# Cross- and Context-Aware Attention Based Spatial-Temporal Graph Convolutional Networks for Human Mobility Prediction

ZHAOBIN MO, Department of Civil Engineering and Engineering Mechanics, Columbia University, USA
HAOTIAN XIANG, Department of Electrical Engineering, Columbia University, USA
CHENG-CHUN CHANG, Department of Electrical Engineering, Columbia University, USA
XUAN DI*†, Department of Civil Engineering and Engineering Mechanics, Columbia University, USA

The COVID-19 pandemic has dramatically transformed human mobility patterns. Therefore, human mobility prediction for the "new normal" is crucial to infrastructure redesign, emergency management, and urban planning post the pandemic. This paper aims to predict people's number of visits to various locations in New York City using COVID and mobility data in the past two years. To quantitatively model the impact of COVID cases on human mobility patterns and predict mobility patterns across the pandemic period, this paper develops a model CCAAT-GCN (**C**ross- and **C**ontext-**A**ttention based Spatial-Temporal **G**raph **C**onvolutional **N**etworks). The proposed model is validated using SafeGraph data in New York City from August 2020 to April 2022. A rich set of baselines are performed to demonstrate the performance of our proposed model. Results demonstrate the superior performance of our proposed method. Also, the attention matrix learned by our model exhibits a strong alignment with the COVID-19 situation and the points of interest within the geographic region. This alignment suggests that the model effectively captures the intricate relationships between COVID-19 case rates and human mobility patterns. The developed model and findings can offer insights into the mobility pattern prediction for future disruptive events and pandemics, so as to assist with emergency preparedness for planners, decision-makers and policymakers.

CCS Concepts: • **Computing methodologies** → **Neural networks**.

Additional Key Words and Phrases: Human Mobility Prediction, COVID-19, Cross-attention, Context-aware Attention, Graph Neural Network, Graph Convolution.

**ACM Reference Format:**
Zhaobin Mo, Haotian Xiang, Cheng-Chun Chang, and Xuan Di. 2023. Cross- and Context-Aware Attention Based Spatial-Temporal Graph Convolutional Networks for Human Mobility Prediction . *J. ACM* 37, 4, Article 111 (August 2023), 21 pages. https://doi.org/XXXXXXX.XXXXXXX

*Corresponding Author
†This author is also affiliated with Data Science Institute, Columbia University, USA

Authors' addresses: Zhaobin Mo, Department of Civil Engineering and Engineering Mechanics, Columbia University, 500 W. 120th St., Mudd 628A, New Yrok, NY, USA, zm2302@columbia.edu; Haotian Xiang, Department of Electrical Engineering, Columbia University, 500 W. 120th St., Mudd 1310, New Yrok, NY, USA, hx2341@columbia.edu; Cheng-Chun Chang, Department of Electrical Engineering, Columbia University, 500 W. 120th St., Mudd 1310, New Yrok, NY, USA, cc4900@columbia.edu; Xuan Di, Department of Civil Engineering and Engineering Mechanics, Columbia University, 500 W. 120th St., Mudd 630, New Yrok, NY, USA, sharon.di@columbia.edu.


0004-5411/2023/8-ART111 $15.00
https://doi.org/XXXXXXX.XXXXXXX

## 1 INTRODUCTION

The COVID-19 pandemic has dramatically transformed human mobility patterns, including the type of visited locations, check-in time of locations, and preference over origin-destination distances [16]. Such a trend consequently induces a shift in travel mode choice, like the rising trend in telecommuting [18] and constantly lower subway ridership [33]. Therefore, human mobility prediction for the “new normal" is crucial to infrastructure redesign, emergency management, and urban planning post the pandemic.

How do we predict a nonstationary spatiotemporal pattern, given that the “new" normal demonstrates a quite different pattern from the “old" normal? To tackle such a challenge, we need to rely on nonstationary features such as COVID cases. This paper aims to predict people’s number of visits to various locations in New York City using various data in the past two years. The developed model and findings can offer insights into the mobility pattern prediction for future disruptive events and pandemics, which will in turn assist with emergency preparedness for transportation planners, decision-makers and policymakers.

Some studies on COVID-19 focus on predicting the evolution of the pandemic without accounting for the underlying mobility patterns [4, 30, 35, 38, 39]. However, pandemic evolution and human mobility are highly correlated. A majority of studies have examined the impact of human mobility on the pandemic case number. Reversely, COVID cases also affect people’s travel desire, thus impacting overall visitation frequencies to various places. Statistical analysis accompanied by data visualization [40] demonstrates strong evidence for the impact of COVID cases on mobility. For example, the impact of COVID-19 on human mobility patterns is analyzed in New York City using statistical analysis and spatial visualization [15]. Comparing the number of visits in 2019 and 2020, this study finds that there is a strong correlation between the number of visits and the trend of the newly reported COVID-19 cases. It finds that most locations have the lowest numbers of visits in the first half of April 2020, when COVID-19 explodes. [17] uses Twitter data to analyze the change of human mobility patterns during COVID-19 in New York City. A general decreasing trend is observed in the Twitter user activity after mid-March, while hospitals witnesses a significant increase of Twitter users after mid-March. [36] analyzes anonymized mobile phone data for six states in the United States and measures the individual daily traveling distance. It detects daily travel distance drops across all six states in March 2020, with New York falling from 5.2 km on March 2 to only 31 meters on March 23. [5] leverage users’ check-in data (i.e., geo-location check-in information) obtained from Twitter to study the trend of travel patterns post the pandemic in New York City. It finds that users’ gyration decreases by 35% after the stay-at-home order. [6] utilizes individual tweets and user demographics to study people’s attitudes toward different travel modes during the pandemic, shown to be consistent with the changes in the ridership of each travel mode.

The aforementioned studies primarily use statistical analysis to investigate the COVID-19 impact on human mobility. However, little research has been done to quantitatively model the influence of COVID-19 cases on human mobility patterns. A quantitative model is essential for simulating different scenarios and assessing the potential pandemic impact on human mobility patterns. This allows for the exploration of various hypothetical situations, enabling researchers to project potential pandemic challenges and develop strategies for similar situations in the future [3].

To quantitatively model the impact of COVID cases on human mobility patterns and predict mobility patterns in time and space, this paper develops a deep learning model, namely, **C**ross- and **C**ontext-**A**ware **At**tention based Spatial-Temporal **G**raph **C**onvolutional **N**etworks (CCAAT-GCN). Graph convolutional networks (GCN) capture the spatial evolution of the number of visits to each location. Attention mechanism, including temporal and spatial attention, aims to model the intricate relationships and patterns in the data. Temporal attention captures the temporal

dependencies and variations over time, while spatial attention captures the spatial interactions and dependencies among different locations. Building upon the GCN framework, the cross-attention module specifically models the correlation between COVID-19 cases and the number of visits, allowing for a comprehensive understanding of their mutual influence. Moreover, the context-attention module learns to incorporate relevant contextual features, such as regional demographics or socioeconomic factors, to enhance the prediction accuracy and interpretability of the model. The proposed model is validated using SafeGraph data[1] in New York City from August 2020 to April 2022.

The rest of this paper is organized as follows. Section 2 presents the related work and highlights our contributions. Section 3 provides the problem statement. Section 4 fleshes out the framework of our proposed CCAAT-GCN. Section 5 introduces the COVID-19 and mobility datasets. Section 6 details the experiments and presents the results. Section. 7 concludes our work and projects future research directions.

## 2 RELATED WORK

In this section, we first introduce the approaches developed for human mobility prediction, including the time-series methods, Markov-based methods, deep-learning based methods, and graph neural networks. Then, we will point out the limitations of existing studies and identify research gaps. The contribution of this paper will be highlighted thereafter.

### 2.1 Time-series methods.

The time-series method is a statistical technique that is commonly used to analyze and model data that is collected over a period of time. This method involves examining and interpreting the patterns and trends present in the data to make forecasts and predictions about future values. These methods include Autogressive (AR), Moving Average (MA), and Autogressive Integrated Moving Average (ARIMA) [24]. In [8], a multivariate nonlinear time series model is used to predict social interactions. In [22], an improved ARIMA-based method is proposed to predict the human mobility in the hotspots. The improved ARIMA combines ARIMA with a prior distribution of the passenger's locations, achieving better prediction accuracy than the original ARIMA. In [34], a time-series based method that uses Gibbs sampling is proposed to predict future human locations. [19] proposes to use a seasonal ARIMA model to predict human mobility.

### 2.2 Markov-based methods

Markov-based methods are a type of probabilistic model that predicts the future states of a system based on its current state. This kind of method assumes that the probability of moving from one state to another depends only on the current state and not on any of the previous states. [27] applies Markov predictors to predict the next location with extensive Wi-Fi mobility data. $O(0)$ Markov predictors are used, that is, this model simply returns the most frequently seen locations from historical trajectories. [25] applies the Hidden Markov Model (HMM) to predict the human mobility trajectory, with self-adaptive parameters that change according to the objects' moving speed. A multilevel Markov-based approach to predict the future location of people, the effectiveness of which is verified by geotagged tweets data. A hybrid Markov-based model is proposed in [26] to predict the next location, which considers the spatio-temporal similarity of human mobility patterns. [31] propose a hidden Markov model to extract travellers' activity patterns.

