$PM_{2.5}$-GNN: A Domain Knowledge Enhanced Graph Neural Network For $PM_{2.5}$ Forecasting

Shuo Wang*
shawnwang.tech@gmail.com
School of Systems Science, Beijing Normal University
Beijing, China

Yanran Li*
csyli@comp.polyu.edu.hk
Department of Computing, The Hong Kong Polytechnic University
Hong Kong, China

Jiang Zhang†
zhangjiang@bnu.edu.cn
School of Systems Science, Beijing Normal University
Beijing, China

Qingye Meng
mengqingye@caiyunapp.com
ColorfulClouds Pacific Technology Co., Ltd.
Beijing, China

Lingwei Meng
lingweim@princeton.edu
Atmospheric and Oceanic Sciences Program, Princeton University
Princeton, NJ, USA

Fei Gao
Philip.sss@mail.bnu.edu.cn
School of Systems Science, Beijing Normal University
Beijing, China

ABSTRACT

When predicting $PM_{2.5}$ concentrations, it is necessary to consider complex information sources since the concentrations are influenced by various factors within a long period. In this paper, we identify a set of critical domain knowledge for $PM_{2.5}$ forecasting and develop a novel graph based model, $PM_{2.5}$-GNN, being capable of capturing long-term dependencies. On a real-world dataset, we validate the effectiveness of the proposed model and examine its abilities of capturing both fine-grained and long-term influences in $PM_{2.5}$ process. The proposed $PM_{2.5}$-GNN has also been deployed online to provide free forecasting service.

CCS CONCEPTS

• **Information systems → Spatial-temporal systems**.

KEYWORDS

air quality prediction, graph neural network, spatio-temporal prediction

ACM Reference Format:

Shuo Wang, Yanran Li, Jiang Zhang, Qingye Meng, Lingwei Meng, and Fei Gao. 2020. $PM_{2.5}$-GNN: A Domain Knowledge Enhanced Graph Neural Network For $PM_{2.5}$ Forecasting. In *28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '20), November 3–6, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3397536.3422208

1 INTRODUCTION

In the past few decades, the rapid development of industry has caused severe air pollution problem. In particular, particles smaller than 2.5μm ($PM_{2.5}$) have been paid special attention as they play an important role in atmospheric visibility reduction, human health, acid deposition and the climate.

However, it is non-trivial to accurately predict $PM_{2.5}$ concentration for several reasons. Firstly, $PM_{2.5}$ concentration is characterized by a complex process, starting from emission generated by pollution sources to transport and diffusion influenced by meteorological and geographical information. It is thus a necessity to make good use of domain knowledge when modeling this temporal and spatial process. Secondly, these aforementioned $PM_{2.5}$ factors often take a wide-range and long-lasting effects. As reported by [2], $PM_{2.5}$ can be transported hundreds of kilometers in 72 hours. The ability of handling long-term dependencies is then another key to prediction accuracy. We summarize these $PM_{2.5}$ characteristics as: (1) **domain-knowledge sensitivity**; and (2) **long-term dependency**.

Existing work often structure $PM_{2.5}$ data into graphs and adopt graph-based approaches to capture the process. [4] establishes an undirected graph and computes the spatial similarity by utilizing the neighborhood information within a single city. [5] also operates on an undirected graph and derives the hidden spatial dependencies mainly based on a Graph Convolution Network. Nevertheless, none of these approaches is capable of explicitly incorporating domain knowledge like wind directions, which are critical for modeling $PM_{2.5}$ transport process. Partially due to this kind of incapability, these existing approaches fail to predict $PM_{2.5}$ concentration in a wide range for a long period of time.

In this work, we develop $PM_{2.5}$-GNN, which takes into account the aforementioned two $PM_{2.5}$ characteristics. Notably, we establish a *directed* graph where nodes are cities and edges representing city-to-city interactions. Since $PM_{2.5}$ prediction is sensitive to domain knowledge, we innovatively consider both meteorological and geographical information to precisely model its temporal and spatial process. The meteorological information for each city is characterized as node and edge features, while geographical knowledge among cities are encoded into graph structures. Based on the constructed graph, we learn $PM_{2.5}$ spatial transport among cities by leveraging a knowledge-enhanced Graph Neural Network, and

*Both authors contributed equally to this paper.
†Jiang Zhang is the correspondence author of this paper.

