-occurrence strength. Here, we define the co-occurrence strength as the number of co-occurrence events between the selected region and other regions. We extract center points of regions to aggregate regions to points, and the relative positions of vertexes are the mapping of regions' geographical locations. Thence, users can obtain spatial attributes of regions to explore the spatial distribution of co-occurrence patterns. In addition, combined with regional functions data, vertexes in the topology view are divided into six categories, corresponding to six regional functions. Users can select only one or more regional functions to display and to explore the global co-occurrence distribution of a concrete kind of regions or co-occurrence interaction of several kinds of regions. In the design of the topology views, we filter out regions without co-occurrence events, i.e., isolated vertexes, to help users focus on vertexes with the high co-occurrence strength without interference. Topology view is implemented by Echarts, an open-source, web-based, cross-platform framework that supports the rapid construction of interactive visualization [34]. Parameter k in the topology view and other views means the number of origins involved in co-occurrence events.

The topology view provides co-occurrence interactions among regions. We design a novel visual form, circular view [Fig. 5(a)], to display the regional distribution and temporal distribution of co-occurrence patterns in the global scope. The entire circle is divided into 542 sectors (corresponding to 542 regions in the region division), and each sector represents a region. Then, the sector is further divided into 48 subsectors (corresponding to 48 time intervals), and each subsector represents the co-occurrence information of a region at a time interval. The regions in the topology view are consistent with the elements in R_a of the co-occurrence event definition, and we say that they have co-occurrence events. The regions displayed in the circular view are consistent with the elements in R_b, that is, the places where co-occurrence events occur. In a co-occurrence event, we need pay attention to not only regions which have the co-occurrence event but also regions where the co-occurrence event occurs. Such a subsector represents the number of co-occurrence events that occur in the region within a certain time interval, and we use color to represent the quantity. The darker the color, the more co-occurrence events occur in the region. If there is no co-occurrence event during the time interval, the subsector is blank.

Furthermore, we attach a subview to present co-occurrence events that occur in a region. In the subview, we use a line to represent regions that participate in a co-occurrence event. A line is divided into multiple sublines (the number of sublines depends on the number of regions participating in the co-occurrence event), i.e., each subline represents a region. The length of sublines is the distance between regions, and the color denotes the flow. In this way, if users find an interesting region in the circular view, they can click on it to generate its subview to further observe its co-occurrence events and to understand co-occurrence strength or density.

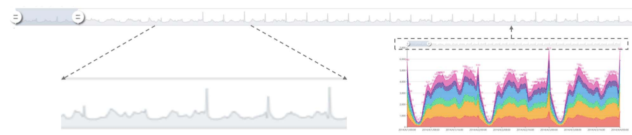


Fig. 6. Thumbnail of the statistical explorer.

B. Regional Explorer

When conducting global exploration, users may find an interested co-occurrence event or region, which requires a more detailed display. We apply cyclic heatmap view to present co-occurrence patterns of specific regions. In this view, the co-occurrence pattern of a region is distributed with time increasing in a clockwise direction, as shown in Fig. 5(c). We divide the whole circle into 48 sectors (a day has 48 time intervals), and each sector represents co-occurrence patterns during the time interval. Each sector is subdivided into multiple subsectors that represent the co-occurrence between the selected region and other regions, which co-occur with the selected region. The color of subsectors denotes the number of co-occurrence events. The darker the color, the more co-occurrence events occur. In this way, users can get the co-occurrence distribution of a concrete region in a day and understand characteristics of co-occurrence patterns by views, like time variation of co-occurrence event density and regions that co-occur with the selected region most closely.

