Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.083824

ARTICLE

# A Spatial-Temporal Traffic Prediction Algorithm Based on Non-Negative Tensor Factorization

**Xiaoxiong Yang[1,2], Dingde Jiang[1,*], Yi Zhang[1] and Zhihan Lyu[3]**

[1]School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
[3]School of Computer Science and Technology, Xidian University, Xi'an, China

*Corresponding Author: Dingde Jiang. Email: jiangdd@uestc.edu.cn

Received: 11 April 2026; Accepted: 07 May 2026; Published: 15 June 2026

**ABSTRACT:** The advancement of communication technology has made traffic engineering a critical issue in network systems. The traffic matrix is essential data that supports traffic engineering. The functionality of routing planning, network monitoring, and other modules within intelligent network management systems relies heavily on the network traffic matrix. However, real-time measurement of the network traffic matrix is costly and often suffers from missing or anomalous values. Consequently, long-term network traffic prediction presents significant challenges. Existing methods often fail to comprehensively address the multidimensional characteristics of traffic and the computational costs of the algorithms. To address these issues, we propose an efficient traffic prediction algorithm based on tensor factorization. First, we introduce a non-negative tensor factorization algorithm that accounts for link errors. This algorithm captures the spatial-temporal characteristics of traffic from different modes, thereby enhancing prediction efficiency. Next, we integrate the tensor factor matrix with a seasonal differential autoregressive moving average model in the temporal mode to identify traffic trends and complete the traffic prediction. Experimental results based on real data demonstrate that our algorithm performs exceptionally well in multi-step predictions and in capturing abnormal fluctuations.

**KEYWORDS:** Flow; tensor factorization; multidimensional features; traffic prediction

## 1 Introduction

The widespread adoption of mobile devices has made the Internet of Things (IoT) a crucial infrastructure. The large-scale and heterogeneous characteristics of network systems, such as space-air-ground integrated networks, vehicular networks, and edge computing networks, are becoming increasingly significant. Concurrently, the volume of network traffic data is growing exponentially, leading to issues such as traffic congestion, routing chaos, and challenges in network management [1]. To effectively maintain the network, operators need to grasp the distribution of network traffic. Traffic collection is a fundamental component of network management systems. The origin-destination (OD) flow represents the traffic between any origin and destination node within the network. Link loads can be monitored using tools such as the simple network management protocol (SNMP). However, in practice, measuring OD flows directly poses significant challenges. The overhead of measuring OD flows using tools like NetFlow and sFlow is quite high. Furthermore, the actual measured OD flow data is often incomplete and may contain missing information. Consequently, developing a method to obtain a usable flow matrix at a low cost has become an urgent issue.

To address the aforementioned challenges and enhance network planning, predicting the network traffic matrix (TM) has become a crucial task in network systems. Traffic forecasting is defined as an inference task based on historical data. Traffic estimation is defined as the reverse mapping of OD traffic from link load. By collecting data on all OD flows within the network and organizing it according to sampling time, one can derive the traffic matrix. The TM serves as the foundation for resolving issues such as route engineering [2] and anomaly detection [3]. Network traffic exhibits characteristics that are multi-modal and non-linear, which significantly increases the time and space complexity associated with computational processing of the network traffic matrix. Most early prediction algorithms relied on group intelligent optimization or statistical methods [4]. These algorithms typically operate on vector or matrix forms, and the constraints imposed by data dimensionality hinder the ability to simultaneously capture the potential relationships among various traffic data patterns. Consequently, this limitation affects the accuracy of predictions and increases the likelihood of falling into local optimization traps [5]. Deep reinforcement learning (DRL) has demonstrated exceptional performance in a variety of complex fields. Examples include automated control, autonomous driving decision-making, and dynamic network resource orchestration. These algorithms offer significant advantages in handling high-dimensional nonlinear features and long-range dependencies. With the rapid advancement of neural network technology, DRL has gained widespread application in traffic prediction problems [6]. However, challenges such as limitations in Euclidean space and the risk of overfitting can impede the performance of these algorithms. Furthermore, DRL-based approaches often entail lengthy training cycles and relatively high operational costs.

Tensor is an extension of vectors and matrices to higher orders, allowing for the preservation of the underlying structure of complex data [7]. Tensor factorization decomposes raw data into a series of elementary component factors. It is widely utilized in various fields, including signal processing and machine learning, and has become a common tool for processing high-dimensional complex data. In the context of traffic matrix data, tensor factorization effectively captures its spatial and temporal characteristics [8]. Therefore, we introduce tensor factorization into the network traffic matrix prediction problem to make predictions based on the multidimensional features of the data, such as time, space, and periodicity. The specific contributions of this paper are as follows:

- We propose a physical constraint embedded non negative tensor factorization framework. This algorithm preserves the spatial and temporal characteristics of the data while reducing its dimensionality under non-negative constraints, thereby enhancing its nonlinear processing capabilities.
- We propose a decoupled spatio-temporal framework that combines tensor decomposition with seasonal autoregressive modeling. This framework effectively separates spatial physical constraints from temporal evolution patterns. This design enables the model to derive time-series components, improve forecasting efficiency, and perform multi-step traffic flow forecasting.
- In this paper, we conduct a substantial number of simulation experiments utilizing real data sets. Compared to other algorithms, the algorithm proposed in this paper can realize lower prediction errors.

The rest of the paper is organized as follows. Section 2 describes the related work on tensor factorization and traffic prediction. Section 3 addresses the network traffic prediction problem. Section 4 describes the proposed algorithm. Section 5 describes the testing environment and results of the algorithm. Finally, Section 6 gives the conclusion.

## 2 Related Works

In this section, we review the research work related to this paper.

### *2.1 Tensor Factorization*

A tensor is a modeling framework for multidimensional flow data. Through tensor factorization and reconstruction, the dimensionality of the data can be reduced while preserving its nonlinear characteristics [9]. The authors of literature [10] proposed an enhanced tensor decomposition model based on a Bayesian algorithm, which effectively fills in missing values in traffic flow data. Literature [11] evaluated the effectiveness of several tensor factorization algorithms for completing traffic flow data and demonstrated the feasibility of indirectly completing missing data through predictive modeling. To address the issues of missing values and outliers in the dataset, an online robust tensor recovery algorithm was introduced in literature [12]. This algorithm achieves lower error rates without incurring additional overhead. Literature [13] proposed a tensor ring factorization-based method for data completion that adapts to various network traffic data missingness under continuous changes. However, the above algorithm does not achieve the prediction of future data. In response, reference [14] also introduced a traffic data prediction algorithm based on CANDECOMP/PARAFAC (CP) decomposition and the Holt-Winters method. This algorithm is efficient in detecting outliers and making predictions.

