Satformer: Accurate and Robust Traffic Data Estimation for Satellite Networks

Anonymous Authors[1]

Abstract

The operations and maintenance of satellite networks heavily depend on traffic measurements. Due to the large-scale and highly dynamic nature of satellite networks, global measurement encounters significant challenges in terms of complexity and overhead. Estimating global network traffic data from partial traffic measurements is a promising solution. However, the majority of current estimation methods concentrate on low-rank linear decomposition, which is unable to accurately estimate. The reason lies in its inability to capture the intricate nonlinear spatio-temporal relationship found in large-scale, highly dynamic traffic data. This paper proposes Satformer, an accurate and robust method for estimating traffic data in satellite networks. In Satformer, we innovatively incorporate an adaptive sparse spatio-temporal attention mechanism. In the mechanism, more attention is paid to specific local regions of the input tensor to improve the model's sensitivity on details and patterns. This method enhances its capability to capture nonlinear spatio-temporal relationships. Experiments on small, medium, and large-scale satellite networks datasets demonstrate that Satformer outperforms mathematical and neural baseline methods notably. It provides substantial improvements in reducing errors and maintaining robustness, especially for larger networks. The approach shows promise for deployment in actual systems.

1. Introduction

As a potential complement to terrestrial networks, satellite networks are envisioned to provide broadband connectivity with seamless coverage and in a cost-effective manner. Internet service and content providers are interested in satellite networks due to their wide international coverage and lower entry costs in rural and underdeveloped areas (Zhang et al., 2023).

Traffic measurement plays a critical role to support fundamental network functions and upper-layer services in satellite networks, such as network monitoring, routing, intrusion detection, traffic engineering, and performance diagnosis (Akhlaghpasand & Shah-Mansouri, 2023). Timely and accurate traffic characteristics beyond basic metrics are undoubtedly beneficial for applications such as routing and traffic engineering.

However, it is tricky and costly to collect massive traffic data by measuring all transmission pairs directly (Xie et al., 2015), since traffic data is naturally distributed throughout the entire network. In order to support the traffic characteristics of the emerging mega-constellations, there is an urgent need to explore cost-effective traffic measurement methods. Traffic data estimation is a viable approach for large-scale satellite networks, where the global traffic data can be estimated accurately according to partial traffic sampling and measurement(Silva et al., 2014).

Since the inherent highly dynamic nature of spatio-distance and orbital positions in satellite networks, traffic volumes and patterns are time-varying. This poses significant challenges for both accurate and robust traffic estimation, including the need to maintain robustness in the face of varying sequence lengths (Ma et al., 2021). Hence, the primary challenge in estimating satellite networks traffic data lies on capturing the complex and nonlinear spatio-temporal correlations effectively.

Unfortunately, most efforts in traffic data estimation focus solely on low-rank linear decomposition, which cannot effectively capture the nonlinear spatio-temporal correlations in large-scale and dynamic traffic data, leading to inaccurate estimations. Therefore, developing a novel approach is crucial for enhancing traffic estimation performance to effectively extract and utilize the complex and nonlinear spatio-temporal correlations in inter-satellite traffic data.

For large-scale and highly dynamic satellite networks traffic data, we propose Satformer, a novel neural network architecture designed for accurate and robust traffic estimation. Systematically, Satformer constructs encoder-decoder components with stacked spatio-temporal modules to capture complex spatio-temporal correlations in traffic data

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

effectively. An adaptive sparse spatio-temporal attention mechanism (ASSIT) is adopted within each module, which can extract key features from a large number of sparse inputs. At the same time, we use a graph embedding module to effectively process non-Euclidean data through a Graph Convolutional Network (GCN).These components enhance the ability of Satformer to capture and exploit the nonlinear and complex information presented in traffic data. Additionally, the transfer module is incorporated to disseminate global context information throughout the model.

Our contributions are as follows:

- We design ASSIT, which adopts a multi-head self-attention structure. It can learn the correlation representation of traffic data at different spatio and temporal scales. We add a sparsity threshold to the attention matrix so that a large number of sparse inputs can be processed efficiently. By dynamically adjusting the threshold value, ASSIT adapts to the sparsity levels of various datasets, thereby enhancing the model's inference efficiency.

- To process non-Euclidean structured data, we introduce a graph embedding module within each module via GCN. Since the graph embedding module learns the relationship between nodes and neighbors adequately, it can extract the local and global information of nodes from non-Euclidean structured data. It improves the ability of the model to extract nonlinear spatio-temporal correlation.

- We add a transfer module to the Satformer framework, which can blend and reshape the traffic representation learned by the previous modules, conveying a global temporal and spatio perspective, while also helping to strengthen the generalization ability of the model on different types of datasets

This paper is structured as follows. Section 2 surveys relevant research. Section 3 explains our proposed Satformer methodology. Section 4 presents experimental verification and comparisons. Section 5 makes a discussion and section 6 concludes the paper.

2. Related Works

We provide a review of the existing work on network traffic estimation. Existing traffic data estimation methods can be mainly divided into matrix completion based, tensor completion based and neural network based methods.

Matrix completion (MC) methods have found widespread application in the estimation of traffic data. Some algorithms, such as the convex relaxation method based on minimum nuclear norm approximation (Cai et al., 2008) and matrix factorization-based methods (Koren et al., 2009), leverage the linear spatiotemporal characteristics of traffic data to infer missing values. Generally, these methods are concise, which may lead to inaccurate estimation on large-scale traffic data.

As an extension of matrix completion, the goal of tensor completion aims to reconstruct low-rank tensors based on sparse observations of their entries. Several studies have adopted tensor completion, including recent works (Xie et al., 2018b; Wang et al., 2020; Jiang et al., 2020).To achieve higher accuracy in traffic data estimation, these works propose the use of tensor completion methods, which can more comprehensively capture spatio-temporal features in traffic data, effectively. An typical work of such a method is LTC (Xie et al., 2018a), which leverages the strong local correlation of the data to identify and complete each subtensor with low rank. However, many traffic estimation algorithms based on tensor completion rely on CP or Tucker decompositions, commonly using inner products as interaction functions. This approach can often reduce estimation performance to some extent due to its limited ability to capture both linear and nonlinear correlations in traffic data.

In recent years, deep learning methods have shown notable advancements in traffic network analysis. Notably, research such as NTF (Wu et al., 2019) and (Ouyang et al., 2022) have explored the application of deep learning models, including Recurrent Neural Networks (RNNs), to achieve adaptive grouping and prediction of traffic tensors within large-scale networks. Noteworthy among these efforts is CoSTCo (Liu et al., 2019), which incorporates two convolutional layers to extract features from stacked embeddings, enhancing awareness of network dynamics through the acquisition of complex spatio-temporal features. Recent studies (Li et al., 2023; Cai et al., 2023) employ meta-learning and other algorithms,alongside attention mechanisms, to dynamically adapt to rapid changes in traffic patterns within the network. However, current deep learning models may focus more on global features, while neglecting the local and hidden spatio-temporal correlations in traffic data, which may lead to suboptimal estimation effects

3. Estimation Model: Satformer

3.1. System Model & Problem Definition

In satellite networks, inter-satellite traffic data can be modeled as a time-space matrix, which reflects the data volume to be transmitted between all node-node pairs over satellite networks. For problem statement of traffic estimation over satellite networks, we introduce the following symbols: N: Number of satellites, T: Discrete time steps, we define the inter-satellite traffic matrix $\mathcal{Y} \in R^{I \times J \times T}$, where $\mathcal{Y}_{ijt}$ represents the data transmission from satellite i to satellite

j at time step t. The t-th layer of this matrix represents a discrete time step.

