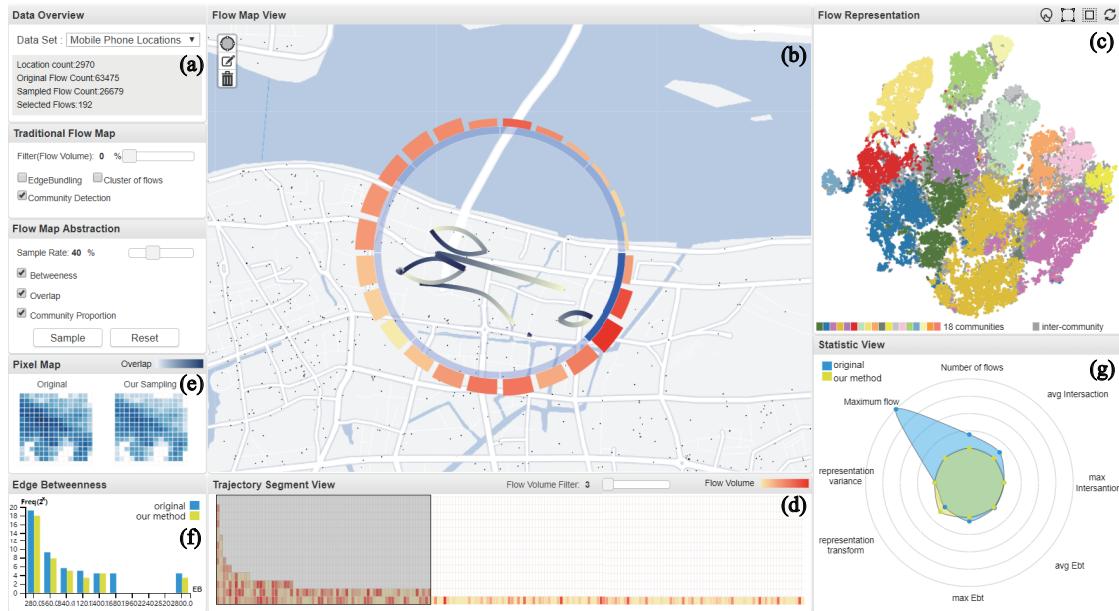


# Visual Abstraction of Large Scale Geospatial Origin-Destination Movement Data

Zhiguang Zhou, Linhao Meng, Cheng Tang, Ying Zhao, Zhiyong Guo,  
Miaoxin Hu and Wei Chen, *Member, IEEE*



**Fig. 1.** The visualization interface for a mobile phone location dataset. (a) A control view. (b) A map view supporting the presentation of OD flows and an interactive definition of flow wheels to focus on local areas of interest. (c) A word embedding view to characterize interactions of OD flows, in which each point corresponds to an OD flow. (d) A matrix view showing continuous trajectory segments, and allowing users to interactively highlight flows of interest. (e-g) present intermediate information in the course of sampling, guiding users to achieve a desired visual abstraction of large scale OD movements.

**Abstract**—A variety of human movement datasets are represented in an **Origin-Destination(OD)** form, such as taxi trips, mobile phone locations, etc. As a commonly-used method to visualize OD data, flow map always fails to discover patterns of human mobility, due to massive intersections and occlusions of lines on a 2D geographical map. A large number of techniques have been proposed to reduce **visual clutter of flow maps**, such as filtering, clustering and edge bundling, but the correlations of OD flows are often neglected, which makes the simplified OD flow map present little semantic information. In this paper, a characterization of OD flows is established based on an analogy between OD flows and natural language processing (NLP) terms. Then, an iterative multi-objective sampling scheme is designed to select OD flows in a vectorized representation space. To enhance the readability of sampled OD flows, a set of meaningful visual encodings are designed to present the interactions of OD flows. We design and implement a visual exploration system that supports visual inspection and quantitative evaluation from a variety of perspectives. Case studies based on real-world datasets and interviews with domain experts have demonstrated the effectiveness of our system in reducing the visual clutter and enhancing correlations of OD flows.

**Index Terms**—Visual abstraction, human mobility, origin-destination, flow map, representation learning

## 1 INTRODUCTION

With the development of location acquisition technologies, such as GPS devices on vehicles and in mobile phones, huge volumes of geospatial datasets have been collected. Thus, unprecedented opportunities are provided to get insights into human mobility, which is of great value in a variety of application domains, ranging from urban planning to social behavior analysis.

Geospatial OD movement data is a common form to describe human mobility, with a great deal of records aggregated into flows by pairs of OD locations, such as air traffics, taxi trips and mobile phone locations. In visualization community, a set of techniques have been designed to present the human mobility patterns, like OD matrices and flow map. In an OD matrix, the rows and columns correspond to OD

- Zhiguang Zhou, Zhiyong Guo and Miaoxin Hu are with School of Information, Zhejiang University of Finance and Economics. E-mail: {zghzhou1983, zyguo1998, humiaoxin}@zufe.edu.cn.
- Linhao Meng and Wei Chen are with State Key Lab of CAD & CG, Zhejiang University. E-mail: alice.menglh@gmail.com, chenwei@cad.zju.edu.cn. Wei Chen is the corresponding author.
- Cheng Tang is with Information School, Zhejiang Sci-tech University. E-mail: tangcheng3310@gmail.com.
- Ying Zhao is with Central South University. E-mail: zhaoying@csu.edu.cn.

Manuscript received 31 Mar. 2018; accepted 1 Aug. 2018.

Date of publication 16 Aug. 2018; date of current version 21 Oct. 2018.

For information on obtaining reprints of this article, please send e-mail to:

reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TVCG.2018.2864503

locations and the visual mapping of cells represents flow magnitudes, while the lack of spatial context information limits its further applications. Different from OD matrices, flow map is directly drawn on the geographical map, and pairs of OD locations are connected with straight or curved lines. The flow magnitude information is presented by means of the color or width mapping of lines. With the increasing size of geographical data, a large amount of interactions and occlusions of lines largely disturb the visual perception of flow map, in which patterns of human mobility are hardly discovered. Therefore, an appropriate abstraction of flow map is required to enable user easily capture features of mass mobility behaviors.

Many techniques have been proposed to simplify the visualization of flow map, such as filtering, bundling and clustering. Filtering methods present major flows with magnitudes larger than a user-specified threshold. The close flows are spatially grouped and aggregated in edge bundling schemes. The clustering methods extract the inherent patterns from massive OD movement data and select a subset of flows to represent the major patterns. The above abstraction schemes simplify massive flow map and enhance pattern discovery from different perspectives, which just consider the spatial and magnitude information of single flows. In fact, there are plenty of interactions among different flows, which are important for flow map to visually present the patterns of human mobility. For example, two air traffic OD flows are formed by the flights from Shanghai to Tokyo and Tokyo to New York. The interactions of two OD flows will be close enough when a large amount of tourists go on trips from Shanghai to New York by way of Tokyo. Therefore, a good characterization of the interactions of OD flows will do great favors for the abstraction of OD movement data, and the abundant semantic features will be largely enhanced in the flow map.

In this paper, we propose a visual abstraction approach for the exploration of large scale geospatial OD movement data. Firstly, the analogy between OD flows and NPL terms is established and real-valued vectors are trained to characterize interactions of OD flows based on a Word2Vec model. Then, a *t-SNE* method is employed to transform high-dimensional vectors into 2D coordinates, and an adaptive sampling model is designed in the projection view to select OD flows with abundant semantic interactions, meanwhile the spatial distribution and network topology are further considered to optimize the sampled flows. In order to enhance the readability of flow map, we design an interactive flow wheel, by means of which the users can easily focus on the local regions of interest, and intuitively capture the semantic patterns based on the abstracted glyphs. A visual abstraction visualization system for massive OD flow map is finally implemented enabling users to interactively explore the patterns of human mobility and a set of quantitative comparisons are integrated to evaluate the confidence of sampled flows. Case studies based on real-world datasets and interviews with domain experts are conducted to demonstrate that our system is able to reduce the visual clutter and enhance correlations of OD flows. To the best of our knowledge, this system is the first to take interactions of OD flows into the abstraction of flow map.

The major contributions of our work are summarized as follows:

- An analogy between massive human OD movements and NPL terms is established and a representation model is used to quantify the implicit interactions of OD flows based on massive human movements.
- An iterative multi-objective sampling model is designed to select OD flows in a vectorized representation space, with the semantic interactions of OD flows well retained.
- A set of visual designs are integrated into a user-friendly visualization framework, enabling users to easily achieve a correlation-enhanced flow map free from visual clutter, and further explore human mobility patterns of interest.

