Understanding Urban Dynamics via Context-Aware Tensor Factorization with Neighboring Regularization

Jingyuan Wang, Junjie Wu, Ze Wang, Fei Gao, and Zhang Xiong

Abstract—Recent years have witnessed the world-wide emergence of mega-metropolises with incredibly huge populations. Understanding residents mobility patterns, or urban dynamics, thus becomes crucial for building modern smart cities. In this paper, we propose a Neighbor-Regularized and context-aware Non-negative Tensor Factorization model (NR-cNTF) to discover interpretable urban dynamics from urban heterogeneous data. Different from many existing studies concerned with prediction tasks via tensor completion, NR-cNTF focuses on gaining urban managerial insights from spatial, temporal, and spatio-temporal patterns. This is enabled by high-quality Tucker factorizations regularized by both POI-based urban contexts and geographically neighboring relations. NR-cNTF is also capable of unveiling long-term evolutions of urban dynamics via a pipeline initialization approach. We apply NR-cNTF to a real-life data set containing rich taxi GPS trajectories and POI records of Beijing. The results indicate: 1) NR-cNTF accurately captures four kinds of city rhythms and seventeen spatial communities; 2) the rapid development of Beijing, epitomized by the CBD area, indeed intensifies the job-housing imbalance; 3) the southern areas with recent government investments have shown more healthy development tendency. Finally, NR-cNTF is compared with some baselines on traffic prediction, which further justifies the importance of urban contexts awareness and neighboring regulations.

Index Terms—Urban dynamics, tensor factorizations, urban planning, spatio-temporal pattern, GPS trajectory

1 Introduction

As reported by the World Bank,[1] at the end of 2016 more than 53 percent population of the world, i.e., about 3.7 billion people, lived in cities; about 36 mega-metropolises worldwide had a population of more than 10 million. Huge urban populations bring great challenges such as traffic jams, educational/medical resource scarcity, environmental pollution, etc. Understanding the behavioral patterns of residents in a city, or *urban dynamics* for short, therefore becomes an important yet urgent demand for urban planning and public policy making from a smart city perspective. Fortunately, the widely adopted *mobile crowd sensing* (MCS) technologies [1], such as GPS, mobile phones, and location-based services, give us an unprecedented opportunity to access to enormous and perhaps unbounded human mobility data, which combined with urban infrastructure data offer a "rich ore" for discovery of urban dynamics.

In general, mining urban dynamics from MCS data has three requirements. The first one is to model *multi-source heterogeneous data*, which consist of mobility records of residents such as the origins and destinations, the travel time, the purposes, and the surroundings hidden in different data sources such as GPS trajectories, urban contexts, and city maps. The second requirement is to capture *long-term evolutions*, which is critically important for urban planners to understand the evolving rules of cities so as to make proper urban planning. The last one is to find urban dynamics with *good interpretability*—an obscure urban dynamic is useless to decision making in real-world application scenarios. Despite of rich literature in applying matrix/tensor factorizations to model urban heterogeneous data, most of them aim to generate patterns to improve the predictive accuracy of traffic volumes [2], [3], [4], but leave pattern explanation to luck. It is not until recently that a few works begin to take the understanding of urban dynamics as the primary research task, and the representative ones include the earlier rNTD model using Tucker factorizations [5], the city spectrum modeling using CP factorizations [6], and still some using single source data [7], [8], [9] or for discovering urban functional zones only [10], [11]. These excellent works, however, cannot meet all the above-mentioned requirements simultaneously.

In this paper, we propose a **N**eighbor-**R**egularized **c**ontext-aware **N**on-negative **T**ensor **F**actorization model (NR-cNTF)

1. http://data.worldbank.org/

- *J. Wang is with the Beijing Advanced Innovation Center for Big Data and Brain Computing, and with School of Computer Science and Engineering, Beihang Unversity, Beijing 100191, China. E-mail: jywang@buaa.edu.cn.*
- *J. Wu is with the School of Economics and Management, and with the Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beihang University, Beijing 100191, China. E-mail: wujj@buaa.edu.cn.*
- *Z. Wang, and Z. Xiong are with the MOE Engineering Research Center of Advanced Computer Application Technology, and with the School of Computer Science and Engineering, Beihang Unversity, Beijing 100191, China. E-mail: {ze.w, xiongz}@buaa.edu.cn.*
- *F. Gao is with the Microsoft Research Asia, Beijing 100080, China. E-mail: feiga@microsoft.com.*

Manuscript received 20 Sept. 2017; revised 1 Aug. 2018; accepted 22 Apr. 2019. Date of publication 0 . 0000; date of current version 0 . 0000.
(Corresponding author: Junjie Wu.)
Recommended for acceptance by X. Lin.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2019.2915231

to discover explainable and evolving urban dynamics from multi-source heterogeneous urban data. In the NR-cNTF model, we introduce the concepts of data space and pattern space and describe the relations between urban data and urban dynamics. The Tucker factorization is then introduced with the POI-based (Point-Of-Interests) urban contexts to factorize the Origin-Destination-Time (ODT) tensor into spatial, temporal, and spatio-temporal patterns of good interpretability. Moreover, a neighboring regularization that incorporates geographically neighboring relations is introduced into our model to further improve the explainability of spatial patterns. Finally, a simple yet effective pipeline initialization approach is designed to capture the long-term evolutions of urban dynamics.

We conduct extensive experiments on a real-life data set that contains the GPS trajectories of over 20,000 taxies and over 400,000 POI records of Beijing from 2008 to 2015. The first scenario of the experiments is to verify the ability of NR-cNTF in disclosing true urban dynamics and obtain managerial insights via NR-cNTF. The results indicate that: 1) NR-cNTF accurately captures four kinds of mobility rhythms and seventeen spatial communities of Beijing; 2) the rapid development of Beijing in the CBD area, is indeed at the expense of severer job-housing imbalance and therefore is unsustainable in a long run; 3) the southern areas of Beijing are experiencing unprecedented growth with the recent government investments, and most importantly they have shown more healthy development tendency. The second scenario of the experiments is to testify the prediction power of NR-cNTF, which is compared with some baselines on traffic prediction. The results demonstrate the superiority of NR-cNTF in tensor completion, which further justifies the importance of adopting urban contexts and neighboring regulations in NR-cNTF.

2 Problem Formulation

In this section, we formulate *urban dynamics discovery* as a context-aware tensor factorization problem. Table 1 lists the math variables to be used, which are divided into two categories, i.e., data-space variables and pattern-space variables, according to their observability. Variables in the data space are observable from real-world human mobility, while variables in the pattern space are latent but crucial for understanding urban dynamics.

Throughout the paper, we use lowercase symbols such as a, b to denote scalars, bold lowercase symbols such as $\mathbf{a}$, $\mathbf{b}$ for vectors, bold uppercase symbols such as $\mathbf{A}$, $\mathbf{B}$ for matrices, and calligraphy symbols such as $\mathcal{A}$, $\mathcal{B}$ for tensors.

Data-Space Variables. The primary variable in data space is a *data tensor*. Assume there are M urban zones in a city, and N time slices in a day. Let r_{xyz} denote the *resident travel intensity* from an origin zone $x \in \{1,\ldots,M\}$ to a destination zone $y \in \{1,\ldots,M\}$ within a time slice $z \in \{1,\ldots,N\}$. A third-order tensor $\mathcal{R} \in \mathbb{R}^{M\times M\times N}$ is then defined by having r_{xyz} as the (x,y,z) element. Intuitively, $\mathcal{R}$ contains the original information about urban dynamics, which can be obtained from urban vehicle and resident trajectory data. Another variable in data space is an *urban-context similarity matrix* $\mathbf{W} \in \mathbb{R}^{M\times M}$. The (p,q) element of $\mathbf{W}$, i.e., w_{pq}, is a coefficient that describes the similarity between urban zones p and q using, e.g., points of interest (POI) data.

TABLE 1
Notation Definition

Space	Variable	Definition
Data Space	$\mathcal{R}$	the data tensor
	r_{xyz}	the (x,y,z) element of $\mathcal{R}$
	$\mathbf{W}$	the urban context matrix
	w_{pq}	the (p,q) element of $\mathbf{W}$
Pattern Space	$\mathcal{C}$	the pattern tensor
	c_{ijk}	the (i,j,k) element of $\mathcal{C}$
	$\mathbf{O},\mathbf{D},\mathbf{T}$	the pattern projection matrices
	$\mathbf{o}_x,\mathbf{d}_x,\mathbf{t}_x$	the xth row vectors of $\mathbf{O},\mathbf{D},\mathbf{T}$
	$\mathbf{o}_{:i},\mathbf{d}_{:i},\mathbf{t}_{:i}$	the ith column vectors of $\mathbf{O},\mathbf{D},\mathbf{T}$
	o_{xi},d_{xi},t_{xi}	the (x,i) elements of $\mathbf{O},\mathbf{D},\mathbf{T}$

Pattern-Space Variables. The variables in pattern space include a *core tensor* and three *pattern projection matrices*. Assume there are I origin spatial patterns (OSP), J destination spatial patterns (DSP), and K temporal patterns (TP) hidden inside the data tensor $\mathcal{R}$. We define $\mathbf{O} \in \mathbb{R}^{M\times I}$ as a spatial projection matrix that projects M origin zones into I OSP's. Similarly, $\mathbf{D} \in \mathbb{R}^{M\times J}$ is defined as another spatial projection matrix that projects M destination zones into J DSP's. The matrix $\mathbf{T} \in \mathbb{R}^{N\times K}$ is a temporal projection matrix that projects N time slices to K TP's. The elements of $\mathbf{O}$, $\mathbf{D}$ and $\mathbf{T}$ are denoted as o_{xi}, d_{yj} and t_{zk}, respectively, indicating the projection intensities from the urban zones x, y and time slice z to OSP i, DSP j and TP k, $1 \le i \le I$, $1 \le j \le J$, $1 \le k \le K$. We define a third-order tensor $\mathcal{C}$ as a core tensor that describes the dynamics of resident travels among temporal and spatial patterns. The (i,j,k) element of $\mathcal{C}$, i.e., c_{ijk}, denotes the intensity of resident travels from OSP i to DSP j within TP k.

2.1 Construction of Data Tensor

We here explain how to construct the data tensor $\mathcal{R}$ using real-life GPS trajectory data of Beijing Taxies. To this end, we first segment the Beijing city map into M urban zones. In the literature, quite a few methods including the grid based, morphology based, road networks based, and administrative boundaries based methods [12], [13] can fulfill this task. Here we adopt a Traffic Analysis Zones (TAZ) map provided by Beijing Municipal Committee of Transport[2] to segment Beijing into $M = 651$ zones. Finally, since resident behaviors in city life are often cyclical every day, we divide one day into $N = 24$ time slices (one hour per slice). The above procedure determines the three modes of $\mathcal{R}$.

We then compute the element values of $\mathcal{R}$. Note that the taxi GPS data are often organized as a set of quintuples in the form as $\langle vid,\ time,\ longitude,\ latitude,\ state\rangle$, where vid is the unique ID of a taxi, $(longitude,\ latitude)$ is the location of the taxi, and $state$ informs whether the taxi is carrying any passengers at time $time$. We first obtain all taxi-based passenger travels by removing the records with "no passengers" state. Then an *origin-destination-time* record is constructed for each travel by picking up the first and last records of the travel and then extracting the origin and destination coordinates and the travel starting time. We collect the travel ODT records of all workdays in a month as a data set. The monthly total amount of travels that depart from TAZ x in time slice z and

2. http://www.bjjtw.gov.cn/

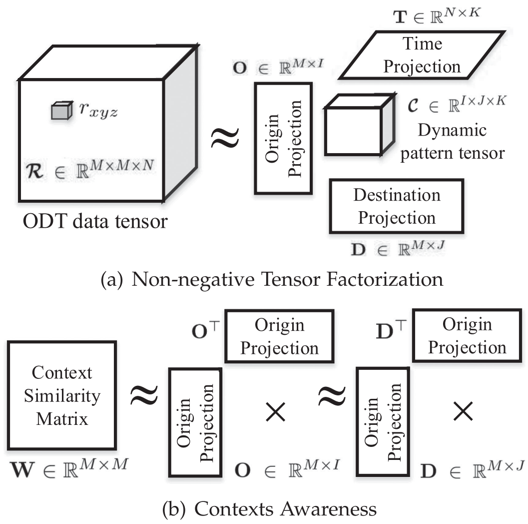


Fig. 1. Model framework of cNTF.

arrive at TAZ y is recorded as $\tilde{r}_{xyz}$. As reported in [14], the travel volumes between different urban zones usually follow a long-tail distribution. Therefore, we adopt the $\log$ function to rescale $\tilde{r}_{xyz}$ as

$$r_{xyz} = \log\left(1 + \tilde{r}_{xyz}\right), \tag{1}$$

which is finally used as the (x, y, z) element of $\mathcal{R}$.