Considering the influence of spatio distance and transmission delay in satellite networks, we can adjust the inter-satellite traffic by introducing a weight matrix. Let $W \in R^{N\times N}$ be the weight matrix representing spatio distance and transmission delay, where $\mathcal{W}_{ij}$ denotes the weight from satellite i to satellite j. Thus, the adjusted inter-satellite traffic data matrix can be represented as $\dot{\mathcal{X}} = \mathcal{Y} \odot \mathcal{W}$, where $\odot$ denotes element-wise multiplication.Taking into account these factors, the mathematical modeling of inter-satellite traffic data can be expressed as follows:

$$\dot{\mathcal{X}}_{ijt} = \mathcal{Y}_{ijt} \cdot \mathcal{W}_{ij} \tag{1}$$

where $i, j = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. This model considers the spatio distance and transmission delay between satellites, allowing the traffic data matrix to more accurately reflect the actual communication scenarios in the satellite networks. In the process of sampling and recovering inter-satellite traffic data, we begin by introducing the sampling matrix S, the sampled data $\mathcal{X}$, and the nonlinear estimation function F. The sampling process can be expressed using mathematical notation: $\mathcal{X} = \dot{\mathcal{X}} \odot S$.

This process retains elements in the inter-satellite traffic matrix $\dot{\mathcal{X}}$ where the corresponding positions in the sampling matrix S are 1, while setting other positions to zero, resulting in the sampled data matrix $\mathcal{X}$. To recover complete traffic data from the sampled data, we introduce a nonlinear estimation function F. This function involves a complex nonlinear mapping to better estimate actual traffic data.

$$\tilde{\mathcal{X}} = F(\mathcal{X}) \tag{2}$$

where $\mathcal{X}$ represents sampled data, and $\tilde{\mathcal{X}}$ is the recovered data obtained through the non-linear estimation function F.

3.2. Satformer Overview

We design Satformer, a tensor completion model designed for the accurate and robust estimation of global traffic data in satellite networks. Illustrated in Figure 1, Satformer is structured as an encoder-decoder architecture, with both components featuring multiple spatio-temporal modules. Residual connections interlink these modules to prevent neural network degradation. Each spatio-temporal module comprises a Graph Embedding layer and a Satformer block. The key of Satformer to improve the estimation accuracy is that it can extract features efficiently and accurately from a large number of sparse satellite networks traffic data. This is achieved through adaptive sparse spatio-temporal attention inside each Satformer block, facilitating the estimation of traffic data. A transfer module facilitates the seamless transmission of features from the encoder to the decoder. The encoder encodes the input traffic information, while the decoder is tasked with estimating the missing traffic data. The subsequent section provides a detailed description of each module.

3.3. Spatio-Temporal Module

Satformer utilizes spatio-temporal modules to extract spatio-temporal features from input tensors; this module primarily consists of graph embedding components and Satformer blocks.

Graph Embedding: The Spatio-Temporal module serves the goal of extracting spatio-temporal features from input tensor. Considering the inherent high sparsity of observed traffic data in real-world, it becomes imperative to represent tensors as low-dimensional vectors. Through the learning of embedded representations for nodes, the model inherently captures both structural and semantic information of nodes within the graph. This capability enables the model to comprehend relationships between nodes more effectively, facilitating the extraction of meaningful features from the $\mathcal{X} \in R^{I\times J\times T}$. Each OD pair corresponds to an origin node, a destination node, and the traffic of the OD-pair. To address the non-Euclidean nature of the data, particularly the spatio relationships within each OD pair, we employ Graph Embedding through Graph convolutional neural network (GCN). This approach allows the model to effectively handle non-Euclidean data, enhancing its capacity to capture and utilize the structural information present in the tensor $\mathcal{X} \in R^{I\times J\times T}$.

In Satformer, each Spatio-Temporal module contains a GCN model. A GCN model contains two layers of convolutional layer, the feature propagation rule can be stated as follows:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{3}$$

$$Z = f(X, A) = \sigma(\tilde{A}\mathrm{ReLU}(\tilde{A}XW^{(0)})W^{(1)}) \tag{4}$$

where, $H^{(l)}$ signifies the node embedding matrix for layer l, $\tilde{A} \in R^{I\times J}$ represents the adjacency matrix with self-connections. $\tilde{D} \in R^{I\times I}$ denotes the degree matrix, which is a diagonal matrix with each element on the diagonal representing the sum of the corresponding row in $\widehat{A}$. The weight matrix for layer l is denoted as $W^{(l)} \in R^{I\times M}$, and $H^{(l+1)} \in R^{I\times J\times M}$ represents the node embedding matrix for layer $l + 1$. $W^{(0)} \in R^{K\times L}$ denotes the weight matrix from the input layer to the hidden layer, and $W^{(1)} \in R^{K\times L}$ denotes the weight matrix from the hidden layer to the output layer. Both $\sigma(\cdot)$ and *ReLU* are activation functions employed in the model. $Z \in R^{I\times M\times K}$ represents the output embedding tensor.

Satformer Block: As show in Figure 1 right, in each Satformer block, we use a layer normalization at the begining to normalize the input embedding tensor. We then apply an

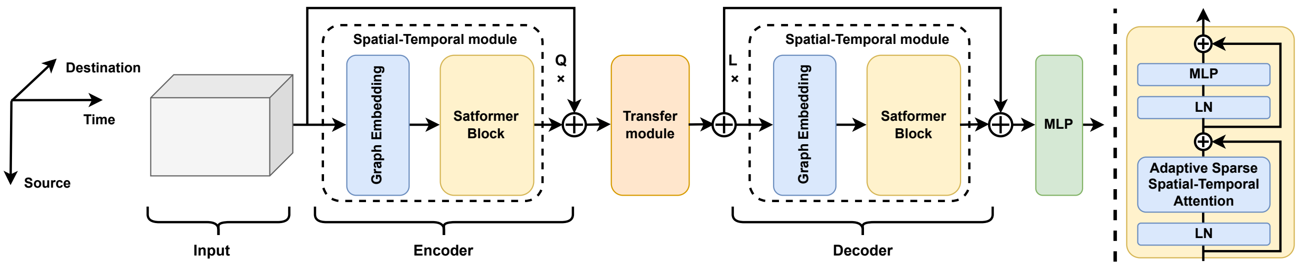


Figure 1. **Left:** The overall architecture of our Satformer. **Right:** Details of a Satformer block.

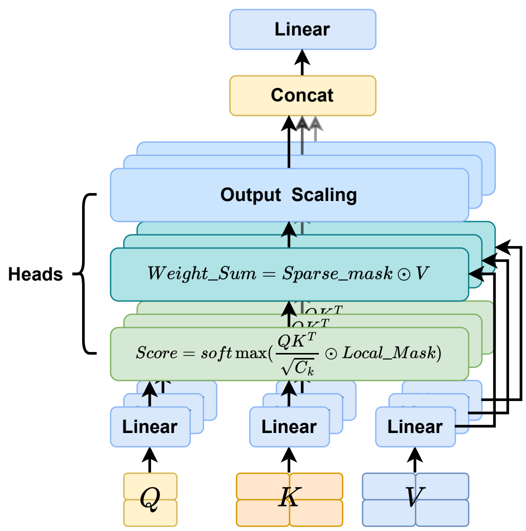


Figure 2. Adaptive Sparse spatio-Temporal Attention

ASSIT mechanism and a 2-layer MLP module for sparse spatio-temporal feature modeling and per-location embedding, respectively.

In the domain of communication network tensor completion, the spatio-temporal relationships among traffic data are complex, and it is necessary to model these relationships effectively. Traditional attention mechanisms, with their intensive nature, may encounter challenges related to high computational complexity and difficulties in capturing global relationships in such intricate scenarios. Several works proposed different sparse attention mechanism to mitigate such issue either relay on static patterns or skip computations in specific regions. As shown in Figure 2, in this work, we explore an adaptive, sparse spatio-temproal mechanism. A detailed descriptions are as follows:

Given an input embedding feature tensor $\mathcal{X} \in R^{I \times M \times K}$. First we divide the tensor slice into several local regions, each of which has a size of $D \times D$. In our module, Q (Query), K (Key), and V (Value) respectively represent query, key, and value, which are used to calculate attention weights and generate the final output(Vaswani et al., 2017). Then we calculate $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\boldsymbol{V}$ tensor with linear projections:

$$\boldsymbol{Q} = \mathcal{X}W^q \quad \boldsymbol{K} = \mathcal{X}W^k \quad \boldsymbol{V} = \mathcal{X}W^v \tag{5}$$

where W^q, W^k and W^v are projection weights for query $\boldsymbol{Q}$, key $\boldsymbol{K}$ and value $\boldsymbol{V}$ respectively.

