Exploring Graph Structure in Graph Neural Networks for Epidemic Forecasting

Ching-Hao Fan[1*] **Sai Supriya Varugunda**[1*] **Lijing Wang**[1]
[1] New Jersey Institute of Technology
{cf322, sv247, lijing.wang}@njit.edu

Abstract

Graph neural networks (GNNs) that incorporate cross-location signals have the ability to capture spatial patterns during infectious disease epidemics, potentially improving forecasting performance. However, these models may be susceptible to biases arising from mis-specification, particularly related to the level of connectivity within the graph (i.e., graph structure). In this paper, we investigated the impact of graph structure on GNNs for epidemic forecasting. Multiple graph structures are defined and analyzed based on several characteristics i.e., dense or sparse, dynamic or static. We design a comprehensive ablation study and conduct experiments on real-world data. One of the major findings is that sparse graphs built using geographical information can achieve advanced performance and are more generalizable among different tasks compared with more complex attention-based adjacency matrices.

1 Introduction

Epidemic diseases, such as seasonal flu or COVID-19, have placed a heavy social and economic burden on our society. For instance, the COVID-19 pandemic has caused over 770 million confirmed cases and over 6.9 million deaths globally[2]. Timely and accurate spatial and temporal forecasts of the epidemic dynamics is particularly crucial for developing effective interventions and marshaling limited healthcare resources. Existing methodologies for forecasting include 1) Mechanistic causal methods [2, 7, 17]; 2) Data-driven methods including statistical methods [8, 11, 13] and 3) deep learning methods [4, 3, 12]. Given the spatial and temporal variations in the epidemic dynamics, interests in graph neural networks (GNNs) have gained as they provide a framework for incorporate cross-location signals for spatiotemporal epidemic forecasting [5, 16, 9, 6, 15].

GNNs considering cross-location signals can model the impact of one location on other locations during the epidemics of infectious disease, which can lead to improved forecasting performance, but could be impacted by model mis-specification biases, especially the level of connectivity in a graph (i.e., graph structure). The connectivity of two locations can be represented by their geographic distance. However, non-adjacent areas may also have potential dependencies due to human mobility activity. Thus, beyond geographical adjacency information, model-generated adjacency information is adopted to estimate the spatiotemporal correlations of two locations, e.g., gravity network [18]. During COVID pandemic, mobility flow is often used [9, 1, 16] due to that it is real and dynamic, however, it is usually not publicly accessible. During the COVID-19 pandemic, attention mechanism [14] is widely used to learn implicit dynamic correlations of any two locations and has shown improved performance in forecasting disease dynamics [5, 6, 15]. The empirical results have shown that attention adjacency which is dynamic and dense often performs better than geographical adjacency which is static and sparse. Despite the superior performance of attention-based GNNs presented

[*] co-first authors
[2] Source: https://covid19.who.int/ as of September 27, 2023

Temporal Graph Learning Workshop @ NeurIPS 2023, New Orleans.

in the previous works, it is yet to investigate the impact of graph structure on GNNs for epidemic forecasting in a comprehensive manner. More interestingly, Zhang et al. in [19] proved that graph sparsification with importance sampling can improve GNNs' learning performance, which seems contradictory to the empirical results.

In this paper, we try to investigate the following questions while using GNNs to forecast epidemic dynamics: 1) what's the impact of dynamic/static network structures for capturing the spatial correlations of disease dynamics? 2) is the geographical information important for capture the disease dynamics? 3) does attention capture implicit spatial patterns of disease dynamics for future predictions? 4) sparse or dense graph, which one is better? We answer the questions through comprehensive experiments on real-world epidemic data.

Contributions To the best of our knowledge, this paper provides the first comprehensive study to examine the influence of graph structure of GNN models on epidemic forecasting performance. (1) We define multiple graph structures characterized by several properties, i.e., dense or sparse, geography or attention. (2) We design a comprehensive ablation study to explore the impact of different graph structures. We also investigate sparse learning of GNNs for epidemic forecasting. (3) We conduct experiments and discuss results using real-world epidemic data. (4) Interesting findings are found and analyzed.

