

# Discovering Fine-grained Spatial Pattern from Taxi Trips: Where Point Process Meets Matrix Decomposition and Factorization

Junbiao Pang, Jing Huang, Xue Yang, Zuyun Wang, Haitao Yu, Qingming Huang, and Baocai Yin

**Abstract**—As increasing volumes of urban data are being available, new opportunities arise for data-driven analysis that can lead to improvements in the lives of citizens through evidence-based policies. In particular, taxi trip is an important urban sensor that provides unprecedented insights into many aspects of a city, from economic activity, human mobility to land development. However, analyzing these data presents many challenges, e.g., sparse data for fine-grained patterns, and the regularity submerged by seemingly random data. Inspired by above challenges, we focus on Pick-Up (PU)/Drop-Off (DO) points from taxi trips, and propose a fine-grained approach to unveil a set of low spatio-temporal patterns from the regularity-discovered intensity. The proposed method is conceptually simple yet efficient, by leveraging point process to handle sparsity of points, and by decomposing point intensities into the low-rank regularity and the factorized basis patterns, our approach enables domain experts to discover patterns that are previously unattainable for them, from a case study motivated by traffic engineers.

**Index Terms**—Spatio-temporal Pattern, Fine-grained Pattern, Low-rank Regularity, Matrix Factorization, Taxis Trip, Point Process

## I. INTRODUCTION

For the first time in history, more than half of the population is living in urban areas. Meanwhile, keeping public services

Manuscript received March 22, 2017; revised July 20, 2017; accepted November 02, 2017. This work was supported in part by Natural Science Foundation of China under Grant 61672069, Grant 61472387, Grant 61332016, Grant 61620106009 in part by the Beijing Post-Doctoral Research Foundation, by the Beijing Municipal Commission of Education under Grant KM201610005034, and by the Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (PHR). (Corresponding author: Junbiao Pang, Haitao Yu.)

J. Pang and J. Huang are with the College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China (e-mail: junbiao\_pang, jhuang@bjut.edu.cn).

X. Yang and H. Yu are with the Beijing Transportation Information Center, Beijing 100161, China (e-mail: {yangxue,yuhaitao}@bjjtw.gov.cn).

Z. Wang is with the Beijing Municipal Transportation Law Enforcement Corps, Beijing 100044, China (e-mail: wangzuyun@bjjzf.gov.cn).

Q. Huang is with the University of Chinese Academy of Sciences, Chinese Academy of Sciences (CAS), Beijing 100049, China, and the Institute of Computing Technology, CAS, Beijing 100190, China (email: qmhuang@ucas.ac.cn).

B. Yin is with Advanced Invocation Center for Future Internet Technology, Dalian University of Technology, Dalian 116024, China, and Beijing University of Technology, Beijing 100124, China (e-mail: ybc@dlut.edu.cn).

of cities effective, efficient, and sustainable in a green way is among the most emergent undertakings for human. In the recent past, policy makers faced significant difficulties in obtaining the data to sense the dynamics of a city and evaluate policies in practice. As a contrast, data are now abundant with the development of pervasive computing, meeting the vision of a smart city in which sensors immediately perceive the current status and further interact accordingly, see, e.g., [1] [21]. Therefore, the challenge is how to make sense of these data.

One particular yet important type of these data is taxi trip. Taxi is a significant transportation in urban areas since it nearly supplies with the flexible door-to-door service. Therefore, a thorough understanding of taxi trip is an important way to sense a city at different spatial scales, e.g., from a street to a district. In Beijing, China, for instance, each day about 50,000 taxis carry over 0.4 million passengers. As plotted in Fig. 2(a), the number of Pick-UPS (PUs)/Drop-Offs (DOs) per day varies roughly from 1,500 to 10,000 per 15 minutes in Beijing. There is an obvious regularity in Fig. 2(a) – the plotted lines are very similar to that of each day. Spatially, some points in Fig. 2(b) concentrate around Beijing Railway Station (indicated by “□”); while at the morning peak/evening one in Figs. 2(c)/(d), most points happen in several areas (indicated by “□”, “○”, and “▽”); interestingly, in Fig. 2(e), most points still happen in centralized business district indicated by “○”. A natural question is what spatial characteristics can be uncovered from PUs and DOs. To unveil the spatio-temporal patterns sensed by taxi trips, there are two main challenges as follows:

1) *Sparsity of PUs/DOs at a fine-grained scale*: Usually, only a small number of points happen in a fine-grained spatio-temporal grid. Therefore, it is necessary to introduce new statistical tool to deal with this problem.

2) *The mixture of regularity and disparity*: Regularity represents the commonness observed from multiple time slices; while, disparity represents the specificity of each time slice. It is reasonable to assume that the complex spatio-temporal point distribution consists of the regularity and the disparity.

Therefore, it is natural to ask how to unveil the mixed patterns from these seemingly random points without suffering the sparsity problem. There are several potential benefits behind this motivation: fine-grained patterns at a small spatio-temporal scale, interesting patterns from the regularity, and unexpected patterns from the disparity. Principally, fine-grained

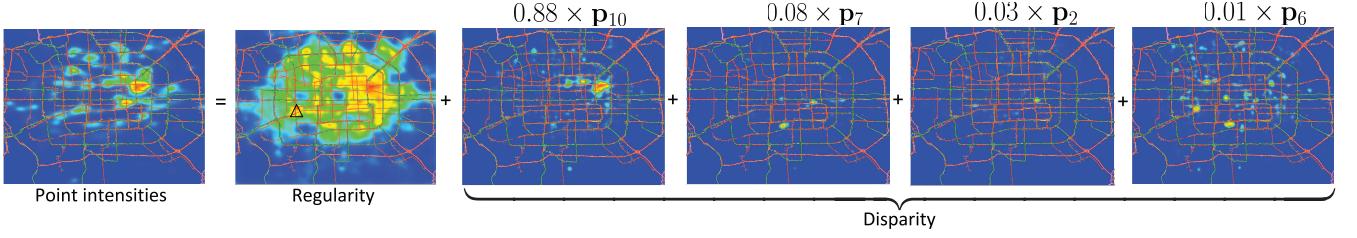


Fig. 1: Decomposing the intensities of PUs from 3:00 a.m. to 3:15 a.m. into a regularity intensity and a disparity one, which is further factorized into four weighted spatial basis patterns. The regularity intensity indicates that some areas are always very active (*i.e.*, the intercity bus station indicated by “ $\Delta$ ”), while the point intensities do not visually indicate this. Moreover, at this time slice, the most active area in  $p_{10}$ , with a weight 0.88, is the intersection of a highway to the airport and a highway to the downtown (best viewed in pdf).

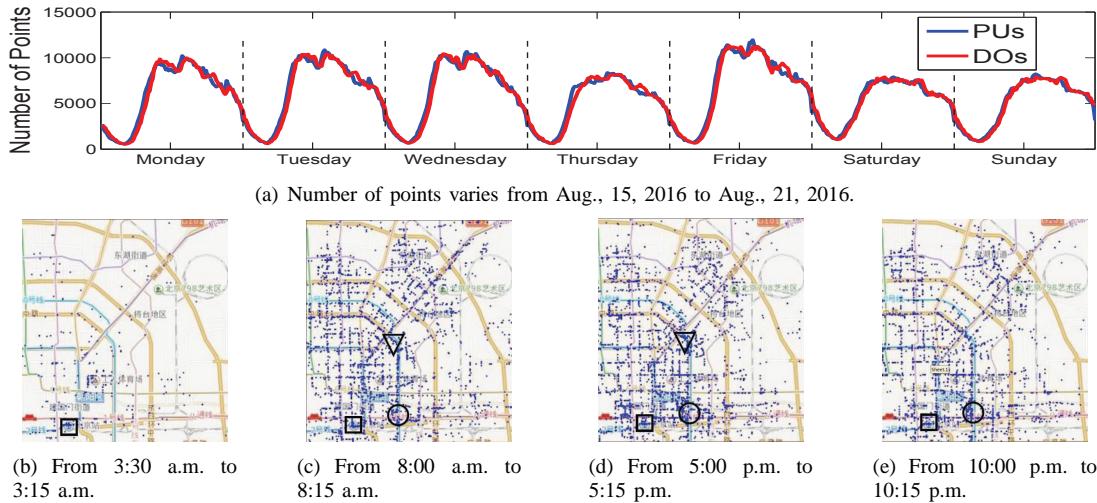


Fig. 2: The spatio-temporal distribution of PUs/DOs in Beijing, China. The spatial distribution of PUs in the north-east area of Beijing is illustrated from (b) to (e) (best viewed in color).

analysis requires to divide the spatio-temporal volume into a small grids, aggravating the sparsity problem; however, the frequently used method barely handle the sparsity problem, *e.g.*, discretized counts [18].

