


Article

A Higher-Order Graph Convolutional Network for Location Recommendation of an Air-Quality-Monitoring Station

Yu Kang ^{1,2,3,4}, Jie Chen ¹, Yang Cao ¹, and Zhenyi Xu ^{1,2,*} 

¹ Department of Automation, University of Science and Technology of China Hefei, Hefei 230026, China; kangduyu@ustc.edu.cn (Y.K.); jaychan@mail.ustc.edu.cn (J.C.); forrest@ustc.edu.cn (Y.C.)

² Hefei Comprehensive National Science Center, Institute of Artificial Intelligence, Hefei 230088, China

³ Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China

⁴ State Key Laboratory of Fire Science, University of Science and Technology of China, Hefei 230027, China

* Correspondence: xuzhenyi@mail.ustc.edu.cn

Abstract: The location recommendation of an air-quality-monitoring station is a prerequisite for inferring the air-quality distribution in urban areas. How to use a limited number of monitoring equipment to accurately infer air quality depends on the location of the monitoring equipment. In this paper, our main objective was how to recommend optimal monitoring-station locations based on existing ones to maximize the accuracy of a air-quality inference model for inferring the air-quality distribution of an entire urban area. This task is challenging for the following main reasons: (1) air-quality distribution has spatiotemporal interactions and is affected by many complex external influential factors, such as weather and points of interest (POIs), and (2) how to effectively correlate the air-quality inference model with the monitoring station location recommendation model so that the recommended station can maximize the accuracy of the air-quality inference model. To solve the aforementioned challenges, we formulate the monitoring station location as an urban spatiotemporal graph (USTG) node recommendation problem in which each node represents a region with time-varying air-quality values. We design an effective air-quality inference model-based proposed high-order graph convolution (HGCNInf) that could capture the spatiotemporal interaction of air-quality distribution and could extract external influential factor features. Furthermore, HGCNInf can learn the correlation degree between the nodes in USTG that reflects the spatiotemporal changes in air quality. Based on the correlation degree, we design a greedy algorithm for minimizing information entropy (GMIE) that aims to mark the recommendation priority of unlabeled nodes according to the ability to improve the inference accuracy of HGCNInf through the node incremental learning method. Finally, we recommend the node with the highest priority as the new monitoring station location, which could bring about the greatest accuracy improvement to HGCNInf.

Keywords: air-quality inference; spatiotemporal interaction; graph convolutional networks; station location recommendation; incremental learning



Citation: Kang, Y.; Chen, J.; Cao, Y.; Xu, Z. A Higher-Order Graph Convolutional Network for Location Recommendation of an Air-Quality-Monitoring Station. *Remote Sens.* **2021**, *13*, 1600. <https://doi.org/10.3390/rs13081600>

Academic Editor: Biswajeet Pradhan

Received: 9 February 2021

Accepted: 16 April 2021

Published: 20 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, with economic growth, environmental problems have become increasingly prominent and air pollution is receiving unprecedented attention [1–4]. The air-quality index (AQI) provides a number used by government agencies to communicate to the public how polluted the air is currently. As the AQI increases, an increasingly large percentage of the population is likely to experience increasingly severe adverse health effects [5–7]. To compute this AQI, an air-pollutant concentration from a monitor or model is required, such as carbon monoxide (CO), carbon dioxide (CO₂), hydrocarbons (HC), nitrogen oxides (NO_x), solid particulate matter (PM_{2.5} and PM₁₀), etc. In order to reflect the air quality and its development trends in a timely and accurate manner, we need accurate air-quality-monitoring equipment.

However, it is unrealistic to establish a large number of monitoring equipment in different areas of the city due to the expensive construction and later maintenance costs [8]. Based on the premise of a limited number of monitoring equipment, a reasonable monitoring station location layout is the first basic link to accurately infer air quality. On the other hand, with the rapid development of urbanization and the continuous growth in the number of motor vehicles, the number of established monitoring equipment is too few to accurately reflect the air-quality distribution in an entire urban area.

Thus far, there is few research on the optimization methods of regional air-quality-monitoring networks and there is also no uniform standard for recommending optimal monitoring equipment locations. Therefore, we need to design a framework that could recommend optimal station locations to establish new monitoring equipment on the basis of the existing ones, which can bring about the greatest accuracy improvement to the inference model. In this way, it not only meets the accuracy requirements for air-quality inference but also saves economic costs.

In recent years, the research on the location selection is mainly divided into the following two categories: knowledge-driven approaches and data-driven approaches.

The knowledge-driven approaches use mathematical models [9,10] and physics knowledge [11] to solve the location recommendation problem through computational simulation. In addition, the United States, Europe, and Japan have also established regional air-quality-monitoring networks around photochemical smog pollution and aerosol pollution [12–14]. These methods all adopted a series of optimization techniques to determine the layout of network sites, including statistical analysis methods such as correlation analysis and cluster analysis [15,16] or mathematical methods such as multi-objective optimization [17,18]. In order to reach a stable state, the simulation process not only requires complex system programming but also consumes a lot of computing power. Simplifications and stationarity assumptions that may be unrealistic in modeling further degrade the model efficiency.

Recently, data-driven approaches for location recommendation have been developed based on fixed monitoring stations [19] or Taxi GPS trajectory [20,21] but without considering the road network spatial structure, such as road segment length, road types (highways, main roads, and streets), POI density, etc. Kang et al. [22] transformed the station recommendation into a graph problem, which is committed to covering the urban area with the least number of monitoring stations but without considering the influences of complex external factors, which might lead to geographically non-smooth values in air-quality distribution. Hsieh et al. [23] studied label propagation [24] on graphs and considered external influential factors but failed to capture the spatiotemporal correlation between nodes in graph.

In this paper, we want to solve a practical problem: how to recommend optimal station locations to establish air-quality-monitoring equipment based on existing ones to maximize the accuracy of the air-quality-inference model, so as to reflect the air-quality distribution and its development trend in a timely manner and accurately. This task is challenging for the following reasons:

(1) **The air-quality distribution is affected by many complex external influential factors (such as weather, traffic volume, land use, etc).** For commercial centers with heavy traffic flow, the air quality is often worse than that in parks or lakes. There will be geographical non-smooth data due to the influence of external influential factors, and it is difficult to obtain accurate air-quality distribution in unobserved areas through interpolation-based methods. For example, Figure 1 is a real-time record of a certain day at the Beijing Air-Quality-Monitoring Station. We can find that the PM_{10} concentration data are not smooth. The monitoring stations in the red circle are very geographically close, but the PM_{10} concentration varies greatly throughout the year. The most likely reason for this result is that the monitoring stations with low PM_{10} concentration are close to parks or lakes and that the stations with high PM_{10} concentration are located near commercial centers or main roads with heavy traffic.

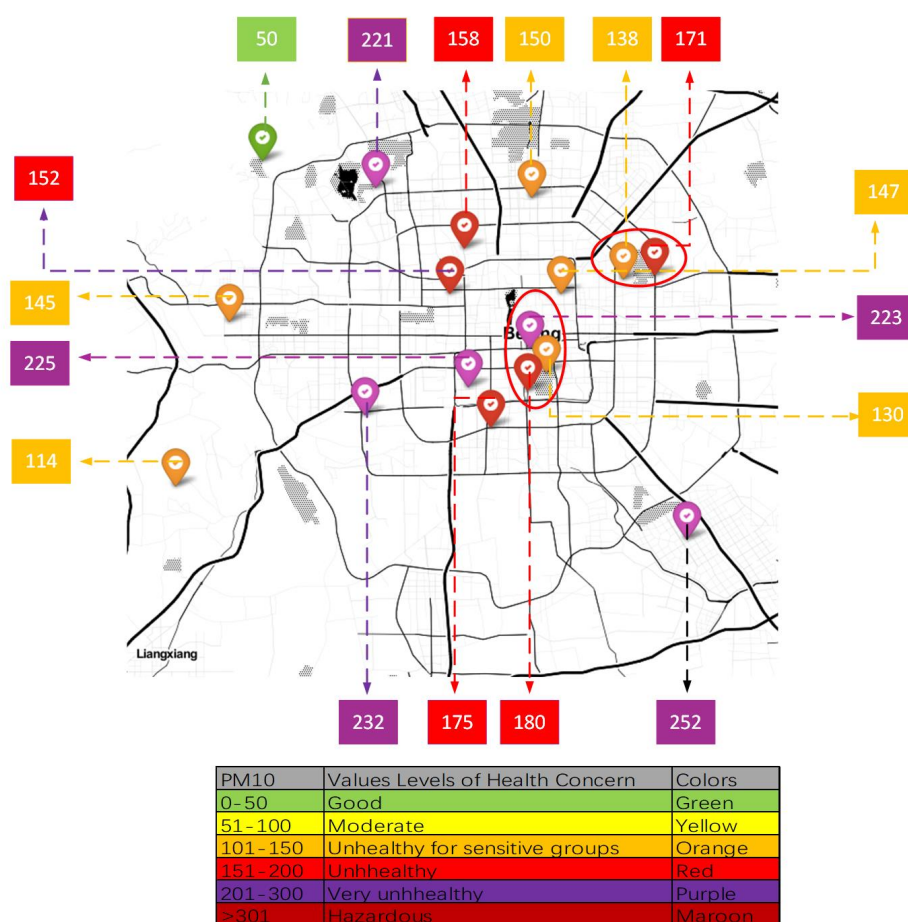


Figure 1. Beijing air-quality distribution on a certain day. The monitoring stations in the red circle are very geographically close, but the PM_{10} concentration varies greatly throughout the year.

(2) **There is spatiotemporal interaction between different nodes in the urban spatiotemporal graph:** At the same time, the air-quality value of a node is affected by other nodes; in the same space, the air-quality values of the same node at different times are correlated.

(3) **Complex correlation between the monitoring station location recommendation model and the air-quality inference model.** The task of the monitoring station location recommendation model is to select optimal locations to establish new monitoring stations to maximize the accuracy of an air-quality-inference model for inferring the air-quality distribution of an entire urban area. There is a close relationship between two models.

Aiming to solve the main research problems of this paper, the current research methods either fail to consider the urban spatial structure or the influence of external factors on air-quality distribution; the first problem is how to accurately infer air-quality distribution. However, the performance of an air-quality-inference model has a crucial impact on the subsequent monitoring-station location recommendation. The next most direct question is how to associate the monitoring-station location recommendation model with the air-quality-inference model to recommend optimal station locations. The existing methods cannot capture the correlation between nodes because of the performance limitation of the model, so the recommended station location is not optimal.

In order to solve the aforementioned challenges, we formulate monitoring station locations as an urban spatiotemporal graph (USTG) node recommendation problem in which each node represents a region with time-varying air-quality values and propose a two-step learning framework.

The main contributions of this research are as follows:

(1) We propose a variant of graph convolutional network (GCN) called higher-order graph convolutional network (HGCN). The ordinary GCN can capture the one-hop neighbor node's spatiotemporal correlation. We improve the ordinary graph convolution network and design a high-order graph convolution network. Compared with an ordinary (one-hop) GCN, HGCN has a larger receptive field, which could capture more and higher-order neighbor node information.

(2) We designed an accurate air-quality-inference model (HGCNInf) based on the proposed HGCN to infer urban air-quality distribution. We applied the graph convolutional network to the field of air-quality monitoring for the first time. By modeling the urban area as an USTG, the HGCNInf can accurately infer the air quality of an entire urban area: the HGCN can effectively capture the spatiotemporal interactions of the air-quality distribution, and a fully connected neural network is used to extract the external influential factor features.

