Graph-Tensor Singular Value Decomposition for Data Recovery

Lei Deng[1], Haifeng Zheng[1], Xiao-Yang Liu[2], Xinxin Feng[1]

[1]Fuzhou University, Fuzhou, China; [2]Columbia University, New York, USA.

Introduction

In real-world scenarios, data are commonly generated with graph structures, e.g., sensory data in transportation networks, user profiles in social networks, and traffic trace data in Internet. Incomplete graph data limits further data analysis. Hence recovering missing graph data becomes crucial. In this work, we exploit graph-tensor decomposition strategies for data recovery. Experimental results show that the recovery performance of graph data can be significantly improved by adopting graph-tensor decomposition.

Graph-Tensor

- **Graph-tensor:** Given a graph where each node has a data matrix, we can obtain a graph-tensor by stacking the data matrices in the third dimension.
- **Graph-tensor product [1]:** The graph-tensor product between $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is represented as $\mathcal{C} \in \mathbb{R}^{n_1 \times n_4 \times n_3} = \mathcal{A} *_g \mathcal{B}$. The (i,j)-th tube $\mathcal{C}(i,j,:)$ can be computed by $\mathcal{C}(i,j,:) = \sum_{k=1}^{n_2} \mathcal{A}(i,k,:) \star \mathcal{B}(k,j,:)$, where $\star$ denotes the graph convolution between two tensor tubes.
- **Graph-tensor singular value decomposition (GT-SVD) [1]:** GT-SVD is developed to capture the topological feature of graph data. The GT-SVD of a graph-tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is given by

$$\mathcal{T} = \mathcal{U} *_g \mathcal{S} *_g \mathcal{V}^\top. \quad (1)$$

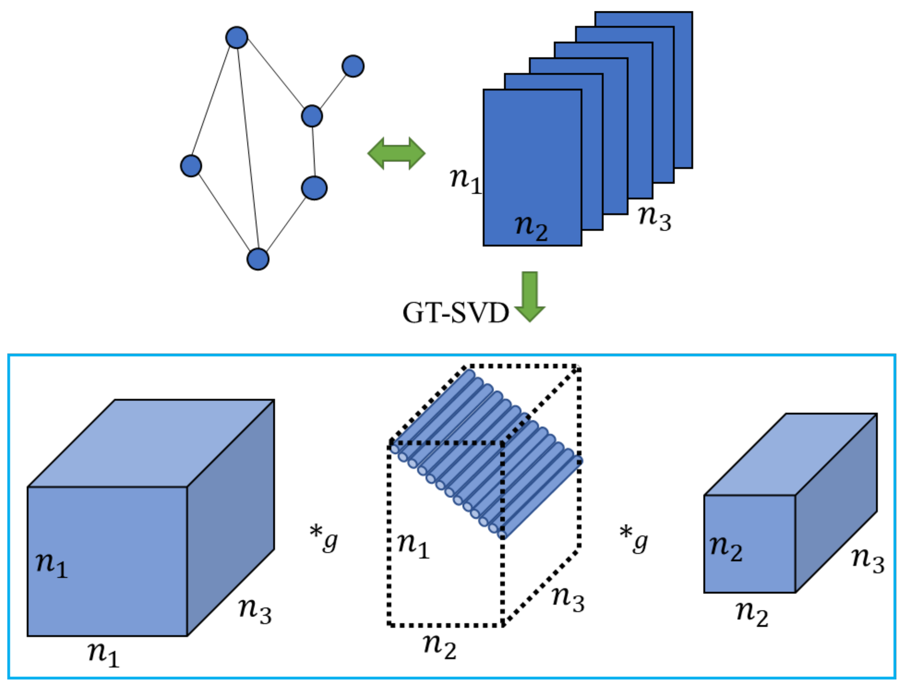

Fig. 1: An illustration of GT-SVD.

GT-SVD

Input: Graph-tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, Laplacian matrix $L \in \mathbb{R}^{n_3 \times n_3}$.
Output: $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$.
$\widetilde{\mathcal{T}} = \mathrm{GFT}(\mathcal{T})$.
for $k = 1: n_3$ **do**
  $[\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}] = \mathrm{svd}(\widetilde{\mathcal{T}}^{(k)})$.
  $\widetilde{\mathcal{U}}^{(k)} = \boldsymbol{U}, \widetilde{\mathcal{S}}^{(k)} = \boldsymbol{S}, \widetilde{\mathcal{V}}^{(k)} = \boldsymbol{V}$.
end
$\mathcal{U} = \mathrm{IGFT}(\widetilde{\mathcal{U}}), \mathcal{S} = \mathrm{IGFT}(\widetilde{\mathcal{S}}), \mathcal{V} = \mathrm{IGFT}(\widetilde{\mathcal{V}})$.