Since tensor factorization is effective in capturing features such as mutations and spatial-temporal relationships, much of the research utilizes it for data completion. This demonstrates its feasibility and serves as inspiration for traffic prediction.

### *2.2 Traffic Prediction*

Traffic management is an important function of network management systems. Consequently, traffic prediction has emerged as a key research direction for future intelligent network systems. In the context of software-defined networks (SDN), reference [15] proposed a traffic prediction algorithm that leverages multi-task learning. This algorithm utilizes the spatial-temporal characteristics of traffic to enhance prediction accuracy. Literature [16] proposed a network traffic prediction algorithm tailored for SDN-optical network architecture, employing a graph convolutional layer to capture bursty correlations between nodes. Additionally, reference [17] designed a dynamic graph neural network that accounts for missing values, addressing the traffic flow prediction problem under conditions of incomplete data. Literature [18] proposed a novel attention mechanism and a weighted loss function that enable graph neural networks to effectively perform OD flow prediction. The authors of literature [19] utilized canonical correlation analysis to extract features from the original dataset and developed a multi-view subspace learning method for traffic prediction. Literature [20] explored the dynamic features of the data to enhance nonlinear processing capabilities and designed a multi-scale traffic prediction algorithm based on a deep echo state network. However, this method requires large-scale data support.

Traditional heuristic algorithms rely on predetermined rules to capture the multidimensional features of traffic data. Researchers are investigating new methods, such as deep learning and reinforcement learning, to enhance the prediction capability. However, artificial neural networks require a substantial amount of data for training and impose additional demands on computational resources [21]. Therefore, this paper employs non-negative tensor factorization to reduce computational complexity, enhance the algorithm's nonlinear processing capabilities, and achieve efficient multi-step traffic prediction.

## 3 Problem Explanation

In this section, we describe the traffic prediction problem model and the associated evaluation metrics. This provides a theoretical basis for subsequent algorithm design and testing.

### 3.1 Modeling of the Flow Forecasting Problem

The underlying network system comprises network nodes and links, which can be represented by $G = \{N, L\}$. Here, denotes the set of network entity nodes, such as sensors and routers, while represents the links between these nodes. In this paper, we define the Traffic Matrix (TM) as a representation of the traffic between any pair of source-destination nodes, as follows

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^{|N|^2} \\ x_2^1 & x_2^2 & \cdots & x_2^{|N|^2} \\ \vdots & \vdots & \ddots & \vdots \\ x_T^1 & x_T^2 & \cdots & x_T^{|N|^2} \end{bmatrix}, \tag{1}$$

where $x_t^n \in X_t$ denotes the value of the $n$-th OD flow in the $t$-th time slice. The definition of the traffic matrix contains a diagonal term. These items represent the traffic from the node to itself. A network topology with $n$ nodes contains $n^2$ OD flows. The OD flows in the network must satisfy a non-negativity constraint, i.e., $x_t^n \geq 0$. Typically, the traffic matrix is not readily available. However, link load, which represents the flow on each segment of link $l \in L$, is more accessible and cost-effective to obtain. In addition to utilizing measurement tools, link load can be derived from the routing matrix and OD flows, as shown in Eq. (2).

$$Y_t = RX_t, \tag{2}$$

where $Y_t = (y_t^1, y_t^2, \ldots, y_t^{|L|})$ denotes the link load at moment $t$ and $R$ denotes the routing matrix. The routing matrix $R$ is a binary matrix that illustrates the relationship between OD flows and link loads. This matrix can be derived directly from the network configuration information. The objective of the traffic prediction problem is to forecast the future data $x_{t+1}^n$, which can be defined as follows, utilizing the known $m$ pieces of traffic data $\{x_t^n, x_{t-1}^n, \cdots, x_{t-m+1}^n\}$:

$$\begin{cases} \hat{X}_{t+1} = f(X_t, X_{t-1}, \cdots, X_{t-m+1}) \\ s.t. \qquad Y_t = RX_t \\ \qquad x_t^i \in X_t \geq 0, i = 1, 2, \ldots, N. \end{cases}, \tag{3}$$

where $f(x)$ denotes the prediction algorithm used in this paper.

### 3.2 Algorithmic Evaluation Indicators

In this paper, the accuracy of the algorithm is assessed using multiple errors between the predicted and true values.

$$ASAE = \frac{\sum_t^T |\hat{x}_t^n - x_t^n|}{T}. \tag{4}$$

$$ATAE = \frac{\sum_i^N |\hat{x}_t^i - x_t^i|}{N}. \tag{5}$$

Eq. (4) is used to calculate the average spatial absolute error, effectively representing the estimation error of the algorithm across various OD flows. This metric verifies the effectiveness of the algorithm in dealing with the spatial correlation of the flow matrix. Conversely, Eq. (5) calculates the average temporal absolute error, which reflects the estimation error of the algorithm at different time points $t$. This metric is able to verify the effectiveness of the algorithm in terms of temporal correlation.

$$SRE(n) = \frac{\left\|\hat{x}_T^n - x_T^n\right\|_2}{\left\|x_T^n\right\|_2}. \tag{6}$$

$$TRE(t) = \frac{\left\|\hat{x}_t^N - x_t^N\right\|_2}{\left\|x_t^N\right\|_2}. \tag{7}$$

Eqs. (6) and (7) are used to calculate the spatial relative error (SRE) and the temporal relative error (TRE), respectively. The relative error eliminates the effect of flow amplitude on the error.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{x}_t^i - x_t^i\right)^2}. \tag{8}$$

Eq. (8) represents the root mean square error (RMSE), which is used to measure the general precision of the algorithm predictions.

## 4 Network Traffic Prediction Algorithm

In the field of multidimensional data analysis, tensor factorization can effectively extract key information from the data. In this paper, we propose the MTF algorithm, a network traffic prediction algorithm based on multimodal non-negative tensor factorization. First, the algorithm develops a non-negative tensor factorization model to extract the temporal pattern factor matrix from the original traffic data. Subsequently, a regression model is employed to perform the network traffic prediction.