SIGSPATIAL '20, November 3–6, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-8019-5/20/11.
https://doi.org/10.1145/3397536.3422208

capture the temporal diffusion process using a Recurrent Neural Network. Combing these two modules, the proposed $PM_{2.5}$-GNN is effective in modeling both transport and diffusion process by exploiting extensive domain knowledge. As a result, on a real-world dataset, we are able to accurately predict $PM_{2.5}$ concentrations of multiple city clusters for 72 hours in the future. This demonstrates the advantage of the proposed approach in characterizing $PM_{2.5}$'s long-term dependency.

In brief, we highlight our contributions as follows:

- We regard the necessity of domain knowledge in $PM_{2.5}$ prediction, and incorporate them into graph-structured data.
- We develop $PM_{2.5}$-GNN, a novel $PM_{2.5}$ prediction model to explicitly model the long-term dependency by utilizing domain knowledge.
- We introduce a large-scale real-world dataset, namely KnowAir, where we validate the effectiveness of our model through performing 72-hour $PM_{2.5}$ prediction. We release the dataset and code[1], and deploy $PM_{2.5}$-GNN in an online website[2].

2 PRELIMINARIES

$PM_{2.5}$ concentration prediction is formulated as a spatio-temporal sequence prediction problem. Let $X^t \in \mathbb{R}^{N \times 1}$ denote the $PM_{2.5}$ concentrations at time step t, where N is the number of nodes. We define a *directed* graph $G = (V, E)$, where V is the set of nodes representing cities, and E is the set of edges denoting potential interactions among cities. Let $P^t \in \mathbb{R}^{N \times p}$ and $Q^t \in \mathbb{R}^{M \times q}$ denote the nodes' and edges' attribute matrices respectively at time step t, where p, q are the corresponding attribute numbers, and $M = |E|$ is the edge size.

To enhance model's prediction ability, the key centers in how to better utilize domain knowledge. Our idea is to explicitly encode domain knowledge, e.g., meteorological information, into the attribute matrices and graph structures. Following previous work [7], we also exploit domain knowledge from the near future, which can be acquired from weather forecasting services. Formally, for any starting point t, we feed the observed $PM_{2.5}$ concentrations X^t at current time t, next T steps of attribute matrices $\left[P^{t+1}, \ldots, P^{t+T}\right]$ and $\left[Q^{t+1}, \ldots, Q^{t+T}\right]$, as well as the graph structure G into our model. We achieve this by iterating T steps of $PM_{2.5}$-GNN. In this way, the $PM_{2.5}$ concentration problem is framed as:

$$\left[X^t; P^{t+1}, ..., P^{t+T}; Q^{t+1}, ..., Q^{t+T}; G\right] \xrightarrow{f(\cdot)} \left[\hat{X}^{t+1}, ..., \hat{X}^{t+T}\right] \quad (1)$$

For training, we are to minimize the error between predicted values $\left[\hat{X}^1, \ldots, \hat{X}^T\right]$ and ground truth $\left[X^1, \ldots, X^T\right]$ using Mean Squared Error (MSE) loss:

$$\text{MSE Loss} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{N} \sum_{i=1}^{N} \left(\hat{X}_i^t - X_i^t \right)^2 \right) \quad (2)$$

3 METHODOLOGY

In this section, we firstly construct the graph consisting of the studied cities and their potential interactions as shown in Figure 1. Then, we present the proposed approach to model the spatial and temporal dependencies of $PM_{2.5}$. A more detailed description may be found in the full paper [6].

3.1 Graph Construction

As analyzed before, $PM_{2.5}$ is influenced by a number of factors. To improve prediction accuracy, we establish our graph by incorporating domain knowledge as attributes on nodes and edges, then learn transport and diffusion process based on the graph.

