Spatial-Temporal Graph Data Mining for IoT-enabled Air Mobility Prediction

Yushan Jiang, Shuteng Niu, Kai Zhang, Bowen Chen, Chengtao Xu, Dahai Liu and Houbing Song, *Senior Member, IEEE*

***Abstract*—Big data analytics and mining have the potential to enable real-time decision-making and control in a range of Internet of Things (IoT) application domains, such as the Internet of Vehicles, the Internet of Wings and the Airport of Things. The prediction toward air mobility, which is essential to the studies of air traffic management, has been a challenging task due to the complex spatial and temporal dependencies in air traffic data with highly nonlinear and variational patterns. Existing works for air traffic prediction only focus on either modeling static traffic patterns of individual flight or temporal correlation, with no or limited addressing of the spatial impact, namely the propagation of traffic perturbation among airports. In this paper, we propose to leverage the concept of graph and model the airports as nodes with time-series features and conduct data mining on graph-structured data. To be specific, firstly, Airline On-Time Performance (AOTP) Data is preprocessed to generate a temporal graph dataset, which includes three features: the number, average delay, and average taxiing time of departure and arrival flights. Then a spatial-temporal graph neural networks model is implemented to forecast the mobility level at each airport over time, where a combination of graph convolution and time-dimensional convolution is used to capture the spatial and temporal correlation simultaneously. Experiments on the dataset demonstrate the advantage of the model on spatial-temporal air mobility prediction, together with the impact of different priors on adjacency matrices and the effectiveness of temporal attention mechanism. Finally, we analyze the prediction performance and discuss the capability of our model. The prediction framework proposed in this work has the potential to be generalized to other spatial-temporal tasks in IoT.**

***Index Terms*—Internet of Things, air mobility, air traffic management, spatial-temporal prediction, graph neural networks, multivariate time-series prediction.**

I. Introduction

With the development of Internet of Things (IoT) [1]–[5], existing fundamental research efforts are being made to facilitate the development of healthcare, civil infrastructure, aeronautics and so on [6]–[9]. In particular, a variety of IoT-based applications and services are being developed to transform the aviation industry. They have the potential to enable safer and more efficient operations, better customer experience, and thus a significant increase in revenue. On the other hand, more and more data become available with the emergence and rise of information systems and data gathering platforms, which have enabled big data analytics and mining for different areas in the IoT ecosystem [10]–[13], as well as a set of algorithms and solutions for more dependable IoT based information managements and data-driven decision making [14]–[17]. For the smart aviation industry, a huge volume of aviation data is gathered to address various tasks toward intelligent air transportation systems. Among these tasks, the prediction toward air mobility has been a key component in the research of air traffic management and the aviation industry.

Originally, air mobility referred to the ability to transport military troops and supplies in and out of combat areas by means of aircraft [18]. In recent years, this concept has been gradually introduced to the aviation industry, which can be described as a safe and efficient aviation transportation system. Moreover, air mobility in aviation is an important research topic toward the development of Advanced Aerial Mobility (AAM) [19], a newly emerging field that aims at safe and responsible operation in an integrated National Airspace System (NAS). Air mobility in aviation is crucial to multiple stakeholders including passengers, airlines, airports, and air traffic management authorities, as a deterioration of air mobility can lead to significantly undesired consequences like severe flight delays, cancellations of flights, and congestion of airports, disrupting the aviation system and resulting in a huge economic loss. Therefore, to achieve efficient airport operation and air traffic management, accurate prediction toward air mobility in dynamic air transportation networks is a requisite. Based on the prediction, timely monitoring and interpose can be executed to capture the perturbation in the air traffic and mitigate the negative impact before the operation efficiency deteriorates.

However, the prediction toward air mobility in aviation remains challenging due to the highly uncertain and dynamic patterns in air traffic data. Most of the existing works based on machine learning methods either focus on extracting static air traffic patterns or only treat it as a regular time-series prediction problem, whose model capacity is unable to address the interdependence of different spatial locations. Therefore, in order to address this challenge, the key is to capture the spatial-temporal dependencies so that the useful features can be extracted effectively and prediction performance can be improved. Recently, researchers have tried to extract spatial-temporal features by composite deep learning models, which

Yushan Jiang, Kai Zhang, Bowen Chen, Chengtao Xu, and Houbing Song are with the Security and Optimization for Networked Globe Laboratory (SONG Lab), Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronatuical University, Daytona Beach, FL 32114 USA e-mail: jiangy2;shutengn;zhangk3;chenb2;xuc3@my.erau.edu; h.song@ieee.org.

Shuteng Niu is with The University of Texas School of Biomedical Informatics, Houston, TX77030 USA e-mail: shutengn@gmail.com.

Dahai Liu is with the College of Aviation, Embry-Riddle Aeronatuical University,Daytona Beach, FL 32114 USA e-mail: dahai.liu@erau.edu.

Corresponding Author: H. Song

Manuscript received April 1, 2021; revised XXX, 2021.

integrate the Recurrent Neural Networks (RNN) with Convolutional Neural Networks (CNN) due to its excellent ability to spatial-related tasks. Nevertheless, the modeling of spatial dependencies based on CNN methods narrows down the input data to be Euclidean, in other words, grid-partitioned, which limits the modeling methods and thus hinders the learning of spatial-temporal correlations in the actual air transportation system.

In recent years, the rise of graph learning has aroused the other formulation of spatial dependencies in traffic prediction tasks. Without the prior knowledge of Euclidean inputs, more complex relationships within the transportation network can be addressed based on graph-structured formulation. A variety of studies on road networks were conducted and achieved success where graph-based learning models were developed to model spatial-temporal dependencies [20]–[23]. On the other hand, there have been many studies of graph theory based on aviation networks [24], [25], which also suggests the feasibility of utilizing graph neural networks to capture the spatial-temporal correlations for the prediction toward aviation air mobility. In our work, we formulate the air transportation network of the United States as a graph, where each commercial airport is treated as a node whose features are a time-series measuring the air mobility level, as shown in Figure 1. Besides, a flight on-time dataset is transformed to a graph dataset, including an adjacency matrix and the aforementioned features. Moreover, we conduct graph data mining based on the obtained dataset and apply a composite model based on spatial-temporal graph neural networks to predict the measurements of air mobility at each airport at a macro-level. The contributions of our paper are summarized as follows:

- To the best of our knowledge, this is the first research which attempts to apply graph neural networks to predict aviation air mobility at a macro level, where multiple measurements are considered, and spatial-temporal dependencies are captured.
- We provide a detailed prepossessing of the raw dataset and construction of spatial-temporal graph dataset.
- We performed extensive ablation experiments to evaluate different priors on adjacency relationships within the air transportation network and verify the effectiveness of temporal attention on our prediction model. In addition, we analyze the model in terms of prediction performance and discuss the model capability.