2 Preliminaries

Pre-existing GNN models To investigate the influence of graph structure on GNN predictive models, we adopt a pre-existing GNN model [5] proposed for influenza-like-illness (ILI) forecasting using both geographical adjacency and attention coefficients to build the graph. This model embeds the input time series of each node via long-short-term-memory (LSTM) layers and use the embedding to learn attention matrix which later is combined with geographical adjacency matrix. The temporal embedding is propagated over the graph using the learned location-aware attention matrix on a two-layer GNN. The details of the model can be found in [5]. We keep the model framework but modify the location-aware attention module to consider different mechanism for building graph adjacency matrix. This model is used as our base model. However, the experimental design and analysis can be applied to any GNN-based models.

Sparse learning for improved GNNs The previous empirical results demonstrate that GNN with attention adjacency (dense graph) can capture implicit spatial patterns than geographical adjacency (sparse graph) leading to improved epidemic forecasting performance. However, a recent study [19] theoretically analyzes sparse learning on two-layer GNNs for classification tasks. The insights of this study indicate that graph sparsification and importance sampling can improve GNN performance with a guarantee. This is contrary to the empirical results. We will design multiple dense and sparse graphs on our base model to explore sparse learning for time series epidemic forecasting using GNNs.

3 Exploring Graph Structure in GNNs

3.1 Problem formulation

We formulate the epidemic forecasting problem as a multivariate time series forecasting problem using a GNN model. We assume N locations in total and define a graph on the N locations as $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of N nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. The nodes are connected via an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, where an element a_{ij} represents the connectivity level between node i and node j. $a_{ij} = 0$ means no connection between two nodes, otherwise, they are connected by an edge with weight a_{ij}. Note that $\mathbf{A}$ can be an asymmetric matrix, i.e., a_{ij} is not necessary to be equal to a_{ji}. A node is associated with a time series of disease dynamics, e.g., the COVID-19 confirmed case counts for the past T days. For any node $v \in \mathcal{V}$, the node's hidden feature after one graph convolutional network (GCN) layer is updated by aggregating neighbor's features where neighbors are determined by adjacency matrix $\mathbf{A}$, i.e., weighted sum of connected nodes' hidden features. The forecasting objective is to predict an epidemiological target (e.g., COVID-19 counts) at future time $T + h$ for N regions where h denotes the horizon time.

In this paper, our aim is to examine the impact of different adjacency matrices on model performance. More specifically, we define adjacency matrix $\mathbf{A}$ using: 1) *sparse* + *static* geographical adjacency $\mathbf{S}$;

Table 1: Notations of adjacency matrix and their descriptions

Notation	Description	State	Density
$\mathbf{S}$	Geographical adjacency matrix. $s_{ij} = 1$ if location i is geographically adjacent to location j; otherwise, $s_{ij} = 0$.	Static	Sparse
$\mathbf{R}$	Randomly shuffled geographical adjacency matrix. The number of $r_{ij} = 1$ is the same with $\mathbf{S}$.	Static	Sparse
$\mathbf{V}$	Adjacency matrix obtained by a gravity model [18]. $v_{ij} = \frac{p_i p_j}{(l_{ij}+1)^2}$ and row normalized where p_i,p_j are the population size of location i and j, l_{ij} denotes as harversine distance. Refer our git repository for the details.	Static	Dense
$\mathbf{D}$	Adjacency matrix of an unweighted complete graph. $d_{ij} = 1, \forall i, j$.	Static	Dense
$\mathbf{B}$	Learned attention matrix. b_{ij} is learned through GNN training and row normalized.	Dynamic	Dense
$\mathbf{M}$	Graphical + learned attention adjacency matrix. $\mathbf{M} = f(\mathbf{S}, \mathbf{B})$ and $f(\cdot)$ is defined by Equation (6) in [5].	Dynamic	Dense
$\mathbf{C}$	Graphical + learned attention adjacency matrix. $\mathbf{C} = \mathbf{B} \odot \mathbf{S}$ where $\odot$ denotes element-wise product.	Dynamic	Sparse