We then consider introducing a local attention mechanism when calculating the attention score $\boldsymbol{\alpha}$ to make the model pay more attention to each local regions in the input tensor. This improvement is designed to sharpen the model's attention specifically on local regions within the input sequence. The goal is to augment the expressiveness and robustness of the model by enabling it to capture and leverage more nuanced details and patterns present in localized segments of the input data. The implementation involves incorporating a position-related weight when calculating attention scores. The local attention in each region is operationalized through the use of a two-dimensional mask matrix $\Psi \in R^{\boldsymbol{D} \times \boldsymbol{D}}$, wherein, elements inside a defined center window H are retained, while elements in other positions are set to zero. The size of the center window H is a hyperparameter of the model, and its optimal value is determined through experiments on different datasets. The calculation of attention score $\boldsymbol{\alpha}$ can be denoted as follows:

$$\boldsymbol{\alpha} = softmax(\frac{QK^T}{\sqrt{C_k}} \odot \Psi) \tag{6}$$

$$\boldsymbol{\alpha_s} = softmax(W^s \mathrm{ReLU}(1 - W^r \alpha) \odot \boldsymbol{V}) \tag{7}$$

where $\odot$ is an element-wise product and C_k is scaling factor.

To regulate the sparsity of the attention scores and channel the model's focus onto specific portions of the input, an adaptive sparse regularization term is introduced. This involves applying L1 regularization to each element of the attention score matrix. The utilization of ReLU operations ensures that the attention scores remain non-negative. Thus the sparse mask can be denoted as: $\mathrm{ReLU}(1 - W^r \alpha)$. Finally we apply the weighted Value to the attention score, while introducing additional learnable parameters to allow the model to adaptively learn the weighted sum of each position, resulting in the final output α_s, as shown in the Equation 7. Where, W^r is the weighted matrix of L1 regularization, W^s is the scaling matrix, both W^r and W^s are trainable parameters.

3.4. Transfer Module

The conventional information transfer between the encoder and decoder typically relies on the output of the last layer of encoder. However, this approach may fall short in adequately conveying global context information, particularly when dealing with input tensors spanning a large number of time slices. The accumulation of errors over time can become a challenge. Consequently, it is necessary to add a module between encoder and decoder to effectively transfer the information. Satformer incorporates a self-attention-based transfer module between the encoder and the decoder. This module leverages Self-Attention, enabling the seamless transfer of globally contextual information learned in the encoder to the decoder. This augmentation empowers the decoder to more comprehensively consider information from the entire input sequence when generating output for each time slice, thus enhancing the estimation accuracy of missing values. Moreover, the transfer module enables the model to integrate spatio-temporal information in a more fine-grained manner, improving its adaptability to patterns across different temporal and spatio scales.

3.5. Loss Function

During the training stage, the primary objective is to minimize the discrepancy between the actual and predicted traffic data. To achieve this, the loss function employed by Satformer is the mean square error (MSE), as expressed in Equation 8. Additionally, to curtail the growth of model weights and mitigate the risk of overfitting, a penalty term is incorporated into the loss function.

$$L(\theta) = \frac{1}{|\bar{\mathrm{A}}|} \sum\nolimits_{(i,j,t)\in\bar{\mathrm{A}}} (\chi_{ijt} - \tilde{\chi}_{ijt}) + \lambda \sum\nolimits_{i} (\theta_i) \quad (8)$$

where $\bar{\mathrm{A}}$ denotes the set of observed traffic data, χ_{ijk} and $\hat{\chi}_{ijk}$ are the truth and estimated traffic data respectively, θ represents all trainable parameters in Satformer, λ is weight decay coefficient.

4. Experiments

4.1. Experimental Settings

Datasets. To assess the performance of Satformer, we employ it on three real-world satellite networks: Iridium, Telesat, and Starlink, thereby evaluating its capabilities across varying network scales: small-scale, medium-scale, and large-scale environments. Given the ongoing construction and utilization of many satellite networks, acquiring actual traffic data proves to be challenging. Thus, we generate corresponding traffic datasets using real satellite parameters and ground station coordinates. Similar methods have been used in many previous studies, and the specific details of this process are explained in the Appendix A. The traffic data collection interval was 1 second for all three datasets.

- **Iridium (Kassas et al., 2023)**: The Iridium constellation comprising a total of 66 satellites uniformly distributed across 6 orbital planes. For our experimentation, we focus on the initial six periods, encompassing 36,000 time slices.
- **Telesat (Pachler et al., 2021)**: It collects traffic data from the Telesat constellation which has a total of 298 satellites distributed in 26 orbital planes. We select the first five periods about 31500 time slots in our experiment.
- **Starlink (Ma et al., 2023)**: The traffic data recording originates from the Starlink constellation, comprising 1584 Low Earth Orbit (LEO) satellites evenly dispersed across 72 orbital planes. The first six periods about 32400 time intervals in our experiment.

For all three datasets, we divided the original dataset into a training set and a test set in an 8:2 ratio using the time slice partitioning method. We then used the training set for model training and the validation set for model validation and tuning. Subsequently, we constructed the test set by randomly masking portions of the training and validation sets that were not used for training. This approach ensures that the model is trained and validated on distinct segments of the data, which can help prevent overfitting and improve the model's ability to generalize to new, unseen data.

Baselines. For comparative analysis against our Satformer model, we select the following baseline models: two mathematical tensor completion models, namely HaLRTC and LATC, and two state-of-the-art neural network-based tensor completion models, CoSTCo and DAIN.

- **HaLRTC (Liu et al., 2012)**: A prototypical high-accuracy low-rank tensor completion algorithm utilizes the Alternating Direction Method of Multipliers (ADMMs) to attain precise outcomes, effectively managing dependencies among various constraints.
- **LATC (Chen et al., 2022)**: It introduces a novel regularization term, integrating temporal variation, into a third-order tensor completion model.
- **CoSTCo (Liu et al., 2019)**: An innovative Convolutional Neural Network (CNN)-based model developed for tensor completion to overcome the limitations associated with traditional low-rank tensor factorization approaches.
- **DAIN (Oh et al., 2021)**: This method explicitly crafted to enhance the accuracy of neural tensor completion methods when predicting missing values within

sparse, multi-dimensional datasets.

- **SPIN** (**Marisca et al., 2022**): An attention-based architecture using spatiotemporal graphs and autoregressive models for effectively reconstructing missing data in sparse, multivariate time series.
- **STCAGCN** (**Nie et al., 2023**): A graph-based deep learning method for traffic volume estimation by utilizing a graph attention-based speed pattern-adaptive adjacency matrix and a customized temporal attention mechanism.

Evaluation Metrics. Two widely employed metrics are applied to evaluate the estimation performance of Satformer. The calculation equations for these metrics are presented as follows:

- **Normalized Mean Absolute Error (NMAE)**:

$$NMAE = \frac{\sum_{(i,j,t)\in\bar{\mathrm{A}}} |\chi_{ijt} - \tilde{\chi}_{ijt}|}{\sum_{(i,j,t)\in\bar{\mathrm{A}}} |\chi_{ijt}|} \tag{9}$$

- **Normalized Root Mean Squared Error (NRMSE)**:

$$NRMSE = \sqrt{\frac{\sum_{(i,j,t)\in\bar{\mathrm{A}}} |\chi_{ijt} - \tilde{\chi}_{ijt}|^2}{\sum_{(i,j,t)\in\bar{\mathrm{A}}} \chi_{ijt}^2}} \tag{10}$$

where χ_{ijk} and $\tilde{\chi}_{ijk}$ represent the truth value and estimated value, $\bar{\mathrm{A}}$ denotes the set of unobserved traffic data. For both two metrics, the smaller they get to 0, the better the estimation performance of the model.

