

# Joint Mobility Pattern Mining with Urban Region Partitions

Jing Lian  
Tsinghua-Berkeley Shenzhen  
Institute, Tsinghua University  
lianj16@mails.tsinghua.edu.cn

Yang Li  
Tsinghua-Berkeley Shenzhen  
Institute, Tsinghua University  
yangli@sz.tsinghua.edu.cn

Weixi Gu  
University of California, Berkeley  
guweixigavin@gmail.com

Shao-Lun Huang  
Tsinghua-Berkeley Shenzhen  
Institute, Tsinghua University  
shaolun.huang@sz.tsinghua.edu.cn

Lin Zhang  
Tsinghua-Berkeley Shenzhen  
Institute, Tsinghua University  
linzhang@tsinghua.edu.cn

## ABSTRACT

Mobility pattern mining answers the fundamental question of where people are likely to go from a given location. It plays an important role in city planning, public transport management and location-based mobile applications. Among these applications, many concern the mobility pattern over contiguous spatial regions as a whole. Traditional ways of mobility pattern mining either result in trip clusters with overlapped origin and destination regions, or require an extra step to partition the city into discrete regions, which may not be optimal for mobility patterns extraction. In this paper, we present a region-aware mobility pattern mining framework to jointly extract trip clusters while maintain non-overlapping partitions of trip origins and destinations. We developed *kernelized ACE*, a novel extension to a classic algorithm in statistics to compute the optimal mobility clusters under spatial constraints. Experimental results using Beijing taxi trip data show that our approach outperforms other methods with only  $\sim 0.3\%$  spatial overlap and 86.43% origin-destination correlation. Our case studies on New York City's and Beijing's taxi datasets also yield insightful findings that reveal city-scale mobility patterns and propose potential improvement for public transportation.

## CCS CONCEPTS

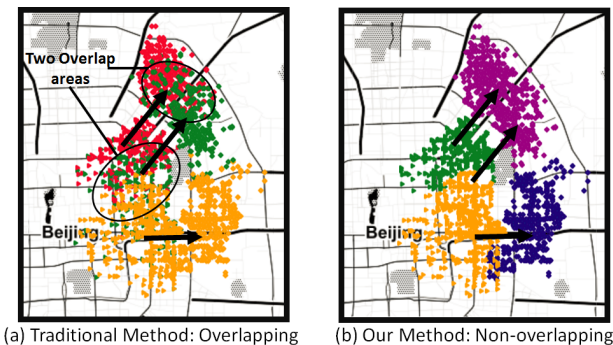
• **Information systems**  $\rightarrow$  *Clustering*; • **Applied computing**  $\rightarrow$  *Transportation*;

## KEYWORDS

mobility pattern, region partition, co-clustering

## 1 INTRODUCTION

Understanding urban mobility is an important socio-economic challenge. It plays a critical role in city planning [18], public transport management [12, 13, 19], as well as many mobile services emerged recently, such as location-based advertisement [1, 14]. In essence, it aims to answer the fundamental question: where people are most



**Figure 1: An example of mobility patterns in OD trip data using a traditional method and our method. In a), mobility patterns are represented by three trip clusters in different colors, which contain significant spatial overlap. Arrows show the direction of trips in each cluster. In b), the same OD trips are assigned to two origin regions (green, yellow) and two destination regions (purple, blue) without overlap. OD patterns are represented by arrows between the regions.**

likely to go from one place to another. With the advent of ubiquitous sensing technology, we are engulfed with massive amount of location tracking data from various sources such as cellular network data [3], geo-tagged social media data [25], and vehicle GPS data [11]. However, it is still a significant challenge to disentangle such noisy, high dimensional data into useful information, even for trip data that only contain Origin-Destination (OD) locations. It is therefore necessary to extract such high-level, well-structured information, a.k.a. mobility patterns, from the uncertain trip data, such that it can be easily interpreted by human or machines to make better decisions for further applications.

Traditional ways of mobility pattern extraction are based on clustering OD trips as points in high dimensional space [2, 33]. This approach has a limitation that, when clusters are viewed by the spatial distribution of their origins and destinations, regions corresponding to different clusters tend to overlap (see Figure 1a). Many city planning applications focus on how people migrate in certain regions [29]. For trips originating or arriving at the overlapping areas, ambiguities of which region they belong to would be an issue. Some methods address this overlap problem by a two-stage process. They first partition the continuous space of OD locations into a finite discrete set of non-overlapping regions (e.g. grid cells [6, 16, 31],

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MobiQuitous'2018, November 2018, New York City, United States

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

or point-of-interests (POI) [10, 15]); then extract mobility patterns among those regions. However, not only does aggregating trips in each region lose fine-grained information, more importantly, there is no guarantee that the partition would be optimal for mobility pattern extraction.

To tackle these problems, we propose the **region-aware mobility pattern mining** framework, which jointly finds the optimal origin and destination region partitions, while extracting mobility patterns at the same time. Informally, we formulate the problem as learning feature representations of trip origins and destinations that guarantee both objectives. For the clustering of trips, we adopt the *HGR maximal correlation* [24], a well-defined dependence measure to represent the shared information between trip origins and destinations. Meanwhile, we add spatial constraints to minimize the amount of overlap between origin and destination regions of different clusters. Using the optimal OD features, we are able to partition the city into non-overlapping origin and destination regions, which also provide the optimal solution for mobility patterns extraction, as shown in Figure 1b.

To solve this optimization problem efficiently, we propose the *Kernelized Alternating Conditional Expectation (KACE) algorithm* based on a statistical technique. The original ACE algorithm [21] is an efficient way to compute HGR maximal correlation for discrete random variables. However, it does not impose any constraints on the extracted features, such as the spatial constraints of OD clusters in our case. Moreover, it cannot be used on continuous data. Therefore, we extended ACE algorithm with an optimal kernel to handle continuous data with spatial constraints.

We evaluated our algorithm on taxi trip data, one of the most popular representative form of mobility data [32]. On a taxi dataset from Beijing, China, we obtained mobility regions with only 0.33% and 0.22% overlaps for origins and destinations respectively but achieved 86.43% OD correlation. Compared with other methods, our method achieves minimum overlaps, concentrated regions, highly correlated patterns and full spatial coverage. We also found insightful findings from case studies of both New York City and Beijing. For instance, our analysis discovered a demand of metro service in Jackson Heights area in New York City. In Beijing's case, we found very low mobility interactions of suburban districts with other regions, which could reveal potential social problems.

We summarize the contributions of our paper below:

- (1) A novel region-aware mobility pattern mining framework. It ensures optimality for both mobility pattern extraction and OD region partition without overlap.
- (2) A kernel-based extension to the ACE algorithm for extracting maximal correlation features. KACE is the first algorithm to solve the HGR maximal correlation problem with continuous input and feature constraints.
- (3) A thorough evaluation of KACE and case studies of the evening commute patterns in Beijing and New York City with real taxi data. Comparisons with both traditional methods and state-of-the-art clustering algorithms show the versatility of our approach in disentangling complicated mobility patterns. We achieved 86.43% OD correlation with only  $\sim 0.3\%$  overlap and  $\sim 3km$  average in-cluster distance, which

reduces overlap by up to ten times comparing to other approaches.

The remaining parts of this paper are organized as follows. Section 2 introduces related work, Section 3 formulates the problem mathematically, Section 4 illustrates the proposed algorithm kernelized ACE, Section 5 tests the algorithm on New York City and Beijing taxi data, and Section 6 points out the findings of Beijing and NYC, and the paper concludes in Section 7.