C. Statistical Explorer

In previous views, no matter global explorer or regional explorer, they are all qualitative observation of co-occurrence patterns. Then, it is quite necessary to analyze co-occurrence patterns from the quantitative point of view, that is, statistical analysis. Two most important attributes of co-occurrence patterns are time and space. We divide the two attributes into multiple dimensions. Time attributes include the gradual refinement of months, weeks, days, or even hours. Space attributes include two dimensions, a type of functional regions and a specific region. We utilize two views, stacked view and line view, to present above co-occurrence statistics. Stacked view shows the statistical information of a certain type of functional regions, and users can select one or more functional regions [Fig. 5(e)]. Time dimensions are implemented by selecting a time scale, such as 30 min or one day. Users also can customize the time range displayed in the view by sliding time axis. In order to provide the guidance for time selection, we attach the overall thumbnail to the view, as shown in Fig. 6. Line view [Fig. 5(f)] is to refine the above statistical information to a specific region with similar time selection function to facilitate users to observe co-occurrence patterns of a specific region. Both stacked view and line view include the statistics of co-occurrence events and statistics of inflow and outflow.

D. Location Explorer

Human mobility has a high degree of spatial dependence, so maps cannot be ignored in the view design.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

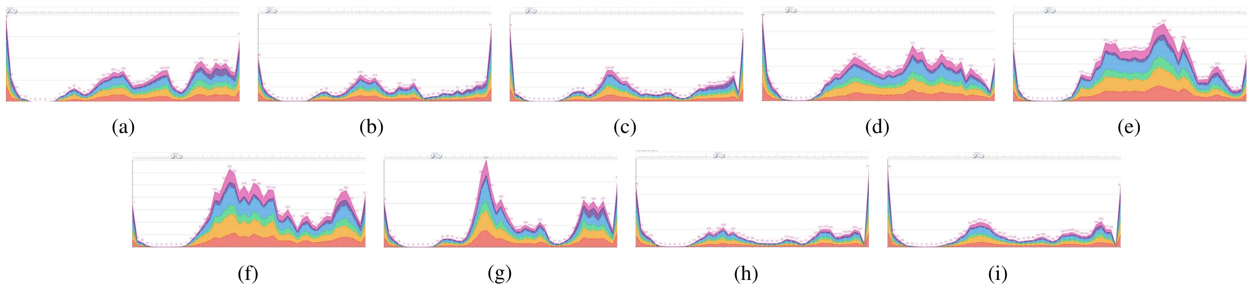


Fig. 7. Co-occurrence events' number of residential area (red), workspace (orange), education (green), business (blue), public service (purple), and scenery spot (pink). (a) 1st April. (b) 2nd April. (c) 3rd April. (d) 4th April. (e) 5th April. (f) 6th April. (g) 7th April. (h) 11th April. (i) 12th April.

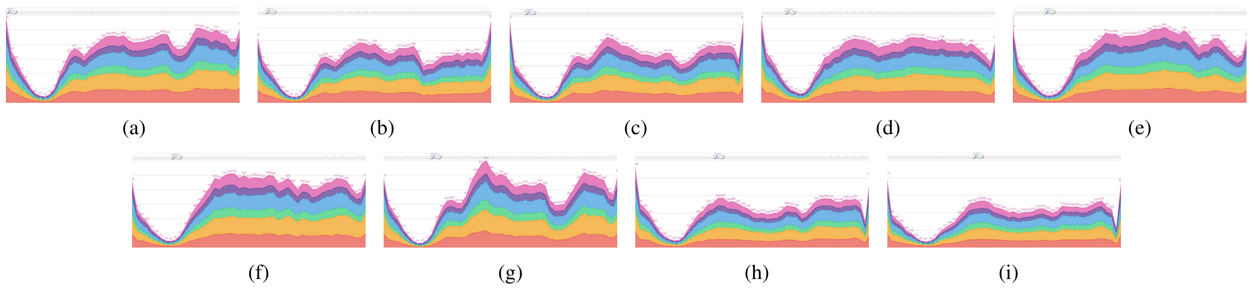


Fig. 8. Flow of residential area (red), workspace (orange), education (green), business (blue), public service (purple), and scenery spot (pink). (a) 1st April. (b) 2nd April. (c) 3rd April. (d) 4th April. (e) 5th April. (f) 6th April. (g) 7th April. (h) 11th April. (i) 12th April.