[1]https://www.safegraph.com/academics

## 2.3 Deep-learning based methods

In this section, we introduce the deep learning models that are used before the emergence of graph neural networks. These models have proven their ability to capture human mobility patterns. They also serve as the components of the graph neural networks to be covered in the next subsection.

**RNN**. Recurrent neural networks (RNN) have emerged as a powerful computational model capable of capturing temporal dependencies in sequential data, making them well-suited for forecasting human mobility patterns. Long short-term memory (LSTM) and gated recurrent unit (GRU) are widely used recurrent units. [23] proposes a two-layer LSTM network to predict traffic flow. In [10], multiple GRUs are stacked to capture long-range dependencies in mobility trajectories. [10] incorporates a learnable user embedding into the LSTM to consider users preferences while predicting human mobility.

**CNN**. Convolutional neural network (CNN) has been applied to a variety of applications such as image classification, etc. It is mainly used to capture the spatial correlations within different locations. Limited by the convolution operator, it can only be used for grid-distributed data. [44] divides the geolocations as grids so that CNN can be used to capture the spatial patterns for mobility prediction. [11] embeds the human trajectories into feature matrices, where the CNN can be used. Despite the limitation, several variants of CNN have been applied to human mobility prediction. A variant of CNN, Gated Temporal Convolutional Networks (Gated TCN) [2] is used to capture the temporal pattern. Gated TCN is a deep learning architecture that has been proposed for modeling sequential data with long-term dependencies. The model is based on the idea of dilated convolutions, which enables the network to effectively capture both short-term and long-term patterns in the data.

**Attention**. Attention mechanisms have become a popular technique in machine learning and natural language processing (NLP) in recent years. Attention mechanisms allow neural networks to selectively focus on parts of the input data that are most relevant to the task at hand. This selective focus is achieved by assigning weights to different parts of the input data, which are then used to compute a weighted sum of the input data. [10] combines the attention mechnism with recurrent networks for mobility prediction, where historical trajectories are handled by the attention mechanism to extract mobility patterns, and a GRU handles current trajectories. [12] proposes a variational attention model to predict human mobility. The variational encoding captures latent features of recent mobility, followed by an attention mechanism to learn the attention on the historical latent features. [9] proposes a decentralized attention-based human mobility prediction method, allowing more efficient training for personalized prediction. Those attention mechanisms only use the historical mobility data to predict its future values, and thus are also called self-attentions.

## 2.4 Graph Neural Networks

With the rapid development of deep learning, graph neural networks (GNN) have emerged as a powerful tool in modeling spatial and temporal patterns in human mobility. In this newly emerged domain, spatial-temporal graph neural network has shown its effectiveness in this task, and becomes the state-of-the-art genre of this method. We summarize those methods in Table 1. As most works explicitly split their models as spatial and temporal components, we follow the same manner and explain each component separately.

For the spatial components, most studies adopt the Graph Convolutional Network (GCN) [1, 4, 7, 13, 14, 20, 21, 35, 37, 39, 41, 42], which is a powerful framework for analyzing and processing graph-structured data. While CNN excels in grid-like data such as images, GCNs offer a specialized approach to capture and model complex relationships within graph data. GCNs leverage the

connectivity patterns of nodes in a graph to propagate information and extract meaningful features. By employing localized and adaptive filters, GCNs can effectively capture both local and global structural information from the graph. This makes GCNs well-suited for human mobility prediction. From Table 1, we can see that many studies combine GCN with self-attention to capture the spatial patterns [7, 14, 20, 35, 42]. When integrating self-attention mechanisms into GCNs for capturing spatial patterns in human mobility prediction, a common approach is to employ the Graph Attention Network (GAT) architecture [32]. In the GAT model, attention mechanisms are incorporated to assign importance weights to different nodes in the graph based on their relevance to the prediction task. This is achieved by computing attention scores that reflect the importance of each region's neighbourhood in relation to the central region. The attention scores are then used to weigh the feature representations of neighbouring regions during prediction.

For the temporal components, the most used method is the Gated TCN [4, 13, 20, 21, 37, 41]. By leveraging dilated convolutions and gate mechanisms, Gated TCNs can effectively capture both short-term and long-term temporal patterns in human mobility data. The dilated convolutions allow the network to process a wide range of temporal contexts, while the gate mechanisms enable the network to focus on relevant temporal features and disregard noise or irrelevant information. Similar to the spatial components, self-attention can also be incorporated into Gated TCN to capture the temporal mobility pattern [13, 20]. By applying self-attention after the temporal convolutions in the Gated TCN, the attention mechanism assigns attention weights to different temporal features, allowing the model to focus on relevant information and capture intricate temporal patterns. The attention weights are calculated based on the relationships between different time steps, enabling the model to assign higher weights to important time steps.

Table 1. GNN-based methods for human mobility prediction.

| Model | Spatial Component | Temporal Component |
|---|---|---|
| STGCN [41] | GCN | Gated TCN |
| MepoGNN [4] | GCN | Gated TCN |
| SAB-GNN [39] | GCN | LSTM |
| AGCRN [1] | GCN | GRU |
| STFGNN [21] | GCN | Gated TCN |
| HGCN [13] | GCN | self attention + Gated TCN |
| Graph WaveNet [37] | GCN | Gated TCN |
| HGARN [29] | self-attention | self-attention + LSTM |
| GCDAN [7] | self attention + GCN | self-attention |
| STAR [42] | self-attention + GCN | self-attention |
| CausalGNN [35] | self-attention + GCN | RNN |
| DSTAGNN [20] | self-attention + GCN | self-attention + Gated TCN |
| ASTGCN [14] | self-attention + GCN | self-attention |
| This paper | self-, cross-, and context-aware attention + GCN | self- and cross-attention + TCN |

## 2.5 Contributions of this paper

Those methods, however, have two main drawbacks:

- Lack of interpretability. Existing modeling methods, such as attention-based GNNs, prioritize prediction accuracy and lack interpretability. However, an interpretable model is crucial in

understanding the relationship between COVID-19 and mobility patterns while also making accurate predictions.

- Lack of contextual features. Most previous studies on modeling human mobility patterns during COVID-19 are autoregressive, in the sense that they only use historical mobility data to predict the future. These studies have neglected to incorporate contextual features, such as information about regional population and income, which may hold valuable insights that affect mobility patterns during the pandemic. However, as these static context features remain constant over time, integrating them into dynamic mobility data poses a non-trivial challenge. In an effort to address this problem, [44] concatenates the static context with the dynamic traffic feature and feeds the concatenated vector directly into the model. However, as the static context remains unchanged over time, this simple concatenation method can hinder the training process, because the model must learn to distinguish between the dynamic and static features.

Our proposed CCAAT-GCN integrates spatial and temporal information into a graph-based framework that captures the complex interdependencies between different regions and periods. Specifically, it uses the cross-attention mechanism to model the mutual influence between the COVID-19 pandemic and human mobility, where the calculated cross-attention scores serve to interpret this mutual influence. Additionally, we use the context-aware attention mechanism to better incorporate static information, such as regional income, population, and points of interest (POI) in predicting the mobility dynamic patterns. Furthermore, we ensemble multiple adjacency matrices together to better capture the spatial patterns. Those adjacency matrices include both the static ones that are calculated based on distance or inter-nodal correlations, together with the adaptive ones that are learned by our proposed model. We evaluate our approach on a large-scale mobility dataset, the SafeGraph dataset, during the COVID-19 pandemic.

The main contributions can be summarized as follows:

(1) We introduce a novel framework of CCAAT-GCN for mobility prediction considering COVID-19 impact, and use a real-world dataset for validation.
(2) We propose to use cross-attention mechanism to enhance model interpretability by explicitly modeling the mutual influence between COVID-19 and human mobility. Learning the interaction between these two critical factors can provide interpretable insights into the relationships between COVID-19 dynamics and mobility patterns, enabling a more nuanced understanding of the mutual interaction between public health and mobility movement. (highlight that we can not only interpret, but learn the correlation; w.r.t our Fig. 7(a).)
(3) We further increase the model interpretability by using the context-aware attention mechanism. By attending to relevant contextual information, such as regional population and income, it enables a better representation of the underlying social and economic factors that influence human mobility.

## 3 PROBLEM STATEMENT

In this section, we formally formulate the problem of predicting human mobility considering the COVID-19 pandemic. Before that, we first define the preliminaries.

**Traffic Networks.** We define the traffic network as an undirected graph $G = (V, E, A)$. $V = \{v_i\}_{i=1}^{N}$ represents the set of nodes, where $N = |V|$ as the number of nodes; $E$ represents the set of edges; $A \in \mathbb{R}^{N \times N}$ denotes the adjacency, which is a square matrix that describes the relationships between the nodes in the graph. In $A$, each row and column correspond to a node in the graph, and the entries $A_{ij}$ in the matrix indicate the presence or absence of edges between the nodes.