In this paper, we combine *point process intensities* with matrix *decomposition* and *factorization* to unveil these hidden patterns at a fine-grained granularity. First, an enormous volume of literature avoid the sparsity problem by empirically defining a proper spatio-temporal grid, there is little work to estimate the number of points for *any* location. Second, it is a reasonable idea to decompose the distribution of points into the regularity and the disparity; this is significant departure from the previous approaches: clustering directly on PUs/DOs [18] [7]. Third, the disparity is further factorized into a set of basis patterns, which facilitate to understand the relationship among different time slices. In summary, we desire a method to decompose the regularity and the disparity from the estimated point intensities.

Point Process Intensities (PPIs) [8] refer to the techniques of estimating intensities of *unobserved* locations from the spatial correlations of the observed ones. The technique has been used in neuroscience [9], finance [2] and forestry [14]. The “Point intensities” in Fig. 1 shows a simple example, explaining the ability of PPI: fitting an intensity for the location which has

not observed an point. Therefore, the sparsity of PUs/DOs are naturally handled with the help of PPI.

After obtaining the location-wise intensities from PPI, the key question is therefore how to decompose the regularity and the disparity from the seemingly random intensities. One important assumption adopted here is that the regularity is composed from multiple simpler ones. Therefore, the point intensities from multiple time slices are decomposed into the nonnegative low-rank term [29] and the nonnegative sparse residual one [12]. The rank indicates the number of patterns shared among different time slices; while the sparse residual represents the special spatio-temporal characteristics of each time slice.

Based on the proceeding steps, the main contributions of this paper are summarized as follows:

- 1) To the best of our knowledge, this paper is the first to investigate PPI for the fine-grained analysis. This technique can be considered as a potential tool for other point-related traffic problems, *e.g.*, point data from cell phones [37].
- 2) By aiming to unveil the regularity of PUs/DOs from taxi trips, to our best knowledge, we first decompose point intensities into the low-rank commonness and the sparse disparity, presenting a comprehensive series of

TABLE I: Averaged statistics of PUs/DOs every 15 min.

Type of Point	Min/Max	Mean±Std	#Point/KM <sup>2</sup>
PU	561/11,971	5,969.20±3,128.88	6.78
DO	519/11,339	5,903.48±3,142.81	6.70

experiments to illustrate the benefits of this technique. We find that low-rank is efficient to characterise this regularity.

3) By factorizing the sparse disparity into basis patterns, we could understand the correlations among different time slices. We find that these basis patterns significantly uncover the spatial activities of human; meanwhile, the weights of these patterns facilitate the comparisons between different time slices. Consequently, the disparity is presented in a comparable approach to better understand traffic problems.

## II. RELATED WORK

Motivated by different applications, researchers utilize taxi trips in diverse approaches. Technically, these literatures can be categorized into two categories: 1) point mapping which mainly unveils information from the distribution of PUs/DOs; and 2) moment mapping which concerns mobile patterns from the paired PU-DO data.

**Point mapping:** Point mapping from taxi trips could be used for: 1) rebuilding road map [6]; 2) analyzing traffic hot spots [22] [34] [13] [31]; 3) detecting flawed urban planning [41]; 4) recommending routes for avoiding traffic jams [39] [11] [7] [32]; 5) suggesting a driving strategy for taxi drivers to find more passengers [18] [20]. In summary, point mapping focuses on spatial patterns from the distributions of PUs/DOs. An intuitive yet consensual solution is *directly* clustering on the frequencies of PUs/DOs. For instance, Pan *et al.* [31] utilized DBSCAN clustering to find hot spots from one year PUs/DOs. Ding *et al.* [11] used a hierarchical agglomerative clustering method to discover hot spots in Shanghai. Moreover, these works need an important precondition: enough points are accumulated to discover patterns at a coarse granularity, *e.g.*, points from one year were used to identify hot spots [11].

**Moment mapping:** The second category uncovers intra-urban human mobility patterns for visual exploration. For instance, Ding [10] *et al.* investigated different mobility patterns of taxi drivers at two income levels and analyzed their distinct behaviors. Zheng [40] *et al.* systematically investigated spatial trajectories from a wide spectrum of perspectives and disciplines. Wood [38] *et al.* proposed a visual analytic method for exploration of movement data at both an aggregated and an individual level. Stolte [35] *et al.* used both PU/DO positions and attributes associated with the movement.

**Summarization:** Here we just give a briefly review of the most related work. This paper belongs to the point mapping. Generally speaking, for the point mapping, the most severe problem of clustering on the frequencies of PUs/DOs is the dilemma between fine-grained analysis and the limited number of points: a fine-grained pattern occurs at a small scale;

meanwhile, a small scale statistically consumes a large amount of points [3].

## III. TAXI GLOBAL POSITIONING SATELLITE TRIPS

The taxi GPS trip data set used in this paper comes from the Beijing Transportation Information Center<sup>1</sup>. Beijing is one of the largest cities in China. The taxi GPS trips were generated over of a week (from Aug., 15, 2016 to Aug., 21, 2016). Approximated 53,000 taxis with GPS devices were used to build this data set. The area in the Fifth Ring Road, *i.e.*, the longitudes from 116.213469 to 116.559137 and the latitudes from 39.763576 to 40.031272, is considered as the spatial boundary with about 880.81 square kilometers, since most of urban activities have taken placed in this area. There are some error GPS data due to the blocked signal transmission by high buildings or overpasses. We removed these error GPS data, if the number of digits after the point is less than 6. Transition of the meter reflects that passengers are picked up or set down. Therefore, the PU/DO locations are identified from these transition data. This leads to 4,011,302 PU points and 3,967,140 DO ones within the Fifth Ring Road.

To better understand the statistics of PUs/DOs, we partition the whole city into small spatio-temporal grids, *i.e.*, the 1 squared kilometer and the 15 min. interval. Therefore, we obtain  $4 \times 24 \times 7 = 672$  time slices. As illustrated in Table I, the PU/DO points are very sparse for each time slice, *e.g.*, averaged 6.7 for each square kilometer. Besides, averaged minimal number of points are far smaller than that of the maximal one, *e.g.*, from 561 to 11,971. It indicates that the distributions of these points from different time slices are very dissimilar. The values of the standard deviations also consistently verify this characteristic.

## IV. DECOMPOSE POINT INTENSITIES INTO LOW-RANK REGULARITY AND FACTORIZED DISPARITY

This paper aims at proposing an approach which is expected to be a general tool for point data from the transportation-related problem. Therefore, from the viewpoint of algorithms, we use PUs/DOs from taxi trips as a case study.

### A. Problem Statement

**Problem 1 (Intensity = Regularity + Disparity):** Given a series of point intensities  $\lambda_t \in \mathbb{R}^{N \times M}$ , ( $t = 1, \dots, T$ ), sampled from a  $N \times M$  grids at the  $t$ -th time slice, the problem is how to decompose the intensities  $\lambda_t$  into the regularity  $B$  ( $B \in \mathbb{R}^{N \times M}$ ) and the disparity  $H$  ( $H \in \mathbb{R}^{N \times M}$ ), which is further factorized into a sum of basis patterns  $\mathbf{p}_k$ , ( $\mathbf{p}_k \in \mathbb{R}^{N \times M}$ ,  $k = 1, \dots, K$ ) as follows:

$$\begin{aligned} \lambda_t &= B + H \\ &= B + \sum_{k=1}^K c_k \mathbf{p}_k, \end{aligned} \tag{1}$$

where  $c_k$  ( $c_k \in \mathbb{R}^+$ ) are the nonnegative coefficients.

<sup>1</sup>This data is released publicly for any non-commercial research. For any interested one, please contact the corresponding authors.

TABLE II: Some important notations in this paper.

Notation	Definition
$\lambda \in \mathbb{R}^{N \cdot M \times 1}$	The intensities reshaped from $\lambda \in \mathbb{R}^{N \times M}$
$\Lambda \in \mathbb{R}^{N \cdot M \times T}$	Intensity matrix from $N \times M$ spatial grids with $T$ time slices
$H_t \in \mathbb{R}^{N \times M}$	Intensity matrix from a $N \times M$ spatial grid
$\mathbf{H} \in \mathbb{R}^{T \times N \cdot M}$	Reshaping $T$ matrices $H_t$ into a matrix $\mathbf{H}$
$B_t \in \mathbb{R}^{N \times M}$	Regularity matrix for the $t$ -th time slice
$\mathbf{B} \in \mathbb{R}^{T \times N \cdot M}$	Reshaping $T$ matrices $B_t$ into a matrix $\mathbf{B}$

The decomposition scheme in (1) is general to model the co-occurrence of the regularity and the disparity. More specially,  $p_k$  serve as the time-aware spatial patterns which indicate the active regions of human; coefficients  $c_k$  reflect correlations among different time slices. However, solving (1) is a non-trivial task.