2) *sparse* + *static* randomly shuffled geographical adjacency $\mathbf{R}$; 3) *dense* + *static* adjacency matrix of a weighted complete graph $\mathbf{V}$, where the weight is computed using a gravity model[18] to estimate the static mobility flow between two locations; 4) *dense* + *static* adjacency matrix of an unweighted complete graph $\mathbf{D}$; 5) *dense* + *dynamic* learned attention adjacency $\mathbf{B}$; 6) *dense* + *dynamic* learned attention adjacency combined with graphical adjacency $\mathbf{M}$; 7) *sparse* + *dynamic* learned attention adjacency truncated by geographical adjacency $\mathbf{C}$; A more detailed description is shown in Table 1.

3.2 Ablation study design

To answer the questions proposed in section 1, we carefully design an ablation study on those adjacency matrices using our base GNN model. It is to be noted that the base GNN model is the same for all ablations except for the adjacency matrix to aggregate the node's features, i.e., $\mathbf{A}$ is replaced by different adjacency matrices from Table 1. Note that we consider GNN models using geographical adjacency $\mathbf{S}$ as the baseline since this is the most traditional graph structure for epidemic forecasting.

I. To examine whether geographical information is important in constructing GNN models, we compare models using geographical adjacency $\mathbf{S}$ with models using randomly shuffled geographical adjacency $\mathbf{R}$. The motivation of using $\mathbf{R}$ is to remove geographical information while keep the same graph complexity with $\mathbf{S}$.

II. To explore the impact of dynamic and static network structures for capturing the spatial correlations of disease dynamics, we compare models using attention adjacency $\mathbf{B}$, gravity adjacency $\mathbf{V}$, and geographical adjacency $\mathbf{S}$.

III. To investigate sparse learning of GNNs for time series epidemic forecasting, we compare models using dense unweighted adjacency $\mathbf{D}$ with models using sparse unweighted adjacency $\mathbf{S}$ and $\mathbf{R}$. $\mathbf{S}$ is a sparse graph with geo-based sampling, and $\mathbf{R}$ is a sparse graph with random sampling.

IV. The combination of static geographical adjacency $\mathbf{S}$ and dynamic attention adjacency $\mathbf{B}$ is explored using $\mathbf{M}$ (a dynamic strategy) and $\mathbf{C}$ (a static strategy). $\mathbf{C}$ is a sparse geo-attention graph, while $\mathbf{M}$ is a dense geo-attention graph.

V. To align the learned attention-based adjacency matrix with the actual geographical adjacency, we add a reconstruction loss term $\mathcal{L}_g = \|\mathbf{A} - \mathbf{S}\|^2$ (introduced in [10]) to the prediction error loss $\mathcal{L}_e$ of the base GNN model, thus $\mathcal{L} = \mathcal{L}_e + \lambda \mathcal{L}_g$. $\mathcal{L}_g$ performs like a geo-based regularization term penalizing on the complexity of the adjacency matrix $\mathbf{A}$. Here λ is a hyperparameter that used to scale the regularization term. We add $\mathcal{L}_g$ when using $\mathbf{M}$ and $\mathbf{B}$.

3.3 Experiment settings

Data and metrics We use two real-world datasets for experiments: 1) **ILI**: weekly ILI positive case counts at US state level, collected from CDC ILINet[3] ranging from week 40, 2010 to week 12, 2023; 2) **COVID**: daily COVID-19 confirmed case counts at US state level, collected from JHU COVID-19 repository[4], ranging from March 2020 to March 2023. There are 651 ILI time points and

[3]https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html

[4]https://github.com/CSSEGISandData/COVID-19

Table 2: Forecasting performance on COVID-19 daily new confirmed cases and ILI weekly cases at US state level. The performance is reported as the average of 3 random trials with standard deviation.

