A Graph Based Deep Learning Framework for Predicting Spatio-Temporal Vaccine Hesitancy

Sifat Afroj Moon
Biocomplexity Institute & Initiative
University of Virginia
Virginia, USA
sifatmoon@virginia.edu

Rituparna Datta
Dept. of Computer Science
University of Virginia
Virginia, USA
hht9zt@virginia.edu

Tanvir Ferdousi
Biocomplexity Institute & Initiative
University of Virginia
Virginia, USA
tanvir@virginia.edu

Hannah Baek
Biocomplexity Institute & Initiative
University of Virginia
Virginia, USA
yc3rz@virginia.edu

Abhijin Adiga
Biocomplexity Institute & Initiative
University of Virginia
Virginia, USA
abhijin@virginia.edu

Achla Marathe
Dept. of Computer Science
Dept. of Public Health Sciences
University of Virginia
Virginia, USA
achla@virginia.edu

Anil Vullikanti
Biocomplexity Institute & Initiative
Biocomplexity Institute & Initiative
University of Virginia
Virginia, USA
vsakumar@virginia.edu

***Abstract*—Predicting vaccine hesitancy at a fine spatial level assists local policymakers in taking timely action. Vaccine hesitancy is a heterogeneous phenomenon that has a spatial and temporal aspect. This paper proposes a deep learning framework that combines graph neural networks (GNNs) with sequence module to forecast vaccine hesitancy at a higher spatial resolution. This integrated framework only uses population demographic data with historical vaccine hesitancy data. The GNN learns the spatial cross-regional demographic signals, and the sequence module catches the temporal dynamics by leveraging historical data. We formulate the problem on a spatial graph where nodes are zip codes. We consider three variants of the graph based on three different criteria: geographic adjacency, distance, and mobility from an acitivty-based social contact network. Our framework effectively predicts the spatio-temporal dynamics of vaccine hesitancy at the zip-code level when the mobility network is used to formulate the graph. We use our combined model for two tasks: 1) spatial prediction and 2) temporal prediction. In the spatial prediction task, we partition the zip codes into two sets: known and unknown. We utilize the vaccine hesitancy data from the known zip codes to predict the vaccine hesitancy levels for the unknown zip codes at time t. Temporal prediction forecasts vaccine hesitancy for each zip code at the next time point t+1. Experiments on the real-world vaccine hesitancy data from the All-Payer Claims Database (APCD) show that our framework can outperform a range of baselines for both tasks. Our study finds that only historical time series data for vaccine hesitancy levels without spatial consideration is insufficient to learn the hesitancy pattern.**

***Index Terms*—Graph neural network, prediction, clustering, claim data, vaccine hesitancy**

I. Introduction

Highly contagious diseases, such as measles, are regarded as vaccine-preventable (VPD) because of the availability of the Measles, Mumps, and Rubella (MMR) vaccine, which has a very high efficacy rate. Measles is preventable using high rates of immunization. The MMR vaccine is required by public schools in most parts of the world, including the US, and measles was declared as "eliminated" from the US in 2000 [1]. Unfortunately, immunization rates are declining for many childhood vaccines, and outbreaks of measles and other VPDs have been occurring regularly in recent years across the world. For instance, there was a large outbreak in New York in 2019, which caused over 900 cases [2]. In 2021, Nigeria had over 10,000 cases [3], and there were 128,000 deaths due to measles worldwide [4]. The risk of measles, and other vaccine-preventable diseases, has significantly been exacerbated due to the COVID-19 pandemic [5].

There are a number of reasons behind the drop in immunization rates, and hesitancy is the leading among them. Even before the pandemic, though the MMR vaccine coverage was quite high ($\sim$ 95%) for kindergarten children

This work is partially supported by National Institutes of Health (NIH) Grant R01GM109718, NSF (National Science Foundation) grants IIS-1955797, ACI-1443054, OAC-1916805, NSF Expeditions in Computing Grant CCF-1918656, CCF-1917819, US Centers for Disease Control and Prevention 75D30119C05935, DTRA (Defense Threat Reduction Agency) subcontract/ARA S-D00189-15-TO-01-UVA, and a collaborative seed grant from the UVA Global Infectious Disease Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsoring agencies.

nationally [6] (which is a high enough rate to reach herd immunity), it was not evenly spread geographically, and there were significant pockets of undervaccination [7]–[10]. During the pandemic, significant drops in routine immunizations have been reported [11]–[13] In 2020 and 2021, over 27 and 25 million children were estimated to have missed their first dose of the measles vaccine, respectively [12], [14]. Measles is now viewed as an imminent global threat [13].

Vaccine hesitancy is a growing concern in public health [15], and predicting vaccine hesitancy at the higher spatial resolution is considered a fundamental problem, as heterogeneous vaccine coverage significantly increases the risk of outbreaks [16], [17]. One of the significant challenges in understanding the extent of hesitancy and how it is spreading is the limited availability of surveillance data on declining immunization rates, especially at finer spatial resolutions. Surveys on immunization rates are often available at a coarse resolution (e.g., a state) [11], [12], which does not help identify specific under-immunized regions for intervention. Many states provide immunization rates for schools through the School Immunization Survey (SIS) reports [18]. However, they only consider specific age groups (e.g., 4-6 years) and even in this age group, fail to cover a large population (e.g., those who are home-schooled).

The focus of our paper is to develop methods to predict vaccine hesitancy at the zip code level; we refer to this as the VACCHESITANCY problem. In this work, we investigate vaccine hesitancy among kids aged between 0-6 years. Prior works have used this information as a reflection of parental vaccine intention [19]. Kids in this age range are expected to receive a set of mandatory vaccines including MMR (Measles, Mumps, and Rubella), HepB (Hepatitis B), and DTaP (Diphtheria, Tetanus, and Pertussis). A novel aspect of our work is the use of an extensive insurance claims dataset for Virginia that includes all insurance claims for over 5 million individuals over a five years period.

Analyzing vaccine hesitancy has been an active area of research. Many works have focused on understanding hesitancy and identifying the responsible factors using social media data [20]. However, these data contain notable biases from demographic variations in platform preferences and the information individuals opt to share. On the other hand, some recent data-driven approaches use detailed individual-level data, making it harder to generalize. We discuss the relevant works on vaccine hesitancy modeling in Section II.

Our main contributions are as follows:

- We develop a novel approach, VH-GNN, for the VACCHESITANCY problem by combining a GNN and a recurrent neural network (RNN) using demographic data and historical hesitancy data, along with detailed population mixing data in the state. The GNN captures the spatial aspect of vaccine hesitancy by learning the impact of neighboring zip codes with respect to population-level mixing. The RNN learns temporal dynamics by leveraging historical hesitancy data. In the rest of the paper, we refer to the combined framework as VH-GNN; Vaccine Hesitancy predicting Graph Neural Network. Figure 1 shows the VH-GNN architecture.
- We train and evaluate the VH-GNN using the large-scale insurance claims data set for the state of Virginia, mentioned earlier, and show that our model outperforms a number of baselines, leading to a substantial reduction in prediction errors ranging from 18.40% to 43.4%.
- Through an ablation study, we demonstrate the effectiveness of the combined framework in improving model performance. In particular, we find that spatial structure from the detailed population-level mixing is very significant in forecasting vaccine hesitancy. We explore other kinds of connectivity and spatial structure too, but do not find them to be as predictive.
- While the performance of VH-GNN is generally superior, we find there are some zip codes (denoted by set V_L) where the prediction error is high. We identify several features which characterize the zip codes in V_L such as size of population in the target age group, vaccine hesitancy percentage, medicaid insurance percentage, hispanic population percentage.
- In order to understand the structure of the solution from VH-GNN and the true hesitancy level datasets, we use the Moran's-I and isolation indices, which are metrics for quantifying spatial clustering. We find that Moran's-I is high and the isolation index is low, indicating a similar clustering structure between the solution predicted by VH-GNN and the actual claims dataset.