### *4.1 Non-Negative Tensor Factorization Algorithm*

Tensor factorization accounts for the spatial and temporal correlations within the data and holds significant potential for managing high-dimensional datasets. From the perspective of network traffic prediction and data analysis, non-negative tensor factorization aligns more closely with practical characteristics than tensor factorization. Non-negative tensor factorization (NTF) incorporates non-negative constraints, which effectively capture the nonlinear features present in the data. A high-order tensor can be viewed as composed of several one-dimensional factor matrices. In this paper, historical flow data is structured as a three-dimensional tensor of size $12 \times 12 \times 2016$, as illustrated in Fig. 1. Specifically, we construct a three-dimensional tensor, where the first dimension represents the source node, the second dimension represents the target node, and the third dimension represents the time slice. Given a Z-dimensional tensor, tensor factorization decomposes $T$ into a sum of outer products of vectors. In this study, we propose a non-negative tensor factorization model that accounts for link errors. First, the non-negative tensor model can be expressed as Eq. (9).

$$t_{(i_1,g_1),\ldots,(i_z,g_z)} = \sum_{r=1}^{R} f^1_{(i_1,g_1)r} f^2_{(i_2,g_2)r} \cdots f^z_{(i_z,g_z)r} + e_{g_1,\ldots,g_z}, \tag{9}$$

where $t$ is an element in $T$. $x^z_{(i_z,g_z)r}$ is the $g$-th row and $r$-th column element of the $z$-th factor matrix. $e$ denotes the error. Here, $R$ is an integer greater than zero, representing the rank of the tensor. The resulting tensor factorization of $T$ can be expressed as follows:

$$T_{i_1,\ldots,i_z} \approx \sum_{r=1}^{R} F^1_{i_1 r} \circ F^2_{i_2 r} \circ \cdots \circ F^Z_{i_z r}, \tag{10}$$

where $F^z_{i_1 r} \in \mathbf{F}^z$ and $\circ$ denotes the outer product of the tensor. $\mathbf{F}^z = [F^z_1, \cdots, F^z_r, \cdots, F^z_R]$ denotes the $z$-th factorization matrix of the tensor. In the problem described in this paper, the factor matrices $\mathbf{F}^1$, $\mathbf{F}^2$, and $\mathbf{F}^3$ correspond to the source, target, and time modes, respectively. Physically, $\mathbf{F}^1$ and $\mathbf{F}^2$ characterize the spatial topology and node pairs, while $\mathbf{F}^3$ captures the underlying temporal evolution patterns. The factor matrix for the tensor factorization can be determined by the following optimization problem.

$$\min_{X^1, X^2 \cdots X^Z} \left\| T - \left[\left[ F^1, F^2 \cdots F^Z \right]\right] \right\|_F^2 + \lambda \sum_{t=1}^{T} \| y_t - R x_t \|. \tag{11}$$

$$\left[\left[ \mathbf{F}^1, \mathbf{F}^2, \cdots, \mathbf{F}^Z \right]\right] = \sum_{r=1}^{R} F^1_{i_1 r} \circ F^2_{i_2 r} \circ \cdots \circ F^Z_{i_z r}. \tag{12}$$

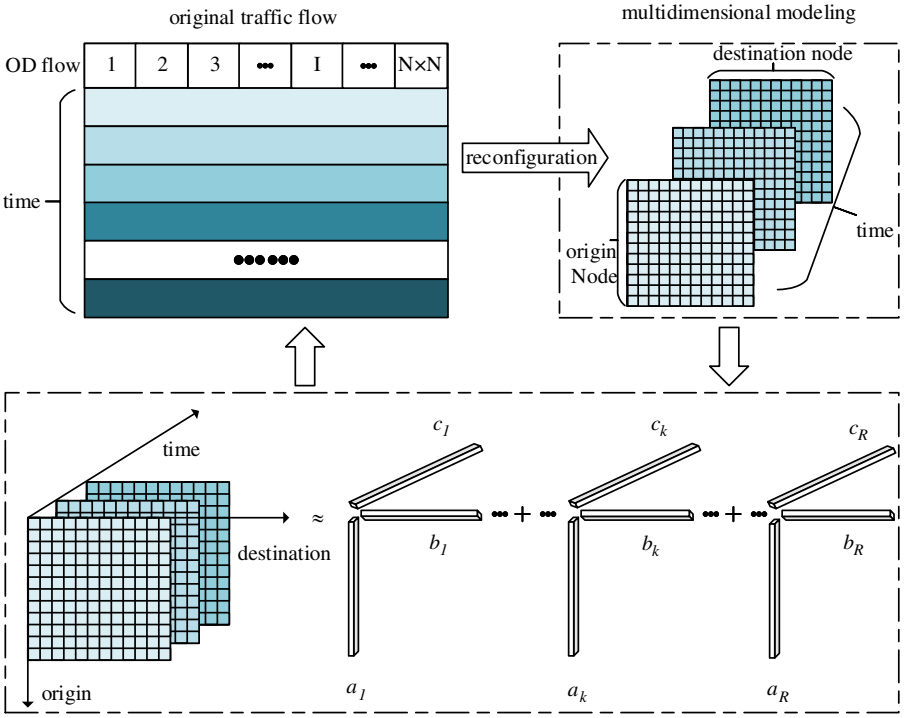


**Figure 1:** Tensor factorization model.

To bridge the gap between abstract tensor components and physical network constraints, we introduce the link-error term into the objective function. This term ensures that the reconstructed OD flows from the tensor factor matrices still satisfy the fundamental routing conservation law.

In the solution process, each $\mathbf{F}^z$ is initially assigned random positive values. Subsequently, two definitions are provided.

**Definition 1:** $S_{i_j, I} = T_{i_1}, \ldots, _{i_z}$. Fix the dimension $j$ and combine the indices of the other $z-1$ dimensions into a new index $I$. Expand the original tensor $X$ into a two-dimensional matrix along the currently processed dimension $j$ to facilitate matrix operations.

