

Human-centric Computing and Information Sciences

July 2021 | Volume 11



www.hcisjournal.com



Hum. Cent. Comput. Inf. Sci. (2021) 11:31.

<https://doi.org/10.22967/HCIS.2021.11.031>

Received February 18, 2021; accepted July 18, 2021; published July 30, 2021

Visual Analysis of Steady-State Human Mobility in Cities

Lei Shi^{1,*}, Zhichun Guo², Tao Jiang³, Ruiyu Fang⁴, Yang Chen⁴, Ye Zhao⁵, Xiatian Zhang⁶, and Min Dai^{6,*}

Abstract

Cities are living systems where urban infrastructures and their functions are defined by and evolved with population behavior. Visualizing, profiling, and comparing the mobility behavior of the population has been challenging because of the enormous population size in modern cities (tens of millions in our dataset). This paper proposes a steady-state visual analysis of human mobility that abstracts the longitudinal trajectory of each city resident into five information-theoretic metrics: Fluidity, vibrancy, Commutation, diversity, and density (FACET). The metrics characterize the long-term mobility behavior of residents concerning municipal structures and points of interest in the city and can also be aggregated to profile underlying city regions. Based on the steady-state analysis method, we develop a visualization system, namely UrbanFACET, which provides a multifaceted panorama of human mobilities in cities and helps to compare urban functions among cities and time. We evaluate the proposed method and system through case studies in real-world big cities. Our result demonstrates the effectiveness of the steady-state analysis in several target domains such as urban planning, business site configuration, and city security surveillance.

Keywords

Urban Data Visualization, Human Mobility Analysis, Information Theory

1. Introduction

Analyzing human trajectories in modern cities illuminates the path towards a thorough understanding of mobility in the urban environment. Governments and corporations have benefited a lot in their daily jobs, including urban planning, business site configuration, and public security surveillance. Each individual human trajectory is often modeled as interleaving segments of stays and travels [1]. The travel part of trajectories has been intensively studied under the topic of movement analysis and visualization. On the other hand, the stay part of human trajectories is often neglected, accounting for up to 90% of location information in the urban trajectory dataset. Visualizing stay segments of human trajectory in cities becomes more challenging as it can involve billions of records as in our data.

* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Corresponding Author: Lei Shi (leishi@buaa.edu.cn) and Min Dai (min.dai@talkingdata.com)

¹School of Computer Science and Engineering and SKLSDE, Beihang University, Beijing, China

²Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

³Institute of Software, Chinese Academy of Sciences, Beijing, China

⁴School of Computer Science, Fudan University, Shanghai, China

⁵Department of Computer Science, Kent State University, Kent, OH, USA

⁶Department of Data Science, Talking Data Inc., Beijing, China

In the literature, trajectory aggregation methods have been widely applied in visually exploring human movement data, e.g., trajectory clustering, origin-destination (OD) flows [2], and mobility networks [3]. These methods could significantly reduce visual clutter by displaying a vast amount of travel segments in human trajectories, yet can not be directly applied to the visualization of stay segments. First, trajectory aggregation methods are generally designed to illustrate short-term human displacements. Second, the aggregation methods do not explain the semantic interaction between human movements and the underlying socioeconomic and urban demographic factors. By hiding movement details while preserving the semantic context of human mobility, the visualization of stay segments could better present complex human mobility patterns in urban environments.

This work targets a class of less investigated yet critical research problems: How to summarize the stay segments of individual human movement in the urban context? How to aggregate and visualize the summarization result for the massive amount of trajectories? How to visually correlate human mobility at the population level with urban semantic factors to address various analytics tasks in modern cities?

To solve the research mentioned above problems, we introduce the concept of steady-state human mobility, which is defined to characterize the long-term stay distribution of individual human trajectory concerning urban factors. To further illustrate steady-state mobilities in cities, we design elaborate visual representations by adapting and integrating state-of-the-art visualization techniques for multivariate and multi-class spatial data. The proposed methods are implemented in a visualization system called UrbanFACET (Fig. 1). The system accomplishes three low-level urban analytics tasks that could answer questions previously impossible, or at least inaccurate: How do the steady-state mobilities of a large population distribute in a modern city (*the overview task*)? Could a city region's socioeconomic and demographic properties be visually inferred from the mobility of residents and visitors in that region (*the profiling task*)? How do steady-state human mobilities vary across city and time (*the comparison task*)? Achieving these tasks brings potential value to (1) urban planners and public security officers to improve their situation and risk awareness about the city and (2) a broad spectrum of business people to determine their enterprise site selection or commercial billboards.

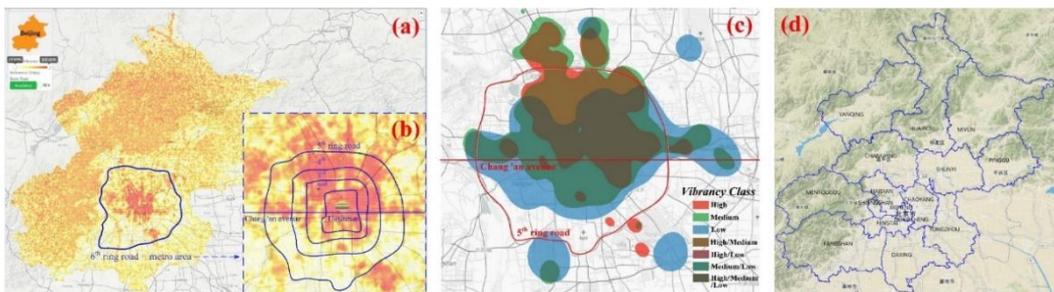


Fig. 1. UrbanFACET visualization of steady-state user mobilities in Beijing:

(a) the full-scale distribution of the vibrancy metric; (b) the metro area inside the 6th ring road is enlarged and annotated in (a); (c) the metro area by the multi-layer vibrancy class visualization, red/green/blue indicate the high/medium/low vibrancy classes; (d) the terrain map of Beijing, overlaid with district boundaries.

Analyzing and visualizing steady-state human mobility over large-scale trajectory data is nontrivial. We make two contributions in this work to tackle these challenges.

- We propose a suite of information-theoretic metrics, namely Fluidity, vibrancy, Commutation, diversity. Together with the population density, they capture the steady-state human mobility of an individual trajectory from multiple FACETs. To correlate spatial locations in the trajectory with the urban context, the classical mobility entropy definition [3] is improved by integrating with

newly introduced semantic distributions. To profile city regions with human mobility, a novel decomposition of the entropy equation is developed to separate the mobility of each stay segment from the whole trajectory.

- We design UrbanFACET, a multifaceted mobility visualization system to illustrate spatial and temporal patterns of steady-state human mobilities in cities. Notably, the Splatterplots visualization for large-scale multi-class point data is adapted in our system from the perspective of data sampling, contouring, and color blending. An elegant flower-shaped glyph design is proposed to represent multivariate mobility metrics in a single view. The visualization approach is evaluated in user experiments in comparison to alternative multi-class data visualization techniques. The system is also implemented with scalable processing and visualization techniques that undertake large-scale human trajectory data. The methods include grid-based data binning, multi-core processing, and kernel density estimation (KDE)-based smoothing for efficient feature extraction, metric computation, and visualization over tens of millions of trajectories.

2. Related Work

2.1 The Visualization of Human Movement

Traditionally, human movements in cities are illustrated on geospatial maps by trajectory visualization techniques. Because of the urban dataset scale, individual trajectories are only captured in their entirety under specific circumstances [4, 5]. For example, Nweke et al. fused the multi-sensor data collection on individuals to detect their personalized human activity in intelligent environments [4]. Most other approaches apply some aggregation methods to reduce visual clutter from visualization of extensive trajectory data. They can be classified into cluster-based, density-based [6], and OD aggregation methods [2].

Existing aggregation-based trajectory visualization techniques, though viable for presenting massive trajectory dataset, focus on short-term human movement during individual intra-city travels. In comparison, our work aims to visualize long-term human mobility indicated by their distribution at stay locations. Additionally, existing approaches often focus on spatial mobility patterns (e.g., OD maps and mobility networks), but our method enables the discovery of semantic patterns concerning specific urban analytic tasks.