II. RELATED WORK

Research on vaccine hesitancy prediction can be divided into two major categories: data-driven studies and model-based studies. Recent data-driven vaccine hesitancy studies explore different machine learning models, such as neural networks, random forest, logistic regression, recursive partitioning, and support vector machines, to find local vaccine hesitancy hotspots or to predict individual decisions [21]–[23]. These studies do not consider the spatial aspect of vaccine hesitancy. However, vaccine refusal has a spatial clustering nature, and the immunization status of kids shows correlations in the same neighborhood, schools, or jurisdictions [7], [24], [25]. In addition, these studies have used detailed socio-economic data, including siblings' vaccination history and private medical history, which are inaccessible in many cases due to privacy concerns. In this work, using only features developed at the zip code level, such as population size, gender, race, and insurance type, we are able to achieve high performance.

Mollalo and Tattar [26] study the spatial distribution of vaccine rates in the US based on social vulnerability index using a multiscale geographically weighted regression model [26]. Their work identifies important covariates and shows that their importance varies across space. A recent spatial mathematical model of opinion dynamics with reinforcement explains the occurrence of vaccine hesitancy. Mathematical models often do not consider heterogeneous social connectivity.

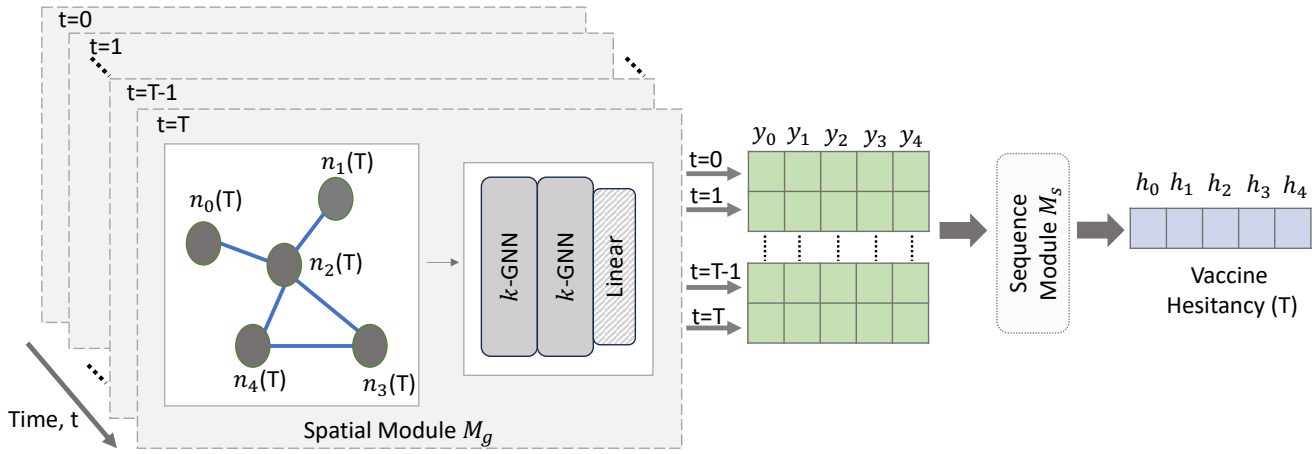


Fig. 1: Architecture of the spatio-temporal graph-based node-level regression learning for the prediction of vaccine hesitancy.

To address the spatial neighbor impact and temporal dynamics of vaccine hesitancy, we use a graph-based deep learning framework. It combines Graph Neural Network (GNN) with a sequence module for a node-level prediction task. GNN is a deep-learning tool specialized to handle graph data [27]. GNN is widely used in different domains for graph-related prediction tasks, such as node classification, link prediction, and graph classification. GNNs have demonstrated good prediction capabilities for spatial data, such as house price estimation, understanding election results [28], and weather forecasting. In this graph-based research, nodes represent zip codes, and edges represent the connectivity among zip code pairs.

Patient refusal or vaccine hesitancy sentiment is changing with time. To learn the temporal aspect of the vaccine hesitancy levels for a location, we propose a sequence module to handle time series data for each zip code. Our framework uses a recurrent neural network structure with Gated recurrent units (GRUs) [29] as the sequence module.

Spatio-temporal graph learning has been used in recent times to forecast traffic flow [30], disease prevalence [31], etc. Prior works have used different combinations of GNN, RNN, or Convolutional Neural Networks (CNN) to perform spatio-temporal forecasting tasks [32], [33]. Traffic forecasting is an example of spatio-temporal modeling. Yu et al. show the potential of graph-based learning frameworks for timely and accurate traffic forecasts with comparisons between CNN and RNN. We extended these concepts to predict vaccine hesitancy for a zip code in the next time step. For spatio-temporal learning, we use a static network, where nodes are zip codes, and edges are connections among pairs of zip codes. Nodes have time-varying features as attributes. Using demographic and historical vaccine hesitancy data, we forecast the vaccine hesitancy percentage of a node or a zip code.

III. Methods

A. Preliminaries

Our goal is to learn vaccine hesitancy, measured by the percentage of patients refusing to take the vaccine within a zip code. As inputs, we use historical vaccine hesitancy data and demographic features of a zip code (which includes age, gender, race, ethnicity, and insurance type), and spatial zip code level connectivity information of several types.

Let $G(V, E, A)$ denote a graph, where V denotes the set of N zip codes. We consider three types of connectivity information in defining the edges: geographic adjacency, the distance between zip codes, and mixing through an activity-based network (described below in Section III-C). Let $A^{N\times N}$ denote the weighted adjacency matrix. An entry of A, $a(i, j)$, represents the connection from node i to j. A node i is associated with a feature vector $x_i(t)$, which is time-dynamic. here, $X^{T\times N\times k}$ is the feature matrix at time t, and k is the number of features.

The VaccHesitancy problem. Let $h_i(t)$ denote the fraction of patients indicating hesitancy in zip code i at time t (which will be specified later in Section IV); let $\mathbf{h}(t)$ denote the vector of hesitancy levels. The VaccHesitancy problem we involve learning $\mathbf{h}(T)$ using historic hesitancy levels $\mathbf{h}(t')$, demographic characteristics of zip codes in V, and the connectivity network G.

B. Framework

An analysis of the hesitancy levels $\mathbf{h}(t)$ over time from the dataset reveals spatial heterogeneity and correlations (presented in the Appendix), which motivates our GNN approach based on spatial structure. Our VH-GNN framework (Algorithm 1) has two major modules: 1) a spatial module, and 2) a sequence module (Figure 1). The spatial module consists of graph forming and graph-based spatial dependency learning, described in Section III-C.