**Definition 2:** Construct a basis matrix $U_{I,r}$ to represent the product of the other dimension factor matrices.

$$U_{I,r} = F^1_{i_1 r} \circ \cdots \circ F^{j-1}_{i_{j-1} r} \circ F^j_{i_j r} \circ \cdots \circ F^Z_{i_z r}. \tag{13}$$

In order to enhance the accuracy of multi-step traffic predictions, we take into account the link error during the updating process of the factor matrix.

$$gl = \lambda R^T sign(Rx - y), \tag{14}$$

where $x$ denotes the OD flow after reconstruction using the current factor matrix, $y$ represents the corresponding link load, and $\lambda$ is the regulation factor. Eq. (14) is employed to calculate the error gradient of the OD flow in relation to the known link flow. In this paper, the expansion of the tensor into a third dimension corresponds to the time factor matrix. When updating the factor matrix along the time dimension, we compute the link error gradient. The reconstruction results from the tensor factorization are optimized by backpropagating the link flow error to the time factor. The MTF algorithm uses link errors to correct the time factor in real time, enhancing its ability to identify and track abnormal fluctuations. This feedback mechanism enables the model to accurately capture traffic bursts rather than treating them as random noise. This process is illustrated in the following equation.

$$gt_r = gl \odot (F_r^1 \circ F_r^2). \tag{15}$$

Next, the factor matrices are optimized sequentially using the multiplicative update rule, as demonstrated in Eq. (16).

$$F^j \leftarrow F^j \odot \frac{S \cdot U + O(gt_r, j)}{F^j \cdot U^T \cdot U + \varepsilon}. \tag{16}$$

$$O(gt_r, j) = \begin{cases} gt_r, & j = TimeD. \\ 0, & else. \end{cases}. \tag{17}$$

In Eq. (16), $\odot$ denotes element-wise multiplication and $\cdot$ denotes matrix multiplication. The rule incorporates $\varepsilon = 10^{-16}$ to prevent division by zero. Through element-wise multiplication and division, non-negativity is ensured, and the optimal solution is approximated incrementally. It is important to note that we only consider link errors during the update of the factor matrix for the time model. In the data utilized in this paper, the time dimension is the third dimension. Therefore $TimeD = 3$.

In order to balance the paradigms of the individual factor matrices and prevent severe anomalies in individual values, they are normalized using Eq. (18).

$$F_{i_j r}^j \leftarrow F_{i_j r}^j \frac{\left(\prod_{i=1}^z K_r^j\right)^{\frac{1}{z}}}{K_r^j}. \tag{18}$$

$$K_r^j = \sqrt{\sum_{i_j} \left(A_{i_j r}^j\right)^2}. \tag{19}$$

The factor matrices for each dimension are updated sequentially using the multiplicative update rule described above. The stability of the final algorithm is ensured through normalization.

### *4.2 SARIMA Forecasting Model Based on Tensor Factorization*

For the selection of a time series forecasting model, this algorithm employs the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. Since the time factor matrix $\mathbf{F}^3$ extracted via tensor decomposition exhibits highly pronounced periodic characteristics, SARIMA is capable of accurately capturing these linear evolution patterns through explicit seasonal, differencing, and autoregressive terms. Furthermore, compared to deep learning models that require large-scale training datasets and incur significant

computational overhead, SARIMA demonstrates higher computational efficiency and faster convergence when processing low-dimensional factor sequences obtained through tensor decomposition. The model expression is

$$\phi_p(L)\Phi_P\left(L_s\right)(1-L)^d(1-L_s)^D X_t = \theta_q(L)\Theta_Q\left(L_s\right)_t. \tag{20}$$

where $s$ represents the seasonal cycle, $L$ denotes the delay operator, and $L_s$ indicates the seasonal delay operator. The parameter $p$ refers to the autoregressive order, which signifies that the current value depends on the values from the previous $p$ time steps. Based on the model, the value of $\hat{X}_{t+h}$ can be predicted for the next $h$ steps.

$$\hat{X}_{t+h} = \sum_{i=1}^{p}\phi_i\hat{X}_{t+h-i} + \sum_{i=1}^{P}\Phi_i\hat{X}_{t+h-i\times s} + \sum_{i=1}^{q}\theta_{i\,t+h-i} + \sum_{i=1}^{Q}\Theta_{i\,t+h-i\times s}. \tag{21}$$

After obtaining the factor matrix $\hat{F}^Z$ for the future time step, tensor reconstruction is performed by combining the other factor matrices $F^1, F^2, \cdots, F^{Z-1}$ obtained from the historical decomposition. As a result, the predicted value of OD flow at the future time is derived. The predictive performance of this algorithm stems primarily from the deep synergy between tensor decomposition and seasonal forecasting models. Tensor decomposition acts as a spatio-temporal filter, projecting the raw traffic data onto a low-rank factor space to isolate underlying trends from random fluctuations and link errors. During the forecasting phase, these factors are then extrapolated along the time axis.

### *4.3 Complexity Analysis*

The computational cost of the non-negative tensor factorization stage in a single iteration is approximately $O(IJKR)$, where $I, J, K$ represent the sizes of the tensor's dimensions, respectively. Since the prediction stage operates only on the $R$ latent factor sequences obtained from the decomposition, its specific prediction complexity is $O(R \cdot K \cdot m^2)$, where $K$ is the sequence length and $m$ is the model order parameter. Compared to the $O(N^2 \cdot K \cdot m^2)$ complexity required for direct prediction on the original traffic matrix, this algorithm reduces computational requirements by a factor of $N^2/R$. Since the number of OD flows $N^2$ in large-scale networks far exceeds the latent rank $R$, this inherent dimensionality reduction ensures that the algorithm possesses excellent scalability when processing massive amounts of data.

In addition, the MTF framework supports streaming data through incremental gradient updates. It adapts to evolving topologies by dynamically adjusting the routing matrix $A$. This spatio-temporal decoupling design ensures both deployment efficiency and physical consistency in dynamic environments.

## 5 Simulation Result and Analysis

To validate the performance of the proposed algorithm, quantitative experiments are designed in this section to evaluate the algorithm.

