Spatio-Temporal Identity Multi-Graph Convolutional Network for Traffic Prediction in the Metaverse

Haihan Nan, Ruidong Li, *Senior Member, IEEE*, Xiaoyan Zhu, *Member, IEEE*, Jianfeng Ma, *Member, IEEE*, and Kaiping Xue, *Senior Member, IEEE*

***Abstract*— The metaverse is at the forefront of the next-generation internet application, where billions of users seamlessly immerse themselves in a hybrid reality of physical-virtual worlds and switch between virtual environments thanks to reliable resource allocation and synchronization. However, the exponential growth of users and computationally intensive applications make joint optimization of multiple indicators challenging. Therefore, predicting user behavior is pivotal in assisting the optimization process. Although graph neural networks have demonstrated remarkable performance in traffic prediction, most existing schemes link nodes based on their distances and require significant computational resources, limiting their generalization and deployment in the metaverse. To solve this problem, we propose an efficient Spatio-temporal Identity Multi-graph convolutional network Framework (SIMF) for application-level traffic prediction in the metaverse. In the SIMF, we design a spatio-temporal embedding layer and multi-graph convolutional module to jointly capture spatio-temporal correlations among nodes (avatars) and reduce the dependence on topology information, which is more consistent with the real relationship between avatars in the metaverse. We conduct extensive experiments to evaluate the SIMF, which show that our proposed framework achieves superior accuracy even without graph information while maintaining low time complexity, making it suitable for traffic prediction in the metaverse.**

***Index Terms*— Metaverse, traffic prediction, graph neural network, digital twin, artificial intelligence.**

I. Introduction

IN the quest to explore the next-generation Internet paradigm, the concept of the metaverse has once again emerged into the spotlight, following its first mention in the science fiction novel *Snow Crash* in 1992 [1]. The metaverse is envisioned to be a boundless, eternal, and self-sustaining surreality space where individuals can seamlessly connect and transition between the physical world and virtual realms. Through advanced infrastructure and enabling techniques, people can fully immerse themselves in the metaverse, bringing an entirely consistent experience with reality or even transcending it. For example, Virtual Reality (VR) technology can render characters in the metaverse with unique or unrelated appearances, known as avatars with varying characteristics. Meanwhile, the virtual environment can be sampled from the actual physical surroundings or created as completely virtual scenes, which break free from the constraints of spatio-temporal limitations.

The metaverse offers an unparalleled experience that transcends that of the traditional Internet. However, it also comes with a concomitant need for increased bandwidth, lower latency, and more computational power to support real-time rendering and synchronization between physical and virtual worlds. To deliver a realistic view with high Quality of Service (QoS) and Quality of Experience (QoE), the virtual world necessitates a bit rate of up to 1Gbps and a motion-to-photon delay of no more than 20ms [2]. Although innovative infrastructure and enabling techniques can expedite the realization of the metaverse, such as 6G providing a peak data rate of 10TB and blockchain guaranteeing data quality while preserving security and privacy [3], they fail to keep pace with the exponential increase in users' demand. Therefore, it is more worthy of consideration to explore ways to maximize resource utilization under current conditions, where traffic prediction is one of the essential technologies.

Various resource allocation frameworks have been widely discussed to facilitate data synchronization for the metaverse. In particular, Han et al. [4] considered a resource allocation problem in which IoT devices are hired by virtual service providers (VSPs) to achieve synchronization. The rewards of device owners were adapted to an equilibrium state based on evolutionary game theory, which helped in the realization of digital twins. To reduce the computational threshold of seamless VR rendering, Xu et al. [5] introduced an incentive mechanism based on reinforcement learning that could achieve an equilibrium between social welfare and auction information exchange costs. Ng et al. [6] proposed a stochastic optimal resource allocation framework that could effectively minimize the cost of various resources without actual knowledge of users' demands. While the existing resource allocation

Manuscript received 15 March 2023; revised 1 August 2023; accepted 31 August 2023. Date of publication 21 December 2023; date of current version 1 March 2024. This work was supported in part by the Fundamental Research Funds for the Central Universities, Innovation Fund of Xidian University under Grant YJSJ23012; in part by the National Natural Science Foundation of China under Grant 61772406 and Grant 61941105; and in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 23H03380. *(Corresponding authors: Ruidong Li; Xiaoyan Zhu.)*

Haihan Nan is with the School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi 710071, China, and also with the Institute of Natural Sciences, Kanazawa University, Kanazawa 9201164, Japan (e-mail: hhnan@stu.xidian.edu.cn).

Ruidong Li is with the Institute of Natural Sciences, Kanazawa University, Kanazawa 9201164, Japan (e-mail: liruidong@ieee.org).

Xiaoyan Zhu is with the School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi 710071, China (e-mail: xyzhu@mail.xidian.edu.cn).

Jianfeng Ma is with the School of Cyber Engineering and the Shaanxi Key Laboratory of Network and System Security, Xidian University, Xi'an 710071, China (e-mail: jfma@mail.xidian.edu.cn).

Kaiping Xue is with the School of Cyber Science and Technology, USTC, Hefei 230027, China (e-mail: kpxue@ustc.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2023.3345389.

Digital Object Identifier 10.1109/JSAC.2023.3345389

schemes can reduce network costs and improve QoE by motivating users to share idle computing resources or allocating appropriate resources in advance, their performance may be dramatically exacerbated as multiple resources coexist in a highly dynamic demand environment. Hence, it is crucial to obtain users' requirements ahead of time to maximize the utilization of multiple resources and QoS while minimizing the total system costs.

While traffic prediction techniques forecast future demand by analyzing historical data, they are highly adaptable to the diverse requirements of each layer within the metaverse. On the one hand, computing, networking, and storage resources can be promptly offloaded to edge servers according to the predicted demands in the physical world. On the other hand, virtual service providers can dynamically adjust the network resources required for rendering avatars and virtual environments in the virtual world based on the predicted flow. Obviously, traffic prediction plays a crucial role in enhancing the metaverse's performance.

However, the integration of existing traffic prediction techniques with the unique characteristics of the metaverse presents a significant challenge. Traditional statistical models and machine learning-based traffic prediction methods have exhibited unsatisfactory performance due to their inability to consider the spatial correlations among nodes, especially avatars and their relationships. This limitation severely impedes their deployment in the metaverse. Despite this, graph neural networks have demonstrated the ability to capture spatio-temporal correlations among avatars and achieve better performance based on the graph topology information. However, graph neural networks typically represent devices as nodes and establish edge connections between them based on their physical distances, which fails to reflect the intrinsic characteristics of the metaverse and poses considerable challenges to their implementation, particularly with regard to establishing connection relationships among all avatars or constructing large-scale graphs based on avatars' physical distances.

To tackle the issues faced by existing traffic prediction techniques, we first define avatars and their relationship based on the graph structure, which better aligns the actual distribution of users in the metaverse and aids in the extraction of spatial dependencies. Subsequently, we propose a novel Spatio-temporal Identity Multi-graph convolutional network Framework (SIMF) for application-level traffic prediction in the metaverse. SIMF initially extracts spatio-temporal identity information to effectively distinguish the spatial and temporal characteristics divergences between avatars, followed by constructing multiple adjacency matrices from topological information, node features, and other dimensions to further refine the spatio-temporal correlation of traffic, which overcomes the limitation of graph neural networks relying solely on nodes' spatial distances. Besides, parallel fully-connected layers are adopted to avoid error accumulation. The contributions of this paper are summarized as follows:

- We define a potential metaverse architecture in which avatars and their relationships can be represented by a graph structure and propose a spatio-temporal traffic prediction framework that evolved from the graph neural network to achieve accurate prediction of user traffic with a small time cost.
- We extract spatio-temporal identity information and design multi-graph matrices to capture the spatio-temporal characteristics of different nodes (avatars), which can also take effect without prior graph topology information.
- We use three real-world datasets to simulate avatars' activities in various sub-metaverse worlds, corresponding to different statuses in the real world. The experiment results demonstrate that our SIMF framework significantly enhances the prediction accuracy with lower time complexity, illustrating the effectiveness of our proposed framework.

