TelCoVis: Visual Exploration of Co-occurrence in Urban Human Mobility Based on Telco Data

Wenchao Wu, Jiayi Xu, Haipeng Zeng, Yixian Zheng,
Huamin Qu, *Member, IEEE*, Bing Ni, Mingxuan Yuan and Lionel M. Ni, *Fellow, IEEE*

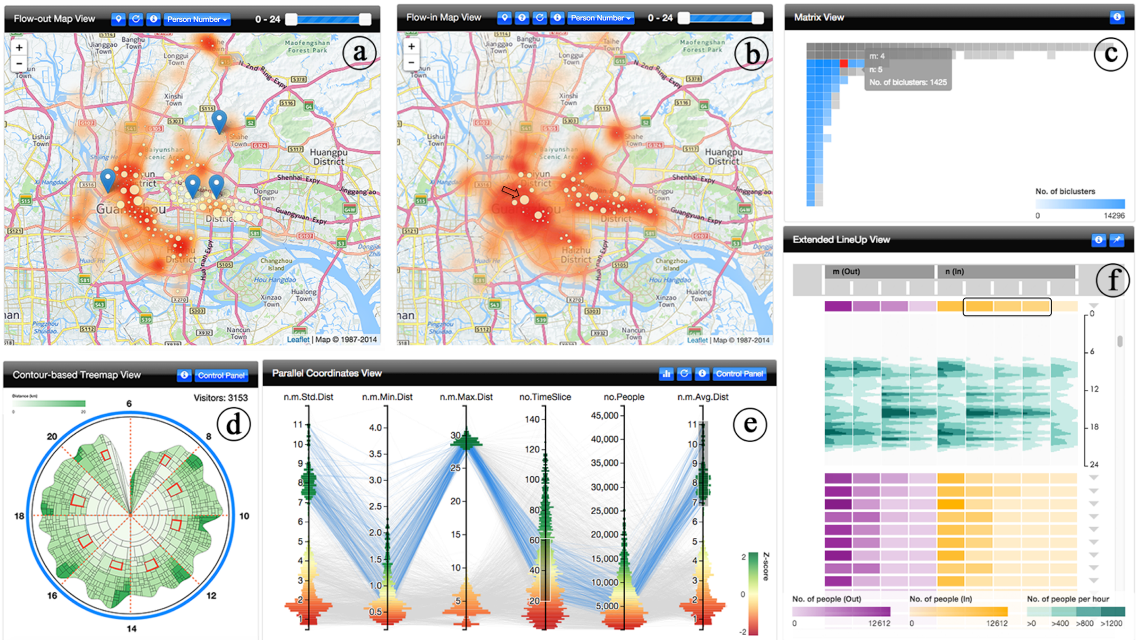


Fig. 1: A visual analytics system for the exploration of co-occurrence in human mobility of Guangzhou based on telco data. Map View (a-b) supports an intuitive region-based exploration of co-occurrence within spatial context. Contour-based Treemap View (d) provides a visual signature characterizing the spatial and temporal distribution of human mobility at a certain place, which facilitates analysts to gain insights into a co-occurrence pattern and generate explanatory hypotheses. Matrix View (c) provides an overview of correlations in co-occurrence through biclustering. Parallel Coordinates View (e) enables an efficient quantitative analysis based on a wide range of attributes of biclusters. And Extended LineUp View (f) conveys the diversity in those biclusters.

Abstract—Understanding co-occurrence in urban human mobility (i.e. people from two regions visit an urban place during the same time span) is of great value in a variety of applications, such as urban planning, business intelligence, social behavior analysis, as well as containing contagious diseases. In recent years, the widespread use of mobile phones brings an unprecedented opportunity to capture large-scale and fine-grained data to study co-occurrence in human mobility. However, due to the lack of systematic and efficient methods, it is challenging for analysts to carry out in-depth analyses and extract valuable information. In this paper, we present TelCoVis, an interactive visual analytics system, which helps analysts leverage their domain knowledge to gain insight into the co-occurrence in urban human mobility based on telco data. Our system integrates visualization techniques with new designs and combines them in a novel way to enhance analysts' perception for a comprehensive exploration. In addition, we propose to study the correlations in co-occurrence (i.e. people from multiple regions visit different places during the same time span) by means of biclustering techniques that allow analysts to better explore coordinated relationships among different regions and identify interesting patterns. The case studies based on a real-world dataset and interviews with domain experts have demonstrated the effectiveness of our system in gaining insights into co-occurrence and facilitating various analytical tasks.

Index Terms—Co-occurrence, human mobility, telco data, bicluster, visual analytics

1 Introduction

With increasing availability of location-acquisition technologies, including GPS devices on vehicles and in mobile phones, huge volumes of data tracking human mobility have been collected, which provides an unprecedented opportunity to study human mobility patterns. Among them, our key interest lies in developing an efficient, effective and comprehensive solution to analyze the co-occurrence pattern (i.e. people from two different regions visit the same place during the same time span), making it possible for extensive applications with high so-

- *Wenchao Wu, Jiayi Xu, Haipeng Zeng, Yixian Zheng, and Huamin Qu are with Hong Kong University of Science and Technology. E-mail: {wwuag, yzhengaj, huamin}@cse.ust.hk, {tsanghaipeng, jiayi.xu64}@gmail.com.*
- *Bing Ni and Mingxuan Yuan are with Noah's Ark Lab, Huawei Technologies Investment Co. Ltd. Email: {Ni.Bing, Yuan.Mingxuan}@huawei.com.*
- *Lionel M. Ni is with University of Macau. Email: ni@umac.mo.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier no. 10.1109/TVCG.2015.2467194

cial and business values in modern society. For example,

- Analyzing the co-occurrence of people from different regions helps urban administrators better cope with threats of contagious diseases, and help social scientists gain insights and better model human social contact (e.g. contact between the rich and poor).
- Knowing the kind of people (e.g. which neighborhood and how far they live) visiting a shopping mall or dining in a restaurant during a specific time period allows shop owners to tailor well-targeted promotions and make better business decisions.
- Investigating the number of people co-occurring at a metro station or a highway entrance during peak hours facilitates decision making for urban planning (e.g. whether the capacity of the transportation system in a region needs to be expanded).

However, understanding the co-occurrence pattern in human mobility is a challenging task. Most of current studies [25, 38] have limitations due to the data (e.g. limited granularity and coverage) and analytical methods (e.g. conventional statistics, pure data mining) used. First of all, both spatial and temporal components are involved in addition to multi-dimensional and multivariate natures. Furthermore, the correlations in co-occurrence (i.e. people from multiple regions visit different places during the same time span) is complex and dynamic, while analyzing such correlations is essential for real world applications. Due to these complexities, a fully automatic analysis of human mobility is difficult, requiring considerable experience and profound knowledge in various fields. Therefore, analysts seek the help of visual analytics so as to take full advantage of both advanced computational power and human cognitive abilities.

In this paper, we propose a visual analytics system, TelCoVis, to analyze co-occurrence in human mobility based on telco data, a new type of all-in-one mobile phone data recording the details of data exchange (e.g. calls, messages, internet usage) between mobile phones and cell stations. The large coverage and fine granularity of the data make it possible for an in-depth analysis. We discuss specific design requirements for the exploration and how they have shaped the design of TelCoVis. In order to provide an intuitive representation to all possible users who may not have much knowledge on information technology, TelCoVis integrates well-established visualization techniques with new designs to present information in a succinct way, enables the combination of different pieces of information, and simplifies the identification of correlations. To the best of our knowledge, this system is the first to provide an array of visualizations that can be combined to gain insights into different aspects of co-occurrence in human mobility and handle various analytical tasks.

The major contributions of this work can be summarized as follows:

- The design and implementation of a comprehensive visual analytics system with two combined exploring schemes to investigate the co-occurrence pattern in human mobility based on the large-scale telco data;
- Several visualization designs enhanced with new features, including a contour-based treemap and an extended LineUp-style chart with diversity encoding, to reveal spatio-temporal characteristics and facilitate multi-perspective exploration;
- Case studies based on a real-world dataset that demonstrate how our system can lead to interesting insights and facilitate analysis for different applications.