4.2. Model Parameter Selection

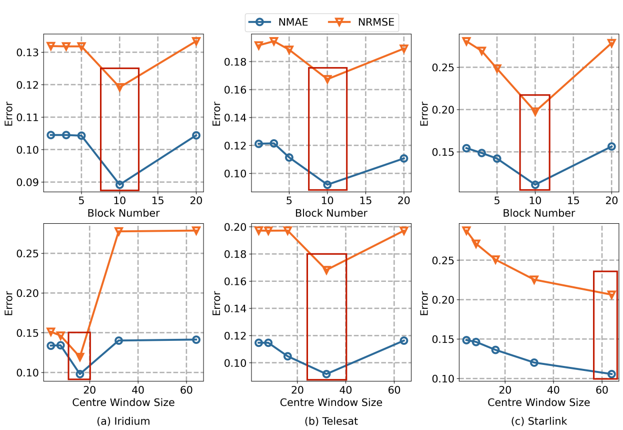


Figure 3. Analysis of hyper-parameters

Impact of module number. The module number indicates how many spatio-temporal modules should be contained in Satformer. It significantly impacts the computational efficiency and accuracy of Satformer. We adjust the number of modules from 1 to 20 and record the NMAE and NRMSE for each dataset, set the sample ratio to 2% and maintain the other hyperparameters constant. As shown in Fig.3, we observe that the error of Satformer continually decreases as the number of modules increases, up to a point (10 for all three datasets), after which it begins to increase. The reason is that increasing the number of modules enhances Satformer's ability to extract spatio-temporal features, while an excessive number of modules can lead to model overfitting and increased complexity. Consequently, we select a module number of 10 for all three datasets.

Impact of centre window size. The centre window size indicates how many elements should be retained in a mask matrix. It also significantly influences the accuracy of Satformer. It is another crucial hyperparameter of Satformer. The results in Figure 3 show that, for Iridium and Telesat, the estimation performance of Satformer continues to improve until it starts to decrease at a certain point (16 for Iridium and 32 for Telesat). Conversely, for Starlink, the estimation performance continues to increase. It can be observed that the window size is proportional to the dataset's scale. For small and medium-scale datasets, the model may extract useful information with a small center window size, thereby improving computational efficiency. However, for large-scale datasets, a larger center window size enables the model to extract more features while maintaining high accuracy. Accordingly, we set the center window size to 16 for Iridium, 32 for Telesat, and 64 for Starlink.

4.3. Performance Comparison with Baselines

Compare Satformer with mathematical baselines. Table 1 provides a summary of the experimental results for our Satformer and the mathematical tensor completion baselines, HaLRTC and LATC. Performance evaluations, measured by NMAE and NRMSE, are conducted across three datasets with sampling ratios ranging from 2% to 10%. Our Satformer consistently outperforms the mathematical tensor completion algorithms, achieving significant improvements. Notably, even at the minimal 2% sampling ratio, Satformer maintains proficient performance, with NMAE values recorded as 0.098, 0.1017, and 0.1402 for the Iridium, Telesat, and Starlink datasets, respectively. In comparison, the leading mathematical models exhibit higher NMAE values of 0.2782, 0.2723, and 0.3784 under the same 2% sampling ratio. The observed performance enhancement in Satformer quantifies at 1.83×, 1.67×, and 1.69× for the respective datasets. Similar trends are also observed in NRMSE. These results indicate that mathematical models based on Alternating Direction Method of Multipliers or reliant on strong assumptions struggle to capture the complex spatio-temporal characteristics. In contrast, neural network-based models such as Satformer demonstrate formidable nonlinear representation capabilities, enabling effective extraction of spatio-temporal features from traffic data.

Table 1. Estimation Performance of Satformer Compared with Baselines

Models	NMAE on Iridium					NRMSE on Iridium				
	2%	4%	6%	8%	10%	2%	4%	6%	8%	10%
HaLRTC	0.2782	0.2252	0.2044	0.1935	0.1886	0.3926	0.3381	0.3074	0.2888	0.2778
LATC	0.581	0.5809	0.5809	0.5809	0.5808	0.6009	0.5998	0.5997	0.5997	0.5996
Improve%	183.88%	131.45%	114.25%	115.24%	131.98%	228.54%	184.84%	165.46%	161.59%	175.32%
CoSTCo	0.1629	0.1623	0.16	0.1588	0.1435	0.5664	0.5644	0.5646	0.5621	0.5574
DAIN	0.1159	0.1156	0.1150	0.1144	0.1126	0.1435	0.142	0.1391	0.1377	0.127
SPIN	0.1206	0.1185	0.1175	0.1170	0.1158	0.1302	0.1310	0.1291	0.1229	0.1181
STCAGCN	0.1064	0.1059	0.1058	0.1049	0.1046	0.1847	0.1622	0.1523	0.1435	0.1203
Satformer	**0.098**	**0.0973**	**0.0956**	**0.0899**	**0.0813**	**0.1195**	**0.1187**	**0.1158**	**0.1104**	**0.1009**
Improve%	8.57%	8.84%	10.67%	16.69%	28.67%	8.95%	10.36%	11.49%	11.32%	17.05%
Models	NMAE on Telesat					NRMSE on Telesat				
	2%	4%	6%	8%	10%	2%	4%	6%	8%	10%
HaLRTC	0.2723	0.2723	0.259	0.2538	0.2267	0.5518	0.4402	0.421	0.3968	0.3632
LATC	0.6193	0.6181	0.6129	0.6031	0.6002	0.6367	0.6367	0.6367	0.6367	0.6368
Improve%	167.75%	172.3%	166.74%	157.66%	150.22%	196.35%	140.81%	138.12%	127.26%	118.53%
CoSTCo	0.2256	0.2182	0.2013	0.1898	0.1864	0.6996	0.6716	0.6482	0.6033	0.5852
DAIN	0.1387	0.1345	0.1328	0.1297	0.1211	0.2687	0.2679	0.2538	0.2499	0.2476
SPIN	0.1298	0.1286	0.1278	0.1274	0.1273	0.2378	0.2365	0.2353	0.2347	0.2344
STCAGCN	0.1488	0.1474	0.1457	0.1412	0.1393	0.2198	0.2184	0.2173	0.2184	0.2170
Satformer	**0.1017**	**0.1**	**0.0971**	**0.0985**	**0.0906**	**0.1862**	**0.1828**	**0.1768**	**0.1746**	**0.1662**
Improve%	27.63%	28.6%	31.62%	29.34%	33.66%	18.05%	19.47%	22.91%	25.09%	30.57%
Models	NMAE on Starlink					NRMSE on Starlink				
	2%	4%	6%	8%	10%	2%	4%	6%	8%	10%
HaLRTC	0.3784	0.3392	0.3116	0.282	0.2558	0.6148	0.4796	0.4398	0.4116	0.3778
LATC	0.5738	0.5733	0.5737	0.5437	0.5348	0.5984	0.5982	0.5937	0.5938	0.5928
Improve%	169.9%	18.81%	20.29%	27.25%	38.50%	20.08%	19.63%	20.12%	24.73%	25.87%
CoSTCo	0.2553	0.2479	0.2466	0.2462	0.2428	0.6635	0.6548	0.6531	0.6519	0.6498
DAIN	0.237	0.2231	0.2346	0.2233	0.2172	0.431	0.4036	0.4114	0.4189	0.4186
SPIN	0.2398	0.2353	0.2353	0.2352	0.2216	0.3989	0.3961	0.3959	0.3942	0.3919
STCAGCN	0.1944	0.1891	0.1802	0.1754	0.1685	0.3644	0.3611	0.3653	0.3644	0.3625
Satformer	**0.1402**	**0.1376**	**0.1349**	**0.1316**	**0.1223**	**0.2754**	**0.2722**	**0.2656**	**0.2645**	**0.2607**
Improve%	38.66%	37.43%	33.58%	33.28%	37.78%	32.32%	32.66%	37.54%	37.77%	39.05%