The tubal-rank $\mathrm{rank}_g(\mathcal{T})$ is defined as the number of non-zero singular tubes of $\mathcal{S}$.

- **Graph-tensor completion:** Graph-tensor with strong topological structure is normally low-tubal-rank , implying strong spatial-temporal correlations in the graph-tensor, which can be exploited for data recovery. The graph-tensor completion problem can be formulated as

$$\min_{\mathcal{X}} \|\mathcal{X}\|_{\text{gTNN}}, \text{ s.t. } \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{T}), \quad (2)$$

where $\|\mathcal{X}\|_{\text{gTNN}}$ is the graph-tensor nuclear norm [1], defined by the sum of singular values of all the frontal slices of $\widetilde{\mathcal{X}}$.

Application-1: Traffic Data Imputation

In an intelligent transportation system, sensors are deployed on the two sides of the highways for collecting traffic information.

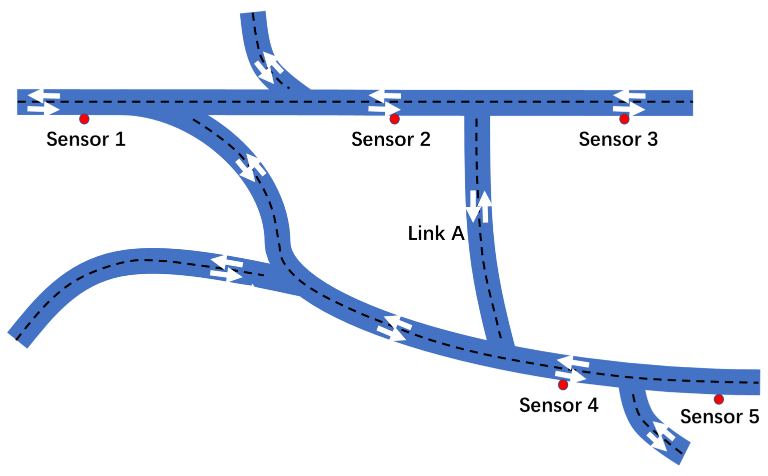

Fig. 2: An illustration of the transportation network.

- **Graph-tensor representation is more effective:** The topological structure of traffic sensors can be represented by a directed graph.

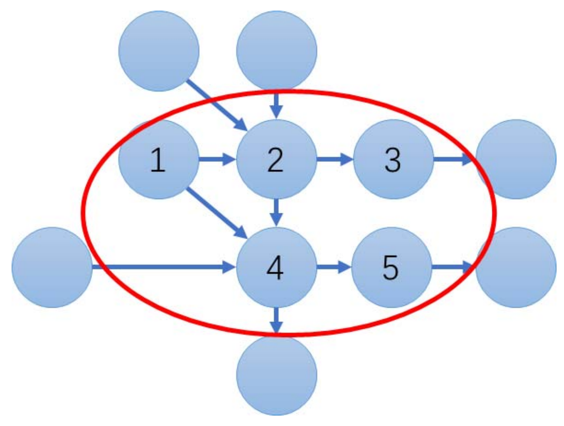

Fig. 3: The topological graph of sensors in the transportation network.

GT-SVD is able to capture more energy compared with the traditional t-SVD.

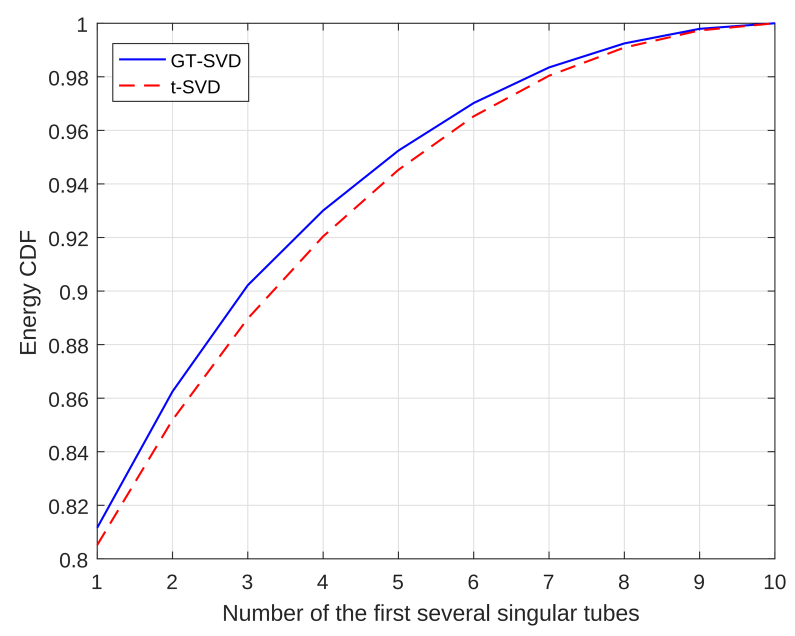

Fig. 4: The energy CDFs of singular tubes captured by t-SVD and GT-SVD.