**Dynamic and Contextual Features.** We use $\mathbf{x}_t^i \in \mathbb{R}^F$ to denote the dynamic feature, i.e., features changing according to time, where $i \in \{1, \cdots N\}$ and $F$ is the length of the feature. The dynamic feature used in this paper includes the COVID-19 case rates and regional number of visits. Apart from the dynamic features, each node also has static features that do not change over time. Although constant, those static features can serve as the context for model prediction. To distinguish between the dynamic and static features, in the remainder of this paper, we use feature to stand for the dynamic feature, and context to account for the static feature. We use $\mathbf{c}^i \in \mathbf{R}^C$ to denote the context, where $i \in \{1, \cdots, N\}$ and $C$ is the length of the context vector for each node. The context used in this paper includes regional population, average income, and points of interest (POI). After defining the feature and context for each node, we use $\mathbf{X}_t = (\mathbf{x}_t^1, \mathbf{x}_t^2, \cdots, \mathbf{x}_t^N)^T \in \mathbb{R}^{N \times F}$ to denote the values of all nodal features at time $t$, where $T$ stands for the vector transpose to make it a column vector, and $\mathbf{C} = (\mathbf{c}^1, \mathbf{c}^2, \cdots, \mathbf{c}^N)^T$ to denote the all the nodal context.

**Problem**. With all preliminaries introduced above, we are ready to define the problem of predicting human mobility, i.e., the future regional number of visits. Given the historical nodal feature of previous $\tau$ time window, $\mathbf{X}_{(t-\tau+1):t} = [\mathbf{X}_{t-\tau+1}, \cdots, \mathbf{X}_t]$, and the nodal context $C$, we aim to learn a function $f$ to predict the future $\tau'$-length mobility sequence $\mathbf{Y}_{(t+1):(t+\tau')} = [\mathbf{Y}_{t+1}, \cdots, \mathbf{Y}_{(t+1):(t+\tau')}]$.

$$[\mathbf{X}_{(t-\tau+1):t}; \mathbf{C}] \xrightarrow{f} \mathbf{Y}_{(t+1):(t+\tau')}, \tag{1}$$

where $\mathbf{Y_t} = (y_t^1, \cdots y_t^N)$ stands for all the nodal numbers of visits; $y_t^i \in \mathbb{R}$ is the number of visit of node $i$ at time $t$, with $i \in \{1, \cdots, N\}$. To better distinguish between the mobility and the COVID-19 case rates, we use $\mathbf{X}_t^{mob}$ and $\mathbf{X}_t^{cov}$ to represent the mobility and COVID-19 case rates, respectively, where $\mathbf{X}_t^{mob}$ and $\mathbf{X}_t^{cov} \in \mathbb{R}^{N \times 1}$, and $\mathbf{X}_t = [\mathbf{X}_t^{mob}; \mathbf{X}_t^{cov}]$. Thus, Eq. 2 can be revised as:

$$[\mathbf{X}_{(t-\tau+1):t}^{mob}; \mathbf{X}_{(t-\tau+1):t}^{cov}; \mathbf{C}] \xrightarrow{f} [\mathbf{X}_{(t+1):(t+\tau')}^{mob}], \tag{2}$$

where $\mathbf{X}_{(t-\tau+1):t}^{mob}$ and $\mathbf{X}_{(t-\tau+1):t}^{cov} \in \mathbb{R}^{N \times \tau}$.

## 4 METHODOLOGY

### 4.1 Overview of CCAAT-GCN

The framework of CCAAT-GCN is shown in Fig. 1. Now we will introduce the overview of this framework from top to bottom and from left to right. In the upper left, the contextual features include POI, income, and population of each ZIP code region, forming a contextual feature graph that does not change through time. The contextual feature graph is then fed into the context-aware attention component in the spatial-temporal block (ST Block). In the middle, the dynamic features include weekly confirmed case rates and mobility data, each forming a spatial-temporal tensor. These two dynamic features are fed into the cross-attention component (CrosAtt), which is followed by a series of ST Blocks. Each ST Block consists of a temporal attention component (TAtt), a spatial attention component (SAtt), a context-aware attention component (CtxtAtt), a GCN, and a temporal convolutional layer (TConv). The ST Blocks are followed by a 1D-convolutional layer (Con1d) for the final transformation, after which the future mobility data is predicted. At the bottom, the Multigraph considers three different metrics and generates three different adjacency matrices. Then, the averaged adjacent matrix is fed into the GCN of the ST Block to conduct K-order Chebyshev polynomial approximation.

In the remainder of this section, we will first introduce the details of all the attention mechanisms mentioned in Fig. 1. Then, we will introduce the details of the ST Blocks and the Multigraph. The loss function will follow.

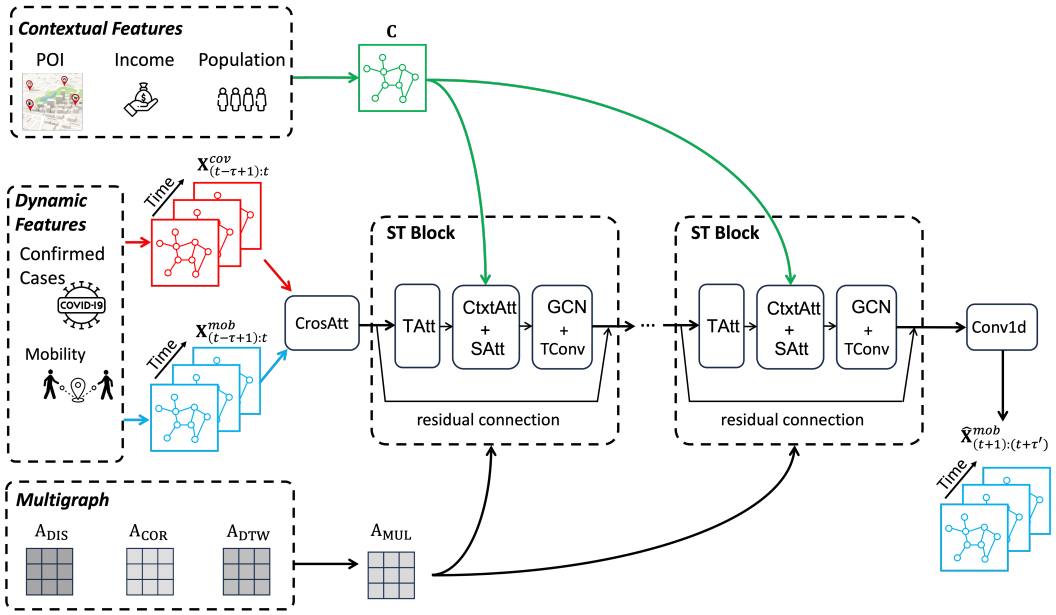



Fig. 1. Framework of the proposed CCAAT-GCN.

## 4.2 Attention Module

We introduce the attention mechanisms in the order of appearance in Fig. 1, which is cross-attention, temporal attention, context-aware attention and spatial attention.

*4.2.1 Cross-Attention.* Cross-attention (or cross-modal attention) mechanism [28] is a technique used in deep learning models to capture the relationship between inputs from two different modes, e.g., images and audios. It allows the model to learn how to selectively attend to different parts of the inputs, which can be useful in tasks such as natural language processing, computer vision, and speech recognition.

Mathematically, given the mobility feature $\mathbf{X}^{mob}_{(t-\tau+1):t}$ and the confirmed COVID-19 cases feature $\mathbf{X}^{cov}_{(t-\tau+1):t}$, we first calculate their embedding using the self-attention mechanisms as follows,

$$\begin{cases} \mathbf{Z}^{mob} = \text{softmax}\left(\dfrac{(\mathbf{Q}_{cr}\mathbf{X}^{mob}_{(t-\tau+1):t})^T\mathbf{K}_{cr}\mathbf{X}^{mob}_{(t-\tau+1):t}}{\sqrt{d}}\right)\mathbf{V}_{cr}\mathbf{X}^{mob}_{(t-\tau+1):t} \\ \mathbf{Z}^{cov} = \text{softmax}\left(\dfrac{(\mathbf{Q}'_{cr}\mathbf{X}^{cov}_{(t-\tau+1):t})^T\mathbf{K}'_{cr}\mathbf{X}^{cov}_{(t-\tau+1):t}}{\sqrt{d}}\right)\mathbf{V}'_{cr}\mathbf{X}^{cov}_{(t-\tau+1):t} \end{cases}, \tag{3}$$

where $\mathbf{Q}_{cr}$, $\mathbf{Q}'_{cr}$, $\mathbf{K}_{cr}$, $\mathbf{K}'_{cr}$, $\mathbf{V}_{cr}$ and $\mathbf{V}'_{cr}$ are learnable matrice; softmax function normalizes the similarity scores into a probability distribution over the keys; $d$ is the number of columns of $(\mathbf{X}^{mob}_{(t-\tau+1):t})^T$, which is the number of nodes $N$. The encoded features are then used to calculate the cross-attention:

$$\begin{cases} \mathbf{E}^{mob\to cov}_{cr} = \text{softmax}\left(\dfrac{(\mathbf{Z}^{mob})^T\mathbf{Z}^{cov}}{\sqrt{d}}\right)\mathbf{Z}^{cov} \\ \mathbf{E}^{cov\to mob}_{cr} = \text{softmax}\left(\dfrac{(\mathbf{Z}^{cov})^T\mathbf{Z}^{mob}}{\sqrt{d}}\right)\mathbf{Z}^{mob} \end{cases}. \tag{4}$$

These two cross attentions are summed to get the final cross attention output: $\mathbf{E}_{cr} = \mathbf{E}_{cr}^{mob \to cov} + \mathbf{E}_{cr}^{cov \to mob}$. In this way, the cross-attention mechanism allows a model to selectively attend to different parts of the input and output sequences, depending on the context of the current query. The final cross-attention output is then fed into the temporal attention component.