The rest of this paper is structured as follows: The related work is reviewed in Section 2. Section 3 presents the analysis tasks and the system overview. The visual designs and the visual analytics method

are detailed in Section 4. Case studies in addition to domain-expert interviews are discussed in Section 5 and finally our conclusions are drawn in Section 6.

## 2 RELATED WORK

The visual analytics of geospatial flows receives substantial attentions in visualization community [4, 25, 28]. Three relevant topics are covered in this section, including OD data visualization, flow map simplification and word embedding technology.

### 2.1 Visualization of OD movement data

The geospatial OD movement is often termed as ‘flow’, which refers to the movements between pairs of locations [4, 14]. Existing studies for flow data visualization can be classified into three categories, such as OD matrix [17], OD map [47, 48] and flow map [42]. In an OD matrix, the rows and columns correspond to the locations of origins and destinations, and the cells are shaded based on the magnitudes of movement, such as speed, duration and volume. A major drawback of OD matrices is the lack of spatial context [47]. Marble et al. reordered the columns and rows of OD matrices to preserve the spatial information of locations, while little spatial context can be retained due to dimension reduction [31]. In order to enhance the spatial perception of OD matrices, OD maps are designed in which the locations are regularly laid out aiming at the minimization of distortions of relative spatial positions [47, 48]. Similar to matrices, each location is represented by a matrix cell. The difference is that a small matrix is embedded within the cell to present the magnitude information by means of colors [50]. The relationships and flow magnitudes of locations can be achieved through multiple small matrices by means of OD matrix and OD maps, but it is difficult to visually present a large amount of matrices especially for massive OD movement data [3].

Different from the indirect visualization techniques based on OD matrix, flow map is a direct flow data visualization method, in which each pair of OD locations is connected with a straight or curved line [22, 23, 29, 38, 41]. Holten et al. conducted an evaluation on a variety of possible representations of directed links [21]. The flow magnitudes can be represented by proportional widths or colors [42, 48]. As the origins and destinations can be located at any distance, massive intersections and occlusions of lines will occur in the flow map, leading to more visual clutter especially with the increasing size of OD movement data. In our work, flow map is employed to visualize OD movement data, and an iterative multi-objective sampling strategy is designed to reduce visual clutter of flow map, meanwhile the interactions of OD flows are retained as much as possible.

### 2.2 Flow map simplification

A large number of approaches have been proposed to reduce visual clutter of flow map. Filtering is a commonly-used method for flow map simplification, by means of which only the flows with those magnitudes larger than a specified threshold will be presented on the geographical map [15, 18, 37, 42, 51]. For example, Wang et al. filtered out route links with smaller flow volumes, and presented major flows to avoid visual clutter [45]. Ferreira et al. proposed an interactive visualization mechanism to support origin-destination queries [14]. Edge bundling is another strategy for the simplification of flow map, in which close OD flows are spatially grouped and merged [9, 11, 13, 35]. The visual clutter can be further reduced by removing the middle parts of OD flows [8, 34]. However, the bundling results will hardly convey insightful information of flows, when the radial flows from/to one location prevail over others. Besides, undesired ambiguities are easily produced in the geographical map with edge bundling, which will mislead the impression of actual flows and mobility [44]. Clustering methods can also be used to simplify flow map, such as hierarchical clustering and kernel-based density estimation. Zhu and Guo proposed a hierarchical clustering scheme to aggregate close flows, and presented abstracted flow patterns aiming at the reduction of visual clutter [51]. They further extracted inherent patterns from massive flows via kernel-based density estimation [19]. Andrienko et al. proposed a relevance-aware density-based trajectory clustering method for air traffic analy-

sis [2]. The visual clutter of flow map is reduced by means of these methods, while the spatial context and detailed connections are lost.

Sampling is a widely-used method in computer graphics community. Stochastic sampling is able to reduce the visibility of aliasing by replacing it with incoherent noise. Better results can generally be obtained with blue noise sampling [20, 36, 49]. Ahmed designed a push-pull optimization algorithm for the blue noise sampling by enforcing spatial constraints on given point sets, which offers flexibility for trading-off among different targets [1]. Sun et al. proposed a frequency analysis method to generate line segment samples with fine properties that can effectively reduce variance and aliasing artifacts [40]. In recent years, sampling has been well designed to reduce visual clutter for visual exploration of large scale datasets [24]. Chen presents a visual abstraction scheme that employs a hierarchical multi-class blue noise sampling technique to show a feature-preserving simplification and support a visual inspection and quantitative analysis [10]. Liu presented a warpping function and blue noise sampling based approach that extracts a representative subset of an ensemble of points while maintaining the statistical spatial distribution of full ensemble [27]. Guo and Zhu designed a flow selection method to filter out duplicate information in the smoothed flows [19].

Different from the above flow filtering and sampling techniques that focus on flow magnitude or spatial distribution, our approach extracts the relationship between flows by means of representation learning and conducts a semantic-based sampling scheme to retain meaningful flows as much as possible.

### 2.3 Word Embedding

In the field of NLP, word embeddings are widely adopted to project semantic meanings into a geometric space. Compared with One-hot representations, distributed representation is a compact, dense and low-dimensional representation, with each embedding dimension representing a latent feature of the word [43]. Mikolov et al. proposed an efficient word embedding algorithm named 'Word2Vec', building a neural network to map words into real-number vectors, with the desideratum that words sharing similar meanings should be represented close to each other in the vector space [33]. The Word2Vec algorithm has extended beyond parsing sentences in the wild. It can be applied in various NLP tasks including sentiment analysis [39], entity linking [6], machine translation [32] and so on.

Word embedding techniques can also be applied in a variety of other fields. Berger et al. designed a visualization scheme named Cite2Vec to support a dynamically exploration of document collections [5]. Node2Vec is an algorithmic framework for node and edge prediction tasks, which defines a mapping between original nodes and a set of low-dimensional features, with the network neighborhoods of nodes preserved as much as possible [16]. Based on an analogy between transportation elements and NLP terms, Word2Vec model can also be employed to quantify the implicit traffic interactions among roads [26]. Similarly, we encode OD flows as words and trajectories as paragraphs. A Word2Vec model is further applied to obtain vectorized representation of OD flows and an adaptive sampling method is designed to reduce visual clutter of OD flow map and select OD movements with abundant interactions and human mobility features.

## 3 TASK ANALYSIS AND SYSTEM OVERVIEW

In this section, we firstly describe the real-world datasets used in this paper. Subsequently, we summarize the analytic tasks identified from the interviews of domain experts, and then present the workflow of the proposed visual analytics system.

### 3.1 Data Description

In this paper, two real-world datasets are used to demonstrate the effectiveness of our system, including the bicycle sharing system and the mobile phone locations. The bicycle sharing system is a traditional OD movement data, in which 5800 bicycles are borrowed and returned across 580 stations. The time span of the dataset is from January 2013 to December 2016, and 10 millions of usage records are contained including a variety of attributes such as the borrowing time, returning

time, the latitude and longitude information of origin and destination stations, and the user information. According to the records of human mobility, a flow map is drawn with 580 locations and 64 thousands of OD flows. The other dataset records the mobile phone locations in the span of a week. There are 2755 cell stations located in the urban center, recording massive OD trajectories of about 1.1 millions of persons. The corresponding flow map is drawn with 2755 stations and 1.3 millions of flows. The visual clutter is heavily generated in the flow map for the two datasets due to overdrawn of massive OD flows, which makes users hardly perceive the patterns of human mobility.

### 3.2 Task Analysis

After a detailed discussion with domain experts in the form of structured interviews, we compiled a list of analytical tasks for the abstraction of large scale OD movement data.

**T.1 Simplification of flow map:** How to design a visual abstraction of massive OD movements and reduce the visual clutter of flow map? Different from traditional filtering, bundling and clustering strategies, how to sample a subset of OD flows that can better represent the original massive flows as much as possible, with a variety of inherent features taken into consideration, such as spatial distribution, network topology and semantic correlation?

**T.2 Characterization of flow interaction:** The OD flows are generated based on a mass of human movements. In the traditional flow map, only the origins, destinations and magnitudes of an independent OD flow are considered. How to characterize the interactions of OD flows based on massive human mobility and retain the semantic features among flows in the simplified flow map? How to visualize and interact with the characterization of flow interaction?

**T.3 Visualization of flow abstraction:** The visualization of flow map is so simple that a large amount of features of interest are concealed in the OD flows, even in a simplified flow map. How to provide an intuitive visualization interface and embed an aesthetically appealing glyphs into the geospatial map, enabling users to easily perceive the semantic interactions of local flows, in addition to the spatio-temporal magnitude information?

