Beijing City Lab

Wang J, Gao F, Cui P, Li C, Xiong Z, 2014, Discovering urban spatio-temporal structure from time-evolving traffic networks. Beijing City Lab. Working paper #39

Discovering Urban Spatio-temporal Structure from Time-evolving Traffic Networks

Jingyuan Wang[1], Fei Gao[1], Peng Cui[2], Chao Li[1], and Zhang Xiong[1]

[1] School of Computer Science and Engineering, Beihang University
[2] Department of Computer Science and Technology, Tsinghua University

Abstract. The traffic networks reflect the pulse and structure of a city and shows some dynamic characteristic. Previous research in mining structure from networks mostly focus on static networks and fail to exploit the temporal patterns. In this paper, we aim to solve the problem of discovering the urban spatio-temporal structure from time-evolving traffic networks. We model the time-evolving traffic networks into a 3-order tensor, each element of which indicates the volume of traffic from i-th origin area to j-th destination area in k-th time domain. Considering traffic data and urban contextual knowledge together, we propose a regularized Non-negative Tucker Decomposition (rNTD) method, which discovers the spatial clusters, temporal patterns and relations among them simultaneously. Abundant experiments are conducted in a large dataset collected from Beijing. Results show that our method outperforms the baseline method.

Keywords: urban computing, time-series data mining, knowledge discovery

1 Introduction

Understanding the urban structure is important for urban planning, transportation management, epidemic prevention, and location based business. However, our knowledge in this area is limited. Nowadays, the rapid growth of information infrastructure collect huge volumes of trajectory data, such as GPS trajectories, mobiles and IC card records, from which we can build a “from-where-to-where”traffic network that indicates the volume of traffic from one origin area to another destination area. The traffic networks give us a great chance to study the detail of urban structures [2]. In this paper, we consider the problem of discovering urban spatio-temporal structure from time-evolving traffic networks. The goal is to discover spatial clusters of urban areas, temporal patterns of urban traffic and their correlations simultaneously, which we refer to urban spatio-temporal structure.

Significant progresses have been made on the problem about finding interesting structures from networks in the field of community detection [1]. For example, researchers have found protein clusters that have same specific function within the cell from protein-protein interaction networks [3]. Social circles can be mined from social networks [7]. The basic intuition shared by this methods is a find community that have more edges "inside"the community than edges "outside". However, in urban situation, challenges arise because the traffic network is different from network in previous research. First, the traffic network is time-evolving and dynamic. Traffic in the morning is apparently different from that in the evening. Second, the structure is also related with urban contextual information. A working area owns different structure from a residential area. Then, the structure we want to find need to consider not only edge information, but also the time and urban contexture attribution.

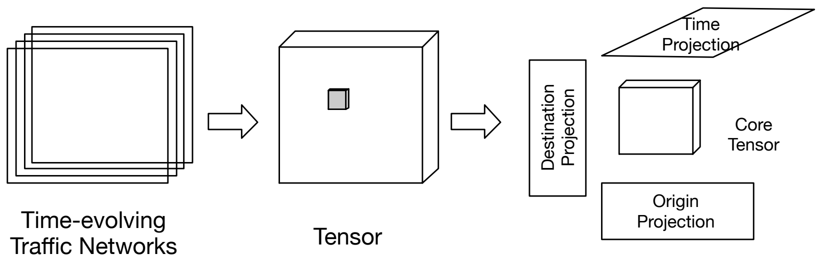


Fig. 1. Framework of rNTD

To solve these challenges, in this paper we model the time-evolving traffic networks into a 3-order *origin-destination-time* tensor [10], each element (i, j, k) of which represents the volume of traffic from the i-th origin area to the j-th destination area in the k-th time domain. Considering traffic data and urban contextual information together, we propose a regularized Non-negative Tucker Decomposition (rNTD) method to solve the problem, which decomposes the original tensor to three projection matrix and a core tensor. The projection matrix in *origin* and *destination* mode indicate which cluster an area belongs to. The projection matrix in *time* mode shows temporal patterns. And the core tensor gives the correlation between spatial clusters and temporal patterns. The framework of rNTD is shown in Fig. 1. We apply Alternating Proximal Gradient to optimize the rNTD problem, and carry out intensive experiments to demonstrate the effectiveness of our proposed method.