(3) We analyze and correlate the air-quality-inference model with the monitoring-station location recommendation model. By using the convolution weight parameters of HGCNInf, we design a GMIE based on the correlation degree between nodes in USTG reflecting the air-quality spatiotemporal changes, marking the recommendation priority of unlabeled nodes according to the ability to improve the inference accuracy of HGCNInf iteratively through the node incremental learning method. The recommended optimal node could bring about the greatest accuracy improvement to HGCNInf.

(4) We evaluate our model using Beijing air-quality data from 1 January 2015 to 31 December 2015, and the experimental results show that our approach far outperforms the state-of-the-art baseline methods.

2. Study Area and Data Requirement

2.1. Study Area

Beijing is China's capital city (Figure 2), located at $40.250583^{\circ}\text{N}$ and $116.462611^{\circ}\text{E}$, on the northwestern edge of the North China Plain. The average altitude of Beijing is 43.5 m above sea level. Beijing is located in a warm temperate and semi-humid area. The climate is a warm temperate and characterized by semi-humid continental monsoons. The average annual precipitation in the plain area is about 600 mm, and the annual average temperature is about 12.9°C .

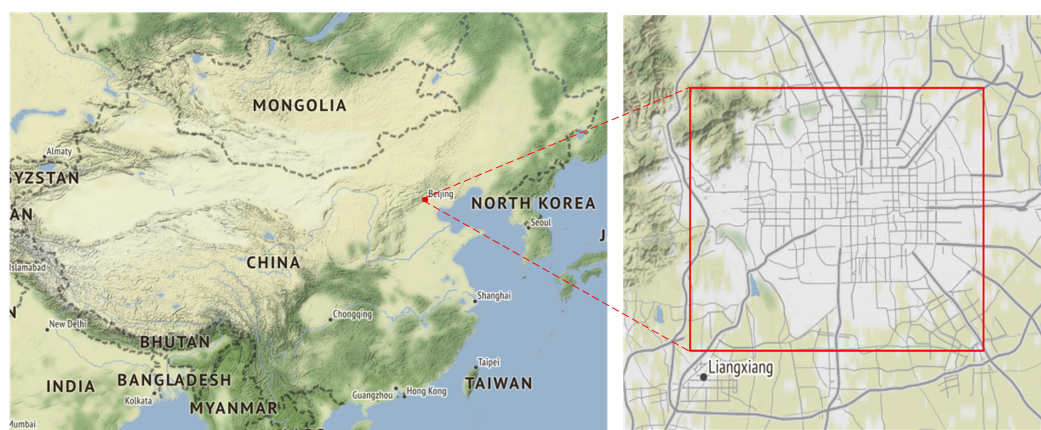


Figure 2. The location of the Beijing air-quality experimental sites.

The main sources of air pollution in Beijing are industrial exhaust gas and automobile exhaust emissions. Our study area covers the Haidian, Shijingshan, Fengtai, Xicheng, and Chaoyang districts, which are the main urban districts in Beijing. The air-quality data collected from the above areas can reflect the overall air-quality condition in Beijing.

2.2. Data Requirement

We evaluated our proposed model with Beijing air-quality data, which comes from the Beijing Municipal Ecological and Environment Monitoring Center. Air-quality data are collected every hour, and the time span is from 1 January 2015 to 31 December 2015. The data are from 17 monitoring stations, and a total of 17×8760 data points were collected. In the experiment, we divided the main urban area of Beijing into 30×38 sub-regions ($1 \text{ km} \times 1 \text{ km}$), in which 17 sub-regions have established monitoring stations, while the air-quality distribution of the remaining 1123 sub-regions are unknown. The external factors include urban meteorological data, the number of points of interest (POIs) in the sub-region, and the geographic road network information in the sub-region (such as the length of highways and main roads, the number of intersections, etc.) [25]. These data features are represented as $F = \{f_1, f_2, \dots, f_m\}$, where f_m represents the m th data feature. The types of POI are shown in Table 1.

Table 1. The types of POIs.

1.	Parks	6.	Companies
2.	Schools	7.	Hotels
3.	Stadium	8.	Supermarkets
4.	Restaurants	9.	Vehicle repair stations
5.	Business center	10.	Gas stations

Based on urban geographic information, the urban area is divided into disjoint grids. Each grid is a sub-area of $1 \text{ km} \times 1 \text{ km}$ and is associated with time-varying air quality values. As shown in Figure 3.

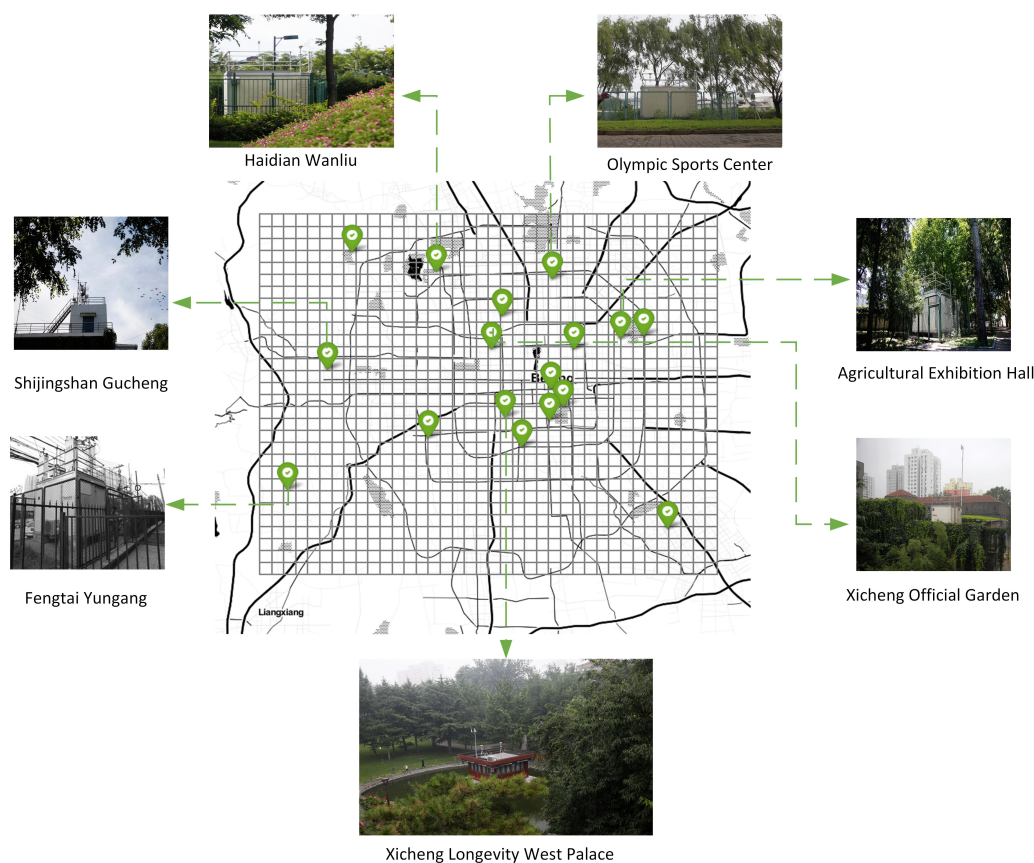


Figure 3. Division of urban areas based on geographic information.

3. Preliminary and Methodology

3.1. Preliminary

3.1.1. Air-Quality Index

The air-quality index (AQI) is a number used by government agencies to communicate to the public how polluted the air is currently. As the AQI increases, an increasingly large percentage of the population is likely to experience increasingly severe adverse health effects. To compute the AQI, an air-pollutant concentration from a monitor or model is required. The function used to convert the air pollutant concentration to AQI varies by pollutants and is different in different countries. The air-quality index values were divided into ranges, and each range was assigned a descriptor and a color code.

3.1.2. Graph Convolutional Networks

Convolutional Neural Networks (CNNs) are mainly effective on Euclidean domain data with regular spatial structures. Generalizing convolutional neural networks on arbitrary graph-structured data is a hot topic in recent researches. Bruna et al. [26] generalized the convolution operator to a general graph via the graph-Laplacian spectrum. By transforming both the filter and the signal to the Fourier domain, convolutions on the graphs can be calculated by multiplying the filter and the signal and then by transforming the result back into the discrete domain. As a meaningful translation operator cannot be expressed in the node domain [26], Defferrard et al. [27] defined the convolution operator $\ast \mathcal{G}$ on graph in the Fourier domain. A signal x filtered by $g_w = \text{diag}(w)$ is as follows:

$$g_w \ast \mathcal{G}x = g_w(L)x = g_w(U\Lambda U^T)x = Ug_w(\Lambda)U^Tx. \quad (1)$$

where U is a matrix of eigenvectors of the normalized graph Laplace matrix L and Λ is a diagonal matrix of eigenvalues of L .

However, evaluating Equation (1) is computationally expensive, as its computational complexity is $\mathcal{O}(N^2)$. To circumvent this problem, the filter $g_w(\Lambda)$ is well approximated by a truncated expansion of K-order Chebyshev polynomials suggested in [28]. Based on [28], Defferrard et al. [27] designed fast localized convolutional filters on graphs to reduce the computational complexity.

$$\begin{aligned} g_w \ast \mathcal{G}x &= g_w(L)x \approx U \sum_{k=0}^{K-1} W'_k T_k(\tilde{A}) U^T x \\ &= \sum_{k=0}^{K-1} W'_k T_k(\tilde{L}) x \end{aligned} \quad (2)$$

where $T_k(\tilde{L})$ is a K-order polynomial about \tilde{L} , $\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N$, and λ_{\max} is the largest eigenvalue of the normalized graph Laplacian matrix L . The computational complexity is reduced from $\mathcal{O}(N^2)$ in Equation (2) to $\mathcal{O}(N)$.

Furthermore, Kipf et al. [29] designed a multi-layer graph convolution network for semi-supervised learning via a first-order approximation of localized spectral filters on graphs. The definition is as follows:

$$Z = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} XW \quad (3)$$

where Z is the convolved signal matrix, $\tilde{A} = A + I_N$ is the graph adjacency matrix with added self-connections, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, and W is a trainable weight matrix. The computational complexity of this convolution operation is the same as $\tilde{A}X$.

As a typical type of graph, based on the spectral graph convolutions in [29], Yu et al. [30] designed a spatiotemporal graph convolutional network for traffic forecasting. Xu et al. [31] designed a spatiotemporal graph multi-fusion network for urban mobile source emission prediction. Sun et al. [32] designed multi-view graph convolutional networks to predict citywide crowd flows in irregular regions. Especially, Cui et al. [33] designed a traffic

higher-order graph convolutional recurrent neural network for network-scale traffic learning and forecasting.

3.1.3. Urban Spatiotemporal Graph

We take each grid element as a node and to construct an urban spatiotemporal graph $G = (V; \varepsilon; A)$ (USTG), where V is the set of nodes associated with the time-varying air-quality value; $|V| = N$ denotes the number of grid nodes; ε is the set of edges, indicating the connectivity between the nodes; and $A \in R^{N \times N}$ is the graph adjacency matrix. In particular, each vertex $v_i \in V$ in the USTG is a three-dimensional representation; the first two dimensions represent the geographic location information of the node, and the third dimension represents the air-quality value that changes with time. In USTG, V can be represented as $V = L \cup U$, where L represents the node set with air-quality-monitoring stations and U represents the node set without monitoring stations. We take the nodes in L as labeled nodes and the nodes in U as unlabeled nodes. Generally speaking, $|L| \ll |U|$. The spatial correlation of the spatiotemporal graph is reflected by the geographical distance characteristics between the nodes, and the temporal correlation is reflected by the air-quality value of different times associated with the nodes.