	ILI						COVID					
	RMSE			PCC			RMSE			PCC		
Graph	1	3	5	1	3	5	1	7	14	1	7	14
$G_{\mathbf{S}}$	375.82 (±19.25)	557.71 (±9.62)	688.16 (±30.11)	0.9015 (±.014)	0.7227 (±.008)	0.4461 (±.101)	1712.37 (±101.41)	1137.98 (±4.48)	1195.12 (±6.98)	0.5306 (±.072)	0.7388 (±.000)	0.7247 (±.001)
$G_{\mathbf{R}}$	401.36 (±6.53)	581.85 (±100.45)	753.69 (±157.96)	0.8741 (±.008)	0.6347 (±.167)	0.3888 (±.156)	1920.07 (±292.58)	1233.03 (±2.56)	1215.98 (±8.46)	0.3367 (±.222)	0.7056 (±.000)	0.7205 (±.001)
$G_{\mathbf{V}}$	393.41 (±9.27)	588.95 (±47.75)	711.41 (±20.71)	0.8871 (±.016)	0.6593 (±.035)	0.4191 (±.075)	1839.37 (±162.30)	1159.29 (±15.84)	1237.18 (±31.33)	0.3490 (±.192)	0.7373 (±.002)	0.7226 (±.002)
$G_{\mathbf{D}}$	394.45 (±19.67)	575.76 (±91.86)	616.54 (±12.53)	0.8690 (±.869)	0.8405 (±.008)	0.6389 (±.020)	1867.91 (±164.56)	1170.59 (±38.21)	1264.63 (±1.81)	0.2979 (±.154)	0.7247 (±.018)	0.7005 (±.000)
$G_{\mathbf{B}}$	366.70 (±20.29)	588.91 (±47.69)	710.67 (±13.29)	0.9042 (±0.01)	0.6508 (±.058)	0.3990 (±.043)	1764.99 (±35.66)	1160.16 (±3.22)	1263.61 (±23.73)	0.4566 (±.015)	0.7360 (±.001)	0.71483 (±.003)
$G_{\mathbf{M}}$	375.03 (±10.77)	567.18 (±45.45)	718.37 (±6.92)	0.8863 (±.034)	0.6786 (±.029)	0.3879 (±.021)	1761.58 (±60.76)	1150.56 (±11.86)	1269.20 (±28.73)	0.4612 (±.07)	0.7359 (±.000)	0.7177 (±.002)
$G_{\mathbf{C}}$	376.47 (±16.13)	571.56 (±65.39)	716.09 (±12.53)	0.8998 (±.005)	0.6796 (±.071)	0.4408 (±.107)	1630.78 (±216.87)	1148.70 (±11.71)	1264.94 (±17.14)	0.5489 (±.104)	0.7348 (±.000)	0.7152 (±.001)
$G_{\mathbf{B}} + \mathcal{L}_g$	398.59 (±56.66)	590.19 (±6.08)	779.99 (±23.59)	0.8597 (±.051)	0.6691 (±0.00)	0.3414 (±.075)	1972.77 (±85.59)	1178.89 (±85.22)	1242.17 (±2.39)	0.2538 (±.075)	0.7304 (±.019)	0.7126 (±.003)
$G_{\mathbf{M}} + \mathcal{L}_g$	343.37 (±30.58)	546.01 (±15.28)	667.67 (±19.08)	0.8967 (±.011)	0.6509 (±.006)	0.4703 (±.003)	1837.77 (±107.72)	1137.91 (±5.50)	1191.21 (±2.51)	0.3926 (±.096)	0.7254 (±.000)	0.7259 (±.000)

1082 COVID time points in our data. Root Mean Squared Error (RMSE) and Pearson's Correlation (PCC) are used to evaluate the forecasting performance.