- **Missing traffic imputation based on GT-SVD [1]:**

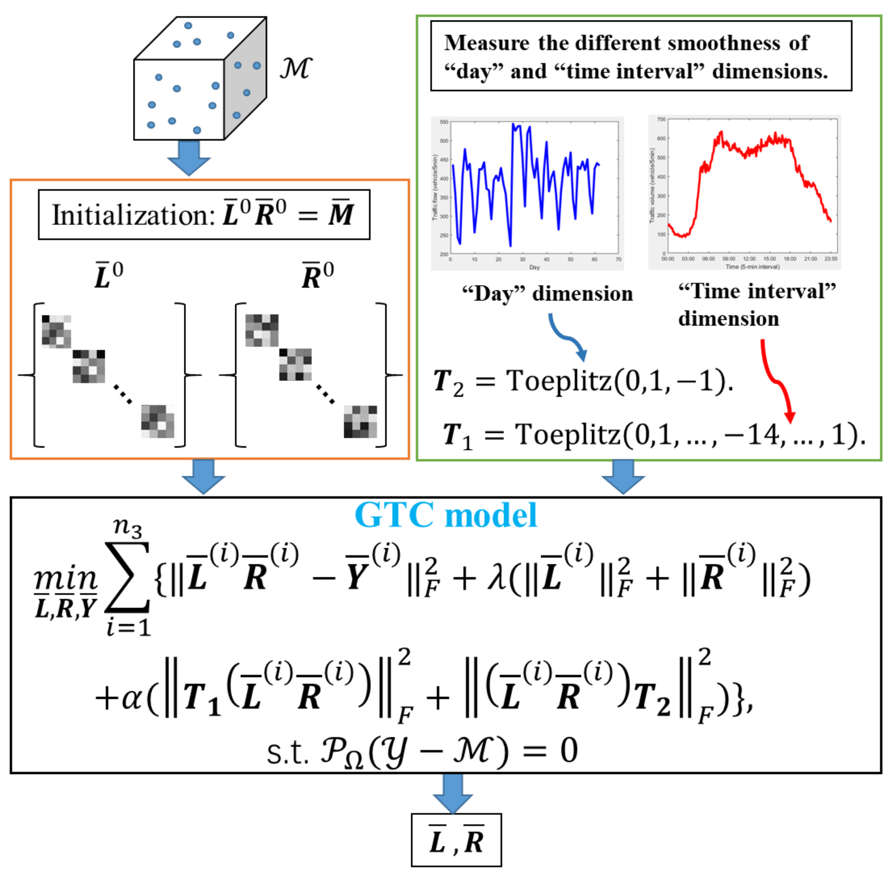

Fig. 5: The overall framework of the proposed algorithm.

- **Experimental results:**

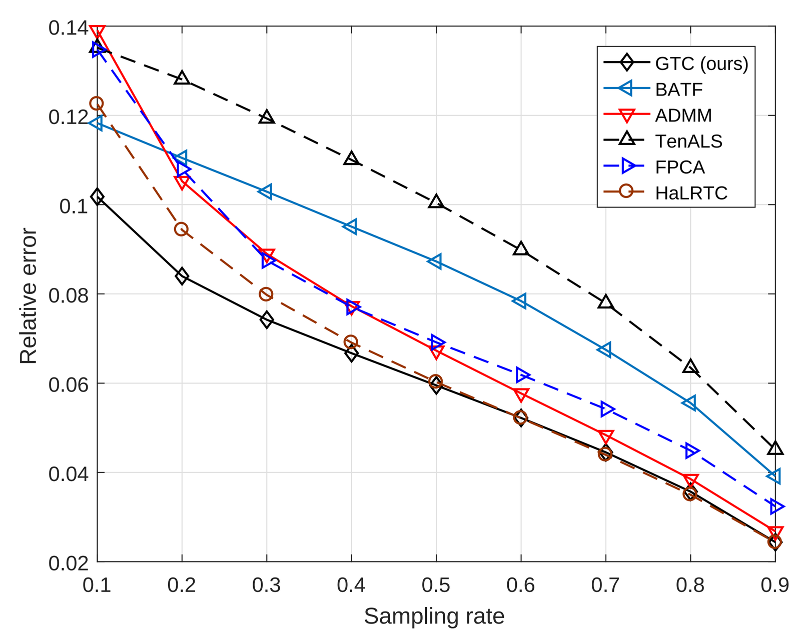

Fig. 6: Performance comparison.