*4.2.2 Temporal attention.* We revise the framework of calculating temporal attention in [14] by adding the learned cross-attention $\mathbf{E}_{cr}$ into the framework. The equation is shown below,

$$\begin{cases} \mathbf{E}'_t = (\mathbf{U}_1 \mathbf{E}_{cr})^T \mathbf{U}_2 \mathbf{U}_3 \mathbf{E}_{cr} \\ \mathbf{E}_t = \text{softmax}\left(\mathbf{V}_t \cdot \sigma\left(\mathbf{E}'_t + \mathbf{b}_t\right)\right) \end{cases}, \tag{5}$$

where $\mathbf{V}_t$, $\mathbf{b}_t$ , $\mathbf{U}_1$, $\mathbf{U}_2$, and $\mathbf{U}_3$ are learnable parameters; $\sigma$ is the sigmoid function. The first step in the calculation of the temporal attention involves transforming the cross-attention matrix $\mathbf{E}_{cr}$ using learnable parameters $\mathbf{U}_1$, $\mathbf{U}_2$, and $\mathbf{U}_3$. This transformation, denoted as $\mathbf{E}'_t$, captures the interdependencies among different temporal slices of the data. To obtain the final temporal attention matrix $\mathbf{E}_t$, we calculate the dot product between the learnable parameter $\mathbf{V}_t$ and the transformed sum from the previous step. We then apply the softmax function to normalize the attention scores across all temporal slices, ensuring that the weights sum up to one.

*4.2.3 Context Attention.* Context-attention mechanism [43] is a technique used in deep learning models to improve the performance of natural language processing tasks, such as machine translation and text summarization. The attention mechanism allows the model to selectively focus on specific parts of the input, while the context-attention mechanism takes into account the context of the input in order to further improve the model's attentional capacity. This mechanism assigns different weights to different parts of the input based on their relevance to the context, allowing the model to better capture the meaning of the input and produce more accurate predictions. For example, in the task of machine translation, the context-attention mechanism can be used to weigh different words in the source sentence according to their importance to the translation of the target sentence. Overall, the context-attention mechanism is a powerful tool for improving the accuracy and interpretability of deep learning models in natural language processing tasks. First, the embedding of the contextual feature $\mathbf{C}$ is calculated as

$$\mathbf{Z}^{ct} = \text{MLP}(\mathbf{C}), \tag{6}$$

where MLP is the multiple layer perceptron. Then context attention can be calculated as

$$\mathbf{E}_{ct} = \text{softmax}\left(\frac{(\mathbf{Q}_{ct}\mathbf{E}_t)^T \mathbf{K}_{ct}\mathbf{E}_t\mathbf{Z}^{ct}}{\sqrt{d}}\right)\mathbf{V}_{ct}\mathbf{Z}^{ct}, \tag{7}$$

where $\mathbf{Q}_{ct}$, $\mathbf{K}_{ct}$ and $\mathbf{V}_{ct}$ are learnable matrices.

*4.2.4 Spatial attention.* We revise the framework of calculating temporal attention in [14] by adding the learned context-aware attention $\mathbf{E}_{ct}$ into the framework. The equation is shown below,

$$\begin{cases} \mathbf{E}'_s = (\mathbf{M}_1 \mathbf{E}_t)^T \mathbf{M}_2 \mathbf{M}_3 \mathbf{E}_t \cdot \mathbf{E}_{ct} \\ \mathbf{E}_s = \text{softmax}\left(\mathbf{E}_t \cdot \mathbf{E}_{ct} \cdot \sigma\left(\mathbf{E}'_s + \mathbf{b}_s\right)\right) \end{cases}, \tag{8}$$

where $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_3$, and $\mathbf{b}_s$ are learnable parameters. We first transform the temporal attention matrix $\mathbf{E}_t$ using the learnable parameters $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$. This transformation, denoted as $\mathbf{E}s'$, captures the spatial dependencies between different locations at the same temporal slice, while considering the contextual information $\mathbf{E}ct$. Next, we apply a sigmoid activation function $\sigma$ to the sum of $\mathbf{E}'_s$ and a bias term $\mathbf{b}_s$. This step enhances the discriminative power of the attention

mechanism by assigning importance weights to different spatial features based on their relevance to the prediction task. To obtain the final spatial attention matrix $\mathbf{E}s$, we calculate the element-wise product between the temporal attention matrix $\mathbf{E}t$, the contextual attention $\mathbf{E}_{ct}$, and the transformed sum from the previous step. We then apply the softmax function to normalize the attention scores.

In contrast to conventional methods of calculating spatial attention, we incorporate both cross-attention $\mathbf{E}_{cr}$ (previously utilized for calculating $\mathbf{E}_t$) and context-aware attention $\mathbf{E}_{ct}$. In our experiments, we will demonstrate how this integration of cross- and context-aware attention aids in learning an interpretable attention matrix.

### 4.3 Multigraph Module

The complex spatial features of human mobility cannot be captured completely by relying on a single graph, thus we propose a multi-graph mechanism. This subsection defines the adjacency matrix to characterize the spatial-temporal relationship of human mobility from multiple perspectives, including inter-regional distance and correlation.

*4.3.1 Distance-based graph.* We use the inter-nodal distance to compute the adjacency matrix of the distance of the distance-based graph [41]. The equation is depicted as follows,

$$(A_{DIS})_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), & i \neq j \text{ and } \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \geq \epsilon \\ 0, & \text{otherwise} \end{cases}, \tag{9}$$

where $d_{ij}$ is the distance between regions $i$ and $j$; $\sigma^2$ and $\epsilon$ are thresholds to control the distribution and sparsity of matrix $A_{DIS}$. We use the centroids of the regions to calculate their distance.

*4.3.2 Correlation-based graph.* We use Pearson correlation coefficient [44] between time series mobility data of a node pair to calculate the nodal correlation. We use $A_{COR}$ to denote the adjacent matrix of the correlation-based graph, which is depicted as

$$(A_{COR})_{ij} = \frac{\sum_{t=1}^{T}(\mathbf{x}_t^i - \bar{\mathbf{x}}^i)(\mathbf{x}_t^j - \bar{\mathbf{x}}^j)}{\sqrt{\sum_{t=1}^{T}(\mathbf{x}_t^i - \bar{\mathbf{x}}^i)^2}\sqrt{\sum_{t=1}^{T}(\mathbf{x}_t^j - \bar{\mathbf{x}}^j)^2}}, \tag{10}$$

where $\mathbf{x}_t^i$ represents the nodal feature of region $i$ at time $t$; $T$ is the historical time interval.

*4.3.3 Dynamic-time-warping based graph.* Dynamic Time Warping (DTW) is an algorithm for comparing and aligning time series data [21]. It measures the similarity between two sequences by finding the optimal alignment that minimizes the total distance between corresponding points. The algorithm calculates a distance matrix using the Euclidean or other distance measure and then applies dynamic programming to find the optimal alignment path.

By computing the DTW between each node's temporal sequences in mobility data, we can obtain a graph that represents temporal correlations. The DTW-based adjacency matrix, denoted as $A_{DTW}$, is calculated as

$$(A_{DTW})_{ij} = d_{ij} + \min\{(A_{DTW})_{i-1,j}, (A_{DTW})_{i,j-1}, (A_{DTW})_{i-1,j-1}\}, \tag{11}$$

where each entry of the adjacent matrix represents the accumulated distance at position $(i, j)$; the minimum value among the three neighbouring regions is used to update the matrix.

Finally, the adjacency matrix of the Multigraph, denoted as $A_{MUL}$, is calculated by averaging the above-mentioned adjacency matrices $A_{MUL} = (A_{DIS} + A_{COR} + A_{DTW})/3$

### 4.4 Graph Convolutional Network

We follow [41] to conduct convolution along the graph. Given the adjacency matrix of the Multigraph $A_{MUL}$, we define the normalized Laplacian matrix of the Multigraph as $L = I - D^{-1/2}A_{MUL}D^{-1/2} \in \mathbb{R}^{N\times N}$, where $I$ is a unit matrix and $D$ is a diagonal degree matrix with $D_{ii} = \sum_j (A_{MUL})_{ij}$. We use the K-order Chebyshev polynomials to approximate the graph convolution operator $*_G$ as follows:

$$g_\theta *_G x = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k \left(T_k(\tilde{L}) \odot \mathbf{E}_s\right) x, \tag{12}$$

where the parameter $\theta \in \mathbb{R}^K$ is the polynomial coefficients vector; $\mathbf{E}_s$ is the spatial attention matrix; $\odot$ is the Hadamard product; $\tilde{L} = \frac{2}{\lambda_{\max}} L - I$, with $\lambda_{\max}$ being the maximum eigenvalue of the Laplacian matrix. The recursive definition of the Chebyshev polynomial is $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, where $T_0(x) = 1$ and $T_1(x) = x$. Using the K-order Chebyshev polynomials approximation, each node is updated by the information of the K neighbouring nodes.