2 Related Work

This section provides an overview of related research work. We focus on three most relevant topics: visualization of mobile phone data, visualization of movement, and bicluster visualization.

Visualization of mobile phone data In recent years, mobile phone data has been widely studied for various applications, such as estimating population distribution [7, 9, 22], analyzing local events [2, 37] and investigating mobility patterns [8, 32]. A systematic survey was provided by Calabrese [6] recently and we will mainly focus on relevant visualization techniques. Density maps have been used to plot the density of mobile phone usage and help identify "hot" spots [9]. The evolution of density may also be shown by employing a density map with color encoding the changes [22, 28]. Researchers have also discretized the spatial and temporal dimension into regions or bins, and then used traditional visualization techniques, such as bar charts [7], node-link diagram [20] and map mashup [7, 22], to present statistical results. In this paper, we propose more comprehensive visualization techniques for visual exploration and analysis rather than presentation tools. In addition to depicting aggregated attributes, Andrienko et al. [2] presented a suite of interactive visual analytics methods for reconstructing past events from mobile phone call records. Shen et al. [30] presented MobiVis to present social and spatial information in one heterogeneous network. And Pu et al. [27] developed a visualization system for visual analysis of mobile phone user groups in a cell station as well as patterns of handoff phone call records. In this paper, we analyze a new type of all-in-one mobile phone data, telco data, to address a different set of analytical tasks identified by our domain experts which contributes to various real world applications.

Visualization of movement Visualization of co-occurrence in human mobility falls into the general topic of visualization of movement, which has been widely studied in the visualization community. Andrienko et al. [1] discussed various characteristics of movement data, and categorized visualization techniques into three major categories, including direct depiction, summarization and pattern extraction. Direct depiction techniques present movement directly. This type of techniques includes plotting paths as polylines [1] or stacked bands [36], representing origin and destination of trajectories as points [12], and depicting spatial and temporal information together with space time cube [3]. However, when handling large and complex datasets, visual clutter is a major defect. Summarization techniques present movement based on statistical calculations, so that analysts can get an overall understanding of the movement and investigate aggregated patterns. This type of techniques includes density map [39], multivariate glyph [23, 42], and flow map [17]. Andrienko et al. [1] pointed out that summarization techniques could help to reduce the hidden uncertainties of data in spatial and temporal coverage. Therefore, we adopt them in our system along with other visualization techniques to provide a novel solution to a comprehensive analysis of co-occurrence in human mobility. Pattern extraction techniques present extracted patterns of movement to analysts for interpretation and further investigation. Many patterns, such as the interchange pattern [43] and the group movement pattern [4], have been studied. In our paper, we focus on the co-occurrence pattern in human mobility which are involved in many important application scenarios, but have not received much attention from researchers in the community.

Bicluster visualization In this paper, we also study correlations of co-occurrence in human mobility. In particular, we want to investigate people from several regions co-occur at a set of different places. Biclustering, which has been extensively used in bioinformatics, provides a potential solution to ease the process of analysis. It is a popular and efficient data mining technique to identify coordinated relationships between groups of entities of different types. Through biclustering, we can bundle origins and destinations of human mobility into coordinated sets called biclusters. Sun et al. [35] summarized the design options for bicluster visualization, including matrix-based visualizations, parallel coordinates and node-link diagrams, and surveyed their state-of-the-art. We briefly discuss some recent work in these three categories. Sun et al. [34] and Fiaux et al. [13] developed matrix-based visualizations, which is easy for analysts to perceive. However, in their methods, the combinatorial nature of biclusters is very likely to cause a large number of row and column replications in order to lay out all biclusters within a big matrix, which may be confusing. Gorg et al. [15] adopted parallel coordinates to organize entities into a list, which makes it easy for entities to be identified and selected. Nonetheless, it is a tedious and time-consuming task for analysts to identify biclusters of interest by trying different combinations of entities one after another through interactions. In addition, node-link diagrams [29] may also be used to visualize biclusters. This type of approaches is able to convey an overview of biclusters, while it may be difficult for analysts to find a specific entity because of random placement. Although these techniques can be used for bicluster visualization, none of them can support intuitive and efficient exploration of biclusters within spatio-temporal context which is essential for our study. Hence, our system integrates several well-established visualization techniques with several novel designs to depict co-occurrence patterns in human mobility along with contextual information.

3 Data and Analytical Tasks

In this section, we first describe the data, and summarize a set of analytical tasks that guide the design of our system. Further, we present the method for data preprocessing.

3.1 Data Description

According to the report of international telecommunication union [19], the global penetration rate of mobile phones (i.e. the percentage of active mobile phone users within the population) reached 96% in 2014. With the popularization of mobile phones and the rapid development of data gathering techniques, telecommunication operators collect enormous amounts of data every day, called telco data, offering us unprecedented information resources to study human mobility in terms of the large coverage and fine-grained resolution. Hence, we cooperated with one of the biggest telecommunication operators in China and used the data collected by their operation supporting system (OSS). The data contains all data exchange records between each mobile phone and cell stations when mobile phone users make phone calls, send messages or connect to the Internet. Compared with mobile phone call detail record (CDR) data widely studied previously [2, 27], the popularity of smart phones with various mobile apps running both in the background and foreground nowadays has significantly increased the density of the telco data, which makes it possible for a more comprehensive and in-depth analysis of human mobility.

In this work, we extract useful information from each record of the telco data, including an encrypted unique mobile phone identity ID_{phone}, a timestamp $time$, a cell station identity ID_{base} and the corresponding location of the cell station $< longitude, latitude >$. Then the extracted data can be viewed as a kind of trajectory data of mobile phone users with a large coverage and fine granularity. The dataset covers the whole day of October 21st, 2013 with a 33 GB data size, containing records of over 8.6 million users on 24789 cell stations mapped to 9472 unique locations (i.e. $< longitude, latitude >$), which can cover most of the urban area in Guangzhou, China.

3.2 Task Abstraction

We followed a user-centered design process to develop and improve our visual analytics system iteratively. We worked closely with four domain experts for more than seven months. Two of them are long engaged in the research on sociology and media studies, and the other two telecommunication engineers helped to maintain the collected data and enabled efficient access to the large-scale telco data in a well-structured form using a MapReduce system. The domain experts suggested that a visual analytics system is urgently needed to facilitate analysts in exploring and exploiting potentials in the telco data. Hence, we held weekly meetings and exchanged emails regularly with experts to gather and refine design requirements, present prototypes and collect feedback.

Through frequent exchange of views with the experts, we compiled a list of basic analytical tasks helping us better understand the problem domain, identify challenges faced by target users and inspire the subsequent visual designs:

- T.1 **Global Exploration:** Analysts need to grasp an intuitive overview of co-occurrence patterns of different regions. First, they need to check the distribution of regions based on the number of regions that they co-occur with during a specific time period (T.1.1). Regions with more or fewer co-occurrences with other regions can be discovered intuitively for further exploration. In addition, the distribution of places where co-occurrences happen is also of interest for analysis (T.1.2).
- T.2 **Insight Exploration:** For an identified interesting co-occurrence pattern, analysts wish to form preliminary hypotheses about possible causes. Spatio-temporal characteristics of human mobility at the place where a co-occurrence happens, such as the number of people visiting the place during different periods of a day and the spatial distribution of the regions those people coming from, could be analyzed to generate explanatory hypotheses.
- T.3 **Correlation Exploration:** The major task for correlation exploration in co-occurrences is to extract and visualize the biclusters of co-occurrences in human mobility (Section 3.3) in an intuitive way and navigate analysts to find ones of their interest. Therefore, an exploration from multiple aspects and at different scales needs to be supported, including an overview of all biclusters (T.3.1), a comprehensive analysis based on different attributes (T.3.2), and an exploration of diversity in biclusters (T.3.3).
- T.4 **Detail Exploration:** Appropriate exploration of details needs to be done for practical applications. Particularly, analysts need to explore the regions that co-occur with a specific set of regions (T.4.1), or the places where a specific set of regions co-occur with each other (T.4.2). Meanwhile, biclusters related to a specific set of regions or places may also need to be visualized to facilitate analysis for real world applications (T.4.3).

