

RESEARCH ARTICLE

Multi-region infectious disease prediction modeling based on spatio-temporal graph neural network and the dynamic model

Xiaoyi Wang^{1,2}, Zhen Jin^{1,2*}

1 Complex Systems Research Center, Shanxi University, Taiyuan, Shanxi, China, **2** Key Laboratory of Complex Systems and Data Science of Ministry of Education, Shanxi University, Taiyuan, Shanxi, China

* jinzhn@263.net



OPEN ACCESS

Citation: Wang X, Jin Z (2025) Multi-region infectious disease prediction modeling based on spatio-temporal graph neural network and the dynamic model. PLoS Comput Biol 21(1): e1012738. <https://doi.org/10.1371/journal.pcbi.1012738>

Editor: Jessica M. Conway, Pennsylvania State University, UNITED STATES OF AMERICA

Received: July 10, 2024

Accepted: December 18, 2024

Published: January 9, 2025

Copyright: © 2025 Wang, Jin. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The code of data process and model construction are available at: github.com/wwwxxxxyyy/M-Graphormer.

Funding: ZJ was funded by Key Programs of the National Natural Science Foundation of China (12231012) and Joint Fund Priority Support Program of the National Natural Science Foundation of China(U23A20331). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Human mobility between different regions is a major factor in large-scale outbreaks of infectious diseases. Deep learning models incorporating infectious disease transmission dynamics for predicting the spread of multi-regional outbreaks due to human mobility have become a hot research topic. In this study, we incorporate the Graph Transformer Neural Network and graph learning mechanisms into a metapopulation SIR model to build a hybrid framework, Metapopulation Graph Transformer Neural Network (M-Graphormer), for high-dimensional parameter estimation and multi-regional epidemic prediction. The framework effectively solves the problem that existing models may lose some hidden spatial dependencies in the data when dealing with the dynamic graph structure of the network due to human mobility. We performed multi-wave infectious disease prediction in multiple regions based on real epidemic data. The results show that the framework is capable of performing high-dimensional parameter estimation and accurately predicting epidemic transmission dynamics in multiple regions even with low data quality. In addition, we retrospectively extrapolate the temporal evolution patterns of contact rate under different interventions implemented in different regions, reflecting the dynamics of intervention intensity and the need for flexibility in adjusting interventions in different regions. To provide early warning of infectious disease transmission, we retrospectively predicted the arrival time of infectious diseases using data from the early stages of outbreaks.

Author summary

In this study, we developed a new method to predict how infectious diseases spread across multiple regions by considering human movement patterns. Our approach combines advanced graph-based deep learning techniques with a classic disease model, creating a powerful framework called M-Graphormer. This framework helps estimate complex disease parameters and predict epidemic trends even when data is scarce or of low quality. By using real-world data from multiple regions, we show that our model can accurately forecast the spread of infectious diseases and adjust for interventions like social distancing or travel restrictions. Additionally, we use our model to understand how different

Competing interests: The authors have declared that no competing interests exist.

intervention strategies impact the disease's progression over time. This research is important because it provides tools for public health authorities to anticipate future outbreaks, make better-informed decisions, and implement timely interventions to control the spread of disease. Our findings have the potential to improve the management of global health crises and offer early warnings of epidemic risks.

Introduction

Over the past three years, the global pandemic of COVID-19 has attracted widespread attention worldwide. With the rapid development of modern transportation, population mobility is accelerating, and disease pathogens are more likely to spread in different regions, so accurately predicting the spread trend of epidemics has become an important task in epidemic prevention and control. A large number of studies have focused on the impact of human mobility on the spread of infectious diseases [1–6], and exploring multi-regional transmission patterns under human mobility can lead to a better understanding of the dynamics of epidemic transmission and the adoption of appropriate control and preventive measures.

Prediction of the spread of infectious diseases in multiple regions can be roughly divided into: infectious disease dynamics methods [7–17], data-driven methods for deep learning [18–24] and deep neural networks combined with infectious disease dynamics methods [25–30].

Infectious disease dynamics aims to establish a mathematical model that can reflect its changing patterns based on the occurrence, development, and environmental changes of the disease. El et al. [7] established a stochastic multi-regional infectious disease model to study the dynamics of infection when infectious diseases occurs in areas connected by any type of anthropological campaign. Fan et al. [8] established a multi-regional SQEIR model to study the impact of global and local interventions on the spread of infectious diseases. In addition, several studies have modeled different multi-regional dynamics based on COVID-19 and evaluated the performance of the models [9, 10]. The metapopulation model is also widely used in multi-regional infectious disease prediction, Feng et al. [11] derived a deterministic non-closed generic model to describe the propagation of diseases on metapopulation networks. Citron et al. [12] compared the dynamics of infectious disease spread in metapopulation models under different population mobility patterns. Das et al. [13] studied the impact of mobility on the spread of COVID-19 by incorporating limited medical resources, isolation, and suppressive behavior of healthy individuals into a metapopulation model. In addition, there are many studies that have established different metapopulation models to study different types of infectious diseases [14–17]. When predictions are made using dynamical models, the epidemiological parameters included in them are usually kept constant [31], whereas in reality they may vary over time, which may lead to the problem of excessive cumulative error when making short-term or long-term predictions. Additionally, the prediction results often rely on model assumptions, which may not fully cover the complexity and uncertainty of the real situation.

Data-driven deep learning is a nonlinear mathematical tool with powerful learning capabilities and is an effective tool for solving many complex problems and tasks. As a typical time series problem, Doni et al. [18] used recurrent neural networks to apply a deep learning approach based on Long Short Term Memory (LSTM) to study the impact of dengue cases in India. Chandra et al. [19] applied LSTM and its three variants for short term COVID-19 prediction in India. In addition, there are many more works for single region infectious disease prediction [20–22]. As mentioned earlier, the spread of infectious diseases usually involves multiple regions, and some studies model it as a spatio-temporal graph structure, using graph

neural networks (GNN) and its variants for prediction. La et al. [23] designed a spatio-temporal graph neural network based on cross-location attention, which combines recurrent neural network (RNN) and GNN to capture spatio-temporal dependencies in data for infectious disease prediction. Kapoor et al. [24] proposed a spatio-temporal graphical neural network that learns the complex dynamics inherent in disease modeling and uses the model to predict COVID-19 daily additions of new cases from fine-grained mobility data. However, these studies have ignored the mechanism of infectious diseases, which may lead to a poor understanding of disease spread and impact, and make the results of the modeling poorly interpretable.

The deep neural network combined with the infectious disease dynamics method aims to make the neural network follow the rules of infectious disease dynamics in the learning process, integrating real-time data and complex infectious disease spreading patterns. Kharazmi et al. [25] embedded the integer-order SIR model, fractional-order SIR model, and delay model into the physical information neural network (PINN), so that the model can identify transient parameters and data-driven fractional difference operators. He et al. [26] embedded a SIR model into PINN to identify intervention intensity during the COVID-19 pandemic. Gao et al. [27] proposed a spatio-temporal graphical attention network for infectious disease prediction, which uses a graphical attention network to capture the spatio-temporal trends of disease dynamics and embedded the SIR model in the loss term to enhance the long term prediction accuracy and interpretability of the results. Wang et al. [28] designed an attention-based dynamic GNN module to capture spatial and temporal disease dynamics and provide epidemiological context for node embedding via a dynamic model. It learns spatio-temporal embedding in the latent space of graph input features and epidemiological context and combines them by using a mutual learning mechanism based on graph-based nonlinear transformations. Cao et al. [29] combined GNN into a metapopulation model to explicitly learn infectious disease parameters and potential infectious disease spread graphs from heterogeneous data end-to-end. Mao et al. [30] combined the hybrid gravity metapopulation model into the spatio-temporal graph attention network to adaptively define interactions between regions and help the model learn the dynamics of infectious disease transmission. Most spatio-temporal graph neural networks use GNN or graph attention networks (GAT) when capturing the spatial dependence of node data. In reality, the graph structure formed by population mobility is constantly changing. GNN aggregates node information by an adjacency matrix or an assignment adjacency matrix, which cannot handle dynamic graph structure and needs to transform the dynamic graph structure into a static graph for processing. GAT aggregates node information by similarity measures between nodes, which does not rely on graph structure at all, although it can deal with dynamic graph problems. This results in the loss of some hidden spatial dependencies in the data when dealing with dynamic maps using both neural networks. Ying et al. [32] proposed a Graph Transformer Neural Network (Graphormer) based on the Transformer [33] architecture, which employs three novel graph structure encoding that can effectively solve the above problems (i.e., it has a great advantage for mining the complex relationships of dynamic graph structures).

The main objective of this study is to solve some of the problems of existing studies as described above and to combine observed data from multiple sources with deep learning and epidemiological modeling so that the neural network follows the rules of infectious disease dynamics during the learning process, integrating real-time data and complex infectious disease transmission patterns. To this end, we incorporate the Graphormer and graph learning mechanisms into the metapopulation SIR model to build a hybrid framework, Metapopulation Graph Transformer Neural Network (M-Graphormer), for high-dimensional parameter estimation and multi-regional epidemic prediction. In addition, we extend the three graph structure encoding included in the Graphormer and use a metapopulation model with migration,

birth, and death terms to enable the hybrid framework to take full advantage of exploiting dynamic graph structures and to be more consistent with the transmission patterns of multi-regional infectious diseases. M-Graphormer is unique in its ability to model complex relationships between regions, its sensitivity to spatial and temporal variations, and its effectiveness in handling dynamic graph structures. It not only explicitly learns high-dimensional epidemiological parameters from heterogeneous data from multiple sources in an end-to-end manner and simultaneously predicts the disease transmission status in multiple regions, but also exhibits excellent performance when data quality is low. In addition, we further retrospectively inferred the temporal evolution patterns of contact rate under different interventions implemented in different regions and predicted infectious disease arrival times.

Materials and methods

This study simultaneously predicts the number of daily new cases in multiple regions. Use $\mathcal{G} = [G^1, G^2, \dots, G^T]$ to represent the dynamic spatial network, where $G^t = (\mathcal{V}, \mathcal{E}^t)$ is the graph at timestamp t with $\mathcal{V}$ denoting the set of N nodes and $\mathcal{E}^t$ being the set of directed weighted edges, respectively. For every pair of connected nodes $i, j \in \mathcal{V}$ at timestamp t , $w_{ij}(t) \in R$ denotes the weight of the directed edge from source node i to target node j at timestamp t . We use

$$\mathcal{A} = [A^1, A^2, \dots, A^T] \in R^{N \times N \times T}, \quad (1)$$

to denote the weighted adjacency matrix sequence. In particular, a weighted adjacency matrix is constructed from mobility data between nodes. To obtain a sparse graph structure, the mobility data threshold is set to 100. Then the edge weight from source node i to target node j is:

$$w_{ij}(t) = \begin{cases} F_{ji}(t), & \text{if } F_{ji}(t) \geq 100, \\ 0, & \text{otherwise,} \end{cases} \quad t = 1, 2, \dots, T. \quad (2)$$

where $F_{ji}(t)$ denotes the mobility data between nodes at timestamp t , i.e., the number of people moving from source node i to target node j . It is worth noting that in this paper we do not use additional edge features and only use mobility data to construct the graph structure. Use $\mathcal{X} = [X^1, X^2, \dots, X^T] \in R^{N \times U \times T}$ to represent the node feature matrix, where X^t is the node feature matrix at timestamp t and U is the number of node features. The input node features include the daily number of new cases, the daily movement changes, *the day of the week*, and daily risk rate. The daily movement change records the change in the range of movement of people compared to the baseline period. The purpose of using the day of the week as an input node feature is to consider different trends and patterns within a week. In some cases, the spread of disease may be affected by different days of the week. For example, people may be more likely to gather or engage in certain behaviors on weekends, which may affect the spread of the disease. Risk rate is an important variable in the transmission process of infectious diseases, indicating the number of new cases in a susceptible population at a given time, and plays a key role in understanding the dynamics of disease transmission within each region. Using this variable as a node feature helps the model to better capture the spatio-temporal dynamics of disease spread. For infectious disease prediction, the objective is to learn a function $f(\cdot)$, which uses the weighted adjacency matrix $A^{t-T_{in}:t}$ and the node feature matrices $X^{t-T_{in}:t}$ of historical T_{in} days as inputs, to predict the number of new cases per day for the next T_{out} days. Where t denotes the timestamp, for example, when t denotes 20 November, $X^{t-T_{in}:t}$ denotes the node feature tensor matrix for 20 November and the T_{in} days before. This problem can be expressed

as follows:

$$[X^{t-T_{in}:t}, A^{t-T_{in}:t}] \xrightarrow{f(\cdot)} Y^{t+1:t+T_{out}}, \quad (3)$$

where $Y^{t+1:t+T_{out}} = [Y^{t+1}, Y^{t+2}, \dots, Y^{t+T_{out}}] \in R^{N \times T_{out}}$ denotes the prediction of daily new cases for all nodes in the next T_{out} days. Y^{t+1} denotes the daily new case prediction for all nodes at timestamp $t + 1$.