## 2 RELATED WORK

Several areas of studies are related to our work, including mobility pattern mining, region partition using GPS data and co-clustering methods.

**Extracting mobility OD patterns.** One type of approaches clustered OD trips to get mobility patterns. Zhu et al. modified density based clustering algorithm DBSCAN in mobility mining setup to extract popular OD pairs [33]. However, it leads to overlap between origin clusters or destination clusters. Moreover, it can only find salient and precise patterns, which does not cover the whole city. Another type of work adopted a two-stage process by discretizing a city into non-overlapping sets first and extracting OD patterns on that set. Gambs et al. clustered adjacent areas in terms of POI information as home, work and others before applying a modified Markov Model of predicting people's next location [10]. Kang et al. analyzed Beijing's taxi data based on traffic analysis zones (TAZ) and found the inter-TAZ network has a gravitational structure [15]. Tang et al. used DBSCAN to cluster taxi pick-ups and drop-offs first and then used discretized OD clusters to calibrate a statistic mobility model [27]. Cao et al. proposed a grid-based hierarchical clustering algorithm to discover frequent spatial patterns [6]. All of these work extracted mobility findings on discrete set of regions, which are POI regions [10, 15], O/D clusters [27] or uniform grids [6, 31] rather than original continuous data, which limit the size, shape of the clustering results and have no guarantee of the optimal discretization and findings' granularity.

**Region partition using GPS data.** Studies of partition city into non-overlapping regions based on people's mobility or region functions have similarities to our work. Zheng et al. detected flawed urban planning areas using taxi trajectories [32]. Yuan et al. discovered regions of different functions using points of interests and mobility information [30]. Both of these work partitioned the urban area boundaries based on existing road network, while our method establishes regions based on mobility patterns learned from data. Liu et al. revealed sub-regions of Shanghai by identifying sub-networks of taxi trip network [18]. The regions are densely intra-connected but with less inter-connections, thus the identified regions cannot fully reveal the spatial dynamics of a city.

**Co-clustering methods.** Co-clustering is the problem of simultaneously clustering two types of correlated data, which has been applied to gene expression [20], text mining [7] and image analysis [17]. Fewer work has applied co-clustering to mobility pattern mining. Kuo et al. used non-negative tensor factorization (NTF) based co-clustering algorithm to establish an OD flow [16]. However, it still has the discretization issue that binning GPS samples into grids before applying NTF. We adopted feature-based co-clustering algorithm without discretization to overcome information loss of

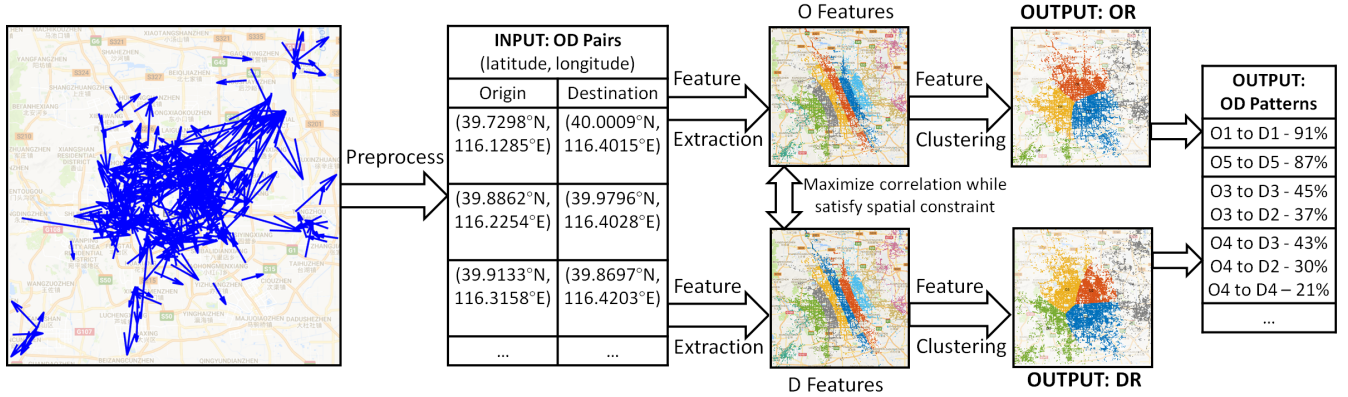


Figure 2: Flow chart of our algorithm

discretization. A classical feature-based co-clustering algorithm is to extract features using kernel canonical correlation analysis (KCCA) and to cluster them by K-means [4]. Our approach extracted features by ACE algorithm [21], which extracts non-linear features like KCCA but without the need of specifying a kernel. Nie et al. proposed a state-of-art multi-view clustering algorithm MLAN [23], a generalization of co-clustering, which extracted an optimal distance metric from data itself and co-clustered data under the optimal metric. However, it only focused on achieving maximum correlation without trade-offs on cluster size thus their metric led to big clusters as shown in our experiments. Our method, on the contrary, has trade-offs between OD correlations and region sizes thus results to small meaningful regions with good correlations.

### 3 PROBLEM FORMULATION

The key step in combining mobility region partition and mobility pattern mining into one problem is to maximize OD correlation with spatial constraints. According to Rényi [24], Hirschfeld-Gebelein-Rényi maximal correlation is the most optimal correlation measure in statistics. It quantizes correlations of random variables to a measure between 0 and 1, and it is comparable regardless of different data size. The formal definition of HGR maximal correlation is shown below.

**DEFINITION 1.** *Hirschfeld-Gebelein-Rényi Maximal Correlation.* Maximal correlation between jointly distributed random variables  $X$  and  $Y$  is defined as:

$$\rho(X; Y) \triangleq \sup_{\substack{f: X \rightarrow \mathbb{R}, g: Y \rightarrow \mathbb{R} \\ \mathbb{E}[f(X)] = \mathbb{E}[g(Y)] = 0 \\ \mathbb{E}[f^2(X)] = \mathbb{E}[g^2(Y)] = 1}} \mathbb{E}[f(X)g(Y)]$$

where the supremum is taken over all Borel measurable functions. Furthermore,  $0 \leq \rho(X; Y) \leq 1$ ,  $\rho(X; Y) = 0$  if and only if  $X$  is independent of  $Y$ ,  $\rho(X; Y) = 1$  if there is a strict dependence between  $X$  and  $Y$ , i.e.  $Y = f(X)$  or  $X = g(Y)$ .

In our mobility pattern mining setup, we want to find good features  $f$  and  $g$  of origins and destinations respectively for each trip according to the definition. The maximal correlation of 0 happens only when resulted origin and destination regions are completely

independent which means origins contain no information of destinations and vice versa. While the maximal correlation of 1 indicates the following three fully correlated scenarios.