3.3 Data Modeling

In order to model the telco data to support the identified analytical tasks, we perform the following three steps. In step 1 *Population Mapping*, we map users to different regions based on their mobility and form several user groups for further exploration. In step 2 *Co-occurrence Events Extraction*, we construct a co-occurrence graph to extract co-occurrence events and organize them in a clear structure to support visual analysis in our system. Finally, in step 3 *Correlations Mining*, we transform the constructed co-occurrence graph into a binary matrix so that biclustering algorithms can be applied to mine the correlations in co-occurrence efficiently. In the rest of this section, we discuss these three steps in details.

3.3.1 Population Mapping

To account for the characters of the data, appropriate aggregations need to be done, so as to compensate for the uncertainties in spatial and temporal coverage [1]. Therefore, in the first step, we do population mapping [9] to map the registered users to different regions and form different user groups based on human mobility. We divide 24 hours of a day into regular time intervals, each with 5 minutes in between. Then, for each time interval, we determine individual user's location based on the cell station he/she connects to for the longest time during the interval. If a mobile phone user does not have any records during that interval, the location will be marked as *unknown*. In this way, each user's movement on that day can be represented as a vector $\mathbf{V}_{interval}$ with 288 dimensions. In order to preserve the temporal density for an effective study, we filter out users whose records cover less than 8 hours. By doing so, about 15% of users are removed. After that, we map each individual user to a cell station based on the corresponding vector $\mathbf{V}_{interval}$. Further, we aggregate cell stations at the same location (i.e. same $< longitude, latitude >$) to form 9472 unique physical regions through Voronoi tessellation.

3.3.2 Co-occurrence Events Extraction

After population mapping, users are clustered into several groups based on the regions they are mapped to. Based on regions, we further extract co-occurrence from data. Before describing our method, we define the **co-occurrence event** studied in this paper as follows:

If people from region A and region B visit region C during the same time interval, we say *"region A co-occurs with region B at region C"*.

According to the definition, we can model the co-occurrence in a graph structure $G = (V, E)$ with regions as nodes and co-occurrence events as edges. First, we construct a series of subgraphs, each representing a time interval. In each subgraph, an edge is added to a pair of nodes if and only if there is a co-occurrence event between them. For each edge, a weight vector (ω_A, ω_B) is attached, indicating the number of people from two regions involved in the co-occurrence event respectively. Then subgraphs are aggregated to form a complete co-occurrence graph. Note that there could be multiple edges between a pair of nodes and those edges can be grouped based on the places or time intervals of the corresponding co-occurrence events. In this way, we extract all co-occurrence events and organize them in a clear structure to support various analytical tasks for our system.

3.3.3 Correlations Mining

Beyond a single co-occurrence event, we also study the correlations of co-occurrence. In particular, we try to analyze the phenomena that multiple regions co-occur at a set of different places, which is important for many practical applications. In order to detect such correlations, biclustering techniques are adopted.

First, we transform the subgraph of each time interval k constructed in the previous step into a binary matrix. Let $P = \{p_1, p_2, ..., p_n\}$ be the set of places where co-occurrences happen and $R = \{r_1, r_2, ..., r_m\}$ be the set of regions. Then a binary matrix $\mathbf{B}^k (k = 1, 2, ..., 288)$ is generated, where each row corresponds to a place $p_i (i = 1, 2, ..., n)$ and each column corresponds to a region $r_j (j = 1, 2, ..., m)$. The cell (i, j) of the matrix $\mathbf{B}^k$ is marked as 1 if the region r_j co-occurs with other regions at the place p_i during the time interval k, and as 0 otherwise. In this way, the biclustering algorithm [21] can be easily applied to each binary matrix $\mathbf{B}^k$. The minimum number of regions and places that the discovered biclusters must contain is set as 2 and 1 respectively. With the superior accuracy as well as high computational efficiency of the algorithm, the correlations in co-occurrence can be mined effectively and efficiently. Note that we focus on *closed biclusters* [35] in our study to avoid double counting.

4 System Overview

4.1 Design Rationale

Based on the analytical tasks, we further compiled a set of design requirements with our collaborators:

R.1 **Multi-scale Visual Exploration:** The information seeking mantra "Overview first, zoom and filter, then details on demand" have been widely adopted in analyzing complex and large-scale datasets [31] and should be followed in designing our system. First, an overview is intended to provide analysts a broad awareness of the entire dataset and guide them in choosing where to drill down for further exploration. Then analysts can probe the data using zooming and filtering to inspect detailed information.

R.2 **Multi-perspective Joint Analysis:** For a comprehensive analysis of co-occurrence in human mobility, several analytical tasks (T.1-T.4) often need to be performed alternately, and various information should be provided. Our system employs multiple well-coordinated or linked views showing different data perspectives, thus enabling analysts to combine them efficiently for a multi-perspective joint analysis.

R.3 **Interactive Pattern Unfolding:** Since analysis of co-occurrence in human mobility requires a trial-and-error process, it is crucial for analysts to interact with data directly. Furthermore, to carry out an in-depth analysis, analysts need to compare multiple visualizations side-by-side and interactively manipulate them. They can start with a visualization and incrementally explore associated information conveyed in other linked visualizations. In addition, filtering based on different properties also needs to be enabled in our system.

R.4 **Intuitive Visual Narrative Structure:** A visual narrative structure capable of telling a story intuitively is desired by our collaborators. In this way, they can present their findings more effectively to a wide audience, including government officials, businessmen and sociologists, supported by intuitive visual evidence. Therefore, our system conveys information through well-established and insightful visualizations posing relatively fewer challenges with respect to interpretation. To address the specific needs for our problem, these techniques are further tailored or extended with new features.

4.2 Analytical Pipeline

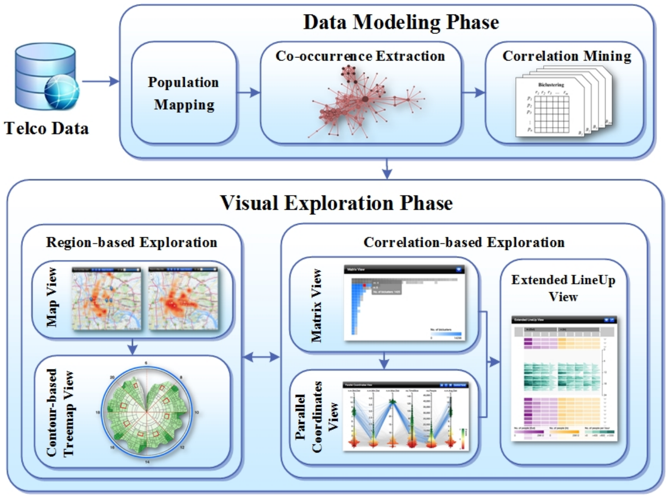


Fig. 2: System pipeline: In the data modeling phase, we perform population mapping, co-occurrence extraction and correlation mining. In the visual exploration phase, five coordinated views are provided to support two basic exploration schemes (i.e. region-based and correlation-based exploration) on co-occurrence in human mobility.

As illustrated in Fig. 2, our system starts with the data modeling phase. We conduct population mapping and extract co-occurrence based on the constructed graph model. In order to study correlations of co-occurrence, we transform the graph into several binary matrices and apply biclustering algorithms for an efficient mining of biclusters. Details of the data modeling phase are presented in Section 3.3. After that, users can perform interactive visual exploration and analysis of co-occurrence in human mobility. Two basic exploration schemes, including region-based exploration and correlation-based exploration, are supported to carry out four kinds of analytical tasks (Section 3.2).

For the region-based exploration, we first provide users with an overview of the co-occurrence intensity for each region in the city (T.1) using Map View, consisting of two linked heat maps (i.e. flow-in map and flow-out map). Users can start with the flow-in map by selecting any regions of interest for detail exploration (T.4). Regions that co-occur with the selected regions will be highlighted on the flow-in map, while the flow-out map shows where the corresponding co-occurrence events happen. Then, in order to gain insights into the causes of such co-occurrence (T.2), we design Contour-based Treemap View, allowing users to explore spatio-temporal characteristics of human mobility at the place where the co-occurrence happens.