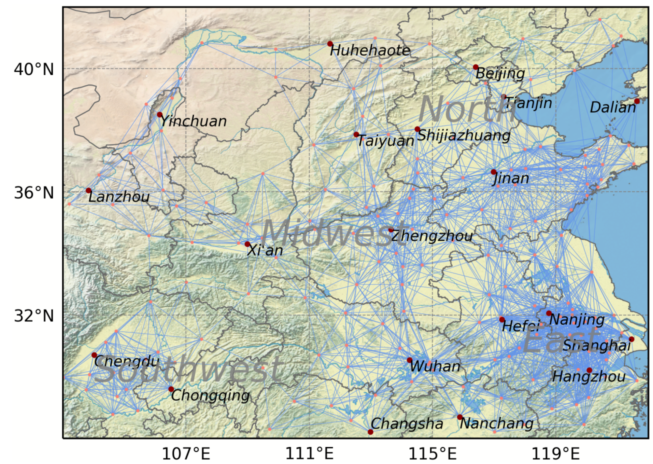

Figure 1: The Studied Area Shown On the Map.

Adjacency Matrix. We need to compute the nodes' correlations to build the adjacency matrix. Intuitively, most aerosol pollutants are distributed in a certain range above the ground. Moreover, the mountains lying along the two cities will hinder $PM_{2.5}$ transport of the pollutants. Based on these intuitions, we constrain the weights in the adjacency matrix through:

$$\begin{aligned} A_{ij} &= H(d_\theta - d_{ij}) \cdot H(m_\theta - m_{ij}), \quad \text{where} \\ d_{ij} &= ||\rho_i - \rho_j|| \\ m_{ij} &= \sup_{\lambda \in (0,1)} \left\{ h(\lambda \rho_i + (1-\lambda)\rho_j) - \max\left\{h(\rho_i), h(\rho_j)\right\} \right\} \end{aligned} \quad (3)$$

where ρ_i is the position (latitude, longitude) of node i, $h(\rho)$ is the altitude of the position ρ, and $||\cdot||$ is the L2-norm of a vector. $H(\cdot)$ is the Heaviside step function,[3] where $H(x) = 1$ if and only if $x > 0$. d_θ and m_θ are the thresholds of distance and altitude, respectively. In this paper, we set $d_\theta = 300km$ for the distance threshold and $m_\theta = 1200m$ for the altitude threshold. That means, $PM_{2.5}$ can transport from one city to another in one step only if the distance between these two cities is less than $300km$ and the mountains between them are lower than $1200m$.

Node Attributes. A place's meteorological characteristics influence how the pollutant will be diffused. Following are the representative meteorological variables we choose as nodes' attributes: *Planetary Boundary Layer (PBL) height*; *K index*; *Wind speed*; *2m Temperature*; *Relative humidity*; *Precipitation*; and *Surface pressure*. These node attributes are related to $PM_{2.5}$ vertical diffusion through dynamic or thermodynamic effects. Besides, we also use hour and day of the week (exact time) as temporal information like [5].

[1] https://github.com/shawnwang-tech/PM2.5-GNN
[2] http://caiyunapp.com/map/
[3] https://en.wikipedia.org/wiki/Heaviside_step_function

Edge Attributes. [2] shows that wind speed and direction have a decisive effect on $PM_{2.5}$ horizontal transport. To incorporate this strong domain knowledge, we using the following variables as edge attributes: *Wind speed of source node* $|v|$; *Distance between source and sink* d; *Wind direction of source node* β; *Direction from source to sink* γ; and *Advection coefficient* S calculated using Equation 4.

$$S = \mathrm{ReLU}\left(\frac{|v|}{d}\cos(|\gamma - \beta|)\right) \tag{4}$$

3.2 $PM_{2.5}$-GNN Model

As illustrated in Figure 2, the proposed $PM_{2.5}$-GNN model consists of two main components. A knowledge-enhanced Graph Neural Network (GNN) is devised to capture pollutants' horizontal transport by leveraging neighboring information and updating nodes' representations. A spatio-temporal GRU is applied after updates to model pollutants' vertical accumulation and diffusion under the influence of weather.