In this paper, we ties together three ideas, Log-Gaussian Cox point Process (LGCP), Nonnegative Low-rank and Sparse Decomposition (NLSD), and Nonnegative Matrix Factorization (NMF) [17], to extract spatial patterns in (1). The procedure is as follows:

1. Fit an intensity surface  $\lambda_t$  for the time slice  $t$  over a city by LGCP;
2. Model the low-rank regularity and the sparse disparity from the matrix  $\Lambda = [\bar{\lambda}_1, \dots, \bar{\lambda}_T]^\top$ , where the intensities  $\bar{\lambda}_t \in \mathbb{R}^{N \cdot M \times 1}$  are the reshaped vector from  $\lambda_t \in \mathbb{R}^{N \times M}$ .
3. Factorize basis patterns  $\mathbf{p}_k$  and the coefficients  $c_k$  with the loss function,  $\lambda_t - B \approx \sum_{k=1}^K c_k \mathbf{p}_k$ .

Some important notations are summarized in Table II.

### B. Handling Sparsity by Log-Gaussian Cox Process

The intensity of point generating process is assumed to vary smoothly over the spatial region. A popular model for such data is the inhomogeneous Poisson Process with a Gaussian Process. Given a set of spatial locations,  $\{\mathbf{x}_i\}_{i=1}^N$ , where each  $\mathbf{x}_i$  could be any points within a designated spatial region  $A \subset \mathbb{R}^2$ , which is presumed to have been partitioned into disjoint regions, i.e.,  $A = A_1 \cup A_i, i = 1, \dots, V$ . Formally, a Cox process is defined via a stochastic intensity function  $\lambda(\mathbf{x}) : A \rightarrow \mathbb{R}^+$ . Thus, the number of points  $N(A_i)$  in a subregion  $A_i$  is Poisson distribution as follows:

$$N(A_i) \sim \text{Poisson} \left( \int_{A_i} \lambda(\mathbf{x}) d\mathbf{x} \right), \quad (2)$$

where  $d\mathbf{x}$  indicates integration with respect to Lebesgue measure over the region. Moreover, ‘‘memoryless’’ assumes that  $N(A_i)$  is independent of  $N(A_j)$  for disjoint regions  $A_i$  and  $A_j$ . This independence is due to the completely independent nature of points in a Poisson process [16].

If we restrict our consideration on a region  $A_i$ , the probability density of a set of  $N$  points fall into this region, conditioned on the intensity function  $\lambda(\mathbf{x})$  is:

$$p(\mathbf{x}_1, \mathbf{x}_N | \lambda(\mathbf{x}_i)) \sim \exp \left\{ - \int_{A_i} \lambda(\mathbf{x}) d\mathbf{x} \right\} \prod_{n=1}^N \lambda(\mathbf{x}_n). \quad (3)$$

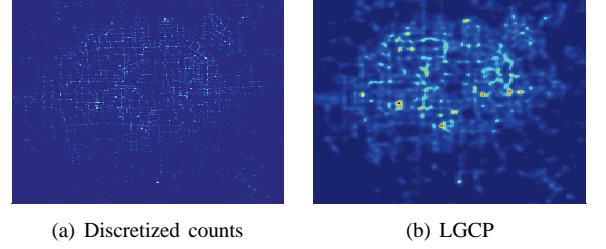


Fig. 3: The interpretation ability of LGCP for the time slice 9:00 a.m.-9:15 a.m.<sup>3</sup>(best view in pdf).

Note that each intensity function  $\lambda(\mathbf{x}_i)$  can be scaled by some non-negative factor and remains a valid intensity function. To model the intensity function  $\lambda(\mathbf{x}_i)$ , a hierachal model [30] is adopted as follows:

$$z \sim GP(0, K) \quad (4)$$

$$\lambda(\mathbf{x}) \sim \exp(z) \quad (5)$$

where both prior  $z$  and intensity  $\lambda(\mathbf{x})$  are two stochastic values, and  $GP(0, K)$  is Gaussian Process [33] model with zero mean and covariance matrix  $K$  calculated by a kernel function  $K(\mathbf{x}_i, \mathbf{x}_j)$ . The advantage of Gaussian Process (4) empower LGCP to estimate the intensity for a given GPS location. Therefore, LGCP enable our method to obtain fine-grained spatial patterns. Fig. 3 shows that LGCP produces more smooth point intensities than discretized counts.

The  $K(\mathbf{x}_i, \mathbf{x}_j)$ , in this paper, is selected as Matérn covariance function<sup>2</sup> with  $\nu = 3/2$  as follows:

$$K_{3/2}(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 \left( 1 + \sqrt{3} \frac{d}{\rho} \right) \exp \left( -\sqrt{3} \frac{d}{\rho} \right). \quad (6)$$

Rather than the exponential covariance kernel,  $k(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 \exp(-\frac{1}{2} \|\mathbf{x}_i - \mathbf{x}_j\|)$ , Matérn with  $\nu=3/2$  isotropically measures the distance between two points.

Though we have formulated a continuous model for the conceptual simplicity, we discretize the whole city into  $N \times M$  grids to gain computational tractability. We fit an intensity surface  $\lambda_t \in \mathbb{R}^{N \times M}$  and obtain the matrix  $\Lambda = [\bar{\lambda}_1, \dots, \bar{\lambda}_T]^\top$ .

### C. Decomposing Regularity by Nonnegative Low-Rank and Sparse Decomposition

To adaptively model the regularity, we decompose the intensity matrix  $\Lambda$  ( $\Lambda \in \mathbb{R}^{T \times N \cdot M}$ ) into the low-rank regularity  $\mathbf{B}$  ( $\mathbf{B} \in \mathbb{R}^{T \times N \cdot M}$ ) and the sparse disparity  $\mathbf{H}$  ( $\mathbf{H} \in \mathbb{R}^{T \times N \cdot M}$ ) as follows:

$$\begin{aligned} & \min_{\mathbf{B} \geq 0, \mathbf{H} \geq 0} \|\mathbf{B}\|_* + \gamma \|\mathbf{H}\|_1 \\ & \text{s.t.: } \Lambda = \mathbf{B} + \mathbf{H}, \end{aligned} \quad (7)$$

where  $\|\cdot\|_*$  is the nuclear norm of a matrix defined by the sum of all singular values;  $\|\cdot\|_1$  is the well-known  $\ell_1$  norm defined by the component-wise sum of absolute values of all

<sup>2</sup> $d = 2 \times R \times \arcsin \sqrt{\sin^2(\frac{\varphi_2 - \varphi_1}{2}) + \cos(\varphi_2) \times \cos(\varphi_1) \times \sin^2(\frac{\phi_2 - \phi_1}{2})}$  where  $\phi$  is longitude,  $\varphi$  is latitude, and  $R$  is the radius of the earth. In this paper,  $R$  is set to be 6378.137 kilometers.

<sup>3</sup>To obtain the high-quality vector graphics, we plotted Fig. 3 with Matlab.

entries;  $\gamma > 0$  is a constant providing a trade-off between the sparse and low-rank components.

The nonnegative constraints,  $\mathbf{B} \geq 0$ , and  $\mathbf{H} \geq 0$ , guarantee both the decomposed matrices still to be point intensities. A matrix with rank  $K$  indicates that this matrix is spanned by a subspace with  $K$  basises. That is, there are  $K$  regularity patterns in the matrix  $\mathbf{B}$ . We term (7) as Nonnegative Low-rank and Sparse Decomposition (NLSD). However, the nonnegative constraints make (7) be complex.

The optimization of (7) can be achieved with augmented Lagrangian methods [4] by introducing auxiliary variables  $\mathbf{Q} = \mathbf{B}$ ,

$$\begin{aligned}\mathcal{L}(\mathbf{Q}, \mathbf{B}, \mathbf{H}) = & \|\mathbf{Q}\|_* + \gamma\|\mathbf{H}\|_1 + \text{trace}(\mathbf{Y}^\top(\mathbf{\Lambda} - \mathbf{B} - \mathbf{H})) \\ & + \frac{\rho}{2}\|\mathbf{\Lambda} - \mathbf{B} - \mathbf{H}\|_F^2 + \text{trace}(\mathbf{Z}^\top(\mathbf{Q} - \mathbf{B})) \\ & + \frac{\rho}{2}\|\mathbf{Q} - \mathbf{B}\|_F^2,\end{aligned}$$

where  $\rho \geq 0$  is called the penalty parameter,  $\|\cdot\|_F$  denotes the Frobenius norm, and the matrices  $\mathbf{Y}$  and  $\mathbf{Z}$  are the dual variables associated with the constraints,  $\mathbf{\Lambda} = \mathbf{B} + \mathbf{H}$ , and  $\mathbf{Q} = \mathbf{B}$ .