The remainder of this paper is structured as follows: A literature review of machine learning in the prediction of air traffic patterns and spatial-temporal characteristics is presented in Section II. We formulate the problem and provide our proposed methodology in Section III. The pre-processing of the dataset to be explored is introduced in Section IV. Performance evaluation, ablation experiments, and model analysis are presented in Section V. Section VI concludes this paper.

II. Related Work

In recent years, more and more aviation data were collected and became available due to the development of aviation information systems, which enabled the research on aviation big data analytics and learning-based data mining for various prediction tasks. To be specific, there are some studies on applying machine learning models to historical flight data, Automatic Dependent Surveillance-Broadcast (ADS-B) data, and weather data for the prediction of micro-level patterns such as on-time performance of flight [26]–[28] and macro-level traffic patterns, including the flight delay and traffic flow of regional airspace and airports, in order to analyze and improve air traffic management [29]–[31]. However, the above works mainly consider the temporal patterns and do not address the spatial-temporal correlation in a joint manner, which is an important characteristic for the mining of complex aviation data, especially for the macro-level air traffic prediction task.

To address this issue, the spatial-temporal correlation is considered from data-level and model-level. In the data-level method [32], the spatial-temporal features are extracted from multi-source datasets consisting of the information of flight path, airspace, and weather for the prediction of air traffic flow at a single airport. In [33], features are extracted in a similar manner and a k-means clustering is applied to origin-destination pairs to generate the abstract spatial features. In model-level methods [34], [35], the air traffic patterns are modeled as a temporal 3D cube where the spatial information is constructed as a actual 2D grid map with multiple flight levels, so that CNN and RNN modules can be utilized to extract the spatial correlation and temporal correlation, respectively.

The traditional deep learning models, which capture spatial dependencies in nearby regions, rely on a prior that the explored patterns are grid-partitioned. This prior does not always hold for the flight-driven air transportation network. Recently, with the development of graph research, the concept of convolution has been generalized from grid-like data to graph-structured data, which includes spectral-based methods [36], [37] and spatial-based methods [38], [39]. These works motivated researchers to reconsider the general spatial-

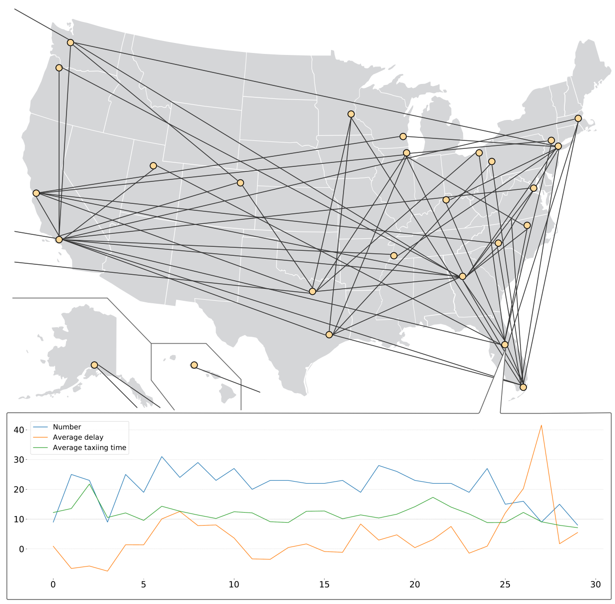

Fig. 1: Visualization of the air transportation network and air mobility measurements at Orlando International Airport(MCO).

temporal tasks and formulate some problems under the context of a graph, which enabled the studies addressing spatial correlations on graph structure. So far, the spatial-temporal graph neural networks have gained success on skeleton-based action recognition [40], sleep stage classification [41] and the forecasting of traffic flow in ground transportation systems [20]–[23]. For the traffic forecasting, these works abstracted the actual road networks or urban as a graph on which each sensor represents a node and provides one or more measurements as the features to be learned. Under this setting, the aggregate strategies of neighborhood information in spatial dimension can be learned and further applied to temporal prediction. While the researches that leverage spatial-temporal graph neural networks toward ground transportation tasks are developing fast, those in air traffic prediction are very limited and underdeveloped. Enlightened by previous researches, we propose a spatial-temporal graph neural network model for predicting macro-level aviation air mobility, where quite a number of airports are considered as nodes with multiple measurements as the input features and targets.

III. Methodology

In this section, we present the problem formulation toward predicting air mobility and introduce the prediction framework based on spatial-temporal graph neural networks. Moreover, the evaluation metrics measuring the performance of the prediction models are presented.

A. Problem Formulation

In this paper, the air transportation network is defined as an undirected graph $\mathbf{G} = (\mathbf{V}, \mathbf{E}, \mathbf{A})$, where $\mathbf{V}$ is a set of nodes with size $|\mathbf{V}| = N$, the number of airports(nodes) in the network, $\mathbf{E}$ denotes the set of edges indicating the connection between two airports, $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the adjacency matrix of graph $\mathbf{G}$ whose elements measures the connectivity of node pairs. As the properties of the air transportation network are different from those of the road network, two different weighted adjacency matrices will be tested based on the distance and the number of flights between departure-and-arrival airports, in addition to the unweighted one. Each airport on the air transportation network will have $\mathbf{F}$ measurements indicating its capacity and air mobility level within a certain period. Each measurement is a time series involving temporal traffic information. To be more specific, the feature matrix is defined as $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_L)^T \in \mathbb{R}^{N\times L\times F}$, where L denotes the number of steps in the entire traffic sequence. For each time step t, $\boldsymbol{X}_t = \left(\boldsymbol{x}_t^1, \boldsymbol{x}_t^2, \ldots, \boldsymbol{x}_t^N\right)^T \in \mathbb{R}^{N\times F}$ denotes all the measurements of all airports. For every element above, $\boldsymbol{x}_t^i = \left(\boldsymbol{x}_t^{i,1}, \ldots, \boldsymbol{x}_t^{i,F}\right)^T \in \mathbb{R}^F$ denotes all the measurements of airport i at time step t and $\boldsymbol{x}_t^{i,j} \in \mathbb{R}$ denotes the value of j-th measurement of $\boldsymbol{x}_t^i$. Besides, we define $\boldsymbol{y}_\tau^i = \boldsymbol{x}_\tau^{i,j}$ as the value of j-th measurement of i-th airport at time τ in the future. In this paper, we consider three measurements: the number, average delay, and average taxiing time of departure-and-arrival flights; and we will discuss them respectively in detail. The future status of all airports at future time step τ can be represented by $\boldsymbol{Y}_\tau = \left(\boldsymbol{y}_\tau^1, \boldsymbol{y}_\tau^2, \ldots, \boldsymbol{y}_\tau^N\right)^T \in \mathbb{R}^{N\times 1}$.