3.1.4. Inference Problem of Air Quality

Problem 1. At time $\{T_i \mid i = 1, 2, \dots, n\}$, based on USTG $G = (V; \varepsilon; A)$, the air-quality distribution $f_l \in R^{|L| \times P}$ of the labeled nodes L is used to infer the air-pollution distribution $f_u^* \in R^{|U| \times P}$ of the unlabeled nodes U . The task of air-quality inference is to learn a function $F(\cdot)$ that maps $f_l \in R^{|L| \times P}$ to $f_u^* \in R^{|U| \times P}$:

$$F(f_l(T_i), G(V; \varepsilon; A)) = f_u^*(T_i). \quad (4)$$

where $|L|$ is the number of labeled nodes, $|U|$ is the number of unlabeled nodes, and P is the number of features associated with each node after one-hot encoding processing.

3.1.5. Location Recommendation for the Monitoring Station

Problem 2. The main objective of monitoring equipment site location recommendation is to choose optimal practical locations to establish new monitoring equipment at candidate locations. The optimal location can bring about the greatest accuracy improvement to the inference model. There is a relatively close relationship between the monitoring station recommendation model and the air-quality-inference model. We need to analyze and correlate these two models and to design an effective algorithm based on a correlation degree to complete the monitoring-station location recommendation.

3.2. Methodology

In this section, we introduce the air-quality-inference model and the monitoring-station location recommendation model. We first introduce the conception of higher-order graph convolutional networks. Then, we design an air-quality-inference model based on the improved HGCN, which is applied effectively to the field of atmospheric environment monitoring. Next, we analyze the air-quality inference model and correlate it with the air-quality-monitoring station location recommendation. We design a greedy algorithm for minimizing information entropy based on the correlation degree between nodes in USTG, marking the recommendation priorities of unlabeled nodes according to the ability to improve the inference model accuracies using the incremental learning method to complete the monitoring-equipment station location recommendation.

The proposed two-step air-quality-inference learning framework is shown in Figure 4.

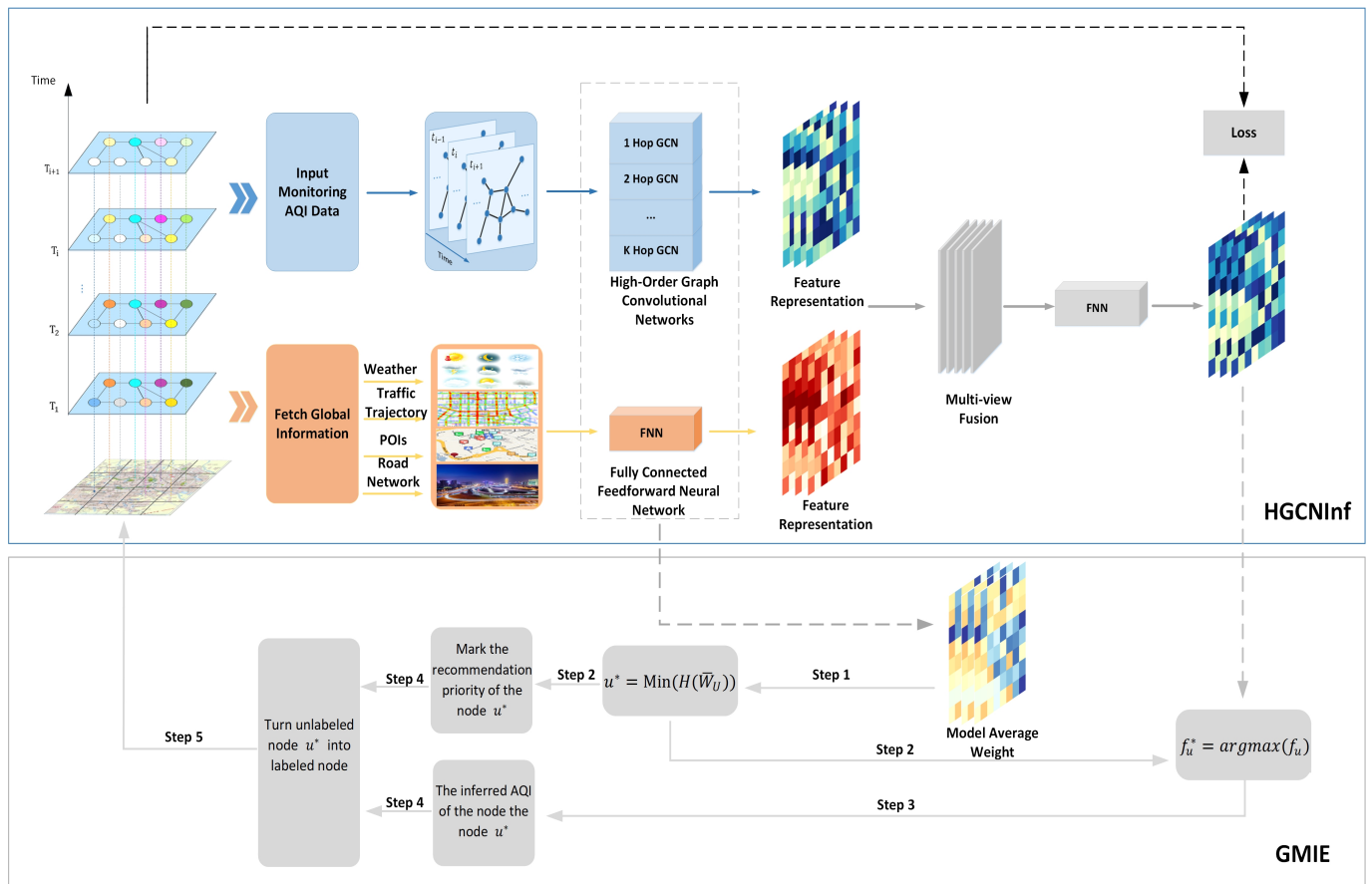


Figure 4. The general framework. (1) Air-quality-inference model: HGCNInf consists of a higher -order graph convolution network (HGCN) and several fully connected neural networks (FNN). By inputting historical monitoring data and by fetching global information, the HGCN is used to capture the spatiotemporal interactions of air-quality distribution and the FNN is used to extract the features of external factors. (2) Monitoring-station location recommendation model: by using the convolution weight parameters of HGCNInf, we designed a GMIE based on the correlation degree between nodes in USTG reflecting the air-quality spatiotemporal changes, marking the recommendation priorities of unlabeled nodes according to the ability to improve the inference accuracy of HGCNInf via the node incremental learning method.

(1) **Air-quality inference model.** We propose an effective air-quality inference model based on the proposed higher-order graph convolution network (HGCNInf) to infer the air-quality distribution, consisting of a higher-order graph convolution network (HGCN) and several fully connected neural networks. By inputting historical monitoring data and by fetching global information, the HGCN is used to capture the spatiotemporal interactions of air-quality distribution and the fully connected neural network is used to extract external influential factor features, so that HGCNInf can accurately infer the air-quality distribution of an entire urban area.

(2) **Monitoring-station location recommendation model.** HGCNInf takes the spatiotemporal interactions of pollutant distribution and external influential factors into account, in which the model convolution weight parameters actually represent the correlation degree between the nodes in USTG that reflect the spatiotemporal changes in air quality. The high correlation between a node and other nodes indicates that the air quality of the node is easier to infer. Therefore, for the station location recommendation, it should be chosen in such a node that has a low correlation with other nodes. In this paper, we introduce the information entropy index to evaluate the correlation degree between nodes. The node with the lower information entropy indicates that it has a higher correlation degree with the remaining nodes. Next, we designed a greedy algorithm for minimizing information entropy (GMIE) based on the correlation degree between nodes in

USTG, marking the recommendation priorities of unlabeled nodes according to the ability to improve the inference accuracies of HGCNInf through the node incremental learning method to complete the monitoring station location recommendation.

3.2.1. Higher-Order Graph Convolutional Network

Some recent researches [34–36] have shown that, for graph node semi-supervised learning, random walk statistics can be used to effectively learn feature representations because it could preserve the completeness of the graph structure. Under special conditions, if the activation function of the first layer in the graph convolutional network is an identity function, then the multi-layer GCN model can also learn random walks. For the 2-layer GCN model proposed in [29], we assume that the activation function of the first layer is an identity function:

$$\sigma(\hat{A}XW^{(0)}) = \hat{A}XW^{(0)} \quad (5)$$

In this way, the GCN model can be further simplified to the following form:

$$\begin{aligned} Z = f(X, A) &= \text{softmax}(\hat{A}(\hat{A}XW^{(0)})W^{(1)}) \\ &= \text{softmax}(\hat{A}^2XW^{(0)}W^{(1)}) \\ &= \text{softmax}(\hat{A}^2XW^{(2)}) \end{aligned} \quad (6)$$

where $W^{(2)} = W^{(0)}W^{(1)}$ is expressed as the weight parameter of \hat{A}^2 , $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$. The two-layer one-hop graph convolutional network is converted into a one-layer two-hop convolutional network. Furthermore, \hat{A}^2 can be expressed as follows:

$$\begin{aligned} \hat{A}^2 &= (\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}})(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}) \\ &= \tilde{D}^{-\frac{1}{2}}\tilde{A}(\tilde{D}^{-1}\tilde{A})\tilde{D}^{-\frac{1}{2}} \\ &\propto \tilde{D}^{-1}\tilde{A} \end{aligned} \quad (7)$$

We can find that the 2-hop GCN network under certain conditions actually learns a one-step random walk, which can transfer the information of the node to the higher-order neighbor nodes. For the 2-layer GCN model proposed in [29], the flow of the input signal through the graph edges is reduced because the information features have gone through matrix multiplication (average neighbor node information) and a nonlinear activation function $\sigma(\cdot)$. Therefore, it is not possible to directly take higher-order operations on \hat{A} based on the original multi-layer GCN network; otherwise, it causes network performance degradation. In order to make better use of the information of high-hop neighbor nodes, we increase the order of the graph convolution network while reducing the layer numbers of the network and then fuse the features extracted from each order network.

$$\hat{A}^k = \tilde{D}^{-\frac{1}{2}}\tilde{A}\left[(\tilde{D}^{-1}\tilde{A})^{(k-1)}\right]\tilde{D}^{-\frac{1}{2}} \quad (8)$$

$$GC^k = \text{GCN}(\hat{A}^k, X) = \hat{A}^kXW^{(k)} \quad (9)$$

$$\begin{aligned} Z = f(X, A) &= GC^{\{k\}} \\ &= \text{softmax}([GC^1, GC^2, \dots, GC^k]) \end{aligned} \quad (10)$$

Thus, the newly designed higher-order graph convolutional network uses fewer layers and parameters but can achieve better results.

3.2.2. The Air-Quality Graph Convolution

We first define the graph higher-order adjacency matrix. The one-hop neighbor matrix of the USTG is the adjacency matrix A , and the k -hop adjacency matrix can be obtained by the k th power of A . As shown in Figure 5, the multi-hop neighbor with respect to a node (red star) is the nodes that can be reached by the black edges in the graph.