*Solving for Q:* when other variables are fixed, the subproblem w.r.t.  $\mathbf{Q}$  is:

$$\min \|\mathbf{Q}\|_* + \frac{\rho}{2}\|\mathbf{Q} - \mathbf{B} + \mathbf{Z}/\rho\|_F^2, \quad (8)$$

which can be solved by the Singular Value Threshold method [5]. More specially, let  $\mathbf{U}\Sigma\mathbf{V}^\top$  be the SVD form of  $(\mathbf{B} - \mathbf{Z}/\rho)$ , the solution to (8) is as follows:

$$\mathbf{Q} = \mathbf{U}\mathcal{S}_{1/\rho}(\Sigma)\mathbf{V}^\top, \quad (9)$$

where  $\mathcal{S}_\alpha(\mathbf{X}) = \max(\mathbf{X} - \alpha, 0) + \min(\mathbf{X} + \alpha, 0)$  is the shrinkage operator [4].

*Solving for H:* The sub-problem w.r.t.  $\mathbf{H}$  can be simplified as:

$$\|\mathbf{H}\|_1 + \frac{\rho}{2\gamma}\|\mathbf{H} - \mathbf{B} - \mathbf{\Lambda} - \mathbf{Y}/\rho\|_F^2, \quad (10)$$

which has a closed from solution,  $H = \mathcal{S}_{\gamma/\rho}(\mathbf{B} + \mathbf{\Lambda} + \mathbf{Y}/\rho)$ .

*Solving for B:* with other variables being fixed, we updated  $\mathbf{B}$  by solving

$$\mathbf{B} = \arg \min_{\mathbf{B} \geq 0} \frac{1}{2} \left\| \mathbf{B} - \frac{1}{4}(\mathbf{\Lambda} + \mathbf{H} + \mathbf{Q} + \mathbf{Y}/\rho + \mathbf{Z}/\rho) \right\|_F^2,$$

which can be solved in a closed form solution,

$$\mathbf{B} = \mathcal{P}_+ \left( \frac{1}{4}(\mathbf{\Lambda} + \mathbf{H} + \mathbf{Q} + \mathbf{Y}/\rho + \mathbf{Z}/\rho) \right), \quad (11)$$

where  $\mathcal{P}_+(\mathbf{X}) = \max(\mathbf{X}, 0)$  projects the elements of a matrix  $\mathbf{X}$  into nonnegative subspace. Since the objective (7) is convex subject to nonnegative constraints, and all of its subproblems can be solved exactly based on the existing theoretical results [28], NLSD converges to global optima.

#### D. Factorizing Disparity by Nonnegative Matrix Factorization

To compare the basis patterns and the possible correlations among different time slices, NMF is proposed to factorize the disparity matrix into different basis patterns as follows:

$$\mathbf{H} = \mathbf{C}^\top \mathbf{P}, \quad (12)$$

where  $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_K]^\top$  ( $\mathbf{P} \in \mathbb{R}^{K \times N \cdot M}$ ) is the basis pattern matrix, and matrix  $\mathbf{C}$  ( $\mathbf{C} \in \mathbb{R}^{T \times K}$ ) is the nonnegative weight matrix. Note that NMF is a special case of tensor factorization to discover latent patterns in different application scenarios [25] [27] [24] [26]. Each vector of  $\mathbf{H}$  can be reconstructed from the weight and the basis patterns as follows:

$$\mathbf{h}_i = \sum_{k=1}^K c_{ik} * \mathbf{p}_k. \quad (13)$$

This naturally results in  $K$  spatial basis patterns  $\mathbf{p}_k$  ( $\mathbf{p}_k \in \mathbb{R}^{1 \times N \cdot M}$ ) and the weights for each pattern  $c_{ik}$ . Due to the superposition property of Poisson Process [16] and the nonnegativity of the basis and the coefficients, the basis patterns can be interpreted as archetypal intensities used to describe disparity for each time slice.

In this paper, we solve the optimal matrices  $\mathbf{C}$  and  $\mathbf{P}$  by the squared Frobenius norm,  $\mathbf{C}, \mathbf{P} = \arg \min_{\mathbf{C} \geq 0, \mathbf{P} \geq 0} \|\mathbf{H} - \mathbf{C}^\top \mathbf{P}\|_F^2$ . Due to the nonnegativity, the basis  $\mathbf{p}_k$  tends to be spatially disjoint, exhibiting a more “parts-based” decomposition than clustering approach. Because the restrictive of nonnegative factorization disallows negative values to cancel out positive ones. Consequently, this restriction in practice leaves a sparser and often more interpretable basis patterns [17]. It is also noted that NMF tends to generate partially overlapped patterns. In terms of efficiency, NMF can be efficiently optimized by the fast method [19] with about 4 times empirically faster than the traditional approach [17].

The proposed approach is totally different from Probabilistic Tensor Factorization (PTF) [36] in motives and techniques: 1) PTF organizes the multi-dimension data into the interactive and latent clusters, our method aims to discover both the regularity and the disparity spatial patterns from point data. 2) PTF utilizes the tensor decomposition on the discretized tensor, our method applies NMF on the sparse disparity. Conceptually, NMF in our method can be replaced by PTF in order to find more interactions among latent factors.

## V. CASE STUDY

The experimental setup on the self-collected data set in Section III is as follows: after the fitting of LGCP, we interpolated the area of the Fifth Ring Road in Beijing into the  $200 \times 200$  grids. i.e., about  $100 \times 200$  squared meters per block. Therefore, the resulting  $\mathbf{\Lambda}$  is a  $672 \times 40,000$  matrix. The number of grids is empirically determined according to the spatial resolution required from different scenarios, as discussed in [30].  $100 \times 200$  in LGCP results in a 0.023 square kilometers, which nearly correspond to a small community in Beijing.

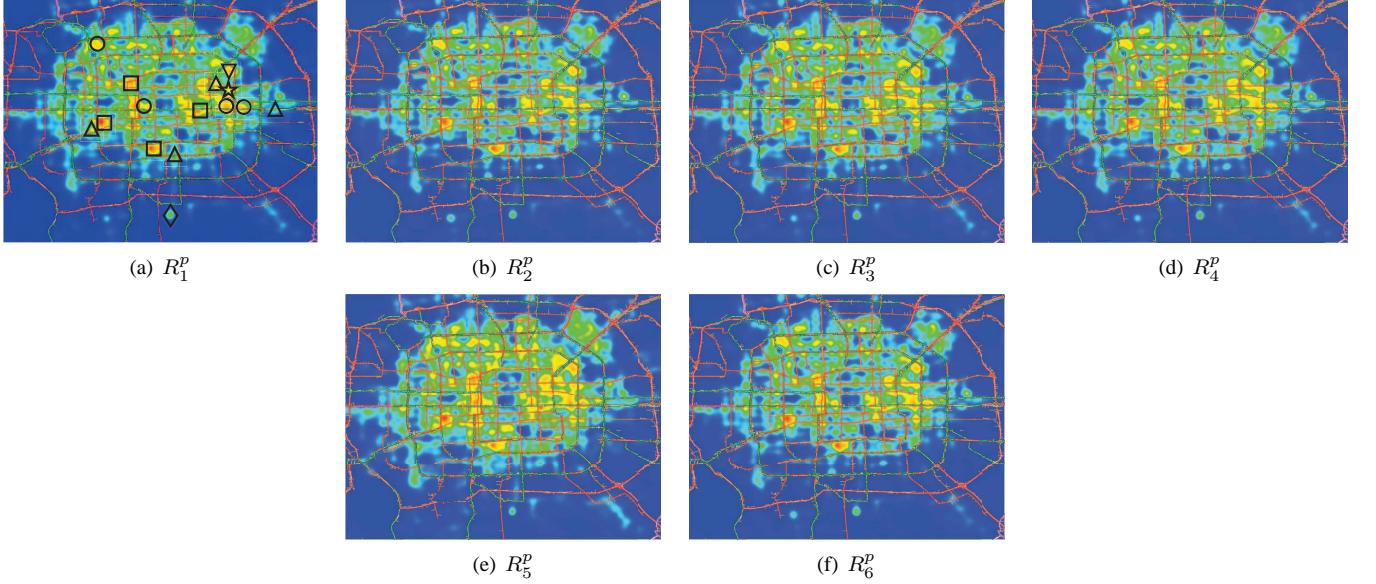


Fig. 4: Visual comparison of different types of regularities decomposed by NMF for PUs. In (a), “ $\Delta$ ” indicates the location of an intercity bus station, “ $\square$ ” represents the location of a railway station, “ $\circ$ ” means the centralized business area, “ $\nabla$ ” represents the intersection of a highway and a road to downtown, “ $\diamond$ ” is the airport, and  $\star$  indicates a popular destination for shopping, dining and entertainment.