In this study, to solve the above problem, we built a hybrid framework, Metapopulation Graph Transformer Neural Network (M-Graphormer), for high-dimensional parameter estimation and prediction of infectious diseases, and its overall framework is shown in Fig 1. It consists of three main components: Spatio-Temporal Layer(ST Layer), Graph and Contact Prediction Layer, and Dynamic model. The specific implementation methods of these three parts will be introduced in detail in the following sections.

A single ST Layer contains only one Spatio-Temporal Extractor(ST Extractor), which consists of Graphormer Layer and Gated Temporal Convolution (Gated TCN) Layer, as shown in Fig 2A. The node features $X^{t-T_{in}:t}$ are first input to the Gated TCN to capture the temporal information of the data, and then passed to the Graphormer to capture the spatial information of the data. For convenience, we define the Gated Temporal Convolution Layer and the Graphormer Layer as the Gated TCN operator and the Graphormer operator, as follows:

$$H_T^l = \text{GatedTCN}(H^l), \quad (4)$$

$$H_{ST}^l = \text{Graphormer}(H_T^l), \quad (5)$$

where H^l is the input to the l^{th} ST Extractor and $H^l = X^{t-T_{in}:t}$ when $l = 0$, H_T^l is the hidden representation containing temporal information, which is used as the input to the Graphormer Layer, and H_{ST}^l is the hidden representation containing temporal and spatial information. By stacking multiple ST Layers (i.e. stacking multiple ST Extractors), the M-Graphormer is able to deal with spatial dependencies at different temporal levels, as shown in Fig 2B. The bottom Graphormer receives short-term temporal information and the top Graphormer receives long-term temporal information. Densely connecting different ST Extractors using a gating system, which can extract important information from the previous ST Extractor to pass on to the next ST Extractor:

$$H_D^l = \begin{cases} X, & \text{if } l = 0, \\ H_D^{l-1} + H^l, & \text{otherwise,} \end{cases} \quad (6)$$

$$H^{l+1} = H_{ST}^l \odot \sigma(H_{ST}^l) + H_D^l \odot (1 - \sigma(H_{ST}^l)), \quad (7)$$

where H_D^l is used to store the information from previous layers and σ denotes the sigmoid activation function. The core idea of gated dense connections is to connect each layer to all previous layers. This connection method allows information to flow more efficiently in the network, thereby alleviating the gradient vanishing problem. The output of each layer not only depends on the input of the current layer, but also utilizes the features of the previous layer, thereby increasing the reuse rate of features. Then, the outputs of different ST Extractors are connected by skip connections, fusing the information of different scales and obtaining the contact and propagation probability matrix $\beta \in R^{N \times T_{out}}$ for all patches on T_{out} days through the output layer, and the contact rate influence factor matrix $C_F = [C_F^1, C_F^2, \dots, C_F^{T_{out}}] \in R^{N \times T_{out}}$, where $C_F^t = [C_{F_1}^t, C_{F_2}^t, \dots, C_{F_n}^t]$ represents the contact rate influence factor of all nodes at

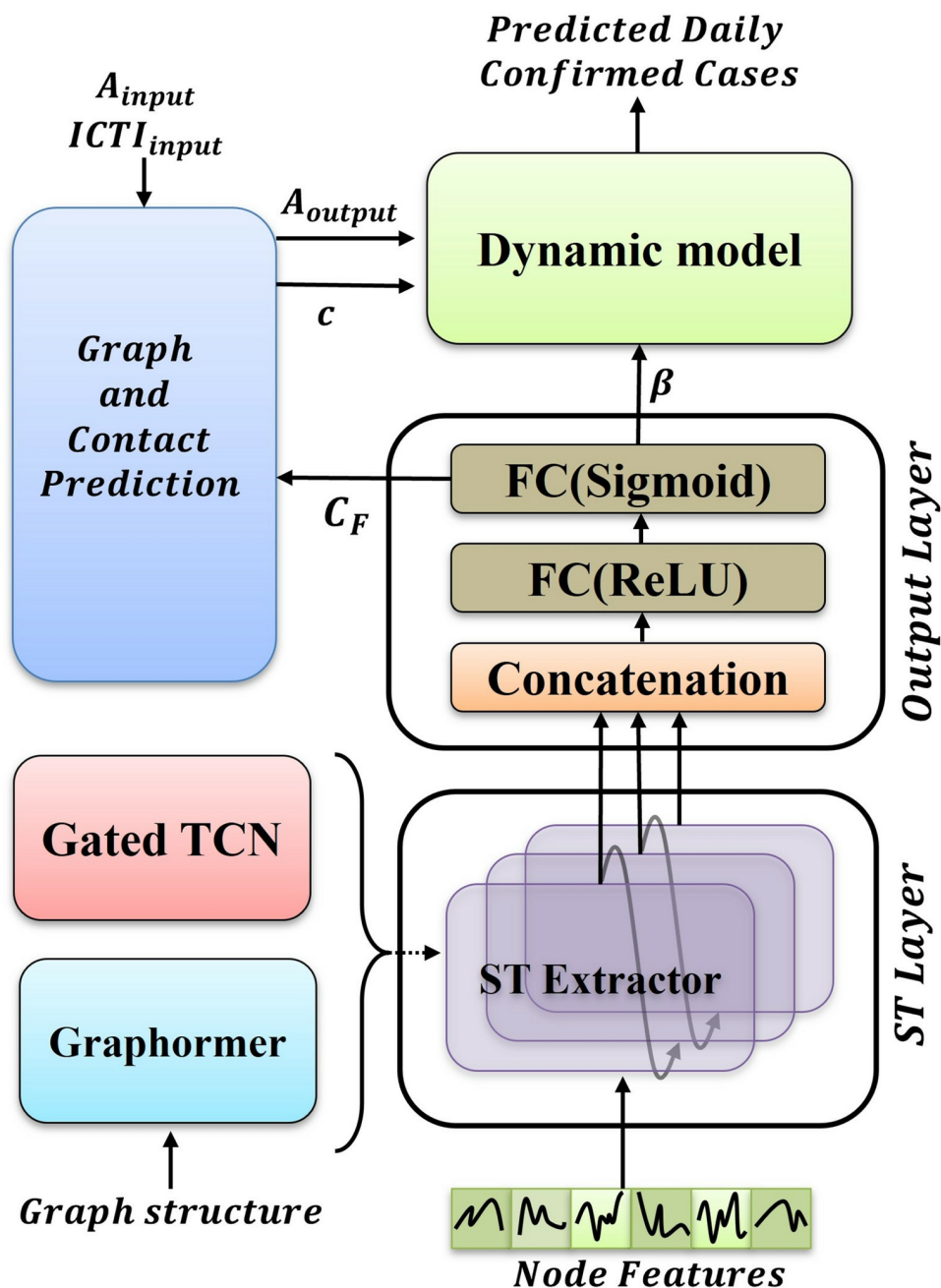


Fig 1. The framework of M-Graphormer. In the output layer, FC represents the fully connected layer, and ReLU and Sigmoid represent the ReLU activation function and Sigmoid activation function respectively.

<https://doi.org/10.1371/journal.pcbi.1012738.g001>

timestamp t . The contact rate influence factor matrix C_F is input into the Graph and Contact Prediction Layer to obtain the contact rate matrix $c \in R^{N \times T_{out}}$, and is input along with β into the dynamic model. Using the predicted β , c and $A^{t+1:t+T_{out}}$, S , I , R , I^{cum} can be iteratively updated by the dynamic model:

$$[S_n^t, I_n^t, R_n^t, I_n^{cum^t}] \rightarrow [S_n^{t+1}, I_n^{t+1}, R_n^{t+1}, I_n^{cum^{t+1}}]. \quad (8)$$

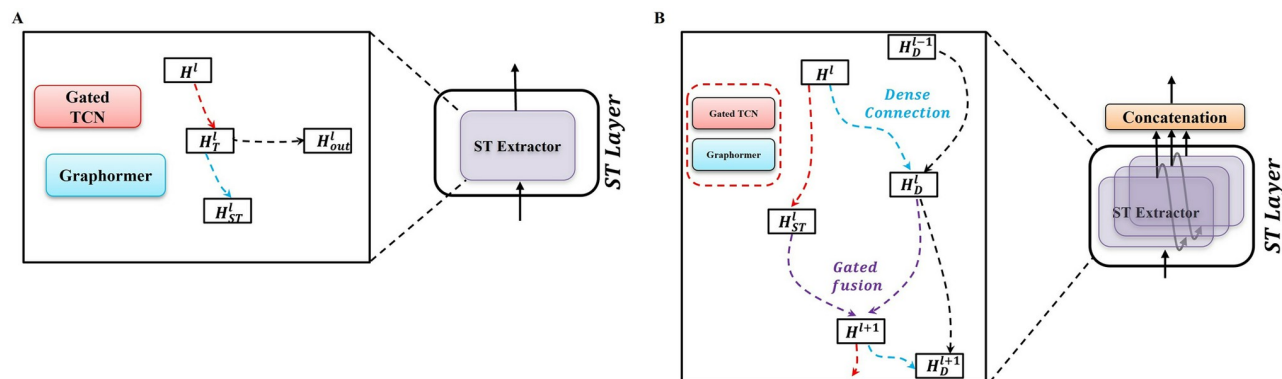


Fig 2. Framework of Spatio-Temporal Layer. A: Framework of Single-layer ST Extractor. B: Framework of Multi-layer ST Extractor.

<https://doi.org/10.1371/journal.pcbi.1012738.g002>

We use the Mean Square Error as the loss function of the M-Graphormer, which is defined as:

$$Loss = \frac{1}{NT_{out}} \sum_{i=1}^N \sum_{j=1}^{T_{out}} (\hat{I}_i^{cum^{t+j}} - I_i^{cum^{t+j}})^2 + \frac{1}{NT_{out}} \sum_{i=1}^N \sum_{j=1}^{T_{out}} (\hat{I}_i^{new^{t+j}} - I_i^{new^{t+j}})^2, \quad (9)$$

where $\hat{I}_i^{cum^{t+j}}$ and $I_i^{cum^{t+j}}$ are the predicted and true cumulative number of confirmed cases at time step $t+j$ for patch i , respectively, and $\hat{I}_i^{new^{t+j}}$ and $I_i^{new^{t+j}}$ are the predicted and true daily number of new cases at time step $t+j$ for patch i , respectively.