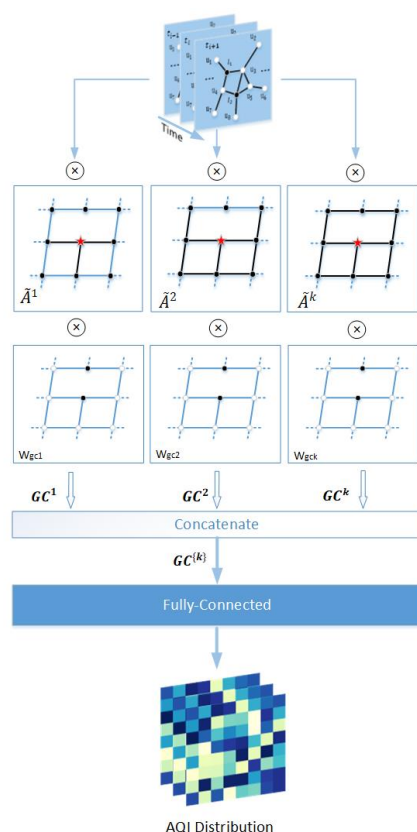


Figure 5. The architecture of proposed higher-order convolutional network. The multi-hop neighbor with respect to a node (the red star) is the nodes that can be reached by the black edges.

We apply the proposed higher-order graph convolutional network to the field of air-quality monitoring. The problem of air-quality inference is specifically described in Section 3.1.4 and the proposed HGCNInf framework is shown in Figure 6.

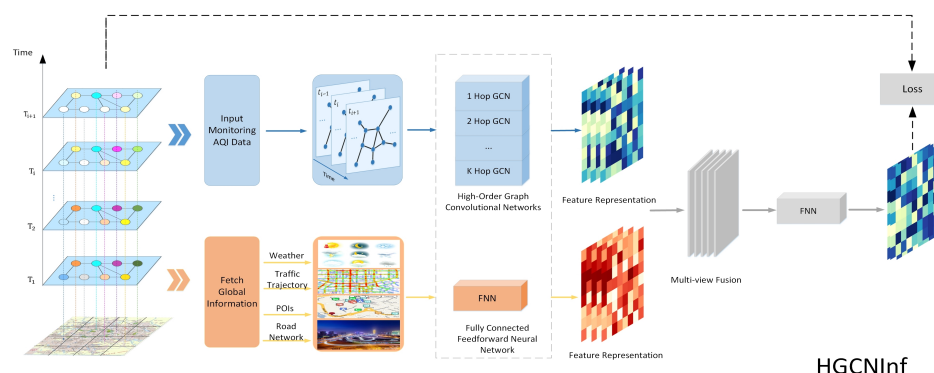


Figure 6. The framework of the air-quality-inference model. The specific structure of the higher-order graph convolutional network is shown in Figure 4.

In order to make the subsequent air-quality monitoring station location recommendation model more reasonable and interpretable, we modified the graph convolution network GC^k described in Equation (9) to the following form.

$$GC^k = GCN(\hat{A}^k, X) = \hat{A}^k W^{(k)} X \quad (11)$$

where $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \in R^{N \times N}$, D is the degree matrix of A , and $W^{(k)}$ is the convolution weight parameter. The graph input signal $X = \begin{bmatrix} f_l \\ f_u \end{bmatrix} \in R^{N \times P}$, where $f_l \in R^{|L| \times P}$ is the air-quality distribution of the labeled nodes, initializing the air-quality distribution of unlabeled nodes to $f_u = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \in R^{|U| \times P}$.

Then, we concatenated the K-hop features extracted from the graph signal, defined as follows:

$$GC^{\{k\}} = [GC^1, GC^2, \dots, GC^k] \quad (12)$$

where $GC^{\{k\}} \in R^{N \times (K \times P)}$ is a set of K-hop features extracted by graph convolutional network. We fed $GC^{\{k\}}$ into a fully connected network f_{fc1} to jointly train

$$f_{fc1} : R^{N \times (K \times P)} \rightarrow R^{N \times P} \quad (13)$$

$$GC = f_{fc1}(GC^{\{k\}}, W_{fc1}) \quad (14)$$

where air-distribution feature $GC \in R^{N \times P}$ and the trainable parameter matrix $W_{fc1} \in R^{P \times (K \times P)}$.

We use the fully connected network f_{fc2} to extract external factor features, defined as follows:

$$f_{fc2} : R^{N \times M} \rightarrow R^{N \times P} \quad (15)$$

$$GC_f = f_{fc2}(f_m, W_{fc2}) \quad (16)$$

where the trainable parameter matrix $W_{fc2} \in R^{P \times M}$; external factor features $GC_f \in R^{N \times P}$; and $f_m \in R^{N \times M}$, which is the M external factors.

We concatenated the air-quality distribution feature $GC \in R^{N \times P}$ with the external factor feature $GC_f \in R^{N \times P}$ and then fed them into the fully connected network f_{fc3} , defined as follows:

$$f_{fc3} : R^{N \times P} \rightarrow R^{N \times P} \quad (17)$$

$$GC_{all} = f_{fc3}((GC + GC_f), W_{fc3}) \quad (18)$$

where the trainable parameter matrix $W_{fc3} \in R^{P \times P}$ and general feature $GC_{all} \in R^{N \times P}$.

Further, the air quality distribution $F_u \in R^{|U| \times P}$ of unlabeled nodes is obtained, defined as follows:

$$F = \text{softmax}(GC_{all}) \quad (19)$$

$$F = \begin{bmatrix} F_l \\ F_u \end{bmatrix} \in R^{N \times P} \quad (20)$$

where $F_l \in R^{L \times P}$ is the air-quality distribution output of the labeled nodes and $F_u \in R^{U \times P}$ is the air-quality distribution output of the unlabeled node.

We minimized the cross-entropy loss of the output F_l and the known training label Y . The cross-entropy loss is defined as follows:

$$H(Y, F_l) = -\text{diag}(V_l) \sum [Y * \log(F_l)] \quad (21)$$

where $*$ represents the Hadamard product and where $\text{diag}(V_l)$ is a diagonal matrix where elements are set to 1 when $(i, i) \in V_l$ and 0 otherwise. In order to make the graph convolution features more stable and interpretable, we added regularization terms to the loss function.

(1) We added the L1-norm of the graph convolution weight matrix to the loss function as a regular term to make these weight matrices as sparse as possible. L1 regularization is defined as follows:

$$R_1 = \|W_{gci}\|_1 = \sum_{i=1}^k |W_{gci}| \quad (22)$$

(2) In order to limit the differences between adjacent features extracted from graph convolution, the convolution feature regularization term based on the L2 norm is added to the loss function. L2 regularization is defined as follows:

$$R_2 = \|GC^{\{k\}}\|_2 = \sqrt{\sum_{i=1}^{k-1} (GC^i - GC^{i+1})^2} \quad (23)$$

Thus, the total loss function is defined as follows:

$$\text{Loss} = H(Y, F_l) + \alpha \cdot R_1 + \beta \cdot R_2 \quad (24)$$

where the parameters α and β are used to control the weight of the regularization term.

Next, we minimized the model loss function to train the network as follows:

$$\begin{aligned} \min \text{Loss} = & \min [-\text{diag}(V_l) \sum [Y * \log(F_l)] \\ & + \alpha \sum_{i=1}^k |W_{gci}| \\ & + \beta \sum_{i=1}^{k-1} (GC^i - GC^{i+1})^2] \end{aligned} \quad (25)$$

When the network training was completed, we found the value with the highest quantization probability from the air-quality distribution F_u and defined f_u^* as the inferred air-quality value. The definition is as follows:

$$f_u^* = \text{argmax}(F_u) \quad (26)$$

3.2.3. Air-Quality-Monitoring Station Location Recommendation

The main objective of this section is use the proposed HGCNInf to complete the location recommendation for a air-quality-monitoring station. The ultimate goal is to recommend an optimal station location from candidate locations (this paper defaults to all unlabeled nodes) that could bring about the greatest accuracy improvement to the inference model. Therefore, we need to correlate the air-quality-inference model with the monitoring-station location recommendation model. For the trained higher-order graph convolutional network, its weight parameters reflect the correlation degree between nodes. Since the designed HGCNInf takes into account both the spatiotemporal interactions of air pollutant distribution and the complex external factors, its convolution weight parameter

represents the correlation degree between the nodes reflecting the spatiotemporal changes in air-quality distribution. This is of great significance to the monitoring-equipment location selection.

In the USTG, a higher correlation degree between a node and the remaining nodes indicate that the air quality of the node is easily inferred. Therefore, the monitoring station should be selected at a node that has the lowest correlation with the remaining nodes. We introduced the information entropy index to evaluate the correlation degree between nodes. The smaller the information entropy value, the higher the correlation degree between nodes. We used the convolution weight parameters of the trained HGCNInf to recommend optimal monitoring station locations that could also bring about the greatest accuracy improvement to HGCNInf. In this way, the two problems are well correlated.

Greedy algorithm based on information entropy minimization

We designed a greedy algorithm for minimizing information entropy (GMIE) based on the correlation degree between nodes in USTG to recommend optimal station locations. HGCNInf is used as the driving model through the incremental learning method, iteratively marking the recommendation priority of unlabeled nodes.

The basic framework is shown in Figure 7, and the specific steps are as follows from step 1 to step 6:

1. Calculate the average convolution weight parameter of the trained HGCNInf. The average convolution weight parameter is defined as follows:

$$\bar{W} = \frac{1}{k} \sum_{i=1}^k \tilde{A}^i |W_{gci}| \quad (27)$$

where k is the order of the higher-order graph convolutional network and \tilde{A} is the standardized adjacency matrix. Next, calculate the information entropy of the average convolution weight parameter of each node in the unlabeled node U , which is defined as follows:

$$H(\bar{W}_u) = - \sum_{u \in U} \bar{W}_u \log \bar{W}_u \quad (28)$$

2. Find the unlabeled node $u' \in U$ with the smallest information entropy value in $H(\bar{W}_u)$, and mark its recommendation priority as the lowest. Based on the air-quality distribution f_u inferred from HGCNInf, associate the inferred air-quality value with the node u' and then turn the node u' from the unlabeled nodes into the labeled nodes to constitute a new graph input signal data f'_l . Remove the node $u' \in U$ from the unlabeled node U , and the remaining unlabeled nodes are represented as U' .

3. Based on the new graph input data f'_l , drive HGCNInf to retrain and update the network convolution weight parameters, that is, update the correlation degree between the nodes in USTG reflecting the spatiotemporal changes in air-quality distribution and then derive the new air-quality distribution f'_u of unlabeled nodes. As described in step 1, new average convolution weight parameters \bar{W}' and information entropy $H(\bar{W}_u)'$ can be calculated.

4. Find the unlabeled node $u'' \in U'$ with the smallest information entropy value in $H(\bar{W}_u)'$, and mark its priority as the second lowest. Based on the new air-quality distribution f'_u inferred from the air-quality inference model, associate the inferred air-quality value with the node u'' , and then, turn the node u'' from the unlabeled nodes into the labeled nodes to constitute a new graph input signal data f''_l . Remove the node $u'' \in U'$ from the unlabeled node U' , and the remaining unlabeled nodes are represented as U'' .