Once the graph convolution operations have successfully captured the information from neighbouring nodes in the spatial dimension, we further enhance the node's signal by stacking a standard convolution layer in the temporal dimension. This step allows us to merge the information obtained from neighbouring time slices and update the node's signal accordingly. A final 1×D convolution with a non-linearity neural network is involved to get the final prediction $\hat{\mathbf{X}}^{mob}_{(t-\tau+1):t}$. We use the mean squared error (MSE) as our loss function:

$$\mathcal{L} = \frac{||\mathbf{X}^{mob}_{(t-\tau+1):t} - \hat{\mathbf{X}}^{mob}_{(t-\tau+1):t}||^2}{N\tau}, \tag{13}$$

where $||\cdot||^2$ is the $\mathbb{L}_2$ norm.

## 5 DATA

To model human mobility during the COVID-19 pandemic, we utilize the SafeGraph dataset, which collects location data from mobile devices through apps installed on users' phones. SafeGraph data provides detailed information on the movement of people between different locations, including residential, commercial, and recreational areas. The dataset covers a large geographical area and is available at a high spatial and temporal resolution.

We aggregate the SafeGraph data into a weekly time frame, as it provides a good balance between granularity and data availability. The weekly resolution also aligns well with the weekly reporting frequency of the COVID-19 case data. We preprocess the SafeGraph data to extract the relevant features, including the number of visitors, duration of visits, and distance traveled between different locations.

Our original data has collected the weekly number of visits and other common features in 172 regions of New York from 2020/08/10-2022/04/18. We use the z-score method to normalize the feature number of visits and take the normalized data as the model input.

In addition to the SafeGraph data, we also utilize COVID-19 case data from the Centers for Disease Control and Prevention (CDC) to study the impact of the pandemic on human mobility. The CDC data provides weekly counts of COVID-19 cases, hospitalizations, and deaths across different regions in the United States. We combine the CDC data with the SafeGraph data to investigate the relationship between human mobility patterns and COVID-19 outbreaks. The contextual features, including income, population, and POI are from New York City open data.

Table 2. Summary of datasets

| | |
|---|---|
| **SafeGraph** | |
| # of ZIP code | 172 |
| # of Weeks | 90 |
| Time Span | 2020/08/10 - 2022/04/18 |
| **Point of Interest (POI)** | |
| # of POIs | 18,912 |
| Types of POIs | residential(16.%), education(20%), culture(3%), transportation(6.1%), social services(8.7%), recreational(17.2%), commercial(5.5%), government(4.5%), religious institution(8.4%), water(1.6%), public safety(3.3%), health services(1.5%), miscellaneous(3.5%) |
| **Median Household Annual Income (unit: Dollars per year)** | |
| Range | [31,536, 243,571] |
| **ZIP Code Population** | |
| Range | [1,783, 111,344] |

## 6 EXPERIMENT RESULTS

In this section, we first introduce our experiment setting, including baselines and evaluation metrics. Then we will present the results with our proposed model versus baselines using the real-world dataset mentioned in Sec. 5.

**Baselines and Evaluation Metrics**. We compare our model with the following baselines.

- ASTGCN. The Attention-based Spatial-temporal Graph Convolutional Network (ASTGCN) is a powerful deep learning model designed to capture both spatial and temporal dependencies in graph-structured data.
- STGCN. The Spatial-temporal Graph Convolutional Network (STGCN) model is composed of several spatial-temporal convolutional blocks and one fully-connected output layer. In each spatial-temporal convolutional block, there are two gated sequential convolution layers for capturing temporal dependency and one spatial graph convolution layer in between for capturing spatial dependency and this is like the "sandwich" structure.
- GraphWaveNet. GraphWaveNet is a graph convolutional neural network (GCN)-based model for graph classification tasks that utilize a WaveNet architecture for encoding graph signals. It operates on the graph in the spectral domain, using a variant of the graph Fourier transform to transform node features into a graph spectral domain representation.
- STFGCN. The Spatial-temporal Fusion Graph Convolutional Network (STFGCN) is a deep learning model designed specifically to capture spatial and temporal patterns in graph-structured data. STFGCN combines the power of graph convolutions and temporal fusion techniques to effectively model and predict the numbe of visits.
- LSTM. A vanilla temporal LSTM.
- Historical Average (HA). The historical average numbers of visits are used as the prediction of the corresponding future number of visits.
- Autogressive (AR). The standard autoregression model.

The mean absolute error (MAE), the root mean squared error (RMSE), and the relative error (RE) are used to measure the performance of models.

**Experiment Settings**. The first 50 weeks of data are selected as the training set, the next and last 20 weeks of data are selected as the validation and test sets, respectively. The historical time window size $\tau = 4$. The learning rate is 0.0001. Batchsize is set to 16. We use 0.1 dropout rate in the attention layer. The number of filters in the 1D-convolutional layer is 16.

### 6.1 Performance Comparison

The results comparison is shown in Table. 3. We focus on predicting the number of visits for 1, 2, and 3 weeks ahead. Multiple prediction horizons are considered to evaluate the model ability to capture short- and long-term patterns accurately.

Our CCAAT-GCN model outperformed the baselines across all prediction horizons, achieving the lowest RMSE and MAE values. This indicates that CCAAT-GCN effectively captured the complex spatiotemporal dynamics inherent in human mobility data. The incorporation of attention mechanisms and graph convolutional operations in CCAAT-GCN enabled it to effectively leverage both spatial and temporal information, resulting in improved prediction accuracy.

Comparing CCAAT-GCN to the other baselines, we observed that HA and AR, which rely solely on historical averages or autoregressive models, demonstrated relatively poor performance. LSTM, a popular recurrent neural network, showed competitive results but was outperformed by CCAAT-GCN. STGCN, STFGCN, Graph WaveNet, and ASTGCN, which incorporate spatial and temporal dependencies, achieved comparable performance, but CCAAT-GCN consistently exhibited superior accuracy.

The results emphasize the effectiveness of our proposed CCAAT-GCN model in capturing and predicting human mobility patterns. The combination of attention mechanisms and graph convolutional operations within CCAAT-GCN enables comprehensive modeling of the spatial and temporal aspects of the data, leading to more accurate and reliable predictions.

Table 3. Evaluation of different models using real-world mobility data (SafeGraph)

| Method | 1 week ahead | | 2 weeks ahead | | 3 weeks ahead | |
|---|---|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| HA | 0.25 | 0.19 | 0.86 | 0.48 | 0.65 | 0.33 |
| AR | 0.40 | 0.16 | 0.36 | 0.15 | 0.38 | 0.16 |
| LSTM | 0.34 | 0.17 | 0.36 | 0.17 | 0.42 | 0.19 |
| STGCN | 1.26 | 0.91 | 1.27 | 0.91 | 1.26 | 0.92 |
| STFGCN | 0.55 | 0.23 | 0.56 | 0.24 | 0.57 | 0.24 |
| Graph WaveNet | 0.28 | 0.10 | 0.31 | 0.11 | 0.32 | 0.12 |
| ASTGCN | 0.17 | 0.09 | 0.20 | 0.11 | 0.24 | 0.13 |
| CCAAT-GCN | 0.13 | 0.06 | 0.13 | 0.06 | 0.17 | 0.07 |

### 6.2 Convergence Analysis

The convergence analysis of our proposed model is presented in Fig. 2, which illustrates the training and validation errors as a function of the training iterations. It can be observed that both curves exhibit a gradual decrease in error over time, indicating the model's learning progress. After approximately 3000 epochs, the training and validation errors reach a convergence point, suggesting

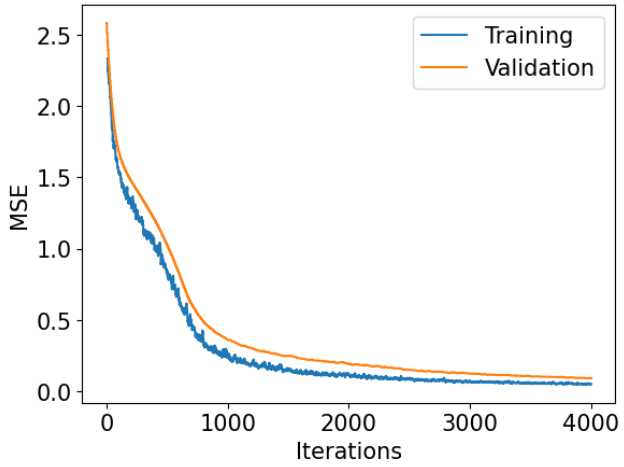



Fig. 2. Training and validation error.

that the model has effectively learned the underlying patterns in the data. This convergence indicates that further training iterations may not significantly improve the model's performance or reduce the error rate. The convergence of the training and validation errors signifies the stability and reliability of our model.