Based on above description, the problem of air mobility prediction is formulated as follows: For a specific time step t, given all the historical measurements of all airports over current and past k steps: $\boldsymbol{X}_{t-k}, \boldsymbol{X}_{t-(k-1)}, \ldots, \boldsymbol{X}_t$, the objective is to predict the most likely status of all airports over future m steps: $\boldsymbol{Y}_{t+1}, \boldsymbol{Y}_{t+2}, \ldots, \boldsymbol{Y}_{t+m}$ as,

$$\hat{\boldsymbol{Y}}_{t+1}, \ldots, \hat{\boldsymbol{Y}}_{t+m} = \underset{\boldsymbol{Y}_{t+1}, \ldots, \boldsymbol{Y}_{t+m}}{\operatorname{argmax}} P(\boldsymbol{Y}_{t+1}, \ldots, \boldsymbol{Y}_{t+m} | \boldsymbol{X}_{t-k}, \ldots, \boldsymbol{X}_t) \tag{1}$$

B. Spatial-Temporal Graph Neural Networks

To capture the spatial-temporal correlation for a better prediction for air mobility, the Spatial-Temporal Graph Neural Networks is leveraged, which mainly consists of node-level graph convolution layers to extract spatial features and time-dimensional convolution layers to extract temporal features. The basic framework of Spatial-Temporal Graph Neural Networks is shown in Figure 2.

1) Graph Convolution Network for Modeling Spatial Dependency: One of the key components for air traffic forecasting is to address the complex spatial dependency, namely to capture the propagation of traffic perturbation among airports. As mentioned above, the air transportation network organizes as a graph, whose structure and related properties can be analyzed by spectral graph convolution. With a graph $\mathbf{G}$ and corresponding adjacency matrix $\mathbf{A}$, the degree matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$ of $\mathbf{G}$ is obtained to describe the degree of each node, which is measured by the number of edges attached to each node. Therefore it is a diagonal matrix with $\mathbf{D}_{i,i} = \sum_j \mathbf{A}_{i,j}$. Laplacian matrix, the other important concept representing the graph with many useful properties, is introduced as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, whose symmetric normalized form is more often used as $\mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$, where $\mathbf{I} \in \mathbb{R}^{N\times N}$ is an identity matrix. As the Laplacian matrix is a positive semi-definite and symmetric matrix, the eigenvalue decomposition can be performed such that it can be rewritten by $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{\Lambda} = \operatorname{diag}([\lambda_0, \ldots, \lambda_{N-1}]) \in \mathbb{R}^{N\times N}$, λ denotes the eigenvalue and $\mathbf{U}$ denotes the orthogonal matrix formed by corresponding eigenvectors. The eigenvalue can be treated as the frequency component, and the matrix serves as a basis of Fourier space, such that Fourier and inverse Fourier transform on graph are introduced as $\hat{f} = \mathbf{U}^T f$ and $f = \mathbf{U}\hat{f}$, where f is the graph signal, i.e. $\boldsymbol{X}_t$ under our context, the traffic information of all airports at time step t. Therefore, the spectral convolution on a graph is defined as the inverse Fourier transform of the product of the learn-able kernel and feature signals in the spectral domain, which is given by $\mathbf{U}(\mathbf{U}^T g)\mathbf{U}^T f = \mathbf{U} g_\theta \mathbf{U}^T f$, where g and g_θ denote the kernel in time and spectral domain respectively.

In the first generation of graph convolutional networks, a single g_θ is often rewritten as a diagonal matrix with N learnable elements. When the number of nodes is large, the number of parameters is even larger for a feature matrix with multi-channels($\mathbf{F}$ in our case), which is not computational efficient. Moreover, performing eigendecomposition on a large graph leads to high computational complexity, and local connectivity is not considered as the whole graph feeds as an input.

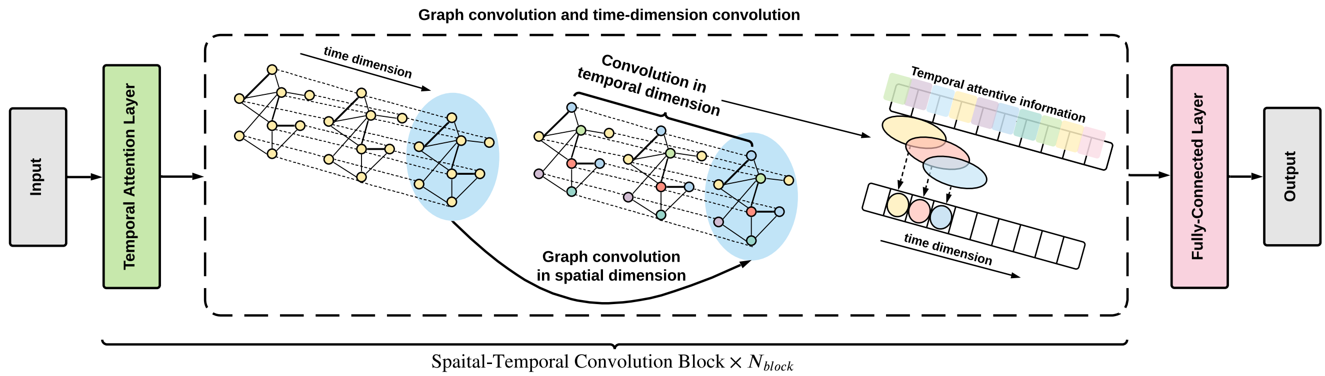

Fig. 2: The basic framework of Spatial-Temporal Graph Neural Networks which consists of the input layer, several spatial-temporal convolution blocks, a fully-connected layer and the output layer.