5. Repeat step 1 to step 4 above until the recommendation priority of all unlabeled nodes U is marked.

6. Furthermore, input graph signal data at other moments and repeat step 1 to step 5.

Finally, analyze the results at multiple times and average the recommendation priority of unlabeled nodes, and then recommend the Top-N nodes with the highest average recommendation priority as the new monitoring station location. These Top-N sites with the highest average recommendation priority have the highest average information entropy, that is, the lowest correlation degree with the remaining nodes. Therefore, newly established monitoring equipment at these Top-N sites could bring about the greatest accuracy improvement to HGCNInf, so that it can reflect the air quality and its development trends in a timely and accurate manner.

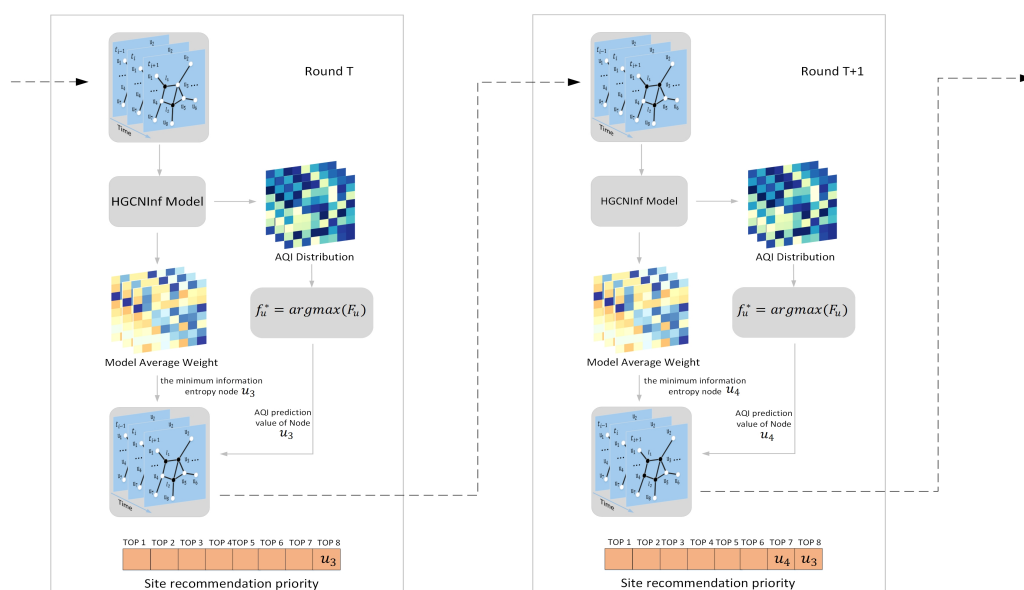


Figure 7. The basic framework of the monitoring-station location recommendation model.

4. Experiments and Results

4.1. Data Processing

The graph input data in this paper are the PM₁₀ values of Beijing and the complex external influential factor data. We used one-hot encoding to process PM₁₀ data and used Z-Score Normalization to process the external influential factor data, the factor data were linearly transformed to be mapped into the interval [0, 1].

The adjacency matrix of the Beijing urban spatiotemporal graph was calculated by using the threshold Gaussian kernel weighting function based on the distance between nodes [30,37]. The calculation formula is as follows:

$$W_{ij} = \begin{cases} \exp\left(-\frac{\text{dist}(p_i, p_j)^2}{\sigma^2}\right), & \exp\left(-\frac{\text{dist}(p_i, p_j)^2}{\sigma^2}\right) \geq \varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (29)$$

where $\text{dist}(p_i, p_j)^2$ represents the Euclidean distance between the nodes p_i and p_j , and σ^2 and ε are the thresholds used to control the distribution and sparsity of the adjacency matrix. In this experiment, σ^2 and ε are set to 10 and 0.05 respectively.

4.2. Experimental Settings

4.2.1. Evaluation

In order to evaluate and compare the performance between different models, we used two commonly used metrics: (1) Mean Absolute Percentage Error (MAPE); (2) Root Mean Squared Error (RMSE). They are defined as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| * 100\% \quad (30)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (31)$$

where \hat{Y}_i and Y_i are the predicted values and the real values from selecting labeled nodes for model validation, and n is the number of all predicted nodes.

4.2.2. Baselines

We compared HGCNInf and GMIE with the following baseline models.

(1) HA: Historical average. HA takes the average value of air-quality data as the predicted value.

(2) SVR: Support vector regression [38]. SVR uses the historical AQI data of each monitoring station as training data to infer the air-quality value of unlabeled nodes.

(3) AQInf: Affinity-based AQI inference model [23]. For air-quality inference problem, this is the current state-of-the-art method. It is used to infer the air-quality distribution of the unlabeled nodes by using the AQI data of the labeled ones and by combining the external influential factors.

(4) GCNInf: An air-quality inference model based on graph convolutional networks (GCNInf). We built a two-layer graph convolutional networks as described in [29]. Similar to AQInf, GCNInf was used to infer the air-quality distribution of unlabeled nodes by using the AQI data of the labeled ones and by combining with the external influential factors.

All of the neural network model implementations are based on Pytorch version: 1.1.0, and evaluation and training are based on CPU Intel(R) Core(TM) i5-4210U CPU at 1.70GHz, NVIDIA GeForce 840M.

4.3. Experimental Results

4.3.1. The Effectiveness of HCNInf

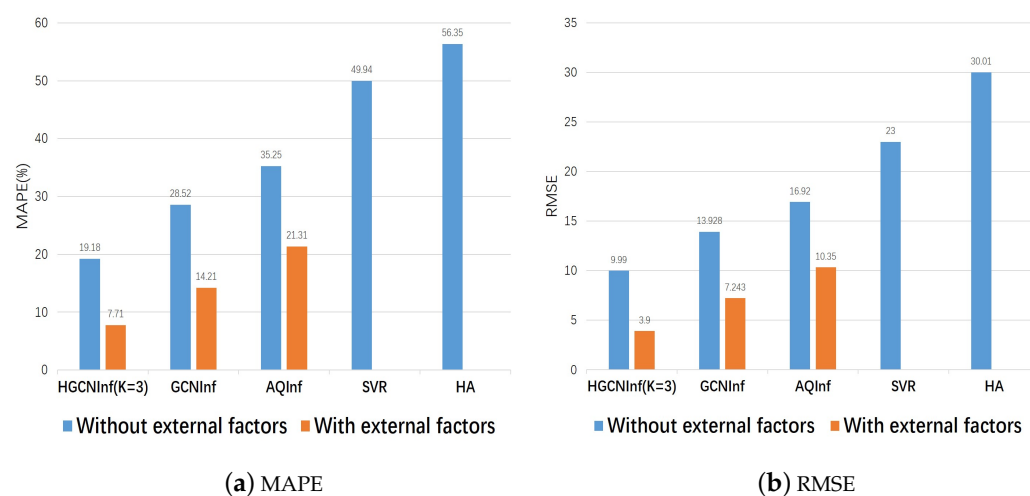
In the experiment, since we only have real data of the labeled nodes (the sub-regions with monitoring stations), for the unlabeled nodes (the sub-regions without monitoring stations), there are no real data to verify the inference accuracy of each model. Therefore, we proposed to conduct cross validation by randomly selecting 12 nodes as labeled nodes from the 17 labeled ones, which generated 12×8760 training data, and the remaining 5 nodes were used to verify the inference accuracy of each model, which generated 5×8760 verification data.

(1) The performance comparison of air-quality-inference models

We trained the above air-quality-inference models separately three times and then took the average values as the final results. Table 2 and Figure 8 show the performance comparison of our proposed HGCNInf and other baseline models on the inferred Beijing air-quality distribution in two metrics. It can be found that HGCNInf is far superior to other baseline models both in the metric MAPE and RMSE.

Table 2. The performance comparison of inference models.

Model	Beijing Air-Quality Data Evaluation			
	MAPE (%)		RMSE	
	Without External Factors	With External Factors	Without External Factors	With External Factors
HA	56.35	/	30.01	/
SVR	49.94	/	23.00	/
AQInf	35.25	21.31	16.92	10.35
GCNInf	28.52	14.10	13.93	7.24
HGCNInf	19.18	7.71	9.99	3.90

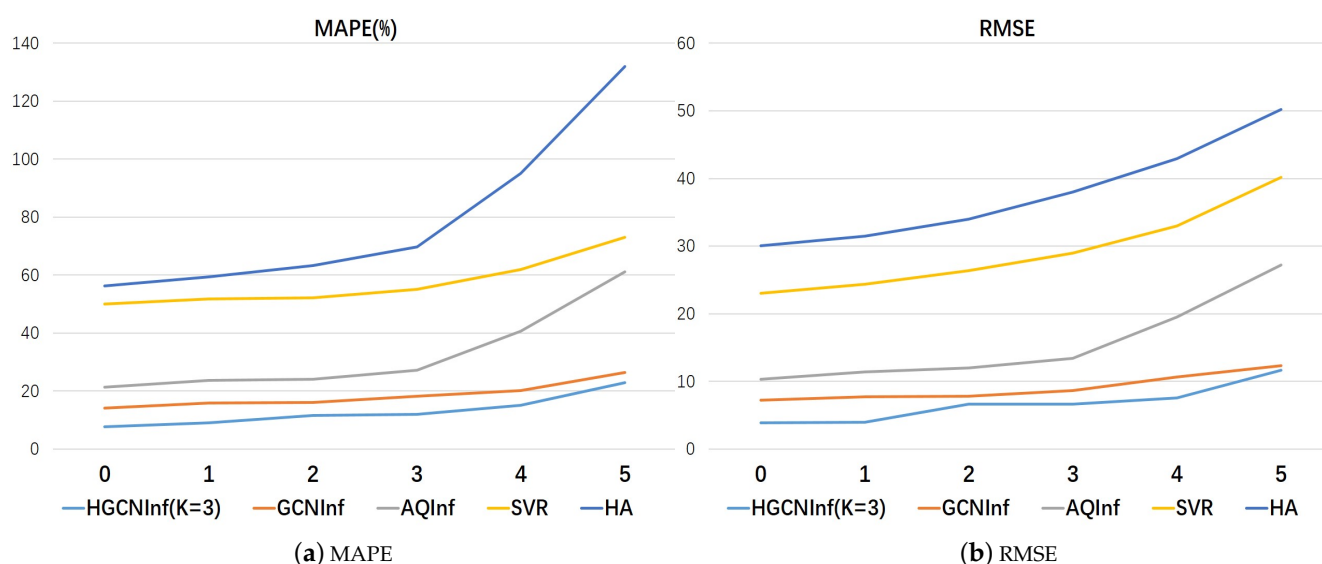
**Figure 8.** The performance comparison of each inference model.

In addition, comparing the results of the air-quality-inference models with external factors and the models without external factors, we can find that, for the same model, after considering complex external influential factors, the error in the air-quality inference decreased significantly. Therefore, the external influential factors play a significant role in improving the inference accuracy of air-quality distribution. We can conclude that the complex external influential factors are closely related to the air-quality distribution, which also explains why the monitoring stations that are geographically close to each other in the red circle differ greatly in the monitoring values throughout the year.

Furthermore, to test the robustness of each air-quality-inference model, we randomly discarded the same labeled nodes and observed the change degree in the air-quality-inference performance. Table 3 and Figure 9 show the inference results of HGCNInf and other baseline models on two evaluation metrics when we gradually randomly discarded the same labeled nodes. From the above chart, we can find that our proposed air-quality-inference model HGCNInf is far more robust than other baseline models.