2.2 Definition of Pattern Tensor

Variables in pattern space include $\mathcal{C}$, $\mathbf{O}$, $\mathbf{D}$, and $\mathbf{T}$, where $\mathcal{C}$ is the core tensor that models the dynamic relations among spatio-temporal patterns in the pattern space, and $\mathbf{O}$, $\mathbf{D}$ and $\mathbf{T}$ are the matrices that project the data tensor $\mathcal{R}$ into the core tensor $\mathcal{C}$. To better understand this, we give formal definitions to the spatial and temporal patterns as follows.

Definition 1 (Spatial Pattern). *A spatial pattern is a vector containing the membership score of each urban zone to this pattern. Assume there are I spatial patterns and M urban zones. The ith spatial pattern is denoted as a vector $\mathbf{v}_{:i} = (v_{1i}, \ldots, v_{Mi})^\top$, where v_{mi} is the membership score of the mth zone to the ith spatial pattern. The spatial projection matrix $\mathbf{V}$ that projects M urban zones to I spatial patterns is then defined as $\mathbf{V} = [\mathbf{v}_{:1}, \ldots, \mathbf{v}_{:I}]$.*

The xth row vector of $\mathbf{V}$, denoted as $\mathbf{v}_x$, is a vector that depicts the membership scores of urban zone x to I different spatial patterns. We assign x to spatial pattern i if $i \in \arg\max_{1\leq j\leq I} v_{xj}$. In this way, we can cluster all urban zones into the I spatial patterns. This implies that a spatial pattern is essentially a *spatial community* consisting of urban zones that function similarly in urban dynamics. For example, most of residents in a residential community leave in the morning and return in the evening. In contrast, for a business community, people arrive in the morning and leave in the evening. Spatial patterns can be further divided into origin spatial patterns and destination spatial patterns. The projection matrix $\mathbf{V}$ is denoted as $\mathbf{O}$ for OSP's and $\mathbf{D}$ for DSP's for differentiation. While $\mathbf{O}$ and $\mathbf{D}$ share the same M urban zones, they might have different numbers of spatial patterns.

Definition 2 (Temporal Pattern). *A temporal pattern is a vector containing the membership score of each time slice within a day to this pattern. Assume there are K temporal patterns and N time slices in a day. The kth temporal pattern is denoted as a vector $\mathbf{t}_{:k} = (t_{1k}, \ldots, t_{Nk})^\top$, where t_{nk} is the membership score of the nth time slice to the kth temporal pattern. The temporal projection matrix $\mathbf{T}$ that projects N times slices into K temporal patterns is then defined as $\mathbf{T} = [\mathbf{t}_{:1}, \ldots, \mathbf{t}_{:K}]$.*

In essence, a temporal pattern describes a *temporal rhythm* of urban dynamics, which might correspond to an event that occurs recurrently everyday, e.g., the morning peak and evening peak in a city. Accordingly, the vector $\mathbf{t}_{:k}$ indicates the dynamic intensity of the rhythm k within a day.

Next, we define a pattern tensor to describe the interrelationships among spatio-temporal patterns.

Definition 3 (Pattern Tensor). *A tensor $\mathcal{C} \in \mathbb{R}^{I\times J\times K}$ is a third-order pattern tensor, if its (i, j, k) element c_{ijk} indicates the intensity of resident travels from OSP i to DSP j in TP k, $1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K$.*

Human behaviors in city life usually have synchronism, which can be described by urban dynamic patterns in $\mathcal{C}$. For example, intuitively, residents living in a residential community commute to business regions synchronously in every morning peak of workdays. So an element c_{ijk} has a high value when the origin spatial pattern i corresponds to a residence community, the destination spatial pattern j corresponds to a business community, and the temporal pattern k corresponds to a morning-peak rhythm.

2.3 Definition of Urban Context

Travel behaviors of residents not only have relations with urban spatial and temporal patterns but also have close relations with the so-called *urban context* [11], [15]. Urban context refers to the surroundings inside an urban zone that can affect the travel behaviors of that zone. One typical type of urban context is the so-called *points of interests* including residential buildings, office buildings, shopping malls, etc. We have the following definition.

Definition 4 (Urban-Context Similarity Matrix). *A matrix $\mathbf{W} \in \mathbb{R}^{M\times M}$ is called an urban-context similarity matrix, whose (p, q) element w_{pq} is a coefficient that measures the POI context similarity between zones p and q, $1 \leq p, q \leq M$.*

In general, $\mathbf{W}$ is a nonnegative and symmetric matrix, which could be used to validate the effectiveness of the spatial patterns found purely from trajectory data. For example, it is intuitive that the travel patterns of urban zones with a mass of office buildings should be very similar, but differ sharply from that of zones filled with residential buildings.

2.4 Problem Definition

We here formulate the *urban dynamics discovery* problem as a tensor factorization problem. The model framework is given in Fig. 1, where the ODT data tensor $\mathcal{R}$, pattern tensor $\mathcal{C}$, and projection matrices $\mathbf{O}$, $\mathbf{D}$, and $\mathbf{T}$ have the following relationship:

$$\mathcal{R} = \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T} + \mathcal{E}, \tag{2}$$

where $\mathcal{E} \in \mathbb{R}^{M\times M\times N}$ is a random error tensor, and $\times_n$ denotes the tensor n-mode product. Eq. (2) implies that the resident travel dynamics hidden inside data tensor $\mathcal{R}$ can be well explained by the latent dynamic patterns given by pattern tensor $\mathcal{C}$. The matrices $\mathbf{O}$, $\mathbf{D}$, and $\mathbf{T}$ express the projection relations between $\mathcal{R}$ and $\mathcal{C}$.

Note that while $\mathcal{R}$ is observable from resident travels data, the pattern tensor $\mathcal{C}$ as well as the projection matrices $\mathbf{O}$, $\mathbf{D}$ and $\mathbf{T}$ are unknown variables. Hence, our task is:

- To infer $\mathcal{C}, \mathbf{O}, \mathbf{D}$ and $\mathbf{T}$ from $\mathcal{R}$;
- To understand urban dynamics using $\mathcal{C}, \mathbf{O}, \mathbf{D}, \mathbf{T}$.

The urban-context similarity matrix $\mathbf{W}$ offers additional information to tensor factorization. Recall the row vector $\mathbf{o}_x$ of the projection matrix $\mathbf{O}$, which contains the membership scores of urban zone x to all the OSP's. It is intuitive that similar urban zones should exhibit similar spatial patterns. Hence, we can measure the similarity of zones x and y by simply having $\mathbf{o}_x\mathbf{o}_y^\top$. Analogously, we can also measure the similarity of zones x and y by employing the information of DSP's in $\mathbf{D}$, i.e., $\mathbf{d}_x\mathbf{d}_y^\top$. Since $\mathbf{W}$ evaluates the similarity between x and y as w_{xy} according to the urban context, we finally have the following relationships between $\mathbf{W}$ and projection matrices $\mathbf{O}$ and $\mathbf{D}$

$$\mathbf{W} = \mathbf{O}\mathbf{O}^\top + \mathbf{E}_O, \text{ and } \mathbf{W} = \mathbf{D}\mathbf{D}^\top + \mathbf{E}_D, \tag{3}$$

where $\mathbf{E}_O$ and $\mathbf{E}_D$ are random error matrices. Note that in Eq. (3), $\mathbf{W}$ is an observable variable and $\mathbf{O}$ and $\mathbf{D}$ are latent ones. In other words, we can use urban context to fine-tune OSP's and DSP's in $\mathbf{O}$ and $\mathbf{D}$, respectively.

In summary, Eqs. (2) and (3) together define a context-aware Non-negative Tensor Factorization (cNTF) problem. Our task is to infer urban dynamics given cNTF.

2.5 Extension to Long-Term Evolution

Long-term evolution is an important characteristic of urban dynamics, which refers to the evolution of urban spatial, temporal and spatio-temporal patterns over time. For example, temporal rhythms of resident travels in a city might change with the developments of public transport, economics, migration, etc.

We use *tensor sequence* to describe the evolution of urban dynamics in both data and pattern spaces. In the data space, we define $\mathcal{R}|_{l=1}^L = \{\mathcal{R}_1, \ldots, \mathcal{R}_L\}$ as a data tensor sequence of length L, where $\mathcal{R}_l$ is the data tensor of the lth year. Suppose we factorize $\mathcal{R}_l$ into $\mathbf{O}_l, \mathbf{D}_l, \mathbf{T}_l$ and $\mathcal{C}_l$ according to Eqs. (2) and (3), then we have the pattern tensor sequence $\mathcal{C}|_{l=1}^L = \{\mathcal{C}_1, \ldots, \mathcal{C}_L\}$, and the corresponding projection matrix sequences $\mathbf{O}|_{l=1}^L$, $\mathbf{D}|_{l=1}^L$ and $\mathbf{T}|_{l=1}^L$, respectively.

The problem is, for any two subsequent years l and $l+1$, the patterns inferred from $\mathcal{R}_l$ might not be comparable to that from $\mathcal{R}_{l+1}$, for they are inferred *separately* to optimize the objectives in Eqs. (2) and (3). Therefore, another task of this study is to infer the long-term evolution of urban dynamics given a data tensor sequence.

3 Model

In this section, we reformulate the cNTF problem from a probabilistic perspective, which results in the exact objective function for urban dynamics discovery.

3.1 Probabilistic Non-Negative Tensor Factorization

We assume the random error of observation $\mathcal{E}$ follows a Gaussian distribution: $\mathcal{N}(0, \sigma_{\mathcal{R}}^2)$, then the conditional distribution over the observed entries in $\mathcal{R}$ is defined as

$$\begin{aligned} &P(\mathcal{R}|\mathcal{C}, \mathbf{O}, \mathbf{D}, \mathbf{T}, \sigma_{\mathcal{R}}^2) \\ &= \prod_{x=1}^{M}\prod_{y=1}^{M}\prod_{z=1}^{N} \mathcal{N}(r_{xyz}|\mathcal{C}\times_o \mathbf{o}_x \times_d \mathbf{d}_y \times_t \mathbf{t}_z, \sigma_{\mathcal{R}}^2). \end{aligned} \tag{4}$$

In order to obtain more evident patterns, we should introduce sparse priors to the variables in pattern space. As a result, we adopt zero-mean Laplace priors for projection matrices

$$\begin{aligned} P(\mathbf{O}|\sigma_O) &= \prod_{x=1}^{M} \mathcal{L}(\mathbf{o}_x|\mathbf{0}, \sigma_O\mathbf{I}_I), \\ P(\mathbf{D}|\sigma_D) &= \prod_{y=1}^{M} \mathcal{L}(\mathbf{d}_y|\mathbf{0}, \sigma_D\mathbf{I}_J), \\ P(\mathbf{T}|\sigma_T) &= \prod_{z=1}^{N} \mathcal{L}(\mathbf{t}_z|\mathbf{0}, \sigma_T\mathbf{I}_K), \end{aligned} \tag{5}$$

and assume zero-mean Laplace priors for the pattern tensor

$$P(\mathcal{C}|\sigma_{\mathcal{C}}) = \prod_{x=1}^{I}\prod_{y=1}^{J}\prod_{z=1}^{K} \mathcal{L}(c_{xyz}|0, \sigma_{\mathcal{C}}). \tag{6}$$