Compare Satformer with neural network-based baselines. Our Satformer outperforms the neural network-based baselines (CoSTCo, DAIN, SPIN, and STCAGCN) across all datasets, achieving the best estimation performance, as shown in Table 1. Notably, even with a 2% traffic data sampling rate, Satformer demonstrates significant improvements compared to the best-performing neural network-based baselines. On the Iridium dataset (66 satellites), Satformer improves NMAE and NRMSE by 8.57% and 8.95%, respectively. As the size of the dataset increases, performance improvements continue and escalate. On the Telesat dataset (298 satellites), Satformer achieves improvements of 27.63% in NMAE and 18.05% in NRMSE. For the Starlink dataset (1584 satellites), Satformer exhibits even more substantial improvements, with NMAE and NRMSE increasing by 38.66% and 32.32%, respectively. These results highlight Satformer's effectiveness in handling large-scale datasets, suggesting potential deployment in real-world satellite networks. The limitations of CoSTCo are evident due to its exclusive reliance on two-dimensional convolution for spatial feature extraction without explicitly modeling temporal features. DAIN falls short by not explicitly modeling interactions between entities, which limits information utilization, despite its combination of information for data augmentation. SPIN's ability to handle sparsity or irregularly sampled data might be limited, which could affect the accuracy of traffic estimation in satellite networks where data is often incomplete. STCAGCN captures time-asynchronous correlations may not fully account for the complex temporal dynamics in satellite network, leading to less accurate estimations. The architecture of STCAGCN cannot ensure the information learned at earlier stages is preserved and utilized in later stages. In contrast, Satformer excels by explicitly incorporating both spatial and temporal features within each module. The graph embedding captures nonlinear information, the Satformer module integrates the ASSIT, and the transfer module seamlessly transmits global contextual information. This comprehensive design enables Satformer to deliver exceptional performance in inter-satellite traffic data estimation, effectively addressing the challenges of large-scale, sparsely populated datasets.

Table 2 presents a comparison of the training and inference times of Satformer with various baseline models. Although SPIN incorporates a sparse attention mechanism, it can be computationally intensive. This may lead to longer processing times. Notably, Satformer demonstrates the fastest training and inference times, which makes it particularly well-suited for real-world deployment. The significant reduction in these times is primarily attributed to the adaptive spatio-temporal attention mechanism, which introduces strategic sparsity in the sampling of input tensors and markedly reduces the number of parameters, offering a substantial time-saving advantage.

Table 2. Training & Inference Time

Dataset	Model	Training (s)	Inference (s)
Iridium	HaLRTC	/	123.9s
	LATC	/	209.2s
	CoSTCo	164.6s	0.200s
	DAIN	255.7s	1.164s
	SPIN	169.43s	0.794s
	STCAGCN	330.9s	0.422s
	Satformer	**80.3s**	**0.082s**
Telesat	HaLRTC	/	520.6s
	LATC	/	748.3s
	CoSTCo	570.4s	0.274s
	DAIN	767.2s	3.722s
	SPIN	923.86s	2.858s
	STCAGCN	824.6	3.872s
	Satformer	**168.9s**	**0.194s**
Starlink	HaLRTC	/	2673.2s
	LATC	/	3350.1s
	CoSTCo	1582.3s	**0.314s**
	DAIN	3254.8s	13.475s
	SPIN	2566.5s	3.291s
	STCAGCN	3993.8s	3.8795s
	Satformer	**879.5s**	0.477s

4.4. Robustness Analysis

To thoroughly evaluate the robustness and performance of Satformer, we conducted a comprehensive comparison of its Normalized Absolute Error (NMAE) and Normalized Root Mean Squared Error (NRMSE) metrics against all six baseline models across three diverse datasets. These datasets encompass varying numbers of time slices, ranging from 100 to 1500, providing a broad spectrum of temporal scales for analysis. The results from our experiments unequivocally demonstrate the superior and consistent performance of Satformer across all datasets, irrespective of the number of time slices involved. Notably, as the number of input time slices increases, Satformer consistently outperforms other models in terms of reliability, maintaining consistently low NMAE and NRMSE indicators. Importantly, even with the escalation of the temporal dimension and the expansion of dataset sizes, Satformer exhibits remarkable stability in its results. These findings underscore the robustness and scalability of Satformer, rendering it not only valuable in theoretical contexts but also highly applicable in real-world engineering scenarios.

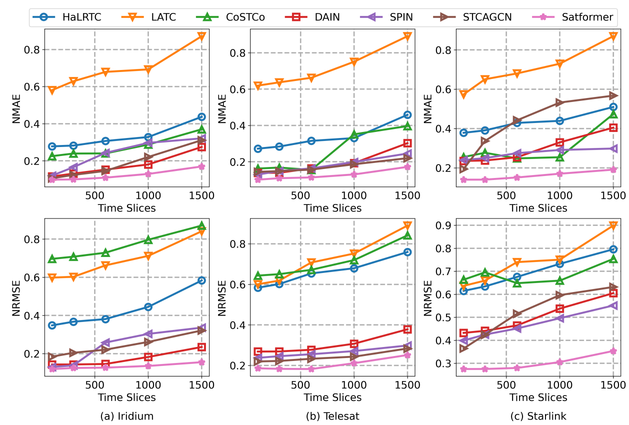


Figure 4. Analysis of robustness

4.5. Ablation Study

To assess the impact of our graph embedding module (GE), ASSIT module, and transfer module on Satformer's performance, we conducted ablation experiments, systematically removing these key components to create various model variants. Specifically, we evaluated the following variants: 1) Satformer-GE omits the graph embedding module; 2) Satformer-AT removes the adaptive sparse spatio-temporal attention mechanism module; 3) Satformer-TM does not incorporate the transfer module.

Experimental results consistently demonstrate the superior performance of the complete model as compared to the variant models from which crucial components have been removed, across three datasets that have varying sampling rates. Notably, the absence of the graph embedding module hinders the effective capture of topological structure information, underscoring its pivotal role in augmenting overall model performance. The significance of the ASSIT module becomes evident in addressing sparsity issues within spatio-temporal data; when it is absent, the model experiences a noticeable decline in performance, validating its effectiveness in uncovering temporal and spatial dependencies. Furthermore, the transfer module plays a crucial role in facilitating the conversion of information between different feature spaces, thereby enhancing the model's capability in feature representation.

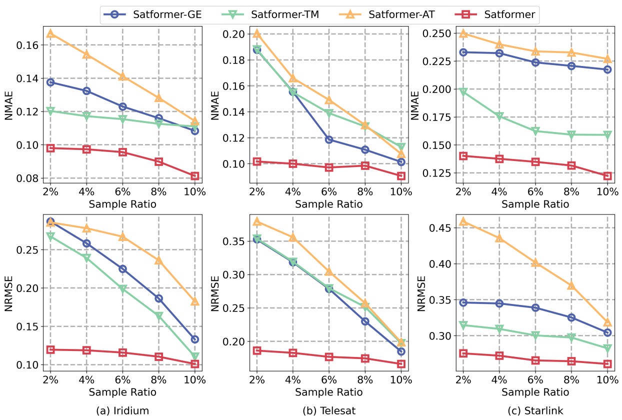


Figure 5. Ablation Study on Satformer

5. Conclusion and Discussion

This paper proposes Satformer, a novel traffic data estimation algorithm for large-scale satellite networks, aiming at fast and accurate estimating global traffic matrix from partial sampling in a cost-effective manner. Motivated by this, we design a region-aware sparse spatio-temporal attention mechanism to concentrates on specific local regions of the input tensor, where the input tensor is embedded in a graph convolutional neural network. Thus, spatio-temporal features from traffic matrix are effectively extracted with computational efficiency and robustness.

Extensive experiments with datasets of varying scales-small, medium, and large have shown that Satformer have significant advantages on both accuracy and efficiency for traffic estimation compared with baselines, particularly in larger networks. Moreover, we analyze the robustness of Satformer under different conditions and further verify the role of each module through ablation studies. The results demonstrate the potential of Satformer for deployment in actual systems.