It's worthwhile to highlight the key contribution of this paper.

– We formulate the urban spatio-temporal structure discovery problem formally with rNTD, and devise an efficient proximal gradient method to solve it.
– The discovered urban spatio-temporal structure is easy to understand and well explained, which can support to solve some urban problem, such as urban planning, traffic jam and so on.
– We conducted intensive experiments on a real dataset collected from Beijing, and the results show that the rNTD can achieve a better performance compared with other competitors.

The rest of the paper is organized as follow: Section 2 summarizes related work. Section 3 gives the formulation of our problem and the method we propose. We give experimental results in Section 4 and conclude in Section 5.

2 Related Work

The research of urban computing have recently received much attention. Yuan *et al.* [14] discover regions of different functions in a city using human mobility data and POI. Zheng *et al.* [17] design an algorithm to detect flawed urban planning with GPS trajectories of taxicabs traveling in urban areas. Zhang *et al.* [15] propose an context-aware collaborative filtering method to sense the pulse of urban refueling behavior. Zheng *et al.* [16] give a novel method to infer the real-time air quality information throughout a city. Different from these study, we explore urban spatio-temporal structure from time-evolving traffic networks.

A great deal of work has been devoted to mining community structure from network [1]. Recently, Wang et al. [11] use Non-negative Matrix Factorization to discover community from undirected, directed and compound networks. Yang *et al.* [13] develop CESNA for detecting communities in networks with node attribution. Kim *et al.* [4] propose a nonparametric multi-group membership model for dynamic networks. Different from the above mentioned work, we mine time-evolving traffic networks by considering node attribution (*urban contextual information*) and temporal dynamic (*time-evolving*) together.

3 Regularized Non-negative Tucker Decomposition

3.1 Problem Formulation

In this section, we will introduce details of our model. First, we formally define the problem of urban spatio-temporal structure discovery. Suppose we have M areas in city with i-th origin area denoted as o_i and j-th destination area denoted as d_j. We split the day time uniformly to K time domain with k-th domain denoted as t_k.

We denote the time-evolving traffic networks data as a 3-order tensor $\hat{\boldsymbol{\mathcal{X}}} \in \mathbb{R}^{M \times M \times N}$, with its (i, j, k)-th entry $\boldsymbol{\mathcal{X}}_{ijk}$ represents the volume of traffic from i-th origin area to j-th destination area in k-th time domain. Since the distribution of the value in tensor $\hat{\boldsymbol{\mathcal{X}}}$ is severely skewed, to reduce the impact of high value, we use the *log* function for scaling the tensor data:

$$\boldsymbol{\mathcal{X}}_{ijk} = log_2 \left(1 + \hat{\boldsymbol{\mathcal{X}}}_{ijk}\right) \tag{1}$$

Then given the time-evolving traffic networks, the spatial and temporal structure discovery problem is converted to finding latent projection matrix in every mode and the correlation core tensor between them simultaneously from the traffic tensor. The projections of mode **O** and **D** summarize the spatial clusters on urban areas, and projections on **T** show the temporal patterns, and correlation between **O**, **D** and **T** give the a compact view of urban traffic. As show in Fig. 1. From the result, we can get the urban spatio-temporal structure.

3.2 Basic Tucker Decomposition

Let $\mathbf{O} \in \mathbb{R}^{M \times m}$ be the latent origin projection matrix, $\mathbf{D} \in \mathbb{R}^{M \times m}$ be the latent destination projection matrix, and $\mathbf{T} \in \mathbb{R}^{N \times n}$ be the latent time projection matrix. Let $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{m \times m \times n}$ be the latent core tensor. We have $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_m\}$, $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_m\}$, and $\mathbf{T} = \{\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_n\}$, where every column vector in projection matrix present the weight of every factor for each area or time accordingly and $\boldsymbol{\mathcal{C}}_{ijk}$ represents the correlation between latent factor $\mathbf{o}_i$, $\mathbf{d}_j$, and $\mathbf{t}_k$.