The dynamic model

We extend the model of metapopulation infectious disease dynamics in [34] to simulate multi-regional infectious disease transmission. It consists of any n patches (also called sub-population), where individuals within each patch are uniformly mixed and subdivided into three compartments: susceptible, infected, and recovered classes, whose numbers are denoted by S_i , I_i , R_i , where i denotes the i^{th} patch. The total number of individuals is $N(t)$, and the i^{th} patch has a total number of individuals of $N_i(t)$, satisfying

$$N(t) = \sum_{i=1}^n (S_i(t) + I_i(t) + R_i(t)) = \sum_{i=1}^n N_i(t). \quad (10)$$

Fig 3 represents the interactions within the two patches and the mobility between patches. All newborns are susceptible and the birth rate is denoted as B_i . Because the intervention is constantly adjusted, the contact rate and the probability of contact transmission vary over time and are denoted as $c_i(t)$ and $\beta_i(t)$, respectively. Infected individuals leave the infected compartment with a recovery rate constant γ and enter the recovery compartment. Recovered individuals are assumed to be fully immune and will not be reinfected for the length of time considered here. All individuals have natural deaths, and the natural death rate is denoted as μ_i , while infected individuals have deaths due to disease and the death rate due to disease is denoted as λ . Individuals are assumed to be mobile between patches by car, train, or plane. Once an individual from patch i arrives at patch j , that individual mixes evenly with the individuals in patch j and is counted as an individual in patch j . $F_{ij}(t)$ quantifies the number of individuals migrating from patch j to patch i , and the total number of mobile individuals in

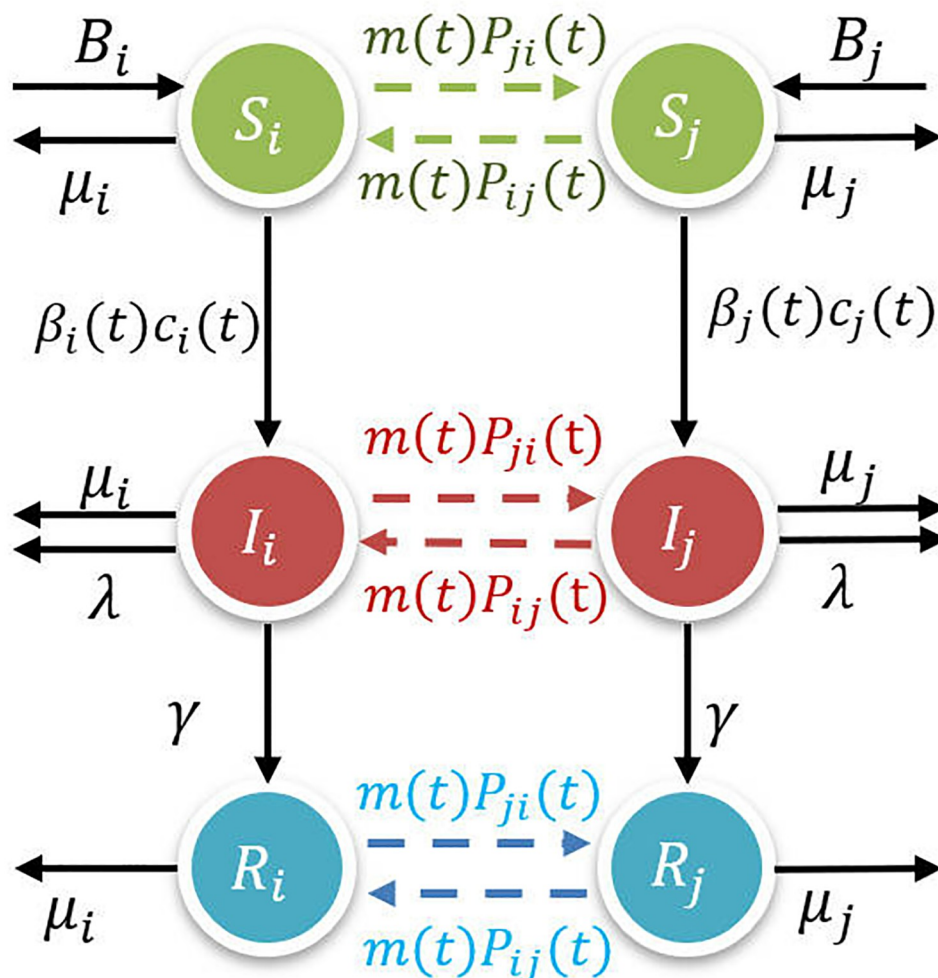


Fig 3. Flow diagram of the metapopulation model (12) showing interactions between patch i and patch j .

<https://doi.org/10.1371/journal.pcbi.1012738.g003>

the system is

$$\Phi(t) = \sum_{i=1}^n \sum_{j=1}^n F_{ij}(t) \quad (i \neq j). \quad (11)$$

Individuals in patch i migrate to patch j at a rate of $m(t)P_{ji}(t)$, where the average migration rate coefficient $m(t) = \frac{\Phi(t)}{N(t)}$, $P_{ji}(t) = \frac{F_{ji}(t)}{\sum_{j \neq i} F_{ji}(t)}$ denotes the proportion of individuals migrating

from patch i to patch j . Therefore, we have the following ordinary differential equation:

$$\begin{cases} \frac{dS_i}{dt} = B_i - \mu_i S_i - \frac{S_i}{N_i} \beta_i(t) c_i(t) I_i + m(t) \sum_{j \neq i} P_{ij}(t) (S_j - S_i), \\ \frac{dI_i}{dt} = \frac{S_i}{N_i} \beta_i(t) c_i(t) I_i - (\mu_i + \lambda + \gamma) I_i + m(t) \sum_{j \neq i} P_{ij}(t) (I_j - I_i), \\ \frac{dR_i}{dt} = \gamma I_i - \mu_i R_i + m(t) \sum_{j \neq i} P_{ij}(t) (R_j - R_i). \end{cases} \quad (12)$$

The special case of the system, i.e., when births and deaths are ignored and all parameters are constants, is modeled as [34]. Here, consider an additional auxiliary compartment to record cumulative cases, and the dynamics of this compartment are driven by the following equations:

$$\frac{dI_i^{cum}}{dt} = \frac{S_i}{N_i} \beta_i(t) c_i(t) I_i. \quad (13)$$

The basic reproduction number R_0 is an important parameter in epidemiology, used to measure the ability of infectious diseases to spread among the population. It defines the average number of individuals that an infected person infects a population of entirely susceptible people without any intervention. The effective reproduction number $R_e(t)$ reflects the spread dynamics of the epidemic and the effectiveness of intervention measures. It indicates how many other people each infected person spreads the virus to given the current intervention measures and stage of the epidemic. The effective reproduction number of patch i is:

$$R_e^i(t) = \frac{\beta_i(t) c_i(t)}{\mu_i + \lambda + \gamma + m(t)} \cdot \frac{S_i(t)}{N_i(t)}. \quad (14)$$

Graphormer Layer

Transformer is a deep learning model architecture designed for processing sequence data. Its core is a self-attention mechanism. Let $H = [h_1^\top, h_2^\top, \dots, h_n^\top]^\top \in R^{n \times d}$ represent the input of the self-attention mechanism, where d represents the hidden dimension and $h_i \in R^{1 \times d}$ is the hidden representation at position i . The input H is passed through three weight matrices $W_Q \in R^{d \times d_K}$, $W_K \in R^{d \times d_K}$ and $W_V \in R^{d \times d_V}$ to obtain the corresponding representation Q, K, V . Self-attention can be calculated using the following formula:

$$Q = HW_Q, K = HW_K, V = HW_V, \quad (15)$$

$$\Psi^{attn} = \frac{QK^\top}{\sqrt{d_K}}, \text{ Attn}(H) = \text{softmax}(\Psi^{attn})V, \quad (16)$$

where Ψ^{attn} is the similarity matrix capturing the similarity between queries and keys, and $\text{softmax}(\cdot)$ represents the *softmax* activation function. In the Transformer model, the attention distribution is calculated through the semantic correlation between nodes. To incorporate the structural information of the graph into the Transformer model, the Graphormer includes three encoding designs: center encoding, spatial encoding, and edge encoding.

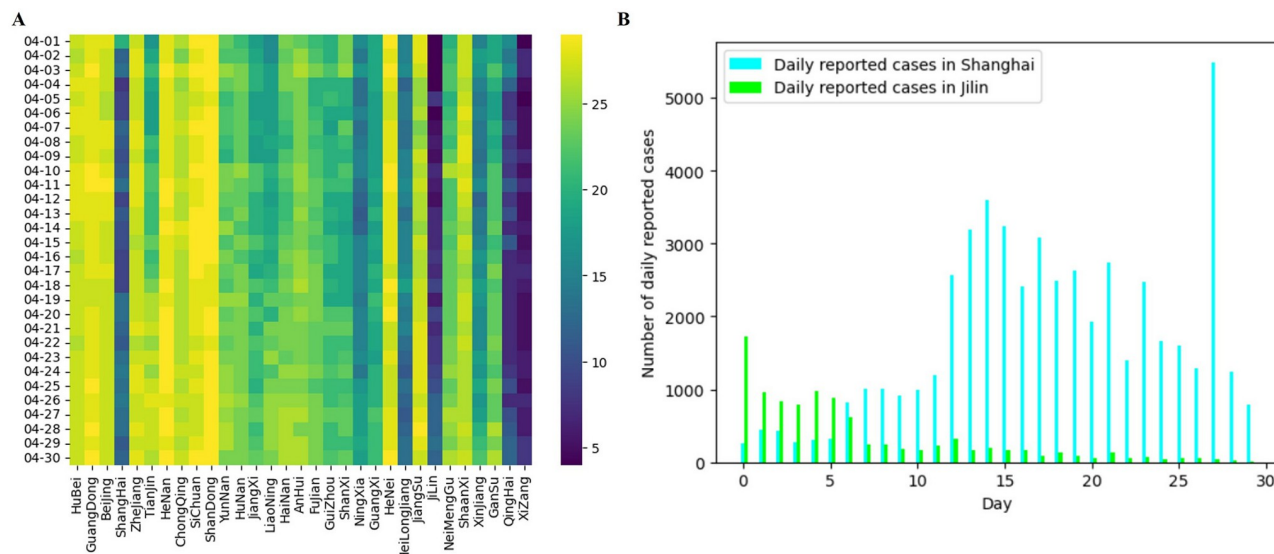


Fig 4. In-degree heat map of Chinese provinces and daily new cases in Jilin and Shanghai in April 2022. A: Heat map of in-degree in April 2022 for each province in China. B: Daily new cases in Shanghai and Jilin, April 2022.