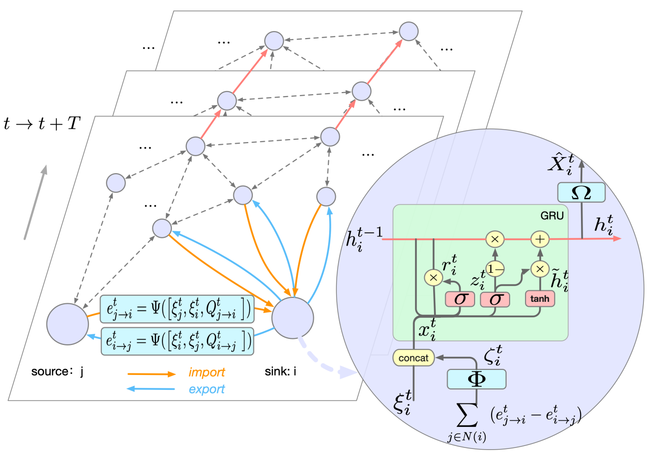

Figure 2: The proposed $PM_{2.5}$-GNN model consisting of a knowledge-enhanced GNN and a spatio-tempo RNN.

Following the paradigm of message passing [1], the knowledge-enhanced GNN learns representations by iteratively aggregating neighboring information on the graph. This iterative process is formulated as in Equation 5, where Ψ and Φ are differentiable functions. At each time step, a node's representation ξ_i^t is obtained by concatenating the previous predicted $PM_{2.5}$ concentrations $\hat{X}_i^{t-1}$ and its current attributes P_i^t. An edge's representation $e_{j\to i}^t$ is calculated by combining its connecting nodes as well as its own attributes.

$$\begin{aligned}
\xi_i^t &= \left[\hat{X}_i^{t-1}, P_i^t\right] && \forall i \in V \\
e_{j\to i}^t &= \Psi(\left[\xi_j^t, \xi_i^t, Q_{j\to i}^t\right]) && \forall \langle j, i\rangle \in E \\
\zeta_i^t &= \Phi(\sum_{j\in N(i)} (e_{j\to i}^t - e_{i\to j}^t)) && \forall i \in V
\end{aligned} \tag{5}$$

Note that under our formulation, edge attributes $Q_{j\to i}^t$ and edge representations $e_{j\to i}^t$ are direction-aware. This kind of formulation enables us to explicitly measure the influence between node neighbors. Given a node i, the transport from its neighbor j approximates to the difference between the import $e_{j\to i}^t$ and export influence $e_{i\to j}^t$, depicted as the orange and blue lines in Figure 2. Then, the spatial correlation ζ_i^t for node i is calculated by summarizing the influences from all its neighbors. After several steps of recursion, each node on the graph is aware of others, and the transport is captured using the encoded domain knowledge.

To learn the temporal diffusion of $PM_{2.5}$, we devise a Recurrent Neural Network (RNN) and apply it over the knowledge-enhanced GNN. In specific, we adopt Gated Recurrent Unit (GRU) as our basic recurrent unit due to its advantage on capturing long-term dependency. At each time step, the GRU cell takes as input the combination of node's representation ξ_i^t and its spatial correlation ζ_i^t. This allows GRU to also consider the spatial transport while modeling the temporal diffusion. We depict the process of our spatio-tempo GRU in Equation 6:

$$\begin{aligned}
x_i^t &= \left[\xi_i^t, \zeta_i^t\right] \\
z_i^t &= \sigma\left(W_z \cdot \left[h_i^{t-1}, x_i^t\right]\right) \\
r_i^t &= \sigma\left(W_r \cdot \left[h_i^{t-1}, x_i^t\right]\right) \\
\tilde{h}_i^t &= \tanh\left(W \cdot \left[r_i^t * h_i^{t-1}, x_i^t\right]\right) \\
h_i^t &= (1 - z_i^t) * h_i^{t-1} + z_i^t * \tilde{h}_i^t
\end{aligned} \tag{6}$$

where W_z, W_r and W are the learnable parameters.

Based on the output of the proposed $PM_{2.5}$-GNN model, we finally predict $PM_{2.5}$ concentration by:

$$\hat{X}_i^t = \Omega(h_i^t) \qquad \forall i \in V \tag{7}$$

where Ω is a Multi-Layer Perceptron (MLP). We summarize the learning procedure of the proposed $PM_{2.5}$-GNN in below.