- From one origin region to one destination region
- From many origin regions to one destination region
- From one origin region to many destination regions

Formally, we define the problem as follows:

#### Region-Aware Mobility Pattern Mining Problem.

Given  $N$  trips represented by OD pairs  $T = \{(x_i, y_i), x_i = (xlat_i, xlon_i), y_i = (ylat_i, ylon_i)\}^1, i = 1, \dots, N\}$  as input, we would like to partition  $N$  origins  $X$  into  $N_x$  origin regions and partition destinations  $Y$  into  $N_y$  destination regions, i.e. each origin point  $x_i$  has an origin region label  $OR(x_i)$  and destination point  $y_i$  with destination region label  $DR(y_i)$  as output, such that

1) The HGR maximal correlation between  $OR$  and  $DR$  is maximized to obtain the optimal mobility pattern, i.e. the correlation between  $f(X)$  and  $g(Y)$  is maximized for  $X$  and  $Y$

$$\max_{\substack{\mathbb{E}[f(X)] = \mathbb{E}[g(Y)] = 0 \\ \mathbb{E}[f^2(X)] = \mathbb{E}[g^2(Y)] = 1}} \mathbb{E}[f(X)g(Y)]$$

2) Points inside a region are close to each other to ensure non-overlapping characteristic. When origins  $x_i$  and  $x_j$  are close to each other geographically, their corresponding features  $f(x_i)$  and  $f(x_j)$  should be similar in value. The same applies to the destination points, i.e.

$$|f(x_i) - f(x_j)| \text{ increases as } \|x_i - x_j\|_2 \text{ increases}$$

$$|g(y_i) - g(y_j)| \text{ increases as } \|y_i - y_j\|_2 \text{ increases}$$

Here,  $f(x_i)$  is the feature value of the  $i$ -th origin point  $x_i$ , likewise  $g(y_i)$  is the feature value of  $i$ -th destination  $y_i$ .

Based on proximity of features  $f(X)$  and  $g(Y)$ , we use average linkage cluster [26], commonly used in geographic areas, to find  $N_x$  ORs,  $N_y$  DRs and mobility patterns between origin and destination regions.

<sup>1</sup>Here  $x_i$  represents trip  $i$ 's origin location, including its latitude  $xlat_i$  and longitude  $xlon_i$ ;  $y_i$  represents trip  $i$ 's destination location, including its latitude  $ylat_i$  and longitude  $ylon_i$ .

## 4 ALGORITHM

### 4.1 Overview

The flow chart of our algorithm is shown in Figure 2. First, we pre-process trajectory data into OD pairs of latitudes and longitudes as input of our algorithm. Then, we extract features of origins and destinations considering both OD correlation and spatial constraints of regions formulated in Section 3. Next, we obtain non-overlapping O/D regions by clustering the origin/destination features separately. As features already contain information of OD correlations, clustering them separately still preserves mobility patterns in the OD data. Finally, we compute the region-aware mobility patterns based on the number of trips between each origin regions (OR) and destination regions (DR). The outputs of our algorithm are origin and destination regions as well as OD patterns from OR to DR, *i.e.* OR and DR label of each O/D point, and patterns with a probability indicating the percentage of trips originating from OR that end at DR.

### 4.2 Kernelized ACE

This section illustrates how to extract features that maximize OD maximal correlations under spatial constraints. According to [21], an efficient algorithm, Alternating Conditional Expectation (ACE), can be used to solve HGR maximal correlation problem and obtain the feature functions  $f(\cdot)$  and  $g(\cdot)$ .

However, this algorithm cannot be applied to our problem directly due to two issues. First, ACE only works with discrete-valued

data rather than continuous-valued data. For example, in the Netflix Prize problem [22] of predicting user ratings for films, the dataset consists of discrete-valued movie IDs and user IDs; while data we care about are continuous GPS coordinates. Second, ACE cannot place additional constraints on features. In the Netflix problem, the movie-user connection is hidden in the data. The similarities between movies are reflected by the ratings of different users but there is no obvious connections between movies themselves. On the contrary, in our problem, not only the correlations between origins( $X$ ) and destinations( $Y$ ) should be considered, the distances between origin points or destination points should also contribute to finding geographically meaningful regions.

The essential reason that ACE cannot deal with continuous-valued data directly is that empirical conditional distributions,  $\hat{P}(Y|X = x)$ , are needed to compute empirical conditional expectations. However, for continuous-valued data, the exact empirical conditional distribution at a specific point is impossible to obtain due to limited data samples and infinite large sample space. Transferring from continuous to discrete data by grouping points together into grids is not an optimal solution because grouping loses information and the resulted ORs and DRs cannot have smooth outlines and arbitrary shapes. Therefore, we need to estimate empirical conditional expectations directly from data samples to skip computing the conditional distributions.

For discrete random variables, ACE expresses the empirical conditional expectations of  $Y$  given  $X$  over  $N$  data samples as Equation (1).

$$\hat{\mathbb{E}}_n[g(Y)|X = x] = \frac{\sum_{j=1}^N g(y_j) \mathbb{1}(x_j = x)}{\sum_{j=1}^N \mathbb{1}(x_j = x)} \quad (1)$$

where  $\mathbb{1}(\cdot)$  is an indicator function.

$$\mathbb{1}(x_j = x) = \begin{cases} 1 & x_j = x \\ 0 & x_j \neq x \end{cases}$$

In the continuous-valued data scenario,  $\sum_{j=1}^N \mathbb{1}(x_j = x)$  can be very small or only 1 so that the expectation loses information contained in the neighborhood of  $x$ . Therefore, we need to apply smoothing techniques, which use values of points nearby for expectation estimation [5]. Intuitively, we replace the indicator function in the numerator of Equation (1) by a kernel  $K(x_j, x)$  and set the denominator to number of samples  $N$  to avoid denominator being 0 shown as Equation (2). It is used to replace Equation (1) in ACE as step 2a in Algorithm 1.

$$\hat{\mathbb{E}}_n[g(Y)|X = x] \leftarrow \frac{\sum_{j=1}^N g(y_j) K(x_j, x)}{N} \quad (2)$$

However, mobility patterns are complicated and extracting only one dimension of feature is not adequate. Therefore, we further extract  $D$ -dimensional features in Algorithm 1. The key idea is to obtain features that are orthogonal to the previously obtained features to avoid redundant information [21]. Therefore, we applied Gram-Schmidt orthogonalization to the feature functions in step 2c.

<sup>1</sup>Average linkage cluster [26], a classical hierarchical clustering method, which is commonly used in geographic areas.

---

**Algorithm 1** D-dimensional kernelized ACE for clustering

---

**Require:** training samples  $\{(x_i, y_i) : i = 1, \dots, N\}$

1. Initialize: randomly generate and regularize  $f_d(x_i), g_d(y_i), i = 1, \dots, N, d = 1, \dots, D$

**repeat**

2a. **Feature iteration:**

$$f_d(x_i) \leftarrow \frac{\sum_{j=1}^N g_d(y_j) K(x_j, x_i)}{N}, K(x_j, x_i) = 1 - \|x_j - x_i\|_2$$

$$g_d(y_i) \leftarrow \frac{\sum_{j=1}^N f_d(x_j) K(y_j, y_i)}{N}, K(y_j, y_i) = 1 - \|y_j - y_i\|_2$$

2b. **Regularize:**  $f_d(\cdot), g_d(\cdot), d = 1, \dots, D$ .

$$f_d(x_i) \leftarrow f_d(x_i) - \frac{\sum_i f_d(x_i)}{N}, f_d(x_i) \leftarrow \frac{f_d(x_i)}{\sqrt{\frac{\sum_i f_d(x_i)^2}{N}}}$$

$$g_d(y_i) \leftarrow g_d(y_i) - \frac{\sum_i g_d(y_i)}{N}, g_d(y_i) \leftarrow \frac{g_d(y_i)}{\sqrt{\frac{\sum_i g_d(y_i)^2}{N}}}$$

2c. **Gram-Schmidt:**

**for**  $d = 1$  to  $D$  **do**

**for**  $k = 1$  to  $d - 1$  **do**

$$f_d(x) \leftarrow f_d(x) - \frac{\langle f_k(x), f_d(x) \rangle}{\langle f_k(x), f_k(x) \rangle} f_k(x)$$

$$g_d(x) \leftarrow g_d(x) - \frac{\langle g_k(x), g_d(x) \rangle}{\langle g_k(x), g_k(x) \rangle} g_k(x)$$

**end for**

**end for**

**until**  $\mathbb{E}[f_{1,2,\dots,D}(x)^T g_{1,2,\dots,D}(y)]$  stops to increase

3. **Output:** Region label of each point  $x_i$  and  $y_i$

3a.  $OR(x_i) \leftarrow$  Average linkage cluster<sup>1</sup> of  $f_1, \dots, D(x_i), i = 1, \dots, N$  with maximum cluster number  $N_x$ ,

3b.  $DR(y_i) \leftarrow$  Average linkage cluster of  $g_1, \dots, D(y_i), i = 1, \dots, N$  with maximum cluster number  $N_y$

---

### 4.3 Kernel Selection

The challenge of kernelized ACE is to choose the most suitable kernel function.