Traditional maps contain complex road network information and point of interest information. In our co-occurrence pattern exploration, we only need to know the spatial locations of regions. Therefore, traditional maps are obviously not applicable and we design a regional boundary map to support the location exploration, as shown in Fig. 5(d). In conjunction with the region division results of Shanghai, we replace the traditional maps in a concise manner, allowing users to focus on the region information rather than the additional information. We use a closed path to describe the contour of a region. We first obtain the dot matrix of region contour and then sort dots so that connecting all dots in sequence to draw lines can form a closed path, which represents region contour, and a concise map is generated. Contour view serves as an auxiliary view to provide the spatial information for the whole design.

E. Interaction

We add rich interaction in our visualization system COOC to encourage users in performing co-occurrence patterns from multiple dimensions and multiple perspectives.

1) Connectivity: In the system, apart from the stacked view, other views are intrinsically connected. If you click on a region in the topology view or circular view, not only the subview will be generated, but also cyclic heatmap view and line view will be converted to the corresponding information.

2) Filtering and Highlighting: Users can choose what to display in the view, such as regional functions, time range, types of co-occurrence events, and types of flow (inflow or outflow). The system highlights the part that users are interested in by providing tips in each view to show more detailed information and highlighting the selected part through the change of display effect.

TABLE IV

DESCRIPTION OF TAXI GPS TRAJECTORY DATA

Field	**Annotation**
TaxiId	Taxi Id
Latitude	Coordinates
Longitude	Coordinates
State	1: occupied, 0: vacant
Date	Date that GPS record was sent
Time	Time that GPS record was sent
Speed	Taxi running speed

VII. CASE STUDIES

To evaluate the system, we carry out case studies based on taxi GPS data set collected from April 1–30, 2015, Shanghai, China. The data set consists of several fields, including state, speed, date, time, and geographical coordinates, as displayed in Table IV. The data set contains 34 billion GPS records and is 619 GB. In this section, we provide how to explore urban co-occurrence patterns from multiple levels of time granularity and space granularity, that is, from monthly, daily, to hourly, and from the global scope, functional regions, to specific regions gradually.

First, we grasp an overview of global co-occurrence patterns in the month. We select co-occurrence event stacked views and flow stacked views of weekdays, holidays, and weekends, respectively, as shown in Figs. 7 and 8. Among the selected days, 1st, 2nd, and 3rd are weekdays, 4th, 5th, and 6th are Qingming Festivals, 7th is the first weekday after holidays, and 11th and 12th are weekends. By comparing Figs. 7 and 8, we can see that the number of co-occurrence events is significantly lower than flow, but flow is less sensitive to different types of days, and time-varying patterns of flow

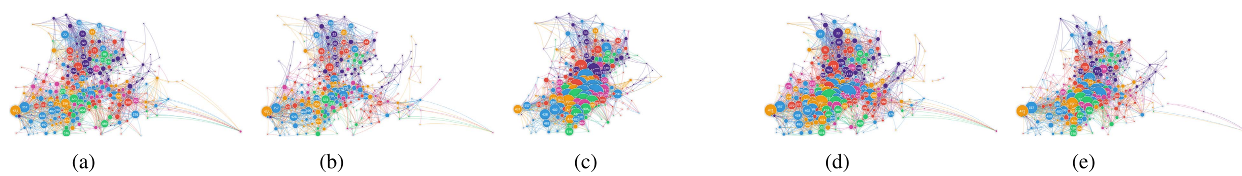


Fig. 9. Topology views of residential area (red), workspace (orange), education (green), business (blue), public service (purple), and scenery spot (pink). (a) 1st April. (b) 2nd April. (c) 5th April. (d) 7th April. (e) 11th April.