Implementation For the base GNN model, we adopt the same hyperparameter setting with that in [5]. The attention matrix $\mathbf{B}$ and $\mathbf{M}$ are learned following the same method in [5]. We set horizon h as 1, 3, 5 for ILI and 1, 7, 14 for COVID. All results are an average of 3 randomized trials using random seeds 41, 42, 43. λ for $\mathcal{L}_g$ is set to 1.5 through cross validation.

4 Forecasting Performance and Analysis

In Table 2, we show forecasting performance on ILI and COVID datasets. We will analyze the results from multiple dimensions aligned with the ablation study design in Section 3.2. Our observations are:

(I) The performance of $G_{\mathbf{S}}$ outperforms $G_{\mathbf{R}}$. This indicates that geographical information is useful when constructing a sparse graph compared with a random graph topology.

(II) Comparing dense graphs $G_{\mathbf{V}}$ and $G_{\mathbf{D}}$ with sparse graphs $G_{\mathbf{S}}$ and $G_{\mathbf{R}}$, we observe that $G_{\mathbf{R}}$ performs the worst while $G_{\mathbf{S}}$ performs the best for most cases. Based on the definition, $G_{\mathbf{S}}$ represents a sparse graph and edges are sampled by geographical information, we call it geo-based sampling. $G_{\mathbf{R}}$ is a sparse graph with random sampling. The results imply that a sparse graph with proper sampling can improve GNN performance. This is consistent with the theoretical analysis from [19]. Model-generated graph structure decreases model performance compared with the baseline $G_{\mathbf{S}}$.

(III) Dynamic graph $G_{\mathbf{C}}$ outperforms static graph $G_{\mathbf{S}}$ for horizon=1, while it does not perform well for long horizons. Similar observations also found by comparing dynamic graphs $G_{\mathbf{B}}$/$G_{\mathbf{M}}$ with static graphs $G_{\mathbf{D}}$/$G_{\mathbf{V}}$. This implies that given the same graph topology, the edges weighted by learned attention coefficients can improve forecasting performance in short-term predictions, but is not helpful in long-term predictions. The learned attention coefficients can capture hidden spatial patterns for the input time series. However, the learned pattern is not generalizable to future time points. This finding is different with the empirical results in the previous work.

(IV) Comparing geo-attention graphs $G_{\mathbf{M}}$, $G_{\mathbf{C}}$ with single geographical graph $G_{\mathbf{S}}$, single attention graph $G_{\mathbf{B}}$, it is observable that the geo-attention graphs perform no better than single ones particularly for large horizons. This indicates that attention matrix incorporating geographical information cannot capture spatial patterns that are generalizable to future time points.

(V) Comparing geo-regularized attention graphs $G_{\mathbf{M}} + \mathcal{L}_g$/$G_{\mathbf{B}} + \mathcal{L}_g$ with no regularization attention grapsh $G_{\mathbf{M}}$/$G_{\mathbf{B}}$, we observe that $G_{\mathbf{M}} + \mathcal{L}_g$ surpasses $G_{\mathbf{M}}$ in performance on both datasets, while $G_{\mathbf{B}} + \mathcal{L}_g$ is inferior to $G_{\mathbf{B}}$ on both tasks. A possible reason is that the hyperparameter λ value is not proper for $G_{\mathbf{B}} + \mathcal{L}_g$. The observation indicates that through delicate tuning of λ value, the forecasting performance can be improved using a geo-based regularization term to penalize on the learned attention-based matrix. However, the performance is sensitive to hyperparameter values.

Overall, for multiple horizons and multiple datasets, using geographical information to construct a sparse graph is the most stable and efficient method in GNN models for epidemic forecasting. It incorporates actual geographical knowledge without adding extra training parameters, leading to improved forecasting performance and superior generalization capabilities among different forecasting tasks. If attention mechanism is applied, a combination of geographical adjacency and attention coefficients should be considered. Adding a geo-based regularization to penalize attention-based coefficients can lead to improved performance. Nevertheless, this approach is hyperparameter-sensitive and lacks generalizability across different forecasting tasks.