For the correlation-based exploration, we visualize biclusters extracted in the data modeling phase which organize the correlations of co-occurrence in a clear structure (T.3). The Matrix View summarizes those biclusters based on their sizes indicating the scale of correlated co-occurrence events. Further, we provide parallel coordinates with Z-score histograms to reveal the distribution of a wide range of related attributes of biclusters in a clear and intuitive way. In addition, we design Extended LineUp View, allowing users to sort biclusters in a linear order and explore their diversity in multiple aspects including the number of people involved and the temporal distribution.

These two exploration schemes can be combined with high flexibility through coordinated or linked views to carry out a comprehensive and in-depth analysis of co-occurrence in human mobility. For example, users can start with Map View to perform the region-based exploration. Once they find a region or a co-occurrence event of interest, they can switch to the correlation-based exploration to analyze the correlated co-occurrence events. Similarly, users can also start with

Matrix View to perform the correlation-based exploration. Once a bi-cluster with interesting patterns is identified, users can further explore related regions through the region-based exploration. In this way, our system simplifies the exploration and interpretation of complicated co-occurrence in human mobility and supports analysts to perform various analytical tasks.

5 Visualization Design

In this section, we describe a set of visualization designs for an in-depth analysis of co-occurrence in human mobility.

5.1 Map View

To support an intuitive and effective region-based exploration, the coordinated relationships between regions that people come from and places where co-occurrence events happen need to be conveyed within spatial context. Meanwhile, due to the scale and noisiness of data, important attributes of co-occurrence, such as the number of regions co-occurring with a specific region or at a specific place, and other meta-data information, should be visually summarized and aggregated to facilitate global exploration (T.1) [11, 44]. In our system, we adopt two linked heat maps, namely flow-out map (Fig. 1a) and flow-in map (Fig. 1b), with rich interactions to support various analytical tasks.

For the flow-out map, we focus on regions where people flow out. We aggregate extracted co-occurrence events to calculate the number of regions that ever have co-occurrence with each region (T.1.1), and then a heat map is generated. The more regions a specific region co-occurs with, the darker the color is. Thus, regions that have more or fewer co-occurrences with other regions can be discovered intuitively. For the flow-in map, we study places where people flow in. Another heat map is generated to show the number of regions which ever co-occur at each place (T.1.2). The darker color means that the corresponding place attracts people from more regions. Note that kernel density estimation is used to derive a smooth representation on the maps. Meanwhile, to provide analysts the flexibility to explore different time periods, our system also includes two interactive time controllers on the top of two maps.

In addition, rich interactions are provided to link the flow-in map and the flow-out map for detail exploration (T.4). Users can click on one or several specific regions on the flow-out map, and regions which have co-occurrence with the selected regions (T.4.1) and places where they co-occur (T.4.2) will be highlighted as nodes on the flow-out map and the flow-in map respectively. Further, we provide alternative encoding schemes for the size of nodes with different attributes (i.e. the number of people involved and the number of regions) to fulfill requirements of various applications. To avoid visual clutter, we adopt a force directed layout similar to Liu et al.'s work [24] and also enable filtering to just keep the regions, places or time periods of interest.

5.2 Contour-based Treemap View

The design goal of this view is to present spatial and temporal characteristics of human mobility at a specific place. As suggested by domain experts, we present the following features including the number of people visiting the place during different period of a day, the distance between the place and the region they come from, and the loyalty of those regions (i.e. the frequency of people from a certain region visiting the place). These features facilitate analysts to gain insights into co-occurrence patterns and generate explanatory hypotheses (T.2).

Based on this design goal, we take consideration of a few alternative designs and evaluate prototypes informally with our collaborators. Radial visualizations are commonly used to encode spatio-temporal information [24], which provides an intuitive visual metaphor. Therefore, we also adopt a radial layout to visualize characteristics of a place during different time periods simultaneously. Our first prototype is implemented using a circular treemap [41]. However, the wasted space undermines the efficacy of this design. Our next idea is to use sunburst tree [33] where the radial space is divided into several segments with each indicating one time period. Then information for each time period is aggregated and shown in different segment. But the major drawback of this design is that it is challenging to choose the length of each time period. Our collaborators point out that a long time period leads to the loss of detailed temporal information which is crucial to distinguish different places, while a short time period damages the legibility of each segment.

Ultimately, we come up with the design of a contour-based treemap. The followings elaborate its construction procedure:

Create temporal contour The first step is to create a contour line to visualize changes in the number of people visiting the place during a day. A series of spikes (e.g. 288 spikes in this paper) are radially arranged around a center clockwise. Each spike represents a time interval in Section 3.3 with its length encoding the number of people visiting the place during that time interval. Furthermore, to generate a more reliable representation for the underlying data, an area-preserving mapping is adopted by taking the square root of the linearly computed length value of each spike. Then the tips of spikes are sequentially connected via a polyline. Note that a gentle smoothing by kernel density estimation might be needed to create a smooth contour line. Thus, the detailed temporal distribution can be conveyed, which helps to distinguish different types of places intuitively.

Segment temporal space Based on the constructed temporal contour, the information of different time intervals can be aggregated into a longer time period for a better legibility, while preserving detailed temporal distribution. We divide the radial space into several segments to indicate different time periods. In order to provide flexibility for different applications, we allow users to interactively select the size of the segment as well as the start/end time, which is a trade off between analysis granularity and aesthetics.

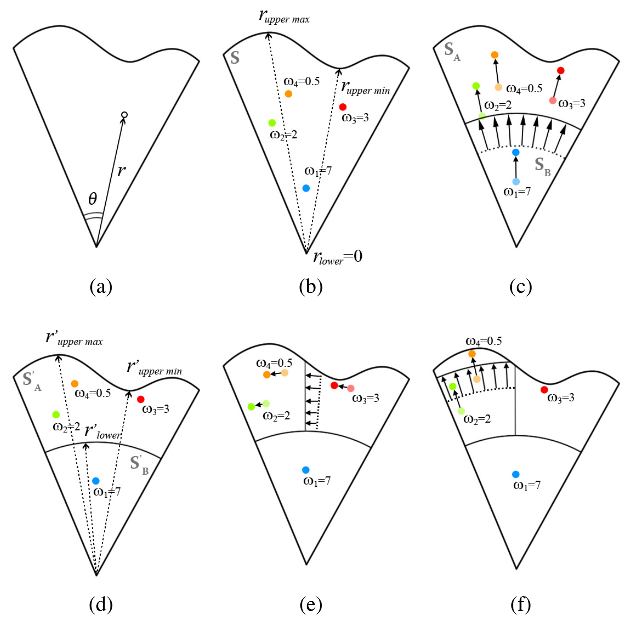


Fig. 3: Overview of the contour-based treemap layout process: (a) Scheme for point position assignment; (b) Initially, four points representing the regions are positioned inside the enclosing segment S. (c)-(d) As the aspect ratio $R > 1$, we bisect S along the direction of arc into two sectors S_A and S_B. Based on the summation of the weights of points in S_A and S_B, scaling is executed so that each sector presents areas proportional to the sum of weights. (e)-(f) This slice-scale process is executed iteratively with radial and arcing cuts, until each point is assigned to a sector with area proportional to its weight.

Assign positions to points The constructed temporal contour together with temporal segments encloses the area to position points. These points represent regions that people visiting the place come from during a certain time period, and provide relative positions for constructing sectors of the treemap. As illustrated in Fig. 3a, for a point in a segment, its distance r from the center and the angle θ measured clockwise from the start of the segment encode the distance and loyalty of the corresponding region (defined in Section 5.2) respectively. Alternatively, r and θ can be used to encode the same attribute (i.e. distance or loyalty) to obtain a more intuitive observation of its distribution. Meanwhile, a weight ω is attached to each point indicating the number of people visiting this place during that period.

Treemap layout algorithm For the layout algorithm, we follow the basic design principles of visualization, including aesthetics, legibility and faithfulness. For aesthetics principle [5], the distribution of the treemap sectors must fully utilize the given area, by avoiding holes and overlapping. The legibility principle is employed to create a proper layout with appropriate aspect ratio of each sector such that information can be clearly presented. Moreover, the faithfulness principle ensures that the spatial and temporal characteristics of human mobility at a specific place can be conveyed correctly.