The remainder of this paper is organized as follows. Section II introduces the metaverse structure and discusses related works on metaverse and traffic prediction. Section III presents a detailed explanation of the SIMF framework and its components for traffic prediction in the metaverse. The performance of the proposed framework on three real-world datasets is evaluated and analyzed in Section IV. Finally, in Section V, we conclude this paper and provide a discussion on future work.

II. Metaverse Architecture and Related Works

In this section, we first introduce the overall architecture of the metaverse and its related work. Subsequently, we provide a succinct overview of the existing traffic prediction methods and their potential utilization in the metaverse.

A. Metaverse Architecture

The metaverse, powered by advanced infrastructure and techniques, provides users with unparalleled experiences and innovative application scenarios that cannot be fulfilled by the traditional physical world. Building upon the existing definition of metaverse structures [7], [8], we further refine and categorize the metaverse architecture based on several key factors, including the physical and virtual worlds, infrastructure and enabling techniques, metaverse applications, and their characteristics.

As shown in Fig. 1, the physical world consists of *users*, *physical environments*, and *physical goods/services*, which correspond to *avatars*, *virtual environments*, and *virtual goods/services* in the virtual world, respectively. In addition, *physical service providers* are responsible for operating and maintaining the infrastructure and providing services in reality, while virtual service providers aim to provide virtual services to avatars in the virtual world. To bridge the gap between users and avatars, users can convert their actions and status to the metaverse or augment reality experience using extended reality (XR) techniques, including but not limited to virtual reality (VR), augmented reality (AR), mixed reality (MR), and holographic technology [9]. Moreover, by using human-computer interaction (HCI) techniques, such as haptic and brain-computer devices, the interaction and multi-modal sensations of avatars can be transmitted back to the physical

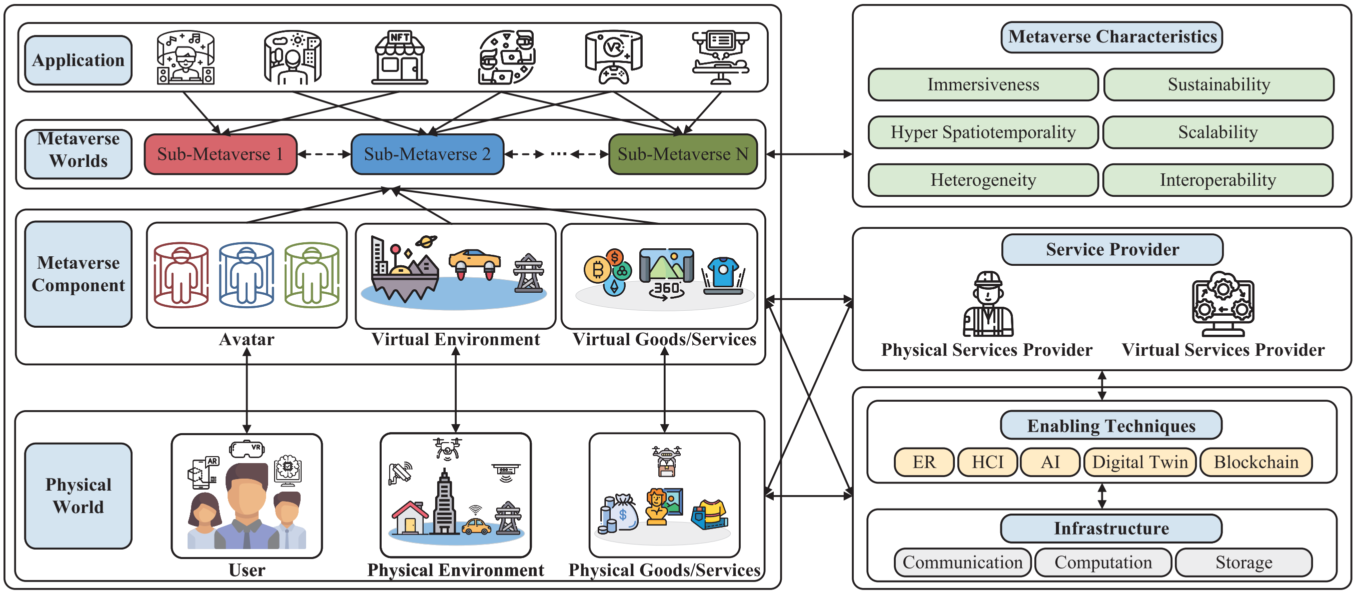

Fig. 1. The architecture of the metaverse.

world with electrical signals, achieving virtual experience feedback [10]. To enhance authenticity and achieve immersive realism in the virtual world, ubiquitous sensors are distributed to collect real-life data and construct a digital environment similar to the physical environment via real-time rendering and synchronization, referred to as digital twin (DT) [11]. In addition, the virtual environment consists of imaged scenes rendered through artificial intelligence. Similarly, various physical goods and services can be transformed into virtual goods and services. Representative cryptocurrencies such as Bitcoin and Ether have gained popularity in the virtual world, facilitating virtual transactions and the circulation of goods [12]. Driven by the novel distributed blockchain-based trading system, users are able to purchase clothing, land, digital creativity [13], etc. Furthermore, avatars can experience virtual services such as virtual tourism with the help of XR techniques.

Supported by physical and virtual service providers, individuals from the real and virtual worlds collectively create a variety of metaverse worlds, each emphasizing different applications and services. Consequently, users and corresponding avatars can engage in various activities through multiple parallel sub-metaverse worlds, such as attending Travis Scott's concert in Fortnite or holding corporate meetings in Horizon Workrooms. Furthermore, the metaverse will accommodate a wide range of applications, including virtual tourism, gaming, and smart factories [14]. These distinct functionalities and applications give rise to six key characteristics of the metaverse: immersiveness, sustainability, hyper spatiotemporality, scalability, heterogeneity, and interoperability [7].

Additionally, different approaches have been proposed to enhance the infrastructure and enabling technologies to meet the requirements for the diverse characteristics of metaverse worlds. Zhou et al. [15] designed a federated learning-based resource allocation algorithm to optimize the weighted sum of energy, delay, and accuracy in the metaverse, significantly easing the constraints of limited communication, networking, and storage resources. Lim et al. [10] focused on several edge intelligent-based frameworks that could considerably reduce overall system expenses. Researchers have also explored the potential combination of enabling techniques within the metaverse [16], [17]. For example, Khan et al. [16] employed machine learning for radio frequency, computing resource allocation, and traffic prediction in the metaverse. Huynh-The et al. [17] concluded that artificial intelligence could assist the metaverse in enhancing immersive experiences and establishing seamless connections with the virtual world. Some works also discussed the application of big data [18], blockchain [5], [19], [20] in the metaverse, and the realization of realistic metaverse systems [14], [21].

B. Traffic Prediction

Traffic prediction has been extensively utilized in fields such as intelligent transportation networks, cellular networks, and smart grids, whereby future trends are predicted based on historical data [22], [23], [24]. While traditional statistical methods such as ARIMA and VAR have been used for this purpose, they often fail to capture the complex nonlinear characteristics of the data [25]. To address this issue, machine learning models like Support Vector Regression (SVR) and long short-term Memory networks (LSTM) have been proposed. Nonetheless, they do not consider the spatial correlation among multiple variables [26].