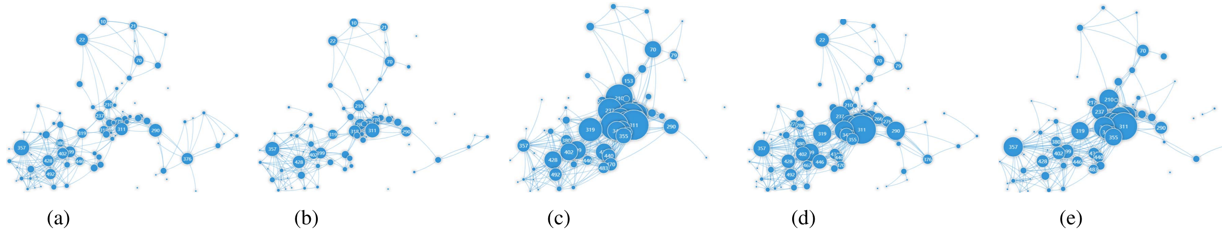


Fig. 10. Topology views of business (blue). (a) 1st April. (b) 2nd April. (c) 5th April. (d) 7th April. (e) 11th April.

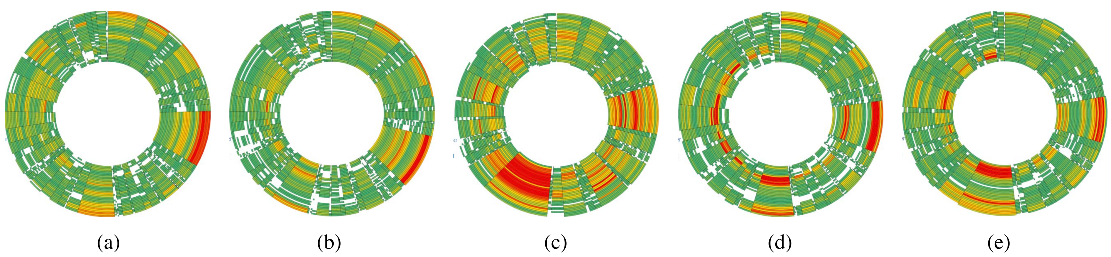


Fig. 11. Circular views of business. (a) 1st April. (b) 2nd April. (c) 5th April. (d) 7th April. (e) 11th April.

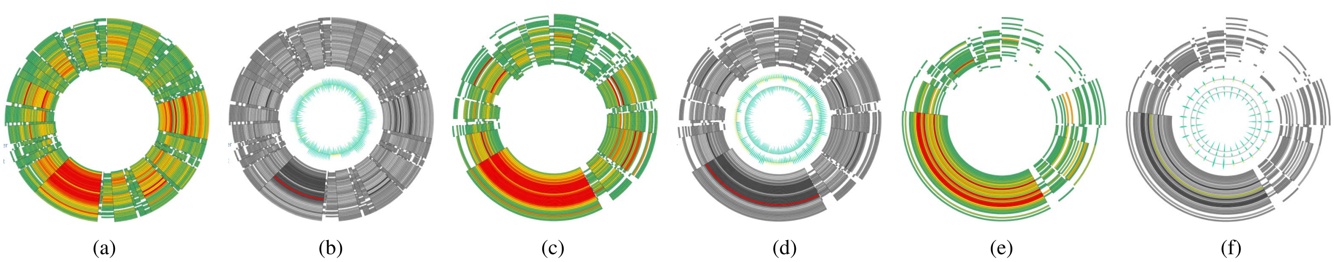


Fig. 12. Circular views of business on 5th April. (a) $k = 2$. (b) 253rd region, $k = 2$. (c) $k = 3$. (d) 253rd region, $k = 3$. (e) $k = 4$. (f) 253rd region, $k = 4$.

are basically similar, that is, flow reaches trough at 4:00 A.M., during days' flow is generally flat with a few slight peaks and troughs. Co-occurrence patterns are more sensitive to different types of days. In co-occurrence event stacked views (Fig. 7), time-varying patterns on 2nd and 3rd are similar, those on 4th, 5th, and 6th are similar, and those on 11th and 12th are similar, i.e., among weekdays, mondays, tuesday, and wednesday; thursday and friday have different co-occurrence patterns, respectively, holidays have special patterns, and patterns on weekends are different from above all. Moreover, co-occurrence density at night is relatively low.