Table 3. The robustness comparison of inference models.

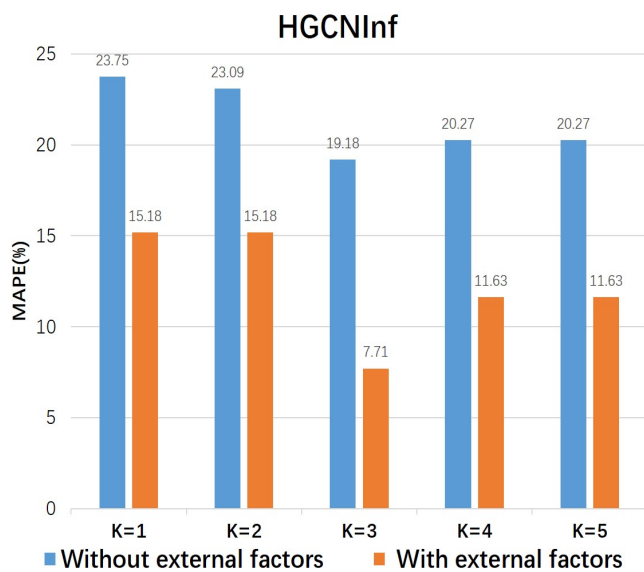
The Robustness of Each Air Quality Inference Model											
Model	HA		SVR		AQInf		GCNInf		HGCNInf		
Metric	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	
Number											
0	56.35	30.01	49.94	23.00	21.31	10.35	14.10	7.24	7.71	3.90	
1	59.45	31.50	51.75	24.40	23.75	11.41	15.82	7.74	9.01	3.98	
2	63.38	34.02	52.15	26.4	24.15	12.01	16.13	7.82	11.63	6.62	
3	69.67	37.99	55.18	29.00	27.18	13.39	18.31	8.61	12.07	6.66	
4	95.05	42.96	62.00	33.00	40.68	19.52	20.27	10.69	15.18	7.52	
5	131.92	50.18	73.01	40.18	61.06	27.18	26.42	12.34	22.94	11.67	

**Figure 9.** The robustness of each air-quality-inference model when gradually reducing the number of labeled nodes.

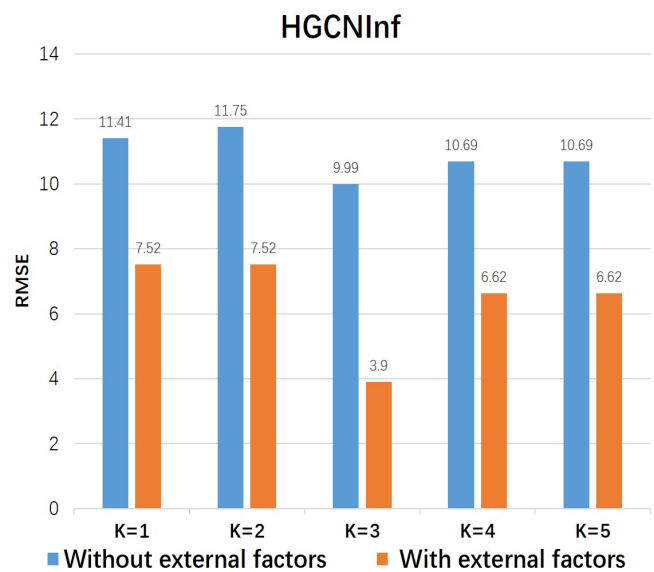
(2) Training efficiency of HGCNInf We considered the influences of different order K of the graph convolution in HGCNInf on inference performance. As shown in the histogram of Figure 10a,b and Table 4, they showed the performance comparison of different orders of HGCNInf. For Beijing's air-quality-distribution data, when we chose the order K of the higher-order convolutional network to be 3, the model HGCNInf had the best performance. Figure 8c and Table 5 show the average training times per 1000 times at each moment when the HGCNInf model takes different orders. The training time increases as the order of the higher-order convolutional network increases. In addition, when the HGCNInf model considers the external influential factors, the time consumed by training increases significantly. Comprehensively considering the model performance and training efficiency, we chose $K = 3$ as the order of the high-order convolutional network in the HGCNInf model, and the results are shown in Table 2.

Table 4. The HGCNInf performance of different orders.

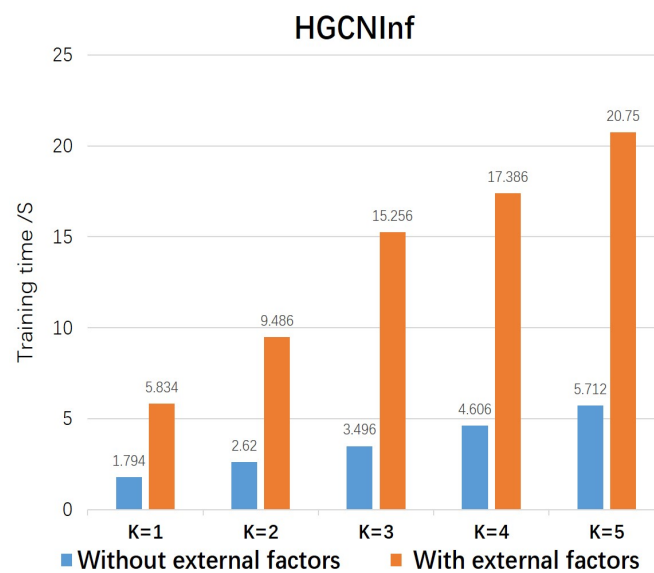
		Order				
		K = 1	K = 2	K = 3	K = 4	K = 5
MAPE (%)	Without external factors	23.75	23.09	19.18	20.27	20.27
	With external factors	15.18	15.18	7.71	11.63	11.63
RMSE	Without external factors	11.41	11.75	9.99	10.69	10.69
	With external factors	7.52	7.52	3.90	6.62	6.62



(a) MAPE



(b) RMSE



(c) Training time

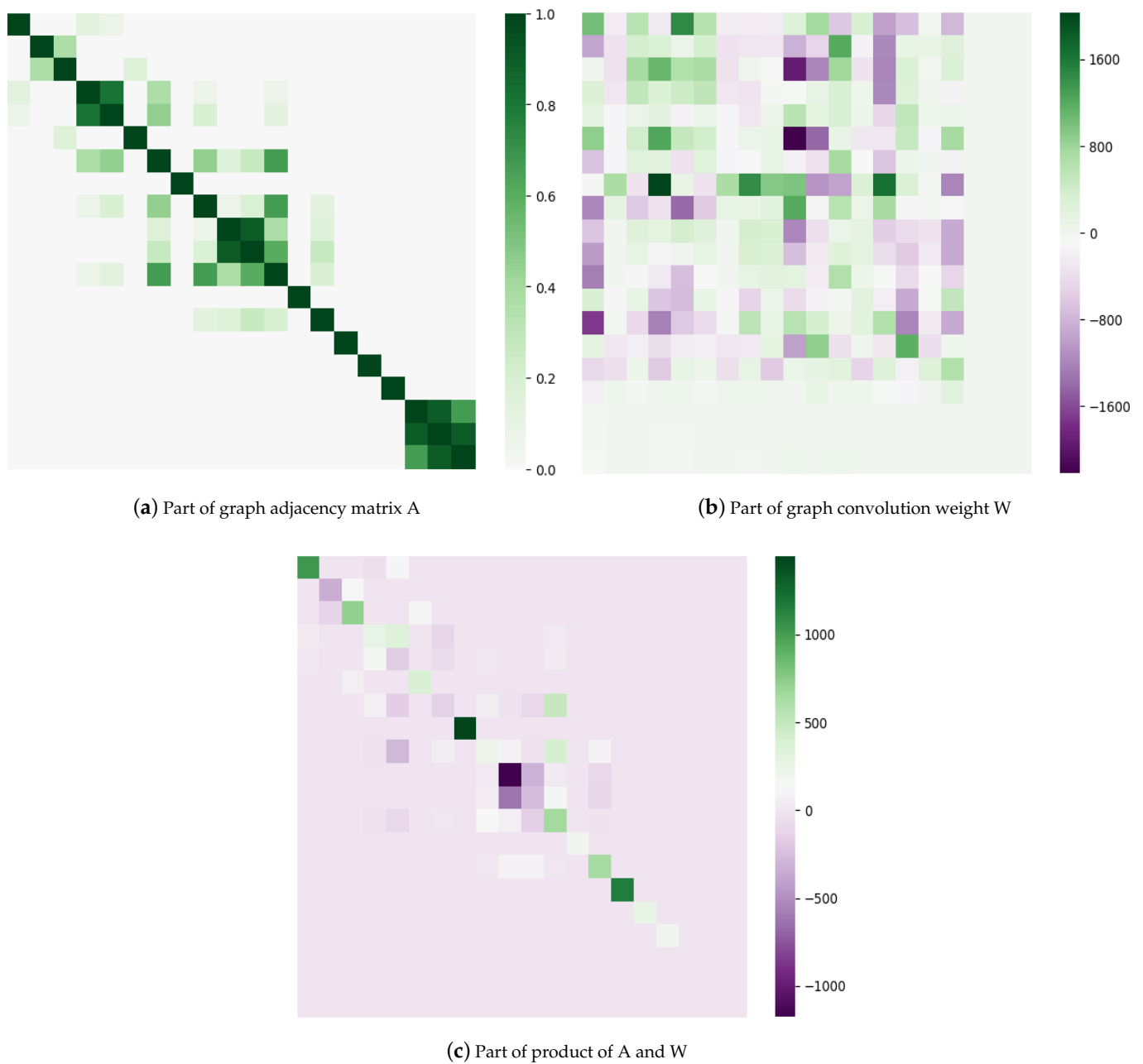
Figure 10. The HGCNInf training efficiency of different orders.

Table 5. The HGCNInf training time of different orders.

Training Time/(s)	K = 1	K = 2	K = 3	K = 4	K = 5
Without external factors	1.79	2.62	3.50	4.61	5.71
With external factors	5.83	9.49	15.26	17.39	20.75

(3) Weight visualization

To better understand the HGCNInf, we visualized a proportion of the average graph convolution weight W , graph adjacency matrix A , and their product of $A \cdot W$, as shown in Figure 11.

**Figure 11.** The HGCNInf weight visualization.

4.3.2. The Effectiveness of GMIE

The main objective of this section is to recommend several optimal locations from the candidate locations to establish new monitoring stations. In this experiment, we verify whether the nodes recommended by our proposed monitoring-station location recommendation model (GMIE) can maximize the inference accuracy of HGCNInf for inferring the air-quality distribution of the entire city. Since there is no ground real air-quality data for unlabeled nodes, it is impossible to verify the specific accuracy improvement that the recommended nodes bring to the air-quality-inference models.

We adopt a cross-validation method to randomly select 7 nodes from 17 labeled nodes as training data, and then, we reserve 5 labeled nodes as candidate sites to establish new monitoring equipment. The remaining 5 labeled nodes are used to verify the performance improvement that the recommended nodes bring to the air-quality-inference models.

Baselines. Since to our knowledge, AQInf+GEM is currently the state-of-the-art method for monitoring-station location recommendation [18], we compared our proposed HGCNInf+GMIE with GCNInf+GMIE and AQInf+GEM. The recommendation priorities of the candidate locations marked by each monitoring-station location recommendation model are shown in the Figure 12.