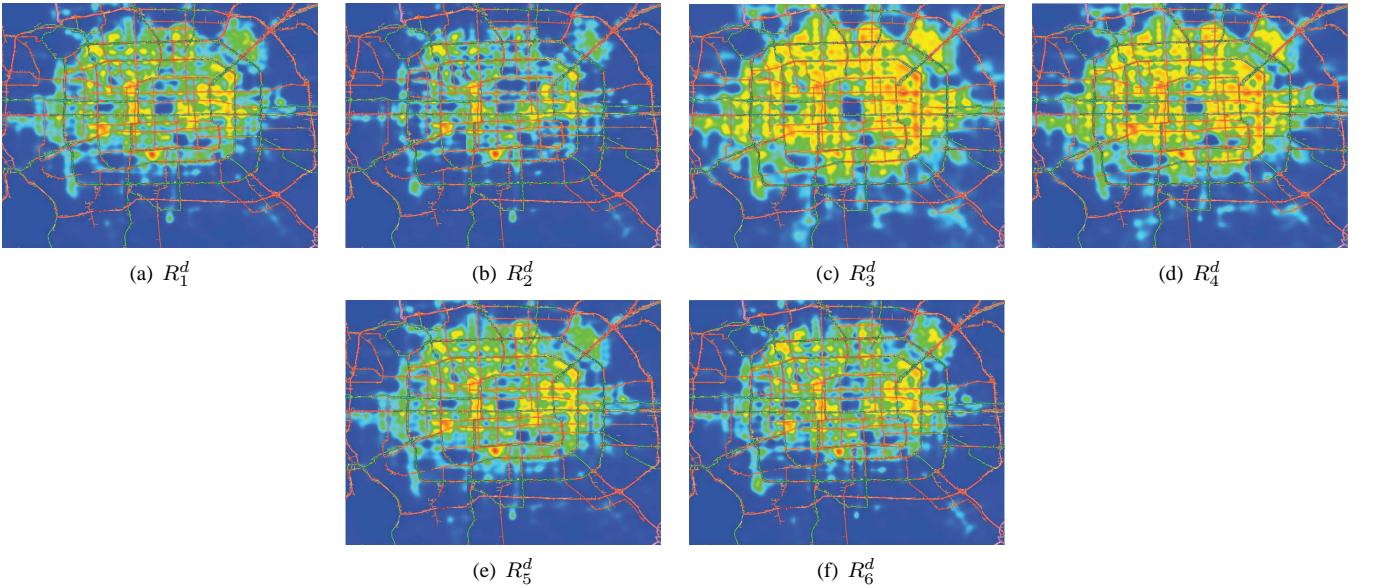


Fig. 5: Visual comparison of different types of regularities decomposed by NMF for DOs.

#### A. Results from the Low-rank Regularity

In our experiments, the rank of  $\mathbf{B}$  is automatically determined as 20. It means that there are 20 types of regularities in a week. To better visualize these regularities, the solved  $\mathbf{B}$  was factorized by NMF as follows:

$$\mathbf{B} = \mathbf{W}^\top \mathbf{R},$$

where  $\mathbf{W}$  ( $\mathbf{W} \in \mathbb{R}^{672 \times 20}$ ) is the nonnegative weight matrix, and  $\mathbf{R}$  ( $\mathbf{R} \in \mathbb{R}^{20 \times 40,000}$ ) is considered as the 20 types of regularities. To identify the importance of these types of regularities, the energy of the  $i$ -th type of regularity is defined

as follows:

$$\text{energy}(i) \propto \sum_j w_{ji}/Z, \quad (14)$$

where  $Z$  is a normalization factor to guarantee that  $\sum_{i=1}^{20} \text{energy}(i)=1$ . Fig. 6 illustrates that the distribution of energies for both PUs and DOs. Therefore, in both Figs. 4 and 5, we visualize these types of regularities whose accumulated energies are larger than 0.9.

As unexpected, these types of regularities seemingly are visually similar to each other in Fig. 4. Considering that each pixel represents an area with 0.023 square kilometers, if

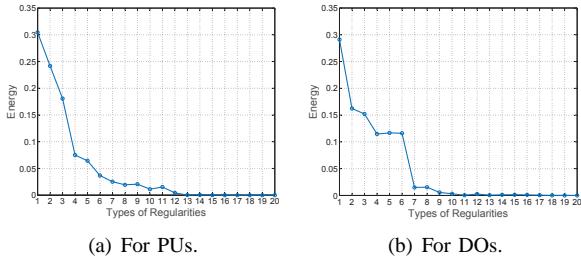


Fig. 6: The distribution of energies for different types of regularities.

readers enlarge these figures, the shapes of these regularities are indeed different to each other; besides, the local intensities are also different, although the local shapes are similar to each other. It means that the travelling requirement across different time slices overall tend to follow the same pattern in Beijing. Moreover, there are two different observations in Figs. 4 and 5:

1. The regions with the intensive PUs/DOs well correspond to the physical-explainable spots. For instance, Figs. 4 and 5 perfectly correspond to the stations, airports, and business areas. It is very useful to understand these visually explainable results in practice.
  2. Interestingly, compared with PUs in Fig. 4, the regularities of DOs are visually more diverse than that of PUs. For example,  $R_2^d$  is far more dissimilar to  $R_1^d$  than  $R_3^d$ . It indicates that the destination of travelling is time-sensitive behaviors in Beijing.

To further understand these regularities, Fig. 7 illustrates the temporal distribution of these regularities for both PUs and DOs. More concretely, in Fig. 7, every point represents one type of regularity from the time slices \*:00 – \*:15, where \* ranges from 0 to 23. The  $j$ -th regularity is selected by the winner-take-out principle as follows:

$$j \leftarrow \max_j\{w_{ij}\}, j = 1, \dots, 20.$$

As expected, the temporal distribution of these regularities illustrates the repeat patterns. More specially, for both PUs and DOs, the patterns between the weekdays (from Monday to Friday) and the weekends (Saturday and Sunday) is nearly similar to each other. For the working days, the transition of temporal patterns from PUs mainly follows “ $R_1^p \curvearrowright R_2^p \curvearrowright R_5^p \curvearrowright R_1^p \curvearrowright R_2^p \curvearrowright R_1^p$ ”, while the transition of patterns for the weekends is relative simple, e.g., “ $R_1^p \curvearrowright R_2^p \curvearrowright R_3^p \curvearrowright R_6^p \curvearrowright R_1^p$ ”. The repetition of the decomposed regularities indicates that a city spatio-temporally follows some hidden rules.

Therefore, our method supplies a new perspective to sense a city in a fine-grained scale, where more application-oriented algorithms can be readily performed on these regularities, such as, DBSCAN for land usage [31],  $k$ -means for passenger hunter [18].

### B. Results from the Factorized Disparity

- 1) Visualization of Basis Patterns:* These regularity-removed intensities can be considered as the “hot spots” in

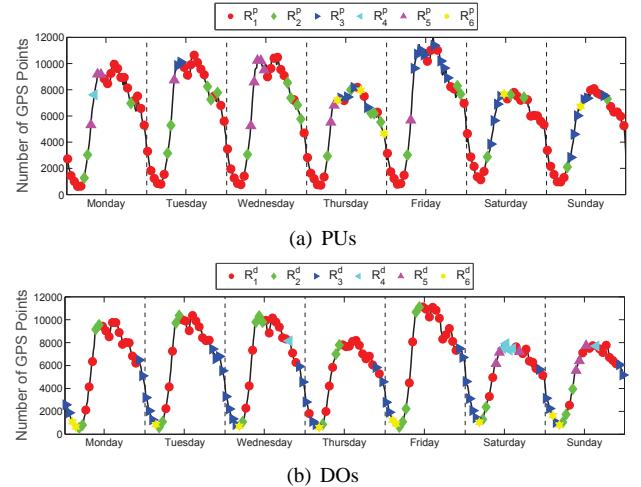


Fig. 7: Visualizing the temporal distribution of the different types of regularities (best viewed in color).

a city. We graphically depict basis patterns  $\mathbf{p}_k$  for  $K=10$  in both Figs. 8 and 9. Considering the point-to-area mapping and the local intensity difference, there is a wide diversity among different basis patterns. For instance, although  $\mathbf{p}_3^p$  and  $\mathbf{p}_4^p$  are seemingly similar, in fact,  $\mathbf{p}_3^p$  represents the intercity coach station (“Liuliqiao station”) which  $\mathbf{p}_4^p$  means the railway station (“Beijing west railway station”). It means that our method successfully discovers the spatially-closed traffic hubs. That is, these basis patterns discovered by our method correspond to the physical-explainable spots.

Moreover, some basis patterns represent a single hot spot; whereas others represent the co-occurred multiple hot spots in Figs. 8 and 9. It is natural that a hot spot simultaneously occurs in different types of disparities.