Application-2: Internet Traffic Trace Data Recovery

Since each origin-destination (OD) pair in the network can be views as a vertex in the graph, then the missing Internet Traffic data can be recovered based on the graph-tensor model.

- **Missing pattern:**

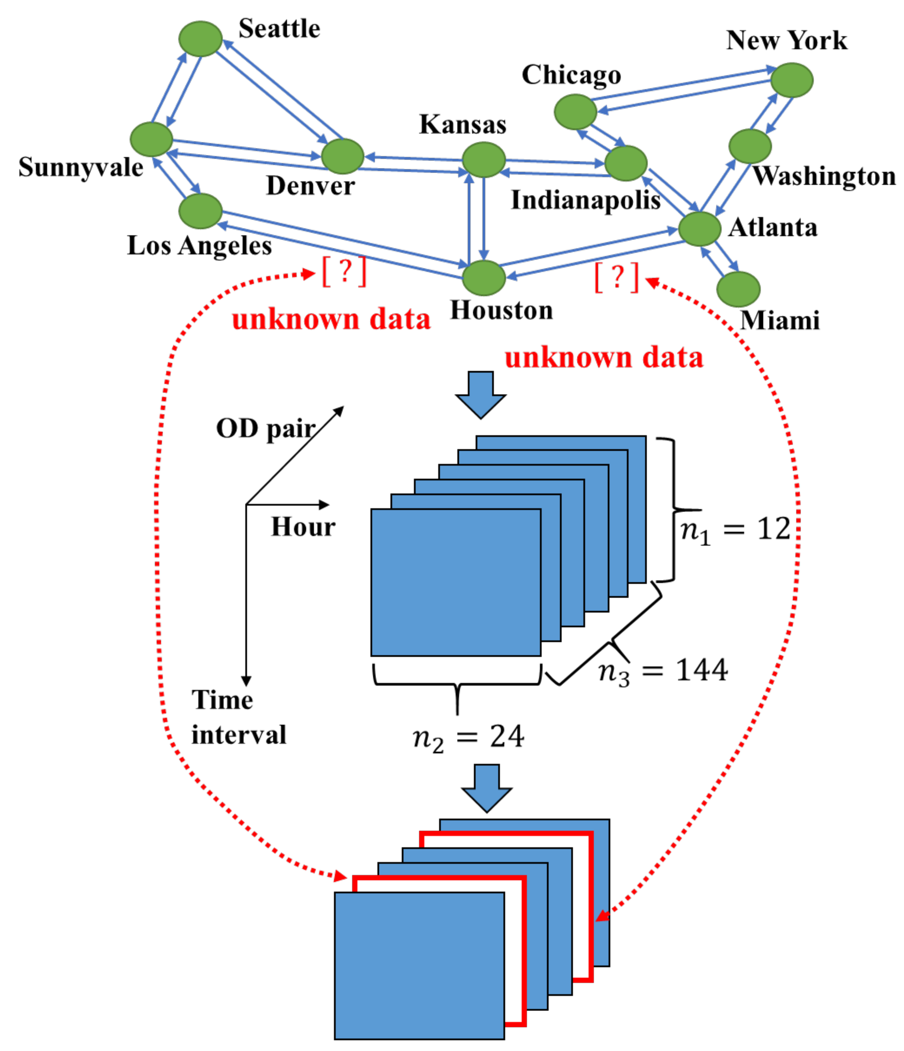

Fig. 7: Illustrations of network traffic with missing data.

- **A deep unfolding based solution:** The iterative ADMM algorithm can be further unfolded into a deep neural network [2].