Our work is also related to visualization methods for multivariate spatial patterns, as a kind of collaborative visualizations based on mobile device data [7]. These techniques can be classified into glyph-based designs (for point or field data), multiple-layer overlays (for multi-class data), and linked views. In glyph-based approaches, classical designs such as textual labels and widget charts are studied. Several techniques, including Bubble Sets, Line Sets, and Kelfusion [8], have been proposed in overlaying multiple map layers. Notable for geospatial maps, Splatterplots integrated data sampling, contouring, and color blending to alleviate the overdraw problem in displaying various groups of data [9]. Finally, both glyph-based designs and multi-layer visualizations can be coordinated in linked views. The multivariate pattern of a selected set of data can be revealed through standalone multi-dimensional data visualization in these linked views.

2.2 The Analysis of Human Mobility

Historically, human mobility in the urban context has been analyzed at multiple scales. The location records can be aggregated in the city scale to reveal the spatiotemporal distribution of urban activity. The OD flow matrix can model the city-level mobility among regions where the region is defined by municipal districts, cell towers, or major roads.

An OD pair in the trajectory usually abstracts each trip to capture an individual's mobility. For

example, Liu et al. [10] analyzed OD pairs of taxi cabs to study the intraurban trip patterns. When the timestamp of OD trips is considered, temporal OD flows can be aggregated from individual OD trips [2]. If each trip is measured continuously by the floating car technique, individual routes or trajectories are available for the detailed human mobility study [11].

Like our work, Song et al. [3] proposed the mobility entropy metric to characterize human mobility from mobile data traffic. The metric has also been applied and adapted in recent literature to generate trajectory data for simulation [12] and analyze the relationship between human mobility and urban factors [13, 14]. Nevertheless, the Mobility Entropy metric is designed to capture the predictability of human mobility in the trajectory level and is unsuitable for the visualization of spatiotemporal mobility patterns in the underlying record level. Meanwhile, mobility entropy is computed chiefly from the distribution of trajectories measured by associated cell towers but does not explicitly encode semantic information by urban socioeconomic factors. This work proposed a suite of new entropy metrics that represent rich semantics of the underlying trajectory by introducing the concept of semantic distributions. The critical literature offering related methods to our work is listed in Table 1. A more in-depth technical comparison is provided in Section 6.

Table 1. Comparison of three categories of related methods in reference to our proposed method

| Category | Ref. | Method | Pros and cons |
|-----------------------------------|----------------|------------------------------------|---|
| Urban trajectory visualization | [21],[22],[23] | Trajectory clustering based method | Only display travel records & movement patterns |
| | [24] | 3D visualization | Travel records only & no spatial semantics |
| | Ours | UrbanFACET | Show semantic info. & stay records |
| Human mobility metrics | [25] | Density-based | Aggregated info., no individual trajectory |
| | [12],[14] | Mobility entropy | No semantic information of the underlying map |
| | Ours | Semantically-enhanced | Both trajectory info and its spatial semantics |
| Multi-class visual representation | - | Multi-layer map | Can not show all the distribution information |
| | - | Multi-layer blending | Inaccurate in showing the class distribution |
| | [8],[26],[27] | Bubble/line sets, etc. | Suffers from the overdraw problem |
| | Ours | Splatterplots design | Resolves the overdraw issue by down-sample |

3. System Overview

3.1 Data Source

Our urban data is provided by TalkingData, a mobile analytics company that keeps the real-time tracking of billions of Chinese smart devices. Because of the popularity of mobile applications such as Fintech payment [15], smart devices have a very high penetration rate in China. There are four fields in each record: the time of recording, location information (longitude and latitude), the unique ID of the smart device, and the localization method (including GPS [16], Wi-Fi [17], base station, and Internet IP).

In this work, we conduct analysis on four datasets extracted from the company’s data repository. Each dataset corresponds to a 90-day collection of one Chinese city, including Beijing (capital of China), Tianjin (one of five national central cities in China), Tangshan, and Zhangjiakou (two major cities in Hebei province surrounding Beijing and Tianjin). These four cities form the so-called national capital region of China. To obtain more accurate results, we only keep the location records collected by GPS and Wi-Fi, from which sources the spatial localization error is kept below 100 m. For the dataset of Beijing, the trajectories of 31.8 million devices are collected, summing up to 8.4 billion location records.

3.2 Task Characterization

We aim to develop a system to analyze steady-state human mobility in modern cities visually. Steady-state mobility is defined as the mobility of urban users indicated by their distribution of stay locations in the longterm. Compared with the mobility measured by short-term travel segments in human trajectories, the steady-state mobility is at least the same important, if not more, in capturing the actual activity of urban users in the city. We consider three classes of urban tasks that the steady-state analysis of urban mobility could complete. These tasks are beneficial in several vital urban applications such as city planning, public security, and business site configuration.

- **Overview** analysis of the steady-state mobility of city residents in the metropolitan scale and the correlation analysis between the mobility distribution and city landmarks, points-of-interest (POIs), and terrain types. These analyses can help understand city residents' high-level mobility characteristics in a single panorama, such as life abundance and commuting patterns. City planners and security officers will gain situation and risk awareness from the static, dynamic, and categorical mobility distributions in cities by completing these tasks.
- **Profiling** city lands by clustering adjacent regions through their multifaceted mobility metrics. The mobility-based land profiles can be correlated with administrative divisions, city landmarks, and POIs to infer the fine-grained land use of cities. Moreover, the up-to-date land profile can be compared with the city planning blueprint to characterize the trend of land use and potential misuses. Both government and business people can identify risks and opportunities for their critical decisions.
- **Comparison** analysis among cities and across time reveals the city-level distinction and the potential temporal pattern on the overall mobility distribution. The mobility panorama can also be compared with regional city demographics, for example, GDP and house price, to infer the contributing factors to the distribution of the mobility metrics. Governments can optimize their city plannings, including facility and transportation management, based on the comparison with more developed cities. For example, Kim et al. [18] proposed a system to simulate agent control signals in modern cities by comparing them with other cities through newspapers.

3.3 Research Problem

To complete these analytics tasks, we study two research problems:

- Which kind of *summarization methods* to apply for abstracting the steady-state mobility of individual trajectory? In the literature, many related methods have been used. For example, the spatiotemporal aggregation of raw location records [19], the OD abstraction, which models the direction and distance of single trips [10], and the mobility entropy measure that represents the predictability of human trajectory [3, 14]. While existing methods might satisfy part of our analytics tasks, we consider three requirements for the summarization: (1) the method should cover the steady-state distribution of human trajectory rather than their movement; (2) the method could correlate human mobility with semantic information of the city such as socioeconomic and demographic factors; (3) the method should be executed efficiently and accurately to adapt to the size of this work.
- Which kind of *visual metaphors* used to represent the mobility of large-scale trajectories? While the design space of geospatial data visualization has been well studied in the literature, the problem in our scenario possesses two specialties: (1) the size of trajectory data in this work prohibits any point-based spatial visualization methods, appropriate data binding, aggregation, and smoothing should be conducted for effective visualization; (2) the abstracted mobility metrics for visualization are multivariate, which calls for the exploration of multivariate spatial data visualization techniques.

3.4 System Pipeline and Implementation

Fig. 2 summarizes the pipeline in UrbanFACET to compute and visualize information-theoretic mobility. The pipeline takes the raw data described in Section 3.1 as input (Fig. 2(a)). In the first stage, all the location records are partitioned into cities according to the administrative boundary (Fig. 2(b)). Secondly, each city's record set is split into user-level trajectories by their device IDs (Fig. 2(c)). For effective visualization, we extract spatiotemporal features from each location record to enrich the semantic representation of user trajectories (Fig. 2(d)). The suite of information-theoretic mobility metrics is computed over the extracted features of each trajectory (Fig. 2(e)). Finally, city-regions are profiled and visualized using multifaceted mobility metrics (Fig. 2(f)).