<https://doi.org/10.1371/journal.pcbi.1012738.g004>

Centrality encoding. Node centrality is important for measuring the importance of a node in the graph and is, therefore, valuable for attention calculations. To use it as an additional signal for the general network, the Graphormer adopts degree centrality, one of the standard centrality measures. Specifically, each node is assigned two real-valued embedding vectors based on its in-degree and out-degree, which are summed with node features as input:

$$h_i^{(0)} = X_i + z_{\deg^-(v_i)}^- + z_{\deg^+(v_i)}^+, \quad (17)$$

where $z^-, z^+ \in R^d$ are learnable embedding vectors specified by the in-degree $\deg^-(v_i)$ and out-degree $\deg^+(v_i)$, respectively. Combining centrality encoding and attention mechanisms, the model is able to consider both semantic relevance and node importance. This helps ensure that the model pays more attention to important nodes during the information dissemination process, thereby improving the model's understanding and expression of the graph structure. For example, regarding the spread of COVID-19 in China, the country has adopted a dynamic zeroing policy. When COVID-19 spreads in a city, there will be fewer cities connected to it as a result of the policy intervention; i.e., the in-degrees and out-degrees will be smaller than they would have been in the city without the outbreak. As shown in Fig 4, the change in-degree corresponding to the course of the epidemic can be clearly seen through the daily new cases in Jilin and Shanghai during this period. Therefore, the center encoding is promoted to the following form, in which the in-degree and out-degree of the node are used as the two-dimensional features of the node to contact the original features. This can avoid the loss of some hidden features caused by feature summation in Eq (17):

$$h_i^{(0)} = [X_i || \deg^-(v_i) || \deg^+(v_i)], \quad (18)$$

where $||$ is the concatenation operator.

Spatial encoding. Spatial encoding in the Transformer is used to represent the positional information of labels in a sequence, allowing the model to distinguish between labels at different positions. In graph data, there is no natural order and we are concerned with the

neighborhood information of nodes rather than linear sequences. Therefore, positional encoding and locality are handled differently in graphs than in sequences. In order to quantify the spatial relationship between pairs of nodes in the graph and enhance the model's ability to understand the spatial layout of nodes in the graph, the Graphormer uses a new spatial encoding method. For any graph, the spatial relationship of node pairs in the graph is measured by the function $\phi(v_i, v_j): V \times V \rightarrow R$. Assign each output value a learnable scalar and use it as a bias term affecting the Query-Key product matrix Ψ^{attn} , thus defining the i^{th} row and j^{th} column elements of the matrix Ψ_{ij}^{attn} as:

$$\Psi_{ij}^{attn} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(v_i, v_j)}, \quad (19)$$

where $b_{\phi(v_i, v_j)}$ is a learnable scalar indexed by $\phi(v_i, v_j)$ and is shared among all layers. The spatial encoding is generalized into the following form:

$$\Psi_{ij}^{attn} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b'_{\phi(v_i, v_j)}, \quad (20)$$

$$b'_{\phi(v_i, v_j)} = \frac{1}{1 + e^{-(k\phi(v_i, v_j) - b)}}, \quad (21)$$

where $\phi(v_i, v_j)$ describes the connectivity between nodes in the graph and sets the output value as the shortest path between node v_i and node v_j . If the two nodes are not connected, the output value of ϕ is set to -1 . Both k and b are learnable parameters and are shared among all layers. Our purpose is to set $b'_{\phi(v_i, v_j)}$ to a decreasing function of ϕ , so that the model pays more attention to the nodes near each node and less attention to distant nodes. Regarding the spread of infectious diseases between cities, the closer the distance between two cities, the greater the probability of spread; that is, the higher the correlation between the epidemics in cities [1, 2].

Edge encoding: Edge features are important for graph representation. Encoding edge features into the network along with node features can significantly improve the model's representation and performance. Especially in graphs containing complex relationships, edge features provide additional information that helps in more accurately modeling the correlations between nodes. The Graphormer uses an edge encoding that, for each node pair (v_i, v_j) , finds the shortest path $SP_{ij} = (e_1, e_2, \dots, e_N)$, the average of the dot product of edge features and learnable embedding is computed and introduced as a bias term into the attention module. From this, the element Ψ_{ij}^{attn} of the i^{th} row and j^{th} column of the matrix Ψ^{attn} is further updated to:

$$\Psi_{ij}^{attn} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(v_i, v_j)} + \varepsilon_{ij}, \quad \text{where } \varepsilon_{ij} = \frac{1}{N} \sum_{n=1}^N x_{e_n} (w_n^E)^T, \quad (22)$$

where x_{e_n} is the feature of the n^{th} edge e_n in SP_{ij} , $w_n^E \in R^{d_E}$ is the n^{th} weight embedding, and d_E is the dimensionality of the edge feature. In our model, the edge features are introduced directly into the attention module via bias terms, which leads to a simple generalization of the following form:

$$\Psi_{ij}^{attn} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b'_{\phi(v_i, v_j)} + e(v_i, v_j), \quad (23)$$

where $e(v_i, v_j)$ denotes the edge features pointing from node v_i to node v_j , i.e., it portrays the effect of mobility between nodes on the spread of the epidemic between two nodes.

Temporal Convolution Layer

Dilated causal convolution [35] is commonly used to deal time series data. Unlike traditional convolution, dilated causal convolution has the property of dilation, whereby the receptive field is increased by skipping a certain number of step values instead of scanning them one by one as in normal convolution, and by increasing the depth of the convolution layers, which allows the network to capture a wider range of input information, thus achieving a larger receptive field. “Causal” means that the convolution operation uses only information from before the current time step and not from the future, maintaining temporal causality. Assuming that $\mathbf{x} \in R^T$ is a one-dimensional sequence and $\mathbf{f} \in R^K$ is a convolution kernel, the dilated causal convolution operation between $\mathbf{x}$ and $\mathbf{f}$ at time step t is denoted as:

$$\mathbf{x} \star \mathbf{f}(t) = \sum_{s=0}^{K-1} \mathbf{f}(s) \mathbf{x}(t - o \times s), \quad (24)$$

where o is the dilation factor. By increasing the dilation factor, the sensory field of the model can be increased, and longer sequence information can be captured efficiently without increasing the number of layers. This is useful for dealing with time-series data with long-term dependencies and can improve the performance of the model while maintaining computational efficiency.

Gated TCN. We use Gated TCN [36] as the Temporal Convolution Layer to capture temporal trends in the node data. The introduction of a gating mechanism helps the network learn and control the flow of information in the time series data more efficiently. It takes the following form:

$$H_T^l = g(\theta_1 \star H^l + b_1) \odot \sigma(\theta_2 \star H^l + b_2), \quad (25)$$

where H^l is the input to the l th layer of the Gated TCN and H_T^l is the output of the l th layer of the Gated TCN, θ_1 and θ_2 are the temporal convolution kernels, b_1 and b_2 are the biases, $g(\cdot)$ is the *tanh* activation function applied to the output, $\sigma(\cdot)$ is the *sigmoid* activation function that forms the gate, and $\odot$ is the corresponding elemental product.

Graph and Contact Prediction Layer

We use the same method to predict intra-city travel intensity (ICTI) and dynamic mobility data for the next T_{out} days. Unlike the general methods, the two modules are made to share a learnable graph. By sharing the learnable graph, the predicted data are viewed as a weighted average of the previous T_{in} days. In addition, the parameters of the graph and the contact rate prediction layer can be updated by gradients in the spatio-temporal layer and in the dynamics model, making the learned results more realistic.

As shown in Fig 5, there are two types of input data. The first type is shown in Fig 5A, which uses the historical T_{in} days of intra-city travel intensity $ICTI^{t-T_{in}:t} \in R^{N \times T_{in}}$. Initialize a learnable time weight matrix $L \in R^{T_{out} \times T_{in}}$, and normalize it to $\tilde{L}$ by rows using the *softmax* function. The normalized time weight matrix $\tilde{L}$ can map the historical T_{in} day intra-city travel intensity data $ICTI^{t-T_{in}:t}$ to the future T_{out} day in-town travel intensity $ICTI^{t+1:t+T_{out}} \in R^{N \times T_{out}}$,

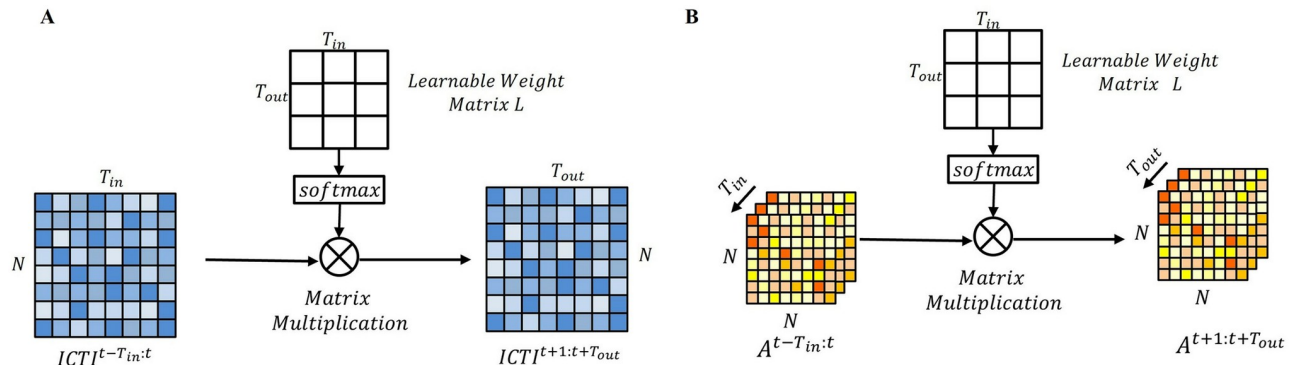


Fig 5. Framework of ICTI and dynamic flow prediction. A: Intra-city travel intensity prediction. B: Dynamic flow (weighted adjacency matrix A) prediction. Here softmax represents the softmax activation function.

<https://doi.org/10.1371/journal.pcbi.1012738.g005>

calculated as follows:

$$\tilde{L} = \underset{j}{\text{softmax}}(L), \quad (26)$$

$$ICTI^{t+1:t+T_{out}} = \tilde{L} ICTI^{t-T_{in}:t}. \quad (27)$$

The second type is shown in Fig 5B, which uses the weighted adjacency matrix (dynamic mobility data) $A^{t-T_{in}:t} \in R^{N \times N \times T_{in}}$ for historical T_{in} days, and uses the same learnable temporal weight matrix and normalized matrix $\tilde{L}$. The normalized temporal weight matrix $\tilde{L}$ can be mapped from the weighted adjacency matrix of the historical T_{in} days to the weighted adjacency matrix of the future T_{out} days $A^{t+1:t+T_{out}} \in R^{N \times N \times T_{out}}$, calculated as follows:

$$A^{t+1:t+T_{out}} = \tilde{L} A^{t-T_{in}:t}. \quad (28)$$

ICTI is a useful indicator of the intensity of human activity and is used as a measure of the effectiveness of interventions. However, ICTI only indicates activity intensity, and more detailed data such as geographic location are not fully utilized to directly infer changes in contact rate. It is often assumed that when people travel less, there is a corresponding reduction in contact rate. However, studies have shown that similar levels of contact are observed at high and low levels of traveling [37, 38]. Inspired by [39], a functional relationship between the intensity of ICTI and the average contact rate was constructed for contact rate prediction:

$$c_i(t) = \ln([PD_i^t \times ICTI_i^t]^{C_{F_i}^t + a}) + b_c, \quad (29)$$

where PD_i^t is the population density of patch i , $C_{F_i}^t \in C_F^t$ is a factor affecting the contact rate of patch i , a is a hyper-parameter and b_c is noise.