In recent years, there has been much research in deep learning-based spatio-temporal traffic prediction models to improve prediction accuracy. For instance, models such as fully connected-LSTM (FC-LSTM) [27] and ConvLSTM [28] are capable of separately modeling spatial and temporal characteristics and better extracting the spatial dependency compared to the LSTM model. However, their performance under non-Euclidean topological structures may significantly decrease. Graph neural networks (GNN) have been proposed

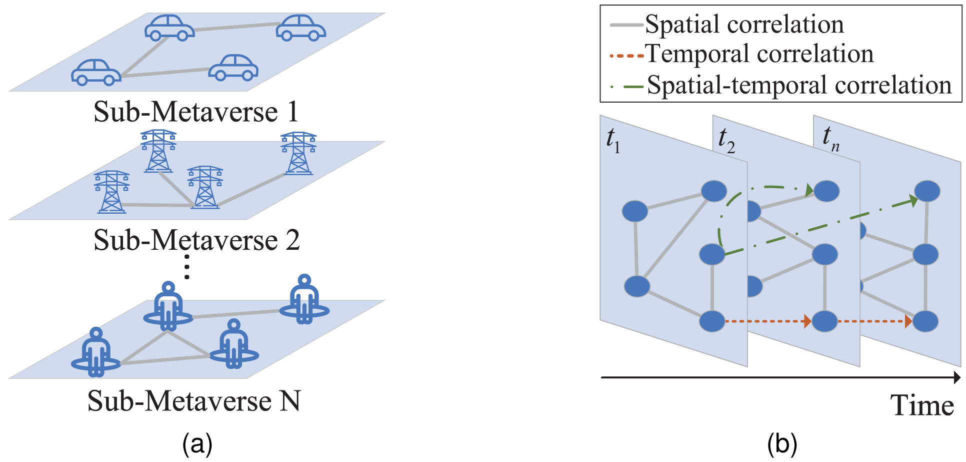

Fig. 2. Spatio-temporal graph structures in the metaverse (a) In each sub-metaverse world, avatars and their relationships are represented as nodes and edges, respectively (b) Nodes are related to themselves and other nodes in three dimensions, including spatial, temporal, and spatial-temporal correlations.

as a way to generalize convolutional networks to graphs in spectral or domains to overcome these models' limitations based on the convolutional network. Li et al. [29] proposed the DCRNN model, which leveraged a graph random walk based encoder-decoder structure to achieve higher accuracy in predicting traffic flows. Meanwhile, the STGCN model reduced computational complexity by utilizing first-order approximations of graph convolutional networks while retaining prediction accuracy [30]. The Graph WaveNet model learned latent spatial dependencies through an adaptive graph matrix and effectively captured spatio-temporal correlations by integrating graph convolution with dilated causal convolution. Additionally, the STSGCN and STFGNN methods utilized a localized spatio-temporal graph to capture spatio-temporal correlations and achieved higher precision with more computational resources [31], [32]. Moreover, Shao et al. [33] distinguished spatio-temporal features through spatio-temporal identity information, which achieved remarkable results without graph information. Although most of these existing traffic prediction techniques accounted for the spatio-temporal correlation among nodes, their reliance on the spatial distance of nodes to establish edge links for traffic prediction makes them ill-suited for the metaverse, where the relationships between avatars are more complex and nuanced than simple spatial distances. At the same time, their performance might sharply deteriorate without detailed graph information, which can be challenging to obtain in the context of the metaverse. Therefore, it is essential to capture the latent spatio-temporal dependencies among avatars and predict their future trends irrespective of the availability of accurate graph topology information. Achieving this objective is crucial for the advancement of an immersive metaverse, which is the main focus of this paper.

III. Proposed Framework

To address the issues identified in Section II-B, we first present a mathematical definition of the problem in the context of the metaverse. Subsequently, we elaborate on the proposed framework SIMF, which includes a spatio-temporal embedding layer and a multi-graph convolutional network module. The proposed framework effectively extracts spatio-temporal correlations even in the absence of graph topology information.

A. Problem Definition

As illustrated in Fig. 2a, the metaverse can be divided into multiple sub-metaverse worlds to handle diverse tasks [7], with each sub-metaverse represented as a weighted directed graph $G = (V, E, \mathbf{A})$, where V denotes the set of nodes, i.e., avatars in each sub-metaverse world, and E represents the set of edge that describes the connectivity among nodes, including but not limited to spatial connectivity, temporal connectivity, attribute connectivity. The adjacency matrix is denoted as $\mathbf{A} \in \mathbb{R}^{N \times N}$. N is the number of nodes in the graph, and $\mathbf{A}_{v_i,v_j}$ measures the correlation between nodes v_i and v_j. Meanwhile, nodes in each sub-metaverse world are linked to each other through three relationships: spatial, temporal, and spatio-temporal correlations (as shown in Fig. 2b). To capture the underlying correlations between nodes and predict future traffic, the traffic prediction problem in the metaverse involves predicting the next T' traffic flows based on T historical spatio-temporal data, which can be formulated as follows:

$$[\mathbf{X}_G^{(t-T+1)}, \ldots, \mathbf{X}_G^{(t)}] \xrightarrow{f} [\mathbf{X}_G^{(t+1)}, \ldots, \mathbf{X}_G^{(t+T')}], \quad (1)$$

where $\mathbf{X}_G^{(t)} \in \mathbb{R}^{N \times C}$ represents the traffic flow of all nodes in the sub-metaverse at time t, and C is the number of features.

B. Spatio-Temporal Embedding Layer

Since traffic flow exhibits spatio-temporal correlation, extracting spatio-temporal correlation information, also called spatio-temporal identity information, can aid in distinguishing samples in both spatial and temporal domains, thereby enhancing prediction accuracy [33]. Inspired by the STSGCN [31] and STID [33] methods, spatial, daily, and weekly information are extracted and concatenated to jointly capture spatio-temporal correlations, which are easily obtained and learned in the metaverse environment through identity and timestamp characteristics and do not require any predefined topology

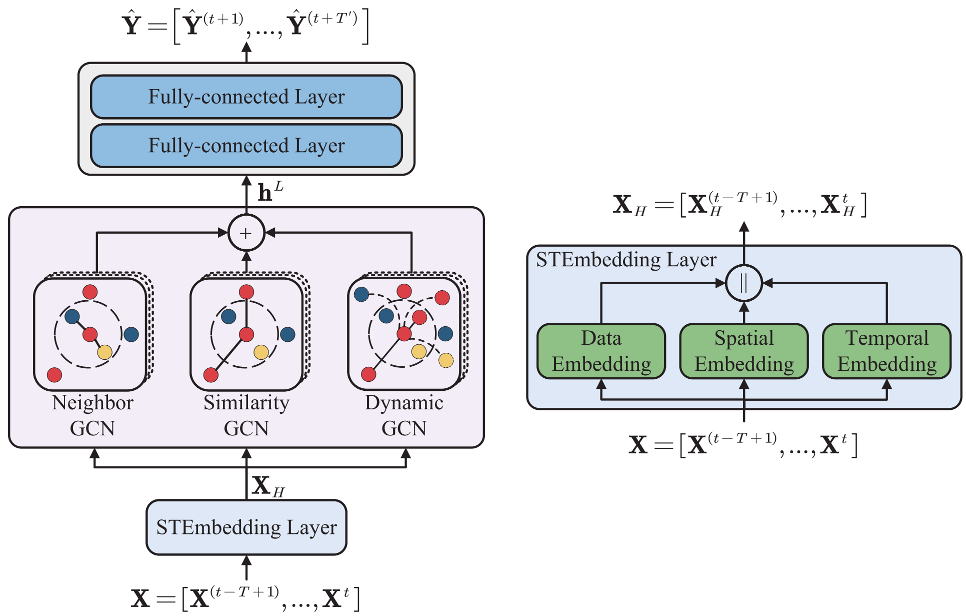

Fig. 3. The framework of SIMF.