5 Conclusion

In this work, we explore different graph structures in GNNs for epidemic forecasting. By adopting a pre-existing GNN model, we build multiple graph structures of different density (dense or sparse) and state (static or dynamic) levels. Geographical information, model-generated gravity, and attention mechanism are considered for constructing graph edge weights. We conduct detailed ablation study on two real-world epidemic datasets. The results indicate that leveraging simple geographical adjacency to build sparse graphs can enhance GNN forecasting performance and is more generalizable among different tasks compared with more complex attention-based dense adjacency matrices.

References

[1] A. Adiga, L. Wang, A. Sadilek, A. Tendulkar, S. Venkatramanan, A. Vullikanti, G. Aggarwal, A. Talekar, X. Ben, J. Chen, et al. Interplay of global multi-scale human mobility, social distancing, government interventions, and covid-19 dynamics. *medRxiv*, 2020.

[2] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos. Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PloS one*, 15(3):e0230405, 2020.

[3] P. Arora, H. Kumar, and B. K. Panigrahi. Prediction and analysis of covid-19 positive cases using deep learning models: A descriptive case study of india. *Chaos, Solitons & Fractals*, page 110017, 2020.

[4] V. K. R. Chimmula and L. Zhang. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals*, page 109864, 2020.

[5] S. Deng, S. Wang, H. Rangwala, L. Wang, and Y. Ning. Cola-gnn: Cross-location attention based graph neural networks for long-term ili prediction. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 245–254, 2020.

[6] J. Gao, R. Sharma, C. Qian, L. M. Glass, J. Spaeder, J. Romberg, J. Sun, and C. Xiao. Stan: spatio-temporal attention network for pandemic prediction using real-world evidence. *Journal of the American Medical Informatics Association*, 28(4):733–743, 2021.

[7] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nature Medicine*, pages 1–6, 2020.

[8] A. Harvey and P. Kattuman. Time series models based on growth curves with applications to forecasting coronavirus. *Covid Economics, Vetted and Real-Time Papers*, (24), 2020.

[9] A. Kapoor, X. Ben, L. Liu, B. Perozzi, M. Barnes, M. Blais, and S. O'Banion. Examining covid-19 forecasting using spatio-temporal graph neural networks. *arXiv preprint arXiv:2007.03113*, 2020.

[10] B. Lu, X. Gan, W. Zhang, H. Yao, L. Fu, and X. Wang. Spatio-temporal graph few-shot learning with cross-city knowledge transfer. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1162–1172, 2022.

[11] F. Petropoulos and S. Makridakis. Forecasting the novel coronavirus covid-19. *PloS one*, 15(3):e0231236, 2020.

[12] A. Ramchandani, C. Fan, and A. Mostafavi. Deepcovidnet: An interpretable deep learning model for predictive surveillance of covid-19 using heterogeneous features and their interactions. *Ieee Access*, 8:159915–159930, 2020.

[13] J. Shaman and A. Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430, 2012.

[14] A. Vaswani, N. Shazeer, et al. Attention is all you need. In *NeurIPS*, pages 5998–6008, 2017.

[15] 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*, volume 36, pages 12191–12199, 2022.

[16] L. Wang, X. Ben, A. Adiga, A. Sadilek, A. Tendulkar, S. Venkatramanan, A. Vullikanti, G. Aggarwal, A. Talekar, J. Chen, et al. Using mobility data to understand and forecast covid19 dynamics. *medRxiv*, 2020.

[17] T. Yamana, S. Pei, and J. Shaman. Projection of covid-19 cases and deaths in the us as individual states re-open may 4, 2020. *medRxiv*, 2020.

[18] W. Yang, D. R. Olson, and J. Shaman. Forecasting influenza outbreaks in boroughs and neighborhoods of new york city. *PLoS computational biology*, 12(11):e1005201, 2016.

[19] S. Zhang, M. Wang, P.-Y. Chen, S. Liu, S. Lu, and M. Liu. Joint edge-model sparse learning is provably efficient for graph neural networks. *arXiv preprint arXiv:2302.02922*, 2023.