Results

Data

We obtained the daily number of new COVID-19 confirmed cases, the daily number of new discharges, and the cumulative number of confirmed cases for a total of 334 days from 1 January 2022 to 30 November 2022, for each province from the provincial health commissions under the National Health Commission of China [40], respectively. Changes in movement

data for each provincial area was generated from the streaming data (activity trajectories) reported by the Beijing Municipal Commission of Health [41] from the 139th confirmed case to the 2,369th confirmed case, which represents the change in the extent of human movement compared to the baseline period. Natural birth rate, natural death rate, resident population and area were obtained from the seventh population census of the National Bureau of Statistics of China [42]. The in-migration (out-migration) size index, in-migration (out-migration) ratio, and intra-city travel intensity of all prefecture-level cities from 1 January 2022 to 30 November 2022 were obtained from Baidu Migration [43], and the number of people who moved in and out of the city per day was inferred and performed based on Baidu's migration size index by the method proposed by [44]. Since Baidu Migration does not provide the intra-provincial travel intensity of each province, we approximate the intra-provincial travel intensity of each province by averaging the intra-city travel intensity of all prefecture-level cities in each province to obtain the intra-provincial travel intensity for each province.

Model calibration

In this study, simultaneous multi-region COVID-19 prediction of the number of new cases per day and simultaneous inference of high-dimensional epidemiological parameters were performed using M-Graphormer. All the data were divided into training, validation and test sets in the ratio of 3:1:1, and all the prediction results were performed in the test set so that the model performance could be better tested. Assuming a recovery rate constant $\gamma = 0.125$, and from [45], the death rate due to disease $\lambda = 0.00008$ is obtained. We show the prediction results for some regions, such as Hubei, Beijing, Shanghai, Hunan, Guizhou, and Shanxi, respectively, in Fig 6, where the daily new prediction results contain two curves for predicting the number of daily new cases in the next 3 days (7 days) using 3 days' (7 days') historical data. Fig 7 is a plot of the cumulative root-mean-square error (RMSE) between the predicted number of new cases per day and the actual reported data for each province/region from 23 September 2022 to 30 November 2022, which is set to 0 because the data used excludes Hong Kong, Macao, and Taiwan Provinces. Fig 8 shows the effective regeneration curves of each provincial region from September 23, 2022 to December 1, 2022.

From Fig 6A–6C, 6G–6I and 7, it can be concluded that the M-Graphormer can obtain predictions for multiple regions simultaneously under conditions of poor data quality and can fit the multi-wave epidemic data well. As can be derived from Fig 6D–6F and 6J–6L, the model can also simultaneously infer contact rate over time under different interventions implemented in different regions. We find that the number of new cases per day and contact rate over time predicted by the model have the same epidemic progression, which can be explained by the fact that China adopts a dynamic zeroing policy, where interventions are stepped up as outbreaks are detected. And the results predicted for specific periods are specifically labeled in Fig 6. For example, in Fig 6B, the outbreak begins in Beijing on 7 November 2022, interventions are intensified, and the corresponding contact rate curve in Fig 6E simultaneously falls off a cliff, consistent with the course of the outbreak in Fig 6B. In Fig 6H, Guizhou has a wave of outbreaks until 30 September 2022 and the outbreak almost disappears between 30 September 2022 and 14 November 2022, corresponding to the contact rate change curve in Fig 6K, where the contact rate is at a very small value on 30 September 2022 and the outbreak almost disappears between 30 September 2022 and 14 November 2022. During this period, the contact rate increases substantially, in line with the course of the epidemic.

All the results show that M-Graphormer can effectively deal with the dynamic graph structure due to human mobility and is able to predict multi-wave outbreaks in multiple regions using heterogeneous data from multiple sources, as well as accurately inferring changes in

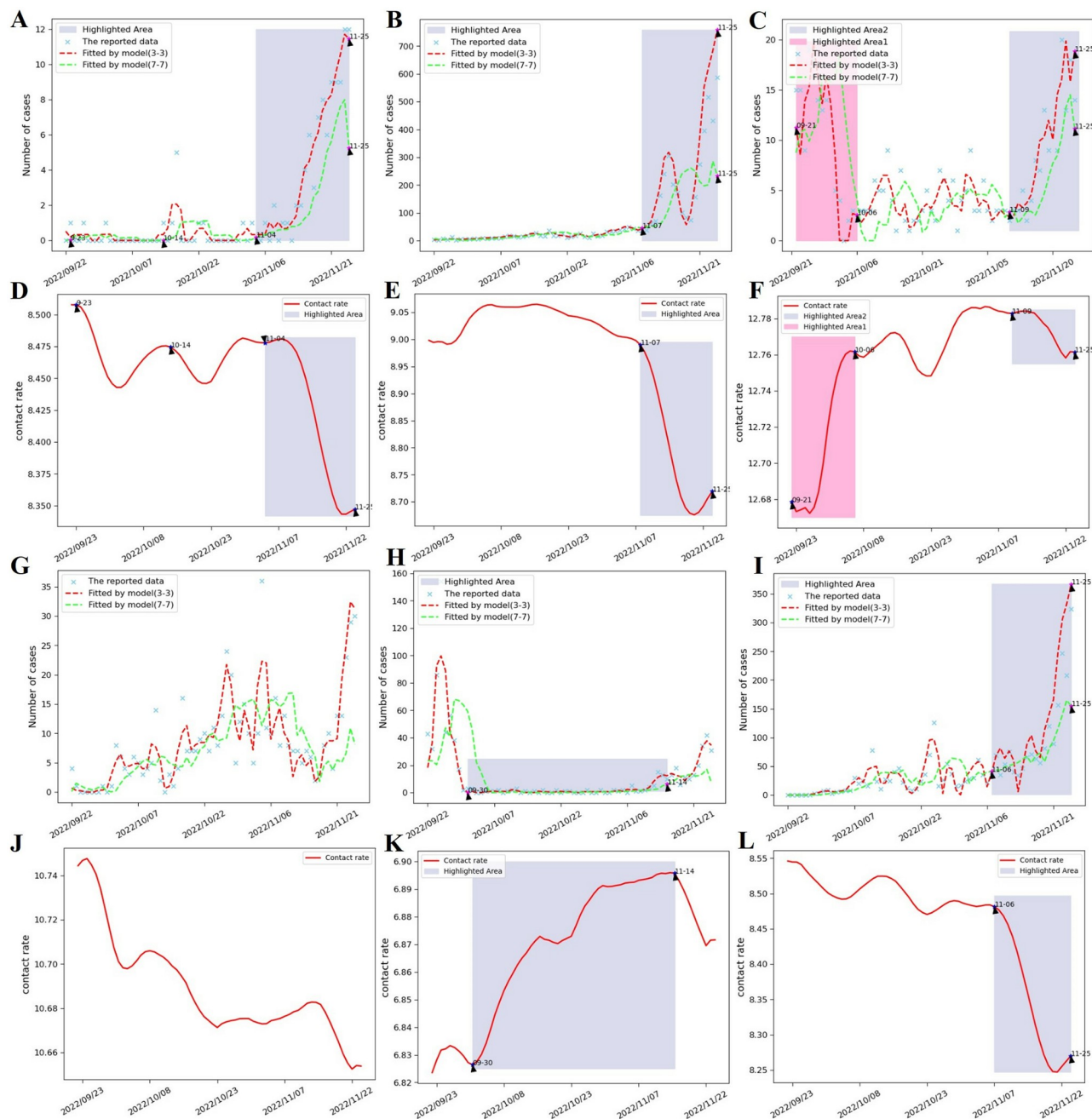


Fig 6. Fitting of daily new case data and time-dependent contact rate inference in Hubei, Beijing, Shanghai, Hunan, Guizhou, and Shanxi provinces based on M-Graphormer. Fig 6A-6C and 6G-6I represent the fitting results of daily new cases in Hubei, Beijing, Shanghai, Hunan, Guizhou, and Shanxi respectively, where the cyan 'x' represents the real reported data, and the red (the green) dotted line indicates that the model inputs 3 days (7 days) of historical data to predict daily new cases in the next 3 days (7 days). Fig 6D-6F and 6J-6L represent the time-dependent contact rate inferences for Hubei, Beijing, Shanghai, Hunan, Guizhou, and Shanxi, respectively. The highlighted areas in the two types of graphs correspond to the specific periods that we have drawn.

<https://doi.org/10.1371/journal.pcbi.1012738.g006>

high-dimensional parameters such as contact rate. The model more accurately restores unobserved epidemic dynamics through limited, sparse and noise-affected data, and completely reconstructs the course of the epidemic. In addition, based on the predicted contact rate and the number of new cases per day, we find that the contact rate shows a decreasing trend in a

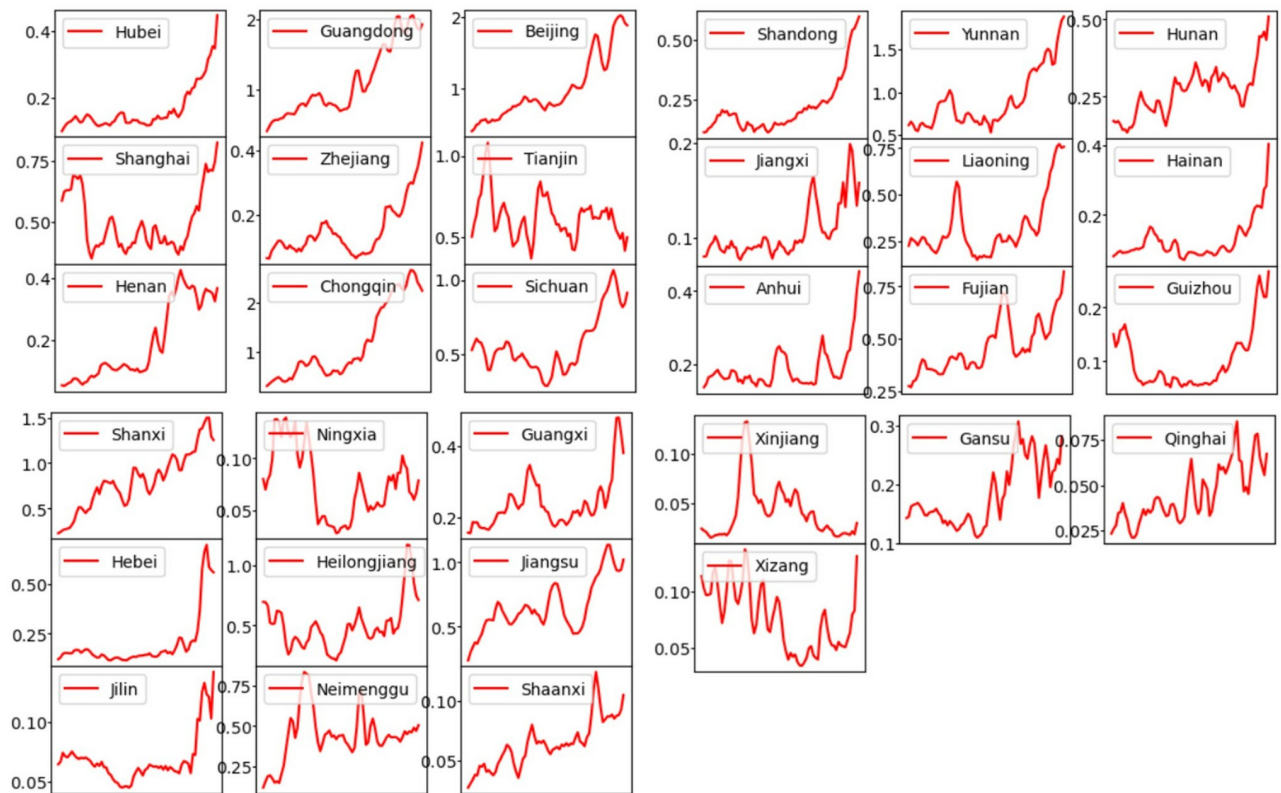


Fig 8. Effective reproduction number curves for each province in China estimated by Eq 14. The x-axis represents the dates, from September 23, 2022 to December 1, 2022.