Following that, we refine the level of time granularity to daily. We choose topology views of the time period from 6:00 to 24:00 on 1st, 2nd, 5th, 7th, and 11th to further explore urban co-occurrence patterns, as shown in Fig. 9. Topology views present that co-occurrence density on weekdays is low and evenly spreads across the city; co-occurrence events are concentrated in urban centers on weekends. What is more, co-occurrence events of business regions represented by the blue color are obviously denser than other functional regions. Urban centers own an explosive co-occurrence density on holidays. As the first weekday after Qingming Festival, the co-occurrence density of 7th is much greater than common weekdays. We further obtain topology views of business regions to acquire the expansion of co-occurrence events on weekends and holidays, as shown in Fig. 10. The co-occurrence patterns of functional regions can be obtained.

Such outbreak of co-occurrence events is also reflected in the circular views (Fig. 11), and the co-occurrence density of the circular view on 5th April is significantly higher than that of other circular views. Here, we notice a region, which shows large co-occurrence strength and through interaction operations, we find that it is the 253rd region. Obviously, the co-occurrence patterns of the 253rd region are desirable to explore further. By clicking this region, we can get Fig. 12. The value of k indicates the number of origins involved in co-occurrence events. The region shows the high co-occurrence strength with different k values. What is more, the lengths of sublines in subviews are relatively short, indicating that distances between regions are short. We further look for the points of interest data of the 253rd region and find that this region is located in the urban center with many residential areas scattered around. Therefore, the co-occurrence events of this region are mostly short distance trips between residential areas to nearby shopping malls.

VIII. Conclusion

In this paper, we present a framework to mine co-occurrence event data from mobile data and to explore urban co-occurrence patterns visually. The framework embraces two modules. The first one is data modeling, which consists of preprocessing, data structure construction, and co-occurrence data extraction. Based on extracted co-occurrence event data, we perform the visualization module, which is composed of global exploration, regional exploration, statistical exploration, and location exploration. What is more, we design a circular view, a novel visual form, to display urban global co-occurrence patterns intuitively. Case studies based on the real taxi data set demonstrate that our visual system can provide interesting insights and analysis of the co-occurrence pattern exploration.

There are several directions to follow in our future work for further research of urban co-occurrence patterns. First, we plan to apply COOC to a variety of real-world data sets to mine the general laws of urban co-occurrence patterns. In addition, we tend to build a general model hidden in co-occurrence patterns. Abnormal detection based on co-occurrence patterns will also be considered in the future work.

References

[1] L. Sun and K. W. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transp. Res. B, Methodol.*, vol. 91, pp. 511–524, Sep. 2016.

[2] F. Li, Z. Li, K. Sharif, Y. Liu, and Y. Wang, "Multi-layer-based opportunistic data collection in mobile crowdsourcing networks," *World Wide Web*, vol. 12, no. 3, pp. 783–802, 2017, doi: 10.1007/s11280-017-0482-9.

[3] D. Chen, "Research on traffic flow prediction in the big data environment based on the improved RBF neural network," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2000–2008, Aug. 2017.

[4] X. Kong, M. Li, T. Tang, K. Tian, L. Moreira-Matias, and F. Xia, "Shared subway shuttle bus route planning based on transport data analytics," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1507–1520, Oct. 2018.

[5] Z. Ning, F. Xia, N. Ullah, X. J. Kong, and X. P. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 16–55, May 2017.

[6] L. Hong, Y. Zheng, D. Yung, J. Shang, and L. Zou, "Detecting urban black holes based on human mobility data," in *Proc. SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2015, p. 35.