Then the posterior distribution of the pattern space variables is given by

$$\begin{aligned} &P(\mathcal{C}, \mathbf{O}, \mathbf{D}, \mathbf{T}|\mathcal{R}, \sigma_{\mathcal{R}}^2, \sigma_{\mathcal{C}}, \sigma_O, \sigma_D, \sigma_T) \\ &\quad = \frac{P(\mathcal{R}|\mathcal{C}, \mathbf{O}, \mathbf{D}, \mathbf{T}, \sigma_{\mathcal{R}}^2)P(\mathcal{C}|\sigma_{\mathcal{C}})P(\mathbf{O}|\sigma_O)P(\mathbf{D}|\sigma_D)P(\mathbf{T}|\sigma_T)}{P(\mathcal{R}|\sigma_{\mathcal{R}}^2)}, \end{aligned} \tag{7}$$

and the log posterior distribution is then calculated by

$$\begin{aligned} &\ln P(\mathcal{C}, \mathbf{O}, \mathbf{D}, \mathbf{T}|\mathcal{R}, \sigma_{\mathcal{R}}^2, \sigma_{\mathcal{C}}, \sigma_O, \sigma_D, \sigma_T) \\ &\propto -\frac{1}{2\sigma_{\mathcal{R}}^2}\sum_{xyz}(r_{xyz} - \mathcal{C}\times_o \mathbf{o}_x \times_d \mathbf{d}_y \times_t \mathbf{t}_z)^2 \\ &\quad -\frac{1}{\sigma_O}\sum_x \|\mathbf{o}_x\|_1 - \frac{1}{\sigma_D}\sum_y \|\mathbf{d}_y\|_1 - \frac{1}{\sigma_T}\sum_z \|\mathbf{t}_z\|_1 \\ &\quad -\frac{1}{\sigma_{\mathcal{C}}}\sum_{xyz}|c_{xyz}|. \end{aligned} \tag{8}$$

Therefore, to obtain the *Maximum A Posteriori (MAP)* estimation of $\mathbf{O}$, $\mathbf{D}$, $\mathbf{T}$ and $\mathcal{C}$ is equivalent to minimizing the object function

$$\begin{aligned} \tilde{\mathcal{J}} = &\frac{1}{2\sigma_{\mathcal{R}}^2}\|\mathcal{R} - \mathcal{C}\times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}\|_F^2 \\ &+ \frac{1}{\sigma_O}\|\mathbf{O}\|_1 + \frac{1}{\sigma_D}\|\mathbf{D}\|_1 + \frac{1}{\sigma_T}\|\mathbf{T}\|_1 + \frac{1}{\sigma_{\mathcal{C}}}\|\mathcal{C}\|_1, \end{aligned} \tag{9}$$

where $\|.\|_F$ is the Frobenius-norm, $\|.\|_1$ is the L1-norm.

TABLE 2
Information of POI Categories

ID	POI category	ID	POI category
1	food & beverage Service	8	education and culture
2	hotel	9	business building
3	scenic spot	10	residence
4	finance & insurance	11	living service
5	corporate business	12	sports & entertainments
6	shopping service	13	medical care
7	transportation facilities	14	government agencies

3.2 Modeling Urban Contexts

We here introduce urban contextual factors into the probabilistic non-negative tensor factorization model. We use a Beijing POI dataset, with the categories given in Table 2.

3.2.1 Computing Urban Contextual Factors

Fig. 2 shows a clear positive correlation between POI quantity and the resident travel volume (including inflow and outflow) for all urban zones of Beijing. Moreover, urban zones in the same community have similar categories of POI's (see Section 3 of *Supplementary Materials*[3], which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TKDE.2019.2915231, for the details). Therefore, we use quantity and categories of POI's in an urban zone to describe urban contextual factors.

Suppose altogether we have H POI categories, and denote n_{ph} as the number of POI's in category h for urban zone p. The fraction of the hth category POI in the zone p is defined as

$$c_{ph} = \frac{n_{ph}}{\sum_{p=1}^{P} n_{ph}}, \tag{10}$$

The fraction of all category of POI in the zone p is then defined as

$$n_p = \frac{\sum_{h=1}^{H} n_{ph}}{\sum_{p=1}^{P}\sum_{h=1}^{H} n_{ph}}, \tag{11}$$

We use the vector $\mathbf{u}_p = (c_{p1}, \ldots, c_{ph}, \ldots, c_{pH}, n_p)^\top$ to describe the POI context of the zone p.

Given the POI context vectors, the similarity of two urban zones p and q can be computed as

$$w_{pq} = \frac{\mathbf{u}_p \cdot \mathbf{u}_q}{\|\mathbf{u}_p\| \cdot \|\mathbf{u}_q\|}, \tag{12}$$

which is the (p, q) element of $\mathbf{W}$.

3.2.2 Incorporating Urban Contextual Factors

Context-aware regularization is an effective tool to fusion contextual information into tensor and matrix factorizations [16], [17]. We introduce urban contextual factors as context-aware regularization using a maximum a posteriori method. Assume the elements of $\mathbf{E}_O$ and $\mathbf{E}_D$ in Eq. (3) follow zero-mean Gaussian distributions, then we have

3. The companion file with the supplementary materials of this paper.

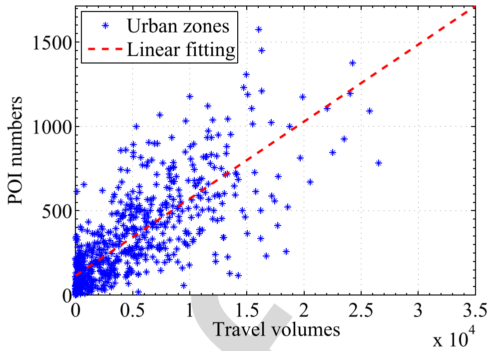


Fig. 2. Validation of urban context correlations.

$$P(\mathbf{W}|\mathbf{O}, \sigma_{WO}^2) = \prod_{p=1}^{M}\prod_{q=1}^{M} \mathcal{N}(w_{pq}|\mathbf{o}_p\mathbf{o}_q^\top, \sigma_{WO}^2), \tag{13}$$

and

$$P(\mathbf{W}|\mathbf{D}, \sigma_{WD}^2) = \prod_{p=1}^{M}\prod_{q=1}^{M} \mathcal{N}(w_{pq}|\mathbf{d}_p\mathbf{d}_q^\top, \sigma_{WD}^2). \tag{14}$$

Let $\Omega = \{\sigma_{\mathcal{R}}^2, \sigma_{WO}^2, \sigma_{WD}^2, \sigma_O, \sigma_D, \sigma_T, \sigma_{\mathcal{C}}\}$. Given the data tensor $\mathcal{R}$ and urban context matrix $\mathbf{W}$, the posterior distribution of $\mathbf{O}$, $\mathbf{D}$, $\mathbf{T}$ and $\mathcal{C}$ is given by

$$\begin{aligned} & P(\mathbf{O}, \mathbf{D}, \mathbf{T}, \mathcal{C}|\mathcal{R}, \mathbf{W}, \Omega) \\ & \propto P(\mathcal{R}|\mathbf{O}, \mathbf{D}, \mathbf{T}, \mathcal{C}, \Omega) P(\mathbf{W}|\mathbf{O}, \Omega) P(\mathbf{W}|\mathbf{D}, \Omega) \\ & \quad P(\mathbf{O}|0, \Omega) P(\mathbf{D}|0, \Omega) P(\mathbf{T}|0, \Omega) P(\mathcal{C}|0, \Omega), \end{aligned} \tag{15}$$

and the log posterior distribution is

$$\begin{aligned} & \ln P(\mathbf{O}, \mathbf{D}, \mathbf{T}, \mathcal{C}|\mathcal{R}, \mathbf{W}, \Omega) \\ & \propto -\frac{1}{2\sigma_{\mathcal{R}}^2}\sum_{xyz}(r_{xyz} - \mathcal{C} \times_o \mathbf{o}_x \times_d \mathbf{d}_y \times_t \mathbf{t}_z)^2 \\ & -\frac{1}{2\sigma_{WO}^2}\sum_{pq}(w_{pq} - \mathbf{o}_p\mathbf{o}_q^\top)^2 - \frac{1}{2\sigma_{WD}^2}\sum_{pq}(w_{pq} - \mathbf{d}_p\mathbf{d}_q^\top)^2 \\ & -\frac{1}{\sigma_O}\sum_x \|\mathbf{o}_x\|_1 - \frac{1}{\sigma_D}\sum_y \|\mathbf{d}_y\|_1 - \frac{1}{\sigma_T}\sum_z \|\mathbf{t}_z\|_1 \\ & -\frac{1}{\sigma_{\mathcal{C}}}\sum_{ijk}|c_{ijk}|. \end{aligned} \tag{16}$$

To maximize the posterior distribution is equivalent to minimizing the sum-of-squared errors function with hybrid quadratic regularization terms, i.e.,

$$\begin{aligned} \min_{\mathbf{O},\mathbf{D},\mathbf{T},\mathcal{C}} \ & \mathcal{J} = \|\mathcal{R} - \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}\|_F^2 \\ & + \alpha\|\mathbf{W} - \mathbf{O}\mathbf{O}^\top\|_F^2 + \beta\|\mathbf{W} - \mathbf{D}\mathbf{D}^\top\|_F^2 \\ & + \gamma\|\mathbf{O}\|_1 + \delta\|\mathbf{D}\|_1 + \epsilon\|\mathbf{T}\|_1 + \varepsilon\|\mathcal{C}\|_1 \\ s.t. \ & \mathbf{O} \geq 0, \mathbf{D} \geq 0, \mathbf{T} \geq 0, \mathcal{C} \geq 0, \end{aligned} \tag{17}$$

where $\alpha = \frac{\sigma_{\mathcal{R}}^2}{\sigma_{WO}^2}$, $\beta = \frac{\sigma_{\mathcal{R}}^2}{\sigma_{WD}^2}$, $\gamma = \frac{2\sigma_{\mathcal{R}}^2}{\sigma_O}$, $\delta = \frac{2\sigma_{\mathcal{R}}^2}{\sigma_D}$, $\epsilon = \frac{2\sigma_{\mathcal{R}}^2}{\sigma_T}$, $\varepsilon = \frac{2\sigma_{\mathcal{R}}^2}{\sigma_{\mathcal{C}}}$. Note that we introduce non-negativity constraints on the variables so as to avoid perplexing negative travel volumes. Eq. (17) indeed formulates the cNTF problem defined in Section 2.4.

$$\begin{tikzcd}
\mathcal{G}_0 \arrow[r, "\text{init}"] & \mathcal{G}_1 \arrow[r, "\text{init}"] & \mathcal{G}_2 \arrow[r, "\text{init}"] & \ldots \mathcal{G}_l \ldots \arrow[r, "\text{init}"] & \mathcal{G}_L \\
 & \mathcal{R}_1\mathbf{W}_1 \arrow[u] & \mathcal{R}_2\mathbf{W}_2 \arrow[u] & \ldots \mathcal{R}_l\mathbf{W}_l \ldots \arrow[u] & \mathcal{R}_L\mathbf{W}_L \arrow[u]
\end{tikzcd}$$

Fig. 3. Pipeline initialization for tensor sequence analysis.

3.3 Neighboring Regularization

Let $\mathcal{SP}_i = \{x : v_{xi} = \max_{1\le j\le I} v_{xj}\}$ denote the ith *urban community* corresponding to the spatial pattern $\mathbf{v}_{:i}$ in the spatial projection matrix $\mathbf{V}$. For the urban zones in $\mathcal{SP}_i$, it is natural to expect that: i) they are geographically neighboring to each other, and ii) their resident mobility behaviors are similar to one another and different from that in other communities. These, however, have not been considered in the above-mentioned cNTF model.

To address these, we here introduce the so-called Neighboring Regularization (NR), which is inspired by the conditional random field based image segmentation method in [18]. Specifically, we model urban community discovery as an image segmentation problem; that is, the community labels of urban zones are modeled as a Markov random field $G(\mathbb{V}, \mathbb{E})$, where $\nu_x \in \mathbb{V}$ is the community label of urban zone x, and $e_{xy} \in \mathbb{E}$ is an undirectional dependency between urban zone x and y. For the latent ν_x, we have an observable matrix $\mathbf{R}_{x::}$ for the origin order of $\mathcal{R}$, or $\mathbf{R}_{:y:}$ for the destination order.