Algorithm 1: $PM_{2.5}$-GNN model

Input : $PM_{2.5}$ observed concentrations X^0 ;
nodes' attributes $\left[P^1, ..., P^T\right]$;
edges' attributes $\left[Q^1, ..., Q^T\right]$;
$G = (V, E)$;
Output: $PM_{2.5}$ predicted concentrations $\left[\hat{X}^1, ..., \hat{X}^T\right]$;

#Initialize:
$h^0 \leftarrow 0$;
$\hat{X}^0 \leftarrow X^0$;
Output_list = [ ];

for $t = 1, ..., T$ **do**
  for $i \in V$ **do**
    ζ_i^t = **GNN**$(\xi_i^t, \{\xi_j^t, Q_{i\to j}^t, Q_{j\to i}^t\}_{j\in N(i)})$ (Equation 5);
    h_i^t = **GRUcell**$(\left[\xi_i^t, \zeta_i^t\right], h_i^{t-1})$ (Equation 6);
    $\hat{X}_i^t$ = **MLP**(h_i^t) (Equation 7);
    Append $\hat{X}_i^t$ into *Output_list*;

4 EXPERIMENTS

We conduct experiments to demonstrate the effectiveness of $PM_{2.5}$-GNN, and analyze whether and to what extent $PM_{2.5}$-GNN is able to resemble $PM_{2.5}$'s characteristics of domain-knowledge sensitivity and long-term dependency. For reproducibility, we release the datasets and codes on GitHub [4]. We refer to the full paper [6] for more details of the experiments because of the page limitation.

Datasets. Following previous work [7], we use two types of data: $PM_{2.5}$ concentrations obtained from Ministry of Ecology and Environment (MEE) [5], and weather forecasting data obtained from climate reanalysis ERA5[6]. We construct a whole 4-year dataset *KnowAir*, which covers in total 184 cities (nodes). To thoroughly investigate the model abilities, the dataset is further spilt into 3 ones as presented in Table 1.

Table 1: Dataset is spilt into 3 sub-datasets.

Sub-dataset	Train	Validate	Test
1	2015/1/1 - 2016/12/31	2017/1/1 - 2017/12/31	2018/1/1 - 2018/12/31
2	2015/11/1 - 2016/2/28	2016/11/1 - 2017/2/28	2017/11/1 - 2018/2/28
3	2016/9/1 - 2016/11/30	2016/12/1 - 2016/12/31	2017/1/1 - 2017/1/31

Compared Models. We compare with **GRU**, and **GC-LSTM** [5]. GC-LSTM is the current state-of-the-art baseline for $PM_{2.5}$ prediction. It integrates LSTM and Graph Convolutional Networks (GCN) [3] to model the temporal and spatial dependency respectively. Different from our $PM_{2.5}$-GNN, the GCN module in GC-LSTM only applies to undirected graph, and no edges' attributes are used. These limit its capacity of explicitly modeling $PM_{2.5}$ transport along the direction of wind flow.

Experimental Settings. For pre-processing, we rescale the features of nodes and edges to have mean of 0 and standard deviation of 1 (unit variance). We initialize the GRU's hidden state h^0 with a zero tensor. Specifically, functions Ψ and Φ in Equation 5 as well as Ω in Equation 7 are MLPs with layer numbers of 2, 1, 1, respectively. All models are trained using RMSprop for 50 epochs with learning rate as 5^{-4}. Early stopping is also adopted on the validation set. We evaluate the model performances using: (1) root mean square error (RMSE) to directly examine the prediction accuracy; and (2) commonly used meteorological metrics, critical success index (CSI), to measure the performance near the pollution threshold $75\mu g/m^3$. For each model, their performances are calculated by averaging each 24 metrics from the total 184 cities, i.e., predicting the concentration per city per 3 hours within 72 hours in total. We repeat 10 times of experiments, and report the mean and standard deviation of metrics in Table 2. Note that the lower RMSE, the higher CSI mean better performance.

Experimental Results. As shown in Table 2, our $PM_{2.5}$-GNN surpasses the compared models on every metric across the sub-datasets. This result demonstrates $PM_{2.5}$-GNN's advantage on exploiting domain knowledge and suggests its reliable prediction ability. Among all, GRU performs the worst because it could only access to the temporal information. This also supports the necessity of neighboring information for $PM_{2.5}$ prediction. Although GC-LSTM could model the temporal and spatial dependency simultaneously, lack of the ability to model the $PM_{2.5}$ transport weakens its performance. We also analyze $PM_{2.5}$-GNN from the perspective of domain knowledge. We focus on PBL height and transport direction information. Recall that the proposed $PM_{2.5}$-GNN encodes PBL height as node features and captures directed transport through the knowledge-enhanced GNN component (Equation 5). Excluding the corresponding node feature and the export influence (remove $e^t_{i\to j}$ from Equation 5), we obtain the ablated results. Clearly, the performances degrade a lot when removing any of them. This ablation study verifies the importance of domain knowledge as well as the motivation of our research.