Interestingly, these multiple hot spots in Figs. 8(e), and (h) (or Figs. 9 (d), (f), and (i)) mean that both the orientations and the destinations of trips occur simultaneously. These patterns with the multiple simultaneously occurred hot spots means that the activity is naturally for some special time slices, *e.g.*, the evening peak.

2) *Spatial Patterns among Time Slices*: The weights specifically provide a concise characterization of these disparities among different time slices. The weight  $c_{ik}$  can be interpreted as the amount time slice  $i$  takes basis patterns  $\mathbf{p}_k$ . Table III compares the normalized weights for a set of the representative time slices. For example, the morning peak (8:00-8:15 a.m.), and the evening peak (5:00-5:15 p.m.).

From these weights of basis patterns, there are some interesting observations:

1. These nonnegative weights of basis patterns are efficient to understand the human activities in a time slice. For instance, the slice 3:00-3:15 a.m. in Table. 8 is mainly described by the pattern  $p_{10}^p$ . It means that this area is very activate, compared with the other time slices. However, for this basis pattern  $p_{10}^p$  at the different time slice (i.e., 10:00-10:15 pm), the corresponding weight is zero. It means that this spatial pattern is not active. As a contrast,  $p_3^p$ ,  $p_4^p$ , and  $p_8^p$  receive the top-3 weights, dis-

TABLE III: Visualization of normalized basis patterns for the several representative time slices.

Basis patterns of PUs	$\mathbf{P}_1^p$	$\mathbf{P}_2^p$	$\mathbf{P}_3^p$	$\mathbf{P}_4^p$	$\mathbf{P}_5^p$	$\mathbf{P}_6^p$	$\mathbf{P}_7^p$	$\mathbf{P}_8^p$	$\mathbf{P}_9^p$	$\mathbf{P}_{10}^p$
3:00-3:15 a.m.	0	0.0325	0	0	0	0.0097	0.0779	0	0	<b>0.8800</b>
8:00-8:15 a.m.	0	0.0794	0	<b>0.1984</b>	0.0218	0.0266	0.0165	<b>0.4081</b>	0	<b>0.2490</b>
0:00-0:15 p.m.	<b>0.3574</b>	0.0377	0	0.1163	<b>0.3523</b>	0.1134	0.0116	0.0112	0	0
5:00-5:15 p.m.	<b>0.1995</b>	<b>0.1947</b>	<b>0.1460</b>	0.0872	<b>0.2021</b>	0	0.0222	0.0704	0.0779	0
10:00-10:15 p.m.	0.1367	0	<b>0.2136</b>	<b>0.2276</b>	0	0.0525	0.0305	<b>0.3128</b>	0.0264	0
Basis patterns of DOs	$\mathbf{P}_1^d$	$\mathbf{P}_2^d$	$\mathbf{P}_3^d$	$\mathbf{P}_4^d$	$\mathbf{P}_5^d$	$\mathbf{P}_6^d$	$\mathbf{P}_7^d$	$\mathbf{P}_8^d$	$\mathbf{P}_9^d$	$\mathbf{P}_{10}^d$
3:00-3:15 a.m.	0.0500	0	0	<b>0.2654</b>	0.0828	<b>0.6018</b>	0	0	0	0
8:00-8:15 a.m.	<b>0.3066</b>	0	0.0543	0	0.0658	0.0702	0.0316	<b>0.4169</b>	0.0545	0
0:00-0:15 p.m.	<b>0.4300</b>	0.1167	0.0475	<b>0.3505</b>	0	0.0006	0.0493	0	0	0.0054
5:00-5:15 p.m.	0	0.1408	<b>0.1830</b>	0	<b>0.3290</b>	<b>0.2141</b>	0.1236	0	0.0097	0
10:00-10:15 p.m.	0	0.0248	0.0811	<b>0.2545</b>	0	0.0582	0.0517	<b>0.2475</b>	<b>0.2822</b>	0

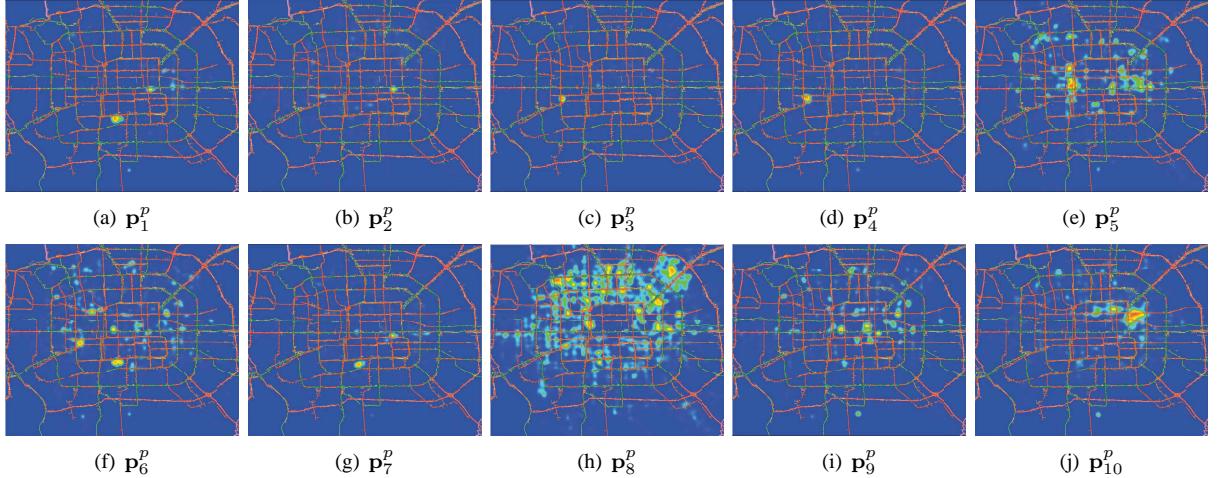


Fig. 8: Visualization of 10 basis patterns discovered from the disparity of PUs.

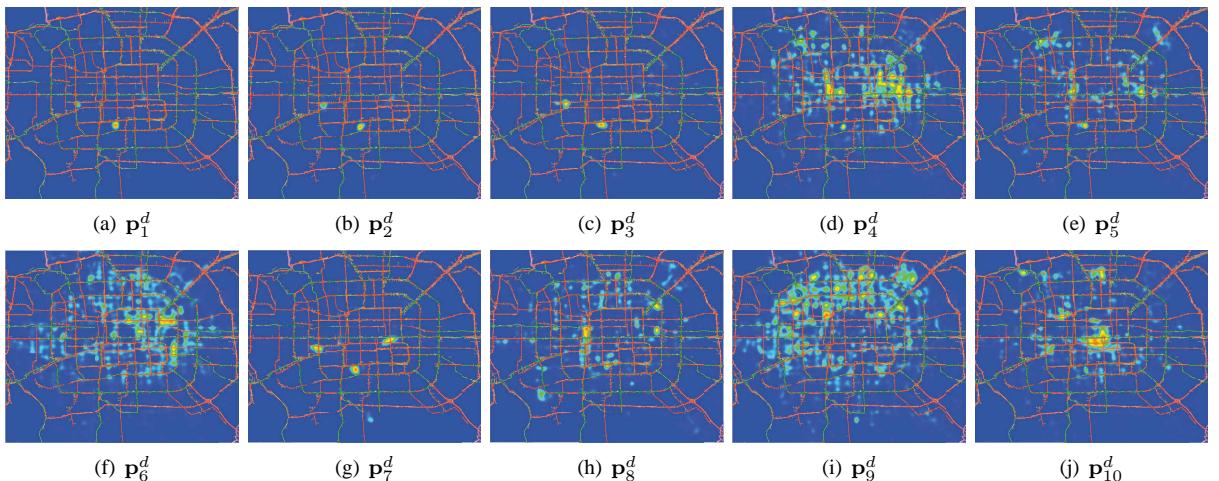


Fig. 9: Visualization of 10 basis patterns discovered from the disparity of DOs.

covering that PUs were occurred from the railway station (“Beijing West Railway Station”), the intercity bus station

(“Liuliqiao Coach Station”), and the scattered-dot pattern across the north of the city. It is very consistent to the