information. Specifically, we transform the original spatio-temporal signals $\mathbf{X}_G \in \mathbb{R}^{N \times C \times T}$ into a high-dimensional embedding representation $\mathbf{X}_{emb} \in \mathbb{R}^{N \times C_h \times T}$, while simultaneously learning a spatial embedding $\mathbf{S}_{emb} \in \mathbb{R}^{N \times C_s}$ and two temporal embeddings: daily temporal embedding $\mathbf{T}^d_{emb} \in \mathbb{R}^{T_d \times C_d}$ and weekly temporal embedding $\mathbf{T}^w_{emb} \in \mathbb{R}^{T_w \times C_w}$ to represent spatial, daily, and weekly correlations, respectively. Here, C_h is the dimension of data embedding. T_d denotes the timestamps in a day, and T_w is equal to 7, representing the length of days in a week. Besides, we use the same hidden dimension d for the spatial, daily, and weekly embeddings, i.e., $C_s = C_d = C_w = d$. Through broadcast and concatenation operations, the spatial and temporal identity information of each node and timestamp are incorporated into various nodes, which is shown in Fig. 3 and can be expressed as:

$$\mathbf{X}_H = \mathbf{X}_{emb} \oplus \mathbf{S}_{emb} \oplus \mathbf{T}^d_{emb} \oplus \mathbf{T}^w_{emb} \in \mathbb{R}^{N \times (C_h + 3d) \times T}. \tag{2}$$

C. Multi-Graph Convolutional Network

Although the spatio-temporal embedding layer effectively captures initial spatio-temporal correlations, the extraction of complex spatio-temporal correlations necessitates the use of graph convolutional network modules to further improve prediction accuracy. However, graph convolutional networks generally need to construct graph matrices based on nodes' topology information, which might cause performance degradation when accurate topology information is absent or even fail when losing distance information about several nodes. To mitigate this problem and enhance the stability and generalization ability of graph convolutional networks, the multi-graph convolutional network aggregates various types of graphs, including neighbor, similarity, and dynamic graphs, to extract correlation from different aspects, such as sequence connectivity and similarity. The weighted adjacency matrix $\mathbf{A}_{adj}$ reflects the proximity of nodes in the light of their distances and connectivity. Similarity graphs are built as undirected graphs based on traffic features and do not require prior knowledge of topological information. Two commonly used similarity graphs are constructed using Pearson correlations and Dynamic Time Warping (DTW) algorithms [32], [34], denoted as $\mathbf{A}_{cor}$ and $\mathbf{A}_{dtw}$, respectively. In addition, dynamic graph $\mathbf{A}_{adp}$ does not require node features or prior topological information at all, which is learned using two separate node embeddings to capture latent spatio-temporal correlations [35]. Since nodes with weaker correlations would affect the capture of spatio-temporal correlations, the graph matrix that filters out smaller values enhances the model's accuracy while reducing system complexity. Based on the filtered graph matrices, the multi-graph convolutional network module can be formed as:

$$\mathbf{h}^{l+1} = MGCN\left(\mathbf{h}^l\right) + \mathbf{h}^l, \tag{3}$$

$$\begin{aligned} MGCN\left(\mathbf{h}^l\right) = \sum_{k=0}^{K} & GCN(\tilde{\mathbf{A}}^k_{adj}, \mathbf{h}^l) + GCN\left(\tilde{\mathbf{A}}^k_{cor}, \mathbf{h}^l\right) \\ & + GCN\left(\tilde{\mathbf{A}}^k_{dtw}, \mathbf{h}^l\right) + GCN\left(\mathbf{A}^k_{adp}, \mathbf{h}^l\right), \end{aligned} \tag{4}$$

$$GCN\left(\mathbf{A}, \mathbf{X}\right) = \begin{cases} \mathbf{D}_A^{-1}\mathbf{A}\mathbf{X}\mathbf{W}_1 + \mathbf{D}_{A^T}^{-1}\mathbf{A}^T\mathbf{X}\mathbf{W}_2, & \text{if } \mathbf{A} = \tilde{\mathbf{A}}_{adj} \\ \mathbf{A}\mathbf{X}\mathbf{W}, & \text{if } \mathbf{A} = \mathbf{A}_{adp} \\ \mathbf{D}^{-1}\mathbf{A}\mathbf{X}\mathbf{W}, & \text{otherwise}, \end{cases} \tag{5}$$

where $\tilde{\mathbf{A}}_{adj}$ and $\tilde{\mathbf{A}}_{dtw}$ are constructed by selecting the smallest distance values for each node in corresponding matrices with a proportion of α, while $\tilde{\mathbf{A}}_{cor}$ chooses the related nodes with largest Pearson correlation coefficient for each node in the correlation matrix with the same ratio. On the basis of filtered matrices, we utilize dual diffusion graph convolution for the weighted adjacency matrix to model the forward and backward transition process. Meanwhile, forward diffusion graph convolution is used to model the undirected graph $\tilde{\mathbf{A}}_{cor}$ and $\tilde{\mathbf{A}}_{dtw}$. $\mathbf{h}^0$ is equal to $\mathbf{X}_H$ while $\mathbf{h}^{l+1}$ denotes the output of the multi-graph convolution network module at layer l. The output dimension of all hidden layers is set to D for ease of computation. To avoid the expensive computational cost brought by spectral graph convolution operation, we also adopt matrix multiplication to obtain the spatio-temporal correlation information directly from the spatial domain. Therefore, the output $\mathbf{h}^{l+1}$ can capture the latent spatio-temporal correlations by adding the convolutional results from multiple graphs.

D. Framework Architecture

Fig. 3 depicts the architecture of the SIMF framework, which is composed of a spatio-temporal embedding layer, stacked multi-graph convolutional network modules, and an output layer consisting of two fully connected layers. Since iterated multi-step prediction would suffer from the error accumulation problem [36], efficient direct multi-step prediction methods are preferred. In light of the well-captured spatio-temporal correlations, a linear layer structure is sufficient for achieving accurately and efficiently traffic prediction. Therefore, we adopt parallel fully-connected layers for the output layer to predict output with a length of T', which can effectively obtain great performance with the simplest structure. Initially, we transform the hidden layer representation $\mathbf{h}^L \in \mathbb{R}^{N\times D\times T}$ to a shape of $\mathbf{h}^O \in \mathbb{R}^{N\times DT}$. Subsequently, we use T' two-layer fully-connected layers to predict the traffic flow, with each having an output dimensionality of 1, expressed as:

$$\hat{\mathbf{Y}}^{(t)} = \mathrm{ReLU}\left(\mathbf{h}^O\mathbf{W}_1 + \mathbf{b}_1\right)\mathbf{W}_2 + \mathbf{b}_2, \tag{6}$$

where $\mathbf{W}_1 \in \mathbb{R}^{NT\times 2D}$ and $\mathbf{W}_2 \in \mathbb{R}^{2D\times 1}$ are the weights of two fully-connected layers, respectively. By concatenating T' one-step predicted traffic value, we can obtain the final prediction result $\hat{\mathbf{Y}} \in \mathbb{R}^{N\times T'}$.