### *5.1 Simulation Environment and Parameter Settings*

The Abilene dataset [22] used in the experiment is a standard benchmark in the field of traffic forecasting. The network topology consists of 12 nodes and 54 directed links, generating a total of 144 OD flows. Traffic data is sampled at 5-min intervals. The experiment utilized data collected over three consecutive weeks, totaling 6048 time slices. This data includes characteristics commonly found in real-world network environments, such as missing values, measurement noise, and abnormal traffic bursts. We selected 2016 data points from the first week as training data for the algorithm, while the remaining 4032 data points

were used to evaluate the algorithm's performance. We set the seasonality parameter to 12. The algorithms selected for comparison include SOFIA [14], TomoG [23], Inverse [24], and SRSVD [25]. TomoG and Inverse represent classical gravitational models and information-theoretic inference methods, respectively. SRSVD uses low-rank singular value decomposition to extract flow structure features, while SOFIA similarly employs tensor decomposition. The specific parameters used in the experimental setup are detailed in Table 1. In this experiment, the rank $R$ was selected via grid search to strike a balance between reconstruction accuracy and computational cost. The regularization parameter $\lambda$ was set to 0.1. The order of the SARIMA model $(p, d, q)$ was automatically selected based on the akaike information criterion. The algorithm employed a random non-negative initialization strategy, with a maximum of 300 iterations and a stopping threshold of $10^{-4}$. The seasonal cycle parameter $s$ is set to 12. This setting meets the requirements for real-time forecasting while avoiding the excessive complexity that would result from using very high-order seasonal parameters. Multi-step forecasting is performed using a recursive approach. The forecasting horizon is set to a series of consecutive time steps.

**Table 1:** Parameter value setting.

| Parameter Name | Value |
|---|---|
| Number of OD flows | 144 |
| Number of links $\|L\|$ | 54 |
| Dimension of the tensor T | $12 \times 12 \times 2016$ |
| Rank $R$ | 25 |
| Seasonal cycle $s$ | 12 |

### *5.2 Simulation Results and Analysis*

In this section, the performance of the proposed algorithm is tested and analyzed based on traffic data in tensor form. By evaluating various algorithms and metrics, we can better visualize the significant performance advantages of the proposed algorithm compared to other methods.

#### *5.2.1 Algorithm Effectiveness Verification*

There are 144 OD flows in the Abilene network topology. In this section, we randomly select four of these flows to analyze their estimation accuracy. Fig. 2 illustrates the prediction results for the 34th and 79th OD flows. As shown in Fig. 2a, in the 34th OD flow, a surge in traffic occurred around time 2250, followed by a slump around time 3750. The changes in the curve indicate that our proposed MTF algorithm effectively captures these two fluctuations, demonstrating high accuracy. As shown in Fig. 2b, our algorithm shows the best performance. The SRSVD algorithm is difficult to deal with high-dimensional nonlinear data and has the worst effect. The 79th OD flow displays a certain periodic pattern overall, but it also contains numerous peaks, with the flow size fluctuating violently and frequently. The MTF algorithm demonstrates strong robustness in the face of traffic fluctuations, indicating superior performance when confronted with complex traffic changes. Other algorithms do not accurately track traffic changes, and there is a large estimation error. Especially when the traffic changes suddenly, the error between the real value and the predicted value is more obvious.

OD flows 102 and 126 are analyzed next. In Fig. 3a, all algorithms show better prediction results. This improvement can be attributed to the more pronounced periodicity of the 102nd OD flow, which exhibits flatter flow fluctuations compared to the 79th flow. Conversely, Fig. 3b illustrates greater fluctuations in the flow. The 126th OD flow experiences a significant drop followed by a leveling off at the 3750th moment.

The MTF algorithm effectively detects the flow trend over time and displays minimal bias. In contrast, the SOFIA, TomoG, Inverse, and SRSVD algorithms exhibit a greater prediction bias. This discrepancy arises because the MTF algorithm is capable of mining multi-dimensional traffic patterns for prediction, resulting in superior outcomes.

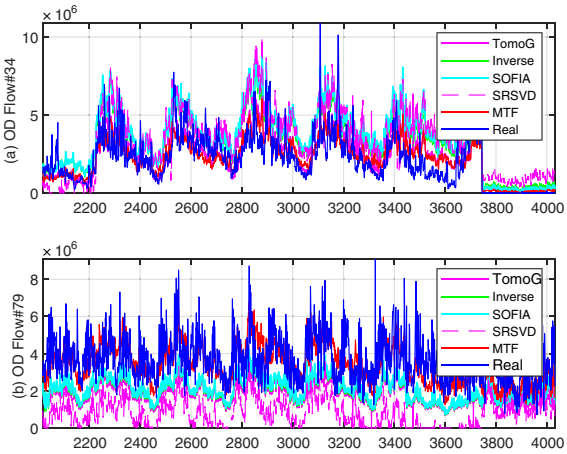


**Figure 2:** Estimation results about OD flows 34 and 79.

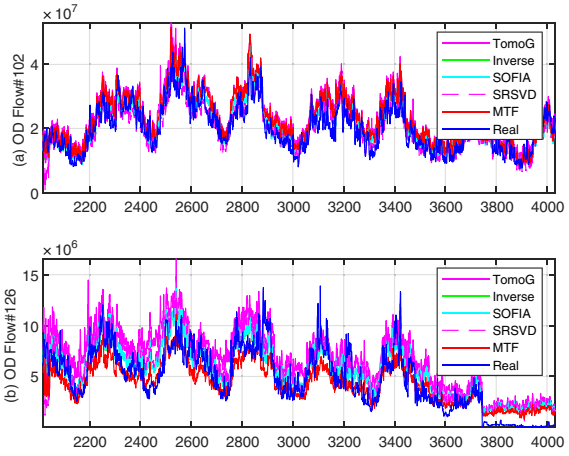


**Figure 3:** Estimation results about OD flows 102 and 126.

Fig. 4 exhibits a comparison of the root mean square error of different algorithms. The MTF algorithm demonstrates the lowest RMSE, which is 21.5%, 27.0%, 28.1%, and 37.9% lower than those of the SOFIA, TomoG, Inverse, and SRSVD algorithms, respectively. The MTF algorithm effectively preserves multidimensional features through pattern extraction. Additionally, it emphasizes temporal patterns to minimize noise interference, resulting in superior performance.