C. Spatial Module

1) Graph Architecture.: The graph G is a static graph without self-loops. We propose three intuitive node connectivity mechanisms to form three variants of G.

- Geographic adjacency graph G_a: If node i and node j share a geographic boundary, they will form an edge in the network G. The weights on each edge is 1.
- Distance-based graph G_d: This is a fully connected graph where the weight on edge (i, j) has weight equal to the inverse of the distance between the centroids of the two end points.
- Mobility graph G_m: This graph represents population movement. We form G_m from an activity-based detailed population-level social contact network H_p [34]–[36]. The nodes in H_p represent individual people, and each node is associated with a location. We aggregate this network at the zip code level to get G_m. If there is any connection from the population of zip code i to zip code j, we form an edge in the graph G_m. This edge has an associated weight equal to the total number of connections from zip code i to j in the graph G_p.

All edge weights are normalized using min-max normalization.

2) Graph-based Spatial Dependency Learning.: Learning the spatial distribution of vaccine hesitancy with demographic characteristics is a major task. We leverage graph neural network (GNN) [37] to learn spatial dependency for each time step through message passing. In this work, node features change with time, but the graph connectivity is static. For each time step, we have one GNN module, which consists of stacking multiple k-GNN layers [38] and a linear layer to perform node-level vaccine hesitancy prediction. We model it as a regression task. The k-GNN is a generalization of graph neural networks based on the k-dimensional Weisfeiler-Leman algorithm (k-WL). This variant of GNN performs message passing directly between subgraphs instead of individual nodes. The l^{th} layer of the first-order GNN is

$$x^l(i,t) = \alpha\Bigg(x^{l-1}(i,t) \cdot W_1^l + \sum_{j \in \mathcal{N}(i)} a(i,j) \cdot x^{l-1}(j,t) \cdot W_2^l + b^l \Bigg) \quad (1)$$

Here, $l > 0$, α represents an activation function (e.g., sigmoid or ReLU). W_1, $W_2 \in \mathbb{R}^{d_{l-1} \times d_l}$ are weight matrices parametrizing GNN layer l, $b^l \in R^{d_l}$ is the parameters of the l-th layer, d_l is the dimension, and $\mathcal{N}(i)$ is the neighborhood of i. The output of the graph embedding module is an intermediate solution $y^{N \times d_{lo}}$ for each time point t. lo is the final layer of the graph embedding module. For T time steps, we merge the y matrix to form $Y^{T \times N \times d_{lo}}$.

D. Sequence Module

To learn the temporal aspect of the vaccine hesitancy for a node i, we use a sequence module. The input of this module is matrix Y, which will predict vaccine hesitancy at time T. This module can be built using any model that can learn sequences; popular choices are moving average, ARIMA, and recurrent neural networks. Recurrent neural networks (RNN) are well-known for predicting sequence data. In this work, we leverage a variant of RNN known as Gated Recurrent Units (GRU) [39]. We also experiment with LSTM (Long Short-Term Memory) and moving average. The fundamental concepts underlying LSTM and GRU models are quite similar. Both employ gated mechanisms to retain extensive long-term information, making them equally proficient for diverse tasks. We find that GRU performs better in the VH-GNN framework. It trains faster with fewer parameters compared to the LSTM variant. Hence, it has better potential to learn from large multidimensional datasets.

E. Optimization

This framework optimizes two modules separately with optimizers opt_1 and opt_2. We use two loss functions to reduce the error between predicted vaccine hesitancy $H(T)$ and the true vaccine hesitancy $\hat{H}(T)$. The first loss function $loss_1$ minimizes the error for the graph learning module at each time step $t = 0, 1, .., T$, and the second loss function $loss_2$ reduces errors for the sequence learning module. We use the mean absolute error (MAE) metric to learn the model parameters.

$$loss_1 = MAE(y(t) - \hat{y(t)}) + \lambda L_r \quad (2)$$

$$loss_2 = MAE(H(T) - \hat{H(T)}) + \lambda L_r \quad (3)$$

$loss_1$ takes into account all time steps, and $loss_2$ takes only the final time step. In this research problem, for any time step t, $\hat{y(t)} = \hat{H(t)}$, as the graph learning module is predicting vaccine hesitancy for that time step from $x_i(t)$. Here, λ is a hyper-parameter, and L_r is the $L2$ regularization term to prevent over-fitting. We optimize two modules separately as we do not want to influence one module's parameters due to the other module's performance. Algorithm 1 details the steps taken.

IV. Experiments and Results

A. Data

We use five years (2016-2020) of the All-Payer Claims Database (APCD) to find the patient refusal levels for each zip code in Virginia.

The data is obtained from VHI (Virginia Health Information). It contains information on paid medical and pharmacy claims for roughly 5 million Virginia residents with commercial, Medicaid, and Medicare coverage across all types of healthcare services. Among other things, it provides information on immunization rates over time, by spatial regions, and by demographics.

International Classification of Disease ICD-10-CM code Z28 is used to filter Patient refusal from medical data. Z28 means, "Immunization not carried out and underimmunization status" [40]. We also analyze the immunization rates as provided in the Virginia Department of Health School Immunization Survey (VDH SIS) reports. We find that vaccine hesitancy is changing in Virginia (Table III). However, we only use APCD for VaccHesitancy, as our targeted population is 0-6 years old, and VDH-SIS does not contain this information. In this work, we use six months as the time unit. We find

Algorithm 1 VH-GNN for Spatio-Temporal Vaccine Hesitancy Learning

Input: Feature matrix X, graph G, GNN module M_g, sequence module M_s, λ, number of training steps $train_s$, hyper-parameters

```
Split the nodes into two sets train, and test randomly.
Initialize model M_g and M_s with random weights and hyper-parameters.
Set optimizer opt_1 and opt_2 with hyper-parameters.
for number of training steps train_s do
    for for each time step t do
        if t < T then
            y(t) ← M_g(G, X(t))
        else
            y(t) ← M_g(G, X(t).train)
        end if
        compute loss_1 and update M_g using Adam optimizer
    end for
    Form Y from all y(t)
    H(T) ← M_s(Y)
    compute loss_2 and update M_s using Adam optimizer
end for
return loss_2
```

that monthly data is sparse at the zip code level. From a detailed Exploratory Data Analysis, we find that patient refusal and vaccine hesitancy is changing over time. From the spatial analysis, we observe that vaccine hesitancy of a location has similarity with its neighboring areas (Figure 2)

In this paper, 'zip code' refers to a ZIP Code Tabulation Area (ZCTA). A ZCTA corresponds to a geographical representation of a service area for a United States Postal Service (USPS) ZIP Code. This delineation is made publicly available by the US Census Bureau [41]. We use 615 zip codes of Virginia ($N = 615$) out of 1241. Among them, only about 52% zip codes have a population size of more than 1000. We discard zip codes that do not have any entry for kids (aged 0-6) in the APCD data or have a very small population size.