[7] X. Kong, M. Li, J. Li, K. Tian, X. Hu, and F. Xia, "CoPFun: An urban co-occurrence pattern mining scheme based on regional function discovery," *World Wide Web*, pp. 1–26, May 2018, doi: 10.1007/s11280-018-0578-x.

[8] A. Machens, F. Gesualdo, C. Rizzo, A. E. Tozzi, A. Barrat, and C. Cattuto, "An infectious disease model on empirical networks of human contact: Bridging the gap between dynamic network data and contact matrices," *BMC Infectious Diseases*, vol. 13, no. 1, p. 185, 2013.

[9] C.-T. Li and S.-D. Lin, "Social flocks: Simulating crowds to discover the connection between spatial-temporal movements of people and social structure," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 1, pp. 33–45, Mar. 2018.

[10] L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang, "Understanding metropolitan patterns of daily encounters," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 34, pp. 13774–13779, 2013.

[11] W. Wu *et al.*, "TelCoVis: Visual exploration of co-occurrence in urban human mobility based on telco data," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 935–944, Jan. 2016.

[12] D. Perna, R. Interdonato, and A. Tagarelli, "Identifying users with alternate behaviors of lurking and active participation in multilayer social networks," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 1, pp. 46–63, Mar. 2018.

[13] F. Wang, W. Chen, Y. Zhao, T. Gu, S. Gao, and H. Bao, "Adaptively exploring population mobility patterns in flow visualization," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 2250–2259, Aug. 2017.

[14] J. Feng *et al.*, "DeepMove: Predicting human mobility with attentional recurrent networks," in *Proc. World Wide Web Conf.*, Lyon, France, Apr. 2018, pp. 1459–1468.

[15] W. Chen, F. Guo, and F. Y. Wang, "A survey of traffic data visualization," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 2970–2984, Jun. 2015.

[16] W. Chen, Z. Huang, F. Wu, M. Zhu, H. Guan, and R. Maciejewski, "VAUD: A visual analysis approach for exploring spatio-temporal urban data," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 9, pp. 2636–2648, Sep. 2017.

[17] K. Faust *et al.*, "Microbial co-occurrence relationships in the human microbiome," *PLoS Comput. Biol.*, vol. 8, no. 7, p. e1002606, 2012.

[18] V. Srinivasan, S. Moghaddam, A. Mukherji, K. K. Rachuri, C. Xu, and E. M. Tapia, "MobileMiner: Mining your frequent patterns on your phone," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, New York, NY, USA, Sep. 2014, pp. 389–400.

[19] X. Qi, R. Xiao, C.-C. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2199–2213, Nov. 2014.

[20] M. Celik, S. Shekhar, J. P. Rogers, and J. A. Shine, "Mixed-drove spatiotemporal co-occurrence pattern mining," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 10, pp. 1322–1335, Oct. 2008.

[21] M. Celik, S. Shekhar, J. P. Rogers, J. A. Shine, and J. S. Yoo, "Mixed-drove spatio-temporal co-occurence pattern mining: A summary of results," in *Proc. 6th Int. Conf. Data Mining (ICDM)*, Hong Kong, Dec. 2006, pp. 119–128.

[22] K. G. Pillai, R. A. Angryk, J. M. Banda, M. A. Schuh, and T. Wylie, "Spatio-temporal co-occurrence pattern mining in data sets with evolving regions," in *Proc. IEEE 12th Int. Conf. Data Mining Workshops (ICDMW)*, Brussels, Belgium, Dec. 2012, pp. 805–812.

[23] B. Aydin, D. Kempton, V. Akkineni, S. R. Gopavaram, K. G. Pillai, and R. Angryk, "Spatiotemporal indexing techniques for efficiently mining spatiotemporal co-occurrence patterns," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Washington, DC, USA, Oct. 2014, pp. 1–10.