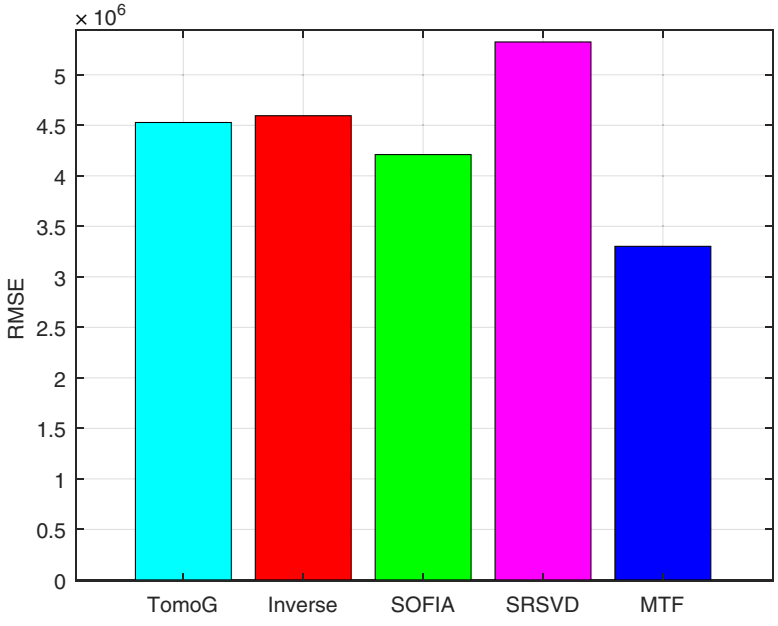


**Figure 4:** Comparison of RMSE.

### *5.2.2 Absolute Error Analysis*

This experiment compares the performance of different algorithms by observing the average absolute error.

Fig. 5 shows the temporal absolute errors of the five algorithms using a cloud and rain plot. The MTF algorithm exhibits the lowest median error and a low number of outliers. This expresses that the method is able to provide more accurate predictions in most cases, demonstrating good stability and robustness. In contrast, the SOFIA algorithm performs sub-optimally, while the Inverse and TomoG algorithms yield similar results. The SRSVD algorithm performs the least effectively. Fig. 6 displays the fluctuations in the average absolute error of the different algorithms at various time points. The MTF algorithm maintains a distinct advantage until time point 3750. Following this, due to a sudden change in traffic volume, the absolute time error of the MTF algorithm increases; however, it remains lower than that of the other algorithms. Compared to the other algorithms, the MTF algorithm has a relatively low average spatial absolute error value, with its curve consistently positioned at the bottom. This is attributed to the MTF algorithm's ability to rapidly reduce data dimensionality and extract key temporal factors while preserving essential spatial information. Based on this, the MTF algorithm is better equipped to capture the trends and periodicity of traffic fluctuations, effectively addressing the uncertainties in network traffic estimation.

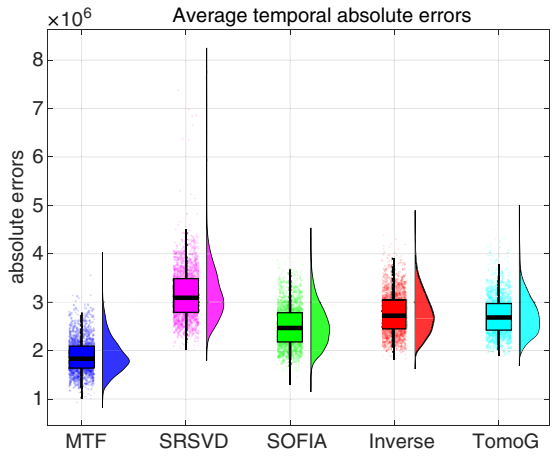


**Figure 5:** Comparison of average temporal absolute error distributions.

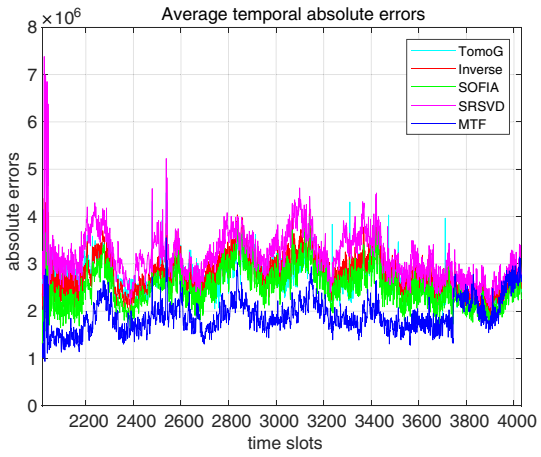


**Figure 6:** Comparison of average temporal absolute error curves.

Fig. 7 presents a comparison of the distribution of mean spatial absolute errors of the different algorithms. The MTF algorithm still shows the best performance, and the rest of the algorithms perform roughly similarly. The MTF algorithm is more concentrated within relatively low ranges and features the lowest upper limit of errors. This further substantiates the assertion that the MTF algorithm outperforms the other four algorithms. Fig. 8 shows the absolute error performance of the algorithms on different OD streams. The SRSVD method has relatively large error fluctuations, with several peaks significantly exceeding those of the other methods. This may be attributed to the fact that the SRSVD method primarily focuses on static data and is unable to effectively capture correlations within the data, resulting in an increase in its mean spatial absolute error. In contrast, the error fluctuations of the MTF algorithm are relatively small and consistently remain at the lowest values overall. Based on this analysis, it can be concluded that the MTF algorithm demonstrates superior performance in terms of estimation accuracy and stability.