To address these issues, we can set $g_\theta \approx \sum_{k=0}^{K} \theta_k \mathbf{\Lambda^k}$, where θ_k is a learn-able parameter. This approximation makes the final convolution result as $\sum_{k=0}^{K} \theta_k \mathbf{L}^k f$. It reduces the computational complexity as Laplacian matrix is used directly without eigendecomposition. Besides, the local connectivity, or K-hop receptive field is introduced by $\mathbf{L}^K$. Based on above method, Chebyshev polynomials are leveraged to further improve the computational efficiency, which is given by $g_\theta \approx \sum_{k=0}^{K-1} \theta_k T_k(\hat{\mathbf{\Lambda}})$ and $\hat{\mathbf{\Lambda}} = \frac{2}{\lambda_{max}}\mathbf{\Lambda} - \mathbf{I}$, where λ_{max} is the maximum eigenvalue of $\mathbf{\Lambda}$. The recursive formulas of Chebyshev polynomials are given by $T_0(x) = 1$,$T_1(x) = x$, and $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ for integer $k > 1$. Therefore, the graph convolution layer used in our paper is:

$$\mathbf{GConv} = \mathrm{GConv}(g_\theta, f) = \sigma(\sum_{k=0}^{K-1} \theta_k T_k(\hat{\mathbf{L}}) f) \tag{2}$$

where $\mathrm{GConv}(\cdot)$ denotes the operation of graph convolution, $\hat{\mathbf{L}} = \frac{2}{\lambda_{max}}\mathbf{L} - \mathbf{I}$, $\sigma(\cdot)$ is the activation function. Under our context, we further extend $f \in \mathbb{R}^{N \times F_{in} \times L_{in}}$ to be the input of a certain layer with time dimension L_{in}, while F_{in} denotes the number of input channel, which is F in the first layer. Thus the layer output $\mathbf{GConv}$ has the dimension of $\mathbb{R}^{N \times F_{out} \times L_{in}}$, where F_{out} denotes the number of convolutional kernels.

2) Time-Dimensional Convolution Neural Network for Modeling Temporal Dependency: In order to capture the temporal correlations of the nodes' features that are aggregated from their neighbors by graph convolution layer, a time-dimensional convolution operation is leveraged to extract the features across the time dimension. Note that this convolution is actually a 1-D convolution if we only consider the temporal dimension, which is similar to other recurrent neural networks such as Long-Short Term Memory (LSTM) [42] and Gated Recurrent Units (GRU) [43]. The temporal convolution layer is described as follows:

$$\mathbf{TConv} = \mathrm{TConv}(\psi, \mathbf{GConv}) = \sigma(\psi * (\mathbf{GConv})) \tag{3}$$

where the layer output $\mathbf{TConv} \in \mathbb{R}^{N \times F_{out} \times L_{out}}$, TConv denotes the layer operation, $*$ denotes the convolution operation used in convolutional neural networks, ψ denotes the learn-able convolution kernel, and L_{out} denotes the length of output after the convolution along time dimension.

A graph convolution layer and a time-dimensional convolution layer form a spatial-temporal convolution block that is able to extract spatial features and temporal features at the same time. We expect to capture a wider range of complex dynamics in the air traffic data by stacking several spatial-temporal convolution blocks. Moreover, we also add a skip-connection between two blocks for the countermeasure of the over-smoothing problem and enhance the representation learning. Finally, a fully-connected layer with an activation function is applied to compile the extracted features by previous blocks for prediction tasks.

3) Attention Based Mechanism: Enlightened by [21], [41] and their success on proposing to apply attention mechanism on their own spatial-temporal prediction tasks, we also attempt to employ the attention based mechanism for the purpose of further extracting temporal attentive information in an adaptive manner as spatial-temporal convolution operation is ongoing. The definition of applied attention is presented as follows:

$$\mathbf{AttScore} = \mathbf{V} \cdot Sigmoid(\mathbf{h_1 W h_2^T} + \mathbf{b}) \tag{4}$$

where $\mathbf{V}$,$\mathbf{W}$,$\mathbf{b}$ are learn-able parameters, $Sigmoid(\cdot)$ denotes the sigmoid function, $\mathbf{h_1}$ and $\mathbf{h_2^T}$ are the product of current input and different learn-able kernels. After attention score matrix $\mathbf{AttScore}$ is calculated, a softmax function is used for element-wise normalization and the temporal attention $\mathbf{Att_t} \in \mathbb{R}^{N \times F' \times L'}$ is obtained, where F' denote the number of channels and L' denotes length of the current input. The temporal attention describes the degree of correlation between two time steps and is directly applied to the input when fed into temporal convolution layer. We leverage above temporal attention mechanism to improve the ability of prediction model to capture valuable temporal information.

C. Evaluation Metrics

To evaluate the performance of the prediction model, the following metrics is used:

1) Mean Absolute Error (MAE):

$$\mathbf{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_t - \hat{y}_t| \tag{5}$$

2) Root Mean Absolute Error (RMSE):

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_t - \hat{y}_t)^2} \tag{6}$$

As shown in equations (5) and (6), MAE is the average absolute error of predictions and actual observations, while RMSE is the root of mean squared difference between predictions and actual observations. Therefore, MAE measures the actual situation of prediction error in a better manner, while RMSE can better evaluate the capability of a prediction model to capture patterns of abrupt perturbation as it is more sensitive to abnormal values.

IV. Data Pre-processing

In this section, we provide a detailed prepossessing of the raw dataset and the construction strategy of the spatial-temporal graph dataset, which consists of the adjacency matrix and time-series feature data.

The dataset that we are exploring in this paper is Airline On-Time Performance (AOTP) data of the year 2016, which is provided by the Bureau of Transportation Statistics. This dataset contains the basic on-time arrival and departure information for each non-stop domestic flight, including the schedule and actual departure and arrival data, the carrier, origin and destination with airtime and non-stop distance.

In the first stage, we aggregate the raw data sets by month and set up the cleansing strategy regarding incompleteness, redundancy, and irrelevance. To be specific, we remove the flight records with the high missing rate of flight departure delay, flight arrival delay, departure time, and arrival time. Besides, the redundant flight records are filtered out. As there is other information that is out of our interest, such as flight schedule, aircraft, and airline information, we also filter these features out and get the following ones for future exploration: day of the month, month, departure and arrival airports, actual departure and arrival time, departure delay and arrival delay, taxiing in and taxiing out time, distance of the flight.