First, we consider the layout algorithm based on Voronoi tessellations. However, the potential concave shape of the temporal contour makes it hard for Voronoi tessellations to be applied. Several approaches are turned down because of the violation of the faithfulness principle. For example, we can split a concave shape into multiple convex parts and apply Voronoi tessellations to each part respectively. However, it is difficult to partition all points into different parts given the proportional relationship between their weights ω and the enclosed area of each part. In addition, concave shapes can also be converted into convex through interpolation. After applying Voronoi tessellations, the interpolated parts need to be removed, which is very likely to change the contour we constructed.

Therefore, in our system, we extend *Nmap* algorithm [10] to fit a contour-based treemap, which is an iterative slice-scale process on the radial space. Fig. 3b-f illustrate the detailed process. In this example, there are 4 points (i.e. p_1, p_2, p_3, p_4) attached with weight values (i.e. $\omega_1, \omega_2, \omega_3, \omega_4$) in a certain segment S. First, we calculate the aspect ratio R as follows:

$$R = \frac{\bar{r}}{\bar{L}_{arc}} = \frac{r_{upper} - r_{lower}}{\theta \cdot (r_{upper} + r_{lower})/2}$$

where θ is the corresponding angle of the segment. r_{upper} and r_{lower} are the radii of the upper and lower bound of the segment. When dealing with the contour, r_{upper} can be calculated as

$$r_{upper} = (r_{upper_{max}} + r_{upper_{min}})/2.$$

If the aspect ratio $R > 1$, we bisect S along the direction of arc into two sectors S_A and S_B, radially otherwise. To find the position of the bisector, the points are sorted in an ascending order based on their radii for the radial bisection, and based on their angles for the arc bisection. Then, they are added one by one to S_A until the difference of weight $|2 \cdot \omega_A - \omega|$ is minimized, where $\omega_A = \sum_{p_i \in S_A} \omega_i$ and $\omega = \sum_{p_i \in S} \omega_i$. The remaining points are assigned to S_B. The position of the bisector is calculated as the average of the largest radius (or angle) of the points in S_A and the smallest one in S_B. Next, the sectors S_A and S_B are scaled to create new sectors S'_A and S'_B with areas proportional to ω_A and ω_B. Note that when bisecting along the direction of arc for a sector related to the contour, we should first test the concavity of the contour. If the contour is concave, the position of the bisector should not exceed $r_{upper_{min}}$ in order to preserve the connectivity of the new sector, otherwise a radial bisection should be done instead. This process is repeated until each sector contains only one point. Then each sector is colored based on the distance between its corresponding region and the place.

Fig. 4 shows the contour-based treemap based on our design. By observing the shape of the contour and the color distribution of containing sectors, users can get an intuitive idea about when there are more people visiting the place and how far those people coming from, so that users can make preliminary hypotheses about the place. Moreover, when users hover on a sector corresponding to a specific region which has people visiting the place during a specific time period, all sectors corresponding to the same region in other segments will be highlighted (Fig. 1d). This helps to correlate time and region as well as provide more details. Further, users can also zoom and click on a segment to mark the location of the corresponding region on the flow-out map, which makes the exploration process convenient.

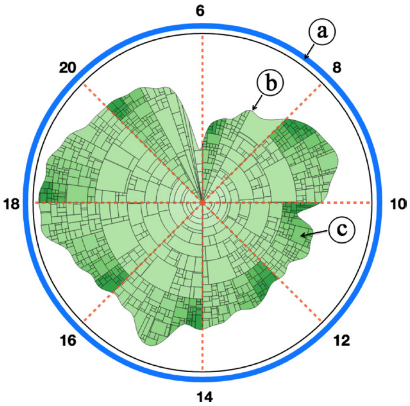


Fig. 4: Contour-based Treemap View: (a) A circular time axis divides a radial space into segments representing time periods (e.g. each for 2 hours); (b) A temporal contour conveys detailed temporal distribution of people visiting the place; (c) Each sector in a segment represents a region which has people visiting this place during the corresponding time period. The size of the sector encodes the number of people, and the color encodes the distance between the region and the place. The darker the color, the longer the distance.

5.3 Matrix View

For an effective and efficient correlation exploration of co-occurrence, the first thing we need is an overview of the extracted biclusters to navigate users for further exploration (T.3.1). As the sizes of those biclusers (i.e. how many regions co-occur at how many different places) are of particular interest to our domain experts, we organize them in the form of an (m, n) matrix, where m indicates the number of regions that co-occur with each other and n indicates the number of places where co-occurrence events happen. After that, a matrix-based visualization is integrated into our system to provide a simple and efficient visual summarization for a large number of biclusters. Fig. 6f shows Matrix View based on our design. The color of each cell (e.g. (m, n)) represents the number of biclusters of the corresponding size (e.g. m regions co-occur at n different places). The darker color indicates a larger number of biclusters. Based on the color distribution of the matrix, analysts can get an overall understanding about the scale of co-occurrence in a city. Furthermore, the detailed information can be obtained by zooming and hovering on a cell to activate a tooltip.

5.4 Parallel Coordinates View

Domain experts have identified several important attributes for a bicluster of co-occurrence, including the size, the frequency, the number of people involved and some other statistical features. They suggest that a quantitative analysis based on these features should be integrated into our system to help users conduct comprehensive exploration of correlations in co-occurrence (T.3.2). Parallel coordinates are adopted in our system as they are widely used for multivariate data visualization and each bicluster can be conveniently presented as a polygonal line. Fig. 1e shows Parallel Coordinates View based on our design.

In order to reveal hidden patterns intuitively, two challenging issues of parallel coordinates, including axis ordering and visual clutter, are particularly considered in our system. For axis ordering, different orders of axes implicitly reveal different aspects of the dataset and the order in which the axes are drawn is critically important for effective visualization. In our system, we employed a correlation-based dimension sorting algorithm [45] to generate a recommended axis order automatically. After that, users can easily drag any axis to make adjustment based on different applications. To mitigate the visual clutter problem caused by a large number of lines rendered, several methods are integrated into our system. First, inspired by angular histogram [14], we aggregate the data by binning techniques, then the density of underling polylines can be visually presented by a histogram along each axis. However, based on users' feedback, rotated histogram bars do not fit their work habits and make them confusing. Thus, our system just uses color of histogram to encode Z-score [26], a standard statistic measurement suggested by domain experts, so as to reveal the data distribution for each dimension and navigate further exploration. When users brush on the axis, the corresponding polylines in the range will be shown by alpha blending [40]. In addition, we also support data filtering through linked views to reduce visual clutter.

5.5 Extended LineUp View

In our system, both Matrix View and Parallel Coordinates View can be used to explore biclusters of co-occurrence from different aspects and at different scales. Furthermore, we need to further analyze the diversity of each bicluster (T.3.3) based on regions where people come from and places where co-occurrence events happen. Thus, it is desirable to show the regions and places contained in each bicluster clearly with statistical information. For each region or place, our collaborators would like to know the number of people involved and the corresponding temporal distribution, which helps them to estimate the importance of the bicluster and form more detailed explanatory hypotheses.

Therefore, we employ a LineUp-style chart [16] to sort all biclusters into a linear order based on their sizes, allowing users to conveniently explore the diversity relative to each other. Fig. 1f shows Extended LineUp View based on our design. Owing to the limited space, we embed colored mosaics into each bar of LineUp to represent regions (as purple mosaics) or places (as yellow mosaics) contained in each bicluster. The intensity of color is determined by the number of people involved. The darker the color, the more people involved. In this way, analysts can get an intuitive overview of the diversity of a large number of biclusters in the limited space.

In the mean time, to analyze the corresponding temporal distribution, horizon graph, a timeline visualization technique shown to be more effective than the standard line chart in a limited space [18], is further embedded. Users can simply click on a row which represents a bicluster, and the row could be expanded vertically with a series of horizon graphs to show the temporal distribution of people involved from different regions or at different places.