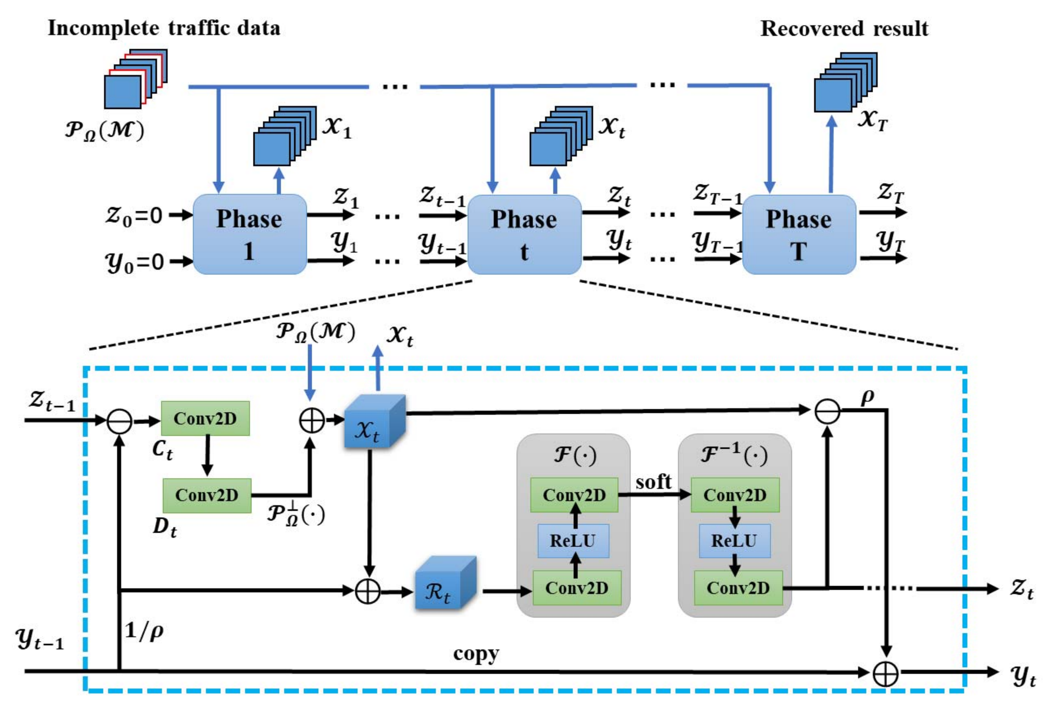

Fig. 8: The overall framework of the proposed deep unfolding network.

- **Experimental results:**

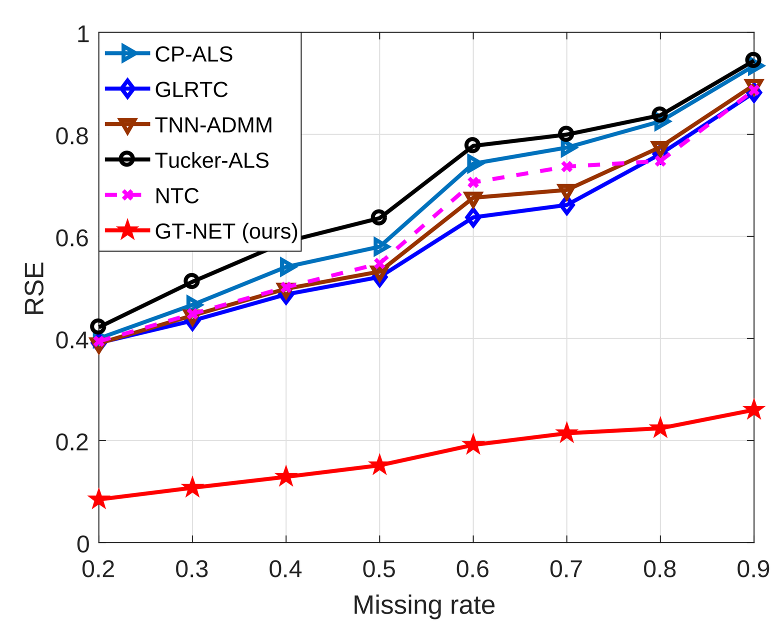

Fig. 9: Recovery error comparison.

Mehods	ours	TNN-ADMM	CP-ALS	Tucker-ALS
Time (ms)	10	1200	1000	500

Conclusions

In this work, we discussed GT-SVD for real-world data recovery applications. Experimental results demonstrated that the GT-SVD based approaches better capture the topological correlations of graph data, and provide higher completion accuracy and run faster than the baselines.

References

[1] L Deng, X Liu, H Zheng, X Feng, and Y Chen. Graph spectral regularized tensor completion for traffic data imputation. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[2] L Deng, X Liu, H Zheng, and X Feng. Deep unfolded graph-tensor nets for network traffic recovery. *submitted to IEEE INFOCOM 2022*.

Contact information:
Lei Deng
denglei0228@gmail.com