Without loss of generality, in what follows, we use the origin order as an example to introduce the neighboring regularization. Suppose $G(\mathbb{V}, \mathbb{E})$ and $\mathbf{R}_{x::}$, $x \in \{1 \ldots M\}$, satisfy the conditional random field hypothesis. Similar to the classical image segmentation task in [18], the optimization objective for community discovery is to maximize a potential function as

$$\zeta = \sum_{x=1}^{M} \psi_x^u(\nu_x) + \sum_{x=1}^{M} \sum_{y\in M_x} \psi_{xy}^p(\nu_x, \nu_y), \tag{18}$$

where M_x is the set of neighbor zones of zone x. $\psi_x^u(\nu_x)$ is the unary potential of the CRF in zone x when the community label of x is set to ν_x, which is defined as

$$\psi_x^u(\nu_x) = -\log \frac{o_{x\nu_x}}{\sum_{i=1}^{I} o_{xi}}. \tag{19}$$

$\psi_{xy}^p(\nu_x, \nu_y)$ is the pairwise potential between zones x and y when the community labels of x and y are set to ν_x and ν_y, respectively; that is

$$\psi_{xy}^p(\nu_x, \nu_y) = \begin{cases} 0, & \text{if } \nu_x = \nu_y, \\ g(x,y), & \text{otherwise.} \end{cases} \tag{20}$$

Note that $g(x, y)$ is a function of the difference between $\mathbf{R}_{x::}$ and $\mathbf{R}_{y::}$, which is defined as a Gaussian kernel as follows:

$$g(x,y) = \exp\left(-\frac{\left\|\mathbf{R}_{x::} - \mathbf{R}_{y::}\right\|_F^2}{2\sigma_{\mathrm{NR}}^2}\right), \tag{21}$$

where σ_{NR} is a parameter suggested in [18]. This actually introduces a penalty for the zones that are adjacent and have similar resident mobility behaviors but are assigned to different communities.

In a nutshell, Eq. (18) introduces the spatial community discovery problem, which could be regarded as a neighboring regularization to cNTF, and thus form the so-called *NR-cNTF* model.

3.4 Modeling Long-Term Evolution

We here introduce a simple yet effective way to model the long-term evolution of spatio-temporal patterns. Let $\mathcal{R}_l$ and $\mathbf{W}_l$ denote the data tensor and POI similarity matrix in the lth year, and $\mathcal{G}_l = \{\mathcal{C}_l, \mathbf{O}_l, \mathbf{D}_l, \mathbf{T}_l\}$ denote the set of latent patterns learnt from the lth year's data, $l = 1, 2, \ldots, L$.

As described in Section 2.5, to factorize every $\mathcal{R}_l$ independently for $\mathcal{G}|_{l=1}^{L}$ is often inappropriate for generating incomparable patterns in successive years. The Dynamic Tensor Analysis (DTA) scheme suggested in [19], [20] cannot fulfill our task either for using $\mathcal{R}_l$ as well as historical data tensors to obtain a "hybrid" $\mathcal{G}_l$, which is not the genuine $\mathcal{G}_l$ we aim to analyze in practice.

We here propose a simple Pipeline Initialization based Tensor Sequence Analysis (PI-TSA) method. In PI-TSA, the factorization results in $\mathcal{G}_l$ are expressed as

$$\mathcal{G}_l = f_{\mathrm{NR-cNTF}}(\mathcal{R}_l, \mathbf{W}_l, \mathcal{G}_{l-1}), \tag{22}$$

where $f_{\mathrm{NR-cNTF}}$ denotes the optimization algorithm for NR-cNTF. Fig. 3 further illustrates PI-TSA via a flow chart. As can be seen, the key of PI-TSA is to set the initial values of the lth year's optimization as the outputs in the (l-1)th step (i.e., $\mathcal{G}_{l-1}$). In this way, the patterns in the (l-1)th year can be "inherited" by the patterns in the lth year, and only the information of $\mathcal{R}_l$ and $\mathbf{W}_l$ is used for pattern discovery in the lth year.

Algorithm 1. Block Coordinate Descent Procedure

Require: Data sets $\{\mathcal{R}, \mathbf{W}\}$, parameters $\{\gamma, \delta, \epsilon, \varepsilon\}$
  Initialization: $\left(\mathcal{C}^{(0)}, \mathbf{O}^{(0)}, \mathbf{D}^{(0)}, \mathbf{T}^{(0)}\right)$
  for $s = 1, 2, \ldots$ **do**
    Update $\mathcal{C}^{(s)}$ by solving the problem (23a).
    Update $\mathbf{O}^{(s)}$ by solving the problem (23b).
    Update $\mathbf{D}^{(s)}$ by solving the problem (23c).
    Update $\mathbf{T}^{(s)}$ by solving the problem (23d).
    Apply Algorithm 2 to $\mathbf{O}^{(s)}$.
    Apply Algorithm 2 to $\mathbf{D}^{(s)}$.
    if convergence **then**
      return $\left(\mathcal{C}^{(s)}, \mathbf{O}^{(s)}, \mathbf{D}^{(s)}, \mathbf{T}^{(s)}\right)$.
    end if
  end for

4 Inference

4.1 Basic Optimization

We adopt the Block Coordinate Descent-Proximal Gradient (BCD-PG) algorithm [21], [22] to solve the cNTF problem in Eq. (17). While this function is not jointly convex with respect to $\mathcal{C}$, $\mathbf{O}$, $\mathbf{D}$, and $\mathbf{T}$, it is *block multiconvex* with each one when the other three are fixed. Therefore, as shown in Algorithm 1, we adopt a Block Coordinate Descent (BCD) procedure, which starts from an initialization on $\mathcal{G}^{(0)}$, and then iteratively updates $\mathcal{G}^{(s)}$, $s = 1, 2, \ldots$, by

$$\mathcal{C}^{(s)} = \arg\min_{\mathcal{C}} \mathcal{J}\left(\mathcal{C}, \mathbf{O}^{(s-1)}, \mathbf{D}^{(s-1)}, \mathbf{T}^{(s-1)}\right) + \gamma\|\mathcal{C}\|_1, \tag{23a}$$

$$\mathbf{O}^{(s)} = \arg\min_{\mathbf{O}} \mathcal{J}\left(\mathcal{C}^{(s)}, \mathbf{O}, \mathbf{D}^{(s-1)}, \mathbf{T}^{(s-1)}\right) + \delta\|\mathbf{O}\|_1, \tag{23b}$$

$$\mathbf{D}^{(s)} = \arg\min_{\mathbf{D}} \mathcal{J}\left(\mathcal{C}^{(s)}, \mathbf{O}^{(s)}, \mathbf{D}, \mathbf{T}^{(s-1)}\right) + \epsilon\|\mathbf{D}\|_1, \tag{23c}$$

$$\mathbf{T}^{(s)} = \arg\min_{\mathbf{T}} \mathcal{J}\left(\mathcal{C}^{(s)}, \mathbf{O}^{(s)}, \mathbf{D}^{(s)}, \mathbf{T}\right) + \varepsilon\|\mathbf{T}\|_1. \tag{23d}$$

Let $(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \mathbf{g}_4)$ denote $(\mathcal{C}, \mathbf{O}, \mathbf{D}, \mathbf{T})$ for concision. Using a Proximal Gradient (PG) method, the algorithm updates the ith variable of $\mathcal{G}$ in the sth round as

$$\begin{aligned}\mathbf{g}_i^{(s)} &= \arg\min_{\mathbf{g}_i \geq 0} \left\langle \frac{\partial \mathcal{J}\left(\mathbf{g}_{<i}^{(s)}, \tilde{\mathbf{g}}_i^{(s)}, \mathbf{g}_{>i}^{(s-1)}\right)}{\partial \mathbf{g}_i}, \mathbf{g}_i - \tilde{\mathbf{g}}_i^{(s)} \right\rangle \\ &\quad + \frac{\tau_i}{2}\left\|\mathbf{g}_i - \tilde{\mathbf{g}}_i^{(s)}\right\|_F^2 + \lambda_i\|\mathbf{g}_i\|_1 \\ &= \max\left\{0, \tilde{\mathbf{g}}_i^{(s)} - \frac{1}{\tau_i}\frac{\partial \mathcal{J}\left(\mathbf{g}_{<i}^{(s)}, \tilde{\mathbf{g}}_i^{(s)}, \mathbf{g}_{>i}^{(s-1)}\right)}{\partial \mathbf{g}_i} - \frac{\lambda_i}{\tau_i}\right\},\end{aligned} \tag{24}$$

where $\langle \cdot \rangle$ denotes the inner product, $\mathbf{g}_{<i}^{(s)}$ denotes $\{\mathbf{g}_1^{(s)} \ldots \mathbf{g}_{i-1}^{(s)}\}$, and $\mathbf{g}_{>i}^{(s-1)}$ denotes $\{\mathbf{g}_{i+1}^{(s-1)} \ldots \mathbf{g}_4^{(s-1)}\}$. The variable $\tilde{\mathbf{g}}_i^{(s)}$ is a linear extrapolated point as follows:

$$\tilde{\mathbf{g}}_i^{(s)} = \mathbf{g}_i^{(s-1)} + \omega_i^{(s)}\left(\mathbf{g}_i^{(s-1)} - \mathbf{g}_i^{(s-2)}\right), \tag{25}$$

where $\omega_i^{(s)}$ is an extrapolation weight set according to [22]. The parameter τ_i in (24) is a Lipschitz constant of $\frac{\partial \mathcal{J}(\mathbf{g}_i)}{\partial \mathbf{g}_i}$ with respect to $\mathbf{g}_i$, namely

$$\left\|\frac{\partial \mathcal{J}\left(\mathbf{g}_{i_1}\right)}{\partial \mathbf{g}_{i_1}} - \frac{\partial \mathcal{J}\left(\mathbf{g}_{i_2}\right)}{\partial \mathbf{g}_{i_2}}\right\|_F \leq \tau_i \|\mathbf{g}_{i_1} - \mathbf{g}_{i_2}\|_F, \forall\, \mathbf{g}_{i_1}, \mathbf{g}_{i_2}, \tag{26}$$

and λ_i is the regularization parameter of $\mathbf{g}_i$. Specifically, the gradients of $\mathcal{J}$ with respect to each component are calculated as

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \mathcal{C}} &= 2\left(\mathcal{C} \times_o (\mathbf{O}^\top\mathbf{O}) \times_d (\mathbf{D}^\top\mathbf{D}) \times_t (\mathbf{T}^\top\mathbf{T}) \right. \\ &\quad \left. - \mathcal{R} \times_o \mathbf{O}^\top \times_d \mathbf{D}^\top \times_t \mathbf{T}^\top\right), \\ \frac{\partial \mathcal{J}}{\partial \mathbf{O}} &= 2\left(\mathbf{O}\big(\mathcal{C} \times_d (\mathbf{D}^\top\mathbf{D}) \times_t (\mathbf{T}^\top\mathbf{T})\big)_{(o)}\mathcal{C}_{(o)}^\top \right. \\ &\quad \left. - \left(\mathcal{R} \times_d \mathbf{D}^\top \times_t \mathbf{T}^\top\right)_{(o)}\mathcal{C}_{(o)}^\top - \alpha\big(\mathbf{W} - \mathbf{O}\mathbf{O}^\top\big)\mathbf{O}\right), \\ \frac{\partial \mathcal{J}}{\partial \mathbf{D}} &= 2\left(\mathbf{D}\big(\mathcal{C} \times_o (\mathbf{O}^\top\mathbf{O}) \times_t (\mathbf{T}^\top\mathbf{T})\big)_{(d)}\mathcal{C}_{(d)}^\top \right. \\ &\quad \left. - \left(\mathcal{R} \times_o \mathbf{O}^\top \times_t \mathbf{T}^\top\right)_{(d)}\mathcal{C}_{(d)}^\top - \beta\big(\mathbf{W} - \mathbf{D}\mathbf{D}^\top\big)\mathbf{D}\right), \\ \frac{\partial \mathcal{J}}{\partial \mathbf{T}} &= 2\left(\mathbf{T}\big(\mathcal{C} \times_o (\mathbf{O}^\top\mathbf{O}) \times_d (\mathbf{D}^\top\mathbf{D})\big)_{(t)}\mathcal{C}_{(t)}^\top \right. \\ &\quad \left. - \left(\mathcal{R} \times_o \mathbf{O}^\top \times_d \mathbf{D}^\top\right)_{(t)}\mathcal{C}_{(t)}^\top\right),\end{aligned} \tag{27}$$

where $\boldsymbol{\mathcal{X}}_{(x)}$ denotes the mode-x matricization of tensor $\boldsymbol{\mathcal{X}}$.