B. Data Preprocessing

From the APCD data, we filter all patients' entries of children aged six or below. Then we prepare a data set for each time t for N nodes, which keeps a record of the number of unique kids, the number of unique kids in different genders, the number of unique kids in different races, the number of unique kids in two different medical insurance types (commercial and Medicaid), and the percentage of kids who refuse to take any vaccine at least once. We use "medical insurance type" as a proxy for the income level. We assume that patient refusal at this age represents parental vaccine intention.

At each time step t, the last column of the data set is the target value; vaccine hesitancy $\hat{H(t)}$, other columns are features of nodes. A node i has ten features at a time step t, including male population, female population, Asian population, Black population, White population, and Hispanic population. We use min-max normalization to normalize all columns for each time step t as we update our M_s module. Then, we find the principal components of the features by using Principal Component Analysis (PCA).

C. Experimental Setup

This study uses the GNN version of Morris et al. [38]. We also explored other strategies, such as Graph Convolutional Network (GCN) [42], which can handle weighted static graphs. However, we find that the GNN of Morris et al. performs better in predicting vaccine hesitancy.

We manually adjust the hyper-parameters of the framework, such as the learning rate, the regularization term, the training epoch, and the number of hidden units. We use a learning rate of 0.0005, a training epoch of more than 10,000, and a learning rate of 10^{-4}. We find that using more than two GNN layers overfits the training data while using less than two introduces bias in the system. For module M_g, we experiment on hidden units [64,128,256,512]. For module M_s, we experiment on hidden units [8,16,32,64]. The setup for M_g with 128 hidden units and M_s with 16 hidden units performs better for the VH-GNN. For all GNN layers and GRU layers, we implement 50% dropout to avoid over-fitting.

We use Python 3.8 to implement the framework. We utilize the open-source deep-learning framework PyTorch version 2.0.0 and NVIDIA CUDA 11.4.2 in a Simple Linux Utility for Resource Management (SLURM) system.

D. Evaluation Metrics

The focus of the VACCHESITANCY study is node-level regression tasks. We use mean absolute percentage error (MAPE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) metrics to evaluate the performance of the VH-GNN framework in predicting spatio-temporal vaccine hesitancy at the node level. The well-known evaluation matrics ROC AUC, F1 score, recall, and precision are not appropriate for our estimation task as they are designed for classification tasks.

- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{|test|}\sum(H_{test}(T) - \hat{H_{test}}(T))^2} \quad (4)$$

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{|test|}\sum|H_{test}(T) - \hat{H_{test}}(T)| \quad (5)$$

MAE is only used during model training. For the test data set, $RMSE$ and MAE metrics are utilized to gauge the model's predictive performance. Smaller values of $RMSE$ and MAE indicate better prediction accuracy.

E. Baseline Methods

The performance of our combined graph framework is compared with the following baseline methods:

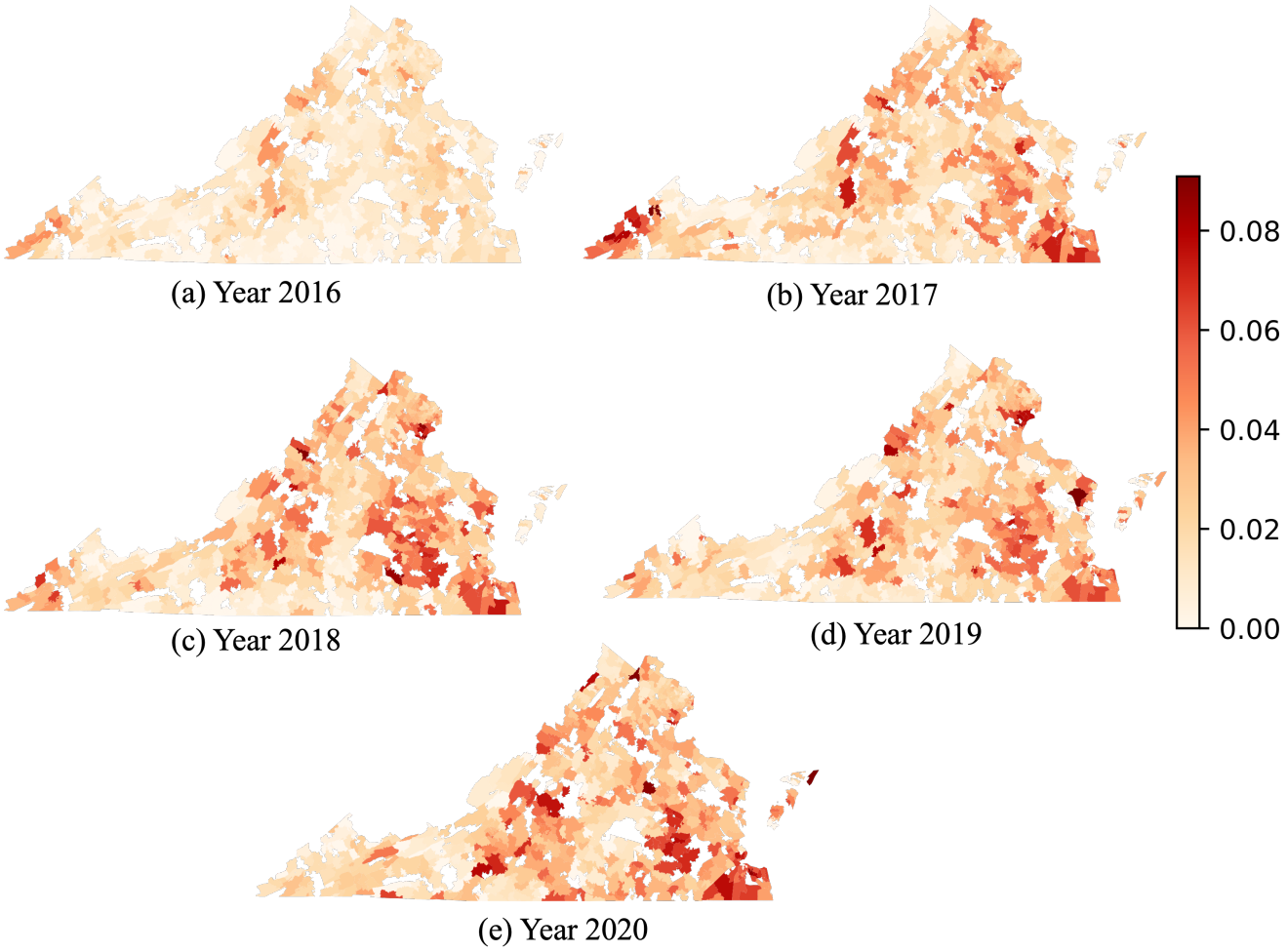


Fig. 2: Vaccine hesitancy rate among kids (age 0-6 years) at the zip-code level in Virginia from the APCD data.

TABLE I: Yearly patient refusal count in Virginia from APCD data.

	Count per year				
	2016	**2017**	**2018**	**2019**	**2020**
Refusal claim any age	92790	197185	249610	297220	261972
Unique refusal patient any age	27947	56844	77679	85003	69303
Refusal claim among age 0-6 years	28301	74506	81542	85935	86844
Unique refusal patient age 0-6 years	7312	15268	18340	18015	17263

TABLE II: Yearly patient refusal percentage in Virginia in the APCD data and in the VDH SIS reports.