In the actual implementation, we apply a multi-core processing technique that computes the mobility metrics of a subset of trajectories in a single CPU core. On a server machine with 12 Intel Xeon E5 CPU cores (1.9GHz with 15MB cache), computing the most costly mobility metric (diversity) over 8 billion records takes about 32 hours, which is feasible as all the computations can be carried out offline.

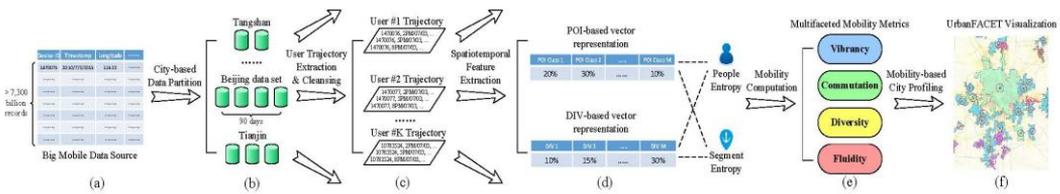


Fig. 2. The analysis pipeline in UrbanFACET.

The trajectory data is first partitioned into city-based data sets. The user-level trajectory and spatiotemporal features are then extracted to compute the multifaceted mobility metrics. Finally, the city regions are profiled by these mobility metrics through multivariate and spatial clustering.

4. Steady-State Mobility Analysis

4.1 Information-Theoretic Mobility Metrics

To visualize steady-state human mobilities, we seek to define specific measures to comprehensively illustrate the steady-state human mobility patterns from their long-term trajectory data. Previous approaches either focus on statistical and predictive models that are not suitable for mobility visualization [1] or only capture the short-term human mobility (e.g., temporal OD flows [2]).

First, we propose a novel mobility measure called **People Entropy**, which is defined by the steady-state distribution of the whole urban trajectory. In this model, human mobility states can be either stay or travel, and these two states are interleaved in each human trajectory. Consider a human trajectory composed of $2N$ segments: $\Gamma = \{\Theta_1, \Lambda_1, \dots, \Theta_N, \Lambda_N\}$, where $\{\Theta_i\}_{i=1, \dots, N}$ denote N stay segments, $\{\Lambda_i\}_{i=1, \dots, N}$ denote N travel segments. For each stay segment Θ_i , the segment length is denoted by T_i . The total length of stay segments is denoted by $T = \sum_{i=1}^N T_i$. We define a *semantic distribution* on each stay segment Θ_i by $Q_i = \langle q_{i1}, \dots, q_{iM} \rangle$. The semantic distribution indicates the property of the stay location in Θ_i concerning certain semantic information having M categories.

Formally, the people entropy of trajectory Γ is defined by:

$$H(P) = - \sum_{j=1}^M p_j \cdot \log p_j \quad \text{where } p_j = \frac{\sum_{i=1}^N q_{ij} \cdot \frac{T_i}{T}}{M} \quad (1)$$

where $P = \langle p_1, \dots, p_M \rangle$ indicates the average semantic distribution of all N stay segments. Note that this

metric is defined on steady-state human mobility where only stay segments are considered. Travel segments refer to the movement between consecutive stay segments and are not counted.

The definition of people entropy can be instantiated only after determining the semantic distribution overstay locations. In this work, we use two distributions related to the characterization of human mobility. First, POI-based resource distribution is used, which indicates the distribution of different types of urban resources available in a stay location (e.g., food and education). Fig. 3(a) lists ten resource types inferred from adjacent POIs on the map. Second, a region distribution is used to indicate the membership of stay location in different city regions (administrative divisions [DIV] in our experiment).

| POI Class | POI Class Type | ID | Name | Time Interval |
|-----------|-------------------------|----|-----------|------------------|
| 1 | Food & Supply | 0 | Morning | 6:00 ~ 9:00 |
| 2 | Entertainment & Leisure | 1 | Forenoon | 9:00 ~ 12:00 |
| 3 | Education | 2 | Noon | 12:00 ~ 14:00 |
| 4 | Transportation | 3 | Afternoon | 14:00 ~ 17:00 |
| 5 | Healthcare & Emergency | 4 | Evening | 17:00 ~ 21:00 |
| 6 | Financial & Bank | 5 | Night | 21:00 ~ 24:00 |
| 7 | Accommodation | 6 | Midnight | 0:00 ~ 6:00 |
| 8 | Office & Commercial | 7 | Weekday | Monday ~ Friday |
| 9 | Natural Landscape | 8 | Weekend | Saturday, Sunday |
| 10 | Factory & Manufacturer | | | |

(a)

(b)

Fig. 3. The list of (a) POI types and (b) the divisions of time intervals used in the feature extraction.

Finally, two mobility metrics are defined using people entropy:

- **Vibrancy** (*people entropy over POI-based resource distribution*) indicates the uncertainty of a user's trajectory accessing different urban resources types. A high vibrancy suggests that the user is able to connect to many types of resources uniformly, while a low vibrancy means that s/he is only able to connect to a few types of resources. We hypothesize that the high vibrancy people might live a more abundant life, while the regions with high vibrancy could have amore robust economy. This hypothesis is roughly validated in Appendix A, where the spatial distribution of vibrancy is shown to be positively correlated with regional GDPs in our dataset (a correlation coefficient between 0.37 and 0.79);
- **Commutation** (*people entropy over DIV-based region distribution*) indicates the uncertainty of a user's trajectory switching among different administrative divisions. As most stay segments of a trajectory locate in the home or working place, a high commutation people suggests s/he needs to commute more frequently in daily jobs. A high commutation region can be interpreted as a home or working place with more high commutation people. Again, we present a validation of the commutation metric in Appendix A, where the distance of displacements in a trajectory is shown to be positively correlated with its metric commutation value using our dataset (a correlation coefficient of 0.34).

To profile city regions with people entropy, we map the entropy value into each stay segment, directly linked to the spatial location. We propose a mapping strategy that defines **Segment Entropy**. The basic idea is to decompose the equation of people entropy on a trajectory and reassign nonuniform entropies to all underlying stay segments while maintaining the weighted average of segment entropies unchanged (still the people entropy $H(P)$).

$$\begin{aligned}
H(P) &= -\frac{1}{T} \cdot \sum_{j=1}^M T \cdot p_j \cdot \log p_j \\
&= -\frac{1}{T} \cdot \sum_{j=1}^M \sum_{i=1}^N q_{ij} \cdot T_i \cdot \log p_j \\
&= -\frac{1}{T} \cdot \sum_{i=1}^N \sum_{j=1}^M q_{ij} \cdot \log p_j \cdot T_i \\
&= \sum_{i=1}^N H(Q_i, P) \cdot \frac{T_i}{T}
\end{aligned}$$

where the segment entropy of the i th stay segment is defined by

$$H(Q_i, P) = -\sum_{j=1}^M q_{ij} \cdot \log p_j \quad (2)$$

Mathematically, the segment entropy is the cross-entropy between distributions Q_i and P by $H(Q_i, P) = D_{KL}(Q_i \| P) + H(Q_i)$. Because Q_i is only related to the i th stay location, the segment entropy mainly describes the relative entropy $D_{KL}(Q_i \| P)$ between distributions Q_i and P , i.e., their KL divergence.

Similarly, the segment entropy is instantiated into two metrics using POI-based resource distribution and DIV-based region distribution.

- **Diversity** (*segment entropy over POI-based resource distribution*) indicates the difference between a stay location's resource distribution and the overall resource distribution of the trajectory. A high-diversity region can be interpreted as having the type of resource that is rarely connected to its visitors. The region can be diversified compared with home and working places and is visited by groups of very different people, e.g., the recreation regions and sports centers.
- **Fluidity** (*segment entropy over DIV-based region distribution*) indicates the difference between a stay location's administrative division and the overall DIV distribution of the trajectory. A high fluidity suggests that people do not live or work in this region. A high fluidity region can be interpreted as the land mixing people from quite different home/working places. The typical high fluidity regions include tourist attractions, commercial centers, and transportation stations.