**T.4 Evaluation of flow abstraction:** The detailed information of original OD flows will be inevitably lost during the course of flow abstraction, even though a variety of constraints are integrated into an interactive multi-objective sampling model. How to conduct an valid evaluation on the effectiveness of our flow abstraction method from objective and subjective perspectives? How to visually present the difference and uncertainty, allowing users to interactively optimize the course of flow abstraction?

### 3.3 System Overview

Fig.2 presents an overview of the proposed system. The OD flows are obtained through the preprocessing of original massive human mobility data. A word embedding model widely used in NLP field is applied to characterize the interactions of OD flows. With the vectorized representations of OD flows, a dimension reduction method is used to visually present the relationship among OD flows (T.2). Then, an adaptive sampling scheme is used to reduce the scale of projected points while retaining the semantic relationship of OD flows. In order to further preserve the visual perception and topological structures of OD flows, three constraints are integrated into the sampling scheme, such as flow intersection, flow importance and community distribution (T.1). We also design a flow wheel to focus on local regions of interest, enabling users to interactively explore the spatio-temporal features (T.3). A set of metrics are further calculated to evaluate the effectiveness of different sampling strategies, and a variety of visualizations are designed to visually present the metrics, such as radar chart, pixel map, bar chart and matrix graph (T.4). A rich set of interactions are integrated into the visualization system for visual abstraction of flow map, enabling users to further highlight and explore patterns of human mobility.

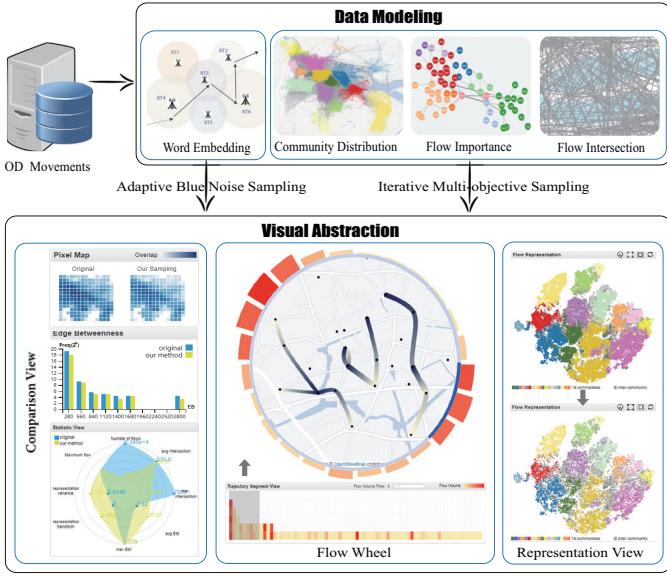


Fig. 2. The pipeline of our system, which is comprised of data modeling, visual abstraction and evaluation.

## 4 VISUAL ABSTRACTION OF MASSIVE OD FLOWS

### 4.1 Flow Representation

In order to highlight the interactions of OD flows in the simplified flow map, an analogy between OD flows and NLP terms is established and an efficient representation learning model is applied to transform OD flows into vectorized representations.

#### 4.1.1 Corpus generation

According to the analogy between OD flows and NLP terms, a mass of OD movement data can be transformed into words and sentences. An OD flow is defined as a word,  $w = (\text{origin}, \text{destination})$ , when a person moves from one location (origin) to another location (destination). A sentence is constituted based on a sequence of OD flows that a person passes in a period of time. The corpus is then generated including all the words and sentences, which will be further trained by means of a distributed representation model Word2Vec.

For example, the cell stations are widely distributed around a city area, which record the temporal locations of persons through GPS devices in the mobile phones. When a person moves across urban areas, the carry-on mobile phone is located by different nearby cell stations. Based on the mobile phone locations, the trajectory of a person can be depicted with a sequence of OD flows:

$$Tr = \{f_1, f_2, \dots, f_m\} \quad (1)$$

The OD movement data is composed of trajectories of persons, and the document is generated as  $D_{tr} = \{Tr_1, Tr_2, \dots, Tr_n\}$ , where  $n$  is the account of persons. Thus,  $D_{tr}$  are the available corpus for further representation learning.

#### 4.1.2 Vectorized representation

In our system, Word2Vec is applied to train the distributed representations of flows. There are two options for the training of word vectors from raw documents. One is a continuous bag-of-words model (CBOW) which predicts target words based on the context, while the other is a continuous skip-gram model which predicts the surrounding words from the target words. Algorithmically, the above two models are similar, while the difference is that CBOW performs better for limited scale of training data, and skip-gram model is suitable for enough training data. As mentioned that a mass of OD movements and human trajectories have been transformed into words and documents, which provide enough training datasets. Therefore, we apply the skip-gram

model to train the OD flows in our experiments. The objective of the skip-gram model is to maximize the average log probability:

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j}|w_t) \quad (2)$$

where  $c$  is the size of training context and  $T$  is the length of training word sequence.  $p(w_{t+j}|w_t)$  represents the probability that  $w_{t+j}$  is the corresponding contextual word given word  $w_t$ , which can be computed by a softmax function:

$$p(wO|wI) = \frac{\exp(v'_{wO}^T v_{wI})}{\sum_{w=1}^W \exp(v'_{wO}^T v_{wI})} \quad (3)$$

where  $v_w$  and  $v'_{wO}$  are the "input" and "output" vector representations of  $w$ , and  $W$  is the number of words in the vocabulary. To enhance the performance of calculation, the skip-gram model uses Negative sampling in which an objective is represented as an alternative to every  $\log P(wO|wI)$  term in the skip-gram objective as follows:

$$\log \sigma(v'_{wO}^T v_{wI}) + \sum_{i=1}^k E_{wi \sim P_n(w)} [\log \sigma(-v'_{wi}^T v_{wI})] \quad (4)$$

The representation results are high-dimensional vectors which can represent the underlying weights or importance of each word. The similarity between two words is measured by the cosine similarity. By training the corpus, we can obtain the representations of OD flows. If flows have similar "contexts" (that is the flows are likely to appear in the same trajectories), similar vectors will be probably achieved by the model.

#### 4.1.3 Dimension reduction

It is a reasonable way to conduct a direct simplification strategy in the original high-dimensional space. However, it is a difficult task to simplify high-dimensional data items by means of traditional sampling strategies [12], especially when the representation of an OD flow is depicted with more than 200 dimensional vectors obtained by Word2Vec model, which is well over the capability of human visual cognition.

As a commonly-used dimension reduction method, t-Distributed Stochastic Neighbor Embedding (*t-SNE*) proposed by Maaten and Hinton [30], can be applied to reduce multiple dimensions of vectorized representations and visualize the OD flows with two-dimensional points. As the *t-SNE* is capable of capturing much of local structure, meanwhile revealing global structure of data items [46], it will do great favors for adaptive sampling operations and preserve more semantic information of OD flows in despite of the loss of accuracy. The distance between each pair of points largely depicts the correlation between OD flows in the high-dimensional semantic space. That is the OD flows would frequently appear in the similar trajectory context, if two points are located close to each other. Thus, we can sample the OD flows in the low-dimensional word embedding space, while the interactions of OD flows will be largely retained. Fig.3 presents the flow map and its corresponding vectorized representations. A community detection method is applied to classify the OD movements into in-community and inter-community flows. A variety of distinctive colors are used to map OD flows and corresponding points enabling users to quickly perceive the urban distribution of human mobility.

## 4.2 Flow Sampling

The visual clutter of flow map for massive OD movement data can be reduced by means of filtering or clustering techniques, but the selected flows only represent the magnitude information or local spatial distribution, with the interactions between flows neglected. In the above subsection, an analogy between flow interactions and NLP terms is established, and the vectorized representations of flows are projected into a low-dimensional space by means of *t-SNE*. In order to retain the interactions of OD flows in the simplified flow map as much as possible, we conduct an iterative blue noise sampling model that estimates

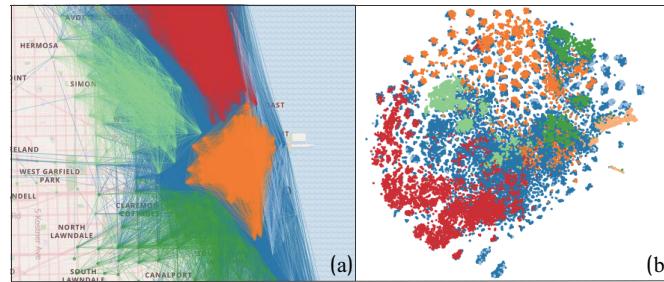


Fig. 3. The visualizations of flow map and corresponding word embeddings based on a bicycle sharing system dataset. (a) presents the flow map. (b) presents the vectorized representations of OD flows. According to the shaded colors of points, we can find that the OD flows in the word embedding space are distributed similar to their community features, which is in accordance with the prior knowledge that the OD flows in a community share similar semantic context.

the point density in the semantic embedding space, and a set of visual perception and network topology constraints are further integrated to resample and optimize the resultant flow map.