Despite Satformer is effective adopted for traffic estimation, deep learning models for traffic estimation remain mostly black boxes. It is quite important to understand the reasons behind inferences in the satellite networking domain. In addition, although Satformer is cost-effective, it is necessary to further reduce its computational complexity, considering the limited computational resources of existing satellites.

Future works should prioritize enhancing computational efficiency. It is also important to explore interpretability and decision basis of our deep learning model for traffic estimation. For example, explanation techniques, such as a local interpretable model-agnostic explanation (LIME) (Bhattacharya, 2022), are able to make a visual analysis of model and analyze the internal working mechanism from specific examples. Additional explanatory tools, such as feature importance analysis, will help users in understanding the model's workings.

Impact statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Akhlaghpasand, H. and Shah-Mansouri, V. Traffic offloading probability for integrated leo satellite-terrestrial networks. Jul 2023.

Bhattacharya, A. *Applied Machine Learning Explainability Techniques: Make ML models explainable and trustworthy for practical applications using LIME, SHAP, and more*. Packt Publishing Ltd, 2022.

Cai, J., Song, S., Zhang, H., Song, R., Zhang, B., and Zheng, X. Satellite network traffic prediction based on lstm and gan. In *2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, volume 3, pp. 175–178. IEEE, 2023.

Cai, J.-F., Candès, E., and Shen, Z. A singular value thresholding algorithm for matrix completion. *arXiv: Optimization and Control,arXiv: Optimization and Control*, Oct 2008.

Chen, X., Lei, M., Saunier, N., and Sun, L. Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):12301–12310, 2022.

Jiang, D., Wang, W., Shi, L., and Song, H. A compressive sensing-based approach to end-to-end network traffic reconstruction. *IEEE Transactions on Network Science and Engineering*, 7(1):507–519, 2020.

Kassas, Z. M., Kozhaya, S., Kanj, H., Saroufim, J., Hayek, S. W., Neinavaie, M., Khairallah, N., and Khalife, J. Navigation with multi-constellation leo satellite signals of opportunity: Starlink, oneweb, orbcomm, and iridium. In *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 338–343. IEEE, 2023.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Koren, Y., Bell, R., and Volinsky, C. Matrix factorization techniques for recommender systems. *Computer*, pp. 30–37, Aug 2009.

Li, Y., Liang, W., Xie, K., Zhang, D., Xie, S., and Li, K. Lightnestle: quick and accurate neural sequential tensor completion via meta learning. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*, pp. 1–10. IEEE, 2023.

Liu, H., Li, Y., Tsang, M., and Liu, Y. Costco: A neural tensor completion model for sparse tensors. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 324–334, 2019.

Liu, J., Musialski, P., Wonka, P., and Ye, J. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–220, 2012.

Ma, S., Chou, Y. C., Zhao, H., Chen, L., Ma, X., and Liu, J. Network characteristics of leo satellite constellations: A starlink-based measurement from end users. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*, pp. 1–10. IEEE, 2023.

Ma, Z., Zhao, Y., Wang, W., and Zhang, J. Demonstration of highly dynamic satellite optical networks supporting rapid reconfiguration. In *2021 17th International Conference on the Design of Reliable Communication Networks (DRCN)*, Apr 2021.

Marisca, I., Cini, A., and Alippi, C. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. *Advances in Neural Information Processing Systems*, 35:32069–32082, 2022.

Nie, T., Qin, G., Wang, Y., and Sun, J. Towards better traffic volume estimation: Jointly addressing the underdetermination and nonequilibrium problems with correlation-adaptive gnns. *Transportation Research Part C: Emerging Technologies*, 157:104402, 2023.

Oh, S., Kim, S., Rossi, R. A., and Kumar, S. Influence-guided data augmentation for neural tensor completion. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 1386–1395, 2021.

Ouyang, Y., Xie, K., Wang, X., Wen, J., and Zhang, G. Lightweight trilinear pooling based tensor completion for network traffic monitoring. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pp. 2128–2137. IEEE, 2022.

Pachler, N., del Portillo, I., Crawley, E. F., and Cameron, B. G. An updated comparison of four low earth orbit satellite constellation systems to provide global broadband. In *2021 IEEE international conference on communications workshops (ICC workshops)*, pp. 1–7. IEEE, 2021.

Silva, J. M. C., Carvalho, P., and Lima, S. R. Computational weight of network traffic sampling techniques. In *2014 IEEE Symposium on Computers and Communications (ISCC)*, Jun 2014.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., and Polosukhin, I. Attention is all you need. *Neural Information Processing Systems,Neural Information Processing Systems*, Jun 2017.

Wang, Q., Chen, L., Wang, Q., Zhu, H., and Wang, X. Anomaly-aware network traffic estimation via outlier-robust tensor completion. *IEEE Transactions on Network and Service Management*, 17(4):2677–2689, 2020.

Wu, X., Shi, B., Dong, Y., Huang, C., and Chawla, N. V. Neural tensor factorization for temporal interaction learning. In *Proceedings of the Twelfth ACM international conference on web search and data mining*, pp. 537–545, 2019.

Xie, K., Wang, L., Wang, X., Xie, G., Zhang, G., Xie, D., and Wen, J. Sequential and adaptive sampling for matrix completion in network monitoring systems. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, Apr 2015.

Xie, K., Peng, C., Wang, X., Xie, G., Wen, J., Cao, J., Zhang, D., and Qin, Z. Accurate recovery of internet traffic data under variable rate measurements. *IEEE/ACM transactions on networking*, 26(3):1137–1150, 2018a.

Xie, K., Wang, L., Wang, X., Xie, G., Wen, J., Zhang, G., Cao, J., and Zhang, D. Accurate recovery of internet traffic data: A sequential tensor completion approach. *IEEE/ACM Transactions on Networking*, 26(2):793–806, 2018b.

Zhang, Y., Gong, Y., Fan, L., Wang, Y., Han, Z., and Guo, Y. Quantum-assisted joint caching and power allocation for integrated satellite-terrestrial networks. Dec 2023.

A. Traffic Generator

The characteristics of traffic load borne by satellite networks are intricate. Satellites predominantly communicate with terrestrial terminals via ground stations. Spatially, the distribution of ground stations is uneven due to factors such as topography, economic considerations, and geopolitical influences. In regions with extreme environmental conditions or economic underdevelopment, such as oceans, deserts, and polar areas, the received traffic is significantly lower compared to more favorable environments, contributing to an uneven spatial distribution of traffic in the satellite network. Furthermore, the global distribution of earth stations spans various time zones, resulting in non-stationary traffic generation at different times. This temporal variability leads to significant traffic variations among stations. Additionally, to ensure link quality between satellites and ground stations and to mitigate the impact of frequent satellite handoffs, ground stations must consider multiple factors, including elevation angle, service time, and signal strength, when selecting communication satellites. This complexity adds to the challenges associated with managing traffic in satellite networks.

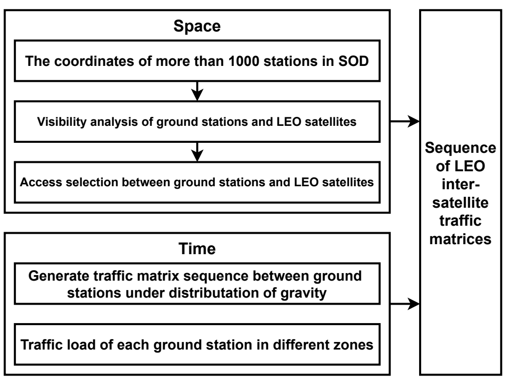


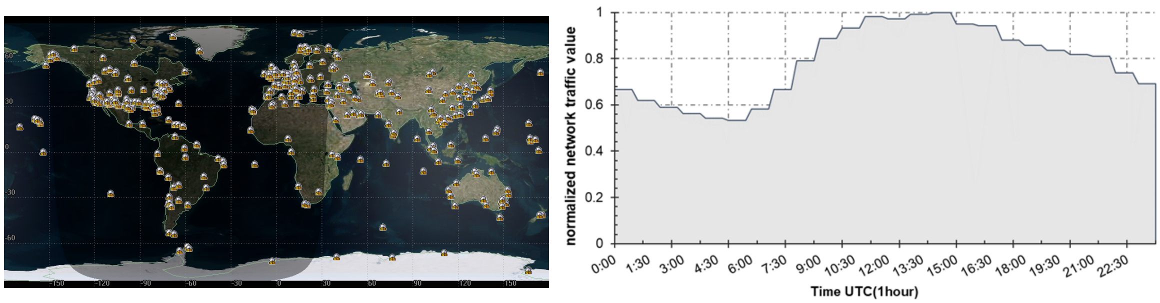
Figure 6. Traffic generation framework.