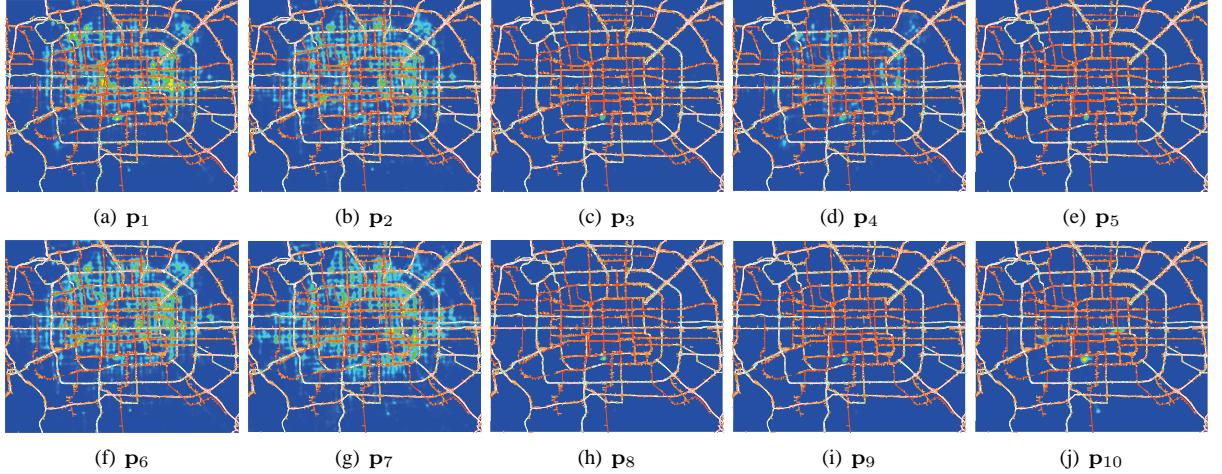


Fig. 10: Visualization of 10 factorized patterns discovered from DOs by FDC.

common knowledge: (a) it is common for people to occur at stations; and (b) the unbalanced travel requirement corresponds to the fact that the development of the north of Beijing is better than that of the south one.

2. By comparing the weights of different time slices, a quantitative understanding of the difference among time slices is possible. For example, if we want to analyze the difference between the evening peak (5:00-5:15 p.m.) and the nightlife time (10:00-10:15 p.m.), we can observe two interesting phenomena: (a)  $p_3^p$  is shared by two time slices; it means that this spot is very active to generate PUs (in fact, it is a railway station); and (b) the most active pattern  $p_8^p$  from the nightlife time means that nightlife mainly occurs in these areas; while  $p_5^p$  (the finance-related business at both “Jingrong Street” area and “Guomao” one) from the evening peak indicates that people in a few spots were getting off duty at this time interval.

### C. Comparisons between Factorizing on Discretized Counts

A naïve approach is applied the matrix Factorization method on Discretized Counts (FDC). Concretely, the number of PUs/DOs is firstly counted in every discretized spatial bin, and then NMF is applied on the temporal-spatial matrix. Compared with our approach, without point process, regularity and disparity decomposition, FDC just factorizes a raw temporal-spatial matrix into a set of low-dimensional subspaces. Following this setting, a temporal-spatial matrix  $\widehat{\Lambda} \in \mathbb{R}^{672 \times 40000}$  (where an entry is the number of PUs/DOs) is factorized by NMF as,  $\min_{C \geq 0, P \geq 0} \|\widehat{\Lambda} - C^\top P\|_F^2$ . Note that the scheme of FDC is a special case of the tensor factorization approach [36].

Compared with basis patterns from the disparity of DOs in Fig. 9, Fig. 10 illustrates that this scheme barely discovers the physical-explainable patterns. It is impossible nearly all areas in a city are hot spots in  $p_1, p_2, p_6$ , and  $p_7$  from FDC in Fig. 10, although  $p_3, p_5, p_8$  and  $p_9$  from FDC in Fig. 10 are very spatial sparse. As a contrast, Fig. 9 generates more sparser hot spots than FDC. For instance, it is seemingly



Fig. 11: An example of how to allocate law enforcers (best viewed in color).

strange that the hot spots from  $p_{10}^d$  in Fig. 9 dominate the north of the city. This perfectly corresponds to the fact that the development of the north of Beijing is better than that of the south area, especially than that of the east-south corner. Meanwhile,  $p_2, p_6$ , and  $p_7$  from FDC in Fig. 9 do not reflect this fact. These comparisons indicate that it is necessary to decompose the regularity and the disparity from the raw data.

### D. An Application with Discovered Patterns

A practical application from the discovered patterns is how to efficiently allocate law enforcers with the aim at inspecting illegal behaviors (*e.g.*, cheating on taximeters) in taxi services. Intuitively, the limited number of law enforcers should be allocated to these spots where passengers have been *intensively* yet *persistently* dropped off or picked up. In terms of our approach, persistency is measured by the frequency of the occurrence of the regularities  $R_*$  in the different time slices. Figs. 4 and 5 indicate that  $R_1^p$  and  $R_1^d$  temporally discover the persistent patterns. While, intensity is measured by the probability of DOs (PUs) visually depicted by the heat map.

Interestingly, except these well-understood spots (*i.e.*, intercity bus stations, railway station, airports, and centralized-business areas), some spots can be manually screened in Fig. 11. For instance, there are many information technology companies in the encircled spot (“Zhongguan Chun”) with

TABLE IV: Running time of each component in our approach.

Component	LGCP	NLSD	NMF
Time(s.)	372.12	165.08	9.04

about 2 squared kilometers. Moreover, thanking for the interpolation ability of LGCP, we can precisely recommend the law enforcers a special location by finding a GPS position with the maximum probability.

#### E. Efficiency

To give readers an intuition about the efficiency, we report the empirical running time for each component. Our experiments are conducted on a server (Windows) with Intel Core CPU with 3.2 GHz and 32 GB main memory. The source code was implemented with Matlab. The NMF was implemented by multiplicative update approach [17]. In optimizing NMF, the stop condition is that the squared Frobenius norm is less than  $1e-5$ . Note that how to fast optimize NMF is out of this paper. Table. IV illustrates that the running time of each component.

Although LGCP nearly costs the 68% computational cost, LGCP can be efficiently optimized by the advanced method, such as, the variational method [23] with a computational complexity  $O(N)$  where  $N$  is the number of points. NLSD can be efficiently optimized by accelerated gradient method [15] with a convergence rate  $O(1/T^2)$ , where  $T$  is the number of iterations. In summary, this paper discusses a roadmap on how to obtain fine-grained patterns from point data. Three key components would be improved with more advanced methods, e.g., [19] [15] [23].

## VI. CONCLUSIONS

In this paper, we presented a new approach that decompose the point intensities into the regularity and the disparity for fine-grained analysis. A key component of this approach is point intensity processes that handle sparse data to uncover fine-grained spatial patterns. We have shown that this approach uncovers interesting spatio-temporal patterns. Another important contribution of this work is factorizing the nonnegative sparse disparity into a set of comparable basis patterns. Therefore, the differences among time slices are quantitatively analyzed. We have present a series of experiments by a case study, using a large data set from over 53,000 taxis in Beijing, which illustrates the capabilities and effectiveness of our approach in sensing a city.

There are several avenues for future work. Our approach just decomposes point intensities into both the regularity and the disparity, supplying a new tool for different applications. Therefore, one possible direction is to combine some application-oriented tasks with our method, such as, finding time-aware hot spots for the passenger hunter task [18]. In addition, we would like to extend this approach into the moment mapping task. Our plan is to extend NMF into the tensor decomposition to allow for such paired data. This paper only requires the nonnegative constraints in factorizing the disparity patterns. How to factorize the disparity into more explainable spatial patterns is also one of interesting directions.

Furthermore, we intend to empirically examine our method on other type of point data, such as, automatic fair collection data from the subway.