<https://doi.org/10.1371/journal.pcbi.1012738.g008>

seeds used were 0,1,2,3,4. The results are recorded in Table 1, where the column ‘Prediction modes’ indicates the number of new cases per day for the next 3 days(7 days) using 3 days(7 days) of historical data. From the results, we can make several observations. *Firstly*, the removal of the dynamical model produced a dramatic decrease in the performance of the model, which confirms that the dynamical model is crucial. *Second*, our full model obtained better performance in the experiments, which verifies that all three modules can significantly improve the performance of our model.

Table 1. Ablation study.

Prediction modes	Model	RMSE	MAE
3-3	w/o SIR	11642.0 ± 411.2	4108.9±243.8
	w/o Eq (29)	177.4±41.7	52.7±7.2
	w/o Three encoding	168.4±29.1	52.5±9.3
	M-Graphormer	167.24±22.1	51.7±3.9
7-7	w/o SIR	13082.8 ± 530.4	4854.8±446.7
	w/o Eq (29)	326.9±63.9	93.2±9.2
	w/o Three encoding	354.9±47.1	92.4±12.9
	M-Graphormer	313.8±38.6	92.2±7.7

<https://doi.org/10.1371/journal.pcbi.1012738.t001>

Table 2. Definition of commonly used parameters of the rate function.

Variables	Description
c_{0i} ($i = 1, 2, 3$)	Contact rate at the initial time
c_{bi} ($i = 1, 2, 3$)	Minimum contact rate under the current control strategies
r_{1i} ($i = 1, 2, 3$)	Exponential decreasing rate of contact rate
η	Interference constant

<https://doi.org/10.1371/journal.pcbi.1012738.t002>

Interpretability analysis of time-dependent parameters

In order to further quantify the pattern of temporal evolution of contact rate under the implementation of different interventions in different regions, the three rate functions mentioned in [26] were introduced. The expressions for these three rate functions are given below:

$$\begin{cases} c^1(t) = (c_{01} - c_{b1})e^{-r_{11}t} + c_{b1}, \\ c^2(t) = (c_{02} - c_{b2})e^{-(r_{12}t)^2} + c_{b2}, \\ c^3(t) = c_{b3} \left[1 + \left(\left(\frac{c_{b3}}{c_{03}} \right)^{-\eta} - 1 \right) e^{-r_{13}\eta t} \right]^{\frac{1}{\eta}}. \end{cases} \quad (30)$$

The main parameters in the Eq (30) are defined in Table 2. The contact rate series inferred in M-Graphormer are used as observations, and they are denoted as $\hat{c}(t)$. Since the rate functions in Eq (30) are all decreasing functions (i.e., quantifying increasing interventions), the highlighted area in Fig 6E indicates that the contact rate in Beijing is in a decreasing trend from 7 November 2022 to 17 November 2022, the highlighted area in Fig 6J indicates that the contact rate in Hunan is in a decreasing trend throughout the interval from 21 September 2022 to 25 November 2022, and the highlighted area in Fig 6L indicates that the contact rate in Shanxi was in a decreasing trend from 6 November 2022 to 17 November 2022, so these three regions and time periods were used as observations separately. Use Θ to denote the parameters to be estimated in the three rate functions, and use the least squares method as the method to estimate the parameters. It's equivalent to solving the optimization problem of:

$$\arg \min_{\Theta} L_c(\Theta) = \sum_{t=1}^T |c^i(t, \Theta) - \hat{c}(t)|^2, i = 1, 2, 3. \quad (31)$$

The optimal parameters obtained by least squares for the three regions and periods are shown in Table 3, and the fitting results are presented in Fig 9A–9C. In order to select the most suitable function to describe the extrapolated results of the contact rate for a specific period, the root mean square error of the fitting results of the three rate functions were calculated separately for each region and are shown in Fig 9D–9F. It can be concluded that for Beijing, the function $c^2(t)$ can be chosen to describe the extrapolated results of the contact rate in that period. For Hunan, either $c^1(t)$ or $c^3(t)$ can be chosen to describe the extrapolated results of the contact rate in that period, and for Shanxi, the function $c^2(t)$ can be chosen to describe the extrapolated results of the contact rate in that period. The above results indicate that, in order to cope with the epidemic, the prevention and control models in different regions should be flexibly adjusted according to their specific conditions, and should not be directly copied from other regions. That is, the prevention and control measures should be tailored according to the actual situation in each region.

Table 3. Definition of commonly used parameters of the rate function.

Variables	Beijing	Hunan	Shanxi	Sources
c_{0i} ($i = 1, 2, 3$)	8.8959	10.7278	8.4397	M-Graphormer
c_{b1}	8.5732	10.6550	7.8714	Estimated
c_{b2}	8.6773	10.6698	8.2412	Estimated
c_{b3}	8.5871	10.6555	7.9507	Estimated
r_{11}	0.1183	0.0342	0.0385	Estimated
r_{12}	0.2234	0.0513	0.1666	Estimated
r_{13}	0.0173	0.0035	0.0068	Estimated
η	8	10	8	Assumed

<https://doi.org/10.1371/journal.pcbi.1012738.t003>

Arrival times: Predictions

Predicting the arrival time of infectious diseases is of great importance in the field of public health and epidemic prevention. It can provide early warning and response opportunities, optimize resource allocation, and develop effective prevention and control strategies in advance. This capability significantly enhances the response capacity of the public health system, mitigates the negative impacts of infectious diseases, and effectively protects public health. The time of arrival is defined as: Assuming that the disease originates in one city, how long does it take for it to appear in another city, i.e. how long does it take for the number of people infected by the disease to reach the threshold κ for the first time in a given city.

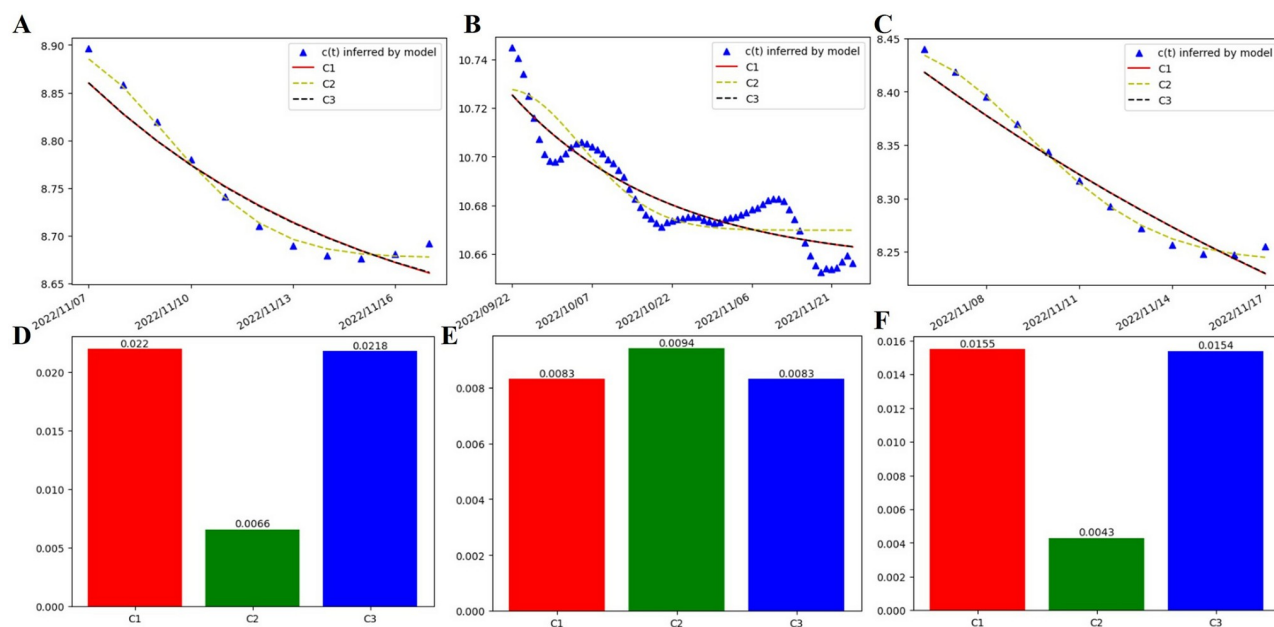


Fig 9. Contact rate inference, rate function fitting and root mean square error in Beijing, Hunan and Shanxi. A–C show the inference and fitting results of time-dependent contact rate in Beijing, Hunan and Shanxi, where the blue triangles represent the M-Graphormer inference results on the contact rate, and the dotted lines and realisations represent the rate function group (30) fitting results. D–F show the root mean square errors corresponding to the fitting results of the rate function group (30) in Beijing, Hunan and Shanxi.

<https://doi.org/10.1371/journal.pcbi.1012738.g009>

Mathematically, the time for the disease to first appear in city i and reach city j is defined as:

$$AT_{ij} = \inf\{t \geq 0 | I_j(t) = \kappa\}. \quad (32)$$

The method proposed by [46] is used to estimate the arrival time of a nonlinear system in terms of a linearized arrival time approximation by linearizing it in the vicinity of the disease-free state. Then the arrival time is estimated as:

$$AT_{ij} = \frac{p_{ij}}{\bar{\beta}c - \bar{\mu} - \lambda - \gamma - m} W \left(\frac{(p_{ij}!)^{\frac{1}{p_{ij}}}}{p_{ij}} \frac{\bar{\beta}c - \bar{\mu} - \lambda - \gamma - m}{m(\rho_{p_{ij}})^{\frac{1}{p_{ij}}}} \left(\frac{\kappa}{\chi_0} \right)^{\frac{1}{p_{ij}}} \right), \quad (33)$$

where $\bar{\beta}c = \frac{1}{n} \sum_k (\beta c)_k$, $\bar{\mu} = \frac{1}{n} \sum_k \mu_k$, p_{ij} denotes the shortest path (minimum number of nodes passed through) from node i to node j . $\rho_{p_{ij}} = \xi_i^T P^{p_{ij}} \xi_j$ describes the probability that an individual starting from node i will be located at node j after p_{ij} steps. χ_0 is the number of infected individuals in node i at the initial moment, ξ_i is the standard Euclidean basis vector and $W(\cdot)$ is the Lambert-W function.