4.2 Neighboring Regularization Optimization

Algorithm 2 shows the optimization process of neighboring regularization. Without loss of generality, we still take the origin order for illustration. In each cNTF optimization iteration, Algorithm 2 regularizes the projection matrix $\mathbf{O}$ through the following steps:

1) *Calculate Unary Potentials*: We first normalize $\mathbf{O}$ as

$$o'_{xi} = \frac{o_{xi}}{\sum_{j=1}^{I} o_{xj}}. \tag{28}$$

Then the unary potential of o_{xi} is $\psi_x^u(i) = -\log o'_{xi}$.

2) *Calculate Pairwise Potentials*: We then calculate the average pairwise potential of $\nu_x = i$ to $\nu_y \in \{j | j \neq i\}$ as

$$Q_{xi} = \sum_{j \neq i}\sum_{y \in M_x} P_{yj} \cdot \psi_{xy}^p(i, j), \tag{29}$$

where M_x is the set of neighbor zones for zone x. P_{yj} in Eq. (29) is a probability of $v_y = j$, which is defined as

$$P_{yj} = \frac{\exp(-\psi_y^u(j))}{Z_y} = o'_{yj}, \tag{30}$$

where $1/Z_x$ denotes the partition function.

3) *Update the Projection Matrix*: Finally, we calculate the total potential of o_{xi} as

$$\zeta_{xi} = \psi_x^u(i) + Q_{xi}. \tag{31}$$

The regularized element is then defined as

$$\tilde{o}_{xi} = \exp(-\zeta_{xi}) \cdot \sum_{j=1}^{I} o_{xj}. \tag{32}$$

For the sth round of iteration in Algorithm 1, we define $\Delta_{NR} = \tilde{o}_{xi}^{(s)} - o_{xi}^{(s)}$, and $\Delta_{cNTF} = o_{xi}^{(s)} - o_{xi}^{(s-1)}$. Algorithm 2 then updates $o_{xi}^{(s)}$ as

$$o_{xi}^{(s)} = \begin{cases} \max\{0, o_{xi}^{(s-1)} + \Delta^{cNTF} + \Delta^{NR}\}, & \text{if } \Delta^{cNTF} \leq 0, \\ o_{xi}^{(s-1)} + \max\{0, \Delta^{cNTF} + \Delta^{NR}\}, & \text{otherwise.} \end{cases} \tag{33}$$

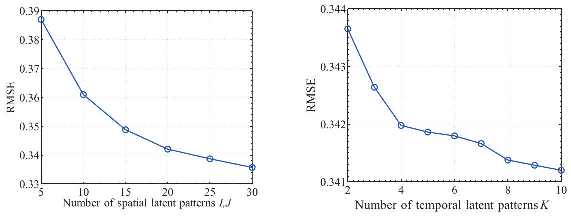
Note that $\tilde{o}_{xi}^{(s)} \leq o_{xi}^{(s)} \Rightarrow \Delta^{NR} \leq 0$, so the update of o_{xi} in Eq. (33) is in the same direction with the gradient of $o_{xi}^{(s-1)}$. Algorithm 2 therefore ensures that the reconstruction error in each iteration is always the same or lower than that in the previous iteration.

Algorithm 2. Neighboring Regularization Optimization

Unary Potentials: $o'_{xi} \leftarrow \frac{o_{xi}}{\sum_{j=1}^{I} o_{xj}}$, $\quad \psi_x^u(i) \leftarrow -\log o'_{xi}$.

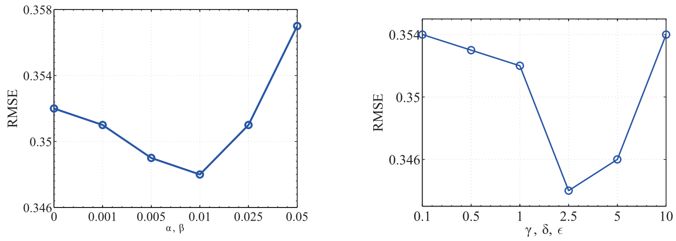
Pairwise Potentials: $\tilde{Q}_{xi} \leftarrow \sum_{j \neq i}\sum_{y \in M_x} \psi_{xy}^p(i, j) o'_{yj}$.

Update the Projection Matrix.



(a) Setting of I, J (b) Setting of K

Fig. 4. Performance with varying dimensionality of pattern space.



(a) POI Regularization (b) L1 Regularization

Fig. 5. Performance with varying POI and L1 regularization coefficients.

5 Experimental Results

In this section, we conduct extensive experiments to evaluate the effectiveness of our methods in learning urban dynamics and gaining managerial insights for urban planning. We also compare our methods with some baselines on traffic prediction, which justifies the modeling of urban contexts and neighboring regulation in NR-cNTF.

5.1 Experimental Setup

5.1.1 Data Sets

Three types of data sets were used in our experiments including taxi trajectory data, POI data, and Traffic Analysis Zone data. The taxi trajectory data set contains the GPS trajectories of 20,000 Beijing taxis collected in November 2008 and November 2015, from which we extracted more than 6 million trips of taxi passengers to present the daily mobility behaviors of residents in Beijing. The POI data set contains more than 400 thousands POI records of Beijing in the years of 2008 and 2015. The Traffic Analysis Zone data set, offered by Beijing Municipal Commission of Transportation, divides the Beijing area within the 5th Ring Road into 651 zones. Using the three data sets, we built two data tensors $(651 \times 651 \times 24)$ and two POI context matrices (651×651) for the years of 2008 and 2015, respectively. In the experiments, we only use data of workdays to construct the data tensor $\mathcal{R}$, so the discovered patterns reflect resident mobility in workdays. Peoples leisure patterns in holiday could be very different from their workday patterns. We have conducted extra experiments on holiday data, and included the results to *Supplementary Materials*, available online, for readers with interests.

5.1.2 Setting of Dimensionality of Pattern Space

The goal of the NR-cNTF model is to find an $I \times J \times K$-dimensional pattern space. How to set I, J, K appropriately, however, is a "tricky" issue. If the dimensionality is too small, we might omit some urban dynamics; if too large, we might obtain many trivial patterns (for the extreme case, if the dimensionality of the pattern space is the same as the data space, the patterns will be meaningless).

In our experiments, we set the parameters carefully so as to make a tradeoff between the reconstruction error and the dimension reduction. The reconstruction error is evaluated by *Root Mean Square Error* (RMSE) defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{x=1}^{M}\sum_{y=1}^{M}\sum_{z=1}^{N}\left(r_{xyz} - \hat{r}_{xyz}\right)^2}{M \times M \times N}}, \tag{34}$$

where $\hat{r}_{xyz}$ is the (x, y, z) element of the reconstructed data tensor. We repeated experiments 10 times with $I = J$ ranging from 5 to 30 and K ranging from 2 to 10. Fig. 4 gives the resultant average reconstruction errors with different parameters, where RMSE reduces sharply at the very beginning but slows down when $I, J \geq 20$ and $K \geq 4$. We therefore set $I = J = 20$ and $K = 4$ as defaults.

5.1.3 Setting of Tradeoff Parameters

In NR-cNTF, the tradeoff parameters α and β are for adjusting the strength of urban context terms, and γ, δ and ϵ for adjusting the strength of sparsity regularization terms. In our experiment, we set the tradeoff parameters using a traverse approach. We vary α and β from 0 to 0.05 and γ, δ and ϵ from 0.1 to 10, respectively, aiming to choose the parameters with the best performances. Fig. 5 exhibits the experimental reconstruction errors with different tradeoff parameters, where each point is averaged on 10 runs. As can be seen, the best performance appears when $\alpha = \beta = 0.01$ and $\gamma = \delta = \epsilon = 2.5$, which become the default settings.

5.2 Discovery of Temporal Patterns

Here, we describe the temporal patterns discovered from Beijing taxi traffic in 2008 and 2015. To facilitate comparison, we first introduce a normalization scheme to the projection matrix $\mathbf{T}$. Specifically, for the kth pattern, we define a mask matrix as $\mathbf{Y}^k \in \mathbb{R}^{N \times K}$, where the element $y_{xi}^k = 1$ when $i = k$, and 0 otherwise. We use the mask matrix to construct a data tensor as

$$\tilde{\mathcal{R}}^k = \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \left(\mathbf{T} \odot \mathbf{Y}^k\right). \tag{35}$$

In Eq. (35), the elements of $\mathbf{T}$ corresponding to the patterns $\neg k$ are multiplied by zero, so $\tilde{\mathcal{R}}^k$ only contains the components of the pattern k. Therefore, the physical meaning of $\tilde{\mathcal{R}}^k$ is a component tensor corresponding to the kth temporal pattern of the data tensor $\mathcal{R}$. Using $\tilde{\mathcal{R}}^k$, we define the *energy* of the temporal pattern k as

$$u_k = \frac{\|\tilde{\mathcal{R}}^k\|_1}{M \times M \times N} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{M}\sum_{z=1}^{N}|\tilde{r}_{xyz}^k|}{M \times M \times N}. \tag{36}$$

The physical meaning of the *energy* u_k is a normalized size of the components corresponding to the temporal pattern k.

In the experiments, we define the re-scaled pattern coefficient $\tilde{t}_{zk}$ as

$$\tilde{t}_{zk} = \frac{t_{zk}}{\sum_{n=1}^{N} t_{nk}} \times u_k. \tag{37}$$

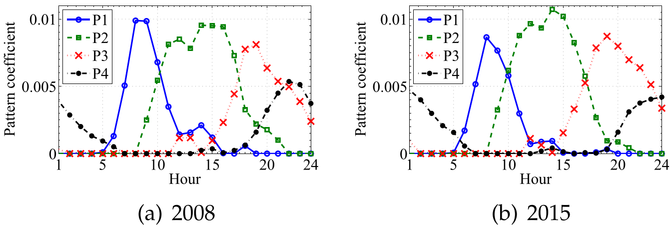


Fig. 6. Temporal patterns in 2008 and 2015.

The physical meaning of $\tilde{t}_{zk}$ is the energy of the temporal pattern k at the time slice z. The vector $\tilde{\mathbf{t}}_{:k}$ is the distribution of u_k over the N time slices, and $\sum_{z=1}^{N} \tilde{t}_{zk} = u_k$. We compare the re-scaled pattern coefficients of different years to demonstrate the changes of temporal patterns of resident mobility from 2008 and 2015.

Fig. 6 shows the four temporal patterns, which indeed correspond to four rhythms of urban traffic:

- *P1: Morning Peak*, with an active range roughly from 6:00 to 11:00.
- *P2: Midday*, with an active range roughly from 9:00 to 18:00.
- *P3: Evening Peak*, with an active range roughly from 16:00 to 24:00.
- *P4: Night*, with an active range roughly from 20:00 to 3:00 of the next day.

To further reveal the evolution of temporal patterns from 2008 to 2015, we plot comparative diagram for each pattern of the two yeas in Fig. 7. The first observation is that the intensity of the morning pattern was decreased significantly from 2008 to 2015 (see Fig. 7a), whereas the evening pattern seems much more stable (see Fig. 7c). We believe the reduction of the morning peak via taxies is due to the rapid development of the metro system in Beijing. During the period from 2008 to 2015, the Beijing metro increased the mileage from 198 to 631 km, which is particularly suitable for the time-rigid morning commute but has less impact to the evening commute with relatively flexible time.