[24] M. Akbari, F. Samadzadegan, and R. Weibel, "A generic regional spatio-temporal co-occurrence pattern mining model: A case study for air pollution," *J. Geograph. Syst.*, vol. 17, no. 3, pp. 249–274, 2015.

[25] S. Barua and J. Sander, "SSCP: Mining statistically significant co-location patterns," in *Proc. Int. Symp. Spatial Temporal Databases*, Berlin, Germany, 2011, pp. 2–20.

[26] Y. Huang, S. Shekhar, and H. Xiong, "Discovering colocation patterns from spatial data sets: A general approach," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1472–1485, Dec. 2004.

[27] Y. Huang, J. Pei, and H. Xiong, "Mining co-location patterns with rare events from spatial data sets," *Geoinformatica*, vol. 10, no. 3, pp. 239–260, 2006.

[28] Y. Yang, T. Dwyer, S. Goodwin, and K. Marriott, "Many-to-many geographically-embedded flow visualisation: An evaluation," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 411–420, Jan. 2017.

[29] M. Sun, C. North, and N. Ramakrishnan, "A five-level design framework for bicluster visualizations," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1713–1722, Dec. 2014.

[30] J. Zhao, M. Sun, F. Chen, and P. Chiu, "BiDots: Visual exploration of weighted biclusters," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 195–204, Jan. 2018.

[31] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.

[32] Q. Lan, D. Zhang, and B. Wu, “A new algorithm for frequent itemsets mining based on apriori and FP-tree,” in *Proc. WRI Global Congr. Intell. Syst.*, May 2009, pp. 360–364.

[33] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 1–12, 2000.

[34] D. Li *et al.*, “ECharts: A declarative framework for rapid construction of Web-based visualization,” *Vis. Inform.*, vol. 2, no. 2, pp. 136–146, 2018, doi: 10.1016/j.visinf.2018.04.011.

Xiangjie Kong (M’13–SM’17) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China.

He is currently an Associate Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored over 70 scientific papers in international journals and conferences (with over 50 indexed by ISI, SCIE). His current research interests include intelligent transportation systems, mobile computing, and cyber-physical systems.

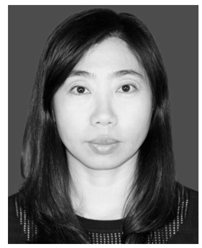
Dr. Kong is a Senior Member of CCF and a member of ACM. He has served as the (Guest) Editor for several international journals and the Workshop Chair or a PC Member for a number of conferences.

Menglin Li received the B.S. degree in software engineering from the Dalian University of Technology, Dalian, China, in 2016, where she is currently pursuing the master’s degree from the Alpha Lab, School of Software.

Her current research interests include big traffic data mining and analysis, human mobility behavior analysis, and smart city development.

Gaoxing Zhao received the B.S. degree in software engineering from the Dalian University of Technology, Dalian, China, in 2018.

His current research interests include big traffic data mining and analysis, urban functional region mining, co-occurrence pattern analysis, and spatiotemporal data visualization.



Huijie Zhang (M’16) received the Ph.D. degree from Jilin University, Changchun, China, in 2009.

She is currently a Full Professor with the School of Information Science and Technology, Northeast Normal University, Changchun. Her current research interests include scientific visualization, information visualization, visual analysis, computer graphics, 3-D model simplification, multiresolution modeling for terrain and 3DGIS, and optimization algorithm.

Dr. Zhang is a member of China Computer Federation.

Feng Xia (M’07–SM’12) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China.

He was a Research Fellow with the Queensland University of Technology, Brisbane, QLD, Australia. He is currently a Full Professor with the School of Software, Dalian University of Technology, Dalian, China. He has authored two books and over 200 scientific papers in international journals and conferences. His current research interests include computational social science, network science, data science, and mobile social networks.

Dr. Xia is a Senior Member of ACM and a member of AAAS. He is the (Guest) Editor of several international journals. He serves as the General Chair, the PC Chair, the Workshop Chair, or the Publicity Chair for a number of conferences.