4.2 Scalable Mobility Computation

The information-theoretic metrics defined above capture the real-world mobility behavior of urban users, independent of the actual measurement of their trajectory. In this part, we describe the method to compute these metrics from the urban trajectory data.

Consider an observed user trajectory with n records at stay locations, which are denoted by $= \{l_1, \dots, l_n\}$. Let the semantic distribution on the i th record be $\tilde{Q}_i = \langle \tilde{q}_{i1}, \dots, \tilde{q}_{iM} \rangle$. The people/segment entropy of y_i is computed by

$$\begin{aligned}
\tilde{H}(\tilde{P}) &= -\sum_{j=1}^M \tilde{p}_j \cdot \log \tilde{p}_j \quad \text{where } \tilde{p}_j = \frac{\sum_{i=1}^n \tilde{q}_{ij}}{\sum_{j=1}^M \sum_{i=1}^n \tilde{q}_{ij}} \quad (3) \\
\tilde{H}(\tilde{Q}_i, \tilde{P}) &= -\sum_{j=1}^M \tilde{q}_{ij} \cdot \log \tilde{p}_j \quad (4)
\end{aligned}$$

where $\tilde{P} = \langle \tilde{p}_1, \dots, \tilde{p}_M \rangle$ indicates the estimated semantic distribution from all n location records.

The semantic distribution on the i th location record (l_i) is computed by extracting M features from the location. Take the POI-based distribution as an example. All POIs on the map are categorized into

tentypes ($M=10$) according to OpenStreetMap POI categories, as listed in Fig. 3(a). The j th-type POI feature extracted from the i th record denoted as \tilde{q}_{ij} , is computed by the probability of the i th record (l_i) belonging to the j th POI type:

$$\tilde{q}_{ij} = \frac{\sum_k \varphi_{0, \sigma_{j,k}^2}(\text{disk}(l_i, \text{POI}_{j,k}))}{\sum_j \sum_k \varphi_{0, \sigma_{j,k}^2}(\text{disk}(l_i, \text{POI}_{j,k}))} \quad (5)$$

where $\varphi_{0, \sigma_{j,k}^2}$ indicates the standard Gaussian probability density function with zero mean and variance of $\sigma_{j,k}^2$. $\text{disk}(l_i, \text{POI}_{j,k})$ indicates the spatial distance between the record l_i and the k th POI in the j th type ($\text{POI}_{j,k}$). This probability is the sum of influences from all j th-type POIs to l_i , before normalized across all POI types. By default, $\sigma_{j,k}$, the influence variance of $\text{POI}_{j,k}$ is set to 1.5 times of its enclosure radius if the POI is area-based to adapt to irregular POI shapes; or set to 100 m if the POI is point-based, according to the average radius of area-based POIs. All these parameters are determined by the domain knowledge and experience of the experts in the mobile analytics company.

In computing these mobility metrics, the main challenge is to process the enormous amount of location records. In UrbanFACET, we apply a grid-based binning technique to resolve the scalability issue. The entire map is partitioned into square cells with a radius of r . The features of each location record are approximated by the features in the center of the cell to which the record belongs. In this way, we only need to extract features for all the cell centers. For the city of Beijing, there are about 0.2 million cells ($r = 100$ m), four magnitudes smaller than the number of records. The computation overhead at each location record is reduced to two multiplications on the longitude/latitude to determine its associated cell. Meanwhile, the feature information of cell centers can be pre-computed offline so that multiple grid settings can be switched online.

4.3 Mobility-based City Profiling

An essential task of UrbanFACET is to profile city regions and their functionalities through the mobility metrics of urban users. We propose three visual profiling methods over the grid-based binning technique. By default, the mobility metric of all location records in a grid cell is averaged to represent the overall mobility of that cell, which is called the *aggregation* method. Note that, for records belonging to the same stay segment of a trajectory, these records are only counted once in the aggregation. A standard heatmap is then applied for visualization (Section 5.1). The aggregation method can effectively display the spatial distribution of the individual mobility metric at the cell level, but it can not drill down to the distribution within the cell.

The second visual profiling method applies the sequential clustering on both the target mobility metric and the recording density in space. In the first step, we compute the distribution of mobility metric values in all records. The distribution is then divided into several range-based classes. For example, in a typical three-class division, the top 20% metric values are categorized into the upper class, the medium 60% into the middle class, and the bottom 20% into the lower class. For each range-based class, a record density distribution on cells is aggregated in which each density measure in a specific cell indicates the number of records belonging to that class in the cell. The sequential clustering result is finally displayed by the proposed multi-layer map visualization (Section 5.2).

The third method applies a three-step *multivariate* data clustering on more than one mobility metric. In the first step, the location records in the same cell are averaged on their multifaceted mobility metrics, same with the aggregation method. A metric vector is obtained for each cell by concatenating the value of these metrics. In the second step, all the cells in the city are clustered based on the pairwise distance between their metric vectors. In the implementation, the Euclidean distance and the k-means clustering algorithm are applied. In the third step, for each k-mean cluster, the spatial cells belonging to

the cluster are processed by DBSCAN to extract spatially connected high-density regions. A flower-shaped glyph displays each cluster's vector of mobility metrics over all the high-density regions belonging to the cluster (Section 5.2).

5. Visualization Design

The UrbanFACET visualization interface (Fig. 4) is designed to fulfill steady-state mobility analysis tasks. First, the overview task, users start by picking the target city, period, and mobility metric in the top-left control panel (Fig. 4(a) and (b)). The distribution of this metric is then displayed in the main panel (Fig. 4(f)) by the aggregation method. The exact value and the probability distribution of the selected metric are shown in the bottom-left detail panel for fine-grained analysis (Fig. 4(c)–(e)). Second, the profiling task, multiple mobility metrics, and the density metric can be overlaid on the main panel by a flower-shaped glyph design and the multi-layer hotspots. Third, the main panel can be split into juxtaposed sub-views to compare cities, periods, or different metrics. In the following, we describe the detailed visualization design in the context of geospatial maps (Section 5.1), mobility-based city profiling (Section 5.2), and interactions for spatiotemporal comparison (Section 5.3).

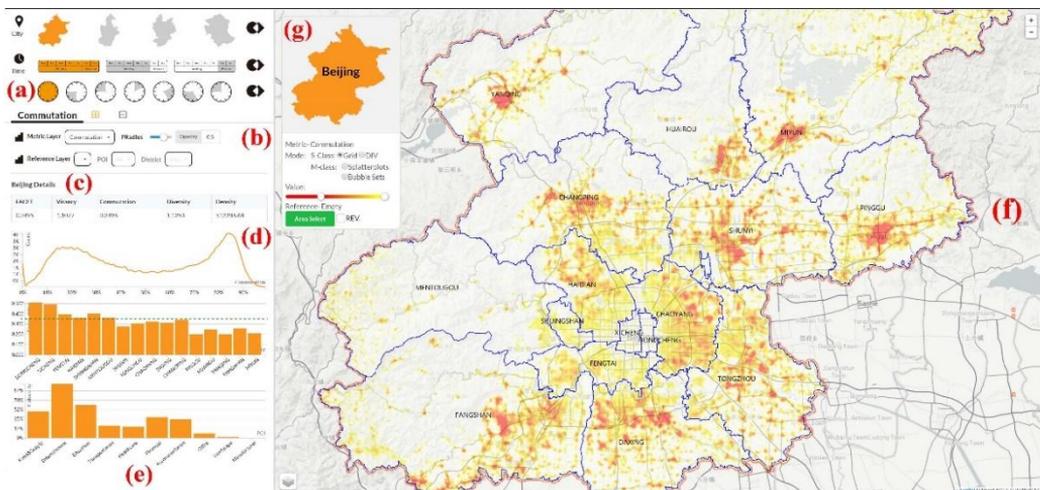


Fig. 4. The visualization interface of UrbanFACET: (a) the selection of city and time; (b) the selection of mobility metrics and the visualization options; (c) detailed information on the selected region; (d) probability density function of the selected mobility metric; (e) average probability distribution on POI/DIV classes to compute the mobility metric; (f) the spatial visualization of mobility metrics (commutation, in this case, the color mapping is inverted); and (g) legend and controller of the map.