Another observation is that the intensity of the midday pattern was increased during the seven years (see Fig. 7b). The main part of travel volume in the midday pattern consists of business travels from one workplace to another, whose destinations are random in essence and therefore cannot count heavily on public transportation systems like metros. Moreover, the fast-rising income in China in recent years might also contribute to the more spending on the relatively expensive taxi service.

The most interesting observation is that the peak time of the night pattern in 2015 came about two hours later than that in 2008 (Fig. 7d). This implies that residents tend to have more travels in the midnight in recent years. The reasons behind this could be complicated, which might include some lifestyle changes in Beijing, such as the more colorful nightlife or the higher overtime working pressures.

To sum up, the NR-cNTF model well captures the temporal patterns hidden inside the Beijing taxi traffic. The evolution of these patterns further unveils the development of Beijing metros and the changes of lifestyle.

5.3 Discovery of Spatial Patterns

Here, we explore the spatial patterns discovered by NR-cNTF. Given any origin or destination pattern $\mathbf{v}_{:i}$ (see Definition 1 in Section 2.2), we first obtain the corresponding urban community $\mathcal{SP}_i$ (see Section 3.3). We adopt the "crisp partition" assumption so that an urban zone will be assigned to one and only one urban community. As a result, among the $I = J = 20$ patterns in our experiment, we obtain 17 urban communities, and the rest three are empty and omitted. Note that we only use destination spatial patterns for illustration below. The origin spatial patterns have the similar results, we don't put them in the paper for concision.

Figs. 8a and 8b visualize the urban communities corresponding to the destination spatial patterns found in 2008 and 2015, respectively. As can be seen, each urban community (filled with a same color) identified by NR-cNTF contains urban zones geographically adjacent to at least one zone in the same community, which agrees with our intuition about functional zoning of a city. In contrast, Fig. 8c shows the 2008 urban communities found by cNTF without neighboring regulation, whose functionalities are less clear due to the geographical discontinuity. For the convenience of discussion, we numbered the communities in Fig. 8b from 1 to 17.

A general observation from Fig. 8 is that the spatial communities of Beijing radially surround the center of Beijing. This character of spatial communities has close relations with the trunk road network structure of Beijing. Fig. 10a shows there are four concentric ring roads surrounding the center of Beijing. As reported in [23], the ring roads provide a basic framework for the city's overall spatial pattern. Affected by the ring roads, we can see that the communities discovered in Fig. 8 also constitute two concentric circles surrounding the center of the Beijing city. Specifically, the communities C1-C10 form the outer circle, and C11-C17 form the inner circle. Fig. 10b plots the trunk road network of Beijing over the communities, from which we can see that many boundaries of the

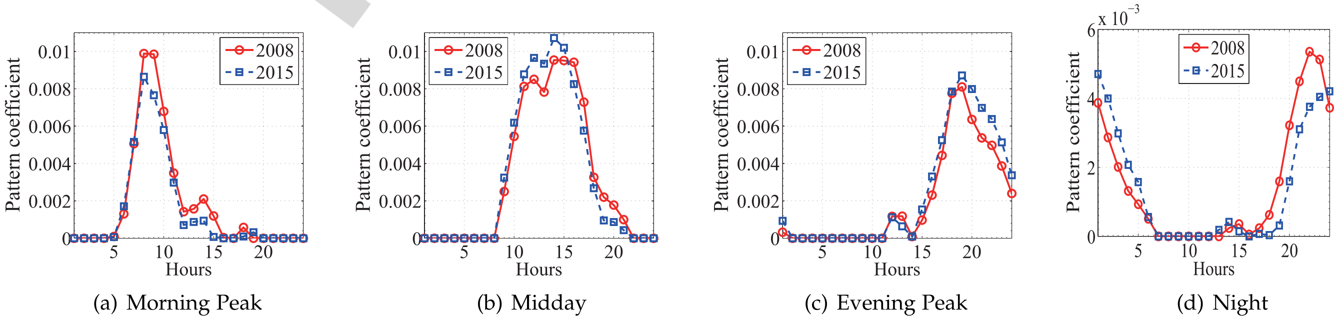
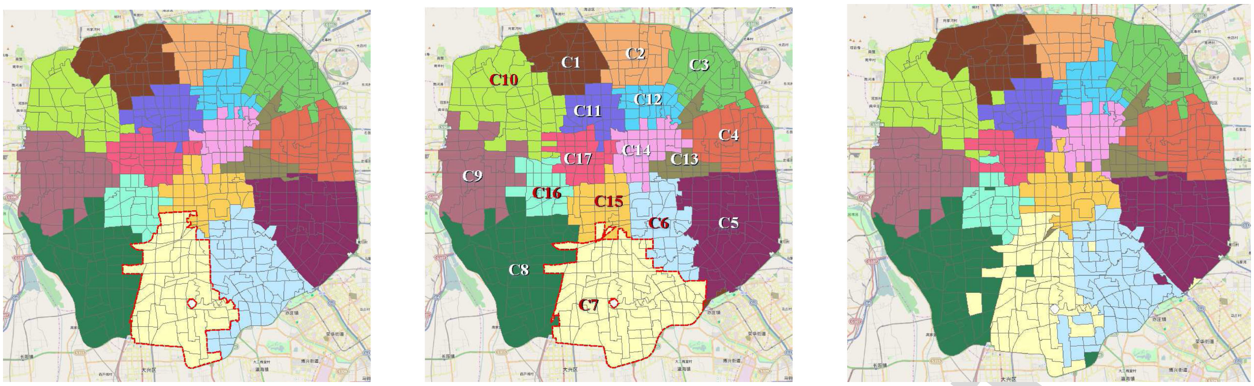
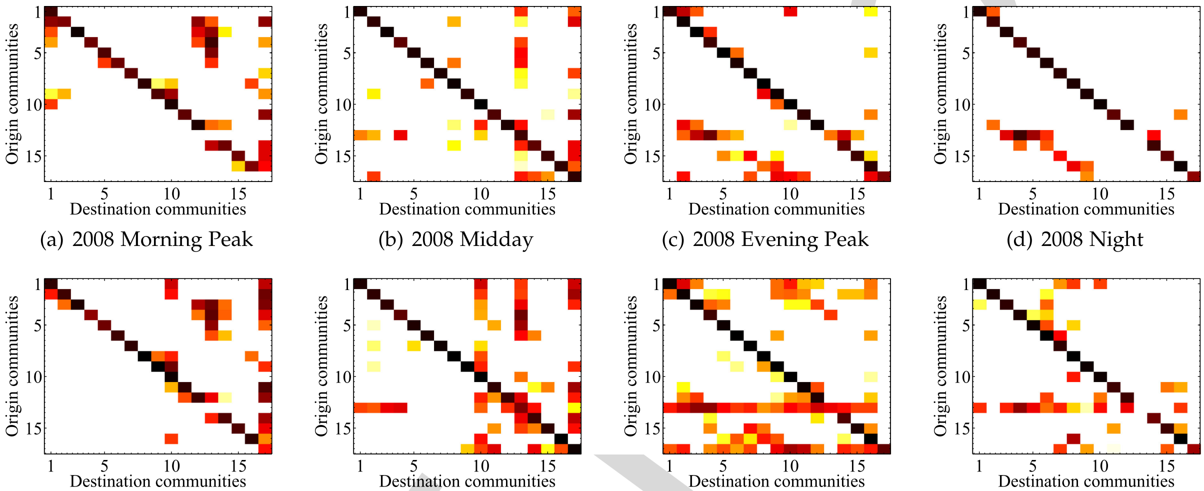


Fig. 7. The temporal patterns comparison between 2008 and 2012.



(a) 2008 DSP's by NR-cNTF (b) 2015 DSP's by NR-cNTF (c) 2008 DSP's by cNTF

Fig. 8. Destination spatial patterns in 2008 and 2015.



(a) 2008 Morning Peak (b) 2008 Midday (c) 2008 Evening Peak (d) 2008 Night

(e) 2015 Morning Peak (f) 2015 Midday (g) 2015 Evening Peak (h) 2015 Night

Fig. 9. Dynamic patterns in 2008 and 2015.

communities overlap with the trunk roads, indicating that the spatial patterns of residential mobility in Beijing are deeply shaped by the urban trunk road network.

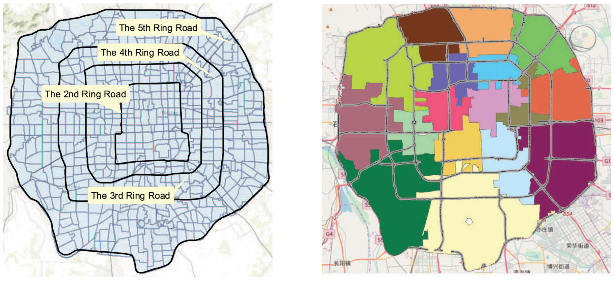
Another observation from Fig. 8 is the interesting evolution of some urban communities in recent years. Let us take a closer look on community C7 located in the south of Beijing, which has an obvious expansion trend from 2008 to 2015. That is, some urban zones that belonged to C6 in 2008 were "absorbed" by C7 in 2015. To understand this, we should trace back to the so-called South Beijing Development Plan (SBDP) issued in 2008, which is a government investment plan in south areas of Beijing, with an executive period from 2010 to 2015 and a total investment of nearly 62.9 billion USD (more information about SBDP could be found in *Supplementary Materials*, available online). The purpose of SBDP is to narrow the development gap between the lagging-behind southern region and other areas of the city. It is interesting that the communities C6 and C7 are just in the investment region of the plan (see Fig. 2 in *Supplementary Materials*, available online, for the evidence). The evolution of C6 and C7 from 2008 to 2015 essentially reflects the great impact of huge economic investments to the real-life development of a city.

To sum up, the above results justify the effectiveness of our NR-cNTF model in uncovering latent and geographically adjacent spatial patterns, as well as their inconspicuous evolutions in recent years.

5.4 Discovery of Urban Dynamics Among Patterns

Here, we use the core tensor $\mathcal{C}$ to explore the urban dynamics, i.e., the interactions among spatial and temporal patterns. We first observe the slice $\mathbf{C}_{::k}$ of $\mathcal{C}$, which reveals the traffic intensity from every origin communities to every destination ones given the temporal pattern k, i.e., a community level origin-destination (OD) matrix in rhythm k.

Fig. 9 visualizes the community OD-matrices in the morning peak, midday, evening peak and night rhythms of 2008 and 2015. A darker color indicates a higher traffic intensity. As can be seen, most energies of the OD-matrices are concentrated in their diagonal lines, implying that most of taxi travels in Beijing actually happened within the same community with relatively short distances. Moreover, the travel demands across communities have a *tidal* phenomenon. That is, in the morning peak, people flowed out from many communities (i.e., residential areas) and flowed in a few ones (i.e., working areas), and the situation was just the reverse in the evening peak and night rhythms. This implies that while the residential areas in Beijing are very dispersed, the workplaces are relatively concentrated. Indeed, it seems from Fig. 9e that C10, C13 and C17 are the three "most



(a) The Ring Roads in Beijing (b) The Trunk Roads in Beijing

Fig. 10. The urban communities and trunk roads in Beijing.

attractive” workplaces in Beijing, which are actually well-known as the Zhongguancun area,[4] Beijing Central Business District (CBD),[5] and Beijing Financial Street,[6] respectively. From this aspect, NR-cNTF indeed generates high-quality patterns for urban dynamics understanding.

We then explore the evolution of traffic intensities from 2008 to 2015 in Beijing. For the comparison purpose, we first concentrate the energies of projection matrices into the core tensor as $c'_{ijk} = c_{ijk} \cdot \sum_x o_{xi} \cdot \sum_y d_{yj} \cdot \sum_z t_{zk}$. The total intensity of inter-community traffic for a community x is then calculated as $I_x^{inter} = \sum_{i \neq x} \sum_k c'_{ixk} + \sum_{j \neq x} \sum_k c'_{xjk}$, and the intra-community traffic intensity for x is given by $I_x^{intra} = \sum_k c'_{xxk}$. Along this line, we can quantify the daily increments of inter- and intra-community traffic intensities from 2008 to 2015, as shown in Fig. 11.