Table 2: Experimental results of $PM_{2.5}$-GNN's compared models, and its different configurations for ablation study. Lack of PBL feature or subtraction (export) component worsens $PM_{2.5}$-GNN's performance.

	Sub-dataset 1		Sub-dataset 2		Sub-dataset 3	
	RMSE	CSI (%)	RMSE	CSI (%)	RMSE	CSI (%)
GRU	21.00 ± 0.17	45.38 ± 0.52	32.59 ± 0.16	51.07 ± 0.81	45.25 ± 0.85	59.40 ± 0.01
GC-LSTM	20.84 ± 0.11	45.83 ± 0.43	32.10 ± 0.29	51.24 ± 0.13	45.01 ± 0.81	60.58 ± 0.14
$PM_{2.5}$-GNN	**19.93 ± 0.11**	**48.52 ± 0.48**	**31.37 ± 0.34**	**52.33 ± 1.06**	**43.29 ± 0.79**	**61.91 ± 0.78**
$PM_{2.5}$-GNN no PBL	20.46 ± 0.18	47.43 ± 0.37	32.44 ± 0.36	51.05 ± 1.15	44.71 ± 1.02	60.64 ± 0.84
$PM_{2.5}$-GNN no export	20.54 ± 0.16	45.73 ± 0.58	31.91 ± 0.32	51.54 ± 1.27	43.72 ± 1.03	61.52 ± 0.95

5 CONCLUSIONS

In this paper, we study a significant problem in real-world that how to precisely predict $PM_{2.5}$ concentrations in the next 72 hours. Observing the two typical $PM_{2.5}$ characteristics, we leverage a GNN to introduce domain knowledge and integrate with a RNN to capture the fine-grained and long-term dependencies during the $PM_{2.5}$ process. We show the success of our approach through extensive experiments and deploy the proposed $PM_{2.5}$-GNN model online in hope to benefit the community. In the future, we plan to study the model interpretability and $PM_{2.5}$ source tracking. It is also promising to generalize our approach to PM_{10} (dust) concentration prediction which has more obvious transport effect.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (NSFC) under the grant number 61673070.

REFERENCES

[1] Matthias Fey and Jan E. Lenssen. 2019. Fast Graph Representation Learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*.

[2] Jianlin Hu, Yungang Wang, Qi Ying, and Hongliang Zhang. 2014. Spatial and temporal variability of PM 2.5 and PM 10 over the North China Plain and the Yangtze River Delta, China. *Atmospheric Environment* 95 (2014), 598–609. https://doi.org/10.1016/j.atmosenv.2014.07.019

[3] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.

[4] Yijun Lin, Nikhit Mago, Yu Gao, Yaguang Li, Yao-Yi Chiang, Cyrus Shahabi, and José Luis Ambite. 2018. Exploiting spatiotemporal patterns for accurate air quality forecasting using deep learning. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 359–368.

[5] Yanlin Qi, Qi Li, Hamed Karimian, and Di Liu. 2019. A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. *Science of The Total Environment* 664 (02 2019). https://doi.org/10.1016/j.scitotenv.2019.01.333

[6] Shuo Wang, Yanran Li, Jiang Zhang, Qingye Meng, Lingwei Meng, and Fei Gao. 2020. PM2. 5-GNN: A Domain Knowledge Enhanced Graph Neural Network For PM2. 5 Forecasting. *arXiv preprint arXiv:2002.12898* (2020).

[7] Xiuwen Yi, Junbo Zhang, Zhaoyuan Wang, Tianrui Li, and Yu Zheng. 2018. Deep distributed fusion network for air quality prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 965–973.

[4] https://github.com/shawnwang-tech/PM2.5-GNN
[5] http://datacenter.mee.gov.cn/websjzx/dataproduct/airproduct/queryAirAround.vm
[6] https://climate.copernicus.eu/climate-reanalysis