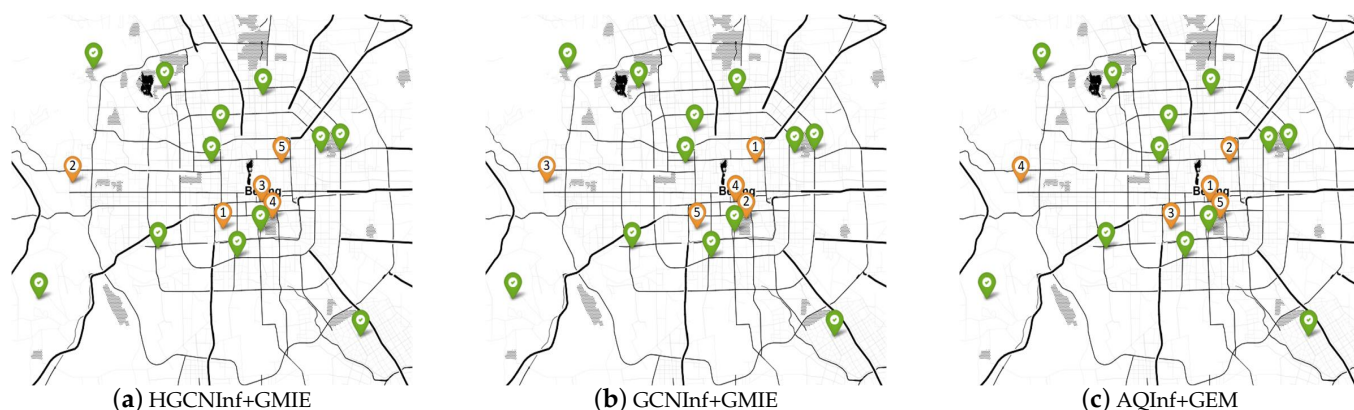


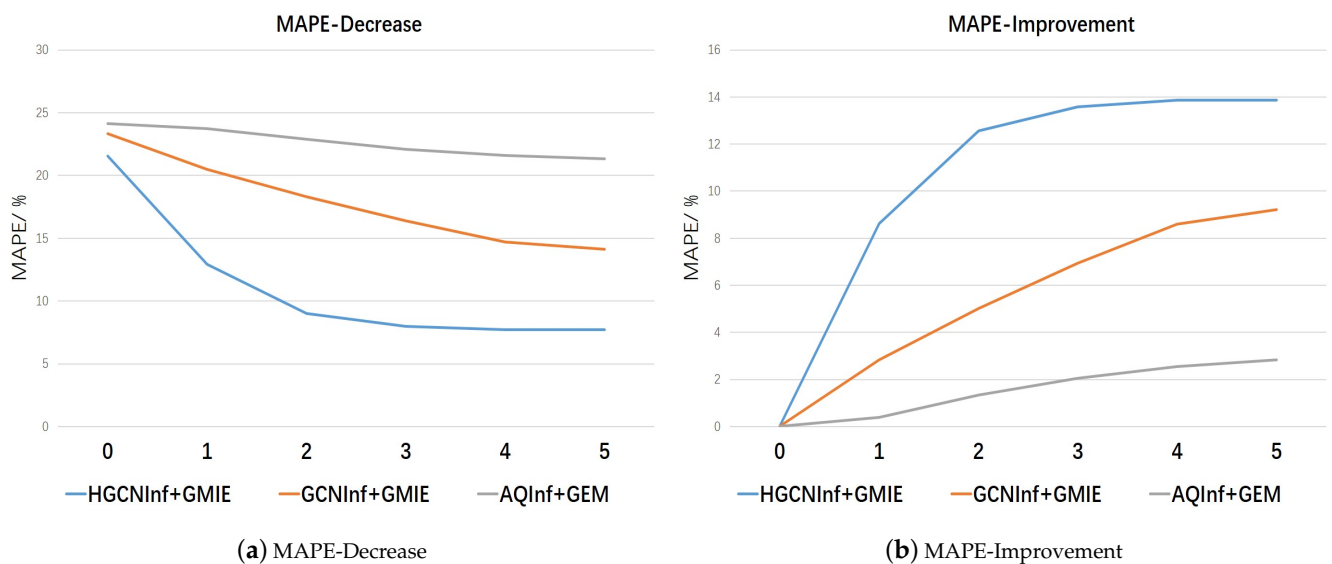
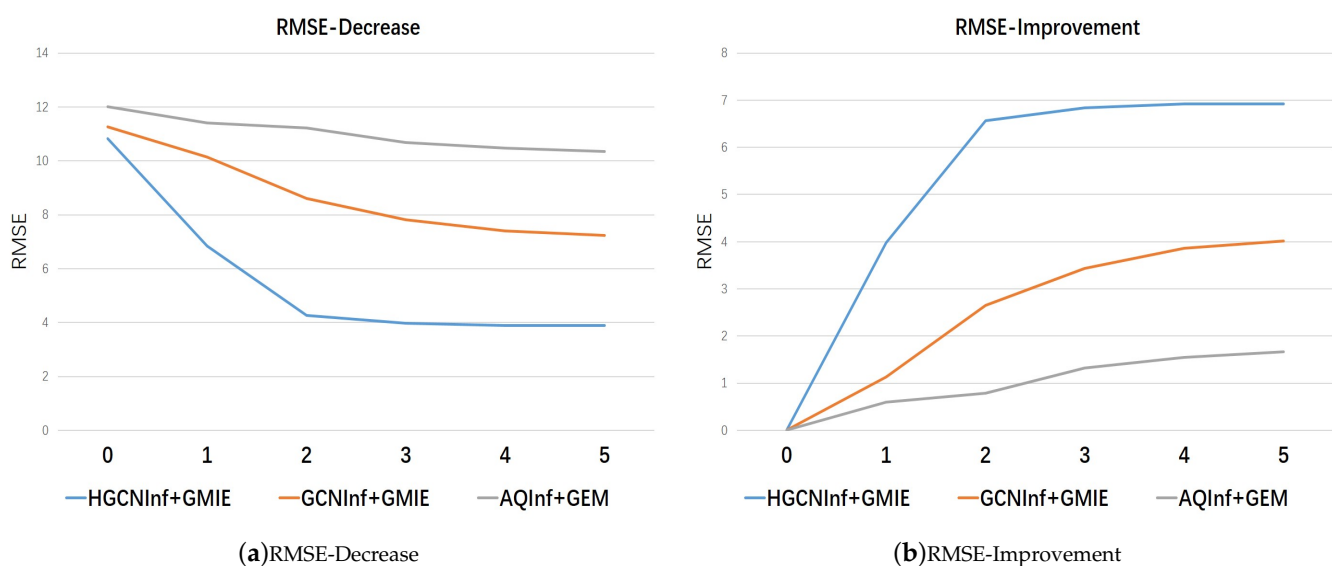
Figure 12. The recommendation results of each model. The nodes marked with orange symbols in the figure are candidate nodes. Each candidate node has a recommendation priority marked by Arabic numerals (the smaller the number, the higher the recommendation priority).

We use MAPE improvement and RMSE improvement to evaluate the effectiveness of monitoring-station location recommendation models. In theory, as the number of recommended station locations increases, the accuracy of the corresponding air-quality inference model should be greatly improved because more station locations are added into the labeled nodes. That is, MAPE and RMSE should be greatly reduced. Since the candidate locations provide the recommendation priority, the node with a higher recommendation priority can bring about greater improvement to the air-quality-inference model. Therefore, when gradually increasing the number of recommended nodes according to the recommendation priority, the declining slopes of MAPE and RMSE should become smaller and smaller or the rising slopes of MAPE improvement and RMSE improvement should become smaller and smaller.

We display how much both MAPE improvement and RMSE improvement models can be obtained by gradually adding more recommended station locations shown in Table 6, and Figures 13 and 14. We can find that the performance of each station location recommendation model is consistent with the theory and that our proposed model's performance has far exceeded the state-of-the-art method, which can bring about the greatest accuracy improvement to HGCNInf for inferring the air quality.

Table 6. The improved MAPE and RMSE results with varying numbers of recommended locations.

The Improved MAPE and RMSE Results with the Increase in the Number of Recommended Nodes							
Model		HGCNInf+GMIE		GCNInf+GMIE		AQInf+GEM	
Number	Metric	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE
0		21.57	10.83	23.31	11.26	24.15	12.01
1		12.94	6.84	20.48	10.13	23.75	11.41
2		9.01	4.27	18.30	8.61	22.80	11.22
3		8.00	3.98	16.37	7.82	22.10	10.69
4		7.71	3.90	14.70	7.40	21.60	10.47
5		7.71	3.90	14.10	7.42	21.31	10.35

**Figure 13.** The MAPE results when gradually increase the number of recommended stations according to its priority.**Figure 14.** The RMSE results when gradually increase the number of recommended stations according to its priority.

5. Discussion

5.1. Air-Quality Inference

The knowledge-driven method is based on the formation mechanism of air pollutants and the actual physical diffusion and migration laws, and then a suitable model is established, and the air-quality distribution in urban areas not covered by the monitoring station is predicted through calculation and simulation. Gibson et al. [39] used a Gaussian plume diffusion model to predict the PM_{2.5} distribution of main points and routes. Rakowska et al. [40] constructed an urban street canyon model to monitor the impact of traffic flow on air quality. These methods are based on a complex mechanism knowledge, and it is difficult to establish a universal model for various environments due to the adopted empirical assumptions. For example, the Gaussian plume diffusion model that is widely used in the field of air pollution assumes that the distribution of air pollutants conform to the Gaussian distribution. However, in an urban street environment, the diffusion of mobile source pollutants on the road is affected by the buildings on both sides of the road, which does not satisfy this model.

Data-driven methods are also constantly developed and have achieved good results in fine-grained air-quality prediction in space. Zheng et al. [8] considered the influence of external factors related to air-quality distribution and proposed a semi-supervised learning method based on the co-training collaborative training framework [41] to infer the fine-grained air-quality distribution in urban space. However, the framework consists of two independent classifiers, and it is difficult to capture the complex interactions between different features. Chen et al. [42] used the K-nearest neighbor strategy to construct the relationship between external data sources (such as road network structure, traffic, etc.) and air-quality distribution. However, there is no exact theoretical standard for the specific value of K. The characteristics of the latest K sites may not be the most effective, and the degree of influence of different sites may change over time, which may lead to inconsistency problems. Hsieh et al. [23] designed an affinity-based air-quality-inference model AQInf (the affinity-based AQI inference model (AQInf)) based on graph label node propagation. However, due to the limitation of its own model performance, it is difficult for AQInf to capture the correlation between graph nodes that reflect the trend for air-quality temporal and spatial changes.

Our proposed HGCNInf consists of a higher-order graph convolution network (HGCN) and several fully connected neural networks (FNN). By inputting historical monitoring data and by fetching global information, the HGCN is used to capture the spatiotemporal interactions of air-quality distribution and the fully connected neural network is used to extract external influential factor features, so that HGCNInf could accurately infer the air-quality distribution of an entire urban area. As shown in Figures 8 and 9, the prediction accuracy and robustness of HGCNInf are far better than those of each benchmark model. The experimental results show that our proposed model is also applicable even when there are very few label nodes available. In addition, the difference in the two evaluation indicators between the air quality inference model without external influencing factors and the external influencing factors added was compared. The same model after considering the influence of external factors improves the accuracy of urban air-quality distribution prediction and has a significant effect. Therefore, we can conclude that complex external influencing factors are closely related to the distribution of air quality, which also explains the phenomenon why the monitoring sites that are very close geographically in Figure 1 have large differences in values throughout the year.