5.1 Mapping User's Mobility Metrics

The map-based visualization in the main panel has two default layers: the base layer and the metric layer. The base layer in the background gives the geospatial information about the city, serving as location references. Several base layer types are available for separate analysis scenarios, for example, road network, terrain, and satellite imaging maps. On the other hand, the metric layer is overlaid on top of the base layer and displays the critical distributional information across the city. A number of metrics can be selected to display (Fig. 4(b)), including the four mobility metrics proposed in Section 4.1, the classical record density metric, and three regional city demographics (GDP, population, house price).

On the metric layer, given the amount of data, it is not feasible to render points of the trajectories individually. In UrbanFACET, we propose a KDE method over the grid-based aggregation technique in

Section 4.3, namely the grid KDE. Take the vibrancy metric as an example, denote the i th grid cell out of all the C cells as G_i , the center of the cell as x_i and the average vibrancy on this cell as c_i . The density function at x by grid KDE is computed as

$$\hat{f}_h(x) = \frac{1}{h} \sum_{i=1}^C c_i \cdot \varphi_{0,1}\left(\frac{x - x_i}{h}\right) \quad (6)$$

where $\varphi_{0,1}$ is the standard normal kernel function, h is the bandwidth parameter controlling the smoothing level. The initial bandwidth is set to twicethe grid cell radius according to [20], and it will change adaptively together with the zoom factor of the map. The grid KDE method fills the gap between discrete points into a spatially continuous density map, providing more pleasing and complete visual results. To avoid an unbalanced visual effect due to skewed metric value distributions, we propose a ranking-based color mapping approach, which maps the percentile of a value (instead of the value itself) to the desired color.

5.2 Profiling Cities with Multifaceted Mobilities

Beyond the display of each individual mobility metric, we further design and implement visualizations to illustrate the result of profiling methods in Section 4.3, which summarize multiple mobility classes or metrics on the same map.

Visualizing multiple mobility classes: The sequential clustering method in Section 4.3 generates three (i.e., high/medium/low) classes on each mobility metric. For each mobility class, a density distribution of the class on the map can be computed. To visualize these distributions, we adopt a multi-layer design. The density distribution of each class is drawn by the grid KDE method in a separate layer, using different color hues. For example, in Fig. 5(c), the high/medium/low vibrancy classes are visualized in red, green, and blue colors, respectively. The key design problem is how to synthesize multiple distribution layers in the same map and still perceive each class's distribution.

This work proposes adapting the Splatterplots design [9] to visualizing multiple metric class distributions. Unlike the original Splatterplots, our adapted version first filters the points based on a threshold and applies point KDE to estimate data densities. Applying thresholding first eliminates the effect that low-density grids may become high-density due to the smoothing. In addition, our Splatterplot uses semi-transparent colors for different metric classes displayed. The overlapping region is drawn by a revised blending method that weighs down both color saturation/brightness and its alpha value. The region with more overlapping classes is shown in a more opaque color. Furthermore, we also provide a contour-based Bubble Sets visualization as an alternative. With Bubble Sets, the distribution of a single class can be better perceived.

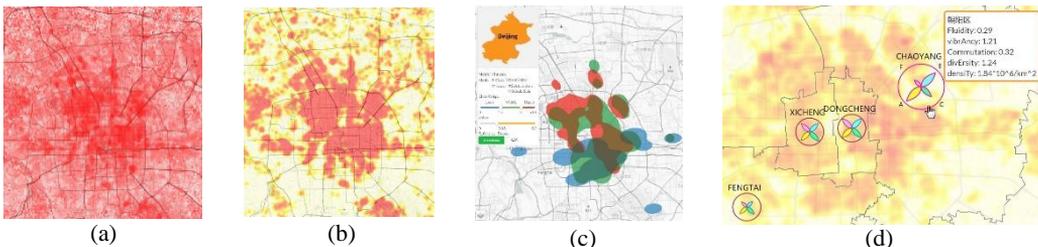


Fig. 5. Metric layer visualizations: (a) grid-based binning with a single-color palette, (b) grid KDE with a two-color palette, (c) multiple metric class visualization using Splatterplots, and (d) flower-shaped glyphs for multifaceted mobility metrics.

Visualizing multiple mobility metrics: The multivariate clustering method in Section 4.3 detects

dense clusters on multiple mobility metrics. These clusters are treated in the same manner with the multiple metric classes as above. Splatterplots are applied to visualize the distribution of these multi-metric clusters and their potential overlaps. In addition, we introduce a flower-shaped glyph design to illustrate the multiple metrics on each cluster, as shown in Fig. 5(d). The glyph is composed of four colored petals and one ring surrounding the petals, which visualizes the 4-tuple of mobility metrics plus density: $\langle \textit{Fluidity}, \textit{vibrancy}, \textit{Commutation}, \textit{diversity}, \textit{density} \rangle$. The area size of the ring is designed to be perceptually proportional to the normalized density metric in the current view. Each petal within the ring represents one mobility metric, with the area size of the petal proportional to the normalized metric value. The largest mobility metric in the view will have its petal stretched to the inner rim of the ring, as shown by the four petals of Dongcheng district in Fig. 5(d). Upon mouse hovering a glyph, all five metrics' detailed values is shown in a pop-up label. The hovered glyph also grows to the same size as the largest glyph in the current view to observe the multifaceted mobility patterns for low-density areas. In some cases, all glyphs are set to the same size only to compare the mobility metrics (Fig.6(a)).

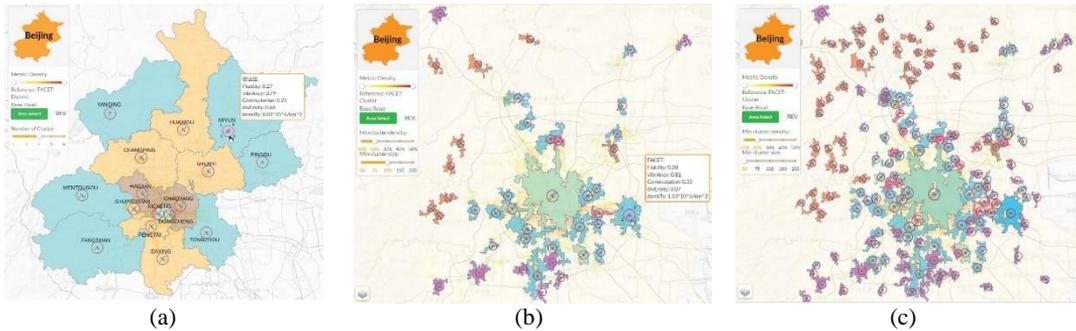


Fig. 6. Profiling the city of Beijing by all four mobility metrics:

(a) the clusters on districts, (b) the clusters on grid cells (medium cluster area), and (c) more clusters on grid cells by setting a smaller cluster area threshold.

5.3 Comparing among Metrics, City and Time

On the comparison task, UrbanFACET relies on a side-by-side interface design that juxtaposes multiple spatial maps as sub-views, which is achieved by the split operation using the “+” button on top of the metric layer configuration in Fig. 4(b). Each time, the current base/metric/reference layer setting is replicated in the newly created sub-view. The new map can be configured in the same manner to a different metric visualization for comparison. By this design, multiple mobility metrics can be compared, the mobility metric and the regional demographics (GDP, population, house price) can also be correlated.

On comparing cities or periods, we provide shortcut buttons as shown in the rightmost column of Fig. 4(a). The system supports the comparison of up to four cities (Beijing, Tianjin, Tangshan, Zhangjiakou) and six periods based on the division in Fig. 3(b), excluding midnight. Switching among cities and filtering by periods are also supported by the iconized selector in the middle column of Fig. 4(a).