In the Eq (12), some of the parameters are allowed to vary with node, and the method in Eq (33) assumes that the parameters are uniform across all patches. Consider how the non-uniformity of these parameters affects the arrival time, i.e., whether this non-uniformity speeds up or slows down the arrival time compared to the average. For convenience, we call dynamical models with identical parameters homogeneous systems, and dynamical models in which some of the parameters vary with the nodes non-homogeneous systems. In non-uniform systems, Eq (33) is no longer effective in estimating the arrival time, and the alternating derivation in [47] is used to estimate the arrival time of the non-uniform system. In this paper, only node pairs with $p_{ij} = 1$ are used to compare arrival times with the uniform system, assuming that the disease first appears at node i and node j is connected to it. Let

$$\Gamma_i = \bar{\beta}c + \omega_i^{\beta c} - \bar{\mu} + \omega_i^{\mu} - \lambda - \gamma, \quad (34)$$

where $\omega_i^{\beta c} = (\beta c)_i - \bar{\beta}c$ and $\omega_i^{\mu} = \mu_i - \bar{\mu}$, the dynamics of the number of infected at node $n = j$ is approximated by:

$$I_j(t) \approx m P_{ji} \chi_0 e^{\Gamma_j t} \int_0^t e^{-\Gamma_j \tau} e^{\Gamma_i \tau} d\tau = m P_{ji} \chi_0 e^{\Gamma_j t} \left[\frac{e^{(\omega_i - \omega_j)\tau}}{\omega_i - \omega_j} \right]_0^t, \quad (35)$$

where $\omega_i = \omega_i^{\beta c} + \omega_i^{\mu}$. Setting Eq (35) to the threshold κ , two estimates of arrival time depending on the size relation between ω_i and ω_j are obtained:

$$\begin{cases} AT_{ij} \approx -\frac{1}{\Gamma_i} \log \left(\frac{1}{m} \frac{\kappa(\omega_i - \omega_j)}{\chi_0 P_{ji}} \right), & \omega_i > \omega_j, \\ AT_{ij} \approx -\frac{1}{\Gamma_j} \log \left(\frac{1}{m} \frac{\kappa(\omega_j - \omega_i)}{\chi_0 P_{ji}} \right), & \omega_j > \omega_i, \end{cases} \quad (36)$$

when $\omega_i > \omega_j$, the infection growth of node j is dominated by the migration of infections from node i . On the contrary, if $\omega_i < \omega_j$, the local infection growth of node j is dominated by node i , which only needs to propagate a small number of initial infections to node j .

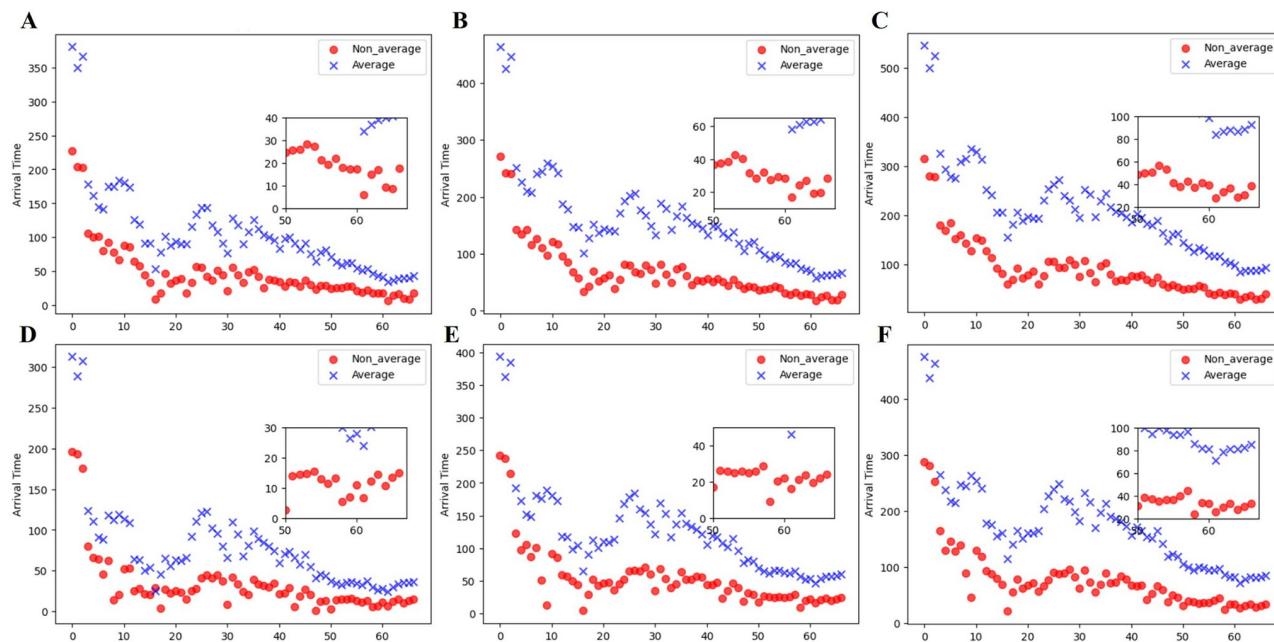


Fig 10. Arrival times of different thresholds κ for the two systems under different parameter groups. The abscissa represents the serial number of different parameter groups. A-C represent the arrival time from Shanxi to Guangdong at thresholds $\kappa = 1, 10, 100$, respectively. D-F represent the arrival time from Shanxi to Beijing at thresholds $\kappa = 1, 10, 100$, respectively. The blue 'x' represents the arrival time of the uniform system, and the red solid circle represents the arrival time of the non-uniform system. The insets in A-F show the minimum arrival time. The x-axis denotes the different parameter group numbers and also the numbering of the date sequence in the test set.

<https://doi.org/10.1371/journal.pcbi.1012738.g010>

We use the parameters inferred by M-Graphormer in the test set to predict arrival times. Since the inferred parameters are time series, all the parameters at each time step are treated as a parameter group. Shanxi is selected as the initial infected province under different parameter groups to predict the arrival time to Guangdong and Beijing, as shown in Fig 10. It shows from left to right the time when Shanxi propagates to Guangdong and Beijing under different parameter sets of the two systems when the threshold $\kappa = 1, 10, 100$, respectively. The results show that as the threshold value keeps increasing, the arrival time also keeps increasing, and the arrival time of the non-uniform system is faster than that of the uniform system, which is consistent with the findings of [47]. We believe that local infection rate are bound to vary due to a variety of factors, so it is reasonable to have considerable variations. For example, for popular provinces such as Beijing and Guangdong, where the population base is large and dense, it is reasonable to expect a rapid spread of the disease, which confirms the findings of [48] that pandemics may occur earlier in large cities than in smaller ones. In addition, the actual arrival times of several cities mentioned in [49] were used as the actual arrival times of the provinces, and various epidemiological parameters and daily additions from 10th January to 20th January 2020 were estimated using the previously trained model. Both Eqs (33) and (36) were used and the arrival times for both systems were successfully estimated as shown in Table 4.

Discussion and conclusion

During the COVID-19 pandemic, China's dynamic zero policy, especially high-frequency large-scale nucleic acid screening and closed management, played an important role in controlling the spread of the epidemic. In order to accurately predict the spreading trend of the epidemic, we retrospectively simulated the course of the COVID-19 epidemic in 31 provincial

Table 4. Prediction of the actual time of arrival for both systems.

Province	Actual arrival time (day)	uniform system	nonuniform system
		Predicted arrival time (day)	Predicted arrival time (day)
Beijing	<23	37	30
Shanghai	<23	39	29
Guangdong	25	29	19
Sichuan	26	33	27
Yunnan	29	41	35
Liaoning	>30	46	28

<https://doi.org/10.1371/journal.pcbi.1012738.t004>

areas of China using the M-Graphormer and inferred high-dimensional epidemiological parameters.

M-Graphormer not only accurately predicts daily new cases and the course of the epidemic, as in Fig 6A–6C and 6G–6I, but also infers high-dimensional epidemiological parameters, as in Fig 6D–6F and 6J–6L. The results show that the use of different interventions in different regions produces different effects, and reveal that predefined functions may not accurately describe the actual contact rate. The highlighted areas in the corresponding contact rate change curves and new cases per day curves for each region in Fig 6 are found to depict the same epidemic progression, i.e., a gradual increase in new cases per day and a gradual decrease in the contact rate, as shown in Fig 6A, 6B, 6D and 6E, etc., which is consistent with the results of the previous studies [50, 51]. We introduced three rate functions from [26] and chose the appropriate function to describe the extrapolated results of the contact rate for Beijing, Hunan and Shanxi for a specific period of time as shown in Fig 9. We predicted the arrival time using two systems of dynamic model, taking Shanxi spreading to Guangdong and Beijing as an example, and used the actual epidemic data to prove the research conclusion in [47] that the arrival time of the non-uniform system is faster than that of the uniform system. At the same time, using the trained model, we successfully predicted the actual arrival time of the epidemic spreading from Wuhan to other provinces at the beginning of the epidemic.

In this study, a hybrid framework, Metapopulation Graph Transformer Neural Network (M-Graphormer), was constructed by incorporating the Graphormer and graph learning mechanisms into the metapopulation SIR model. It enables the neural network to follow the rules of infectious disease dynamics during the learning process, integrating real-time data and complex infectious disease transmission patterns. It also extends the three graph structure encoding included in the Graphormer so that it can be more applicable to epidemiological scenarios. The hybrid framework avoids the loss of certain hidden spatial dependencies in dynamic graph data, and is able to learn high-dimensional epidemiological parameters and simultaneously predict the spreading state of multi-region epidemics in an end-to-end manner using heterogeneous data from multiple sources. At the same time, the model can restore the unobserved epidemic dynamics more accurately and reconstruct the epidemic development process completely with limited, sparse and noise-affected data. In addition, the method can be easily extended to other more complex models of network infectious disease dynamics, allowing a more comprehensive understanding of the impact of different factors on the spread of infectious diseases, which is important for the study of infectious diseases and the development of prevention and control measures.

We believe that the vast majority of infectious diseases spread through human mobility can be effectively applied to this new method. For example, diseases with obvious geographical and temporal transmission characteristics, such as influenza, can be modeled and predicted with

the help of this model. However, it is worth noting that since this method is based on deep learning and is a data-driven model, its performance depends on a large amount of high-quality training data. In order for the model to fully learn effective transmission laws and obtain good prediction results, a sufficient sample size is usually required. If the amount of training data is small, the model may not be able to effectively capture complex flow patterns, resulting in insufficient prediction performance. In addition, due to the particularity of human mobility data, many real-world data (such as population mobility, migration trajectories, etc.) may be difficult to obtain or incomplete, which also brings certain challenges to the application of the model. Therefore, in practical applications, the quality, availability and representativeness of the data will be key factors affecting the effectiveness of the model.

Our study has some limitations. The dynamic model only considered the migration term and did not take into account the effect of short-staying people, such as commuters or travelers, and ignored the effect of measures such as vaccination and quarantine on the spread of infectious diseases. When setting a mobility threshold of 100 for constructing a weighted adjacency matrix to obtain a sparse graph structure, we did not consider the potential impact of different thresholds on the final prediction results, and experiments and evaluations may be needed to determine the mobility threshold that best suits the task requirements. After constructing the sparse graph structure, comparative experiments should be conducted to evaluate the effects of different thresholds on the final prediction results. Because of the lack of fine-grained movement change data, we generated movement change data for all provinces in the country using flow-regulation data from a particular region, which may not accurately reflect the differences and characteristics between provinces due to differences in their level of economic development and cultural background, and may be lacking in the characteristics of the nodes used, such as differences in healthcare resources and changes in GDP in different regions. In predicting the actual arrival times in other provinces at the beginning of the epidemic, using a model trained from the 2022 epidemic data to estimate the epidemiological parameters at the beginning of the epidemic may lead to inaccuracies because the early stages of the epidemic are very different from the later stages of transmission. We leave this for future work.

Author Contributions

Conceptualization: Xiaoyi Wang, Zhen Jin.

Data curation: Xiaoyi Wang, Zhen Jin.

Formal analysis: Xiaoyi Wang, Zhen Jin.

Funding acquisition: Zhen Jin.

Investigation: Xiaoyi Wang, Zhen Jin.

Methodology: Xiaoyi Wang, Zhen Jin.

Project administration: Xiaoyi Wang, Zhen Jin.

Resources: Xiaoyi Wang, Zhen Jin.

Software: Xiaoyi Wang.

Supervision: Zhen Jin.

Validation: Xiaoyi Wang.

Visualization: Xiaoyi Wang, Zhen Jin.

Writing – original draft: Xiaoyi Wang, Zhen Jin.