Recall that spatial constraints in our problem states that features should be close to each other when data points are close. Therefore, features of points close to the specific point of interest are more important than features of points far away. In Equation (2), kernel  $K(x_j, x)$  serves as feature weights, which should satisfy the following two properties.

- (1)  $K(x_j, x)$  is a non-increasing function of  $\|x_j - x\|_2$
- (2)  $K(x_j, x) = 1$  when  $\|x_j - x\|_2 = 0$

Property (2) comes from the meaning of weights, where  $K(x_j, x) = 1$  means feature  $g(y_j)$  of trip  $(x_j, y_j)$  is fully considered while  $K(x_j, x) = 0$  means  $g(y_j)$  is not considered at all. Property (1) satisfies  $|f(x_i) - f(x_j)|$  increases as  $\|x_i - x_j\|_2$  increases.

We list three kernels that satisfy these two properties, which are

- (1) window kernel,  $K_1(x_j, x) = \mathbb{1}(\|x_j - x\|_2 \leq \gamma)$ , parameterized by  $\gamma (\gamma > 0)$ ;
- (2) Gaussian kernel,  $K_2(x_j, x) = e^{-\frac{\|x_j - x\|_2^2}{2\sigma^2}}$ , parameterized by  $\sigma (\sigma > 0)$ ;
- (3) negative linear kernel,  $K_3(x_j, x) = 1 - a\|x_j - x\|_2$ , parameterized by  $a (a > 0)$ .

Here,  $K_3(x_j, x) \in (-\infty, 1]$  is different from the other two kernels  $K_1, K_2 \in (0, 1]$ , since features of points  $x_j$  far away from the point of interest  $x$  have negative effect rather than no effect on  $x$ 's features.

The features  $f$  and  $g$  are not only influenced by different kernels, but also parameter choices. However, traversing every possible parameters and kernels satisfying previous two properties is time-consuming and almost impossible. However, we proved that parameters of negative linear kernel don't affect correlation's results as shown in Theorem 1. Therefore, we chose negative linear kernel,  $K_3(x_j, x) = 1 - \|x_j - x\|_2$ , to be used in Algorithm 1.

**THEOREM 1.** *Different intercepts or slopes for negative linear kernel do not change the results of features, i.e.  $K_1(x_j, x) = 1 - \|x_j - x\|_2$ ,  $K_2(x_j, x) = b - \|x_j - x\|_2$  and  $K_3(x_j, x) = 1 - a\|x_j - x\|_2 (a > 0)$  produce the same features  $f$  and  $g$ .*

**PROOF.** Denote  $\vec{d}_i = [\|x_1 - x_i\|_2, \|x_2 - x_i\|_2, \dots, \|x_N - x_i\|_2]^T$ ,  $\vec{g} = [g(y_1), g(y_2), \dots, g(y_N)]^T$ .

1) Feature invariance over intercepts.

For  $K_1$ , we have

$$\begin{aligned} f_1(x_i) &= \frac{\sum_{j=1}^N g(y_j) K_1(x_j, x_i)}{N} = \frac{\sum_{j=1}^N g(y_j) (1 - \|x_j - x_i\|_2)}{N} \\ &= \frac{0 - \sum_{j=1}^N g(y_j) \|x_j - x_i\|_2}{N} = -\frac{\langle \vec{g}, \vec{d}_i \rangle}{N} \end{aligned}$$

The third equality holds because,

$$\hat{\mathbb{E}}_n[g(Y)] = \frac{\sum_{j=1}^N g(y_j)}{N} = 0$$

For  $K_2$ , we have

$$\begin{aligned} f_2(x_i) &= \frac{\sum_{j=1}^N g(y_j) K_2(x_j, x_i)}{N} = \frac{\sum_{j=1}^N g(y_j) (b - \|x_j - x_i\|_2)}{N} \\ &= \frac{0 - \sum_{j=1}^N g(y_j) \|x_j - x_i\|_2}{N} = -\frac{\langle \vec{g}, \vec{d}_i \rangle}{N} = f_1(x_i), \end{aligned}$$

which implies  $f_1(X) = f_2(X)$ .

2) Feature invariance over slopes.

For  $K_3$ , we have

$$\begin{aligned} f_3(x_i) &= \frac{\sum_{j=1}^N g(y_j) K_3(x_j, x_i)}{N} = \frac{\sum_{j=1}^N g(y_j) (1 - a\|x_j - x_i\|_2)}{N} \\ &= \frac{0 - a \sum_{j=1}^N g(y_j) \|x_j - x_i\|_2}{N} = -a \frac{\langle \vec{g}, \vec{d}_i \rangle}{N} = a \cdot f_1(x_i), \end{aligned}$$

After regularizations (step 2b), we have  $f_1(X) = f_3(X)$ .

Combining these two scenarios, we have  $f_1(X) = f_2(X) = f_3(X)$ . Likewise,  $g(Y)$  also satisfies this theorem.  $\square$

## 5 EXPERIMENTS

### 5.1 Data Description

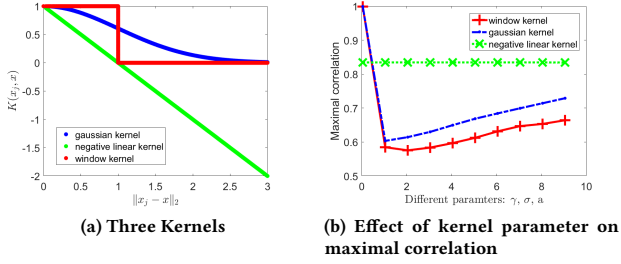
We tested Algorithm 1 on two taxi datasets of Beijing and New York City. Since mobility patterns depend on time-of-day, day-of-week and other temporal factors, one month of weekdays' data in November, 2015 during the morning and evening commute hour 7:00-7:59/17:00-17:59 are considered.

For New York City, we used open source dataset [28] published by the NYC Taxi and Limousine Commission. The data only contains OD pair information with time stamps without intermediate points. We combined trips of both yellow taxis, serving Manhattan exclusionary zone, and green taxis, serving borough areas, to evaluate taxi trips covering the whole city. Besides, only trips longer than 2km were considered for the New York data.