### 6.3 Ablation Study

To investigate the effectiveness of various components in CCAAT-GCN, Fig. 3 presents the results of ablation studies. We evaluate the performance of our proposed CCAAT-GCN model by comparing it with three variations that remove specific components from CCAAT-GCN. These variations include CCAAT-GCN without cross-attention, CCAAT-GCN without context-aware attention, and CCAAT-GCN without both components. The results of the ablation study reveal that our complete CCAAT-GCN model outperformed all the variants in terms of prediction accuracy. When comparing the performance of CCAAT-GCN without cross-attention and CCAAT-GCN without context-aware attention, it is observed that the removal of either component resulted in decreased prediction accuracy. This indicates that both cross-attention and context-aware attention contribute to the model's ability to capture and leverage important information for accurate predictions. This result confirms the importance of the cross-attention and context-aware attention mechanisms in our CCAAT-GCN model. The integration of these components allows the model to effectively capture and leverage relevant information from both spatial and temporal contexts, leading to improved prediction accuracy in human mobility prediction.

### 6.4 Visualization

Fig. 4 provides a heatmap visualization of the relative errors across different ZIP code areas on the map of New York City. The relative error displayed in the heatmap represents the overall relative error for the 3-week-ahead prediction. The results reveal that our CCAAT-GCN model exhibits the best performance among the compared models, demonstrating the lowest overall relative error across various ZIP codes. This indicates that CCAAT-GCN successfully captures the complex spatiotemporal patterns in the human mobility data, resulting in accurate predictions across different regions of New York City. In contrast, the HA model generally exhibits relatively

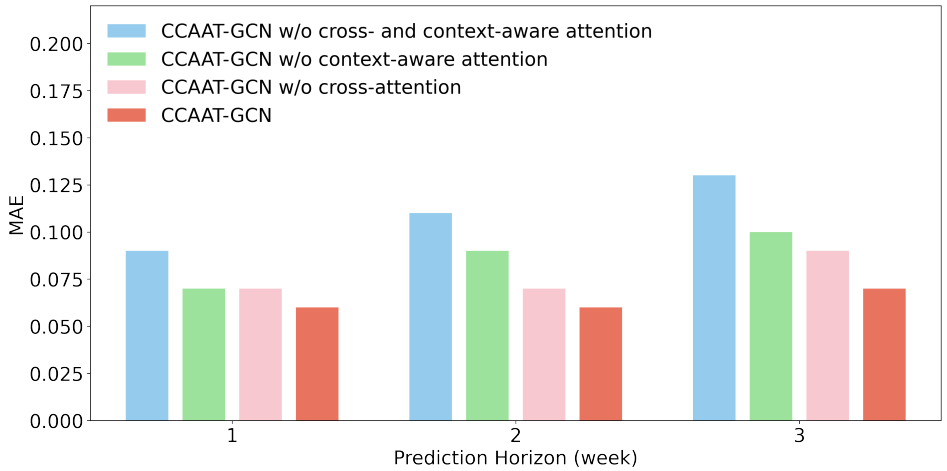



Fig. 3. The result of the ablation study.

higher relative error. This is likely due to the simplistic approach of calculating the historical average, which may not be suitable for long-term predictions with non-stationary data. The HA model's limitations in capturing the dynamics of the human mobility patterns could explain its higher relative error compared to the other models. While STFGCN generally performs well, there are specific regions where its performance is suboptimal. This could be attributed to its inability to handle corner cases effectively, resulting in less accurate predictions in those particular areas. It is worth noting that all models struggle to achieve satisfactory performance in some common areas, such as the right-bottom corner of Staten Island. This can be attributed to the use of relative error as the evaluation metric, which amplifies the impact of high relative error in regions with low visitation frequency. In areas with a limited number of visits, even a slight deviation in predictions can result in a relatively high relative error, affecting the overall performance of the models.

Fig. 5 presents a bar chart depicting the comparison between the predicted and ground-truth number of visits in different ZIP code areas. The x-axis represents the ZIP codes, while the y-axis represents the number of visits. The comparison is focused on our proposed CCAAT-GCN model. The bar chart demonstrates that our CCAAT-GCN model achieves favorable results overall, accurately predicting the number of visits in various ZIP code areas. This indicates the model's ability to capture and learn the underlying patterns of human mobility, allowing it to provide reliable predictions even in areas with a large-scale number of visits. Furthermore, this figure helps explain the previous observation of high relative errors in specific areas. It becomes evident that the areas with high relative errors correspond to those with very low numbers of visits. In such regions, even a slight discrepancy between the predicted and ground-truth values can lead to a significantly high relative error, given the small denominator. This emphasizes the challenge of accurately predicting human mobility patterns in areas with sparse or limited visits. The bar chart highlights the effectiveness of our CCAAT-GCN model in capturing the nuances of human mobility across different ZIP code areas, including those with varying scales of visitation. The model's ability to provide reliable predictions even in areas with a high number of visits contributes to its overall performance and reinforces its suitability for human mobility prediction tasks.

Fig. 6 provides a visual representation of the time-series number of visits spanning 90 weeks, along with the predicted number of visits for the last 20 weeks, for two selected ZIP code areas.

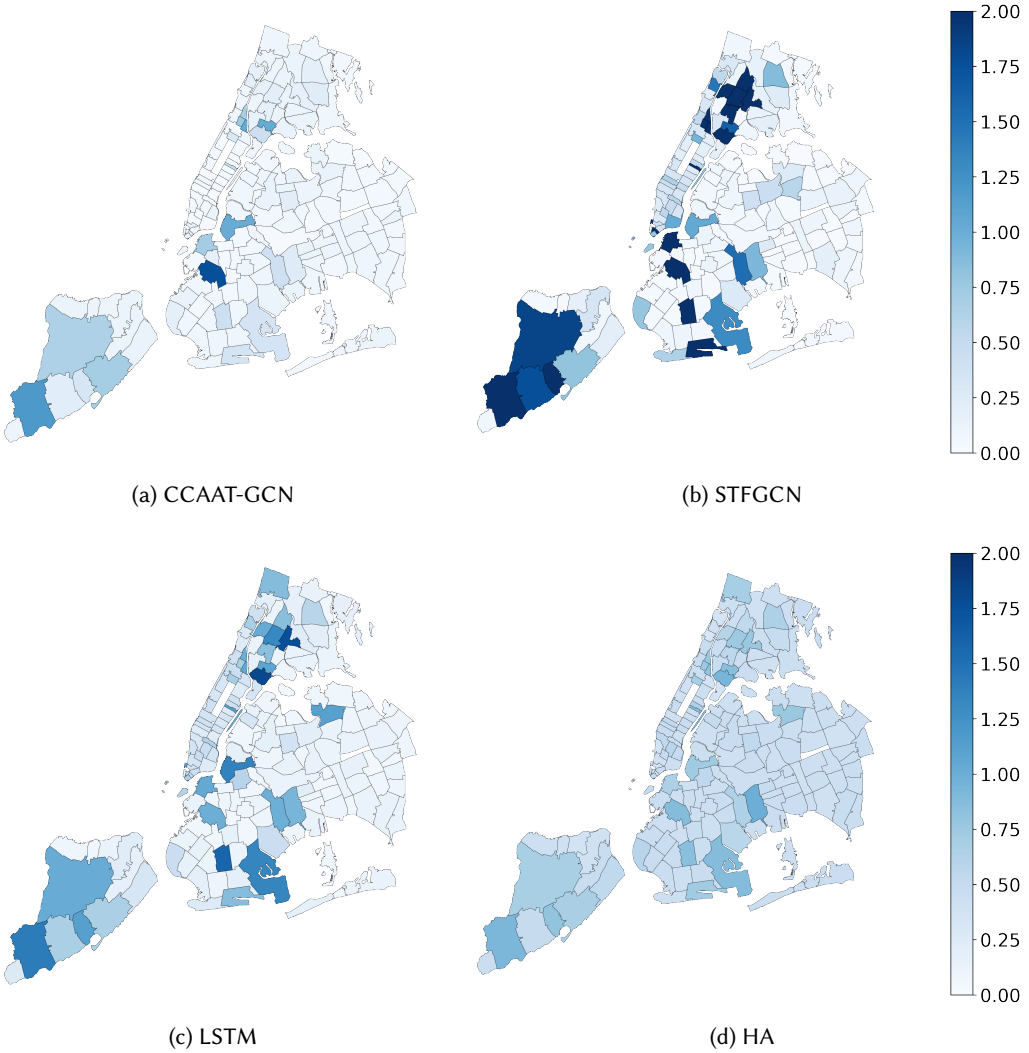



Fig. 4. Relative errors for each ZIP code.