Vaccine hesitancy	**Year**					**Source**
	2016	**2017**	**2018**	**2019**	**2020**	
Patient Refusal	1.55%	3.31%	4.01%	4.08%	3.83%	APCD
State Public School Unimmunized or not Adequately Immunized	3.6%	3.9%	3.7%	3.6%	11.9%	VDH SIS
State Private School Unimmunized or not Adequately Immunized	5.8%	6.7%	6.7%	4.8%	7.6%	VDH SIS

- **Linear Regression with Neighbors (LRN):** We first evaluate our model's performance against the Linear Regression approach, where this baseline method is used to predict the dependent variable $H(T)$ using input features. Linear regression fits a linear model to minimize the residual sum of squared differences between true and predicted values using linear approximation. To make a fair comparison, we provide an extra feature, neighbor's information, for each node i, we calculate $\sum_{j \in \mathcal{N}(i)} h_j(t) a(i,j)$.
- **Multi-layer Perceptron (MLP):** We also experiment with this second benchmark method, Multi-layer Perceptron (MLP), to assess our model's effectiveness. This feedforward neural network architecture is capable of capturing complex non-linear relationships within data.
- **Graph Convolutional Network (GCN):** GCN is our third benchmark method. It uses convolutional architecture to capture both local and global patterns within graph-structured data for semi-supervised learning.
- **Graph Convolutional Network & Gated Recurrent Units (GCN-GRU):** In this fourth benchmark, we replace M_s module with the GCN in the VH-GNN.

Figure 3 shows the mean output performances of all the baselines compared to the VH-GNN framework for the target years 2019 and 2020. The same $train$ and $test$ data sets were used for all models. We always use a batch gradient process to update the model parameters. We employ model-specific hyperparameters to unlock their full potential. Results indicate that VH-GNN outperforms all baselines in both evaluation metrics. It is also evident that the VH-GNN performed better for the year 2019 compared to 2020.

The results are also affected by the choice of node connectivity mechanisms. Table IV shows the performance of VH-GNN using three G variants for 2019. The VH-GNN performs the best when mobility graph G_m is used. Hence, the remaining set of results in this paper for VH-GNN is produced using G_m.

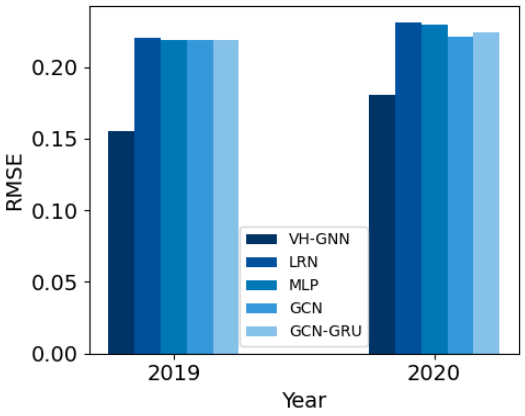
Figure 4 shows predicted values vs true values for 2019. The R^2 value is 0.61.

F. Ablation Study

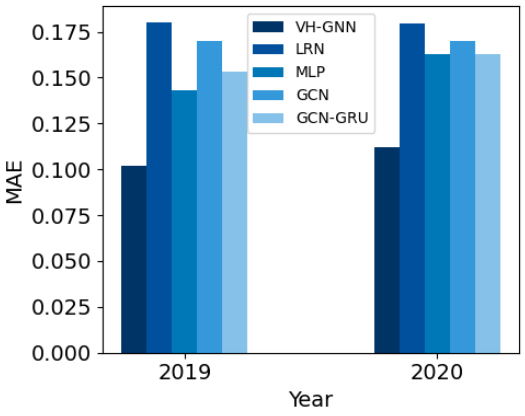
We conduct an ablation study to understand the role of two modules of the VH-GNN.

- **VH-GNN w/o M_s module:** In this setup, we only train the spatial learning module M_g with the $loss_1$ function and evaluate the prediction performance only using M_g.
- **VH-GNN w/o M_g module:** In this setup, we only keep the sequence learning module, which is the GRU model. Here, we train M_s by using $\hat{H_{test}}$ for $t = 0$ to $t = T - 1$ and $\hat{H_{train}}$ for $t = 0$ to $t = T$. It does not consider any graph structure.

Table V shows the prediction performance of the two modules. Ablation study shows the importance of spatial learning for node-level vaccine hesitancy forecasting. Although our combined framework performs better than either of these configurations, the VH-GNN w/o M_s outperforms VH-GNN w/o M_g. This indicates the importance of the spatial component in explaining vaccine hesitancy.



(a)



(b)

Fig. 3: A comparative analysis between our VH-GNN and the baseline methods across 2019 and 2020. Figure 3a and 3b show that our method outperforms alternative approaches, as evident in both the RMSE and MAE metrics.

G. Performance Analysis at the Node Level

The properties of nodes were investigated where VH-GNN performed poorly. For a comparative analysis, the nodes are divided into two sets, one where VH-GNN performs poorly and the other where it performs well. The test nodes are sorted according to the absolute error between predicted and true vaccine hesitancy values, and the nodes were divided into two sets:

- **Nodes with Large Error, V_L:** Top 25% nodes, nodes with large error, where VH-GNN did not perform well.
- **Nodes with Small Error, V_S:** Rest of the test nodes, where VH-GNN performs well.

We investigated features of two sets: V_L and V_S, to see why the VH-GNN framework does not predict well. We find that population sizes, vaccine hesitancy percentages, population percentages with Medicaid insurance, and Hispanic populations differ between these two sets. Table VI reports the

TABLE III

	T+1					T+4				
	R^2	MAE	MSE	MAPE	RMSE	R^2	MAE	MSE	MAPE	RMSE
GCN-LSTM	**0.8813**	**0.0276**	**0.00259**	**29.7731**	**0.05089**	**0.87117**	**0.02940**	**0.00281**	33.1513	**0.05304**
GCN-GRU	0.85914	0.03034	0.00307	31.5569	0.05546	0.8215	0.03283	0.00389	35.183	0.06242
GCN	0.8343	0.031510	0.003616	35.1869	0.06014	0.8358	0.031065	0.00358	**33.1212**	0.059878
LSTM	0.8094	0.03403	0.004161	36.6645	0.06451	0.75083	0.0388	0.00544	41.8586	0.07376
LR	0.68813	0.04536	0.00681	48.871	0.08252	0.66685	0.04751	0.0072	51.190	0.08529
MLP	0.78269	0.03619	0.004747	38.9923	0.06890	0.7024	0.04423	0.00649	45.7539	0.08016

TABLE IV: Prediction performance of VH-GNN across three graphs connectivity mechanisms for the year 2019.