According to the Tucker Decomposition[5], we can factorize the tensor approximately based on the factor matrix and core tensor as

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{C}} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T} \tag{2}$$

where $\times_n$ represents the tensor-matrix multiplication on mode n.

Then, given the observed origin-destination-time tensor $\boldsymbol{\mathcal{X}}$, the objective of this paper is to find the optimal latent projection matrix $\mathbf{O}$, $\mathbf{D}$, $\mathbf{T}$ and core tensor $\boldsymbol{\mathcal{C}}$ by minimizing the following objectives

$$\mathcal{J}_1 = \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{C}} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}\|_F^2 \tag{3}$$

The objective function can be seen as the quality of approximation of tensor $\boldsymbol{\mathcal{X}}$ by the projection matrix $\mathbf{O}$, $\mathbf{D}$, $\mathbf{T}$ and core tensor $\boldsymbol{\mathcal{C}}$. However, as mentioned above, urban structure is also strongly related with urban contextual information, and the sparsity of X makes it very challenging to directly learn the structure from only observed urban traffic network. That's the reason why we need to make full use of the urban contextual information.

3.3 Urban Contextual Regularization

A point of interest(POI), is a specific point location that someone may find useful or interesting. A POI is associated with a geo-position (latitude, longitude) and a POI category, which implies the urban contextual information . Table 1 shows the information of POI category used in this paper. For each area, the number of POI in each POI category can be counted. Then the POI feature vector in area i can be denoted by $v_i = (c_1, c_2, ..., c_P)$, where the P is the number of POI categories, and c_k is the number of k-th category POI. Based on the POI feature vectors, we consider the *consin* distance of POI vectors to measure similarity of two areas.

$$W_{ij} = \frac{v_i \cdot v_j}{\|v_i\| \cdot \|v_j\|} \tag{4}$$

In this way, we can construct the area-area similarity matrix $\mathbf{W} \in \mathbb{R}^{M \times M}$. We further assume that $\mathbf{W}$ can be approximated by the inner product of the latent origin projection matrix and latent destination projection matrix respectively, thus we need to minimize the following objective

$$\mathcal{J}_2 = \|\mathbf{W} - \mathbf{O}\mathbf{O}^\top\|_F^2 \tag{5}$$

Table 1. Information of POI category

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

$$\mathcal{J}_3 = \|\mathbf{W} - \mathbf{D}\mathbf{D}^\top\|_F^2 \tag{6}$$

We also use non-negativity constraints and sparsity regularizer on the core tensor and/or factor matrices. Non-negativity allows only additivity, so the solutions are often intuitive to understand and explain[6]. Promoting the sparsity of the core tensor aims at improving the interpretability of the solutions. Roughly speaking, the core tensor interacts with all the projection matrices, and a simple one is often preferred. Forcing the core tensor to be sparse can often keep strong interactions between the projection matrices and remove the weak ones. Sparse projection matrices make the decomposed parts more meaningful and can enhance uniqueness, as explained in [8].

Finally, by combining $\mathcal{J}_1$, $\mathcal{J}_2$, $\mathcal{J}_3$, together, we can get the latent factor matrix $\mathbf{O}$, $\mathbf{D}$, $\mathbf{T}$ and latent core tensor $\boldsymbol{\mathcal{C}}$ by minimizing the following objective function.

$$\begin{aligned}\mathcal{J} =& \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{C}} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}\|_F^2 + \alpha\|\mathbf{W} - \mathbf{O}\mathbf{O}^\top\|_F^2 \\ &+\beta\|\mathbf{W} - \mathbf{D}\mathbf{D}^\top\|_F^2 + \gamma\|\boldsymbol{\mathcal{C}}\|_1 + \delta\|\mathbf{O}\|_1 + \epsilon\|\mathbf{D}\|_1 + \varepsilon\|\mathbf{T}\|_1 \\ s.t.\ & \boldsymbol{\mathcal{C}} \geq 0, \mathbf{O} \geq 0, \mathbf{D} \geq 0, \mathbf{T} \geq 0\end{aligned} \tag{7}$$

3.4 Optimization

In this section, we will introduce a Alternating Proximal Gradient(APG)[12] method to solve the optimization problem.