IV. Experiments

To validate the suitability and superiority of our proposed framework, we begin by presenting the datasets and baseline methods used for comparison. We then describe experiment settings and provide a detailed analysis of the experimental results. Besides, ablation experiments are carried out to illustrate the effect of different components in the framework.

A. Datasets

We conduct a series of comparison experiments on three real-world datasets covering road traffic and taxi trajectory. Although these datasets are not collected from the metaverse, they are theoretically consistent with the datasets collected in the metaverse, as the virtual world synchronizes with the physical world in real time. The road traffic dataset is collected from the California state highway system and includes PeMS08 and PeMS-BAY [31], [37], with traffic sampled every 5 minutes. The SZ-taxi trajectory dataset is aggregated every 15 minutes in Shenzhen [38]. More statistical details are summarized in Table I.

TABLE I
Description and Statistics of Datasets

Datasets	#Nodes	Sampling Interval	Time range
PeMS08	170	5 mins	2016/7/1-2016/8/31
PeMS-BAY	325	5 mins	2017/1/1-2017/6/30
SZ-Taxi	156	15 mins	2015/1/1-2015/1/31

B. Baseline Methods

We compare our scheme with the commonly used baseline methods in the field of traffic forecasting, including but not limited to traditional statistical models, machine learning models, and graph neural network-based models. The specific details of baseline methods are described as follows:

- ARIMA [39]: Auto-Regressive Integrated Moving Average Model is a classic statistical method for modeling non-stationarity data.
- VAR [40]: Vector Auto-Regressive model can model multivariate time series by capturing their pairwise correlation.
- FC-LSTM [27]: A special type of RNN model that uses a fully-connected peephole LSTM.
- DCRNN [29]: Diffusion Convolutional Recurrent Neural Network integrates graph convolutional network into a GRU-based encoder-decoder architecture model.
- STGCN [30]: A spatio-temporal graph neural network that combines graph convolution and one-dimensional temporal convolution.
- Graph WaveNet [35]: An adaptive model that combines graph convolution and gated graph convolution.
- GMAN [41]: A graph multi-attention network with an encoder-decoder structure.
- STFGNN [32]: Spatial-Temporal Fusion Graph Neural Network fuses multiple graphs to capture spatio-temporal correlation simultaneously.
- STID [33]: Spaital-Temporal Identity method distinguishes spatial-temporal dependencies using spatio-temporal identity information.

C. Experiment Settings

To ensure fairness in the comparison, we partition the datasets into training, validation, and test sets using the same ratio as previous works. Specifically, for the PeMS-BAY dataset, the ratio is 7:1:2, while for the other datasets, it is 6:2:2. Besides, the input and output lengths are 12 for the PeMS datasets, whereas for the SZ-Taxi dataset, they are both 4. Following comparative experiments, we set the hidden dimension of data C_h and spatio-temporal embedding d to 64 and 32, respectively. Meanwhile, we stack two multi-graph convolutional network modules and employ two convolution kernels for each layer, *i.e.*, $L = 2$ and $K = 2$, and the hidden dimension of graph convolutional network D equals 64, which achieve the best performance on the validation sets. The loss function used in the experiments is the Mean Absolute Error (MAE), and the model is trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. To ensure a similar number of elements with the adjacency matrix, the

TABLE II
PERFORMANCE COMPARISON OF THE PROPOSED FRAMEWORK AND BASELINE METHODS ON PEMS-BAY, PEMS-08, AND SZ-TAXI DATASETS

	T	Metric	ARIMA	VAR	FC-LSTM	DCRNN	STGCN	GWNet	GMAN	STFGNN	STID	SIMF
PeMS-BAY	15 min	MAE	1.63	1.74	1.46	1.32	1.36	1.29	1.36	1.36	1.32	**1.29**
		RMSE	3.31	3.09	3.11	2.78	2.93	2.72	2.93	2.88	2.79	**2.68**
		MAPE	3.38	3.59	3.1	2.76	2.88	2.69	2.8	2.85	2.76	**2.69**
	30 min	MAE	2.25	2.33	1.94	1.68	1.83	1.61	1.65	1.71	1.63	**1.61**
		RMSE	4.75	4.15	4.35	3.81	4.29	3.67	3.8	3.85	3.69	**3.56**
		MAPE	5.07	5.02	4.42	3.78	4.2	**3.61**	3.62	3.8	3.68	3.65
	1 hour	MAE	3.19	2.92	2.52	2.03	2.5	1.91	1.91	2.04	1.92	**1.89**
		RMSE	6.47	5.11	5.66	4.7	5.92	4.44	4.44	4.61	4.37	**4.23**
		MAPE	7.69	6.46	6.22	4.79	5.98	4.51	**4.39**	4.77	4.5	4.5
PeMS08	15 min	MAE	18.4	16.87	15.62	14.14	15.35	13.57	13.31	15.04	13.26	**12.73**
		RMSE	27.69	25.19	24.4	22.35	23.79	21.67	22.65	23.32	21.5	**21.38**
		MAPE	11.82	10.84	9.81	9.2	9.95	8.64	8.76	10.14	8.62	**8.46**
	30 min	MAE	23.67	19.82	17.57	15.2	17.84	14.56	13.93	16.35	14.21	**13.78**
		RMSE	35.04	29.31	27.72	24.6	27.59	23.51	24.01	25.63	23.53	**23.37**
		MAPE	15.73	12.87	10.99	9.88	11.24	9.26	9.17	10.57	9.28	**9.16**
	1 hour	MAE	34.75	24.22	21.46	17.13	22.5	16.03	15.13	19.05	15.59	**15.06**
		RMSE	49.2	35.13	33.42	27.76	34.24	25.82	25.93	29.47	25.91	**25.65**
		MAPE	24.16	16.57	13.88	11.17	13.78	10.29	10.15	12.68	10.34	**10.12**
SZ-Taxi	15 min	MAE	3.24	3.64	3.44	3.30	3.21	3.22	3.25	3.20	3.22	**3.18**
		RMSE	4.54	4.88	4.87	4.67	4.53	4.54	4.56	4.53	4.55	**4.49**
		R^2	0.816	0.789	0.806	0.828	0.838	0.839	0.834	0.833	0.834	**0.842**
	30 min	MAE	3.31	3.56	3.52	3.33	3.26	3.26	3.27	3.28	3.27	**3.21**
		RMSE	4.63	4.88	4.98	4.70	4.60	4.59	4.60	4.63	4.61	**4.53**
		R^2	0.801	0.796	0.801	0.825	0.834	0.835	0.832	0.827	0.831	**0.837**
	1 hour	MAE	3.40	3.42	3.63	3.35	3.31	3.31	3.31	3.35	3.31	**3.24**
		RMSE	4.73	4.73	5.12	4.72	4.66	4.66	4.65	4.70	4.67	**4.58**
		R^2	0.801	0.801	0.782	0.822	0.830	0.831	0.828	0.820	0.830	**0.831**

sparsity α of the graph matrices is set to 0.02 for the PeMS08 and SZ-Taxi datasets and 0.03 for the PeMS-BAY dataset.