The solid and dashed lines represent the observed and predicted number of visits, respectively. The figure demonstrates a good agreement between the predicted and observed number of visits, even in scenarios where the patterns of the number of visits exhibit non-stationary behavior. This is particularly evident during the transition from the first 50 weeks to the final 20 weeks, where the number of visits displays varying patterns. The ability of our model to accurately predict non-stationary patterns can be attributed to the utilization of cross- and context-aware attention mechanisms. These mechanisms leverage information from multiple sources, including COVID-19 case rates and contextual features, to enhance the training process and account for distribution shifts. By incorporating these attention mechanisms, our model effectively captures the evolving

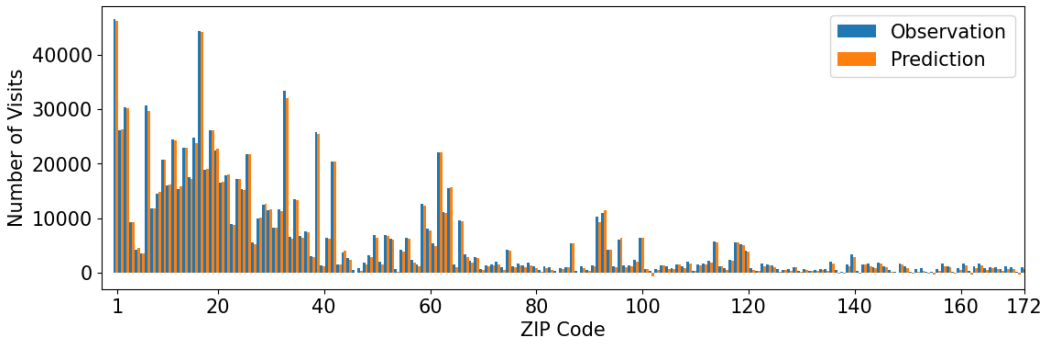



Fig. 5. Prediction results of CCAAT-GCN

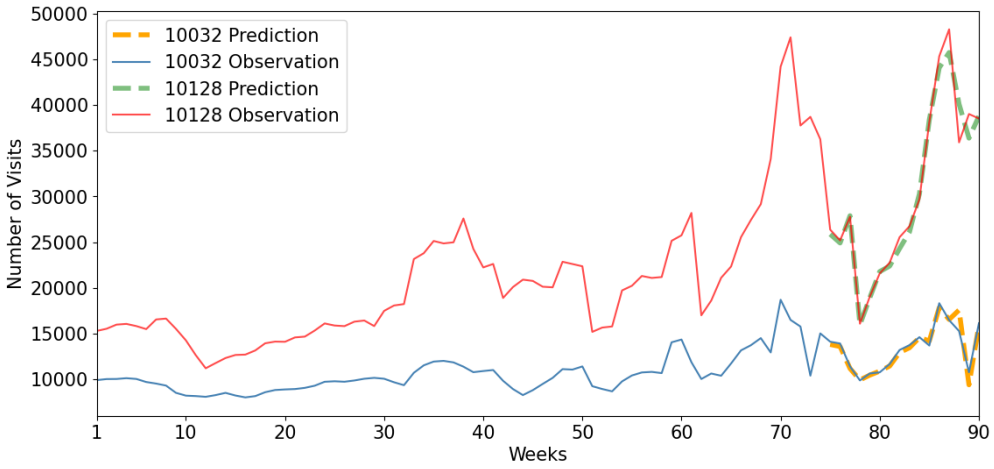
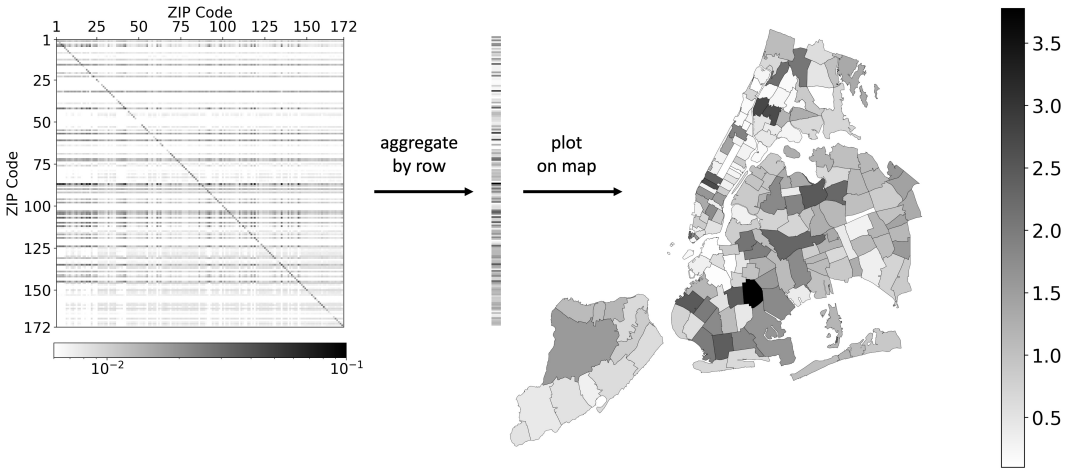



Fig. 6. Predicted number of visits.

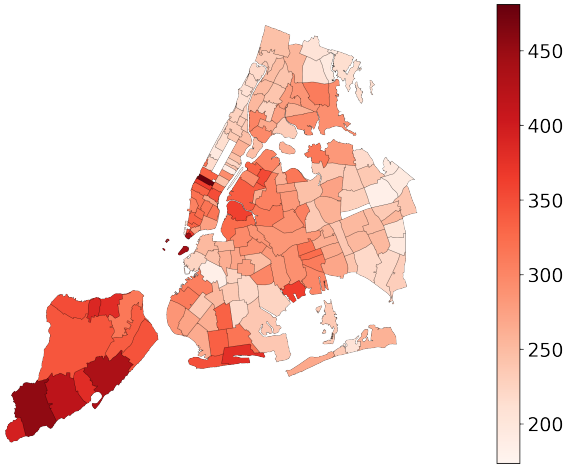
dynamics of human mobility, enabling accurate predictions even in the presence of changing patterns.

Fig. 7(a) illustrates a heatmap depicting the attention matrix acquired from the spatial-attention within our CCAAT-GCN model. The x-axis and y-axis represent different ZIP codes. We can see that certain ZIP codes along the y-axis exhibit notably high attention scores across a majority of ZIP codes along the x-axis. This observation indicates that these specific areas play a critical role in affecting mobility patterns throughout the entire region. This heatmap provides insights into the importance of certain areas in relation to their COVID-19 case rates when predicting the number of visits in other areas. To better interpret the attention scores, we aggregate them along the x-axis and plot them on the map. For comparative analysis, we also plot the 50-week aggregated COVID-19 case rates in Fig. 7(b). Two notable observations are made when comparing these two heatmaps.

(1) In Fig. 7(b), Staten Island, situated in the bottom left corner, exhibits relatively high COVID-19 case rates. However, in Fig. 7(a), this region displays low spatial-attention scores. This discrepancy can be attributed to the island's geographical isolation from other areas. That




(a) Left: average attention matrix learned in CCAT-GCN. Middle: aggregated attention scores in each ZIP code by row; the sum of the attention score represents the influence of a specific ZIP code on others. Right: geospatial visualization of the aggregated attention scores on the map.




(b) Average weekly COVID-19 case rates

Fig. 7. Visualization of the average attention matrix (a) versus average weekly COVID-19 case rates (b), both calculated using data from 2020/08/10 to 07/26/2021.

is to say, despite its high case rates, the impact of COVID-19 case rates in Staten Island on mobility patterns in other regions of New York City is limited. This finding also underscores the capability of the cross- and context-aware-attention mechanism to discover the underlying impact of COVID-19 on mobility, rather than solely relying on the magnitudes of case rates.

(2) The region with the highest attention score in Fig. 7(a), located in the King County of Brooklyn, exhibits a relatively low COVID-19 case rate in Fig. 7(b). After investigating the POI of this region, we find that the Kings County Hospital Center is located here together

883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931

with several other healthcare facilities. This observation suggests that, despite its lower case rates, the region encompassing Kings County Hospital Center plays an important role in influencing mobility patterns in other areas, owing to its concentration of healthcare facilities. Moreover, this observation serves as evidence of the cross- and context-aware attention mechanism's ability to identify crucial areas even in the absence of high COVID-19 case rates.

## 7 CONCLUSION

We developed a model CCAAT-GCN (Cross- and Context-Aware Attention based Spatial-Temporal Graph Convolutional Network) for human mobility prediction, especially during disruptive events like COVID-19. In the past two years during the COVID-19 pandemic, people's mobility patterns have gone through several waves, aligned with the waves of COVID evolution. How do we predict a nonstationary spatiotemporal pattern using deep learning models? To tackle such a challenge, here we include the COVID case number to capture such nonstationarity. Building upon the GCN framework, the cross-attention module specifically models the correlation between COVID-19 cases and the number of visits, allowing for a comprehensive understanding of their mutual influence. Moreover, the context-attention module learns to incorporate relevant contextual features, such as regional demographics or socioeconomic factors, to enhance the prediction accuracy and interpretability of the model.

The proposed model was validated using SafeGraph data in New York City from August 2020 to April 2022. A comprehensive list of baseline models were performed, ranging from various spatiotemporal GCN models to time-series models. Ablation study confirms the importance of the cross-attention and context-aware attention mechanisms in our CCAAT-GCN model. The integration of these components allows the model to effectively capture and leverage relevant information from both spatial and temporal contexts, leading to improved prediction accuracy in human mobility prediction.