From Fig. 11a, it is obvious that the inter-community traffic increased from 2008 to 2015 for almost all communities, with C10 (Zhongguancun area), C13 (CBD area) and C17 (Financial Street area) being the most significant ones. In particular, as shown in Fig. 11b, the Zhongguancun area, a technology hub of Beijing and well-known as the “Chinese Silicon Valley”, gains a highest growth ratio during the seven years, which coincides with the developing priority of Beijing with high-tech industries preference.

Fig. 11c depicts the intra-community traffic intensity of each community from 2008 to 2015. It is interesting that C7 and C15 emerged as the top-2 communities with highest growth in internal traffic. Recall that these two communities are located in the south of the Beijing city, and have benefited from the 30 billion dollar investment of the South Beijing Development Plan. The significant growth of internal traffic implies that these two communities are gaining more active economics, and perhaps are enjoying more sustainable developing pattern—residents can work and rest interchangeably within a small distance. This indeed recommends a potential solution to mitigating the “big city disease” of Beijing: to promote industries and housing in a same community or close ones. This job-housing balance thinking, however, was not the primary choice of Beijing in the past several decades. The development of the CBD area, which we will discuss below, is just the epitome.

In Fig. 12, we study the dynamic patterns of a particular community: the CBD area (C13), which is the central business district of Beijing and shapes the lifestyle of the city deeply. In the figure, the color of a community indicates the traffic intensity of that community from or to the CBD community: the redder the stronger, and the arrows indicate traffic directions between communities. As shown in Fig. 12, CBD is a pure business area, with residents flowing in in the morning and flowing out in the evening. Similar situations can be found from the Zhongguancun (C10) and the Financial Street (C17) communities. This indeed reflects the severe job-housing imbalance in Beijing, which contributes a lot to the city disease such as traffic congestion. Nevertheless, it is more interesting to find the pattern evolution of CBD from 2008 to 2015. From Figs. 12a and 12b, we can find the nearly symmetric incoming and outgoing flows between the CBD community and the communities surrounding CBD in 2008. This symmetry, however, disappeared in 2015, where the outflows from CBD in the evening spread over more communities than that in the morning (see Figs. 12c and 12d). We believe it is Fig. 12d rather than Fig. 12c that revealed all the housing communities for CBD. The possible reason is, for residents living in remote communities, the *long-term*, *timely* and *economic* way commuting to CBD in the morning is to take metro rather than taxi. From this angle, we can conclude that the job-housing imbalance gets even worse with the rapid development of the CBD area from 2008 to 2015.

To sum up, the evolution of urban dynamics indicates the rapid development of Beijing city in recent years. The development pattern, however, is still worrying for the job-housing imbalance status quo, although the southern area has showed some positive changes.

5.5 Quantitative Evaluation

In this section, we evaluate our NR-cNTF model by comparing its data tensor reconstruction error with that of some baseline models, for further explaining why NR-cNTF can work well for understanding the Beijing city. Following the tradition of tensor factorization based studies [4], [20], the *Root Mean Square Error* defined in Eq. (34) is used as an indicator of quality.

In the experiments, we define a sampling tensor $\mathcal{S} \in \mathbb{R}^{M \times M \times N}$, in which the element $s_{xyz} = 1$ when the traffic volume form zone x to zone y in time slice z was sampled, otherwise un-sampled. We then rewrite the objective function in Eq. (17) as

$$\begin{aligned} \arg \min_{\mathcal{C},\mathbf{O},\mathbf{D},\mathbf{T} \geq 0} \ \mathcal{J} = & \|\mathcal{S} \odot (\mathcal{R} - \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T})\|_F^2 \\ & + \alpha \|\mathbf{W} - \mathbf{O}\mathbf{O}^\top\|_F^2 + \beta \|\mathbf{W} - \mathbf{D}\mathbf{D}^\top\|_F^2 \\ & + \gamma \|\mathbf{O}\|_1 + \delta \|\mathbf{D}\|_1 + \epsilon \|\mathbf{T}\|_1 + \varepsilon \|\mathcal{C}\|_1. \end{aligned} \tag{38}$$

The reconstruction error between $\mathcal{R}$ and the reconstructed tensor $\hat{\mathcal{R}} = \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}$ is calculated using Eq. (34).

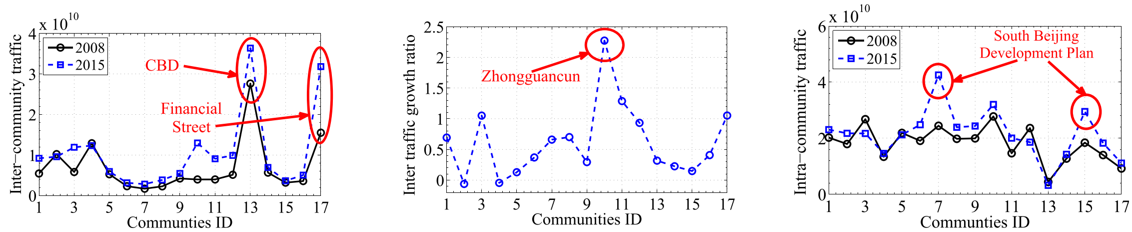
We compare the reconstruction error of NR-cNTF with that of the following baseline methods:

- *Tucker*: Non-negative Tucker Factorization, of which the objective function is

$$\begin{aligned} \arg \min_{\mathcal{C},\mathbf{O},\mathbf{D},\mathbf{T}} & \|\mathcal{S} \odot (\mathcal{R} - \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T})\|_F^2 \\ & + \gamma \|\mathbf{O}\|_1 + \delta \|\mathbf{D}\|_1 + \epsilon \|\mathbf{T}\|_1 + \varepsilon \|\mathcal{C}\|_1. \end{aligned} \tag{39}$$

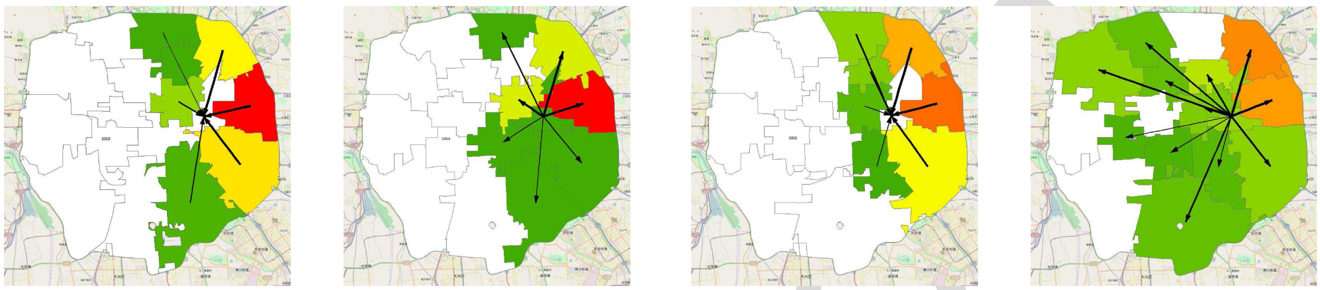
  Compared with our method, *Tucker* does not consider urban context and neighboring regularization.

4. https://en.wikipedia.org/wiki/Zhongguancun
5. https://en.wikipedia.org/wiki/Beijing_central_business_district
6. https://en.wikipedia.org/wiki/Beijing_Financial_Street



(a) Inter-Community Traffic (b) Inter-Community Traffic Growth (c) Intra-Community Traffic

Fig. 11. Inter- and intra-community traffic intensities.



(a) 2008 Morning Peak (b) 2008 Evening Peak (c) 2015 Morning Peak (d) 2015 Evening Peak

Fig. 12. Dynamic patterns from and to the CBD community.

- *CP*: Non-negative CP Factorization, which supposes a joint latent space for each mode by solving an objective function as

$$\arg\min_{\mathbf{O},\mathbf{D},\mathbf{T}} \left\| \mathcal{S} \odot \left(\mathcal{R} - \sum_{m} \mathbf{o}_{:m} \circ \mathbf{d}_{:m} \circ \mathbf{t}_{:m} \right) \right\|_F^2, \\ + \gamma\|\mathbf{O}\|_1 + \delta\|\mathbf{D}\|_1 + \epsilon\|\mathbf{T}\|_1, \tag{40}$$

  where operator $\circ$ represents the vector outer product. In the CP factorization, the latent factor dimensionality for both the spatial and temporal patterns are the same. As a result, we set the number of latent factors $m = 4$ or $m = 20$. The former is the same as the number of temporal patterns for NR-cNTF, and the latter is in accordance with that of spatial patterns.
- *rCP*: Regularized Non-negative CP Factorization, which is a CP factorization with the urban context-aware regularization. The objective function is

$$\arg\min_{\mathbf{O},\mathbf{D},\mathbf{T}} \left\| \mathcal{S} \odot \left(\mathcal{R} - \sum_{m} \mathbf{o}_{:m} \circ \mathbf{d}_{:m} \circ \mathbf{t}_{:m} \right) \right\|_F^2 \\ + \alpha\|\mathbf{W} - \mathbf{O}\mathbf{O}^\top\|_F^2 + \beta\|\mathbf{W} - \mathbf{D}\mathbf{D}^\top\|_F^2 \\ + \gamma\|\mathbf{O}\|_1 + \delta\|\mathbf{D}\|_1 + \epsilon\|\mathbf{T}\|_1. \tag{41}$$

In our experiments, we compared the methods on the data tensor of 2015. The sampling rate varied from 50 to 90 percent. The average $RMSE$ values of ten times repeated experiments are reported in Table 3. From the table, we have the following observations:

- Both NR-cNTF and cNTF performed much better than the baseline methods, indicating the general superiority of the proposed methods.
- NR-cNTF performed nearly the same as cNTF, indicating that the neighboring regularization improves the interpretability of spatial patterns at the very low cost of model deviation from real-world data.
- NR-cNTF/cNTF performed generally better than Tucker, indicating the distinct value of urban contexts for tensor factorization.
- NR-cNTF/cNTF/Tucker performed generally better than rCP4/CP4/rCP20/CP20, implying the advantage of employing Tucker rather than CP based methods. This is not unusual, since the core tensor generated by Tucker factorization contains important information about urban dynamic patterns and improves the model interpretability.

In summary, besides the superior interpretability, NR-cNTF also shows excellent performance in quantitative evaluation on tensor factorization, by employing core tensor, neighboring regulation, and urban contexts. As a natural corollary, NR-cNTF could be used for urban traffic volume prediction when the elements of a data tensor are only partially available.

6 Related Work

Mining knowledge from human mobility data generated in urban areas has attracted many researchers' interests in

TABLE 3
Tensor Reconstruction Performance by RMSE

	50%	60%	70%	80%	90%
NR-cNTF	0.351	**0.344**	**0.343**	**0.342**	**0.341**
cNTF	**0.350**	0.345	**0.343**	**0.342**	**0.341**
Tucker	0.357	0.356	0.353	0.351	0.350
rCP-20	0.351	0.349	0.349	0.347	0.347
rCP-4	0.403	0.401	0.400	0.398	0.396
CP-20	0.353	0.352	0.349	0.348	0.346
CP-4	0.405	0.403	0.401	0.401	0.400

recent years [24], [25]. Various types of "social sensors", such as cell phones [26], GPS terminals [25], and smart bus/metro cards [27], have been adopted to record mobility information of urban residents, based on which many successful applications have emerged for intelligent transportation [28], [29], [30], environmental protection [31], urban planning [10], urban emergency [32], etc. An excellent survey from an urban computing perspective can be found in [24], while [25] provides a survey from a social and community dynamics perspective.

Among the abundant methods for human mobility data mining, tensor factorization/decomposition, like CANDECOMP/PARAFAC (CP) [33] and Tucker factorizations [34], gains particular interests for its distinct ability in modeling multi-aspect heterogeneous big data. Indeed, in city scenarios data samples are always involved with many aspects, such as time, space, human, urban contexts and so on, and therefore are very suitable for tensor factorization based data mining methods [24]. Typical applications of tensor factorization could be classified into two categories. The first category is to reconstruct tensors for predicting unknown values in multi-aspect data sets, such as completing missing traffic data [2], inferring urban gas consumption [3], predicting travel time [4], recommending social tags [35], movies [36] and sightseeing locations [37], [38], and so on.