For convenience of description, We first introduce the basic APG method to solve the problem

$$\min_x F(\mathbf{x}_1, \cdots, \mathbf{x}_s) = f(\mathbf{x}_1, \cdots, \mathbf{x}_s) + \sum_{i=1}^{s} r_i(\mathbf{x}_i) \tag{8}$$

where variable $\mathbf{x}$ is partitioned into s blocks $\mathbf{x}_1, \cdots, \mathbf{x}_s$, f is a differentiable and for each i, it is a convex function of $\mathbf{x}_i$ while all the other blocks are fixed. Each $r_i(\mathbf{x}_i)$ is the regularization item on $\mathbf{x}_i$.

Then at the k-th iteration of APG, $\mathbf{x}_1, \cdots, \mathbf{x}_s$ are updated alternatively from $i = 1$ to s by

$$\mathbf{x}_i^k = \arg\min_{\mathbf{x}_i} \left\langle \mathbf{g}_i^k, \mathbf{x}_i - \mathbf{x}_i^{k-1} \right\rangle + \frac{L_i^k}{2} \|\mathbf{x}_i - \mathbf{x}_i^{k-1}\|_2^2 + r_i\left(\mathbf{x}_i\right) \tag{9}$$

where $\mathbf{g}_i^k = \nabla f_i^k \left(\mathbf{x}_i^{k-1}\right)$ is the block-patial gradient of f at $\mathbf{x}_i^{k-1}$, L_i is a Lipschitz constant of $\nabla f_i\left(\mathbf{x}_i\right)$, namely,

$$\|\nabla f_i\left(\mathbf{x}_{i_1}\right) - \nabla f_i\left(\mathbf{x}_{i_2}\right)\|_F \leq L_i \|\mathbf{x}_{i_1} - \mathbf{x}_{i_2}\|_F, \forall \mathbf{x}_{i_1}, \mathbf{x}_{i_2} \tag{10}$$

In this paper, we consider the non-negative and sparse regularization. Then the Equation 9 has closed form

$$\mathbf{x}_i^k = \max\left(0, \mathbf{x}_i^{k-1} - \frac{1}{L_i^k} \nabla f_i\left(\mathbf{x}_i^{k-1}\right) - \frac{\lambda_i}{L_i^k}\right) \tag{11}$$

Then we return to our problem. Although the objective function is not jointly convex with respect to $\boldsymbol{\mathcal{C}}$, $\mathbf{O}$, $\mathbf{D}$, and $\mathbf{T}$, it is convex with each of them with the other three fixed. We can adopt a block coordinate descent scheme to solve the problem. That is, starting from some random initialization on $\boldsymbol{\mathcal{C}}$, $\mathbf{O}$, $\mathbf{D}$, and $\mathbf{T}$, we solve each of them alternatively with the other three fixed, and proceed step by step until convergence. Specifically, the gradients of the objective $\mathcal{J}$ with respect to the variables are