Graph	**RMSE**	**MAE**
Adjacent Graph G_a	0.2047	0.1487
Distance-based Graph G_d	0.1816	0.1342
Mobility Graph G_m	**0.1554**	**0.1019**

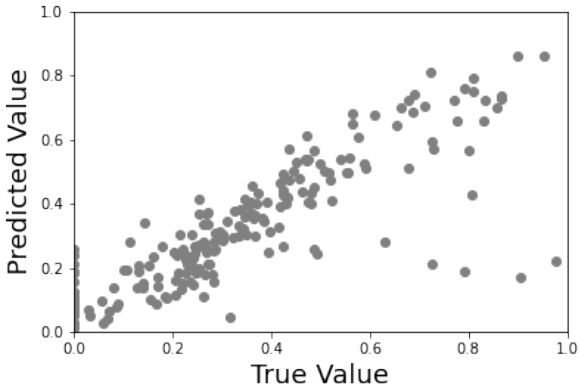


Fig. 4: Predicted vaccine hesitancy and the true vaccine hesitancy in the $test$ set for year 2019.

TABLE V: Ablation study on two setups for the year 2019, 1) VH-GNN w/o M_s module, and 2) VH-GNN w/o M_g module.

Ablation Study Setting	**RMSE**	**MAE**
VH-GNN w/o M_s module	0.2054	0.1434
VH-GNN w/o M_g module	0.4557	0.3820

TABLE VI: Average values of significant features in the set V_L (nodes with large errors) and V_S (nodes with small errors).

Features	V_L	V_S
Kids Population	311.62	873.24
Vaccine Hesitancy Percentage	0.026	0.033
Population Percentage with Medicaid	0.647	0.609
Hispanic Population Percentage	0.006	0.016

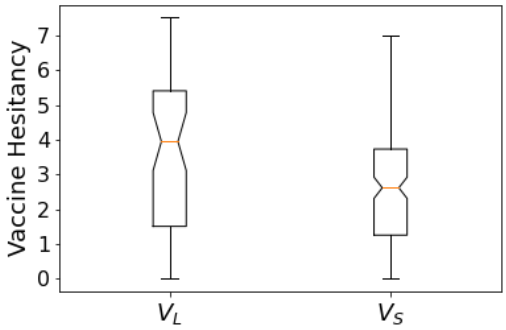


Fig. 5: Vaccine hesitancy percentage in two sets, V_L and V_S.

average of these features for V_L and V_S sets. Table VI shows that the kids' population sizes are significantly different across V_L and V_S. The VH-GNN is prone to have large predictive errors for nodes that have a small population with a high vaccine hesitancy level. Further investigation in Figure 5 shows that VH-GNN also performs well when the vaccine hesitancy percentage is high.

H. Forecasting Performance

We test the VH-GNN as a vaccine hesitancy forecasting tool. For this purpose, we train the VH-GNN until the $T-1$ time step, then we use VH-GNN to forecast vaccine hesitancy percentages for all zip codes at time T. Figure 6 shows the forecast vaccine hesitancy percentage. The mean RMSE and MAE value for this case is 0.1602 and 0.1115.

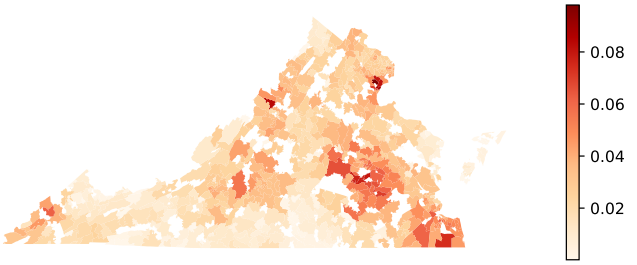


Fig. 6: Vaccine hesitancy forecast for all the zip codes of Virginia in the first half of 2019. Here, colors represent vaccine hesitancy percentages.

I. Evaluation Metrics to Understand Spatial Structure

Understanding spatial structure is essential for the VACCHESITANCY problem. We evaluate the performance of the VH-GNN in capturing spatial structure by using two following clustering measures:

- **Moran's-I:** It is the measure of global spatial autocorrelation. This value ranges from -1 to 1, with 0 indicating no autocorrelation; -1 indicating perfect clustering with dissimilar values, such as clustering of high vaccine hesitancy location with the low vaccine hesitancy location; and 1 indicating perfect clustering with similar values, such as clustering of high vaccine hesitancy locations with high vaccine hesitancy locations. The equation for the Moran's-I for a time t is

$$\frac{N}{\sum_i \sum_j a(i,j)} \frac{\sum_i \sum_j a(i,j)(h_i - \bar{H})(h_j - \bar{H})}{\sum_i (h_i - \bar{H})^2} \quad (6)$$

- **Isolation Index:** It indicates the level of segregation within a specific group or cluster compared to the larger population, with values ranging from 0 (no segregation) to 1 (full segregation). The equation for the isolation index for a time t is

$$\sum_{i=1}^{N} [(\frac{h_i p_i}{\sum_{i=1}^{N} h_i p_i}) h_i] \quad (7)$$

Our predicted value from the VACCHESITANCY produces a Moran's I value of 0.8488 for the year 2019, and the actual true value results in a value of 0.4580. The predicted value and the true value both finds that individuals exhibiting higher levels of vaccine hesitancy are more likely to be situated closely in G_m.

The calculated isolation index from the predicted value and the true value is 0.0466 and 0.0480 for the year 2019, both indicates almost no segregation.

J. Spatio-temporal clustering

Vaccine hesitancy shows a spatial clustering phenomenon. However, the cluster shape and sizes change over time. The statistically significant clusters of higher vaccine hesitant zip codes for five years are presented by dark color in Figure 7. We use a graph-based scan statistics method to identify statistically significant geographical clusters of higher vaccine hesitancy [43]. The scan statistics method is a hypothesis-testing approach for anomaly detection, used in previous studies to detect hotspots and anomalies in spatial distributions [44]. We use a modified Kulldroff's scan statistics method to find statistically significant clusters of zip codes with higher vaccine hesitancy in the adjacent graph G_a.

A cluster $C \subset G_a$ of zip codes in the adjacent network G_a can have an arbitrary shape. We calculate the scan statistic or score function of a cluster of zip codes C as $F(C) = \frac{Pr[Data|H_1(C)]}{[Data|H_0]}$ which is a likelihood ratio of the probability of the observed data (i.e., a certain level of vaccine hesitancy in C) generated under an alternative hypothesis $H_1(C)$, to the probability of the observations under the null hypothesis H_0. We use the Poisson version of the Kulldorff scan statistic, which assumes that the observations are generated from Poisson distribution. The null hypothesis H_0 is generated proportionally from the baseline count $(1-\mu)P_i$, where μ is the state-wide vaccine hesitancy rate and P_i is kids population in a zip code i. The alternative hypothesis of a cluster $H_1(C)$ counts the vaccine hesitancy among nodes outside C; in $V_z - C$, the hesitancy count comes from a rate proportional to the baseline counts. But, for the nodes within C, the counts are generated at a higher rate than expected. We use the Monte Carlo sampling approach to compute the p-value for each cluster. Maximizing $F(C)$ is the objective function. The general dynamic programming method allows us to optimize a large class of parametric and non-parametric scan statistics [45]. We use 0.9 as the cut-off p-value to consider significant clusters.

V. CONCLUSIONS

The VH-GNN framework is able to predict the spatio-temporal aspects of vaccine hesitancy with a combined GNN and RNN structure. Our method crucially uses a very large all payers insurance dataset, and a detailed activity-based synthetic contact network. Our method outperforms several baseline methods in predicting vaccine hesitancy at a zip code level, in terms of the RMSE and MAE evaluation metrics.