The pivotal aspect in assessing the effectiveness of the proposed scheme is constructing a coherent satellite network traffic model that accurately represents the traffic characteristics of the satellite network. The devised traffic generation method, as illustrated in the accompanying figure, takes into account the spatial distribution of ground stations, temporal characteristics, and satellite-ground access. Although this approach primarily focuses on inter-satellite communication and excludes satellite-ground communication, it offers comprehensive consideration of three key factors: spatial distribution of ground stations, temporal dynamics, and satellite-ground access. The culmination of these considerations results in the generation of a sequence of inter-satellite traffic matrices.

spatio distribution of ground stations. Due to limitations imposed by antenna size, equipment volume, and quality, current user communications with satellites predominantly occur through ground stations. The spatial distribution of traffic load in a satellite network is intricately linked to the global geographical distribution of earth stations. In this work, ground stations are strategically positioned based on the coordinates provided by the Standard Object Database (SOD) within the Satellite Tool Kit (STK). The SOD database contains geographical location information for 1,016 Earth stations worldwide, primarily situated in islands, mountainous areas, rural locales, and other remote regions effectively served by satellites. In contrast to alternative assumptions regarding user distribution, leveraging the SOD ensures a more accurate representation of user distribution and density, which is crucial for this work. Consequently, this facilitates an effective depiction of the spatial distribution of traffic load within the satellite network.

temporal dynamics. The temporal variations in traffic load within non-geostationary orbit satellite networks predominantly arise from two factors: the diurnal fluctuations induced by regional local times and the geographic variations in daily traffic patterns influenced by global time zones. In the foreseeable future, the satellite optical network is expected to handle a traffic load comparable to that of the ground network, either matching, proportionally scaling, or exhibiting similar patterns. To ensure accuracy and effectiveness in generating traffic scenarios, this paper employs the four-month average daily traffic change trend from the GEANT network to characterize daily traffic variations. Fig.7b depicts the normalized cumulative traffic load over a 24-hour period, with the highest peak value normalized to 1. Notably, the flow intensity peaks around 12 noon, gradually diminishes, and then experiences a subsequent rise around 5 AM the following day. In addressing geographical variations, for the sake of model simplicity, the local time of a ground station within its respective time zone is incremented by one hour for every 15 degrees of longitude eastward from Greenwich Mean Time. This adjustment is contingent on the specific time zone associated with each ground station.

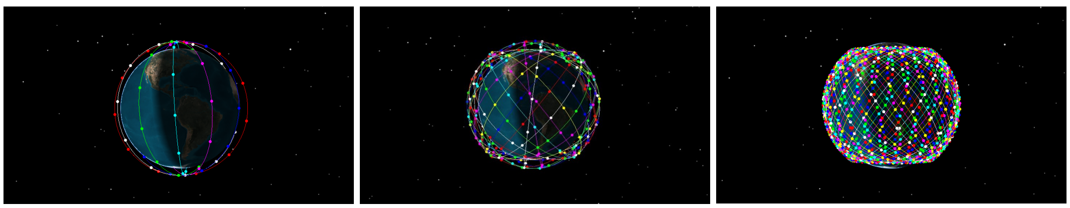
satellite-ground access. The capability of a ground station to establish satellite-ground communication links with multiple satellites within a given time window is influenced by factors such as satellite density and the coverage area of an individual satellite. Different satellite-ground access methods introduce varying effects on the characteristics of traffic load. Utilizing visibility analysis outcomes obtained through the Satellite Tool Kit (STK) and considering conditions for establishing satellite-ground links, the service time offered by all satellites visible to a ground station in each time window is computed. The satellite offering the longest service time is then selected for access, allowing for the flexibility to choose alternative or custom-designed satellite-ground access methods as needed. The ground



(a)

(b)

Figure 7. (a) Global distribution of satellite ground stations. (b) Normalized one-day traffic variation for ground station



(a) (b) (c)

Figure 8. (a)Iridium constellation. (b)Telesat constellation. (c)Starlink constellation

station selects the destination node, following a uniform distribution. The traffic density at the ground station is jointly determined by the local time within the time zone and the spatial distribution of ground stations across each time zone.

Assuming the network time is based on GMT (Greenwich Mean Time) and considering the generation interval $\Delta t(s)$ of the traffic matrix, the computation process for the traffic matrix sequence $\{F_{ter}^t | t \in N^*\}$ between ground stations is as follows:

For any time t, the total traffic D_t sent by all ground stations in the whole network is calculated as Eq.11.

$$D_t = offerload \times B \times n_{ter} \tag{11}$$

where, $offerload$ is the network traffic load, B is the maximum bandwidth of the inter-satellite link, and n_{ter} is the total number of ground stations.

The latitude and longitude coordinates of any ground station i is (x_i, y_i), then the local time $t_m(h)$ of the time zone m of any ground station i can be calculated by Eq.12:

$$t_m = \left\lfloor \frac{t}{3600} \right\rfloor + \left\lfloor \frac{x_i}{15} \right\rfloor \tag{12}$$

Combined with the normalized cumulative traffic load in 24 hours in Fig.7b, the traffic intensity weight of the time zone m where the ground station i is located can be calculated:

$$w_m^t = \frac{w_{t_m}}{w_{total}}, 0 \le m \le 23 \& m \in N^* \tag{13}$$

where, w_{total} is the total traffic load and w_{t_m} is the load at time t corresponding to time zone m.

The traffic $(f_m^t)_z$ that a ground station i located in time zone m needs to send at time t can be calculated by Eq.14 as follows.

$$\left(f_m^t\right)_i = \frac{D_t \times w_m^t}{n_m} \tag{14}$$

where, n_m denotes the number of ground stations in time zone m, and $\sum_{m=0}^{23} n_m = n_{ter}$.

Since the destination node is selected by the ground station according to uniform distribution, the traffic sent from the ground station i to the ground station j at time t can be calculated by Eq.15:

$$F(i,j)^t = U(0.1, 1) * \left(f_m^t\right)_i \tag{15}$$

where, $F(i,j)$ denotes the traffic from the source ground station i to the destination ground station j, and $U(0.1, 1)$ represents the uniform distribution from 0.1 to 1.

By traversing all the ground stations in the network at each time interval, the traffic matrix sequence $\{F_{ter}^t | t \in N^*\}$ between the ground stations can be calculated. According to the inter-satellite visibility analysis results provided by STK, as shown in Fig.8, combined with the satellite-to-ground access method described previously, the inter-satellite traffic matrix sequence $\{F_{sat}^t t \in N^*\}$ can be obtained.

Table 3. Hyper Parameter Settings

Dataset	Model	lr	epochs	batch size
Iridium	CoSTCo	0.001	100	64
	DAIN	0.0001	50	256
	SPIN	0.0008	50	32
	STCAGCN	0.0005	50	32
	Satformer	0.001	200	128
Telesat	CoSTCo	0.001	100	64
	DAIN	0.0001	100	256
	SPIN	0.0008	50	32
	STCAGCN	0.0005	100	32
	Satformer	0.001	200	128
Starlink	CoSTCo	0.001	100	64
	DAIN	0.0001	50	1024
	SPIN	0.0008	100	64
	STCAGCN	0.0005	150	32
	Satformer	0.001	300	128

B. Implemenation Details

Satformer and the neural network-based baselines are implemented in PyTorch, while the mathematical baselines are implemented using Numpy. We evaluated Satformer against the baselines on a server equipped with an NVIDIA RTX 2080Ti GPU, 128 GB DDR4 RAM, and an Intel Xeon Silver 4208 CPU, running the Ubuntu 18.04 operating system. All models are trained for a range of 50 to 300 epochs with the first 5 epochs designated for warmup, and early stopping is adopted during the training process. The Adam optimizer (Kingma & Ba, 2014) is used to optimize our model. A grid search strategy is applied to determine the best learning rate, epochs, and batch size. Based on the results of the grid search strategy, the optimal hyperparameters for the Satformer model and the neural network baselines are presented in Table 3, with the best weight decay determined to be 0.00001.