We use Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) to measure the performance on the PeMS-BAY and PeMS08 datasets and replace the MAPE metric with R^2 score for the SZ-Taxi dataset. The evaluation metrics are represented as $\mathrm{MAE} = \frac{1}{NT'}\left\|\mathbf{Y} - \hat{\mathbf{Y}}\right\|_1$, $\mathrm{RMSE} = \frac{1}{\sqrt{NT'}}\left\|\mathbf{Y} - \hat{\mathbf{Y}}\right\|_2$, $\mathrm{MAPE} = \frac{1}{NT'}\left\|\frac{\mathbf{Y}-\hat{\mathbf{Y}}}{\mathbf{Y}}\right\|_1$, and $R^2 = \frac{1}{T'}\sum_{i=1}^{T'} 1 - \frac{\left\|\mathbf{Y}_{:,i}-\hat{\mathbf{Y}}_{:,i}\right\|_2^2}{\left\|\mathbf{Y}_{:,i}-\bar{\mathbf{Y}}_i\right\|_2^2}$, where $\bar{\mathbf{Y}} \in \mathbb{R}^{T'}$ represents the average value of the ground truth for each prediction step, $\|\cdot\|_1$ and $\|\cdot\|_2$ represent the L1-norm and L2-norm of a matrix or a vector. All experiments are conducted on a Linux server with an Intel(R) Xeon(R) Gold 6132 CPU @ 2.60GHz and two NVIDIA Quadro RTX 5000 GPUs.

D. Experimental Results

Table II presents a comparison of the performance of the proposed SIMF framework with the baseline methods on PeMS-BAY, PeMS08, and SZ-Taxi datasets for predicting the next 15, 30, and 60-minute traffic flow. The results show that our proposed framework generally outperforms the other baselines on most metrics, except for the MAPE on the PeMS-BAY dataset, which is worse than that of the Graph WaveNet and GMAN in the 30-minute and 1-hour range. The traditional statistical methods, ARIMA and VAR, fail to capture complex spatio-temporal correlation, resulting in poor performance, especially for long-term traffic prediction. FC-LSTM performs better than traditional statistical methods with its recursive structure, but they still do not fully consider the impact of spatial correlation. In contrast, graph-based methods generally yield better performance by taking spatio-temporal correlations into consideration, including DCRNN, STGCN, GWNet, GMAN, STFGNN, and our proposed method. Compared to DCRNN and STGCN, our framework incorporates spatio-temporal embedding representation and simultaneously predicts multiple time steps based on fully connected layers, which mitigates the error accumulation problem in long-term sequence prediction. Additionally, SIMF outperforms Graph WaveNet and STFGNN by incorporating additional correlation and DTW graph matrices and utilizing the spatio-temporal dependencies from multiple graphs instead of using a unified spatio-temporal fusion graph, leading to improved accuracy while maintaining lower complexity. Although STID does not contain any graph structure, its learned spatio-temporal identity information has a similar effect to graph neural networks by reducing the indistinguishability of spatial and temporal information. Our framework further enhances the capture of spatio-temporal correlation through a multi-graph convolutional network module on top of a spatio-temporal embedding layer, thereby achieving higher accuracy than STID based on spatio-temporal identity information and GMAN based on an attention mechanism.

To verify the effectiveness of the proposed framework, we compare the training time of the SIMF framework with other baselines. Table III demonstrates that the SIMF framework has a relatively low training time compared to other baselines, only slightly more than that of the STGCN and STID methods. Both FCLSTM and DCRNN contain the recurrent neural network structure and have to generate prediction results step by step, thus consuming more training time. STGCN and Graph WaveNet methods are more effective than FCLSTM and DCRNN based on complete convolutional structures. However, the performance of the STGCN model rapidly

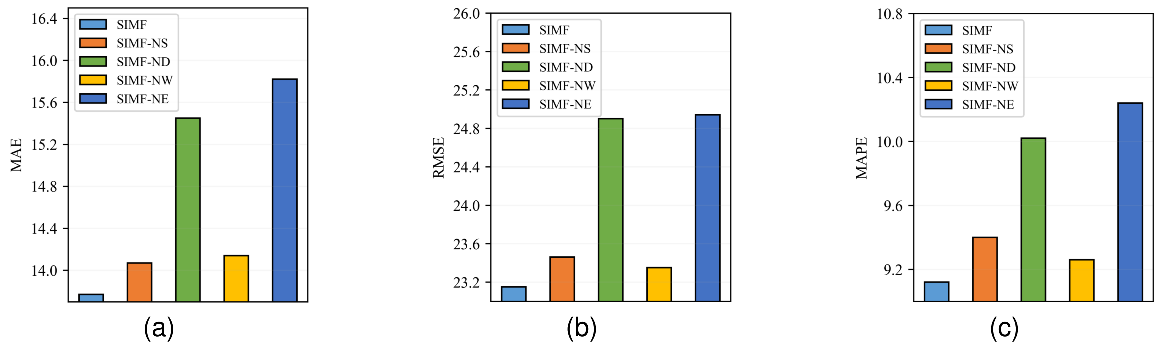

Fig. 4. Comparison of the average performance metrics of SIMF and its variants with different spatio-temporal embedding layers on the PEMS08 dataset (a) MAE (b) RMSE (c) MAPE.

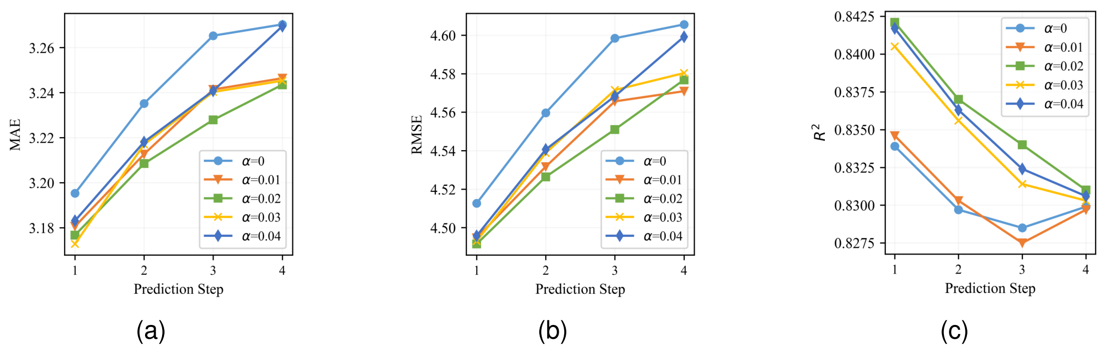

Fig. 5. The effect of graph matrix sparsity α for traffic prediction on the SZ-Taxi dataset (a) MAE (b) RMSE (c) R^2.

TABLE III
THE COMPUTATIONAL TIME ON THE DIFFERENT DATASETS

Methods	Training time(s/epoch)		
	PeMS-BAY	PeMS08	SZ-Taxi
FCLSTM	454.2	71.98	3.65
DCRNN	389.55	54.62	2.6
STGCN	52.33	8.76	1.47
GWNet	127.47	36.91	2.72
GMAN	388.88	47.69	3.16
STFGNN	311.68	41.75	1.64
STID	16.58	3.46	0.7
SIMF	113.72	23.63	1.34

deteriorates with the increase of prediction horizons because of the characteristic of single-step prediction, while the Graph WaveNet model is inferior to our model in both accuracy and efficiency. Moreover, GMAN and STFGNN methods require a large amount of training time due to the influence of the spatio-temporal attention mechanism and spatio-temporal fused graph, and their training time grows faster as the number of nodes and input length increase. In contrast, our approach utilizes the matrix multiplication for the multi-graph convolutional network module and predicts multi-step traffic based on parallel fully-connected layers, achieving a remarkable performance with only 113.72s, 23.63s, and 1.34s training time for each epoch on the PeMS-BAY, PeMS08, and SZ-Taxi datasets, respectively.

E. Ablation Experiments

SIMF captures potential spatio-temporal correlations through its spatio-temporal embedding layer and the multi-graph convolutional network. In order to illustrate the effects of these modules, we have adopted different variants of SIMF.