As all records in the cleansed data are based on individual flights, we further aggregate the information to generate node-level representations so that it can be fed into the prediction model. The first step is to generate adjacency matrices according to different priors. The first kind of adjacency matrix is unweighted, whose element is one if there is a flight between departure-and-arrival airports and otherwise zero. The second one is a weighted matrix based on the flight distance between departure-and-arrival airports. Moreover, we apply the threshold Gaussian kernel [44] to scale the elements:

$$\mathbf{A_{i,j}} = \begin{cases} exp(-\frac{dist(v_i,v_j)^2}{\sigma^2}) & \text{if } dist(v_i,v_j) \le \kappa \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $\mathbf{A_{i,j}}$ denotes the element of adjacency, $dist(v_i, v_j)$ denotes the flight distance between departure airport v_i and arrival airport v_j, σ denotes the standard deviation of flight distances and κ denotes the threshold of flight distance. This kernel is often used in road networks while we wonder if it is still a good prior for air transportation networks. The third kind of adjacency relationship is also described by a weighted matrix based on the other prior whose element is calculated by the total number of flights between departure-and-arrival airports. Instead of distance, the third prior formulates the spatial connectivity in the air transportation network as a factor of flight frequency. We will further explore these priors under the context of air mobility prediction task.

Besides the adjacency matrix, the other input is a time-series feature data. The first step is to gather all departure information and arrival information separately on an airport basis. Besides, the timezone is assigned to each airport, and the local time of departure and arrival in each record is converted under the same standard. Note that the time and date need to be adjusted again if the converted time of some records is beyond the 24-hour time window due to the time difference. Based on converted time and date, each record is assigned with a sequence number. After that, the records are then grouped by airport and sequence number, where the features are summarized in a 30-minute time window. To be more specific, three features are calculated, the average number, delay and taxiing time of departure and arrival flights. Next, processed departure and arrival data are merged by the key of the airport and then fitted in a template data frame with all airports and the whole sequence. Finally the data frame is constructed as a 3-D array $\boldsymbol{X} \in \mathbb{R}^{N\times L\times F} = \mathbb{R}^{285\times 17606\times 3}$, which means the sequence length is 17606, each of which has 3 measurements of 285 airports.

V. Experiments

In this section, we introduce the experiment settings and conduct extensive experiments regarding the comparison with baseline models, different priors on adjacency matrices, and the temporal attention mechanism in our prediction model. A model analysis with prediction visualizations is also presented to discuss the model capability.

A. Settings

In our experiments, we utilize the air traffic information of the past 4 hours to predict that over the next 90 minutes, which means the input is built by chopping the dataset into a 4-D array by 8 units along the sequence dimension, and the output is built by further gathering 3 units right after the end of each input. Moreover, a zero-mean normalization is performed on the input. The dataset is split into 7:2:1 for the training set, evaluation set, and testing set respectively. The threshold κ for distance-based adjacency matrix is the largest value of flight distance. There are $N_{block} = 4$ spatial-temporal convolution blocks, where the max order of Chebyshev polynomials K used in graph convolution layer is set as 3, the kernel size along the temporal axis is 3 in the time-dimension convolution layer. The activation function within each spatial-temporal convolution block is Rectified Linear Unit (ReLU) [45]. For training details, the batch size is 32, Adamax [46] is used as the optimizer with an initial learning rate 0.001 and decay rate 0.003. The loss function of the model output is Mean Squared Error(MSE).

TABLE I: Performance comparison of different models

Model	Number		Average Delay		Average Taxiing	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
CNN	2.026	4.211	7.401	23.321	3.965	6.143
1-D Conv	2.102	5.972	7.469	25.324	3.883	7.448
LSTM	1.518	3.834	7.190	23.945	3.685	6.010
GRU	1.503	3.870	6.938	23.851	3.448	5.872
3-D Conv	1.724	3.959	6.975	23.019	3.498	5.532
Ours	**0.773**	**1.706**	**6.320**	**22.427**	**2.574**	**5.033**

B. Result Comparison with Baseline Models

On the final dataset, we compare our model with 5 baseline models commonly used in air traffic prediction tasks including spatial learning model CNN, temporal learning models 1-D convolution, LSTM and GRU, and spatial-temporal learning model 3-D convolution. Table I demonstrates the average MAE and RMSE of three measurements, number of departure and arrival flights, average delay of departure and arrival flights, average taxiing time of departure and arrival flights, which indicate the air mobility level of an airport within the air transportation network. It can be observed that our model outperforms all 5 baseline models in terms of all evaluation metrics and all measurements. Among these baseline models, the spatial learning model CNN simply addresses the spatial dependencies in the air transportation network as a grid-like relationship and fails to capture temporal patterns. Temporal learning models can only address temporal correlations in a relatively coarse manner and do not incorporate any countermeasure for complex spatial characteristics. By comparison, the spatial-temporal learning model captures both kinds of dependencies, but the manipulation of spatial characteristics is still limited due to the same constraint in CNN. Differently, our model gains advantages on prediction tasks as it jointly addresses the spatial-temporal correlations, where the spatial dependencies in the air transportation network are captured by the graph convolution module and temporal contexts are also involved and captured by the time-dimensional convolution module.

More specifically speaking, for the number and average taxiing time of departure and arrival flights, our model shows a more obvious strength of capturing general traffic patterns and a larger capability to make an accurate prediction with the existence of abrupt perturbations. For average delay of departure and arrival flights that is highly uncertain with multiple complex factors, our model also gains favors by making a prediction with lower errors in terms of absolute magnitude and against perturbations.

C. Ablation Experiments

1) Exploration of different priors on adjacency matrix: To explore the impact of prior knowledge in the air traffic graph learning, we test three different adjacency matrices: unweighted, flight-based, and distance-based. Table II illustrates the performance comparison of our prediction model with different adjacency matrices as the input. According to the results, prior knowledge of the adjacency relationship can

TABLE II: Comparison of different adjacency matrix

Adjacency	Number		Average Delay		Average Taxiing	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Unweighted	0.849	1.840	6.637	22.738	2.761	5.063
Flight	**0.773**	**1.706**	**6.320**	**22.427**	**2.574**	5.033
Distance	0.811	1.816	6.542	22.689	2.682	**5.024**

TABLE III: Comparison of model with/without temporal attention

		Attention				W/O Attention			
		30 min	60 min	90 min	Overall	30 min	60 min	90 min	Overall
Number	MAE	0.753	0.766	0.801	0.773	0.851	0.913	0.941	0.901
	RMSE	1.663	1.698	1.755	1.706	1.970	2.082	2.174	2.077
Delay	MAE	6.297	6.326	6.339	6.320	6.622	6.697	6.764	6.695
	RMSE	22.227	22.487	22.568	22.427	22.590	22.643	22.700	22.644
Taxiing	MAE	2.416	2.642	2.663	2.574	2.809	2.897	2.964	2.883
	RMSE	4.848	5.090	5.157	5.033	5.228	5.257	5.294	5.259

impact the performance of the prediction model. Besides, it can be easily observed that the model with a flight-based adjacency matrix generally makes more accurate predictions by a marginal improvement in terms of number and delay measurements. For the taxiing measurement, the flight-based adjacency still gives rise to a smaller MAE and a slightly inferior result of RMSE compared to the distance-based adjacency matrix.