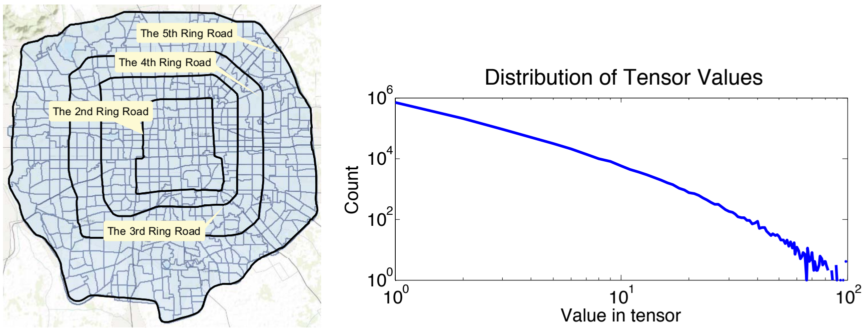
$$\begin{aligned}
\frac{\partial \mathcal{J}}{\partial \boldsymbol{\mathcal{C}}} &= 2\left(\boldsymbol{\mathcal{C}} \times_o \left(\mathbf{O}^\top \mathbf{O}\right) \times_d \left(\mathbf{D}^\top \mathbf{D}\right) \times_t \left(\mathbf{T}^\top \mathbf{T}\right) - \boldsymbol{\mathcal{X}} \times_o \mathbf{O}^\top \times_d \mathbf{D}^\top \times_t \mathbf{T}^\top\right) \\
\frac{\partial \mathcal{J}}{\partial \mathbf{O}} &= 2\Big(\mathbf{O}\left(\boldsymbol{\mathcal{C}} \times_d \left(\mathbf{D}^\top \mathbf{D}\right) \times_t \left(\mathbf{T}^\top \mathbf{T}\right)\right)_{(o)} \boldsymbol{\mathcal{C}}_{(o)}^\top - \left(\boldsymbol{\mathcal{X}} \times_d \mathbf{D}^\top \times_t \mathbf{T}^\top\right)_{(o)} \boldsymbol{\mathcal{C}}_{(o)}^\top \\
&\quad - \alpha\left(\mathbf{W} - \mathbf{O}\mathbf{O}^\top\right)\mathbf{O}\Big) \\
\frac{\partial \mathcal{J}}{\partial \mathbf{D}} &= 2\Big(\mathbf{D}\left(\boldsymbol{\mathcal{C}} \times_o \left(\mathbf{O}^\top \mathbf{O}\right) \times_t \left(\mathbf{T}^\top \mathbf{T}\right)\right)_{(d)} \boldsymbol{\mathcal{C}}_{(d)}^\top - \left(\boldsymbol{\mathcal{X}} \times_o \mathbf{O}^\top \times_t \mathbf{T}^\top\right)_{(d)} \boldsymbol{\mathcal{C}}_{(d)}^\top \\
&\quad - \beta\left(\mathbf{W} - \mathbf{D}\mathbf{D}^\top\right)\mathbf{D}\Big) \\
\frac{\partial \mathcal{J}}{\partial \mathbf{T}} &= 2\left(\mathbf{T}\left(\boldsymbol{\mathcal{C}} \times_o \left(\mathbf{O}^\top \mathbf{O}\right) \times_d \left(\mathbf{D}^\top \mathbf{D}\right)\right)_{(t)} \boldsymbol{\mathcal{C}}_{(t)}^\top - \left(\boldsymbol{\mathcal{X}} \times_o \mathbf{O}^\top \times_d \mathbf{D}^\top\right)_{(t)} \boldsymbol{\mathcal{C}}_{(t)}^\top\right)
\end{aligned} \tag{12}$$

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

4 Experiment

4.1 Data sets

In this section, we conduct extensive experiments to evaluate the effectiveness and show insight of our proposed method based on a real taxi trajectory dataset, which contains more than 3 millions occupied trips generated by taxis of Beijing in one month (November, 2011). According to the report of Beijing Transportation Bureau, the taxi trips occupy over 12 percent of traffic flows on road surface[14]. We split the Beijing map within 5-*th* Ring Road into 651 areas according to the traffic analysis zone as shown in Fig. 2(a), which is the most commonly used unit of geography. We also split 24 time domains according to hours. After the data preprocessing, we built a $(651 \times 651 \times 24)$ traffic tensor. Statistics about tensor data distribution shown in Fig. 2(b) reflects severe skewness. Our datasets also include a POI dataset in year 2011, which contains more than 30 thousands POI records.



(a) Areas in 5-th Ring Road of Beijing split by traffic analysis zone

(b) Distribution of values in tensor

Fig. 2. Description of datasets

4.2 Comparative Method

Besides the proposed rNTD method, we also implement the following methods for comparison.

- **Basic Non-negative Tensor Factorization** (bNTF): As a extension to non-negative matrix factorization, this method suppose a joint latent space for each mode by solving the objective function:

$$\min_{\mathbf{O},\mathbf{D},\mathbf{T}} \|\boldsymbol{\mathcal{X}} - \sum_r \mathbf{o}_r \odot \mathbf{d}_r \odot \mathbf{t}_r\|^2 \tag{13}$$

  where $\odot$ represents the vector outer product. This method can be also seen as the special case of NTD when the core tensor is super diagonal, also known as PARAFAC.
- **Regularized Non-negative Tensor Factorization** (rNTF): By incorporating the urban contextual regularization, this method consider the follow-

ing objective function

$$\min_{\mathbf{O},\mathbf{D},\mathbf{T}} \|\boldsymbol{\mathcal{X}} - \sum_r \mathbf{o}_r \odot \mathbf{d}_r \odot \mathbf{t}_r\|^2 + \alpha\|\mathbf{W} - \mathbf{O}\mathbf{O}^\top\|_F^2 + \beta\|\mathbf{W} - \mathbf{D}\mathbf{D}^\top\|_F^2 \quad (14)$$