#### 4.2.1 Adaptive blue noise sampling

Given the massive projected points in word embedding space that correspond to the OD flows in geographical map, an adaptive blue noise sampling method is applied to reduce the scale of OD flows and retain the distribution of points as much as possible.

In the course of sampling, a poisson disk radius  $r$  at the trial location  $p$  is built adaptively according to conflict checking.  $r$  generalizes the distance constraint in the Poisson disk sampling for a point. Kernel density estimation (KDE) is employed to estimate distribution of 2D projection points. Let  $P = \{p_1, p_2, \dots, p_m\}$  be the data points. Mathematically, the density at location  $p$  is computed as followed:

$$f(p) = \sum_{j=1}^m K_h(p - p_j) \quad (5)$$

where  $K_h(\cdot)$  is a kernel function with bandwidth  $h$ , which determines the smoothing degree of the reconstructed density field. A Gaussian kernel is used in our approach. Then, the radius  $r$  of poisson disk is defined as  $ra/f(p)$ .  $ra$  is a user-defined parameter, which is able to control the sampling rate.

Despite that the adaptive sampling in the word embedding space is able to retain the interactions of OD flows, there are still a set of flows with valuable semantic or spatial features filtered out in the sampled result, along with the reduction of sampling rate. In order to preserve the original geographic shape of the flows, a multi-objective sampling strategy is further proposed with a number of constraints taken into consideration, by means of which the selected flows can be replaced with other flows in the iterative sampling process, to fulfill various user requirements.

#### 4.2.2 Iterative multi-objective sampling

Three objectives are integrated into the adaptive blue noise sampling, including flow intersections in the geographical map, as well as edge betweenness and community distribution in the geospatial network. For each sample  $p_i$ , the underlying mechanism of our approach allows each visited point to be replaced with an optimal point in its optional set. The integration of three objectives is detailed as below.

**Flow intersection optimization:** The intersections and occlusions of OD flows largely disturb the visual perception of flow map. Therefore, the measurement of flow intersections is integrated into the sampling model, aiming at the reduction of intersections of sampled flows. The intersection degree of flow  $p_i$  is measured by:

$$FI(p_i) = \sum_{k=0, k \neq i}^m Is\_Intersection(p_i, p_k) \quad (6)$$

where  $Is\_Intersection(\cdot)$  is a function to judge the intersection between flow  $p_i$  and flow  $p_k$ , and  $m$  is the total account of OD flows.

In the course of sampling on the projected points, an optional point set  $P_i$  are generated for the visited sample point  $p_i$ . In order to optimize the simplification of OD flow map,  $p_i$  will be replaced by its optional points, if  $FI(p_i)$  is larger than  $FI(p_j)$ .

**Flow importance optimization:** A large amount of OD flows form a weighted directed network, in which weight is depicted with flow magnitude and direction is represented by flow orientation. Thus, a set of traditional topological metrics can be applied to measure the importance of OD flows, such as edge betweenness. Therefore, we compute the edge betweenness of flow  $p_i$ :

$$EB(p_i) = \sum_{p_j \neq p_i \neq p_k \in P} \frac{\sigma_{p_j p_k}(p_i)}{\sigma_{p_j p_k}} \quad (7)$$

where  $\sigma_{p_j p_k}$  is the shortest path sets of  $p_j$  and  $p_k$ ,  $\sigma_{p_j p_k}(p_i)$  is the shortest path sets of  $p_j$  and  $p_k$  which pass flow  $p_i$ .

Given the visited point  $p_i$  and its optional point set  $P_i$ ,  $p_i$  will be replaced by  $p_j$  in the resampling course, if  $EB(p_i)$  is smaller than  $EB(p_j)$ . Therefore, the measurement of flow importance can be integrated into the sampling model to guarantee that those important flows can be retained as far as possible.

**Community distribution optimization:** Based on the weighted network formed by OD flows, a modularity-based community detection method [7] is employed to classify the OD flows into different clusters, such as in-community flows and inter-community flows.

In order to retain the community structures in the sampled flow map, we replace the visited sample with its optional points that will make the community distribution largely approximate to the original community distribution. For example, a visited sample  $p_i$  will be replaced by  $p_j$  when  $\Delta_{CD}(p_i) - \Delta_{CD}(p_j) > 0$ . As a measurement of community distribution,  $\Delta_{CD}(p_i)$  is defined as:

$$\Delta_{CD}(p_i) = |CD_{sampling}(p_i) - CD_{original}| \quad (8)$$

where  $CD_{original}$  is the original probability distribution of communities before sampling.  $CD_{sampling}(p_i)$  is the probability distribution of communities after sampling.

Given above mentioned constraints, an iterative multi-objective sampling model is established by means of which the OD flows can be resampled aiming at the optimization of a flow map. The sampling technique is detailed through a pseudocode as shown in Algorithm 1.

### 4.3 Visual Interface

The visual interface of our system consists of four components, such as a geographical map view, a word embedding view, an interactive flow wheel and a set of comparison views for the evaluation of flow map simplification.

#### 4.3.1 Map View

The spatial context of flow map is important for the exploration of patterns of human mobility. Therefore, a major view of the proposed system is a geographical map, in which the OD flows are visualized by means of straight lines. A variety of operations are integrated in this system to reduce the visual clutter of flow map as shown in Fig.1 (a), such as magnitude-based filtering, cluster-based edge bundling and iterative multi-objective sampling, which enhance the visual perception of flow map from different perspectives (T.1).

In order to get deeper insights into the simplified flow map and capture human mobility patterns of interest, we design a user-friendly interface enabling analysts to focus on a local region on geographical map (T.3), which is named "flow wheel". The flow wheel is a circular design that is comprised of three components: (1) The temporal flow magnitude information is encoded by bar charts located outside of the flow wheel. There are 24 bars distributed on the flow wheel and each bar represents an hour in a day. The heights of bars present the account of flows located in the focus area, and the colors present the average magnitude of the flows. We also design a circular ring inside of the

**Algorithm 1:** Multi-objective possion disk sampling

---

**Input:**  $P$ : the multivariate data points;  $r$ : possion disk radius;  $maxI$ : the maximum steps of iterations;  $P''$ : a temporary array for optional sets;

**Output:**  $P'$ : the output samples;

- 1  $kde = GetKdeDistribution(P)$ ;
- 2  $P', P'' = AdaptivePoissonDiscSampling(P, kde, r)$ ;
- 3 //Iteratively optimize sampling results;
- 4  $i = 0$ ;
- 5 **while**  $i <= maxI$  **do**
- 6   **for each**  $p_i \in P'$  **do**
- 7     **for each** neighbor sample  $p_j$  of  $p_i$  in  $P''$  **do**
- 8       **if**  $FI(p_j) < FI(p_i)$  **then**
- 9         Exchange  $p_i$  and  $p_j$ ;
- 10   **for each**  $p_i \in P'$  **do**
- 11     **for each** neighbor sample  $p_j$  of  $p_i$  in  $P''$  **do**
- 12       **if**  $EB(p_j) > EB(p_i)$  **then**
- 13         Exchange  $p_i$  and  $p_j$ ;
- 14   **for each**  $p_i \in P'$  **do**
- 15     **for each** neighbor sample  $p_j$  of  $p_i$  in  $P''$  **do**
- 16       **if**  $\Delta_{CD}(p_i) - \Delta_{CD}(p_j) > 0$  **then**
- 17         Exchange  $p_i$  and  $p_j$ ;
- 18   *i = i + 1*;

---

bar charts, by means of which the users can easily focus on a period of interest. According to the common sense, every three hours are aggregated to classify the day into different time intervals to capture more features of flow interactions. (2) In order to identify human mobility features, we design a matrix map to present the interactions of flows, and provide a threshold for users to filter out flows with smaller magnitude values or fewer interactions. In the matrix map, each column of flows represent a trajectory segments that persons move across. The trajectory segments are arranged in rows enabling users to easily capture the distribution of flow interactions and select trajectories of interest. The colors of cells present the magnitude values of corresponding flows. (3) In order to further highlight the flow interactions in a local area surrounded by the flow wheel, the trajectory segments selected by users in the matrix map are aggregated into meaningful curves. The widths of curves present the co-occurrence volume magnitude values of corresponding flows. The origins of curves are colored in yellow while the destinations of curves are colored in blue, which largely enhance the visual perception of patterns of human mobility. As shown in Fig.1 (b), a flow wheel is interactively defined on the geographical map. The user selects trajectories of interest in the matrix map as shown in Fig.1 (d), and the interactions of corresponding flows are highlighted in the central of flow wheel.