In recent years, more and more works focused on mining explainable latent factors from multi-aspect urban data sets, which form the second category of applications. The focal point here is to use tensor factorization to discover latent lower-dimensional factors from higher-dimensional multi-aspect data sets. For instance, Metafac [39] used CP factorizations to extract latent community structures from various social networks, and [40] proposed a multi-view data clustering and partitioning method based on Tucker factorization. Our study in this paper also falls in this category, with some most related works as follows.

The study [7] used a non-negative matrix factorization, i.e., a second-order tensor factorization, to model taxi trip data, and discovered the latent factors corresponding to three rhythms of resident's daily life. Similarly, matrix factorizations were used for understanding the operational behaviors of taxicabs in cities [8]. In the inspiring work, [5] adopted a regularized non-negative Tucker decomposition (rNTD) to discover residents' mobility patterns in Beijing from an origin-destination-time tensor. Following this idea, [9] proposed a probabilistic tensor factorization method to find mobility patterns of public transaction system passengers from an origin-destination-time-type tensor. CitySpectrum [6] used CP factorizations to mine joint time-day-location patterns of residents after the Great East Japan Earthquake. Some more complex algorithms include NTCoF [41], which is a non-negative tensor co-factorization algorithm for urban events detection from bike trip and check-in data, and HTM [42], which is a hybrid tensor model and uses ACS-tucker decomposition to detect events from traffic data. In recent years, many dynamic tensor factorization algorithms were proposed for time series and stream data mining. For instance, Dynamic Tensor Analysis [19] extended Tucker factorization to process dynamic and stream high-order data, the Facets model [43] combined dynamic graphical models with tensor factorizations for mining co-evolving high-order time series, and FEMA [20] was a flexible evolutionary tensor factorization algorithm to mine dynamic behavioral patterns of multi-facet data sets.

Despite of the wide existence of related works mentioned above, our study in this paper has its own uniqueness. Unlike the previous works, we focus on understanding urban dynamics from multiple aspects, including spatial, temporal, as well as spatio-temporal interactions, with still a pursue to long-term evolution patterns. The results indeed bring some important managerial insights and suggestions to city development of Beijing. The proposed NR-cNTF model takes Tucker factorization as a basic framework, which compared with CP and matrix factorization based models [6], [7], [8], [42] has better interpretability for adopting a core tensor to model relations among latent factors. Compared with the existing Tucker factorization based methods [2], [9], [24], NR-cNTF incorporates urban contexts and neighboring regulation, which improve both the accuracy and interpretability of Tucker factorization greatly. Moreover, we proposed a pipeline initialization approach to analyze the evolution of urban dynamics across several years, which is simple yet practical.

7 Conclusion

In this paper, we proposed a POI context-aware nonnegative tensor factorization model with neighboring regulation (NR-cNTF) for urban dynamics discovery. A simple pipeline initialization method was also introduced to NR-cNTF to facilitate evolution analysis of the dynamics. Experiments on Beijing taxi trajectory and POI data demonstrated the high-quality of the spatial, temporal and spatio-temporal patterns generated by NR-cNTF for city-disease diagnosing and urban planning. The comparative studies with some baselines on traffic prediction further justified the advantage of NR-cNTF in adopting urban contexts and neighboring regulation.

Acknowledgments

J. Wang's work was partially supported by the National Natural Science Foundation of China under the Grant Number 61572059 and the Fundamental Research Funds for the Central Universities. J. Wu's work was partially supported by the National Natural Science Foundation of China (71531001, 71725002, U1636210). Z. Wang's work was partially supported by the Science and Technology Project of Beijing (Z181100003518001). Z. Xiong's work was partially supported by the National Key Research and Development Program of China (2017YFC0820405).

References

[1] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 29–35, Aug. 2014.
[2] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transp. Res. Part C: Emerging Technol.*, vol. 28, pp. 15–27, 2013.
[3] F. Zhang, D. Wilkie, Y. Zheng, and X. Xie, "Sensing the pulse of urban refueling behavior," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2013, pp. 13–22.
[4] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 25–34.

[5] J. Wang, F. Gao, P. Cui, C. Li, and Z. Xiong, "Discovering urban spatio-temporal structure from time-evolving traffic networks," in *Proc. Asia-Pacific Web Conf.*, 2014, pp. 93–104.

[6] Z. Fan, X. Song, and R. Shibasaki, "CitySpectrum: A non-negative tensor factorization approach," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 213–223.

[7] C. Peng, X. Jin, K.-C. Wong, M. Shi, and P. Liò, "Collective human mobility pattern from taxi trips in urban area," *PloS One*, vol. 7, no. 4, 2012, Art. no. e34487.

[8] C. Kang and K. Qin, "Understanding operation behaviors of taxicabs in cities by matrix factorization," *Comput. Environ. Urban Syst.*, vol. 60, pp. 79–88, 2016.

[9] L. Sun and K. W. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transp. Res. Part B: Methodological*, vol. 91, pp. 511–524, 2016.

[10] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.

[11] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 186–194.

[12] Y. Zheng, Y. Liu, J. Yuan, and X. Xie, "Urban computing with taxicabs," in *Proc. 13th Int. Conf. Ubiquitous Comput.*, 2011, pp. 89–98.

[13] N. J. Yuan, Y. Zheng, and X. Xie, "Segmentation of urban areas using road networks," Microsoft, Albuquerque, NM, USA, Tech. Rep. MSR-TR-2012-65, 2012.

[14] X. Liang, X. Zheng, W. Lv, T. Zhu, and K. Xu, "The scaling of human mobility by taxis is exponential," *Physica A: Statistical Mech. Appl.*, vol. 391, no. 5, pp. 2135–2144, 2012.

[15] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.

[16] D. Zhang, F. Zhang, and T. He, "MultiCalib: National-scale traffic model calibration in real time with multi-source incomplete data," in *Proc. 24th ACM SIGSPATIAL Int. Conf. Advances Geographic Inf. Syst.*, 2016, Art. no. 19.

[17] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang, "Diagnosing New York City's noises with ubiquitous data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 715–725.

[18] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 109–117.

[19] J. Sun, D. Tao, and C. Faloutsos, "Beyond streams and graphs: Dynamic tensor analysis," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2006, pp. 374–383.

[20] M. Jiang, P. Cui, F. Wang, X. Xu, W. Zhu, and S. Yang, "FEMA: Flexible evolutionary multi-faceted analysis for dynamic behavioral pattern discovery," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 1186–1195.

[21] Y. Xu, "Alternating proximal gradient method for sparse nonnegative tucker decomposition," *Math. Program. Comput.*, vol. 7, no. 1, pp. 39–70, 2015.

[22] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[23] G. Tian, J. Wu, and Z. Yang, "Spatial pattern of urban functions in the Beijing metropolitan region," *Habitat Int.*, vol. 34, no. 2, pp. 249–255, 2010.

[24] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, 2014, Art. no. 38.

[25] P. S. Castro, D. Zhang, C. Chen, S. Li, and G. Pan, "From taxi GPS traces to social and community dynamics: A survey," *ACM Comput. Surveys*, vol. 46, no. 2, 2013, Art. no. 17.

[26] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in Rome," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011.

[27] L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang, "Understanding metropolitan patterns of daily encounters," *Proc. Nat. Academy Sci. United States America*, vol. 110, no. 34, pp. 13 774–13 779, 2013.

[28] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "T-Drive: Enhancing driving directions with taxi drivers' intelligence," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 220–232, Jan. 2013.

[29] L. Chen, X. Ma, T.-M.-T. Nguyen, G. Pan, and J. Jakubowicz, "Understanding bike trip patterns leveraging bike sharing system open data," *Frontiers Comput. Sci.*, vol. 11, no. 1, pp. 38–48, Feb. 2017. [Online]. Available: https://doi.org/10.1007/s11704-016-6006-4

[30] C. Chen, S. Jiao, S. Zhang, W. Liu, L. Feng, and Y. Feng, "Tripimputor: real-time imputing taxi trip purpose leveraging multi-sourced urban data," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3292–3304, 2018.

[31] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-Air: When urban air quality inference meets big data," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1436–1444.

[32] X. Song, Q. Zhang, Y. Sekimoto, T. Horanont, S. Ueyama, and R. Shibasaki, "Modeling and probabilistic reasoning of population evacuation during large-scale disaster," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1231–1239.

[33] H. A. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, vol. 14, no. 3, pp. 105–122, 2000.

[34] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[35] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, "A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 2, pp. 179–192, Feb. 2010.

[36] J. Tang, G.-J. Qi, L. Zhang, and C. Xu, "Cross-space affinity learning with its application to movie recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 7, pp. 1510–1519, Jul. 2013.

[37] V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang, "Collaborative filtering meets mobile recommendation: A user-centered approach," in *Proc. AAAI Conf. Artif. Intell.*, 2010, pp. 236–241.

[38] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Towards mobile intelligence: Learning from GPS history data for collaborative recommendation," *Artif. Intell.*, vol. 184, pp. 17–37, 2012.

[39] Y.-R. Lin, J. Sun, P. Castro, R. Konuru, H. Sundaram, and A. Kelliher, "MetaFac: Community discovery via relational hypergraph factorization," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 527–536.

[40] X. Liu, S. Ji, W. Glänzel, and B. De Moor, "Multiview partitioning via tensor methods," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1056–1069, May 2013.

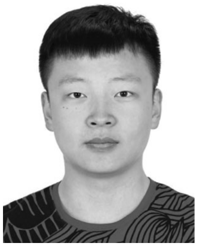
[41] L. Chen, J. Jakubowicz, D. Yang, D. Zhang, and G. Pan, "Fine-grained urban event detection and characterization based on tensor cofactorization," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 3, pp. 380–391, Jun. 2017.

[42] H. Fanaee-T and J. Gama, "Event detection from traffic tensors: A hybrid model," *Neurocomput.*, vol. 203, pp. 22–33, 2016.

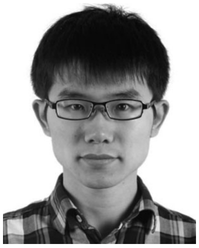
[43] Y. Cai, H. Tong, W. Fan, P. Ji, and Q. He, "Facets: Fast comprehensive mining of coevolving high-order time series," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 79–88.

Jingyuan Wang received the PhD degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China. He is currently an associate professor with the School of Computer Science and Engineering, Beihang University, Beijing, China. His is also the head of the Beihang Interest Group on SmartCity (BIGSCity), and vice director of the Beijing City Lab (BCL). His general area of research is data mining and machine learning, with special interests include smart cities, finance, and healthcare data analytics.

Junjie Wu received the BE degree in civil engineering and the PhD degree in management science and engineering from Tsinghua University. He is currently a full professor with the Information Systems Department, School of Economics and Management, Beihang University, the director of the Research Center for Data Intelligence (DIG), and the vice director of the Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations. His general area of research is data mining and complex networks, with special interests include social, urban, and financial computing. He is the recipient of various national awards in China, including NSFC Distinguished Young Scholars, MOE Changjiang Young Scholars, and MOE Excellent Doctoral Dissertation.



Ze Wang received the BS degree in electronics and information engineering from Beihang University, Beijing, China, in 2017. He is currently working as a research assistant with the School of Computer Science and Engineering, Beihang University. His research interests generally focus on machine learning and data mining.



Fei Gao received the MS degree from the School of Computer Science and Engineering, Beihang University, Beijing, China. He is currently an associate researcher with the Machine Learning Group, Microsoft Research Asia. His research interests include machine learning, distributed system, and data mining.

Zhang Xiong received the BS degree from Harbin Engineering University, in 1982, and the MS degree from Beihang University, in 1985. He is currently a professor and PhD supervisor with the School of Computer Science and Engineering, Beihang University, and the director of the Advanced Computer Science Application Technologies Research Center, MOE of China. His research interests include smart cities and data vitalization.