These results indicate the appropriateness of flight-based description of adjacency relationship within the air transportation network, which actually makes more sense than a distance-based prior and reflects the actual situation of the flight-driven network: the more flights between two airports, the more contribution it will make when traffic pattern propagates. Nevertheless, the distance-based adjacency relationship still leads to a slightly lower MAE and RMSE comparing to the unweighted one, and it seems to be a reasonable prior in some scenarios when traffic events occur: The impact of traffic events may share within a region covering two or more airports that are geographically close to each other, especially when the scales of such events are large.

2) Analysis of temporal attention mechanism: To verify the benefits of the temporal attention mechanism, we also perform ablation experiments on our prediction model with and without temporal attention involved. Table III shows the metrics of two settings for predicting three measurements over 90 minutes with the overall and individual results at each time step. These results illustrate that the attention mechanism on temporal dimension boosts the performance of the prediction model. To be more specific, the model with temporal attention achieves lower errors by an obvious margin for 30-minute, 60-minute and 90-minute traffic prediction, and in terms of all measurements, which indicates the improvement of our model in both short-term and long-term traffic prediction. Besides, with the temporal attention, the performance of our prediction model generally has less tendency to decrease as the time step goes, which suggests a certain degree of robustness regarding the time of prediction. The above results and analysis show the effectiveness of the temporal attention mechanism, where the valuable dynamic temporal information is exploited in an adaptive manner for prediction tasks.

D. *Model Analysis*

To further analyze our prediction model, we select a 400-step time window(200 hours) of Orlando International Airport(MCO) in the testing set and visualize the ground truth and prediction results of different models as shown in Figure 3, Figure 4 and Figure 5. For the purpose of clear visualization, we select a deep learning model GRU beside our model, as it is generally better than other baseline models. We can observe that for the measurements of average taxiing time and the number of departure and arrival flights that have certain periodic patterns, either explicit or not, our model can have a relatively accurate prediction and show the capability to keep tracking of actual air traffic situation. On the other hand, the GRU model can also predict the trend of mobility level changes, but the prediction is more likely to be smoother; namely, it seems not acute enough to capture the abrupt perturbation. It could be due to the lack of handling complex spatial dependencies, and the model prediction could be simply smoothed by the data of smaller airports that have fewer flights and appear to be inactive.

Moreover, although deep learning models like the 3-D convolution and composite CNN-RNN models generally have a certain capability to address spatial-temporal correlation, they have certain limitations due to the constraint of the input. To be specific, the input fed into any deep learning model must be shaped as grid-like data, namely Euclidean data. However, the complex spatial dependencies in some scenarios seem more suitable to be modeled in non-Euclidean space, such as the social network and traffic network. Our model leverages the advantage of graph neural networks, which have no prior on Euclidean input, so that the propagation of traffic status among airports can be better captured by learning to aggregate the features of neighboring airports that are not necessarily neighbors in a geographical sense.

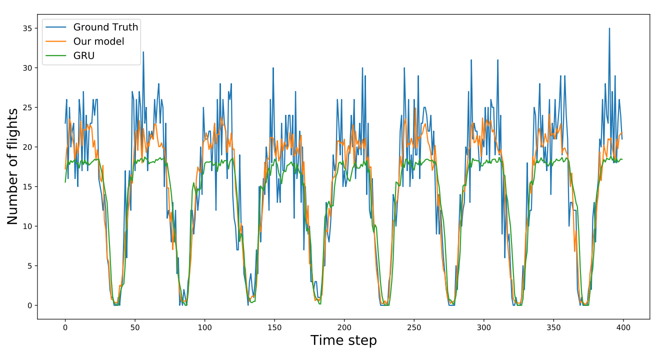

Fig. 3: Visualization results for number of departure and arrival flights at MCO.

For the measurement of average delay of departure and arrival flights that is even more uncertain and complicated, similar conclusions can be drawn for the baseline model. In the meantime, our prediction model can still capture short-period heavy delay patterns and roughly track the trend of fluctuation. Nevertheless, our model shows one drawback, which also reflects in other measurements: our model has limited capability to seize highly variational air traffic patterns, for example, two peak delay around time step 80 and 130, as shown in Figure 5.

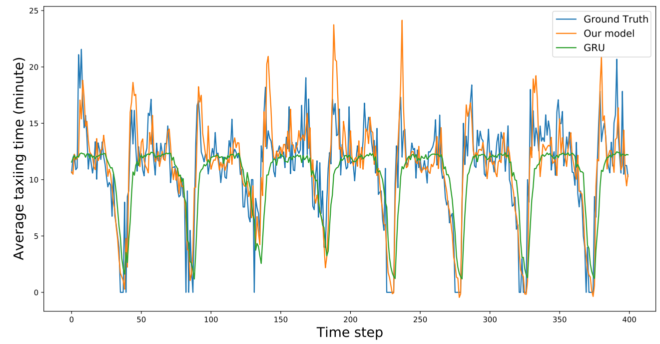

Fig. 4: Visualization results for average taxiing time of departure and arrival flights at MCO.

We would explain such limited prediction capability from two points of view. From the perspective of the model, the stack of graph convolution layers can give rise to an over-smoothing issue, namely over-aggregating the information of the neighborhood for each node. Even if we add a skip-connection between two spatial-temporal blocks to alleviate this issue, the receptive field of a node could still inevitably extend to the whole graph and reduce the diversity of each node within the network which impedes the representation learning of nodes' features. From the data perspective, highly variational air traffic patterns are often caused by multi-factors such as weather, operational interference, mechanical failure, and event-based disturbance, which is not involved in the dataset we explored, thus not learned by our prediction model.