- **Basic Non-negative Tucker Decomposition** (bNTD): This is a variant of our method, but with no consideration about regularization terms by solving:

$$\min_{\boldsymbol{\mathcal{C}},\mathbf{O},\mathbf{D},\mathbf{T}} \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{C}} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T}\|_F^2 \quad (15)$$

We evaluate the quality of patterns we discover by tensor reconstructed error using the *Root Mean Square Error*(RMSE)

$$RMSE = \sqrt{\frac{\sum_{ijk\in\boldsymbol{\mathcal{H}}} \boldsymbol{\mathcal{X}}_{ijk} - \hat{\boldsymbol{\mathcal{X}}}_{ijk}}{|\boldsymbol{\mathcal{H}}|}} \quad (16)$$

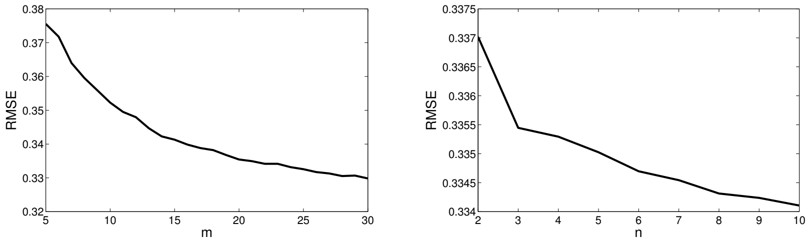
where $\boldsymbol{\mathcal{H}}$ is the set of hidden elements in our experiment and $|\boldsymbol{\mathcal{H}}|$ is its size, $\hat{\boldsymbol{\mathcal{X}}}$ is the reconstructed tensor.

4.3 Parameter Settings

In this section, we report the sensitivity of parameters our method involves, the dimensionality of hidden space m, n, and the tradeoff parameter for urban contextual regularization α, β.

Dimensionality of Hidden Space The goal of our model is to find a $m\times m\times n$-dimensional space for origin areas, destination areas and time. How to set m and n is important for our problem. If m, n are too small, the cluster can not be well represented and discriminated in the latent space. If m, n are too large, the low-rank structure would not capture the relations between different dimensions and the computational complexity will be greatly increased. Thus, we conduct 10 experiments with m ranging from 5 to 30 and n ranging from 2 to 10 on the dataset. The result are shown in Fig. 3, from which we can see that with the increase on the dimension m, n, $RMSE$ will reduce gradually. When $m \geq 20, n \geq 3$, the $RMSE$ reduces rather slow. For the concern of the tradeoff between precision and explainaration, we choose $m = 20$ and $n = 3$ as latent space dimension in our experiments.

Tradeoff Parameters The tradeoff parameters α, β in our method play the role of adjusting the strength of different terms in the objective function. Fig. 4 shows the $RMSE$ when α, β changes from 10^{-5} to 1. When α, β are small, the performance is close to that of bNTD, as we will see in Table 2. However, when α, β are relatively large, the optimization in Equation 7 my be dominated by the urban contextual regularization term, therefore the reconstructed loss term is not properly optimized. The result in Fig. 4 show that the parameter set $\alpha = \beta = 0.01$ produce the best performance. In our following, we just use this parameter setting. Moreover, we also find the best configurations in every comparative methods on our datasets to make sure comparisons are fair.



(a) (b)

Fig. 3. Dimensionality of hidden space.