We plan to extend this work in the following aspects: (1) validate it using different datasets across various disruptive disasters and events, like hurricanes and big events, which could transform human mobility patterns. (2) learn the invariant structure underlying the spatiotemporal mobility patterns for generalization and transfer learning.

## ACKNOWLEDGEMENTS

This work is sponsored by the National Science Foundation under the Smart and Connected Communities (S&CC) program #2218809. We would also like to thank master and undergraduate students from Columbia University who have assisted in data processing, preparation, and baseline model testing throughout the process of the project. They are Bader Mansour A Alaskar, Hanrui Lyu, Lindong Liu, Ruiqi Wang, Srikanth Chandar, and Xuan Lian.

## REFERENCES

[1] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems* 33 (2020), 17804–17815.

[2] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* (2018).

[3] Eddy Bekkers and Robert B Koopman. 2022. Simulating the trade effects of the COVID-19 pandemic: Scenario analysis based on quantitative trade modelling. *The World Economy* 45, 2 (2022), 445–467.

[4] Qi Cao, Renhe Jiang, Chuang Yang, Zipei Fan, Xuan Song, and Ryosuke Shibasaki. 2023. MepoGNN: Metapopulation epidemic forecasting with graph neural networks. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part VI*. Springer, 453–468.

[5] Xu Chen and Xuan Di. 2022. How the COVID-19 Pandemic Influences Human Mobility? Similarity Analysis Leveraging Social Media Data. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2955–2960.

[6] Xu Chen, Zihe Wang, and Xuan Di. 2023. Sentiment Analysis on Multimodal Transportation during the COVID-19 Using Social Media Data. *Information* 14, 2 (2023), 113.

[7] Weizhen Dang, Haibo Wang, Shirui Pan, Pei Zhang, Chuan Zhou, Xin Chen, and Jilong Wang. 2022. Predicting Human Mobility via Graph Convolutional Dual-attentive Networks. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 192–200.

[8] Manlio De Domenico, Antonio Lima, and Mirco Musolesi. 2013. Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing* 9, 6 (2013), 798–807.

[9] Zipei Fan, Xuan Song, Renhe Jiang, Quanjun Chen, and Ryosuke Shibasaki. 2019. Decentralized attention-based personalized human mobility prediction. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–26.

[10] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. 2018. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 world wide web conference*. 1459–1468.

[11] Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. 2020. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3426–3433.

[12] Qiang Gao, Fan Zhou, Goce Trajcevski, Kunpeng Zhang, Ting Zhong, and Fengli Zhang. 2019. Predicting human mobility via variational attention. In *The world wide web conference*. 2750–2756.

[13] Kan Guo, Yongli Hu, Yanfeng Sun, Sean Qian, Junbin Gao, and Baocai Yin. 2021. Hierarchical graph convolution network for traffic forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 151–159.

[14] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 922–929.

[15] Xinchen Hao, Renhe Jiang, Jiewen Deng, and Xuan Song. 2022. The Impact of COVID-19 on Human Mobility: A Case Study on New York. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 4365–4374.

[16] Jizhou Huang, Haifeng Wang, Miao Fan, An Zhuo, Yibo Sun, and Ying Li. 2020. Understanding the impact of the COVID-19 pandemic on transportation-related behaviors with human mobility data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3443–3450.

[17] Yuqin Jiang, Xiao Huang, and Zhenlong Li. 2021. Spatiotemporal patterns of human mobility and its association with land use types during COVID-19 in New York City. *ISPRS International Journal of Geo-Information* 10, 5 (2021), 344.

[18] Rushina Khan and SL Hasan. 2020. Telecommuting: The problems & challenges during COVID-19 (2020). *International Journal of Engineering Research & Technology (IJERT)* 9, 7 (2020), 1027–1033.

[19] S Vasantha Kumar and Lelitha Vanajakshi. 2015. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review* 7, 3 (2015), 1–9.

[20] Shiyong Lan, Yitong Ma, Weikang Huang, Wenwu Wang, Hongyu Yang, and Pyang Li. 2022. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In *International Conference on Machine Learning*. PMLR, 11906–11917.

[21] Mengzhang Li and Zhanxing Zhu. 2021. Spatial-temporal fusion graph neural networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 4189–4196.

[22] Xiaolong Li, Gang Pan, Zhaohui Wu, Guande Qi, Shijian Li, Daqing Zhang, Wangsheng Zhang, and Zonghui Wang. 2012. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science* 6 (2012), 111–121.

[23] Xianglong Luo, Danyang Li, Yu Yang, and Shengrui Zhang. 2019. Spatiotemporal traffic flow prediction with KNN and LSTM. *Journal of Advanced Transportation* 2019 (2019).

[24] Terence C Mills. 1990. *Time series techniques for economists*. Cambridge university press.

[25] Shaojie Qiao, Dayong Shen, Xiaoteng Wang, Nan Han, and William Zhu. 2014. A self-adaptive parameter selection trajectory prediction approach via hidden Markov models. *IEEE Transactions on Intelligent Transportation Systems* 16, 1 (2014), 284–296.

[26] Yuanyuan Qiao, Zhongwei Si, Yanting Zhang, Fehmi Ben Abdesslem, Xinyu Zhang, and Jie Yang. 2018. A hybrid Markov-based model for human mobility prediction. *Neurocomputing* 278 (2018), 99–109.

[27] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. 2006. Evaluating next-cell predictors with extensive Wi-Fi mobility data. *IEEE transactions on mobile computing* 5, 12 (2006), 1633–1649.

[28] Zhaoxin Su, Gang Huang, Sanyuan Zhang, and Wei Hua. 2022. Crossmodal transformer based generative framework for pedestrian trajectory prediction. In *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2337–2343.

[29] Yihong Tang, Junlin He, and Zhan Zhao. 2022. HGARN: Hierarchical Graph Attention Recurrent Network for Human Mobility Prediction. *arXiv preprint arXiv:2210.07765* (2022).

[30] Yudong Tao, Chuang Yang, Tianyi Wang, Erik Coltey, Yanxiu Jin, Yinghao Liu, Renhe Jiang, Zipei Fan, Xuan Song, Ryosuke Shibasaki, et al. 2022. A Survey on Data-Driven COVID-19 and Future Pandemic Management. *ACM computing surveys* 55, 7 (2022), 1–36.

[31] Fernando Terroso-Sáenz, Jesús Cuenca-Jara, Aurora González-Vidal, and Antonio F Skarmeta. 2016. Human mobility prediction based on social media with complex event processing. *International Journal of Distributed Sensor Networks* 12, 9 (2016), 5836392.

[32] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[33] Haoyun Wang and Robert B Noland. 2021. Bikeshare and subway ridership changes during the COVID-19 pandemic in New York City. *Transport policy* 106 (2021), 262–270.

[34] Huandong Wang, Sihan Zeng, Yong Li, and Depeng Jin. 2020. Predictability and prediction of human mobility based on application-collected location data. *IEEE Transactions on Mobile Computing* 20, 7 (2020), 2457–2472.

[35] Lijing Wang, Aniruddha Adiga, Jiangzhuo Chen, Adam Sadilek, Srinivasan Venkatramanan, and Madhav Marathe. 2022. Causalgnn: Causal-based graph neural networks for spatio-temporal epidemic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 12191–12199.

[36] Michael S Warren and Samuel W Skillman. 2020. Mobility changes in response to COVID-19. *arXiv preprint arXiv:2003.14228* (2020).

[37] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. 1907–1913.

[38] Feng Xie, Zhong Zhang, Liang Li, Bin Zhou, and Yusong Tan. 2023. EpiGNN: Exploring spatial transmission with graph neural network for regional epidemic forecasting. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part VI*. Springer, 469–485.

[39] Jiawei Xue, Takahiro Yabe, Kota Tsubouchi, Jianzhu Ma, and Satish Ukkusuri. 2022. Multiwave covid-19 prediction from social awareness using web search and mobility data. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4279–4289.

[40] Chuang Yang, Zhiwen Zhang, Zipei Fan, Renhe Jiang, Quanjun Chen, Xuan Song, and Ryosuke Shibasaki. 2022. Epimob: Interactive visual analytics of citywide human mobility restrictions for epidemic control. *IEEE Transactions on Visualization and Computer Graphics* (2022).

[41] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875* (2017).

[42] Cunjun Yu, Xiao Ma, Jiawei Ren, Haiyu Zhao, and Shuai Yi. 2020. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*. Springer, 507–523.

[43] Rui Yu, Yifeng Li, Wenpeng Lu, and Longbing Cao. 2022. Tri-Attention: Explicit Context-Aware Attention Mechanism for Natural Language Processing. *arXiv preprint arXiv:2211.02899* (2022).

[44] Yingxue Zhang, Yanhua Li, Xun Zhou, Xiangnan Kong, and Jun Luo. 2019. TrafficGAN: Off-deployment traffic estimation with traffic generative adversarial networks. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 1474–1479.