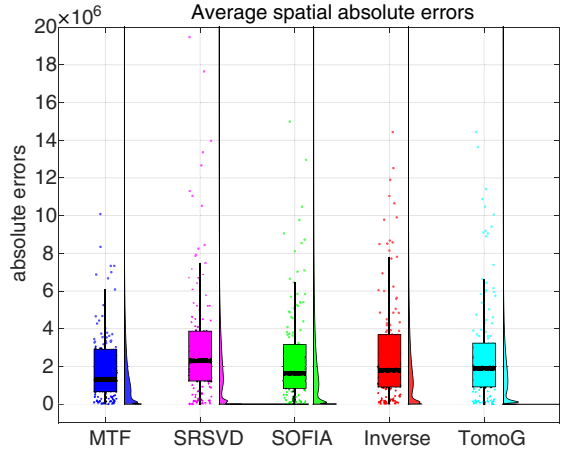


**Figure 7:** Comparison of average spatial absolute error curves.

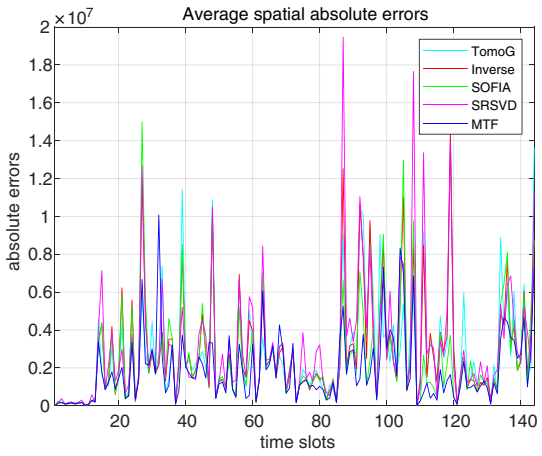


**Figure 8:** Comparison of average spatial absolute error curves.

*5.2.3 TRE and SRE Analysis*

The reliability of the overall prediction trend can be assessed by examining the relative error. Fig. 9 displays the spatial relative errors and the corresponding cumulative distribution functions (CDF) for various algorithms. The TMF algorithm demonstrates the lowest spatial relative error, with a particularly pronounced advantage in the first 40 OD flows. The CDF curves for the SRSVD method are significantly lower than those for the MTF algorithm, while the curves for the other three algorithms are approximately equivalent. When the SRE is 0.77, 90% of the OD flows can be estimated accurately. In contrast, the SOFIA, Inverse, TomoG, and SRSVD algorithms can only predict 86%, 82%, 81%, and 65% of the OD flow data, respectively, under the same conditions. This demonstrates the robust optimization capabilities and convergence mechanisms of the MTF algorithm throughout the overall prediction process. This characteristic allows the MTF method to estimate the network traffic matrix quickly and accurately in complex network environments, thereby providing substantial support for network management and optimization.

The TRE can illustrate the errors of all OD flows at a specific moment. Fig. 10 shows that, during a continuous extrapolation process spanning 4032 time steps, the TRE of the MTF algorithm remained at a low level with minimal fluctuation. No significant accumulation or divergence occurred as the prediction horizon increased. This stability stems from the effective locking of the network's global spatiotemporal structure achieved through tensor decomposition, enabling the model to remain focused on the intrinsic patterns of traffic flow even during long-range predictions.

To summarize, the MTF algorithm shows an excellent level of performance in predicting network traffic. By focusing on temporal patterns, it effectively reduces noise interference. The spatial patterns are preserved through tensor factorization, resulting in more accurate predictions. In both the temporal and spatial dimensions, the MTF method maintains lower error rates and a smoother trend, with a more concentrated distribution of error values. This superiority arises from the MTF algorithm's ability to handle the spatial, temporal, and periodic characteristics of the data, leading to enhanced prediction accuracy and stability. Furthermore, the MTF algorithm employs tensor factorization to reduce data dimensionality, which significantly improves prediction efficiency.

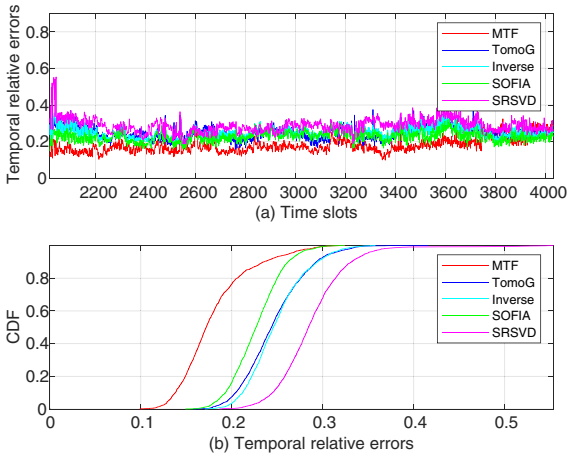


**Figure 9:** Comparison of spatial relative errors.

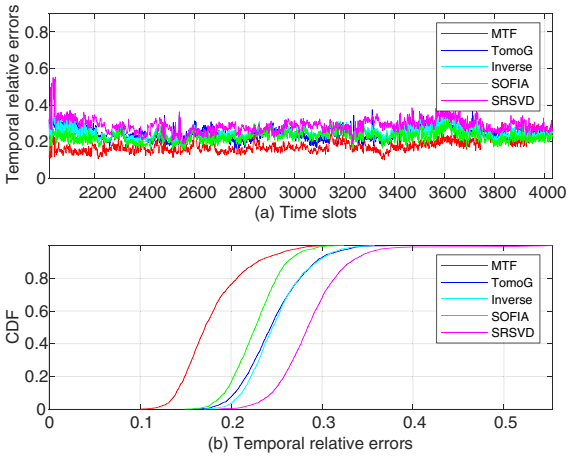


**Figure 10:** Comparison of temporal relative errors.

## 6 Conclusion

In this paper, we propose MTF, a tensor factorization-based traffic prediction algorithm. We integrate the tensor-based factor matrix with SARIMA to extract the multi-modal factors of the data and enhance the nonlinear processing capability. The proposed algorithm achieves efficient and accurate predictions through dimensionality reduction, pattern extraction, and seasonal forecasting. Our experimental results demonstrate the superiority of the MTF algorithm in terms of prediction error and its ability to capture abnormal fluctuations, with a 21.5% reduction in root mean square error compared to other algorithms. However, the algorithm does exhibit some dependence on its parameters.

Future research will focus on adaptability in dynamic topological environments. Additionally, comparing our approach with deep learning predictors in large-scale traffic forecasting tasks is an important direction for our next steps.

**Acknowledgement:** Not applicable.

**Funding Statement:** This work was supported in part by the National Natural Science Foundation of China (No. U20B2070), the Fund Projects (Nos. 315197107, JZDQZX202211-43-JS).

**Author Contributions:** The authors confirm contribution to the paper as follows: Conceptualization, Dingde Jiang and Yi Zhang; methodology, software, Xiaoxiong Yang; validation, Dingde Jiang; formal analysis, Xiaoxiong Yang; resources, Zhihan Lyu; data curation, Zhihan Lyu; writing—original draft preparation, Zhihan Lyu; writing—review and editing, Yi Zhang; visualization, Yi Zhang; supervision, Zhihan Lyu; project administration, Zhihan Lyu. All authors reviewed and approved the final version of the manuscript.