C. Virtual Attention

The adaptive sparse spatio-temporal attention mechanism allows Satformer to focus on specific local regions of the input tensor, which is particularly beneficial for handling the large-scale and highly dynamic nature of satellite networks. We visualize the attention mechanism in Satformer to verify whether the functionality is achieved, as shown in the Fig.9.

Fig.9(a) present the initial traffic matrices for Iridium, Telesat, and Starlink satellite constellations. Each matrix's dimension is determined by the number of satellite nodes, denoted as *N*, with each point representing the volume of traffic between respective node pairs. Fig.9(b) illustrate the traffic matrices after applying a 10% sample rate to the three datasets, demonstrating a clear reduction in data density. Fig.9(c) display the attention map, which are derived from the Satformer module using the attention scores, α_s, as defined by Eq.7. The attention map's dimensions are R ($D \times D$×heads), which is then averaged across the head dimension and scaled to $N \times N$ in the $D \times D$ dimensions for visualization purposes. This scaling process is designed to be intuitive without altering the inherent relationships within the tensors. Finally, Fig.9(d) represent the estimated traffic matrices, which are the reconstructed traffic data from the sampled data.

Upon examining the initial traffic matrix for Iridium constellation, a notable volume of traffic is evident within the red-boxed area. After sampling, only sparse data points remain, yet the attention map successfully captures the significance of this high-traffic area, as indicated by the high attention scores. The estimated traffic matrix aligns well with the actual traffic in this region. Similar observations can be made for the Telesat and Starlink datasets, where the attention mechanism effectively identifies and emphasizes critical traffic areas, leading to accurate estimations within the reconstructed traffic matrices.

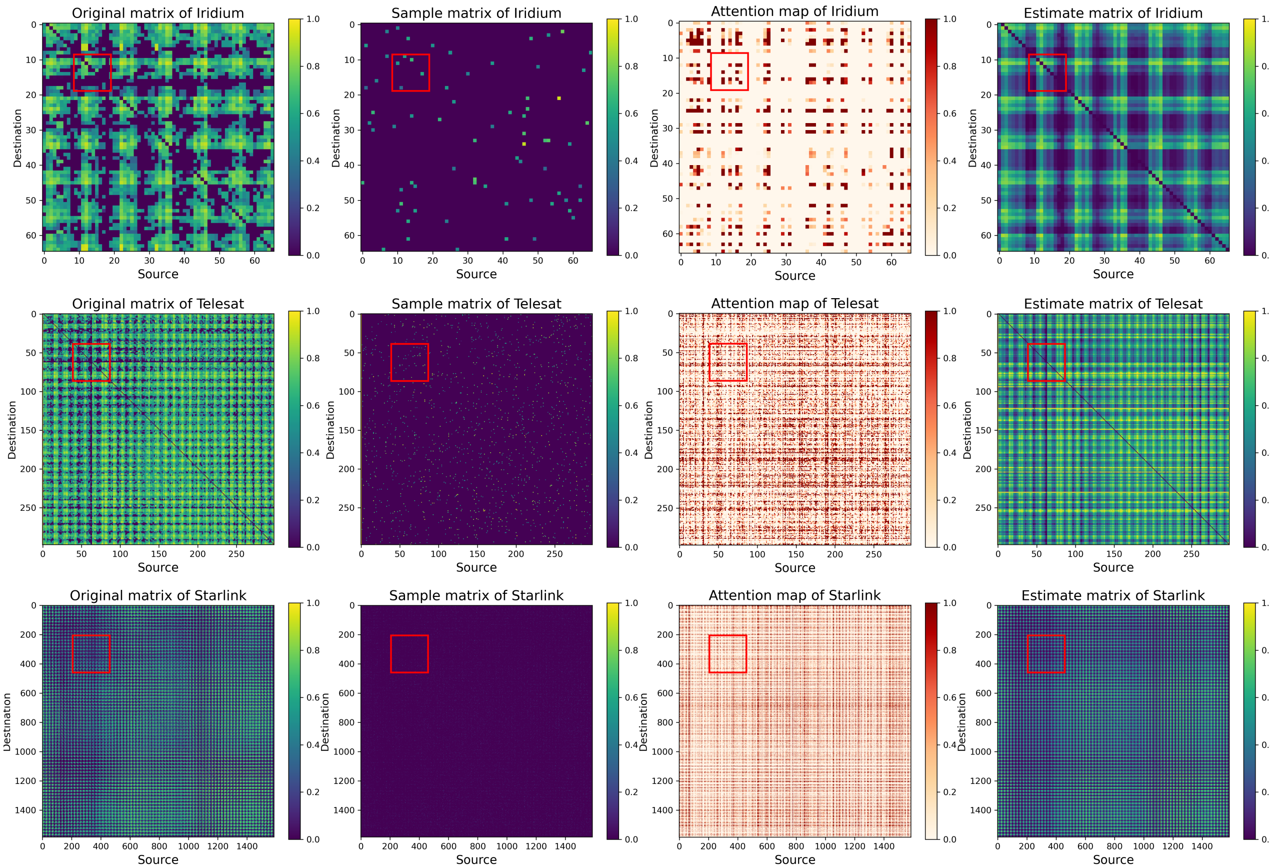
D. General Tensor Completation Tasks

Although Satformer was developed to address traffic data estimation in satellite networks, the core strengths of its methodology endow it with the potential to be applied to other tensor completion tasks requiring the handling of large-scale, sparse, and complex spatio-temporal characteristics. This includes, but is not limited to, social network analysis, environmental monitoring, bioinformatics, and other domains that necessitate the reconstruction and analysis of multidimensional data.

Table 4. Performance Under Foursquare tensor dataset

Models	NMAE				
	2%	4%	6%	8%	10%
CoSTCO	0.2548	0.2513	0.2493	0.2425	0.2402
DAIN	0.2434	0.2412	0.2401	0.2396	0.2368
SPIN	0.1998	0.1973	0.1936	0.1922	0.1913
STCAGCN	0.1999	0.1974	**0.1933**	0.1929	0.1918
Satformer	**0.1996**	**0.1967**	0.1936	**0.1920**	**0.1913**
Models	NRMSE				
	2%	4%	6%	8%	10%
CoSTCo	0.1465	0.1460	0.1454	0.1449	0.1438
DAIN	0.1464	0.1460	0.1453	0.1450	0.1439
SPIN	0.1322	0.1317	0.1308	**0.1293**	**0.1291**
STCAGCN	0.1328	0.1321	0.1309	0.1296	**0.1291**
Satformer	**0.1320**	**0.1311**	**0.1304**	0.1295	0.1294

We use the Foursquare tensor dataset (Oh et al., 2021) as a representative example to evaluate Satformer's performance on general tensor completion tasks. The Foursquare dataset



(a) Original matrices (b) Sample matrices (c) Attention maps (d) Estimate matrices

Figure 9. Visualization of adaptive sparse spatio-temporal attention mechanism

is a point-of-interest tensor defined by user, location, and timestamp. Utilizing this dataset allows researchers and developers to gain insights into the dynamics of geographic social networks and to develop innovative applications and services based on these insights.

The test results on the Foursquare dataset demonstrate the applicability of the Satformer method to other general tensor completion tasks. These positive results indicate that Satformer can serve as a powerful tool across various fields that involve complex spatio-temporal data, including social network analysis, traffic flow forecasting, environmental monitoring, and others. Naturally, additional adjustments and optimizations may be necessary for different application scenarios to achieve optimal performance.