For the Beijing taxi data, we extracted occupied trip information from the raw GPS trajectories of 20,067 taxis with sample rate at around 1 minute. Each sample contains a unique taxi number, time stamp, latitude, longitude, azimuth, speed and occupancy indicator (occupied 1, vacant 0 or stopped -1). We determined pick-up happens when the occupancy indicator turns from 0 to 1 and drop-off happens when occupancy turns from 1 to 0. In addition, we omitted short trips to eliminate incorrectly recorded data. Only trips longer than 3km (measured by OD distance) and lasting more than 1min were considered. Beijing's filtered OD distance is longer than New York's because Beijing has longer average OD distance. More details of the taxi datasets are shown in Table 1.

**Table 1**  
Summary of Beijing and New York City Taxi Dataset

		Beijing	NYC
17:00-17:59	Total Trip Number	118433	213175
	Average OD Distance (km)	3.63	2.95
	OD Filtered Trip Number	54199	127648
7:00-7:59	Total Trip Number	116817	208336
	Average OD Distance (km)	4.71	3.38
	OD Filtered Trip Number	65330	137140



**Figure 3: Three kernels and the effect of window kernel's  $\gamma$ , gaussian kernel's  $\sigma$  and negative linear kernel's  $a$  on resulted maximal correlation**

## 5.2 Kernel Evaluation

We tested three different kernel's performance on a small subset (1000 trips) of Beijing's evening data. Different parameters' effect on the maximal correlation is drawn in Figure 3b. When parameters are small ( $\gamma, \theta = 0.05$ ), the correlations are large. Because if each O/D point forms a cluster, one-to-one dependence is achieved and the correlation is 1 at this extreme scenario. The correlation decreases and then increases for both the window kernel and the Gaussian kernel. With the increasing of  $\gamma$  and  $\theta$ , the radius of O/D region increases. For the extreme scenario that the region is the whole city, all points form a single cluster and the correlation is one. However, negative linear kernel stays at its maximal correlation at 0.8344 regardless of the slope  $a$ . This result verifies Theorem 1 by real experiment - the slope and intercept of negative linear kernel do not change extracted features. Therefore, negative linear kernel is a good choice for identifying the existence of significant mobility patterns for its stability and near-optimal performance. It searches in linear directions for regions while the Gaussian and window kernels search in spherical directions and the results depend largely on the sphere radius.

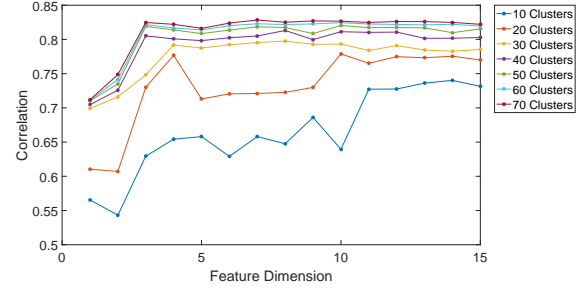
## 5.3 Model Analysis.

One challenge in finding the best global mobility patterns using the kernelized ACE is to choose the optimal region number as well as the optimal dimension of features. In this section, we explore the impact of these parameters on regions' correlations.

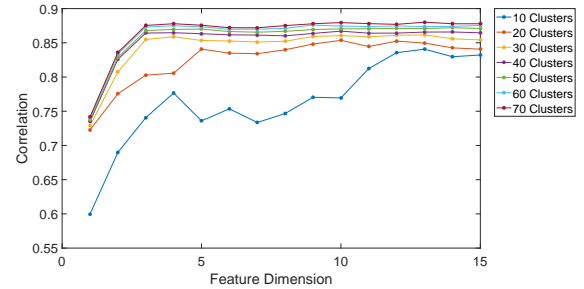
We studied how OD region correlations change with the increase of feature dimensions for different cluster numbers and showed results in Figure 4. We extracted origin and destination regions using the first  $D$  ( $1 \leq D \leq 20$ ) dimensions of KACE feature and set region number from 10 to 70. We first showed Beijing's results in Figure 4a and 4b. The correlation under different number of clusters generally increases with more feature added in, but has a turning point at minimum optimal feature dimension. Specifically, the correlation raises significantly for the first three dimension of features and reach a stable stage (when cluster number is more than 30). Therefore, the minimum optimal feature dimension for obtaining mobility patterns in Beijing should be three, regardless of morning or evening. The results of New York's data show similar findings in Figure 4c and 4d. The difference is that the minimum optimal feature dimension is two for NYC. Therefore, in our case

studies of Beijing and NYC, we used three and two dimension of features for each city respectively.

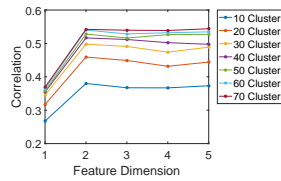
With number of clusters increasing, the regional correlation increases gradually but the increasing step decreases for both Beijing and NYC. It tells us that separating for more clusters benefits only a little for regional correlations. Therefore, we chose cluster number of 40 for both Beijing and NYC to achieve a comparatively higher correlation with fewer clusters to obtain findings easily in our case studies.



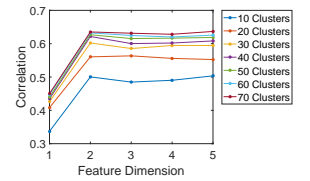
**(a) Beijing, 7:00-7:59**



**(b) Beijing, 17:00-17:59**



**(c) New York City, 7:00-7:59**



**(d) New York City, 17:00-17:59**

**Figure 4: Relationship between regional correlation to number of clusters and feature dimensions with Beijing and New York City data, 7:00-7:59/17:00-17:59, weekdays of November 2015.**

## 5.4 Cluster Performance Evaluation

We compared KACE with both traditional methods and state-of-art co-clustering algorithms on the Beijing dataset using four metrics: **Spatial coverage**, measures the percentage of points identified and clustered by the given method

**Average in-cluster distance**, measures average pair-wise distance within the same O/D region. This is a standard metric for evaluating the quality of clusters. A smaller value is better.

**Table 2**  
**Comparisons with other methods based on Spatial Coverage, in-cluster distance and regional dependency**

Methods	Spatial Coverage	Average Origin in-cluster Distance	Average Destination in-cluster Distance	Regional Correlation	Origin Overlap	Destination Overlap
KACE	100%	2.98km	3.21km	0.8643	0.33%	0.22%
ACE-5 × 5	100%	5.06km	5.21km	0.8068	0.78%	0.64%
ACE-20 × 20	100%	8.48km	8.78km	0.8646	0.75%	0.61%
MLAN	100%	11.82km	12.58km	1	4.43%	4.19%
K-Means++	100%	4.26km	4.42km	1	54.26%	50.75%
DBSCAN	25.75%	0.60km	0.63km	1	39.21%	35.85%

**Regional correlation**, average of the top-five correlations among all pairs of OD regions. We use this metric to measure how well the clustering methods retain trip information.

**Overlap**, estimates the spatial ambiguity among different origin regions or destination regions. This metric is computed using the KNN classification error of cluster labels based on origin or destination coordinates, which should be very small for non-overlapping data [8]. In particular, the KNN model (K=5) was trained and tested on a 9-1 split of the origin or destination points, *i.e.* 10-fold cross-validation classification error is used as overlapping indicator.

For traditional methods, the classic K-Means++ [2] and density-based method DBSCAN [9] are adopted as baselines. We directly clustered OD trips using these two methods by treating trips as points in four dimensional space, *i.e.* two dimensions of origins with latitudes and longitudes plus two dimensions of destinations with latitudes and longitudes. Then evaluated the results in O/D view. For co-clustering methods, we compared with state-of-art multi-view clustering algorithm MLAN [23]. We also compared with the performance of the original ACE algorithm, by discretizing GPS data using 5 × 5 and 20 × 20 grids. For fair comparison, we set the cluster number to 40 for all methods.