5.6 Interactions

Based on the design rationale (R.3), our system should enable analysts to interact with data and facilitate exploration with flexibility. Apart from basic interactions in each view, TelCoVis also supports the following interactions:

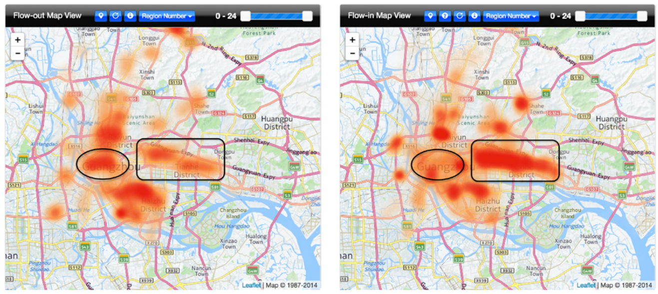
Linking The system supports automatic linking among five proposed views not only to support interactive pattern unfolding (R.3), but also to facilitate multi-perspective joint analysis (R.2). For example, if a specific set of regions on the flow-out map or places on the flow-in map are selected, the cell in Matrix View that contains related biclusters and the corresponding lines in Parallel Coordinates View will be highlighted, and Extended LineUp View will be updated accordingly, which enable users to perform a targeted analysis of correlations in co-occurrence conveniently.

Filtering Filtering enables analysts to focus on important information, especially when handling data of large scale and with uncertainties. TelCoVis allows analysts to interactively eliminate less important information through different views and from different aspects.

Configuration Users can configure our system to choose parameters for different views. For example, in the parallel coordinates, users can choose which attributes shall be shown on display. In addition, in the contour-based treemap, the time period of interest and smoothing parameters can be set for different applications.

6 Case Studies

We implement a web-based system, TelCoVis, using Javascript. To evaluate the system, we carry out four case studies based on telco data collected on Oct. 21, 2013 in Guangzhou, China. The data modeling phase takes about 2.5 hours, and then our system can run in real time.



(a) Flow-out map (b) Flow-in map

Fig. 5: Region-based exploration of co-occurrence in Guangzhou on Oct. 21st, 2013. Regions in darker color in the flow-out map (a) have more co-occurrences with other regions, while more co-occurrences happen at places in darker color in the flow-in map (b).

6.1 Global Exploration of Co-occurrence in the City

With Map View and Matrix View, our system provides analysts with a city-level abstraction of co-occurrence in human mobility (T.1).

For region-based exploration, by comparing the flow-out map and flow-in map, we can see that the region highlighted in a black circle is obviously darker in the flow-out map (Fig. 5a), indicating people from that region often co-occur with people from other regions at other places, while only a few people go there and co-occur at that region (T.1.1). The reason might be that this region is a residential area with few public facilities (e.g. shopping mall). In contrary, another big region highlighted in a black box in the flow-in map (Fig. 5b) is darker. This is the most bustling downtown area in Guangzhou where people across the city come and co-occur there (T.1.2).

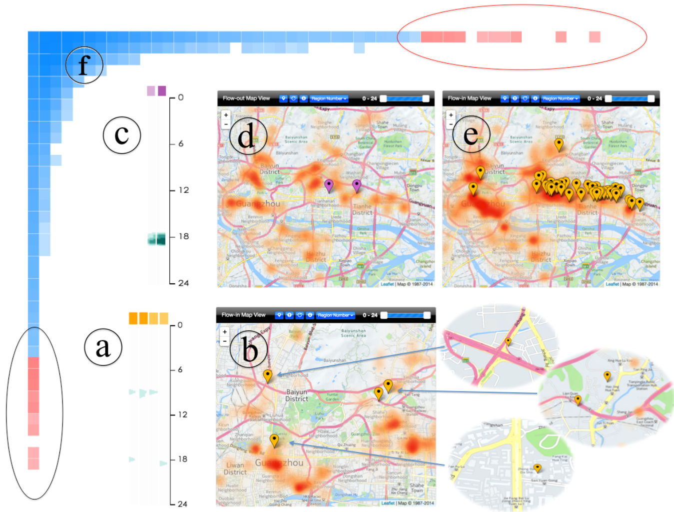
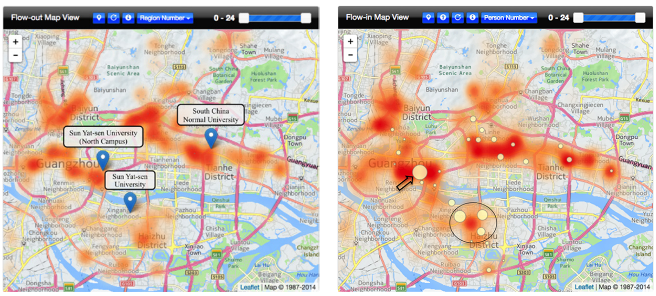


Fig. 6: Correlation-based exploration of co-occurrence in Guangzhou on Oct. 21st, 2013. The outliers (i.e. biclusters with big m or big n) might be caused by traffic congestions during peak hours.

For the correlation-based exploration, in Matrix View (Fig. 6), we can have an overview of correlations of co-occurrence (T.3.1) in the form of an (m,n) matrix w.r.t. the size of biclusters. The distribution looks like a power law curve with few biclusters of big m or big n. This observation supports the hypothesis that it is less likely that a large number of regions co-occur with each other or co-occur at a lot of different places. However, some outliers exist. First, we look into the outliers of big m and small n (i.e. $m \geq 35, n = 1$) (highlighted in black circles). By clicking on those cells, the Extended LineUp View is generated. According to the expanded horizon graphs (Fig. 6a), we can see that such correlated co-occurrences mainly happen during peak hours. Further, we click the yellow mosaics in the Extended LineUp View to highlight the places where these co-occurrence events happen on the flow-in map. We can see that four interchanges of the highways are highlighted (Fig. 6b). Thus, we infer that traffic jams at those interchanges during peak hours might cause such correlated co-occurrences. Then, for the outliers of big n and small m (i.e. $n \geq 40, m = 2$) (highlighted in red circles), a similar exploration process is applied. Surprisingly, we find that only two regions in downtown are involved in these outliers (Fig. 6d), and most co-occurrence events happen near those two places (Fig. 6e) during peak hours in the evening (Fig. 6c). A similar hypothesis can also be made on traffic congestions in downtown area.

6.2 Co-occurrence of College Students

In this case, we are particularly interested in the co-occurrence of college students. Three campuses, including two campuses of Sun Yat-sen University and the campus of South China Normal University, are selected on the flow-out map (Fig. 7a). Then, in the flow-in map (Fig. 7b), all the places where students co-occur are highlighted with nodes of different size encoding the number of students (T.4.2). We can see that college students from those three campuses only co-occur at a few places. Apart from the places near the main campus of Sun Yat-sen University (highlighted in black circles), many students co-occur at another place (highlighted with a black arrow). In order to gain deeper insights (T.2), we click on this node and a contour-based treemap is generated (Fig. 8b). By observing the treemap, we infer that this place could be a popular shopping center (refer to Section 6.4) among youngsters. This hypothesis is verified by local people.



(a) (b)

Fig. 7: Co-occurrence of college students from (a) three main campuses in Guangzhou. (b) The places where they co-occur are highlighted with nodes of different size encoding the number of students.

6.3 Co-occurrence of People of Different Income Level

Our collaborators are keenly interested in the co-occurrence of people of different income levels. In this case study, we choose four regions suggested by our domain experts. Two of them are of high income level, and the other two are of relatively low income level. Fig. 1 shows the visualizations generated by our system. In Map View, by observing the highlighted nodes, we see that those four regions co-occur with many different regions (Fig. 1a) (T.4.1) and at many different places (Fig. 1b) (T.4.2) across the city. However, the sizes of most nodes are small indicating the limited number of people involved, while a few nodes in the flow-in map (Fig. 1b) are larger than others, which indicates more people co-occur at those places. By clicking on those nodes, the corresponding contour-based treemaps are generated (one of them shown in Fig. 1d, corresponding to the node highlighted with a black arrow in Fig. 1b). Based on the treemaps, we infer that these places are likely to be shopping centers (refer to Section 6.4) (T.2).