6. Evaluation with State-of-the-Art Methods

The UrbanFACET system on spatial-temporal data addresses several critical issues of urban data analysis, which also involves a rich set of state-of-the-art visualization and analysis methods. First, there are quite some visualization methods proposed for spatiotemporal urban data, in particular, trajectories. Second, a couple of human mobility metrics similar to ours have been presented. Third, the visual representation of multi-class spatial data is an important problem in multivariate data

visualization. Many designs have been published for various purposes. These state-of-the-arts are briefed in Section 2, and more than ten representative methods are listed in Table 1. Below we detail the advantage of the proposed UrbanFACET techniques in comparison to state-of-the-art, using experiments and in-depth discussions.

6.1 Visualization Methods for Urban Trajectories

The movement of people in urban environments is generally visualized as trajectories on the geospatial map. For a vast number of trajectories, the standard visualizations usually are pretty cluttered. The mainstream techniques resolve the visual clutter by visualizing the clustering result of trajectories [21–23]. For example, in UrbanMotion [21], the city map is partitioned into fixed cells (500m each). Individual movement records are clustered locally within each cell and then connected globally into movement flows for visualization (Fig.7(a)). Using a larger spatial cell partition, the FlowMap approach [22] applied Voronoi tessellation to form a higher-level mobility network (Fig.7(b)). The MapTrix method by Yang et al. [23] introduced the matrix visualization to display the trajectory clusters in the highest city/country region level. The traditional clustering-based trajectory visualization methods focus on the travel part data of urban user's location records but miss the opportunity to display the distribution of stay records in an urban dataset. Another kind of trajectory visualization technique utilizes the 3D space to display spatiotemporal movements, e.g., the display wall metaphor by Tominski et al., which introduces a display wall design to stack trajectory bands on the z-axis [24]. Yet, this approach does not consider the stay records in urban data and the underlying semantic information in the spatial map. In comparison, the proposed urban data visualization method in this work (Fig.7(c)) explores the semantics of underlying spatial maps, e.g., the POI and district class. Our method focuses on displaying the steady-state distribution of stay records in urban data, which the previous methods mostly overlooked.

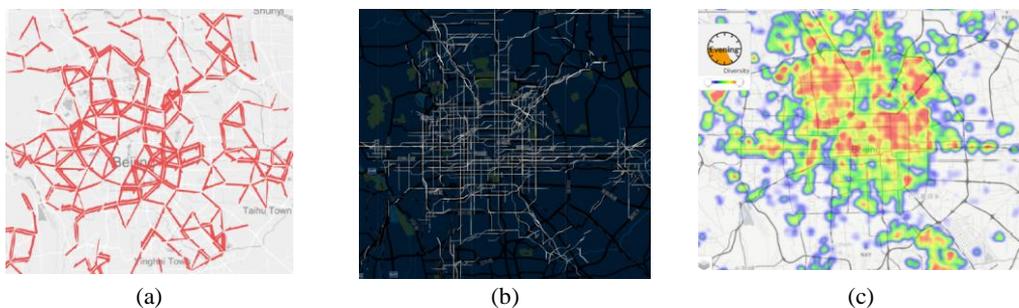


Fig. 7. Comparison of urban data visualization methods:

- (a) UrbanMotion clustering at small cell level, (b) FlowMap by Voronoi tessellation, and (c) our method focusing on stay records with semantic info.

6.2 Human Mobility Metrics

The default method in the literature (e.g., [25]) aggregates the big location dataset into spatiotemporal density distributions on the visualization of human location records. A variety of heat map displays are applied to visualize these density metrics. These approaches are the most straightforward design and can reveal the overall distribution of longitudinal human locations. However, the aggregation-based approach misses the opportunity to correlate mobility information of individual users and their versatile trajectory data. To fill this gap, Song et al. [3] proposed the mobility entropy metric, which bears the same motivation to embed individual trajectory information to describe human mobility data. The

literatures [12, 14] further applied the metric to generate human trajectory data or indicate unique mobile users. Yet, the mobility entropy metric is defined on the distribution of cell tower grids with little semantic information regarding urban analytics tasks. The metric is originally designed to capture the predictability of trajectory, but not its semantics. In comparison, our work defines rich-semantic entropy measures through the introduction of semantic distributions. We also showcase its usage by four metrics as an example (vibrancy, commutation, diversity, fluidity).

First, we study the distribution of four entropy-based mobility metrics over Beijing. The region defines the metropolitan area of Beijing (metro in short) within the 6th ring road (the blue enclosure in Fig. 1(a), enlarged in Fig. 1(b)). The metro is centered at Tiananmen and divided into the northern and southern city by the Chang'an avenue and into five ring-shaped regions by the 2nd to 6th ring roads (Fig. 1(b)). The population density of Beijing spreads radially from the center to the remote city area.

6.3 Visual Representation of Multi-class Spatial Data

Our mobility analysis techniques summarize multiple classes of single/multi-metric from the human trajectory in this work. The classical visual representation of multi-class spatial data applies overlapped Multi-Layer map (Fig.8(a)). The multi-layer map design allows layer switching through mouse clicks but can not show all the distribution information due to overlapping layers. Another alternative by the multi-layer alpha-blending (Fig.8(b)) is not entirely accurate in showing the range of each class's distribution. Advanced techniques such as Bubble Sets [26] (Fig.8(c)), Line Sets [27], and Kelpfusion [8] suffer from the overdraw problem in displaying multiple classes of data. In comparison, the proposed technique adapted from the Splatterplots design (Fig.8(d)) integrates data sampling, contouring, and color blending and resolves the overdraw issue by elegant down-sampling.

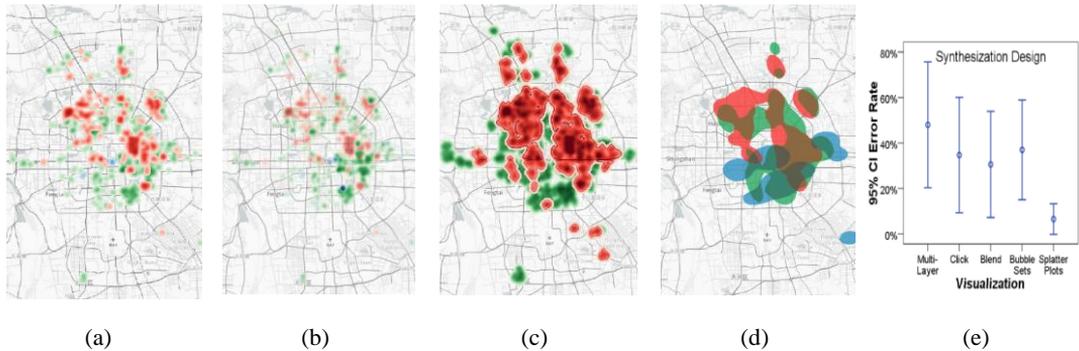


Fig. 8. Comparison of multi-class spatial data visualization:

(a) multi-layer map design, (b) alpha-blending, (c) hotspot bubble sets, (d) the adapted Splatterplot, and (e) the result of user study comparison.

We also conducted a controlled user experiment to evaluate the visualization performance of alternative multi-class spatial data visualization techniques in distinguishing multiple density distributions. Eighteen graduate students were recruited as subjects. Each subject was asked to complete the below evaluation task with all five methods in comparison: multi-Layer map, multi-Layer map allowing user click, multi-layer alpha-blending, Bubble Sets, and the adapted Splatterplots.

Task: *In a map with two classes of distributions, namely class A and B, estimate the ratio of A to B in the area size of the distribution.*

We depict the subject's task error rate using five methods in Fig. 8(e). It can be shown that the Splatterplots design achieves the lowest error rates on average. The Mann-Whitney tests detect statistically significant differences between the error rate of the Splatterplots design with two

alternatives: the baseline multi-layer design ($U = 85.5$, $p = 0.01$) and the bubble sets ($U = 106.5$, $p = 0.045$). The experiment result validates that the application of the adapted Splatplot design helps to perceive multi-class density distributions on the same map view in our entropy-centric scenario. All the other alternatives have severe limitations.