## REFERENCES

- [1] NYC Open Data. [Online]. Available: <https://nycopendata.socrata.com>
- [2] S. Basu and A. Dassios, "A cox process with log-normal intensity," vol. 31, no. 2, pp. 297–302, 2002.
- [3] C. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
- [4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [5] J. Cai, J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [6] L. Cao and J. Krumm, "From gps traces to a routable road map," in *ACM GIS*, 2009, pp. 3–12.
- [7] C. Chen, D. Z. N. Li, and Z. Zhou, "B-planner: Planning bidirectional night bus routes using large-scale taxi gps traces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1451–1465, 2014.
- [8] D. Cox, "Some statistical methods connected with series of events," *Journal of the Royal Statistical Society Series B*, vol. 17, no. 2, pp. 129–164, 1955.
- [9] J. Cunningham, B. Yu, K. Shenoy, and M. Sahani, "Inferring neural firing rates from spike trains using gaussian processes," in *Advances in Neural Information Processing Systems*, vol. 20, 2008, pp. 89–98.
- [10] L. Ding, H. Fan, and L. Meng, "Understanding taxi driving behaviors from movement data," in *AGILE on Geographic Information Science. Lisbon*, 2015.
- [11] L. Ding, M. Jahnke, S. Wang, and K. Karja, "Understanding spatiotemporal mobility patterns related to transport hubs from floating car data," in *International Conference on Location-Based Services*, 2016, pp. 175–185.
- [12] D. Donoho, "For most large underdetermined systems of linear equations, the minimal  $\ell_1$ -norm solution is also the sparsest solution," *Technical report, Stanford University*, 2004.
- [13] D. Guo, X. Zhu, H. Jin, P. Gao, and C. Andris, "Discovering spatial patterns in origin-destination mobility data," *Transactions in GIS*, vol. 16, no. 3, pp. 411–429, 2012.
- [14] J. Heikkilä and E. Arjas, "A nonparametric bayesian approach," vol. 55, no. 3, pp. 738–745, 1999.
- [15] S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *International Conference on Machine Learning*, 2009, pp. 457–465.
- [16] J. Kingman, *Poisson Processes*. Oxford University Press, 1992.
- [17] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6775, pp. 788–791, 1999.
- [18] B. Li, D. Zhang, L. Sun, and C. Chen, "Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset," in *IEEE International Conference on Pervasive Computing and Communications Workshops*, 2011, pp. 63–68.
- [19] L. Li, L. Wu, H. Zhang, and F. Wu, "A fast algorithm for nonnegative matrix factorization and its convergence," in *IEEE Trans. on Neural Networks and Learning Systems*, vol. 25, no. 10, 2014, pp. 1855–1863.
- [20] X. Li, G. Pan, Z. Wu, G. Qi, S. Li, D. Zhang, W. Zhang, and Z. Wang, "Prediction of urban human mobility using large-scale taxi traces and its application," *Journal of Frontiers Computer Sciences*, vol. 6, no. 1, pp. 111–121, 2012.
- [21] X. Liang, X. Zheng, W. Lv, T. Zhu, and K. Xu, "The scaling of human mobility by taxi is exponential," *Physica A: Statistical mechanics and its applications*, vol. 39, no. 5, pp. 2135–2144, 2012.
- [22] S. Liu, Y. Liu, L. Ni, J. Fan, and M. Li, "Towards mobility-based clustering," in *ACM Conference on Knowledge Discovery and data Mining*, vol. 15, no. 4, 2010, pp. 919–928.
- [23] C. Lloyd, T. Gunter, M. Osborne, and S. Roberts, "Variational inference for gaussian process modeled poisson processes," in *International Conference on Machine Learning*, vol. 37, 2015, pp. 89–98.
- [24] X. Luo, M. Zhou, H. Leung, Y. Xia, Q. Zhu, Z. You, and S. Li, "An incremental-and-static-combined scheme for matrix-factorization-based collaborative filtering," *IEEE Trans. on Automation Science and Engineering*, vol. 13, no. 1, pp. 333–343, 2016.

- [25] X. Luo, M. Zhou, S. Li, Z. You, Y. Xia, and Q. Zhu, "A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 27, no. 3, pp. 579–592, 2016.
- [26] X. Luo, M. Zhou, M. Shang, S. Li, and Y. Xia, "A novel approach to extracting non-negative latent factors from big sparse matrices," *IEEE Access*, vol. 4, pp. 2649–2655, 2016.
- [27] X. Luo, M. Zhou, Y. Xia, Q. Zhu, A. Ammari, and A. Alabdulwahab, "Generating highly accurate predictions for missing qos data via aggregating nonnegative latent factor models," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 27, no. 3, pp. 524–537, 2016.
- [28] Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," arxiv.org/abs/1208.3922, 2012, [Online].
- [29] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, p. 891C909, 2008.
- [30] J. Møller, A. Syversveen, and R. Waagepetersen, "Log gaussian cox processes," *Scandinavian Journal of Statistics*, vol. 25, no. 3, pp. 451–482, 1998.
- [31] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li, "Land-use classification using taxi gps traces," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 113–123, 2013.
- [32] C. Peng, X. Jin, K. Wong, M. Shi, and P. Li, "Collective human mobility pattern from taxi trips in urban area," *PLoS ONE*, vol. 7, no. 4, pp. 151–158, 2012.
- [33] C. Rasmussen and C. Williams, *Gaussian Process for Machine Learning*. MIT Press, Cambridge, Massachusetts, 2006.
- [34] Y. Shen, L. Zhao, and J. Fan, "Analysis and visualization for hot spot based route recommendation using short-dated taxi gps traces," *Information*, vol. 6, no. 2, pp. 134–151, 2015.
- [35] C. Stolte, D. Tang, and P. Hanrahan, "Polaris: A system for query, analysis and visualization of multidimensional relational databases," vol. 8, no. 1, pp. 52–65, 2002.
- [36] L. Sun and K. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transportation Research Part B: Methodological*, vol. 91, pp. 511–524, 2016.
- [37] P. Widhalm, Y. Yang, M. Ulm, S. Athavale, and M. Gonzalez, "Discovering urban activity patterns in cell phone data," *Transportation*, vol. 42, no. 4, pp. 597–623, 2015.
- [38] J. Wood, J. Dykes, and A. Slingsby, "Visualization of origins, destinations, and flows with od maps," *Cartographic Journal*, vol. 47, no. 2, pp. 117–129, 2010.
- [39] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *ACM Conference on Geographic Information Systems*, 2010, pp. 99–108.
- [40] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger?" in *International conference on ubiquitous computing*, 2011, pp. 109–118.
- [41] Y. Zhang, Y. Liu, J. Yun, and X. Xie, "Urban computing with taxicabs," in *International Conference on UbiComputing*, 2011, pp. 89–98.



**Junbiao Pang** received the B.S.degree and the M.S. degree in computational fluid dynamics and computer science from the Harbin Institute of Technology, Harbin, China, in 2002 and 2004, respectively, and the Ph.D. from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2011.

He is currently an Associate Professor with the College of Metropolitan Transportation, Beijing University of Technology (BJUT), Beijing, China. He has authored or coauthored approximately 20 academic papers in publications such as the IEEE TRANSACTIONS ON IMAGE PROCESSING, ECCV, ICCV, and ACM *Multimedia*. His research interests include multimedia and machine learning for transportation problem.



**Jing Huang** received the B.S. degree in computer science and technology from North China University of Technology, Beijing, China, in 2014, and is currently working towards the M.S. degree in computer science and technology at Beijing University of Technology, Beijing, China.

Her research interests include machine learning for transportation problem, and information retrieval.



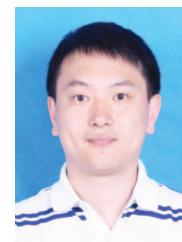
**Xue Yang** received the B.S. degree and the M.S. degree in computer science and technology from the North China University of Water Resources and Electric Power, Zhengzhou, China, in 2001, and the Beijing University of Posts and Telecommunications, Beijing, China, in 2005, respectively.

She is currently a senior engineer at the Beijing Transportation Information Center, Beijing, China. Her research interests include traffic data analysis for the transportation problem.



**Zuyun Wang** received the B.S. degree in Software Engineer from the Capital Normal University, Beijing, China, in 1992, and the M.S. degree in Computer Application Technology from the Beijing University of Technology, Beijing, China, in 2003.

She is currently the director of the Information Center division of the Beijing Municipal Transportation Law Enforcement Corps, Beijing, China. Her research interests include the application of key technologies for intelligent transportation systems.



**Haitao Yu** received the B.S. degree in management information systems from the Beijing Information Science and Technology University, Beijing, China, in 2005, and the M.S. degree in computer software from the Beihang University, Beijing, China, in 2009, and is currently working towards the Ph.D. degree in computer science and technology at Beihang University, Beijing, China.

He is currently the associate dean with the Beijing Transportation Information Center, Beijing, China. His research interests include traffic data analysis for the key technologies in traffic information service.



**Qingming Huang** ( M'04 - SM'08) received the B.S. degree in computer science and the Ph.D. degree in computer engineering from Harbin Institute of Technology, Harbin, China, in 1988 and 1994, respectively.

He is currently a Professor with the University of the Chinese Academy of Sciences (CAS), Beijing, China, and an Adjunct Professor with the Institute of Computing Technology, CAS, China. He has authored or coauthored more than 300 academic papers in prestigious international journals including the IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE TRANSACTIONS ON IMAGE PROCESSING, and top-level conferences such as ACM Multimedia, CVPR, AAAI, IJCAI and VLDB. His research interests include multimedia computing, image processing, computer vision, pattern recognition, and machine learning.



**Baocai Yin** received the M.S. and Ph.D. degree in computational mathematics from Dalian University of Technology, Dalian, China, in 1988 and 1993, respectively.

He is currently a Professor with the Department of Electronic Information and Electrical Engineering, Dalian University of Technology. He is also the Director of Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing, China. He has authored or coauthored more than 200 academic papers in prestigious international journals including the IEEE TRANSACTIONS ON MULTIMEDIA, and top-level conferences such as INFOCOM and ACM SIGGRAPH. His research areas include multimedia, image processing, computer vision, and pattern recognition.