#### 4.3.2 Representation View

In order to construct a compact correlation between the vectorized representations and corresponding OD flows, a word embedding view is provided to represent the semantic difference of OD flows (T.2). Each point in the word embedding view corresponds to a OD flow in the map view and the distance of points represents the semantic interactions of OD flows. Three interactive selection operations are provided to explore the spatial context of 2D projections, including single point selection, local region selection and community selection. The corresponding icons are listed in the upper right corner of word embedding view, as shown in Fig.1 (c). (1) The single point selection allows users to quickly focus on an independent OD flow, and establish a visual connection between the points in word embedding view and lines in map view, which is much helpful for a detailed exploration and comparison of a small number of OD flows. (2) The local region selection is provided to focus on the a cluster of points in word embedding view,

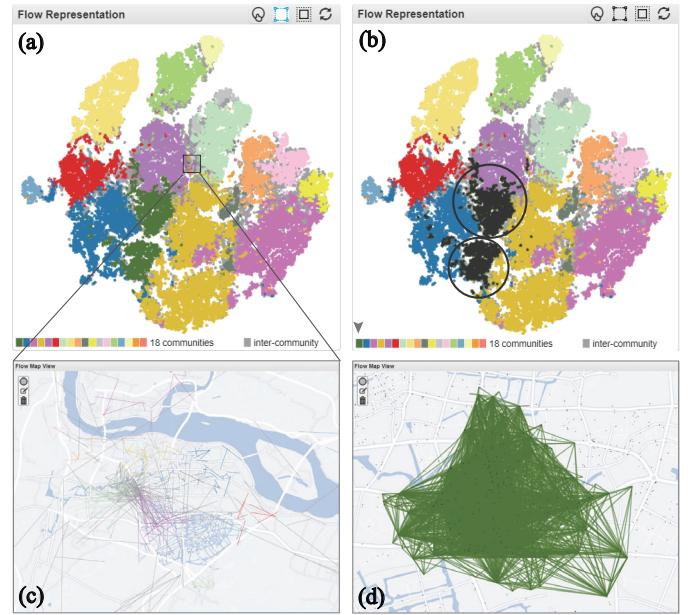


Fig. 4. The different distributions of OD flows in the word embedding view and map view. (a) a word embedding view in which different colors correspond to a variety of communities. (c) a set of OD flows are highlighted in the map view, since the corresponding points are specified in (a). (b) a community of OD flows is selected and the corresponding points are encircled in the word embedding view. (d) highlights the OD flows in the community selected in (b).

allowing users to explore the geographical distribution of OD flows with close semantic relationship. (3) Community detection is able to classify the OD flows into different communities. A set of grids are arranged at the bottom of word embedding view, and they are shaded in different colors and correspond to different communities. Community selection focuses on community features in the geographical network, which is able to demonstrate the effectiveness of our sampling method based on the word embedding view.

#### 4.3.3 Comparison View

In order to guide the sampling course and evaluate the validity of visual abstraction results, a list of metrics are calculated to measure the simplified flow map from different perspectives (T.4). A radar chart is designed to visually present the difference between the original and simplified flow maps, with a set of attributes taken into consideration: (1) The account of flows is used to represent the sampling rate. (2) The average and maximum intersection of flows are used to evaluate the visual clutter of simplified flow map. (3) The average and maximum importance (edge betweenness) are used to evaluate the retaining of topological structures in the simplified flow map. (4) The change of the barycenters and the variance of flows in the word embedding view are used to measure the difference between original distribution and sampled distribution in the vectorized representations. (5) The account of flows that massive persons move across and the co-occurrence magnitude values are larger than a user-specified threshold, is used to evaluate the retaining of flow interactions in the simplified flow map. Fig.1 (g) shows a radar map based on a user-defined sampling objective, by means of which the metric distribution can be easily captured through the shapes of radar chart. Furthermore, a pixel map is designed to present the distribution of flow coverage of different sampling strategies as shown in Fig.1 (e), which can be used to measure the visual perception of simplified flow map. A group histogram is applied to present the retaining of edge betweenness through different sampling schemes as shown in Fig.1 (f), according to which the users can find whether important flows are preserved in the simplified flow map. Based on the measurements and visualizations of different met-

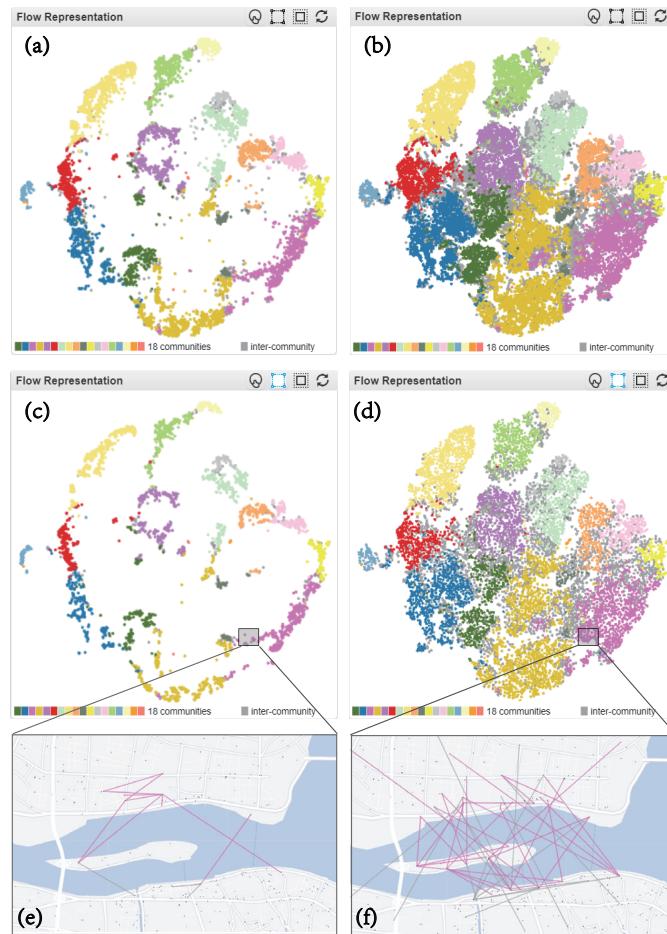


Fig. 5. The simplified results based on different schemes at the sampling rate of 20% (a,b) and 10% (c,d). Filtered projection views are presented in (a) and (c) by means of a magnitude filtering operation. Sampled projection views are presented in (b) and (d) by means of the proposed adaptive blue noise sampling. (e) and (f) are the flow maps in accordance with the points specified by a same local area of interest in the projection views.

rics, we can easily evaluate the effectiveness of simplified flow map from different perspectives.

## 5 EVALUATION

In this paper, a web-based system is implemented using Javascript, supporting a real-time exploration of human mobility patterns. Complex data modelings are conducted in an off-line preprocessing phase such as the representation learning by means of Word2Vec. In order to further evaluate the effectiveness of our system, a set of case studies are carried out based on mobile phone location dataset, and a quantitative comparison is conducted to validate that our multi-objective sampling method is able to reduce visual clutter and retain structures of interest in the simplified flow map.

### 5.1 Case Study

In this subsection, a real-world mobile phone location dataset is used to evaluate the effectiveness of our proposed system.

**Vectorized Representation:** As shown in Fig.3, we find that the OD flows in the word embedding space are distributed similar to their community features. It corresponds to the prior knowledge that those highly related OD flows might share similar semantic context. Therefore, the results largely demonstrate the usability of Word2Vec in the representation of OD flows. In this case, we intend to explore the different distributions of OD flows in the word embedding space and

geospatial space. Fig.4 (a) presents a word embedding space. When a local region of interest is specified, the corresponding OD flows are highlighted in the map view as shown in Fig.4 (c). We can see that the OD flows are distributed all over the map view, which do not present community features or other distribution patterns. Similarly, when a community is specified in the word embedding view as shown in Fig.4 (b), the OD flows are highlighted in both the map view and geographical view. We can see that the OD flows are densely distributed in a local area and present intense community features as shown in Fig.4 (d). However, the corresponding projections are classified as a set of clusters, that are distributed separately in the word embedding view as shown in Fig.4 (b). Therefore, we conclude that the OD flows in the word embedding space are distributed similar but not equivalent to their community features. Sometimes, the distributions of OD flows in the word embedding view present more detailed semantic information compared with community distribution.