7. Case Study

7.1 Overview of City Mobility

We evaluate UrbanFACET using the trajectory data in four cities: Beijing, Tianjin, Tang-shan, and Zhangjiakou. First, we study the distribution of four entropy-based mobility metrics over Beijing. The region defines the metropolitan area of Beijing (metro in short) within the 6th ring road (the blue enclosure in Fig. 1(a), enlarged in Fig. 1(b)). The metro is centered at Tiananmen and divided into northern and southern cities by the Chang'an avenue and into five ring-shaped regions by the 2nd to 6th ring roads (Fig. 1(b)).

Vibrancy: As discussed in Section 4.1, a high vibrancy region means the people or visitors there on average access many different types of resources, thus potentially live a more prosperous life. The overall vibrancy distribution in Beijing is shown in Fig. 1(a). Two high vibrancy regions can be found: (i) the central area in the metro of Beijing; (ii) the western, northern, and northeastern regions outside the metro.

We drill down to these two areas to obtain more findings on the vibrancy distribution. First, in the metro area, we observe an asymmetric pattern quite different from the population density distribution, as shown by the enlarged vibrancy distribution in Fig. 1(b). The core of the high vibrancy region is rectangle-shaped, from the 5.5th ring in the north to the 3rd ring in the south and the 4th ring in the east to the 3.5th ring in the west. In this sense, it can be inferred that the northern people live a more abundant life than the southern people. This finding agrees well with the multiple vibrancy class visualization in Fig. 1(c). The hotspots of high vibrancy people (red area and other overlapping areas such as brown and grey) are primarily located in the northern city. Second, by comparing with the terrain map in Fig. 1(d), we discover that the three high vibrancy regions outside the metro correspond to three major mountains surrounding Beijing in the west, north, and northeast. We sample 200 users who visit these mountain areas and compute their overall record distribution to reason about this result. It can be found that the sampled distribution has almost 70% records in the metro area of Beijing. The location records in the mountains are made mainly by the high-vibrancy travelers who live in the Beijing metro area.

Commutation: The distribution of the commutation metric in Beijing is shown in Fig.9(a). The mountain areas outside the metro have a significant commutation need, which is understandable. Meanwhile, unlike the vibrancy distribution, the commutation in the metro is relatively low. As shown in Fig.9(b), the metro area can be divided into two sub-regions if we take a closer look. The sub-region within the 2.5th ring, mainly the old city districts of Dongcheng and Xicheng, has a higher commutation. We hypothesize that people primarily work in this sub-region, that is the center of Beijing, and do not live there because of the over high house price and inconvenient life in the old city. The second sub-region is between the 2.5th ring and 5.5th ring, where the commutation is among the lowest in Beijing. This phenomenon is that people here can afford the house price near their workplace and find these modern city districts attractive for life. Outliers in this sub-region include the Xiangshan park, a famous recreation site, and Yizhuang, the Beijing Economic Development Area, the workplace for many people.

With an inversed color mapping of the commutation distribution (Fig. 4(f)), more valleys outside the

metro of Beijing are found to be the center of suburb districts (Yanqing, Miyun, Pinggu). The residents there can meet their everyday demands in the district center and do not need to commute to the metro.

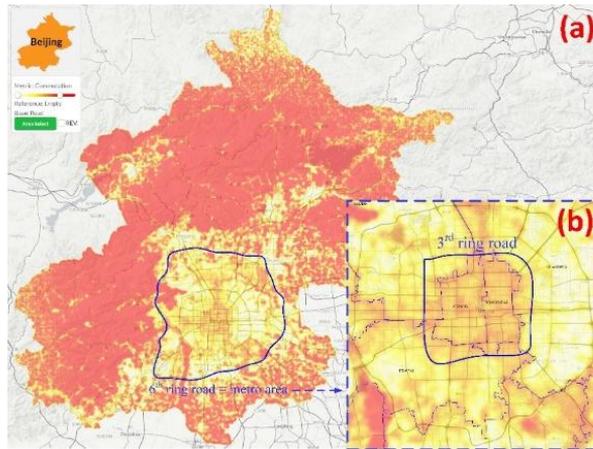


Fig. 9. The commutation metric distribution in Beijing: (a) the full-scale visualization of the city and (b) the metro area is enlarged.

Fluidity: In the fluidity distribution of Beijing (Fig.10(a)), there is a three-layer pattern similar to the commutation distribution. The fluidity is high in the mountain regions outside the metro, except the valley in suburb centers, which validates our hypothesis that the records on the mountain are made by occasional travelers whose bases are in different districts of the metro area. Inside the metro, the region outside the 2.5th ring road mostly has a low fluidity. Some exceptions happen in the airport, Xiangshan Park, and the National Route 107 (the busiest road connecting northern and southern China), where people from different districts come for recreation or transportation. Next, we enlarge the map and switch to multi-layer hotspots to visualize the dense regions in the high-fluidity class. As shown in Fig.10(b), the fluidity hotspots are found to be co-located with several tourist attractions (the Beijing Zoo), the commercial centers (Xidan/Dongdan Shopping regions), etc.

The overview of mobility metrics in Beijing provides several recommendations: for city governors and planners, more investments can be made in the southern area to foster a balanced development of the whole city; for enterprise managers, new sites and offices can be opened between the 2.5th and 5th ring roads so that employees can benefit from fewer commutations; for security officers, more attention should be paid to famous tourist attractions where the risk is higher with visitors from different places.

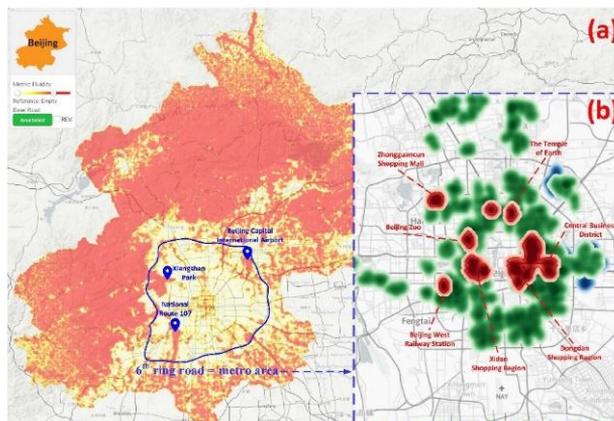


Fig. 10. The fluidity distribution in Beijing:

(a) the full-scale visualization of the city and (b) the metro area is enlarged.

7.2 Profiling City Regions

At the district level, the four mobility metrics can be studied altogether to profile the city's functionality. As shown in Fig.6(a), all districts of Beijing are clustered into four classes. The flower-shaped glyph depicts the four metrics of each district on the map. Dongcheng and Xicheng, the two districts in the inner/old city, have the most significant mobility in all four metrics. They are diversified and fluid. Their residents are vibrant and commuting. Moving outside, Haidian and Chaoyang are also core districts in the Beijing metro. Their vibrancy and diversity are still high, but the commutation and fluidity are low. People there do not need to commute much and tend to be residents.

The city can be profiled in the finest granularity based on regular grids used in the binning technique. By the region clustering in Section 4.3, spatial clusters of grids are detected and visualized. As shown in Fig.6(b), there are six major output clusters indicated by their filled colors. The green cluster encompasses the core area of the Beijing metro plus well-known wealthy regions in the surrounding, for example, Shangdi (China's Silicon Valley), Yizhuang. The blue cluster mainly locates outside the green cluster, featuring a medium vibrancy and low diversity. The pink cluster is similar to the blue cluster but is smaller in size and higher in vibrancy, corresponding to upper-class accommodations for rich people.

Meanwhile, the purple cluster co-locates with the center of suburb districts, having low values in all four metrics. The orange cluster appears mainly in the mountain regions, corresponding to several tourist attractions such as the Great Wall, the Tanzhe Temple, and the Phoenix Mountain. They have a high vibrancy, fluidity, and commutation value, indicating a composite of tourists from different metro districts. The last brown cluster features a high fluidity and vibrancy but medium commutation interpreted as a mixture of business people that commutes less than the tourists. These clusters can be extended by setting a smaller cluster size threshold. As shown in Fig.6(c), more clusters are detected, notably the increased orange tourism sites in the mountain area.