We further explore the correlations of co-occurrence. In Matrix View (Fig. 1c), all cells containing biclusters related to those four regions are colored in blue (T.4.3). Our collaborators are more interested in those biclusters of median size. Thus, we click on a cell of $m = 4, n = 5$, indicating those four regions co-occur at five different places. Then all related biclusters of such size are extracted and displayed as polylines in Parallel Coordinates View (Fig. 1e). Based on the embedded Z-score histograms, we can clearly see that the average number of people involved is about 8,000, and the average number of time intervals during which co-occurrences happen is around 40 (about 3 hours) (T.3.2). Then, we brush on the $no.People$ axis and $no.TimeSlice$ axis to choose the values around average, corresponding to the yellow area on the axis (i.e. Z-score is around 0), so as to focus on the average pattern of people. In addition, we are more interested in the part of biclusters with long travelling distances (corresponding to large values on $n.m.Avg.Dist$ axis) which might lead to some interesting patterns. Based on these choices, Extended LineUp View (Fig. 1f) is updated accordingly for diversity exploration (T.3.3). In the LineUp view, we observe that the color of first two purple mosaics are relatively darker than the other two. By clicking on those mosaics, we find first two mosaics correspond to two regions of low income. We further expand to show horizon graphs, and find two different temporal patterns. More people come from regions of high income level around 3 pm, while coming from regions of low income level during morning and evening peak hours. These two patterns also appear at the places where co-occurrences happen.Thus, we infer that these three places (correspond to three yellow mosaics highlighted with a black box in Fig. 1f) might be shopping centers where people of high incomes go for afternoon tea, while the other two places might be two metro stations where people of low incomes go during peak hours. Through linked views, we further cross verify our hypothesis.

6.4 Exploration of Region Functions

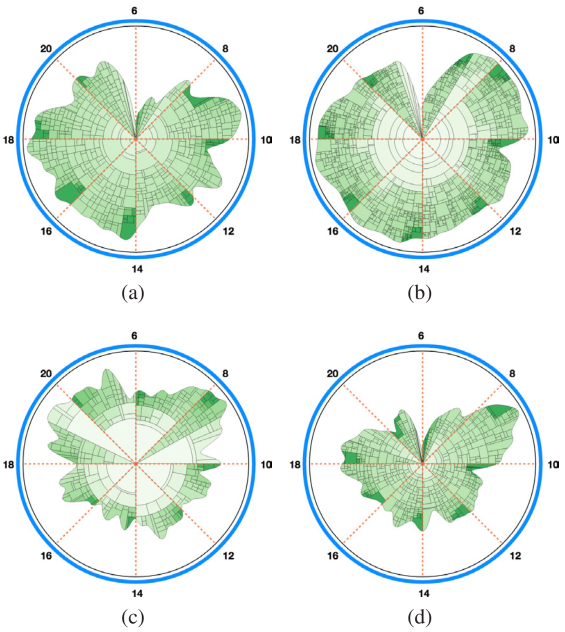
In this case, we explore functions of regions based on the design of contour-based treemap, which provides important insights for domain experts to form explanatory hypotheses (T.2). According to the local people's input, we picked some representative regions of different functions (e.g. shopping centers, office blocks, residential areas and metro stations) in Guangzhou. Note that here we focus on the time span from 6 am to 10 pm. During that period, there are more active mobile phone users, which implies the major time period for people's activities and is essential for the exploration of region functions.

Fig. 8a and Fig. 4 show treemaps for two well-known office blocks (i.e. Zhujiang New Town and White Swan Pond area), while Fig. 8b and Fig. 1d show treemaps for two popular shopping centers (i.e. China International Plaza and Yuefu Plaza). Despite the fact that there are always many people visiting these places, treemaps for shopping centers (Fig. 8b and 1d) have more large sectors in light green, which indicates more visitors coming from nearby regions. Through interactions with those sectors, we found most of them are corresponding to nearby neighborhoods, which implies that shopping centers are likely to be built next to neighborhoods for more customers in Guangzhou. Based on this observation, we can distinguish office blocks and shopping centers effectively. Fig. 8c shows a treemap for a residential area in which the size of the upper part is larger than the lower part, since more people stay in the residential area before 9 am and after 7 pm. In addition, a treemap for a metro station is shown in Fig. 8d, where there are obvious morning and evening peak hours.

In summary, a contour-based treemap can be viewed as a visual signature characterizing the spatial and temporal distribution of human mobility at the corresponding region, which helps analysts get an idea of the major regional function. We plan to further extend and improve it for more applications in the future.

7 Expert Interview

To evaluate the effectiveness of TelCoVis, other than our collaborators, we demonstrated our system and presented our use cases to three domain experts and conducted one-on-one interviews with them to collect their feedback. One of them is from university C (Expert A)



(a) (b) (c) (d)

Fig. 8: Contour-based treemaps, visual signatures of different types of regions based on human mobility: (a) Office block; (b) Shopping center; (c) Residential area; (d) Metro station.

who has worked on sociology and media studies for more than fifteen years, another from an inspection and quarantine bureau in China (Expert B), and the last one is a project manager working on business intelligence (Expert C). Each interview lasted for about 1 hour. All of the experts appreciated the idea of exploring co-occurrence in human mobility based on telco data and were intrigued by our system. Their feedback is summarized as follows:

Interactive Visual Design All our domain experts confirmed that the system is nicely designed according to the problem domain and the characteristics of telco data. Expert C commented "TelCoVis is not simply new visualization designs. Rather, it is a combination of visualization techniques together with navigation and interaction techniques to provide a comprehensive system useful to explore co-occurrence in human mobility. This kind of integrated methodology, visualization coupled with flexible navigation, is becoming more and more prevalent to address challenging problems nowadays." Expert B added "It is an excellent idea to display the information of co-occurrence in multiple visualizations and support flexible exploration schemes through rich interactions. Differing the visualization based on the nature of the information and analytical tasks would greatly enhance users' understanding." In addition, Expert A considered correlation exploration valuable for practical applications. He further highlighted the design of the contour-based treemap and said "This view is aesthetic, intuitive, and useful to explore functions of different regions. It could be interesting to extend it to various applications." Meanwhile, all experts believed that the visualizations adopted in our system can be readily comprehended by users with different background.

Applicability and Improvements All domain experts expressed interests in applying our system to deal with practical problems in their domains. Expert B highlighted "Nowadays, a common way to cope with threats of contagious diseases is through interviews with people infected, which is neither effective nor efficient. With access to real time data, this system has a great potential to be applied to improve the emergency response system for contagious diseases in a revolutionary way." Expert C also commented "TelCoVis is highly suitable to support business intelligence." An example of application might be that the advertiser could pre-select the type of customers he/she is targeted for an advertisement, and TelCoVis would facilitate him/her to make sensible decisions based on where and when those people co-occur. Furthermore, Expert A suggested to apply our system on a dataset covering a longer time span, so that we could have enough information to access routine behaviors of people which is an important aspect in sociology studies. In addition, by integrating other information, such as social media data, more advanced analytical tasks can be carried out for various applications.

8 Discussion

The case studies demonstrate the advantages of TelCoVis to explore co-occurrence in urban human mobility. The combination of the different visualizations enables analysts to integrate various information for an analysis from different aspects and at different scales.

Our research is still in progress. Weaknesses have been observed and will be addressed in the future. First, our system focuses on visualization techniques and lacks sufficient support for automatic analysis. In many cases, we have to examine data manually with our system to discover patterns, which is like to find a needle in a haystack. If more advanced data mining techniques are incorporated, for example, by recognizing people's routine behavior patterns to identify regions for home or work, TelCoVis will be capable of performing more sophisticated analytical tasks. Second, we use color intensity to encode the number of people and biclusters in Extended LineUp View and Matrix View respectively, but users can only distinguish a few color intensities efficiently. Thus, detailed information should be explored through interactions supported by our system. Third, to get the full potential of the parallel coordinates, more advanced correlation metrics and axis ordering algorithms should be explored. Moreover, our system will be faced with scalability issues when the data grows (e.g. when handling data of multiple days, or with large number of regions). To cope with this problem, we can extend the filtering of regions or time periods in the Map View to the whole system. More levels of detail and abstraction can also be introduced to handle this problem. Besides, despite the unprecedented coverage and granularity, we acknowledge that the information derived may underestimate the full repertoire of co-occurrence in human mobility at the spatial and temporal scale considered. Nonetheless, this issue is common to all existing works for tracking human mobility at a fine spatial scale. Additionally, based on our telco data, we do not have enough information to access precise locations of individuals, but our domain experts consider the spatial granularity based on cell stations (i.e. 0.2-0.5 km) as acceptable.