**Adaptive Blue Noise Sampling:** In order to demonstrate the effectiveness of the proposed adaptive blue noise sampling, we compare the simplified results based on different schemes. Fig.5 (a) and Fig.5 (b) present the simplified results at the same sampling rate 20%. Fig.5 (c) and Fig.5 (d) present the simplified results at the same sampling rate 10%. We can find that the filtered points the magnitude values of which are larger than a user-specified threshold are distributed in Fig.5 (a) and Fig.5 (c). They are always clustered and distributed in local regions with a lot of blank space left in the projection view. The reason might be that flows with larger volume often exist in the center of urban area with close enough relationship. As a variance, Fig.5 (b) and Fig.5 (d) present the points sampled by means of our adaptive blue noise sampling strategy, in which the original distribution of points is largely preserved. Therefore, the correlation of OD flows is highly retained in our sampling strategy, which can be further verified by comparison of a local region in the projection view. The corresponding OD flows of local region defined by user are highlighted in Fig.5 (e) and Fig.5 (f). It can be found that many OD flows stride across the river are filtered out in Fig.5 (e) due to their small magnitude values. However, those filtered flows are the important routes that link different areas on both sides of the river. They share abundant correlations with other OD flows that are largely retained based on our sampling strategy as shown in Fig.5 (f). Therefore, it can be concluded that the proposed adaptive blue noise sampling method is able to enhance the visibility of flow map with the semantic relationship among different OD flows largely retained in the simplified result.

**Multi-objective Sampling:** In order to further highlight the usefulness of iterative sampling model with multi-objective constraints integrated, we compared the sampling results under different conditions at a common sampling rate of 10%. Fig.6 (a-c) present the simplified flow maps obtained based on an adaptive blue noise sampling model without considering any constraints. Fig.6 (d-f) present the simplified flow maps with a variety of constraints taken into consideration. The small views at the upper right corner present changes of corresponding attributes after different sampling operations. For example, compared with Fig.6 (a), more important flows shaded in black are well retained as shown in Fig.6 (d), since a measurement of edge betweenness is considered for an iterative course of resampling. The similar conclusion is easily drawn according to the group histograms, in which the proportion of flows with large edge betweenness is well preserved. We also can find that the visual perception of Fig.6 (e) is better than that of Fig.6 (b), which can be further verified through comparison of pixel maps. The reason is that a measurement of flow intersection is integrated into the sampling model for an iterative replace of sampled OD flows. Also, a matrix map is designed to present the difference between sampled and original communities. Each row and column represents a community and the colors of grid cells represent the difference between corresponding communities. In order to preserve community distribution, a measurement of portion change is integrated into the model. Based on an adaptive blue noise sampling, big changes generated make the matrix map colorful as shown in Fig.6 (c). As a variance, small changes are generated in the matrix map of Fig.6 (f), since the constraint of community distribution is integrated

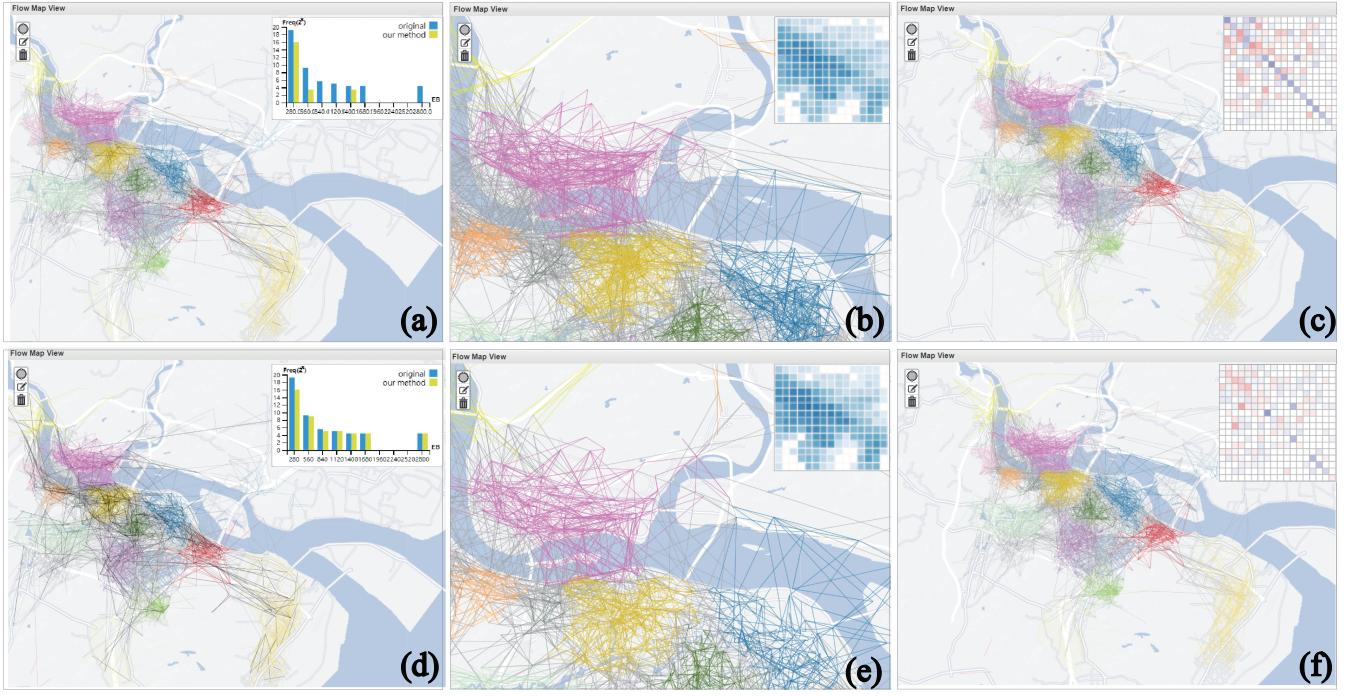


Fig. 6. The comparison for sampled flow maps based on different multi-objective sampling strategies. (a-c) present the sampling results by means of adaptive blue noise sampling without considering any constraint. (d-f) present the different sampling results by means of multi-objective sampling schemes with a variety of constraints taken into consideration. The statistics are also presented in a variety of visualization views at the upper right corner for an insightful comparison.

into the sampling model. The effectiveness of multi-objective sampling can be further demonstrated through quantitative comparisons in the next subsection.

**Interactive Flow Wheel:** In our system, flow wheel is designed to focus on local areas of interest, enabling users to quickly capture the spatio-temporal patterns of human mobility. In order to demonstrate the effectiveness of flow wheel, we invited a user to explore the sampled OD flows on a geographical map by means of flow wheel. He first moved the mouse on the geographical map to find local areas of interest. By specifying the center and radius, a local region was determined and a flow wheel was immediately presented. The radar chart in the lower right corner presented the statistical information of the specified local area. Then, he modified the sampling objective and sampling rate according to the comparison of radar chart, to achieve a desired sampled flow map. According to the bar charts arranged around the wheel, he found that the amounts and average magnitude values of sampled flows in the time intervals from 6:00 am to 9:00am and from 18:00 pm to 21:00 pm were larger than that of other time intervals. When he clicked the segment from 6:00 am to 9:00am in the ring, the OD flows were transformed into aggregation curves. The trajectory segments are displayed in the matrix map. He brushed the trajectories with large amount of segments, and found in the Fig.7 (a) that a lot of flows took the station located in the upper right corner of flow wheel as a destination. When he selected the time interval from 6:00 am to 9:00am, a lot of flows took the station located in the upper right corner of flow wheel as an origin. After an interview with local residents, we learned that the station in the upper right corner was located nearby a big factory. Thousands of workers pass the mobile phone location when they go to work or return from work. This fact is quite identical to the visual exploration results of the user, which further demonstrates the effectiveness of the design of flow wheel.