5.2. Air-Quality-Monitoring Station Location Recommendation

HGCNInf takes the spatiotemporal interactions of pollutant distribution and external influential factors into account, for which the model convolution weight parameters actually represent the correlation degree between the nodes in USTG that reflects the spatiotemporal changes in air quality. Based on the correlation degree, we designed a GMIE that aimed to mark the recommendation priority of unlabeled nodes according to the ability

to improve the inference accuracy of HGCNInf to complete the monitoring-station location recommendation. According to Figures 13 and 14, our proposed model's performance far exceeded the state-of-the-art method, which can bring about the greatest accuracy improvement to HGCNInf for inferring the air quality.

6. Conclusions

In this paper, we formulated the monitoring station location as a urban spatiotemporal graph (USTG) node recommendation problem. We designed an effective air-quality inference model HGCNInf based on a proposed higher-order graph convolution network for inferring the air-quality distribution throughout the city. HGCNInf can not only capture the spatiotemporal interactions of atmospheric pollutant distribution but also extract complex external influential factor features, so that HGCNInf could accurately infer the air-quality distribution of the entire urban area. In addition, HGCNInf can also learn the correlation between the nodes in USTG reflecting the air-quality distribution. Then, we designed a greedy algorithm for minimizing information entropy (GMIE) based on the correlation degree, marking the recommendation priority of unlabeled nodes according to the ability to improve the inference accuracy of HGCNInf to complete the task of recommending the optimal station location. The final recommended optimal station can bring HGCNInf the greatest accuracy improvement. By evaluating real-world Beijing air quality data, HGCNInf and GMIE far outperform the baselines, including the state-of-the-art method.

In the future, we will consider transfer learning methods to improve the applicability of the model to different regions and apply the proposed model to other areas of urban computing, such as traffic flow monitoring, urban regional population flow prediction, etc.

Author Contributions: Conceptualization, Y.K. and J.C.; methodology, Y.K. and J.C.; software, J.C.; validation, Y.C.; formal analysis, J.C. and Z.X.; investigation, J.C.; resources, Y.C. and Z.X.; data curation, J.C.; writing—original draft preparation, Y.K.; writing—review and editing, Z.X.; visualization, J.C.; supervision, Y.C. and Z.X.; project administration, Y.C.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key R&D Program of China under grants 2018YFE0106800, 2018AAA0100800, and 2018YFC0213104; National Natural Science Foundation of China (62033012, 61725304, 61673361, 61773360, and 61872327); Major Special Science and Technology Project of Anhui, China (912198698036, 012223665049); Special Fund for Transformation of Scientific and Technological Achievements of Smart City Research Institute (Wuhu) of University of Science and Technology of China(2019ZX004); as well as the Fundamental Research Funds for the Central Universities under grant (WK2100000013, WK2380000001).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, W.; Shao, L.; Wang, W.; Li, H.; Wang, X.; Li, Y.; Li, W.; Jones, T.; Zhang, D. Air quality improvement in response to intensified control strategies in Beijing during 2013–2019. *Sci. Total. Environ.* **2020**, *744*, 140776. [\[CrossRef\]](#)
2. Shi, K.; Wu, L. Forecasting air quality considering the socio-economic development in Xingtai. *Sustain. Cities Soc.* **2020**, *61*, 102337. [\[CrossRef\]](#)
3. Suman. Air quality indices: A review of methods to interpret air quality status. *Mater. Today Proc.* **2020**, *34*, 863–868. [\[CrossRef\]](#)
4. Xu, Z.; Kang, Y.; Cao, Y.; Li, Z. Deep amended COPERT model for regional vehicle emission prediction. *Sci. China Inf. Sci.* **2021**, *64*, 1–3. [\[CrossRef\]](#)
5. Reames, T.G.; Bravo, M.A. People, place and pollution: Investigating relationships between air quality perceptions, health concerns, exposure, and individual-and area-level characteristics. *Environ. Int.* **2019**, *122*, 244–255. [\[CrossRef\]](#)
6. Yang, X. The relationship between air pollution and human health. *Green Build. Mater.* **2019**, *154*, 53–56.
7. Bai, L.; Shin, S.; Burnett, R.T.; Burnett, R.T.; Kwong, J.C.; Hystad, P.; Donkelaar, A.V.; Goldberg, M.S.; Lacigne, E.; Copes, R.; et al. Exposure to ambient air pollution and the incidence of congestive heart failure and acute myocardial infarction: A population-based study of 5.1 million Canadian adults living in Ontario. *Environ. Int.* **2019**, *132*, 105004. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Zheng, Y.; Liu, F.; Hsieh, H.P. U-air: When urban air quality inference meets big data. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 1436–1444.
9. Qin, Y.; Qian, Y.; Rong, T. Research on the optimization of monitoring site selection based on atmospheric pollutants. *China Environ. Sci.* **2015**, *35*, 1056–1064.

10. Liu, P.; Zheng, J.; Li, Z.; Zhong, L.; Wang, X. Research on Optimal Distribution Method of Regional Air Quality Monitoring Network. *China Environ. Sci.* **2010**, *30*, 907–913.
11. Luo, H.; Feng, H.; Yan, X. Application of air quality prediction model for monitoring site selection. In Proceedings of the 2012 Annual Conference of Chinese Society for Environmental Sciences, Sapporo, Japan, 1–4 August 2012.
12. USEPA. *Fresno Supersite Final Report*; Desert Research Institute for the OAQPS, U.S. Environmental Protection Agency: Research Triangle Park, NC, USA, 2005.
13. Europe Environment Agency. The European Environment. State and Outlook 2005 [EB/OL]. Available online: <http://www.eea.europa.eu/highlights/20051122115248> (accessed on 29 November 2005).
14. Fukushima, H. Air pollution monitoring in East Asia. *Sci. Technol. Trend* **2006**, *18*, 55–63.
15. Raffuse, S.M.; Sullivan, D.C.; McCarthy, M.C. *Ambient Air Monitoring Network Assessment Guidance*; U.S. Environmental Protection Agency: Washington, DC, USA, 2006.
16. Arbeloa, S.F.J.; Caseiras, P.C.; Andres, L.P.M. Air quality monitoring: Optimization of a network around a hypothetical potash plant in open countryside. *Atmos. Environ.* **1993**, *27A*, 729–738. [CrossRef]
17. Kao, J.J.; Hsieh, M.R. Utilizing multiobjective analysis to determine an air quality monitoring network in an industrial district. *Atmos. Environ.* **2006**, *40*, 1092–1103. [CrossRef]
18. Sarigiannis, D.A.; Saisana, M. Multi-objective optimization of air quality monitoring. *Environ. Monit. Assess.* **2008**, *136*, 87–99 [CrossRef] [PubMed]
19. Xu, Z.; Kang, Y.; Cao, Y. Emission stations location selection based on conditional measurement GAN data. *Neurocomputing* **2020**, *388*, 170–180. [CrossRef]
20. Liu, D.; Weng, D.; Li, Y.; Bao, J.; Zheng, Y.; Qu, H.; Wu, Y. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Trans. Vis. Comput. Graph.* **2016**, *23*, 1–10. [CrossRef]
21. Li, Y.; Zheng, Y.; Ji, S.; Wang, W.; U, L.H.; Gong, Z. Location selection for ambulance stations: A data-driven approach. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 3–6 November 2015; pp. 1–4.
22. Kang, Y.; Li, Z.; Zhao, Y.; Qin, J.; Song, W. A novel location strategy for minimizing monitors in vehicle emission remote sensing system. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *48*, 500–510. [CrossRef]
23. Hsieh, H.P.; Lin, S.D.; Zheng, Y. Inferring air quality for station location recommendation based on urban big data. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015; pp. 437–446.
24. Zhu, X.; Ghahramani, Z.; Lafferty, J.D. Semi-supervised learning using gaussian fields and harmonic functions. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 912–919.
25. Vardoulakis, S.; Fisher, B.E.A.; Pericleous, K.; Gonzalez-Flesca, N. Modelling air quality in street canyons: A review. *Atmos. Environ.* **2003**, *37*, 155–182. [CrossRef]
26. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and locally connected networks on graphs. *arXiv* **2013**, arXiv:1312.6203.
27. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In Proceedings of the Advances in Neural Information Processing Systems, Bracelona, Spain, 9–10 December 2016; pp. 3844–3852.
28. Hammond, D.K.; Vandergheynst, P.; Gribonval, R. Wavelets on graphs via spectral graph theory. *Appl. Comput. Harmon. Anal.* **2011**, *30*, 129–150. [CrossRef]
29. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
30. Yu, B.; Yin, H.; Zhu, Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv* **2017**, arXiv:1709.04875.
31. Xu, Z.; Kang, Y.; Cao, Y.; Li, Z. Spatiotemporal Graph Convolution Multifusion Network for Urban Vehicle Emission Prediction. *IEEE Trans. Neural Networks Learn. Syst.* **2021**. [CrossRef]
32. Sun, J.; Zhang, J.; Li, Q.; Yi, X.; Liang, Y.; Zheng, Y. Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks. *arXiv* **2019**, arXiv:1903.07789.
33. Cui, Z.; Henrickson, K.; Ke, R.; Wang, Y. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 4883–4894. [CrossRef]
34. Abu-El-Haija, S.; Perozzi, B.; Al-Rfou, R.; Alemi, A. Watch your step: Learning graph embeddings through attention. *arXiv* **2017**, arXiv:1710.09599.
35. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
36. Perozzi, B.; Al-Rfou, R.; Skiena, S. *DeepWalk: Online Learning of Social Representations*; ACM: New York, NY, USA, 2014; pp. 701–710.
37. Shuman, D.I.; Narang, S.K.; Frossard, P.; Ortega, A.; Vandergheynst, P. The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains. *IEEE Signal Process. Mag.* **2013**, *30*, 83–98. [CrossRef]
38. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
39. Gibson, M.D.; Kundu, S.; Satish, M. Dispersion model evaluation of PM_{2.5}, NO_x and SO₂ from point and major line sources in Nova Scotia, Canada using AERMOD Gaussian plume air dispersion model. *Atmos. Pollut. Res.* **2013**, *4*, 157–167. [CrossRef]

-
40. Rakowska, A.; Wong, K.C.; Townsend, T.; Chan, K.L.; Westerdahl, D.; Ng, S.; Močnik, G.; Drinovec, L.; Ning, Z. Impact of traffic volume and composition on the air quality and pedestrian exposure in urban street canyon. *Atmos. Environ.* **2014**, *98*, 260–270. [\[CrossRef\]](#)
 41. Blum, A.; Mitchell, T. Combining labeled and unlabeled data with co-training. In Proceedings of the 11th Annual Conference on Computational Learning Theory, Madison, WI, USA, 24–26 July 1998; pp. 92–100.
 42. Chen, L.; Cai, Y.; Ding, Y.; Lv, M.; Yuan, C.; Chen, G. Spatially fine-grained urban air quality estimation using ensemble semi-supervised learning and pruning. In Proceedings of the Acm International Joint Conference on Pervasive & Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 1076–1087.