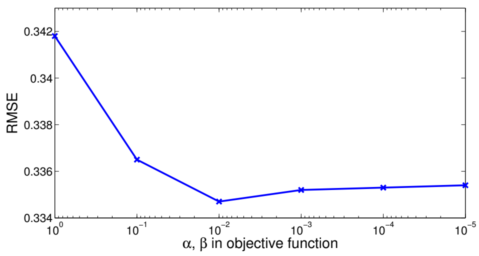


Fig. 4. Varing the city context regularization coefficient α, β

4.4 Experimental Performance

To do the comparision, We randomly select 50% 70% and 90% of the observed entries in tensor as training dataset and compute the reconstruction error of the hidden entries, where we obtain the $RMSE$. We repeat the experiments 10 times and report the average performance of all methods in Table 2,

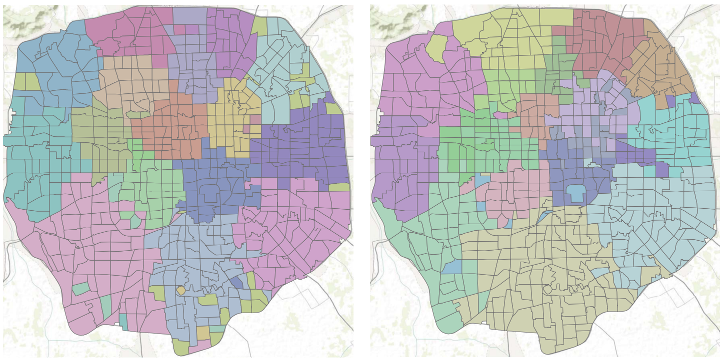
From the Table 2, we can observed that:

Table 2. Experiment performance

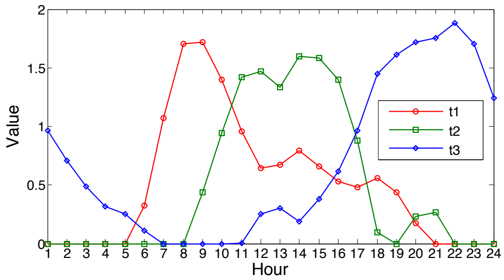
	bNTF	rNTF	bNTD	rNTD
50%	0.3974	0.3952	0.3366	**0.3359**
70%	0.3970	0.3950	0.3361	**0.3352**
90%	0.3963	0.3943	0.3354	**0.3347**

- The comparison between bNTF v.s. bNTD and rNTF v.s. rNTD reveals the advantage of the tensor Tucker decomposition as a method to capture relation from high dimensions data and get stable and compact representation.
- The advantage of rNTD over bNTD, as well as the advantage of rNTF over bNTF, shows shows the importance of urban contextual in our problem.

– Finally, our proposed method, rNTD, which incorporates the spatio-temporal interaction data and urban contextual information together, achieves the best performance in all experimental trials.



(a) Clusters of origin areas (b) Clusters of destination areas



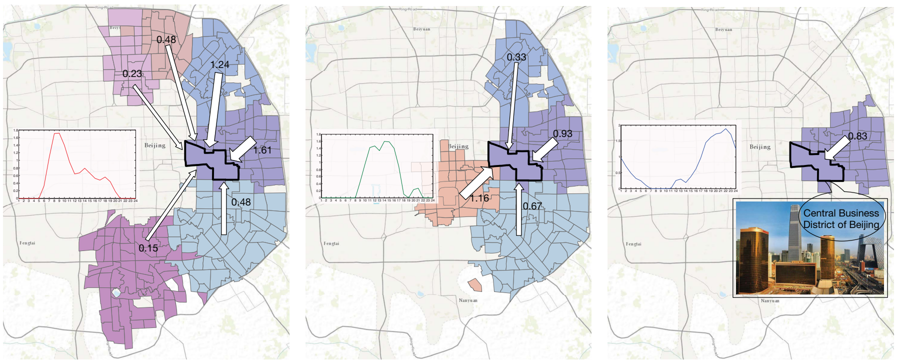
(c) Three temporal patterns

Fig. 5. Spatial clusters of Beijing areas and temporal patterns of Beijing traffic

4.5 Insights

In this section, we will give the empirical study to show the insights that founded by our method. As shown in Fig. 1, the goal of urban spatio-temporal structure discovery is to find a low-rank tensor, from which we can not only identify cluster behaviors on the *origin*, *destination*, *time* modes, but can also detect the cross-mode association.