## 5.2 Quantitative Comparison

The statistical measurements of experimental results are presented in Table.1 to further demonstrate the effectiveness of our method by comparison for a variety of sampling strategies. The different objectives

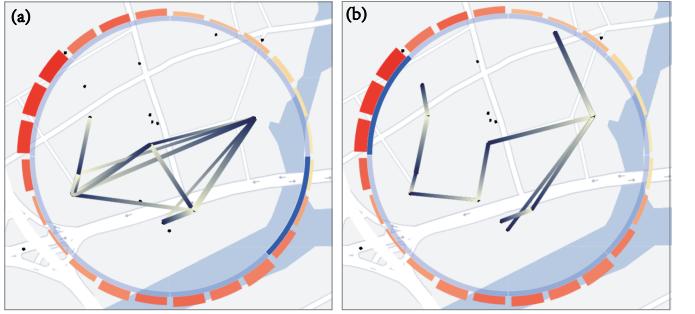


Fig. 7. A flow wheel are interactively specified to focus on a local region of interest. (a) presents the highlighted OD flows generated from 6:00 am to 9:00am. (b) presents the highlighted OD flows generated from 18:00 pm to 21:00 pm. We can easily find the spatio-temporal patterns of human mobility according to the visual mappings of OD flows.

integrated into adaptive blue noise sampling are listed in the first column, including flow intersection (FI), edge betweenness (EB), community distribution (CD) and a synthesized objective composed of all above three objectives (ALL). The sampling rates are listed in the second column and five properties are calculated to measure the difference among sampling strategies, such as average flow intersections (AFI), maximum of flow intersections (MFI), average edge betweenness (AEB), maximum of edge betweenness (MEB), difference between community distributions (DCD). According to the statistics, we can find that when an objective is specified to optimize the course of blue noise sampling, the corresponding attribute will be largely enhanced in the final sampled flow map. For example, the objective of edge betweenness is integrated into the sampling model, the AEB and MEB values are both larger than that of other sampling strategies. Note that, in single objective sampling schemes, the corresponding features can be better retained in the sampled flow maps, but other

Table 1. Comparison for different objectives and sampling rates.

OBJ	Rate	AOF	AFI	MFI	AEB	MEB	DCD
ORI	100%	63476	575.78	7039	2.82	2779.00	0.000
FI	40%	26678	528.28	3874	2.94	2779.00	0.043
EB	40%	26739	599.09	7039	5.56	2779.00	0.033
CD	40%	26850	599.13	5519	3.01	2755.33	0.019
ALL	40%	26697	548.57	4357	4.89	2779.00	0.017
FI	20%	15545	471.31	3408	3.10	2755.33	0.071
EB	20%	15531	604.88	6926	9.02	2779.00	0.044
CD	20%	15476	608.96	6926	3.05	2755.33	0.032
ALL	20%	15500	541.06	3844	7.32	2779.00	0.051
FI	10%	6207	359.85	3408	2.61	1211.97	0.131
EB	10%	6258	607.12	4851	20.93	2779.00	0.066
CD	10%	6172	599.58	4860	3.02	1169.29	0.021
ALL	10%	6241	512.82	3626	13.22	1500.92	0.245
FI	5%	3496	293.85	3051	2.83	2755.33	0.176
EB	5%	3426	609.94	3962	34.21	2779.00	0.090
CD	5%	3338	607.35	3588	3.26	1169.29	0.026
ALL	5%	3392	538.01	3874	20.40	2779.00	0.278

structures are hard to maintain. Although the sampling results considering all of the three constraints are not the best from each view of single constraint, the result is a trade-off for the three objectives. The above comparison and analysis largely demonstrate the validity of our multi-objective sampling strategy.

### 5.3 Expert Interview

The feedback of the domain experts are collected based on one-on-one interviews, which is summarized as follows:

**Visual Design and Interactions:** The experts confirmed that the visual abstraction system was well designed and the interface was quite user-friendly. By means of Word2Vec, massive OD flows can be transformed into vectorized representations, and multi-objective sampling schemes can be defined interactively to simplify the flow map. An interactive flow wheel will enhance the readability of simplified flow map, enabling users to better perceive the patterns of human mobility. They all believed that the system could be applied for an in-depth exploration of large scale human mobility data. The first expert commended that “It is difficult to capture valuable information from flow map due to visual clutter, especially with the increasing size of OD flows. This system is able to transform the OD flows into points, the difference among which largely represent semantic intersections of OD flows in the geographical map. After all, it is really simple to conduct a sampling operation on the point view rather than on the geographical map, because the difference between OD flows on the map view is hard to evaluate. Also, the interactions of OD flows can be better represented in the word embedding view, which largely enhance the effectiveness of the proposed semantic sampling strategy.”

**Usability and Improvements:** The experts also appreciated the design of flow wheel, which enabled users to interactively focus on areas of interest, and a sequence of OD flows with close interactions were described with aesthetical curves. They all agreed that the proposed visualization system will do great favors for the exploration of large scale OD movement datasets. The visual clutter is largely reduced, meanwhile the semantic interactions are better retained. In addition, the experts also provided some valuable comments for the improvements of our system. The first expert mentioned that “It is still difficult enough to understand the significance of vectorized representations. The similar word embedding points represent different OD flows which will be located far away from each other in the graphical map. How to visually present and effectively use the similarity between flows for an in-depth mining of patterns of human mobility is still a problem.” The second expert mentioned that “The multi-objective sampling method is able to reduce visual clutter of flow map with internal features of interest retained, such as spatial distribution and topological structures.

But, it is inevitable that important information will be lost due to the randomness of sampling strategy. How to visually represent the potential uncertainty generated in the course of sampling and guide users to optimize the sampling course is still an interesting problem.”

### 5.4 Discussion

There are still some issues not well resolved in this paper, which will be addressed in the future work. **First**, a representation learning model Word2Vec is applied to characterize the interactions among flows, and a *t-SNE* method is employed to visualize the flows with multiple vectors. The correlations of flows can be visually achieved and highlighted through user-friendly interactions. However, there are still some available functions of Word2Vec that are not integrated into our system, because of the different complexity of human mobility and NLP terms. For example, the correlations of words will be well retained in vectorized representations, such as “King – Man + Woman = Queen”. In the future work, we would like to conduct a high-efficiency query for those flows with equivalence relations, which will be important clues for the simplification of flow map and further exploration of human mobility patterns. **Second**, an iterative multi-objective sampling model is defined to reduce the amount of OD flows based on a set of user-defined constraints. When more and more constraints are integrated into the sampling model, the course of flow map simplification becomes so complex that it is too difficult and time-consuming to achieve a global optimization. In the future work, an energy function would be established and an approximate solving method will be tried to achieve the optimal solution of multi-objective of sampling. **Third**, by means of the proposed sampling model, the OD flows with abundant semantic interactions are well retained in the final simplified flow map. However, uncertainty will be generated in the simplified flow map which largely disturbs the visual perception of actual human mobility, due to that some important flows are inevitably filtered out. In the future work, we will design glyphs to visually present the uncertainty, allowing users to conduct a detailed and careful analysis on those areas with ambiguities.

## 6 CONCLUSION

In this paper, a representation learning model Word2Vec is employed to represent OD flows with high-dimensional vectors. Then, an adaptive blue noise sampling method is conducted to reduce the scale of OD flows with the semantic interactions well retained, and a variety of constraints are further integrated to enhance spatial distribution and network topological features of the simplified flow map. Furthermore, an interactive flow wheel is designed to focus on local areas of interest and a set of comparison metrics are measured and visually presented. Case studies based on real-world datasets, quantitative comparisons in addition to interviews with domain experts are provided to further demonstrate the effectiveness of our system in reducing the visual clutter of flow map and enhancing correlations of OD flows.

### ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments. This work was supported by the National 973 Program of China (No.2015CB352503), the National Natural Science Foundation of China (No.61872314, No.61772456, No.61672538, No.U1609217, No.U1736109), the Natural Science Foundation of Zhejiang Province (No.LY18F020024), the Humanities and Social Sciences Foundation of Ministry of Education in China (No.18YJC910017), the Open Project Program of the State Key Lab of CAD&CG of Zhejiang University (No.A1806) and the First Class Discipline of Zhejiang-A (Zhejiang University of Finance and Economics-Statistics).

### REFERENCES

- [1] A. G. M. Ahmed, J. Guo, D. M. Yan, J. Y. Franceschia, X. Zhang, and O. Deussen. A simple push-pull algorithm for blue-noise sampling. *IEEE Transactions on Visualization & Computer Graphics*, 23(12):2496–2508, 2017.