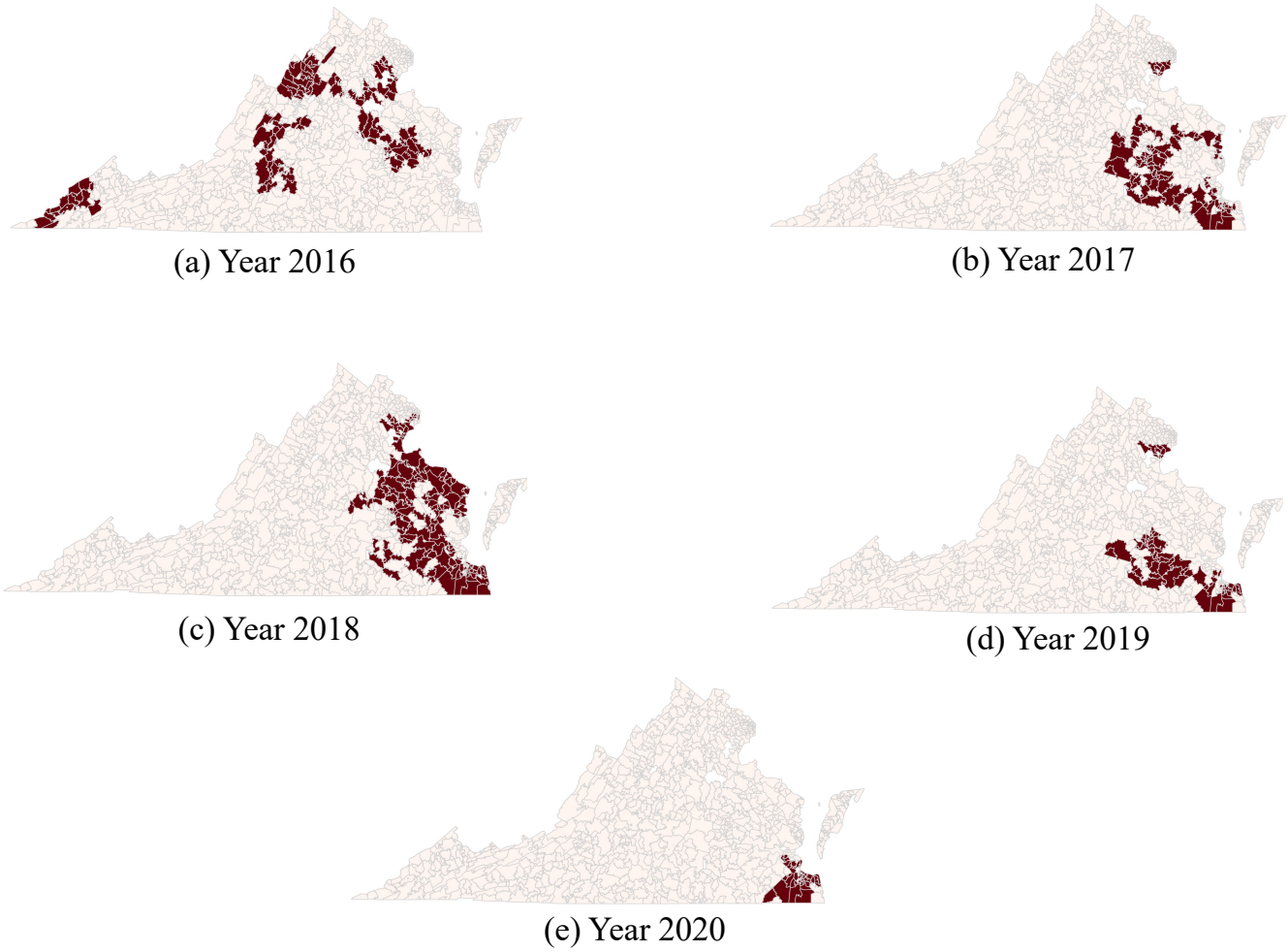
We also demonstrate the model's effectiveness at the node-level data, highlighting the challenges in learning vaccine hesitancy for smaller populations. Although GRU is well-known to handle sequential data, they are computationally expensive and require a lot of data to train. We find that GRU or neural network alone cannot predict vaccine hesitancy at a zip code level. However, a combination of GNN and GRU can learn the spatial and temporal aspects of vaccine hesitancy and can predict patient refusal at a higher spatial resolution.

ACKNOWLEDGMENT

The authors would like to thank members of the Network Systems Science and Advanced Computing Division in the Biocomplexity Institute at the University of Virginia (UVA) for useful discussion and supports.

REFERENCES

[1] Centers for Disease Control, "Measles history," https://www.cdc.gov/measles/about/history.html, 2020.



(a) Year 2016

(b) Year 2017

(c) Year 2018

(d) Year 2019

(e) Year 2020

Fig. 7: Clusters of higher vaccine hesitancy regions in Virginia.

[2] M. Patel, A. D. Lee, N. S. Clemmons, S. B. Redd, S. Poser, D. Blog, J. R. Zucker, J. Leung, R. Link-Gelles, H. Pham *et al.*, "National update on measles cases and outbreaks—united states, january 1–october 1, 2019," *Morbidity and Mortality Weekly Report*, vol. 68, no. 40, p. 893, 2019.

[3] R. Sato, O. A. Makinde, K. C. Daam, and B. Lawal, "Geographical and time trends of measles incidence and measles vaccination coverage and their correlation in nigeria," *Human Vaccines & Immunotherapeutics*, vol. 18, no. 6, p. 2114697, 2022.

[4] W. H. Organization, "Measles fact sheet," 2023, https://www.who.int/news-room/fact-sheets/detail/measles.

[5] M. Thakur, R. Zhou, M. Mohan, A. Marathe, J. Chen, S. Hoops, D. Machi, B. Lewis, and A. Vullikanti, "Covid's collateral damage: likelihood of measles resurgence in the united states," *BMC Infectious Diseases*, vol. 22, no. 1, p. 743, 2022.

[6] R. Seither, K. Calhoun, O. B. Yusuf, D. Dramann, A. Mugerwa-Kasujja, C. L. Knighton, and C. L. Black, "Vaccination coverage with selected vaccines and exemption rates among children in kindergarten—united states, 2021–22 school year," *Morbidity and Mortality Weekly Report*, vol. 72, no. 2, p. 26, 2023.

[7] T. A. Lieu, G. T. Ray, N. P. Klein, C. Chung, and M. Kulldorff, "Geographic clusters in underimmunization and vaccine refusal," *Pediatrics*, vol. 135, no. 2, pp. 280–289, 2015.

[8] J. Cadena, D. Falcone, A. Marathe, and A. Vullikanti, "Discovery of under immunized spatial clusters using network scan statistics," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 28, 2019.

[9] P. Gahr, A. S. DeVries, G. Wallace, C. Miller, C. Kenyon, K. Sweet, K. Martin, K. White, E. Bagstad, C. Hooker *et al.*, "An outbreak of measles in an undervaccinated community," *Pediatrics*, pp. peds–2013, 2014.