Specifically, We first provide four variants of the spatio-temporal embedding layer, named SIMF-NS, SIMF-ND, SIMF-NW, and SIMF-NE, corresponding to four distinct situations, including no spatial embedding, no daily temporal embedding, no weekly temporal embedding, and no spatio-temporal embedding layer. Fig. 4 depicts the performance metrics of different variants on the PeMS08 dataset with a 1-hour prediction horizon. We observe that SIMF achieves superior performance compared to other variants due to its combination of spatial embedding, daily and weekly temporal embedding. It is worth noting that the effect of the daily temporal embedding is greater than that of the spatial embedding, which could be attributed to the multi-graph convolutional network's partial compensation for the spatial embedding's function. In addition, since some weekly information is already captured by the daily signal, spatial embedding has a greater effect on prediction accuracy compared to weekly temporal embedding.

In Fig. 5, we compare the performance of SIMF on the SZ-Taxi dataset with different levels of graph matrix sparsity α. Here, a value of $\alpha = 0$ indicates that the graph matrix is replaced by the identity matrix. We conclude that SIMF performs optimally when the sparsity of $\tilde{\mathbf{A}}_{cor}$ and $\tilde{\mathbf{A}}_{dtw}$ is similar to that of the adjacency matrix, which occurs at $\alpha = 0.2$. This can be attributed to the fact that a lower sparsity coefficient may not capture sufficient spatio-temporal correlations. In comparison, a higher sparsity coefficient may account for the influence of non-correlated nodes, which in turn deteriorates prediction accuracy.

Furthermore, we test SIMF on the PEMS-BAY dataset using various combinations of graphs to examine the impact of multiple graphs, including SIMF with the absence of the

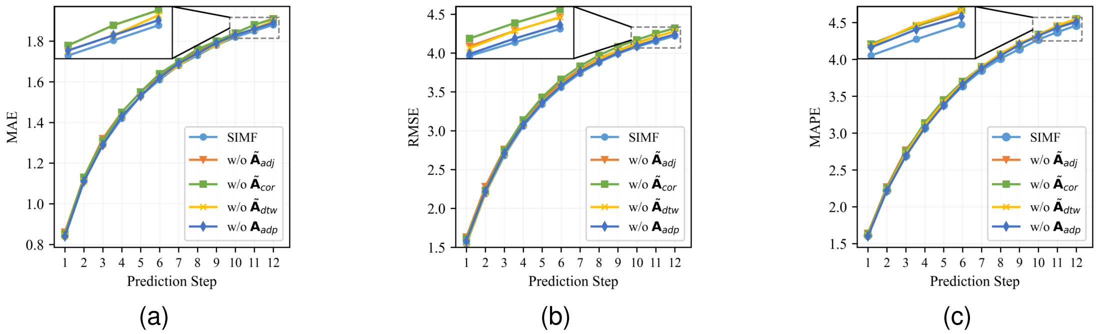

Fig. 6. The effect of different graphs in the multi-graph convolutional networks module for traffic prediction on the PeMS-BAY dataset (a) MAE (b) RMSE (c) MAPE.

adjacency matrix $\tilde{\mathbf{A}}_{adj}$, Pearson correlation matrix $\tilde{\mathbf{A}}_{cor}$, the dtw matrix $\tilde{\mathbf{A}}_{dtw}$, and the adaptive matrix $\mathbf{A}_{adp}$. It can be seen from Fig. 6 that the proposed framework achieves the best performance by leveraging the unique properties of all graph matrices, particularly for long-term prediction. Specifically, $\tilde{\mathbf{A}}_{adj}$ can aggregate the nodes with strong spatial correlation while $\tilde{\mathbf{A}}_{cor}$ and $\tilde{\mathbf{A}}_{dtw}$ can figure out nodes with the highest temporal correlation with the target nodes in time-aligned and time-unaligned manners, respectively. As a supplement to other graph matrices, the dynamic matrix $\mathbf{A}_{adp}$ can learn latent spatio-temporal correlations and strengthen the weights of strongly correlated nodes, further improving the prediction accuracy. In addition, the adjacency matrix $\tilde{\mathbf{A}}_{adj}$ and Pearson correlation matrix $\tilde{\mathbf{A}}_{cor}$ play a more crucial role in capturing spatio-temporal correlations in the multi-graph convolutional network module. Due to the existing graph matrices having well-constructed spatio-temporal correlations, the dynamic matrix $\mathbf{A}_{adp}$ only considers a few spatio-temporal correlated nodes and has less improvement in accuracy compared with other graph matrices. Moreover, as the combination of multiple matrices can jointly capture spatio-temporal correlations, it is feasible to predict traffic without graph topology information or node features. In other words, SIMF can still achieve high accuracy even in the absence of the adjacency matrix, contributing to the construction of spatio-temporal correlations in the metaverse.

V. Conclusion and Future Work

In this paper, we proposed a novel Spatial-temporal Identity Multi-graph convolutional network Framework for spatio-temporal traffic prediction in the metaverse. Our framework begins by introducing the architecture of the metaverse and representing avatars and their relationships as a graph topology within it. We then leverage available identity and timestamp features to extract spatio-temporal identity information so as to capture spatio-temporal correlations. To further extract latent spatio-temporal correlations, multiple graphs are introduced and fused in the multi-graph convolutional network. Experimental results demonstrate that our framework can establish connections among nodes by fusing multiple predefined and dynamic graph matrices, keeping low time complexity and improving prediction accuracy even without prior topology information compared to other baselines. Therefore, our framework can be easily integrated into the metaverse, where most avatars have spatio-temporal dependencies due to the metaverse's hyper-spatiotemporality and interoperability characteristics.

As more and more avatars will access the metaverse for various types of applications, separate traffic prediction for each sub-metaverse will not be sufficient to meet the future stringent performance requirements for immersive access. Hence, it is necessary to improve the proposed framework in several technological aspects to simultaneously satisfy the needs of users and VPS. Specifically, spatio-temporal traffic prediction can be applied not only in the metaverse application layer to assist functions such as the digital twin, but also in the infrastructure layer to optimize resource allocation efficiency by predicting actual resource demands for communication, networking, and storage. Meanwhile, the fusion of other state-of-the-art models that have achieved remarkable results for multi-step time series prediction can further improve the prediction accuracy, such as the Transformer. Besides, the metaverse should pay more attention to the security and privacy of users with the aid of quantum networks. For example, advanced technologies like in-network computation are capable of enhancing user privacy protection by combining with quantum networks for efficient entanglement generation and distribution. In view of the heterogeneity and interoperability characteristics of different sub-metaverse worlds, heterogeneous graph-based models will contribute to higher prediction accuracy and efficiency by integrating heterogeneous individuals from different sub-metaverse worlds into a graph and extracting their associations in a unified way. Furthermore, the integration with edge intelligence networks fulfills the scalability requirements of the metaverse, such as a traffic prediction framework based on federated learning in the metaverse. Finally, it is essential to construct metaverse application-level datasets through advanced software and hardware platforms, e.g., Omniverse and Meta Quest, which significantly improve the scope and performance of applications without jeopardizing the physical world.

References

[1] N. Stephenson, *Snow Crash*. New York, NY, USA: Bantam Books, 1992.
[2] J. Yu, A. Alhilal, P. Hui, and D. H. K. Tsang, "6G mobile-edge empowered metaverse: Requirements, technologies, challenges and research directions," 2022, *arXiv:2211.04854*.

[3] C. T. Nguyen, D. T. Hoang, D. N. Nguyen, and E. Dutkiewicz, "MetaChain: A novel blockchain-based framework for metaverse applications," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2022, pp. 1–5.

[4] Y. Han, D. Niyato, C. Leung, C. Miao, and D. I. Kim, "A dynamic resource allocation framework for synchronizing metaverse with IoT service and data," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 1196–1201.