- [2] G. Andrienko, N. Andrienko, G. Fuchs, and J. M. C. Garcia. Clustering trajectories by relevant parts for air traffic analysis. *IEEE Transactions on Visualization & Computer Graphics*, 24(1):34–44, 2018.
- [3] 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 & Computer Graphics*, 23(9):2120–2136, 2017.
- [4] N. V. Andrienko and G. L. Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.
- [5] M. Berger, K. McDonough, and L. M. Seversky. cite2vec: Citation-driven document exploration via word embeddings. *IEEE Transactions on Visualization & Computer Graphics*, 23(1):691–700, 2017.
- [6] R. Blanco, G. Ottaviano, and E. Meij. Fast and space-efficient entity linking for queries. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pp. 179–188. ACM, New York, USA, 2015.
- [7] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, (10):155–168, 2008.
- [8] I. Boyandin, E. Bertini, and D. Lalanne. Visualizing the worlds refugee data with jflowmap. In *Proceedings of the Eurographics/ IEEE-VGTC Symposium on Visualization*, 2010.
- [9] K. Buchin, B. Speckmann, and K. Verbeek. Flow map layout via spiral trees. *IEEE Transactions on Visualization & Computer Graphics*, 17(12):2536–2544, 2011.
- [10] H. Chen, W. Chen, H. Mei, Z. Liu, K. Zhou, W. Chen, W. Gu, and K. L. Ma. Visual abstraction and exploration of multi-class scatterplots. *IEEE Transactions on Visualization & Computer Graphics*, 20(12):1683–1692, 2014.
- [11] W. Cui, H. Zhou, H. Qu, P. C. Wong, and X. Li. Geometry-based edge clustering for graph visualization. *IEEE Transactions on Visualization & Computer Graphics*, 14(6):1277–1284, 2008.
- [12] M. Ebeida, S. Mitchell, M. Awad, C. Park, L. P. Swiler, D. Manocha, and L. Wei. Spoke darts for efficient high dimensional blue noise sampling. *ACM Transactions on Graphics*, 37(2), 2018.
- [13] O. Ersoy, C. Hurter, F. Paulovich, G. Cantareiro, and A. Telea. Skeleton-based edge bundling for graph visualization. *IEEE Transactions on Visualization & Computer Graphics*, 17(12):2364–2373, 2011.
- [14] 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. *IEEE Transactions on Visualization & Computer Graphics*, 19(12):2149–2158, 2013.
- [15] S. Gao, Y. Liu, Y. Wang, and X. Ma. Discovering spatial interaction communities from mobile phone data. *Transactions in Gis*, 17(3):463–481, 2013.
- [16] A. Grover and J. Leskovec. Node2vec: Scalable feature learning for networks. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864. ACM, New York, NY, USA, 2016.
- [17] D. Guo. Visual analytics of spatial interaction patterns for pandemic decision support. *International Journal of Geographical Information Science*, 21(8):859–877, 2007.
- [18] D. Guo. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization & Computer Graphics*, 15(6):1041–1048, 2009.
- [19] D. Guo and X. Zhu. Origin-destination flow data smoothing and mapping. *IEEE Transactions on Visualization & Computer Graphics*, 20(12):2043–2052, 2014.
- [20] D. Heck, T. Schlömer, and O. Deussen. Blue noise sampling with controlled aliasing. *ACM Transactions on Graphics*, 32(3):1–12, 2013.
- [21] D. Holten, P. Isenberg, J. J. van Wijk, and J. D. Fekete. An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *Proceedings of the 2011 IEEE Pacific Visualization Symposium*, pp. 195–202. IEEE Computer Society, Washington, DC, USA, 2011.
- [22] A. H. Robinson. The 1837 maps of henry drury harness. *Geographical Journal*, 121(4):440–450, 1955.
- [23] H. Johnson, E. S. N. Johnson, and E. S. Nelson. Using flow maps to visualize time-series data: Comparing the effectiveness of a paper map series, a computer map series, and animation. *Cartographic Perspectives*, (30):47–64, 1998.
- [24] B. C. Kwon, J. Verma, P. J. Haas, and C. Demiralp. Sampling for scalable visual analytics. *IEEE Computer Graphics & Applications*, 37(1):100–108, 2017.
- [25] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu. Smartradp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Transactions on Visualization & Computer Graphics*, 23(1):1–10, 2017.
- [26] K. Liu, S. Gao, P. Qiu, X. Liu, B. Yan, and F. Lu. Road2vec: Measuring traffic interactions in urban road system from massive travel routes. *ISPRS International Journal of Geo-Information*, 6(11), 2017.
- [27] L. Liu, A. P. Boone, I. T. Ruginski, L. Padilla, M. Hegarty, S. H. Creem-Regehr, W. B. Thompson, C. Yuksel, and D. H. House. Uncertainty visualization by representative sampling from prediction ensembles. *IEEE Transactions on Visualization & Computer Graphics*, 23(9):2165–2178, 2017.
- [28] S. Liu, W. Cui, Y. Wu, and M. Liu. A survey on information visualization: Recent advances and challenges. *The Visual Computer*, 30(12):1373–1393, 2014.
- [29] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu. Storyflow: Tracking the evolution of stories. *IEEE Transactions on Visualization & Computer Graphics*, 19(12):2436–2445, 2013.
- [30] L. V. D. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2605):2579–2605, 2008.
- [31] D. F. Marble, Z. Gou, L. Liu, and J. Saunders. *Recent Advances in the Exploratory Analysis of Interregional Flows in Space and Time*. 1997.
- [32] T. Mikolov, Q. V. Le, and I. Sutskever. Exploiting similarities among languages for machine translation. *Computer Science*, 2013.
- [33] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pp. 3111–3119. Curran Associates Inc., USA, 2013.
- [34] D. Peng, N. Lu, W. Chen, and Q. Peng. Sideknob: Revealing relation patterns for graph visualization. In *Proceedings of the 2012 IEEE Pacific Visualization Symposium*, pp. 65–72, 2012.
- [35] D. Phan, L. Xiao, R. Yeh, and P. Hanrahan. Flow map layout. In *Proceedings of the IEEE Symposium on Information Visualization*, pp. 219–224, Oct 2005.
- [36] H. Qin, Y. Chen, J. He, and B. Chen. Wasserstein blue noise sampling. *ACM Transactions on Graphics*, 36(5):1–13, 2017.
- [37] A. Rae. From spatial interaction data to spatial interaction information? geovisualisation and spatial structures of migration from the 2001 uk census. *Computers Environment and Urban Systems*, 33(3):161–178, 2009.
- [38] A. H. Robinson. The thematic maps of charles joseph minard. *Imago Mundi*, 21(1):95–108, 1967.
- [39] Z. Su, H. Xu, D. Zhang, and Y. Xu. Chinese sentiment classification using a neural network toolword2vec. In *Proceedings of the 2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI)*, pp. 1–6, Sept 2014.
- [40] X. Sun, K. Zhou, J. Guo, G. Xie, J. Pan, W. Wang, and B. Guo. Line segment sampling with blue-noise properties. *ACM Transactions on Graphics*, 32(4):1–14, 2013.
- [41] W. Tobler. Depicting federal fiscal transfers. *The Professional Geographer*, 33(4):419–422, 1981.
- [42] W. Tobler. Experiments in migration mapping by computer. *Cartography and Geographic Information Science*, 14(2):155–163, 1987.
- [43] J. Turian, L. Ratinov, and Y. Bengio. Word representations: A simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 384–394. Association for Computational Linguistics, Stroudsburg, PA, USA, 2010.
- [44] Y. Wang, Q. Shen, D. Archambault, Z. Zhou, M. Zhu, S. Yang, and H. Qu. Ambiguityvis: Visualization of ambiguity in graph layouts. *IEEE Transactions on Visualization & Computer Graphics*, 22(1):359–68, 2016.
- [45] Z. Wang, T. Ye, M. Lu, X. Yuan, H. Qu, J. Yuan, and Q. Wu. Visual exploration of sparse traffic trajectory data. *IEEE Transactions on Visualization & Computer Graphics*, 20(12):1813–1822, 2014.
- [46] M. Wattenberg, F. Vigas, and I. Johnson. How to use t-sne effectively. *Distill*, 2016.
- [47] J. Wood, J. Dykes, and A. Slingsby. Visualisation of origins, destinations and flows with od maps. *The Cartographic Journal*, 47(2):117–129, 2010.
- [48] J. Wood, A. Slingsby, and J. Dykes. Visualizing the dynamics of london's bicycle-hire scheme. *Cartographica the International Journal for Geographic Information & Geovisualization*, 46(4):239–251, 2011.

- [49] D.-M. Yan and P. Wonka. Gap processing for adaptive maximal poisson-disk sampling. *ACM Transactions on Graphics*, 32(5):1–15, 2013.
- [50] Y. Yang, T. Dwyer, S. Goodwin, and K. Marriott. Many-to-many geographically-embedded flow visualisation: An evaluation. *IEEE Transactions on Visualization & Computer Graphics*, 23(1):411–420, 2017.
- [51] X. Zhu and D. Guo. Mapping large spatial flow data with hierarchical clustering. *Transactions in Gis*, 18(3):421–435, 2014.