Writing – review & editing: Xiaoyi Wang, Zhen Jin.

References

1. Tuite AR, Thomas-Bachli A, Acosta H, Bhatia D, Huber C, Petrusek K, et al. Infectious disease implications of large-scale migration of Venezuelan nationals. *J Travel Med.* 2018; 25(1). <https://doi.org/10.1093/jtm/tay077> PMID: 30192972
2. Kraemer MU, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science.* 2020; 368(6490):493–497. <https://doi.org/10.1126/science.abb4218> PMID: 32213647
3. HaËncean MG, Slavinec M, Perc M. The impact of human mobility networks on the global spread of COVID-19. *Journal of Complex Networks.* 2020; 8(6):cnaa041.
4. Xiong C, Hu S, Yang M, Luo W, Zhang L. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proc Natl Acad Sci U S A.* 2020; 117(44):27087–9. <https://doi.org/10.1073/pnas.2010836117> PMID: 33060300
5. Pan Y, Darzi A, Kabiri A, Zhao G, Luo W, Xiong C, et al. Quantifying human mobility behaviour changes during the COVID-19 outbreak in the United States. *Sci Rep.* 2020; 10(1):20742. <https://doi.org/10.1038/s41598-020-77751-2> PMID: 33244071
6. Kraemer MU, Hill V, Ruis C, Dellicour S, Bajaj S, McCrone JT, et al. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B. 1.1. 7 emergence. *Science.* 2021; 373(6557):889–95. <https://doi.org/10.1126/science.abj0113> PMID: 34301854
7. El Kihal F, Abouelkheir I, Rachik M, Elmouki I. Role of Media and Effects of Infodemics and Escapes in the Spatial Spread of Epidemics: A Stochastic Multi-Region Model with Optimal Control Approach. *Mathematics.* 2019; 7(3).
8. Fan J, Du H, Wang Y, He X. The Effect of Local and Global Interventions on Epidemic Spreading. *Int J Environ Res Public Health.* 2021; 18(23). <https://doi.org/10.3390/ijerph182312627> PMID: 34886355
9. Wei Y, Wang J, Song W, Xiu C, Ma L, Pei T. Spread of COVID-19 in China: analysis from a city-based epidemic and mobility model. *Cities.* 2021; 110:103010. <https://doi.org/10.1016/j.cities.2020.103010> PMID: 33162634
10. Oka T, Wei W, Zhu D. The effect of human mobility restrictions on the COVID-19 transmission network in China. *PLoS One.* 2021; 16(7):e0254403. <https://doi.org/10.1371/journal.pone.0254403> PMID: 34280197
11. Feng S, Jin Z. Moment closure of infectious diseases model on heterogeneous metapopulation network. *Adv Differ Equ.* 2018; 2018(1):339. <https://doi.org/10.1186/s13662-018-1801-x> PMID: 32226451
12. Citron DT, Guerra CA, Dolgert AJ, Wu SL, Henry JM, Sanchez CH, et al. Comparing metapopulation dynamics of infectious diseases under different models of human movement. *Proc Natl Acad Sci U S A.* 2021; 118(18). <https://doi.org/10.1073/pnas.2007488118> PMID: 33926962
13. Das T, Bandekar SR, Srivastav AK, Srivastava PK, Ghosh M. Role of immigration and emigration on the spread of COVID-19 in a multipatch environment: a case study of India. *Sci Rep.* 2023; 13(1):10546. <https://doi.org/10.1038/s41598-023-37192-z> PMID: 37385997
14. Meloni S, Perra N, Arenas A, Gomez S, Moreno Y, Vespignani A. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci Rep.* 2011; 1:62. <https://doi.org/10.1038/srep00062> PMID: 22355581
15. Feng S, Jin Z. Infectious diseases spreading on a metapopulation network coupled with its second-neighbor network. *Appl Math Comput.* 2019; 361:87–97. <https://doi.org/10.1016/j.amc.2019.05.005> PMID: 32287503
16. Danon L, Brooks-Pollock E, Bailey M, Keeling M. A spatial model of COVID-19 transmission in England and Wales: early spread, peak timing and the impact of seasonality. *Philos Trans R Soc Lond B Biol Sci.* 2021; 376(1829):20200272. <https://doi.org/10.1098/rstb.2020.0272> PMID: 34053261
17. Iyaniwura SA, Ringa N, Adu PA, Mak S, Janjua NZ, Irvine MA, et al. Understanding the impact of mobility on COVID-19 spread: A hybrid gravity-metapopulation model of COVID-19. *PLoS Comput Biol.* 2023; 19(5):e1011123. <https://doi.org/10.1371/journal.pcbi.1011123> PMID: 37172027
18. Doni AR, Sasipraba T. LSTM-RNN Based Approach for Prediction of Dengue Cases in India. *Ingénierie des Systèmes d'Information.* 2020; 25(3).
19. Chandra R, Jain A, Singh Chauhan D. Deep learning via LSTM models for COVID-19 infection forecasting in India. *PLoS One.* 2022; 17(1):e0262708. <https://doi.org/10.1371/journal.pone.0262708> PMID: 35089976
20. Guo Y, Feng Y, Qu F, Zhang L, Yan B, Lv J. Prediction of hepatitis E using machine learning models. *PLoS One.* 2020; 15(9):e0237750. <https://doi.org/10.1371/journal.pone.0237750> PMID: 32941452

21. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals*. 2020; 135:109864. <https://doi.org/10.1016/j.chaos.2020.109864> PMID: 32390691
22. Arora P, Kumar H, Panigrahi BK. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos Solitons Fractals*. 2020; 139:110017. <https://doi.org/10.1016/j.chaos.2020.110017> PMID: 32572310
23. La Gatta V, Moscato V, Postiglione M, Sperli G. An Epidemiological Neural Network Exploiting Dynamic Graph Structured Data Applied to the COVID-19 Outbreak. *IEEE Trans Big Data*. 2021; 7(1):45–55. <https://doi.org/10.1109/TBDATA.2020.3032755> PMID: 37981990
24. Kapoor A, Ben X, Liu L, Perozzi B, Barnes M, Blais M, et al. Examining covid-19 forecasting using spatio-temporal graph neural networks. *arXiv preprint arXiv:200703113*. 2020.
25. Kharazmi E, Cai M, Zheng X, Zhang Z, Lin G, Karniadakis GE. Identifiability and predictability of integer- and fractional-order epidemiological models using physics-informed neural networks. *Nat Comput Sci*. 2021; 1(11):744–53. <https://doi.org/10.1038/s43588-021-00158-0> PMID: 38217142
26. He M, Tang S, Xiao Y. Combining the dynamic model and deep neural networks to identify the intensity of interventions during COVID-19 pandemic. *PLoS Comput Biol*. 2023; 19(10):e1011535. <https://doi.org/10.1371/journal.pcbi.1011535> PMID: 37851640
27. Gao J, Sharma R, Qian C, Glass LM, Spaeder J, Romberg J, et al. STAN: spatio-temporal attention network for pandemic prediction using real-world evidence. *J Am Med Inform Assoc*. 2021; 28(4):733–43. <https://doi.org/10.1093/jamia/ocaa322> PMID: 33486527
28. Wang L, Adiga A, Chen J, Sadilek A, Venkatramanan S, Marathe M. CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2022; 36(11):12191–9. <https://doi.org/10.1609/aaai.v36i11.21479>
29. Cao Q, Jiang R, Yang C, Fan Z, Song X, Shibasaki R, editors. MepoGNN: Metapopulation Epidemic Forecasting with Graph Neural Networks. *Machine Learning and Knowledge Discovery in Databases: 2023* 2023//; Cham: Springer Nature Switzerland.
30. Mao J, Han Y, Wang B. MPSTAN: Metapopulation-Based Spatio-Temporal Attention Network for Epidemic Forecasting. *Entropy (Basel)*. 2024; 26(4). <https://doi.org/10.3390/e26040278>
31. Tang L, Zhou Y, Wang L, Purkayastha S, Zhang L, He J, et al. A Review of Multi-Compartment Infectious Disease Models. *Int Stat Rev*. 2020; 88(2):462–513. <https://doi.org/10.1111/insr.12402> PMID: 32834402
32. Ying C, Cai T, Luo S, Zheng S, Ke G, He D, et al. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*. 2021; 34:28877–88.
33. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Advances in neural information processing systems*. 2017; 30.
34. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *science*. 2013; 342(6164):1337–42. <https://doi.org/10.1126/science.1245200> PMID: 24337289
35. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:151107122*. 2015.
36. Dauphin YN, Fan A, Auli M, Grangier D. Language Modeling with Gated Convolutional Networks. In: Doina P, Yee Whye T, editors. *Proceedings of the 34th International Conference on Machine Learning; Proceedings of Machine Learning Research: PMLR*; 2017. p. 933–41.
37. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*. 2020; 368(6498):1481–6. <https://doi.org/10.1126/science.abb8001> PMID: 32350060
38. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health*. 2020; 5(5):e261–e70. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6) PMID: 32220655
39. Huang B, Wang J, Cai J, Yao S, Chan PKS, Tam TH, et al. Integrated vaccination and physical distancing interventions to prevent future COVID-19 waves in Chinese cities. *Nat Hum Behav*. 2021; 5(6):695–705. <https://doi.org/10.1038/s41562-021-01063-2> PMID: 33603201
40. Health Commission of the People's Republic of China. [cited 30 Dec 2022]. Available from: <http://www.nhc.gov.cn/>.
41. Beijing Health Commission. [cited 30 Dec 2022]. Available from: <http://wjw.beijing.gov.cn/>
42. National Bureau of Statistics of China. [cited 22 Oct 2023]. Available from: <https://www.stats.gov.cn/>
43. Baidu migration. [cited 22 Oct 2023]. Available from: <https://qianxi.baidu.com/>

44. Wang C, Yan J. An inversion of the constitution of the Baidu migration scale index. *Dianzi Keji Daxue Xuebao/Journal of the University of Electronic Science and Technology of China*. 2021:616–26.
45. Bai W, Cai H, Sha S, Ng CH, Javed A, Mari J, et al. A joint international collaboration to address the inevitable mental health crisis in Ukraine. *Nat Med*. 2022; 28(6):1103–4. <https://doi.org/10.1038/s41591-022-01828-w> PMID: 35562525
46. Chen LM, Holzer M, Shapiro A. Estimating epidemic arrival times using linear spreading theory. *Chaos*. 2018; 28(1):013105. <https://doi.org/10.1063/1.5002009> PMID: 29390617
47. Armbruster A, Holzer M, Roselli N, Underwood L. Epidemic Spreading on Complex Networks as Front Propagation into an Unstable State. *Bull Math Biol*. 2022; 85(1):4. <https://doi.org/10.1007/s11538-022-01110-7> PMID: 36471174
48. Zhang Y, Zhang A, Wang J. Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China. *Transp Policy (Oxf)*. 2020; 94:34–42. <https://doi.org/10.1016/j.tranpol.2020.05.012> PMID: 32501380
49. Hossain MP, Junus A, Zhu X, Jia P, Wen TH, Pfeiffer D, et al. The effects of border control and quarantine measures on the spread of COVID-19. *Epidemics*. 2020; 32:100397. <https://doi.org/10.1016/j.epidem.2020.100397> PMID: 32540727
50. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. 2020; 584(7820):257–61. <https://doi.org/10.1038/s41586-020-2405-7> PMID: 32512579
51. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*. 2020; 585(7825):410–3. <https://doi.org/10.1038/s41586-020-2293-x> PMID: 32365354