[5] M. Xu, D. Niyato, J. Kang, Z. Xiong, C. Miao, and D. I. Kim, "Wireless edge-empowered metaverse: A learning-based incentive mechanism for virtual reality," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 5220–5225.

[6] W. C. Ng, W. Y. B. Lim, J. S. Ng, Z. Xiong, D. Niyato, and C. Miao, "Unified resource allocation framework for the edge intelligence-enabled metaverse," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 5214–5219.

[7] Y. Wang et al., "A survey on metaverse: Fundamentals, security, and privacy," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 319–352, 1st Quart., 2023.

[8] M. Xu et al., "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 656–700, 1st Quart., 2023.

[9] L.-H. Lee et al., "All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda," 2021, *arXiv:2110.05352*.

[10] W. Y. B. Lim et al., "Realizing the metaverse with edge intelligence: A match made in heaven," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 64–71, Aug. 2023.

[11] S. B. Far and A. I. Rad, "Applying digital twins in metaverse: User interface, security and privacy challenges," *J. Metaverse*, vol. 2, no. 1, pp. 8–16, Jun. 2022.

[12] R. Cheng, N. Wu, S. Chen, and B. Han, "Will metaverse be NextG Internet? Vision, hype, and reality," *IEEE Netw.*, vol. 36, no. 5, pp. 197–204, Sep. 2022.

[13] L.-H. Lee et al., "When creators meet the metaverse: A survey on computational arts," 2021, *arXiv:2111.13486*.

[14] H. Duan, J. Li, S. Fan, Z. Lin, X. Wu, and W. Cai, "Metaverse for social good: A university campus prototype," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 153–161.

[15] X. Zhou, Y. Li, and J. Zhao, "Resource allocation of federated learning assisted mobile augmented reality system in the metaverse," 2023, *arXiv:2301.12085*.

[16] L. U. Khan et al., "Machine learning for metaverse-enabled wireless systems: Vision, requirements, and challenges," 2022, *arXiv:2211.03703*.

[17] T. Huynh-The, Q.-V. Pham, X.-Q. Pham, T. T. Nguyen, Z. Han, and D.-S. Kim, "Artificial intelligence for the metaverse: A survey," *Eng. Appl. Artif. Intell.*, vol. 117, Jan. 2023, Art. no. 105581.

[18] J. Sun, W. Gan, Z. Chen, J. Li, and P. S. Yu, "Big data meets metaverse: A survey," 2022, *arXiv:2210.16282*.

[19] T. R. Gadekallu et al., "Blockchain for the metaverse: A review," 2022, *arXiv:2203.09738*.

[20] Q. Yang, Y. Zhao, H. Huang, Z. Xiong, J. Kang, and Z. Zheng, "Fusing blockchain and AI with metaverse: A survey," *IEEE Open J. Comput. Soc.*, vol. 3, pp. 122–136, 2022.

[21] T. Braud, L.-H. Lee, A. Alhilal, C. B. Fernández, and P. Hui, "DiOS—An extended reality operating system for the metaverse," *IEEE MultimediaMag.*, vol. 30, no. 2, pp. 70–80, Apr./Jun. 2023.

[22] D. A. Tedjopurnomo, Z. Bao, B. Zheng, F. M. Choudhury, and A. K. Qin, "A survey on modern deep neural network for traffic prediction: Trends, methods and challenges," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 4, pp. 1544–1561, Apr. 2022.

[23] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1389–1401, Jun. 2019.

[24] Y. Xu, F. Yin, W. Xu, J. Lin, and S. Cui, "Wireless traffic prediction with scalable Gaussian process: Framework, algorithms, and verification," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1291–1306, Jun. 2019.

[25] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A comparison of ARIMA and LSTM in forecasting time series," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 1394–1401.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[27] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[28] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[29] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*.

[30] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[31] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial–temporal synchronous graph convolutional networks: A new framework for spatial–temporal network data forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 914–921.

[32] M. Li and Z. Zhu, "Spatial–temporal fusion graph neural networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4189–4196.

[33] Z. Shao, Z. Zhang, F. Wang, W. Wei, and Y. Xu, "Spatial–temporal identity: A simple yet effective baseline for multivariate time series forecasting," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2022, pp. 4454–4458.

[34] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intell. Data Anal.*, vol. 11, no. 5, pp. 561–580, Oct. 2007.

[35] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial–temporal graph modeling," 2019, *arXiv:1906.00121*.

[36] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" 2022, *arXiv:2205.13504*.

[37] C. Chen, K. Petty, A. Skabardonis, P. Varaiya, and Z. Jia, "Freeway performance measurement system: Mining loop detector data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1748, no. 1, pp. 96–102, Jan. 2001.

[38] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[39] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transp. Res. Rec.*, no. 722, pp. 1–9, 1979.

[40] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 2020.

[41] C. Zheng, X. Fan, C. Wang, and J. Qi, "Gman: A graph multi-attention network for traffic prediction," in *Proc. Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 1234–1241.

Haihan Nan received the B.S. degree in communication engineering from Xidian University, Xi'an, China, in 2018, where he is currently pursuing the Ph.D. degree in communication and information system. He is also a joint-cultivated Ph.D. student with Kanazawa University, Japan. His research interests include big data analytics in cellular network and traffic prediction.

Ruidong Li (Senior Member, IEEE) received the B.S. degree from the Department of Information Science and Electronic Engineering, Zhejiang University, China, in 2001, and the M.S. and Ph.D. degrees from the Department of Computer Science, University of Tsukuba, Japan, in 2003 and 2008, respectively. His research interests include quantum networks, metaverse, future networks, autonomous aerial vehicles, mobile computing, the Internet of Things, cloud computing, game theory, learning (artificial intelligence), and quantum networks. He was a recipient of the Best Paper Awards for IEEE ICC 2022 and IWCMC 2022. He serves as the Vice Chair for the IEEE Internet Technical Committee (ITC) and served as the Chair for several conferences, such as the General Co-Chair for IEEE MSN 2021 and CPSCom 2021, and the TPC Co-Chair for IWQoS 2021 and IEEE MSN 2020.

Xiaoyan Zhu (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Xidian University, Xi'an, China, in 2000, 2004, and 2009, respectively. She is currently a Professor with the School of Telecommunications Engineering, Xidian University. Her research interests include security and privacy for wireless networks, cloud computing, and big data.

Jianfeng Ma (Member, IEEE) received the B.S. degree in mathematics from Shaanxi Normal University, Xi'an, China, in 1985, and the M.S. and Ph.D. degrees in computer software and telecommunication engineering from Xidian University, Xian, in 1988 and 1995, respectively. From 1999 to 2001, he was a Research Fellow with the Nanyang Technological University of Singapore. He is currently a Professor and a Ph.D. Supervisor with the Department of Computer Science and Technology, Xidian University. He is also the Director of the Shaanxi Key Laboratory of Network and System Security. His current research interests include information and network security, wireless and mobile computing systems, and computer networks.

Kaiping Xue (Senior Member, IEEE) received the bachelor's degree from the Department of Information Security, University of Science and Technology of China (USTC), in 2003, and the Ph.D. degree from the Department of Electronic Engineering and Information Science (EEIS), USTC, in 2007. From May 2012 to May 2013, he was a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, University of Florida. He is currently a Professor with the School of Cyber Science and Technology, USTC. His research interests include next-generation internet architecture design, transmission optimization, and network security. He is a fellow of IET. He serves on the Editorial Board for several journals, including IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT. He served as the (Lead) Guest Editor for many reputed journals/magazines, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, *IEEE Communications Magazine*, and IEEE NETWORK.