Table 2 lists the comparison results. For K-means++ and DBSCAN, regional correlations are at the maximum value 1. However, the resulting clusters have a significant amount of spatial overlap. MLAN has great regional correlation, tolerable overlap but failed to identify detailed regions as the average in-cluster distance is greater than 10km. The original ACE algorithm improves upon the results of MLAN, though it is not as good as KACE. Our method, KACE, outperforms all the other methods for overall performance, which has minimum overlap, best spatial coverage, small region size and good correlation.

## 6 CASE STUDIES

In this section, we looked into spatial features, region partitions and mobility patterns of New York City and Beijing in detail to demonstrate how such regions and patterns reveal important insights for public transport planning and location-based services. We first visualized extracted features on map, *i.e.* spatial patterns, of these two cities in Section 6.1. Then we focused on mobility patterns from two aspects, *i.e.* region partitions and mobility dynamics in Section 6.2. For region partition, we examined the city topology from the origin and destination partitions. For mobility dynamics, we analyzed salient mobility patterns measured by the conditional

transition probabilities of people moving from one region to another. In the end, we compared Beijing and New York in terms of their spatial patterns and mobility patterns in Section 6.3. We found that Beijing and NYC have different city-specific characteristics and Beijing has more correlated mobility patterns than NYC.

### 6.1 Spatial Patterns

We plotted the extracted origin mobility features of Beijing and New York City on each city’s map as shown in Figure 5. We found that the spatial features are non-overlapping in the same dimension and each dimension of the mobility feature captures a unique spatial pattern for each city.

For Beijing’s data, the spatial distribution of Feature Dimension 1 has the biggest variation from Southwest to Northeast (Figure 5a). It implies that trip destination changes the most along this direction. Each parallel stripe represents an origin region with similar destinations in respect to this feature. Feature Dimension 2 spans across Northwest to Southeast of the city, indicating the second axis of mobility variations (Figure 5b). Dimension 3 shows a radial pattern, which correspond to people having different destinations at inner to outer regions of the city centered at Dongdan (Figure 5c).

For New York’s data, the first feature dimension lies along the upper to lower direction (Figure 5d), which aligns perfectly with the major axis of the Manhattan island. The second and third dimensions are represented by concentric patterns centered at the western side and the eastern side of Manhattan (Figure 5e, 5f).

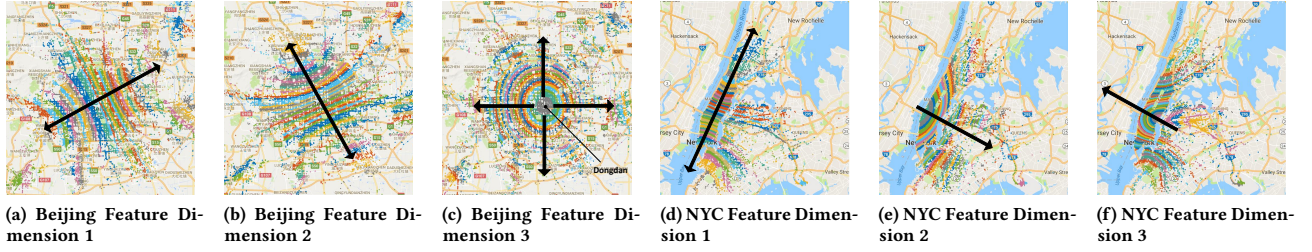
Features of Beijing and New York represent each city’s spatial characteristic and city topology.

### 6.2 Mobility Patterns

For clarity, we only display salient patterns where trips from one origin region have high tendency of moving to a destination region, measured by conditional transition probability  $P(D = d|O = o)$ . Given origin region  $o$  and destination region  $d$ , the conditional transition probability is computed as

$$P_{D|O}(d|o) = \frac{\text{Number of trips started from } o \text{ ended at } d}{\text{Number of trips started from } o \text{ region}}$$

In the case studies, we only display patterns where origin regions have more than 30% trips ended at the same destination region ( $P(D = d|O = o) > 0.3$ ) for Beijing and 20% for NYC. We also denote mobility patterns like “O1 to D2” as pattern “from origin region O1 to destination region D2”.



**Figure 5: Different dimensions of origin features represent different spatial patterns at Beijing (a,b,c) and New York City (d,e,f), 17:00-17:59, weekdays of November 2015.**

### 6.2.1 New York City Findings.

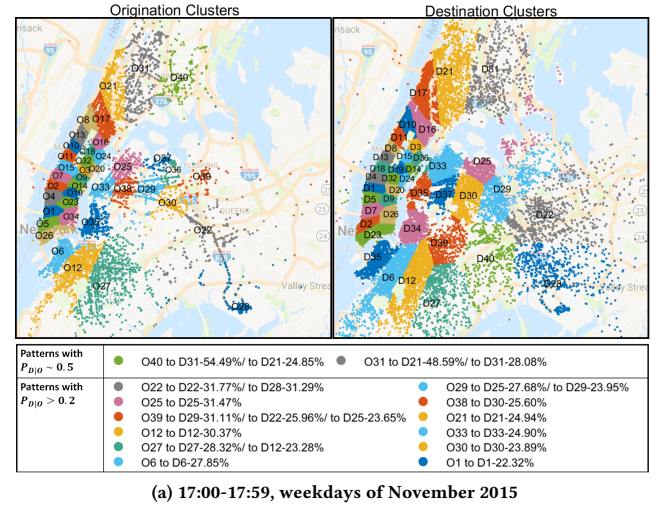
We divided New York City into 40 origin and destination regions with data of workdays in Nov. 2015, 17:00-17:59 shown in Figure 6a. As the size of the NYC dataset is twice that of the Beijing dataset, we replace the linkage clustering algorithm (Step 3 in Algorithm 1) with K-means++, which is more computationally efficient for large dataset [2]. The key findings are shown in the following paragraphs.

**NYC's mobility regions can be split into two types, inner city regions and outer city regions with different patterns.** New York's mobility regions can be split into Manhattan and off-island peripheral areas, Bronx, Brooklyn and Queens. We found that 93.49% people tend to stay on Manhattan without leaving for off-island areas by taxis. One possible reason is that taxis re-entering Manhattan would incur an extra charge. In Figure 6a, the resulting partition of NYC reveals the block city topology in Manhattan.

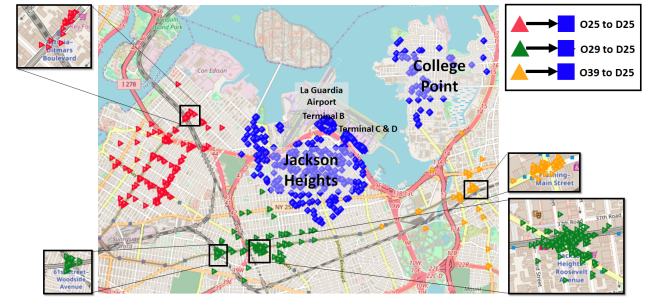
Off-island areas, including Bronx, Brooklyn and Queens, have more concentrated patterns than Manhattan. Among 21 listed salient patterns in Figure 6a, only one pattern, O1 to D1, is on the island of Manhattan, which connects upper area of Lower Manhattan (around SoHo, NoHo and Washington Square Park) to upper Chelsea. We also found that 64.91% of trips end in D1 are within 500m radius of Pennsylvania Station. One possible explanation is that tourists and shoppers go from shopping malls, restaurants and bars to take trains at Penn Station to visit other places in New York. Other patterns originating from off-island areas share similarities that they all point to neighboring regions.