The projection on **O** and/or **D** gives the spatial correlation information among the areas in city. The entries with high values in a columns of **O** and/or **D** imply which cluster an area belongs to. The result is visualized in Fig. 6(a) and Fig. 6(b). Surprisingly, we see that although our method make no use of geography information, spatial cluster are geographically close. In Fig. 6(c), we plot the three columns of projection matrix **T**. Along the time mode, three temporal



(a) Temporal pattern 1 (b) Temporal pattern 2 (c) Temporal pattern 3

Fig. 6. Visualization of core tensor: within the black line is the a destination cluster for case study, which is Central Business District (CBD) of Beijing. Three subfigures represent three temporal patterns respectively. Each subfigure gives the origin cluster that have correlations with CBD. Values are from the core tensor.

patterns are found. This result is similar with the work in [9], where three patterns are explained as home to workspace, workspace to workspace, workspace to others. Although the patterns are similar, our method can find also the corresponding spatial patterns and the correlation between them, thus the better understanding of spatio-temporal structure of city.

We choose the Central Business District (CBD) of Beijing as a destination cluster for case study, which is shown within the black line in Fig. 6. We visualize the origin cluster that have non-zero value with CBD destination cluster in core tensor in three temporal patterns accordingly. From Fig. 6(a), we see that most traffic to CBD, as a typical workplace, reach peak in the morning and fade down soon. In the daytime, most traffic occur in the nearby area. Since CBD is one of the most congested place in Beijing, Fig. 6 give an intuitional guidance for urban planning to solve traffic jam.

5 Conclusion and Future Work

In this paper we investigate the problem of discovering urban spatio-temporal structure from time-evolving traffic network. We model the time-evolving traffic network into a 3-order origin-destination-time tensor. We propose a regularized Non-negative Tucker Decomposition (rNTD) method to solve this problem, which consider traffic data and urban contextual knowledge together. We also propose an alternating proximal gradient algorithm for optimization. Experimental results on a real-world dataset show that our method can significantly outperform the baseline methods.There are many potential directions of this work. It would be interesting to investigate how the POI affect urban traffic, which will give more guideline for urban planning.

References

1. Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.
2. Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
3. Pall F Jonsson, Tamara Cavanna, Daniel Zicha, and Paul A Bates. Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis. *BMC bioinformatics*, 7(1):2, 2006.
4. Myunghwan Kim and Jure Leskovec. Nonparametric multi-group membership model for dynamic networks. In *Advances in Neural Information Processing Systems*, pages 1385–1393, 2013.
5. Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.
6. Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
7. Julian J McAuley and Jure Leskovec. Learning to discover social circles in ego networks. In *NIPS*, volume 272, pages 548–556, 2012.
8. Morten Mørup, Lars Kai Hansen, and Sidse M Arnfred. Algorithms for sparse nonnegative tucker decompositions. *Neural computation*, 20(8):2112–2131, 2008.
9. Chengbin Peng, Xiaogang Jin, Ka-Chun Wong, Meixia Shi, and Pietro Liò. Collective human mobility pattern from taxi trips in urban area. *PloS one*, 7(4):e34487, 2012.
10. Jimeng Sun, Dacheng Tao, and Christos Faloutsos. Beyond streams and graphs: dynamic tensor analysis. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 374–383. ACM, 2006.
11. Fei Wang, Tao Li, Xin Wang, Shenghuo Zhu, and Chris Ding. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 22(3):493–521, 2011.
12. Yangyang Xu and Wotao Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6(3):1758–1789, 2013.
13. Jaewon Yang, Julian McAuley, and Jure Leskovec. Community detection in networks with node attributes. *2013 IEEE 13th International Conference on Data Mining*, 0:1151–1156, 2013.
14. Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 186–194. ACM, 2012.
15. Fuzheng Zhang, David Wilkie, Yu Zheng, and Xing Xie. Sensing the pulse of urban refueling behavior. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 13–22. ACM, 2013.
16. Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. U-air: when urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1436–1444. ACM, 2013.
17. Yu Zheng, Yanchi Liu, Jing Yuan, and Xing Xie. Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 89–98. ACM, 2011.