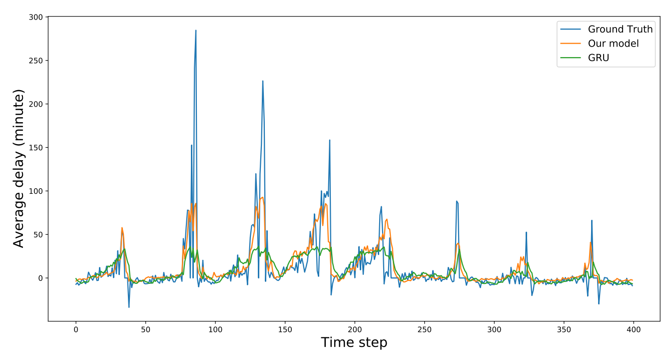

Fig. 5: Visualization results for average delay of departure and arrival flights at MCO.

VI. Conclusion and Future Work

In this paper, we conduct spatial-temporal graph data mining regarding the prediction toward aviation air mobility for airports. Different from existing works that only consider modeling temporal dependencies or model spatial dependencies in a grid-partitioned manner, we formulate the air transportation network in a graph structure and construct a graph dataset from AOTP data. Then we apply spatial-temporal graph neural networks to predict three measurements related to air mobility: number, average delay, and average taxiing time of departure and arrival flights at 285 airports in the United States. The experiment results show the advantage of our model in air traffic prediction tasks. Moreover, we conduct ablation studies to compare different priors on adjacency relationships in the air transportation network and verify the effectiveness of the temporal attention mechanism. Furthermore, we analyze the performance of our prediction model and discuss the model capability. For future work, we suggest involving the weather data and event information for a more accurate spatial-temporal prediction toward air mobility in aviation. The spatial-temporal prediction framework proposed in this paper has the potential to be generalized to other spatial-temporal applications to enable real-time decision-making and control in IoT.

Acknowledgment

This research was supported by the Center for Advanced Transportation Mobility (CATM), USDOT Grant #69A3551747125 and through the National Science Foundation under Grant No. 1956193.

References

[1] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.

[2] S. Jeschke, C. Brecher, H. Song, and D. Rawat, *Industrial Internet of Things*. Cham, Switzerland: Springer, 2017.

[3] H. Song, R. Srinivasan, T. Sookoor, and S. Jeschke, *Smart cities: foundations, principles, and applications*. John Wiley & Sons, 2017.

[4] G. Dartmann, H. Song, and A. Schmeink, *Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things*. Elsevier, 2019.

[5] L. Farhan, R. Kharel, O. Kaiwartya, M. Quiroz-Castellanos, A. Alissa, and M. Abdulsalam, "A concise review on internet of things (iot) - problems, challenges and opportunities," in *2018 11th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, 2018, pp. 1–6.

[6] X. Zheng and Z. Cai, "Privacy-preserved data sharing towards multiple parties in industrial iots," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 968–979, 2020.

[7] K. Kumar, S. Kumar, O. Kaiwartya, Y. Cao, J. Lloret, and N. Aslam, "Cross-layer energy optimization for iot environments: technical advances and opportunities," *Energies*, vol. 10, no. 12, p. 2073, 2017.

[8] Y. Liang, Z. Cai, J. Yu, Q. Han, and Y. Li, "Deep learning based inference of private information using embedded sensors in smart devices," *IEEE Network*, vol. 32, no. 4, pp. 8–14, 2018.

[9] L. Farhan, O. Kaiwartya, L. Alzubaidi, W. Gheth, E. Dimla, and R. Kharel, "Toward interference aware iot framework: Energy and geo-location-based-modeling," *IEEE Access*, vol. 7, pp. 56 617–56 630, 2019.

[10] Z. Lv, H. Song, P. Basanta-Val, A. Steed, and M. Jo, "Next-generation big data analytics: State of the art, challenges, and future research topics," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1891–1899, 2017.

[11] Y. Liu, X. Weng, J. Wan, X. Yue, H. Song, and A. V. Vasilakos, "Exploring data validity in transportation systems for smart cities," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 26–33, 2017.

[12] X. Zheng, Z. Cai, and Y. Li, "Data linkage in smart internet of things systems: a consideration from a privacy perspective," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 55–61, 2018.

[13] J. Granat, J. M. Batalla, C. X. Mavromoustakis, and G. Mastorakis, "Big data analytics for event detection in the iot-multicriteria approach," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4418–4430, 2019.

[14] Z. Cai and Z. He, "Trading private range counting over big iot data," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 144–153.

[15] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, 2018.

[16] Q. Song, B. Zong, Y. Wu, L.-A. Tang, H. Zhang, G. Jiang, and H. Chen, "Tgnet: Learning to rank nodes in temporal graphs," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 97–106.

[17] K. Li, G. Lu, G. Luo, and Z. Cai, "Seed-free graph de-anonymiztiation with adversarial learning," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 745–754.

[18] J. J. Tolson, *Airmobility, 1961-1971*. Department of the Army, 1973, vol. 90, no. 4.

[19] N. A. of Sciences, M. Engineering *et al.*, *Advancing aerial mobility: A national blueprint*. National Academies Press, 2020.

[20] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018, pp. 3634–3640.

[21] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 922–929.

[22] M. Li and Z. Zhu, "Spatial-temporal fusion graph neural networks for traffic flow forecasting," *arXiv preprint arXiv:2012.09641*, 2020.

[23] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2019.

[24] S. Dunn and S. M. Wilkinson, "Increasing the resilience of air traffic networks using a network graph theory approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 90, pp. 39–50, 2016.

[25] G. Lykou, P. Dedousis, G. Stergiopoulos, and D. Gritzalis, "Assessing interdependencies and congestion delays in the aviation network," *IEEE Access*, 2020.

[26] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140–150, 2020.

[27] S. Ayhan, P. Costas, and H. Samet, "Predicting estimated time of arrival for commercial flights," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 33–42.

[28] Y. Jiang, Y. Liu, D. Liu, and H. Song, "Applying machine learning to aviation big data for flight delay prediction," in *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 2020, pp. 665–672.

[29] Y. Liu, Y. Liu, M. Hansen, A. Pozdnukhov, and D. Zhang, "Using machine learning to analyze air traffic management actions: Ground delay program case study," *Transportation Research Part E: Logistics and Transportation Review*, vol. 131, pp. 80–95, 2019.

[30] Z. Yang, Y. Wang, J. Li, L. Liu, J. Ma, and Y. Zhong, "Airport arrival flow prediction considering meteorological factors based on deep-learning methods," *Complexity*, vol. 2020, 2020.