**Trips in suburban districts tend to stay within the respective districts.** From Figure 6a, we observe that two origination regions (O40, O31) have concentrated patterns with nearly 50% of trips go to a destination region. Region O40 corresponds to East Bronx, region O31 is West Bronx. We found that East Bronx has 54.49% possibility of moving inside itself (O40 to D31) and West Bronx has 48.49% probability of inside moving (O31 to D21). Moreover, O40-D21 is the pattern from East Bronx to West Bronx and Inwood. O31 to D31 goes from West Bronx to East Bronx. Therefore, people in Bronx tend to move inside Bronx or travel between East and West Bronx.

**Taxis demand from metro and railway stations to areas without easy metro access is high.** For example, 89.82% of the trips in O22 to D22 have originated from 4 metro stations (Forest Hills-71 Av, 75 Av, Union Tpke-Kew Gardens, Sutphin Boulevard-Archer Avenue-JFK Airport) and one railway station (Jamaica). The destination region D22 mostly consists of areas far away from the



**(a) 17:00-17:59, weekdays of November 2015**



**(b) Detailed Analysis of D25**

**Figure 6: Mobility patterns of New York City with 40 clusters and a detailed analysis of destination region D25.**

metro system, such as Jamaica Hills, Jamaica Estates, Hillcrest, Kew Gardens Hills, Pomonok, Utopia and Fresh Meadows.

The analysis of D25 also shows similar findings. Figure 6b plots three major mobility patterns O25 to D25, O29 to D25 and O39 to D25, with triangles representing origin locations and rectangles representing destination locations. Among these trip clusters, 83.62% of the trips ended at D25 are from metro stations. This result indicates a lack of convenient public transportation from Astoria, Woodside and Elmhurst to Jackson Heights and College Point. Moreover, La

Guardia Airport is also located in D25, and 46.79% of the trips in D25 are rides to the airport. Such findings can assist traffic authorities in New York City identify regions in need of new metro services.

### 6.2.2 Beijing Findings.

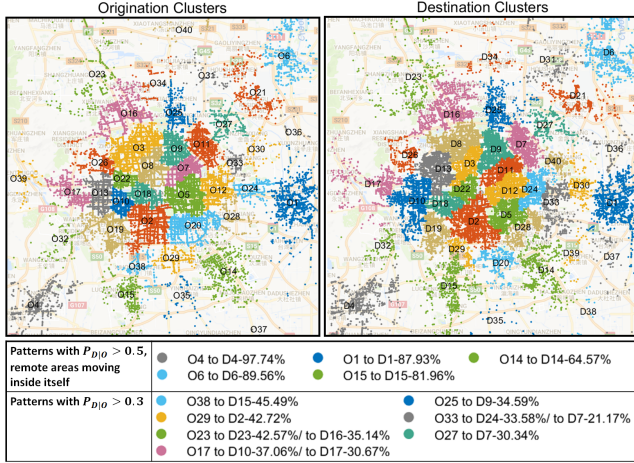


Figure 7: Mobility patterns of Beijing with 40 clusters, weekdays of November 2015, 17:00-17:59

**Beijing has mobility regions from inner to outer city same as its ring expansion city topology.** In Figure 7, we observe that Beijing's mobility regions expand from inner to outer city same as the city's characteristic of expanding outward.

**Suburban districts in Beijing have a tendency of moving inside the region itself by taxis.** Figure 7 illustrates that all patterns with  $P_{D|O} > 0.5$  are from suburban districts and have a tendency ( $\sim 80\%$ ) of staying in the region. Especially Fangshan District (O4 to D4) has 97.74% probability of staying in because it is one of the most remote districts in Beijing and it is a fully functional satellite district. Therefore, people residing in that area have little tendency of moving into the main city. Another reason may be that, these suburb districts are far away and travel by taxis to inner city is expensive.

Moreover, part of Shunyi District (O21) is also far from main city but does not have the pattern of inside moving. It is because that Beijing Capital International Airport is located in O21 and 69.11% trips moving outside the region are originated within 500m radius of the terminals.

Observing this pattern, it is effective to advertise about suburban companies in suburb taxis to promote their products for potential customers, *i.e.* taxi passengers residing locally. While it will be useless to propagandize stores and services in O21 to those who probably would take taxis to go out of the region.

**Major mobility patterns in Beijing end at major residential and recreational areas during 17:00-17:59.** Figure 7 shows 10 patterns with  $P_{D|O} > 0.3$  from 7 origin regions, in which region D7 have two origin regions, O27 and O33. D7 is around Wangjing area, which is a comprehensive area with both residential and recreational functions. Apart from patterns shown in Figure 7, among 16 additional origin regions with  $P_{D|O} > 0.2$ , 4 of which arrived

in D10, a major residential area across Haidian and Shijingshan districts.

### 6.3 Pattern Comparisons between Cities

**Beijing and NYC have different dimension of spatial patterns.**

As shown in Figure 4, the optimal feature dimension for Beijing is three while for NYC is two. This points out that Beijing's mobility patterns are mainly different from Southwest to Northeast, Southeast to Northwest and Inner to Outer city while NYC's mobility regions only depend on South to North and West to East (Figure 5). It makes sense as feature dimension 2 and 3 of NYC are from West to East and East to West respectively. They carry similar spatial patterns thus adding feature dimension three doesn't provide much extra information (Figure 5e, 5f).

**Beijing has a more concentrated mobility pattern than New York in general.** First, the transition probabilities  $P_{D|O}$  is larger than New York's on average in Figure 6a and Figure 7. Second, this is also represented in the OD transition probability heatmap<sup>1</sup> in Figure 8. The transition probability  $P(D = d, O = o)$  is computed as

$$P_{D,O}(d, o) = \frac{\text{Number of trips started from } o \text{ ended at } d}{\text{Total number of trips}}$$

It shows that Beijing's diagonal transition probability (Figure 8b) is much more salient than New York's (Figure 8a), as for the same row or column the diagonal element is more standing out and other elements are small. Moreover, the values of  $P_{D,O}$  in Beijing is much larger than NYC's. Third, Figure 9 illustrates that Beijing's correlation is always larger than New York's for both morning and evening peak hour. Besides, evening peak hour seems to have more concentrated patterns than morning in both cities.

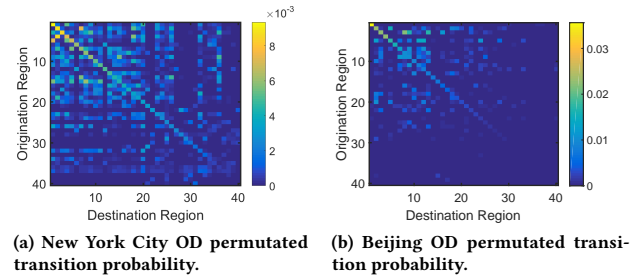
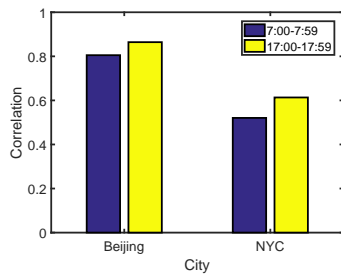


Figure 8: Origination-Destination transition probability heatmap with 40 clusters, weekdays of November 2015, 17:00-17:59.