[10] P. A. Gastañaduy, J. Budd, N. Fisher, S. B. Redd, J. Fletcher, J. Miller, D. J. McFadden, J. Rota, P. A. Rota, C. Hickman, B. Fowler, L. Tatham, G. S. Wallace, S. de Fijter, A. P. Fiebelkorn, and M. DiOrio, "A measles outbreak in an underimmunized Amish community in Ohio," *The New England journal of medicine*, vol. 375, pp. 1343–1354, Oct. 2016.

[11] R. Seither, J. Laury, A. Mugerwa-Kasujja, C. L. Knighton, and C. L. Black, "Vaccination coverage with selected vaccines and exemption rates among children in kindergarten—united states, 2020–21 school year," *Morbidity and Mortality Weekly Report*, vol. 71, no. 16, p. 561, 2022.

[12] K. Causey, N. Fullman, R. J. Sorensen, N. C. Galles, P. Zheng, A. Aravkin, M. C. Danovaro-Holliday, R. Martinez-Piedra, S. V. Sodha, M. P. Velandia-González *et al.*, "Estimating global and regional disruptions to routine childhood vaccine coverage during the covid-19 pandemic in 2020: a modelling study," *The Lancet*, vol. 398, no. 10299, pp. 522–534, 2021.

[13] G. Iacobucci, "Measles is now "an imminent threat" globally, who and cdc warn," 2022.

[14] G. Guglielmi, "Pandemic drives largest drop in childhood vaccinations in 30 years." *Nature*, pp. 253–253, 2022.

[15] S. McGregor and R. D. Goldman, "Determinants of parental vaccine hesitancy," *Canadian Family Physician*, vol. 67, no. 5, pp. 339–341, 2021.

[16] N. B. Masters, M. C. Eisenberg, P. L. Delamater, M. Kay, M. L. Boulton, and J. Zelner, "Fine-scale spatial clustering of measles nonvaccination that increases outbreak potential is obscured by aggregated reporting

data," *Proceedings of the National Academy of Sciences*, vol. 117, no. 45, pp. 28 506–28 514, 2020.
[17] S. A. Moon, A. Marathe, and A. Vullikanti, "Are all underimmunized measles clusters equally critical?" *medRxiv*, pp. 2023–04, 2023.
[18] V. D. of Health, "Virginia student immunization status survey," https://www.vdh.virginia.gov/immunization/sisresultsarchived/, 2021, [Online; accessed 10-Nov-2022].
[19] J. Müller, A. Tellier, and M. Kurschilgen, "Echo chambers and opinion dynamics explain the occurrence of vaccination hesitancy," *Royal Society Open Science*, vol. 9, no. 10, p. 220367, 2022.
[20] Q. C. Nguyen, I. Yardi, F. X. M. Gutierrez, H. Mane, and X. Yue, "Leveraging 13 million responses to the us covid-19 trends and impact survey to examine vaccine hesitancy, vaccination, and mask wearing, january 2021-february 2022," *BMC Public Health*, vol. 22, no. 1, pp. 1–15, 2022.
[21] V. Carrieri, R. Lagravinese, and G. Resce, "Predicting vaccine hesitancy from area-level indicators: A machine learning approach," *Health Economics*, vol. 30, no. 12, pp. 3248–3256, 2021.
[22] S. Chandir, D. A. Siddiqi, O. A. Hussain, T. Niazi, M. T. Shah, V. K. Dharma, A. Habib, A. J. Khan *et al.*, "Using predictive analytics to identify children at high risk of defaulting from a routine immunization program: feasibility study," *JMIR public health and surveillance*, vol. 4, no. 3, p. e9681, 2018.
[23] A. Bell, A. Rich, M. Teng, T. Orešković, N. B. Bras, L. Mestrinho, S. Golubovic, I. Pristas, and L. Zejnilovic, "Proactive advising: a machine learning driven approach to vaccine hesitancy," in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2019, pp. 1–6.
[24] J. E. Atwell, J. Van Otterloo, J. Zipprich, K. Winter, K. Harriman, D. A. Salmon, N. A. Halsey, and S. B. Omer, "Nonmedical vaccine exemptions and pertussis in california, 2010," *Pediatrics*, vol. 132, no. 4, pp. 624–630, 2013.
[25] E. O. Nsoesie, R. J. Beckman, S. Shashaani, K. S. Nagaraj, and M. V. Marathe, "A simulation optimization approach to epidemic forecasting," *PloS one*, vol. 8, no. 6, p. e67164, 2013.
[26] A. Mollalo and M. Tatar, "Spatial modeling of covid-19 vaccine hesitancy in the united states," *International journal of environmental research and public health*, vol. 18, no. 18, p. 9488, 2021.
[27] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.
[28] K. Klemmer, N. S. Safir, and D. B. Neill, "Positional encoder graph neural networks for geographic data," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2023, pp. 1379–1389.
[29] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
[30] M. Li and Z. Zhu, "Spatial-temporal fusion graph neural networks for traffic flow forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 5, 2021, pp. 4189–4196.
[31] L. Wang, A. Adiga, J. Chen, A. Sadilek, S. Venkatramanan, and M. Marathe, "Causalgnn: Causal-based graph neural networks for spatio-temporal epidemic forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 11, 2022, pp. 12 191–12 199.
[32] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *arXiv preprint arXiv:1709.04875*, 2017.
[33] W. Hu, W. Li, X. Zhou, A. Kawai, K. Fueda, Q. Qian, and J. Wang, "Spatio-temporal graph convolutional networks via view fusion for trajectory data analytics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4608–4620, 2022.
[34] S. Eubank, H. Guclu, V. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.
[35] C. L. Barrett, R. J. Beckman, M. Khan, V. A. Kumar, M. V. Marathe, P. E. Stretz, T. Dutta, and B. Lewis, "Generation and analysis of large synthetic social contact networks," in *Proceedings of the 2009 Winter Simulation Conference (WSC)*. IEEE, 2009, pp. 1003–1014.
[36] J. Cadena, D. Falcone, A. Marathe, and A. Vullikanti, "Discovery of under immunized spatial clusters using network scan statistics," *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1–14, 2019.
[37] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *arXiv preprint arXiv:1709.05584*, 2017.
[38] C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, and M. Grohe, "Weisfeiler and leman go neural: Higher-order graph neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 4602–4609.
[39] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 9, pp. 3848–3858, 2019.
[40] ICD10data, "Icd-10-cm codes," https://www.icd10data.com/ICD10CM/Codes/Z00-Z99/Z20-Z29/Z28-, 2023, accessed: 2023-1-06.
[41] U.S. Census Bureau, "ZIP Code Tabulation Areas (ZCTAs)," 2021, [Online; accessed 11-Nov-2022]. [Online]. Available: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html
[42] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
[43] S. Afroj Moon, A. Marathe, and A. Vullikanti, "Are all underimmunized measles clusters equally critical?" *Royal Society Open Science*, vol. 10, no. 8, p. 230873, 2023.
[44] L. Duczmal, M. Kulldorff, and L. Huang, "Evaluation of spatial scan statistics for irregularly shaped clusters," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 428–442, 2006.
[45] J. Cadena, F. Chen, and A. Vullikanti, "Near-optimal and practical algorithms for graph scan statistics," in *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 2017, pp. 624–632.