**Availability of Data and Materials:** Data openly available in a public repository.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhang Y, Zhang P, Jiang C, Wang S, Zhang H, Rong C. QoS aware virtual network embedding in space-air-ground-ocean integrated network. IEEE Trans Services Comput. 2024;17(4):1712–23. doi:10.1109/tsc.2024.3357707.

2. Le VA, Ji Y, Tran HH, Le Nguyen P, Lui JCS. Achieving multi-time-step segment routing via traffic prediction and compressive sensing techniques. IEEE Trans Netw Serv Manag. 2024;21(2):1534–49. doi:10.1109/tnsm.2023.3338622.
3. Bi J, Xu K, Yuan H, Zhang J, Zhou M. Network attack prediction with hybrid temporal convolutional network and bidirectional GRU. IEEE Internet Things J. 2024;11(7):12619–30. doi:10.1109/jiot.2023.3334912.
4. Zou G, Lai Z, Wang T, Liu Z, Li Y. MT-STNet: a novel multi-task spatiotemporal network for highway traffic flow prediction. IEEE Trans Intel Transp Syst. 2024;25(7):8221–36.
5. Xie R, Wen J, Chen X, Xie K, liang W, Xiong N, et al. $M^2$STL: multi-range multi-level spatial-temporal learning model for network traffic prediction. IEEE Trans Netw Sci Eng. 2024;11(5):4315–29.
6. Yang H, Yu W, Zhang G, Du L. Network-wide traffic flow dynamics prediction leveraging macroscopic traffic flow model and deep neural networks. IEEE Trans Intell Transp Syst. 2024;25(5):4443–57. doi:10.1109/tits.2023.3329489.
7. Gao Y, Yang LT, Yang J, Zheng D, Zhao Y. Jointly low-rank tensor completion for estimating missing spatiotemporal values in logistics systems. IEEE Trans Ind Inform. 2023;19(2):1814–22. doi:10.1109/tii.2022.3190549.
8. Thanh LT, Abed-Meraim K, Trung NL, Hafiane A. Robust tensor tracking with missing data and outliers: novel adaptive CP decomposition and convergence analysis. IEEE Trans Signal Process. 2022;70:4305–20.
9. Saragadam V, Balestriero R, Veeraraghavan A, Baraniuk RG. DeepTensor: low-rank tensor decomposition with deep network priors. IEEE Trans Pattern Anal Mach Intell. 2024;46(12):10337–48.
10. Chen X, He Z, Chen Y, Lu Y, Wang J. Missing traffic data imputation and pattern discovery with a Bayesian augmented tensor factorization model. Transp Res Part C Emerg Technol. 2019;104(7):66–77. doi:10.1016/j.trc.2019.03.003.
11. Li Q, Tan H, Wu Y, Ye L, Ding F. Traffic flow prediction with missing data imputed by tensor completion methods. IEEE Access. 2020;8:63188–201. doi:10.1109/access.2020.2984588.
12. Hu Y, Qu A, Wang Y, Work DB. Streaming data preprocessing via online tensor recovery for large environmental sensor networks. ACM Trans Knowl Discov Data (TKDD). 2022;16(6):122. doi:10.1145/3532189.
13. Hao F, Wang Z, Xu Y, Leng S, Fang L, Li F. Missing data completion for network traffic with continuous mutation based on tensor ring decomposition. In: 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD). Piscataway, NJ, USA: IEEE; 2024. p. 151–6.
14. Lee D, Shin K. Robust factorization of real-world tensor streams with patterns, missing values, and outliers. In: 2021 IEEE 37th International Conference on Data Engineering (ICDE). Piscataway, NJ, USA: IEEE; 2021. p. 840–51.
15. Wang S, Nie L, Li G, Wu Y, Ning Z. A multitask learning-based network traffic prediction approach for SDN-enabled industrial Internet of Things. IEEE Trans Ind Inf. 2022;18(11):7475–83.
16. Wang F, Xin X, Lei Z, Zhang Q, Yao H, Wang X, et al. Transformer-based spatio-temporal traffic prediction for access and metro networks. J Light Technol. 2024;42(15):5204–13. doi:10.1109/jlt.2024.3393709.
17. Wang P, Zhang Y, Hu T, Zhang T. Urban traffic flow prediction: a dynamic temporal graph network considering missing values. Int J Geogr Inf Sci. 2023;37(4):885–912. doi:10.1080/13658816.2022.2146120.
18. Zhao Y, Cheng S, Gao S, Wang P, Lu F. Predicting origin-destination flows by considering heterogeneous mobility patterns. Sustain Cities Soc. 2025;118(1):106015. doi:10.1016/j.scs.2024.106015.
19. Kumar A, Vidyapu S, Saradhi VV, Tamarapalli V. A multi-view subspace learning approach to internet traffic matrix estimation. IEEE Trans Netw Serv Manag. 2020;17(2):1282–93. doi:10.1109/tnsm.2020.2983329.
20. Zhou J, Han T, Xiao F, Gui G, Adebisi B, Gacanin H, et al. Multiscale network traffic prediction method based on deep echo-state network for Internet of Things. IEEE Internet Things J. 2022;9(21):21862–74.
21. Lv Z, Cheng Z, Li J, Xu Z, Yang Z. TreeCN: time series prediction with the tree convolutional network for traffic prediction. IEEE Trans Intell Transp Syst. 2024;25(5):3751–66.
22. Traffic datsets: Abilene, GEANT, TaxiBJ. IEEE Dataport. 2004. doi:10.21227/7x3c-5p06.
23. Singhal H, Michailidis G. Structural models for dual modality data with application to network tomography. IEEE Trans Inf Theory. 2011;57(8):5054–71. doi:10.1109/tit.2011.2158474.
24. Tan L, Wang X. A novel method to estimate IP traffic matrix. IEEE Commun Lett. 2007;11(11):907–9.
25. Roughan M, Zhang Y, Willinger W, Qiu L. Spatio-temporal compressive sensing and internet traffic matrices (extended version). IEEE/ACM Trans Netw. 2012;20(3):662–76. doi:10.1109/tnet.2011.2169424.