## 7 CONCLUSIONS

Mobility pattern mining is an important socio-economic problem. However, previous methods of extracting mobility patterns either have overlapping issue or requires an extra step of partitioning the city into discrete regions, which may not be optimal for mobility pattern mining. To address these problems, we proposed a region-aware mobility pattern mining framework by jointly partitions the optimal origin (O) and destination (D) regions while extracting

<sup>1</sup>We numbered region labels in a way that the maximum transition probability mainly lie on diagonal of the probability matrix.



**Figure 9: Correlation comparisons between Beijing and NYC, workdays of November 2015**

mobility patterns. We formulated it as an optimization problem of maximizing OD's correlations with spatial constraints. Kernelized ACE, an improved version of original ACE algorithm, solved the optimization problem efficiently by learning feature representations of trip origins and destinations that guarantee both objectives. Experimental results on Beijing's taxi data show that our approach outperforms all the other methods, including both traditional and state-of-art clustering algorithms. We achieved only  $\sim 0.3\%$  overlap and 86.43% OD correlations, which has best overall performance with minimum overlap. Our case studies on New York City and Beijing also provide insightful findings. In NYC, we identified certain areas that need new metro lines. In Beijing, We found people in suburbs have little interactions with central city, which may reveal certain social problem. From comparisons of both cities, we found that spatial pattern based on mobility is a city-specific characteristic and Beijing has more concentrated mobility patterns than New York City. In the future, we plan to consider POI information and temporal factors as well into extracting semantic time-dependent mobility patterns.

## ACKNOWLEDGMENTS

This work is funded by Shenzhen Municipal Development and Reform Commission, Shenzhen Engineering Laboratory for Data Science and Information Technology, Grant Number: SDRG [2015]1872.

## REFERENCES

- [1] Khaled Ammar, Abdullah Elsayed, Mohamed M Sabri, and Michael Terry. 2015. Busmate: Understanding mobility behavior for trajectory-based advertising. In *Mobile Data Management (MDM), 2015 16th IEEE International Conference on*, Vol. 2. IEEE, 74–79.
- [2] David Arthur and Sergei Vassilvitskii. 2007. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 1027–1035.
- [3] Richard A Becker, Ramon Caceres, Karrie Hanson, Ji Meng Loh, Simon Urbanek, Alexander Varshavsky, and Chris Volinsky. 2011. A tale of one city: Using cellular network data for urban planning. *IEEE Pervasive Computing* 10, 4 (2011), 18–26.
- [4] Matthew B Blaschko and Christoph H Lampert. 2008. Correlational spectral clustering. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 1–8.
- [5] Leo Breiman and Jerome H Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association* 80, 391 (1985), 580–598.
- [6] Qing Cao, Bouchra Bouqata, Patricia D Mackenzie, Daniel Messier, and Joseph J Salvo. 2009. A grid-based clustering method for mining frequent trips from large-scale, event-based telematics datasets. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. IEEE, 2996–3001.
- [7] Inderjit S Dhillon. 2001. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 269–274.
- [8] Chris Ding and Xiaofeng He. 2004. K-nearest-neighbor consistency in data clustering: incorporating local information into global optimization. In *Proceedings of the 2004 ACM symposium on Applied computing*. ACM, 584–589.
- [9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *Kdd*, Vol. 96. 226–231.
- [10] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. 2012. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*. ACM, 3.
- [11] Li Gong, Xi Liu, Lun Wu, and Yu Liu. 2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science* 43, 2 (2016), 103–114.
- [12] Wei Xu, Ming Jin, Zimu Zhou, Costas J Spanos, and Lin Zhang. 2016. Metroeye: smart tracking your metro trips underground. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. ACM, 84–93.
- [13] Wei Xu, Kai Zhang, Zimu Zhou, Ming Jin, Yuxun Zhou, Xi Liu, Costas J Spanos, Zuo-Jun Max Shen, Wei-Hua Lin, and Lin Zhang. 2017. Measuring fine-grained metro interchange time via smartphones. *Transportation research part C: emerging technologies* 81 (2017), 153–171.
- [14] Miao He, Wei Xu, and Ying Kong. 2017. Group recommendation: by mining users' check-in behaviors. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 65–68.
- [15] Chaogui Kang, Yu Liu, and Lun Wu. 2015. Delineating intra-urban spatial connectivity patterns by travel-activities: A case study of Beijing, China. In *Geoinformatics, 2015 23rd International Conference on*. IEEE, 1–7.
- [16] Chia-Tung Kuo, James Bailey, and Ian Davidson. 2015. A framework for simplifying trip data into networks via coupled matrix factorization. In *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 739–747.
- [17] Jingren Liu and Mubarak Shah. 2007. Scene modeling using co-clustering. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 1–7.
- [18] Xi Liu, Li Gong, Yongxi Gong, and Yu Liu. 2015. Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography* 43 (2015), 78–90.
- [19] Jihui Ma, Yang Yang, Wei Guan, Fei Wang, Tao Liu, Wenyuan Tu, and Cuiying Song. 2017. Large-Scale Demand Driven Design of a Customized Bus Network: A Methodological Framework and Beijing Case Study. *Journal of Advanced Transportation* 2017 (2017).
- [20] Sara C Madeira and Arlindo L Oliveira. 2004. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 1, 1 (2004), 24–45.
- [21] Anuran Makur, Fabián Kozynski, Shao-Lun Huang, and Lizhong Zheng. 2015. An efficient algorithm for information decomposition and extraction. In *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*. IEEE, 972–979.
- [22] Netflix. 2009. Netflix Prize Data Set. (2009). <http://archive.ics.uci.edu/ml/datasets/Netflix+Prize>
- [23] Feiping Nie, Guohao Cai, and Xuelong Li. 2017. Multi-View Clustering and Semi-Supervised Classification with Adaptive Neighbours.. In *AAAI*. 2408–2414.
- [24] Alfred Rényi. 1959. On measures of dependence. *Acta mathematica hungarica* 10, 3-4 (1959), 441–451.
- [25] Thiago H Silva, Pedro OS Vaz De Melo, Jussara M Almeida, and Antonio AF Loureiro. 2014. Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Communications* 21, 1 (2014), 42–51.
- [26] Robert R Sokal. 1958. A statistical method for evaluating systematic relationship. *University of Kansas science bulletin* 28 (1958), 1409–1438.
- [27] Jinjun Tang, Fang Liu, Yinhai Wang, and Hua Wang. 2015. Uncovering urban human mobility from large scale taxi GPS data. *Physica A: Statistical Mechanics and its Applications* 438 (2015), 140–153.
- [28] toddwschneider. 2017. Unified New York City Taxi and Uber data. <https://github.com/toddwschneider/nyc-taxi-data/>.
- [29] DESA UN. 2015. World urbanization prospects: The 2014 revision. *United Nations Department of Economics and Social Affairs, Population Division: New York, NY, USA* (2015).
- [30] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 186–194.
- [31] Wangsheng Zhang, Shijian Li, and Gang Pan. 2012. Mining the semantics of origin-destination flows using taxi traces.. In *UbiComp*. 943–949.
- [32] Yu Zheng, Yanchi Liu, Jing Yuan, and Xing Xie. 2011. Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 89–98.
- [33] Bing Zhu, Qixing Huang, Leonidas Guibas, and Lin Zhang. 2013. Urban population migration pattern mining based on taxi trajectories. In *3rd international workshop on mobile sensing: the future, brought to you by big sensor data*. Philadelphia, USA.