8. Conclusion

We present UrbanFACET, a system that visually analyzes the steady-state human mobility in modern cities from large-scale urban trajectory data. The system efficiently visualizes, profiles, and compares human mobility from their long-term stay distributions. Notably, the mobility analysis is explicitly correlated with semantic information in cities, such as socioeconomic and demographic factors. In solving key challenges abstracting and visualizing steady-state human mobility, we have proposed: (1) a suite of information-theoretic metrics to characterize both long-term stay distribution of individual trajectories and their urban semantics; (2) an integrated visualization interface that adapts state-of-the-art multivariate glyph designs and multi-class layer syncretization methods; and (3) a scalable system implementation that utilizes data binning, parallel processing, and KDE-based smoothing to deal with the enormous amount of trajectory data.

In the future, we would like to extend the entropy-based mobility metrics in three directions and answer the following questions. First, the semantic distribution could be defined in hierarchies, e.g., the DIVs of a country/city. How to represent this hierarchical information in mobility metrics? Second, the current entropy metric is defined over probability distributions. A more comprehensive representation is the mobility network of urban trajectory. How to summarize the mobility network using multi-faceted information-theoretic measures? Third, we now treat the

Acknowledgements

Appendix A can be found via https://visdata.github.io/leishi/paper/UrbanFACET_appendix.pdf.
Video demo can be found via <https://visdata.github.io/leishi/video/UrbanFACET.mp4>.

Author's Contributions

Conceptualization, investigation and methodology, LS. Writing of the original draft, LS, YC, YZ. Writing of the review and editing, LS. Data Curation, XZ, MD. Software and Visualization, LS, ZG, TJ, RF. All the authors have proofread the final version.

Funding

We are supported by National Natural Science Foundation of China (No. 61772504 and 62072115) and Fundamental Funds for the Central Universities.

Competing Interests

The authors declare that they have no competing interests.

References

- [1] M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779-782, 2008.
- [2] G. Andrienko, N. Andrienko, G. Fuchs, and J. Wood, "Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 9, pp. 2120-2136, 2016.
- [3] C. Song, Z. Qu, N. Blumm, and A. L. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018-1021, 2010.
- [4] H. F. Nweke, Y. W. Teh, G. Mujtaba, U. R. Alo, and M. A. Al-garadi, "Multi-sensor fusion based on multiple classifier systems for human activity identification," *Human-centric Computing and Information Sciences*, vol. 9, article no. 34, 2019. <https://doi.org/10.1186/s13673-019-0194-5>
- [5] S. A. Khowaja, B. N. Yahya, and S. L. Lee, "CAPHAR: context-aware personalized human activity recognition using associative learning in smart environments," *Human-centric Computing and Information Sciences*, vol. 10, article no. 35, 2020. <https://doi.org/10.1186/s13673-020-00240-y>
- [6] R. Scheepens, N. Willems, H. van de Wetering, and J. J. Van Wijk, "Interactive visualization of multivariate trajectory data with density maps," in *Proceedings of 2011 IEEE Pacific Visualization Symposium*, Hong Kong, China, 2011, pp. 147-154.
- [7] L. Chen, H. N. Liang, F. Lu, K. Papangelis, K. L. Man, and Y. Yue, "Collaborative behavior, performance and engagement with visual analytics tasks using mobile devices," *Human-centric Computing and Information Sciences*, vol. 10, article no. 47, 2020. <https://doi.org/10.1186/s13673-020-00253-7>
- [8] W. Meulemans, N. H. Riche, B. Speckmann, B. Alper, and T. Dwyer, "Kelfusion: a hybrid set visualization technique," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 11, pp. 1846-1858, 2013.
- [9] A. Mayorga and M. Gleicher, "Splatterplots: overcoming overdraw in scatter plots," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 9, pp. 1526-1538, 2013.
- [10] Y. Liu, C. Kang, S. Gao, Y. Xiao, and Y. Tian, "Understanding intra-urban trip patterns from taxi trajectory data," *Journal of Geographical Systems*, vol. 14, no. 4, pp. 463-483, 2012.
- [11] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-drive: enhancing driving directions with taxi drivers'

- intelligence,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 220-232, 2013.
- [12] L. Pappalardo and F. Simini, “Data-driven generation of spatio-temporal routines in human mobility,” *Data Mining and Knowledge Discovery*, vol. 32, no. 3, pp. 787-829, 2018.
- [13] M. Vanhoof, W. Schoors, A. Van Rompaey, T. Ploetz, and Z. Smoreda, “Comparing regional patterns of individual movement using corrected mobility entropy,” *Journal of Urban Technology*, vol. 25, no. 2, pp. 27-61, 2018.
- [14] C. Cottineau and M. Vanhoof, “Mobile phone indicators and their relation to the socioeconomic organisation of cities,” *ISPRS International Journal of Geo-Information*, vol. 8, no. 1, article no. 19, 2019. <https://doi.org/10.3390/ijgi8010019>
- [15] J. Kang, “Mobile payment in Fintech environment: trends, security challenges, and services,” *Human-centric Computing and Information sciences*, vol. 8, article no. 32, 2018. <https://doi.org/10.1186/s13673-018-0155-4>
- [16] M. W. Akhtar, S. A. Hassan, R. Ghaffar, H. Jung, S. Garg, and M. S. Hossain, “The shift to 6G communications: vision and requirements,” *Human-centric Computing and Information Sciences*, vol. 10, article no. 53, 2020. <https://doi.org/10.1186/s13673-020-00258-2>
- [17] T. T. Khanh, V. Nguyen, X. Q. Pham, and E. N. Huh, “Wi-Fi indoor positioning and navigation: a cloudlet-based cloud computing approach” *Human-centric Computing and Information Sciences*, vol. 10, article no. 32, 2020. <https://doi.org/10.1186/s13673-020-00236-8>
- [18] E. Kim, S. Jang, S. Li, and Y. Sung, “Newspaper article-based agent control in smart city simulations,” *Human-centric Computing and Information Sciences*, vol. 10, article no. 44, 2020. <https://doi.org/10.1186/s13673-020-00252-8>
- [19] G. Andrienko, N. Andrienko, H. Bosch, T. Ertl, G. Fuchs, P. Jankowski, and D. Thom, “Thematic patterns in georeferenced tweets through space-time visual analytics,” *Computing in Science & Engineering*, vol. 15, no. 3, pp. 72-82, 2013.
- [20] O. D. Lampe and H. Hauser, “Interactive visualization of streaming data with kernel density estimation,” in *Proceedings of 2011 IEEE Pacific Visualization Symposium*, Hong Kong, China, 2011, pp. 171-178.
- [21] L. Shi, C. Huang, M. Liu, J. Yan, T. Jiang, Z. Tan, Y. Hu, W. Chen, and X. Zhang, “UrbanMotion: visual analysis of metropolitan-scale sparse trajectories,” *IEEE Transactions on Visualization and Computer Graphics*, 2020. <https://doi.org/10.1109/TVCG.2020.2992200>
- [22] N. Adrienko and G. Adrienko, “Spatial generalization and aggregation of massive movement data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 2, pp. 205-219, 2010.
- [23] Y. Yang, T. Dwyer, S. Goodwin, and K. Marriott, “Many-to-many geographically-embedded flow visualisation: an evaluation,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 411-420, 2017.
- [24] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, “Stacking-based visualization of trajectory attribute data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2565-2574, 2012.
- [25] G. Andrienko, N. Andrienko, H. Bosch, T. Ertl, G. Fuchs, P. Jankowski, and D. Thom, “Thematic patterns in georeferenced tweets through space-time visual analytics,” *Computing in Science & Engineering*, vol. 15, no. 3, pp. 72-82, 2013.
- [26] C. Collins, G. Penn, and S. Carpendale, “Bubble sets: revealing set relations with isocontours over existing visualizations,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1009-1016, 2009.
- [27] B. Alper, N. Riche, G. Ramos, and M. Czerwinski, “Design study of linesets, a novel set visualization technique,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2259-2267, 2011.