[31] K. Zhang, Y. Liu, J. Wang, H. Song, and D. Liu, "Tree-based airspace capacity estimation," in *2020 Integrated Communications Navigation and Surveillance Conference (ICNS)*. IEEE, 2020, pp. 5C1–1.

[32] K. Zhang, Y. Jiang, D. Liu, and H. Song, "Spatio-temporal data mining for aviation delay prediction," in *2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC)*. IEEE, 2020, pp. 1–7.

[33] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," *Transportation research part C: Emerging technologies*, vol. 44, pp. 231–241, 2014.

[34] H. Liu, Y. Lin, Z. Chen, D. Guo, J. Zhang, and H. Jing, "Research on the air traffic flow prediction using a deep learning approach," *IEEE Access*, vol. 7, pp. 148 019–148 030, 2019.

[35] Y. Lin, J.-w. Zhang, and H. Liu, "Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation," *Aerospace Science and Technology*, vol. 93, p. 105113, 2019.

[36] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29, 2016.

[37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[38] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[39] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 1025–1035.

[40] S. Yan, Y. Xiong, and D. Lin, “Spatial temporal graph convolutional networks for skeleton-based action recognition,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.
[41] Z. Jia, Y. Lin, J. Wang, R. Zhou, X. Ning, Y. He, and Y. Zhao, “Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification,” in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI*, 2020, pp. 1324–1330.
[42] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[43] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” in *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.
[44] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” *IEEE signal processing magazine*, vol. 30, no. 3, pp. 83–98, 2013.
[45] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 807–814.
[46] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *ICLR (Poster)*, 2015.

Yushan Jiang is a Ph.D. student in the Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University (ERAU), Daytona Beach, Florida. He is a graduate research assistant in the Security and Optimization for Networked Globe Laboratory (SONG Lab). His major research interests include machine learning, spatial-temporal data mining and Internet of Things.

Shuteng Niu received the Ph.D. Ph.D. degree in Electrical Engineering and Computer Science from Embry-Riddle Aeronautical University, Daytona Beach, Florida, in May 2021. He is a postdoctoral research fellow at University of Texas Health Science Center at Houston, School of Biomedical Informatics. His major research interests include machine learning, computer vision, and biomedical informatics.

Kai Zhang received the M.S. degree in Electrical Engineering from Stevens Institute of Technology, NJ, US, in 2019. He is currently pursuing the Ph.D. degree in Electrical Engineering and Computer Science at Security and Optimization for Networked Globe Laboratory, Embry-Riddle Aeronautical University, FL, US. His research interests include intelligent transportation systems and machine learning.

Bowen Chen received his M.S. in Electrical Engineering in 2020 from the Embry-Riddle Aeronautical University (ERAU) Daytona Beach, Florida and B.S. from Civil Aviation University of China (CAUC) in 2017. He was a graduate research assistant in the Security and Optimization for Networked Globe Laboratory (SONG Lab). His major research interests include machining learning and data mining.

Chengtao Xu is currently pursuing his Ph.D. degree in the Department of Electrical Engineering & Computer Science at Embry-Riddle Aeronautical University. He is a graduate research assistant in the Security and Optimization for Networked Globe Laboratory (SONG Lab). He received a master’s degree in Electrical Engineering from Minnesota State University, Mankato, in 2019. His research interests are broadly in the areas of UAV swarm, Advanced Air Mobility, Spectrum Sensing, and Multi-agent System.

Dahai Liu received the Ph.D. degree in Industrial engineering from the University of Nebraska-Lincoln in August 2002, and the M.E. degree in Systems Engineering, B.E., in Management Information Systems from Tsinghua University, Beijing, China, in 1996 and 1993, respectively.

In 2002, he joined the Department of Human Factors and Systems, Embry-Riddle Aeronautical University (ERAU), Daytona Beach, Florida, as an Assistant Professor, became Tenured Associate Professor in 2008 and Professor in 2014. In 2015, he joined the School of Graduate Studies in College of Aviation at ERAU as a tenured Professor. His current research interests include system engineering design and optimization, systems modeling and simulation, human factors engineering and artificial intelligence. He is currently serving the editorial board for International Journal of Industrial Ergonomics, Human Factors and Ergonomics in Manufacturing and Service Industries, and he has been an editorial member for the International Journal of Aviation Psychology from 2009-2019. He is the author of the book “System Engineering: Design Principles and Models” and has published more than 90 journal articles, conference proceedings, book chapters and technical reports. He is the Organizing Chair of the 3rd CATM Annual Symposium in 2019 and conference co-chair of ISIATM 2012. Dr. Liu is a senior member of IIE.

Houbing Song (M’12–SM’14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in August 2012.

In August 2017, he joined the Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University, Daytona Beach, FL, where he is currently an Assistant Professor and the Director of the Security and Optimization for Networked Globe Laboratory (SONG Lab, www.SONGLab.us). He has served as an Associate Technical Editor for IEEE Communications Magazine (2017-present), an Associate Editor for IEEE Internet of Things Journal (2020-present), IEEE Transactions on Intelligent Transportation Systems (2021-present), and IEEE Journal on Miniaturization for Air and Space Systems (J-MASS) (2020-present). He is the editor of seven books, including Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things, Elsevier, 2019, Smart Cities: Foundations, Principles and Applications, Hoboken, NJ: Wiley, 2017, Security and Privacy in Cyber-Physical Systems: Foundations, Principles and Applications, Chichester, UK: Wiley-IEEE Press, 2017, Cyber-Physical Systems: Foundations, Principles and Applications, Boston, MA: Academic Press, 2016, and Industrial Internet of Things, Cham, Switzerland: Springer, 2016. He is the author of more than 100 articles. His research interests include cyber-physical systems, cybersecurity and privacy, internet of things, edge computing, AI/machine learning, big data analytics, unmanned aircraft systems, connected vehicle, smart and connected health, and wireless communications and networking. His research has been featured by popular news media outlets, including IEEE GlobalSpec’s Engineering360, USA Today, U.S. News & World Report, Fox News, Association for Unmanned Vehicle Systems International (AUVSI), Forbes, WFTV, and New Atlas.

Dr. Song is a senior member of ACM and an ACM Distinguished Speaker. Dr. Song was a recipient of 5 Best Paper Awards (CPSCom-2019, ICII 2019, ICNS 2019, CBDCom 2020, and WASA 2020).