9 Conclusion and Future Work

In this paper, a visual analytics system, TelCoVis, is presented to facilitate the exploration of co-occurrence in human mobility based on telco data. Some well-established visualization techniques, such as heat map, matrix, parallel coordinates and LineUp, are extended and integrated into our system. In addition, we have also proposed a novel contour-based treemap as a visual signature characterizing the spatio-temporal characteristics of human mobility at a specific place. Combining advanced visualization techniques with intuitive user interactions, our system not only makes it easier and more efficient for domain experts to perform a series of analyses, but also enables a new way to explore the data from multiple levels and perspectives. Our system has a wide range of applications in various domains, including urban planning, business intelligence and social behavior analysis.

There are multiple venues for future work. First, we plan to extend our system to support analysis of real time telco data for time-critical applications. In addition, we intend to test our system on a larger dataset (e.g. covering multiple days), and conduct controlled experiments with quantitative measurements to collect more feedback from end users for further improvement.

Acknowledgments

The authors thank the anonymous reviewers for their valuable comments. This research was supported in part by HK RGC GRF 16208514, the National Basic Research Program of China (973 Program) under Grant No. 2014CB340304, and a grant from Noahs Ark Lab, Huawei Technologies Investment Co. Ltd.

REFERENCES

[1] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel. *Visual analytics of movement*. Springer Science & Business Media, 2013.

[2] G. Andrienko, N. Andrienko, M. Mladenov, M. Mock, and C. Politz. Discovering bits of place histories from people's activity traces. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 59–66. IEEE, 2010.

[3] G. Andrienko, N. Andrienko, H. Schumann, and C. Tominski. Visualization of trajectory attributes in space–time cube and trajectory wall. In *Cartography from Pole to Pole*, pages 157–163. Springer, 2014.

[4] N. Andrienko, G. Andrienko, L. Barrett, M. Dostie, and P. Henzi. Space transformation for understanding group movement. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2169–2178, 2013.

[5] M. Balzer, O. Deussen, and C. Lewerentz. Voronoi treemaps for the visualization of software metrics. In *Proceedings of the 2005 ACM symposium on Software visualization*, pages 165–172. ACM, 2005.

[6] F. Calabrese, L. Ferrari, and V. D. Blondel. Urban sensing using mobile phone network data: A survey of research. *ACM Computing Surveys (CSUR)*, 47(2):25, 2014.

[7] F. Calabrese and C. Ratti. Real time rome. *Networks and Communication studies*, 20(3-4):247–258, 2006.

[8] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.

[9] P. Deville, C. Linard, S. Martin, M. Gilbert, F. R. Stevens, A. E. Gaughan, V. D. Blondel, and A. J. Tatem. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45):15888–15893, 2014.

[10] F. S. Duarte, F. Sikansi, F. M. Fatore, S. G. Fadel, and F. V. Paulovich. Nmap: A novel neighborhood preservation space-filling algorithm. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2063–2071, 2014.

[11] N. Elmqvist and J.-D. Fekete. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *Visualization and Computer Graphics, IEEE Transactions on*, 16(3):439–454, 2010.

[12] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2149–2158, 2013.

[13] P. Fiaux, M. Sun, L. Bradel, C. North, N. Ramakrishnan, and A. Endert. Bixplorer: Visual analytics with biclusters. *Computer*, (8):90–94, 2013.

[14] Z. Geng, Z. Peng, R. S. Laramee, J. C. Roberts, and R. Walker. Angular histograms: Frequency-based visualizations for large, high dimensional data. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2572–2580, 2011.

[15] C. Gorg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *Visualization and Computer Graphics, IEEE Transactions on*, 19(10):1646–1663, 2013.

[16] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2277–2286, 2013.

[17] D. Guo and X. Zhu. Origin-destination flow data smoothing and mapping. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2043–2052, 2014.

[18] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1303–1312. ACM, 2009.

[19] International Telecommunication Union. *World Telecommunication Development Conference (WTDC-14): Final Report*, 2014. `http://www.itu.int/pub/D-TDC-WTDC-2014`.

[20] C. Kang, S. Gao, X. Lin, Y. Xiao, Y. Yuan, Y. Liu, and X. Ma. Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In *Geoinformatics, 2010 18th International Conference on*, pages 1–7. IEEE, 2010.

[21] A. Király, J. Abonyi, A. Laiho, and A. Gyenesei. Biclustering of high-throughput gene expression data with bicluster miner. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 131–138. IEEE, 2012.

[22] J. M. Krisp. Planning fire and rescue services by visualizing mobile phone density. *Journal of Urban Technology*, 17(1):61–69, 2010.

[23] J. Lanir, P. Bak, and T. Kuflik. Visualizing proximity-based spatiotemporal behavior of museum visitors using tangram diagrams. In *Computer Graphics Forum*, volume 33, pages 261–270. Wiley Online Library, 2014.

[24] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. M. Ni. Visual analysis of route diversity. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 171–180. IEEE, 2011.

[25] 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*, 13(1):185, 2013.

[26] M. L. Marx and R. J. Larsen. *Introduction to mathematical statistics and its applications*. Pearson/Prentice Hall, 2006.

[27] J. Pu, P. Xu, H. Qu, W. Cui, S. Liu, and L. Ni. Visual analysis of people's mobility pattern from mobile phone data. In *Proceedings of the 2011 Visual Information Communication-International Symposium*, page 13. ACM, 2011.

[28] C. Ratti, S. Williams, D. Frenchman, and R. Pulselli. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B Planning and Design*, 33(5):727, 2006.

[29] R. Santamaría, R. Therón, and L. Quintales. Bicoverlapper: a tool for bicluster visualization. *Bioinformatics*, 24(9):1212–1213, 2008.

[30] Z. Shen and K.-L. Ma. Mobivis: A visualization system for exploring mobile data. In *Visualization Symposium, 2008. PacificVIS'08. IEEE Pacific*, pages 175–182. IEEE, 2008.

[31] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[32] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[33] J. Stasko and E. Zhang. Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pages 57–65. IEEE, 2000.

[34] M. Sun, L. Bradel, C. L. North, and N. Ramakrishnan. The role of interactive biclusters in sensemaking. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1559–1562. ACM, 2014.

[35] M. Sun, C. North, and N. Ramakrishnan. A five-level design framework for bicluster visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1713–1722, 2014.

[36] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2565–2574, 2012.

[37] V. A. Traag, A. Browet, F. Calabrese, and F. Morlot. Social event detection in massive mobile phone data using probabilistic location inference. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 625–628. IEEE, 2011.

[38] G. M. Vazquez-Prokopec, D. Bisanzio, S. T. Stoddard, V. Paz-Soldan, A. C. Morrison, J. P. Elder, J. Ramirez-Paredes, E. S. Halsey, T. J. Kochel, T. W. Scott, et al. Using gps technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PLOS one*, 8(4), 2013.

[39] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering. Visual traffic jam analysis based on trajectory data. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2159–2168, 2013.

[40] E. J. Wegman and Q. Luo. *High dimensional clustering using parallel coordinates and the grand tour*. Springer, 1997.

[41] K. Wetzel. Pebbles-using circular treemaps to visualize disk usage. *URL: http://lip. sourceforge. net/ctreemap. html*, 2, 2003.

[42] W. Wu, Y. Zheng, H. Qu, W. Chen, E. Grller, and L. M. Ni. Boundaryseer: Visual analysis of 2d boundary changes. In *Proceedings of IEEE VAST 2014*. IEEE, 2014.

[43] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu. Visualizing interchange patterns in massive movement data. In *Computer Graphics Forum*, volume 32, pages 271–280. Wiley Online Library, 2013.

[44] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins. # fluxflow: Visual analysis of anomalous information spreading on social media. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1773–1782, 2014.

[45] X. Zhao and A. Kaufman. Structure revealing techniques based on parallel coordinates plot. *The Visual Computer*, 28(6-8):541–551, 2012.