Human Mobility Modeling during the COVID-19 Pandemic via Deep Graph Diffusion Infomax

Yang Liu[2], Yu Rong[3], Zhuoning Guo[1], Nuo Chen[1], Tingyang Xu[3], Fugee Tsung[1, 2], Jia Li[1, 2*]

[1]The Hong Kong University of Science and Technology (Guangzhou)
[2]The Hong Kong University of Science and Technology
[3]Tencent AI Lab
yliukj@connect.ust.hk, yu.rong@hotmail.com,{zguo772,nchen022}@connect.hkust-gz.edu.cn,
tingyangxu@tencent.com,{season,jialee}@ust.hk

Abstract

Non-Pharmaceutical Interventions (NPIs), such as social gathering restrictions, have shown effectiveness to slow the transmission of COVID-19 by reducing the contact of people. To support policy-makers, multiple studies have first modeled human mobility via macro indicators (e.g., average daily travel distance) and then studied the effectiveness of NPIs. In this work, we focus on mobility modeling and, from a micro perspective, aim to predict locations that will be visited by COVID-19 cases. Since NPIs generally cause economic and societal loss, such a micro perspective prediction benefits governments when they design and evaluate them. However, in real-world situations, strict privacy data protection regulations result in severe data sparsity problems (i.e., limited case and location information). To address these challenges, we formulate the micro perspective mobility modeling into computing the relevance score between a diffusion and a location, conditional on a geometric graph. we propose a model named *Deep Graph Diffusion Infomax* (DGDI), which jointly models variables including a geometric graph, a set of diffusions and a set of locations. To facilitate the research of COVID-19 prediction, we present two benchmarks that contain geometric graphs and location histories of COVID-19 cases. Extensive experiments on the two benchmarks show that DGDI significantly outperforms other competing methods.

Introduction

The COVID-19 pandemic has been the greatest global public health challenge with over 567 million confirmed cases and over 6.3 million deaths as of 27 July 2022[1]. The outbreak of the COVID-19 not only threatens the public health but also has a devastating impact on economic activity, leading to increased food insecurity, poverty, and socioeconomic inequality (Wang et al. 2020; Laborde et al. 2020; Clouston, Natale, and Link 2021). To control the spread of COVID-19, Non-Pharmaceutical Interventions (NPIs), such as social gathering restrictions, school and business closures, and stay-at-home orders, are implemented via changing people's mobility behaviours (Haug et al. 2020). Therefore, multiple studies (Levin et al. 2021; Hu et al. 2021; Gibbs et al. 2020, 2021) have analyzed the human mobility patterns to support policy-makers. They usually aggregate mobility data, provided by commercial companies, such as SafeGraph [2] and Google [3], and then compute metrics (e.g., daily average fraction of residents staying at home) to model human mobility. Despite their effectiveness, they only focus on a macro aspect of mobility during COVID-19 and ignore the impact of NPIs on individuals. Since NPIs slow the virus spread by limiting human mobility, an in-depth understanding of human behaviours during the COVID-19 pandemic will allow decision makers judiciously and promptly implement interventions.

In this work, we strictly follow the real-world scenario that case diffusions (i.e., location visiting histories) can only be observed when they are tested positive, which abides by the privacy data protection regulation of most countries/regions. Under these settings, we aim to answer the question: *will a location be visited by the COVID-19 cases (e.g., confirmed cases, close contacts, and asymptomatic infections) in the near future?* Addressing the above problem is challenging due to the **data sparsity** in two aspects.

Firstly, due to the privacy policy, the available diffusion data is anonymous, resulting in a limited number of location visiting records of each case. To alleviate such data sparsity, a common solution is to incorporate side information. Thus, we propose to utilize the geometric information, i.e., a geometric graph is constructed where two locations are connected if they are neighbors. To this end, our problem is formulated into computing the relevance score between a diffusion and a location, conditional on a geometric graph. Following the most recent works (Chang et al. 2021b, 2020; Schwabe, Persson, and Feuerriegel 2021), we employ the representation learning framework. That is, we use fixed-length embedding vectors to denote locations, diffusions, and graphs respectively. Figure 1 is an example of the COVID-19 high-risk location prediction, in which a diffusion illustrates a location history of the COVID-19 case, and the geometric graph reflects the distance between locations.

Secondly, most locations have a small number of appear-

*Corresponding author.

[1]who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---27-july-2022

[2]https://www.safegraph.com/academics

[3]https://www.google.com/covid19/mobility/

ances. Generally, in real-world diffusions, common locations (head nodes) can have sufficient appearances while some uncommon locations (tail nodes) can be underrepresented by limited appearances (Anderson 2006). This imbalance poses great challenge to learning unbiased representations, making the learned representations easily dominated by head nodes. To address this limitation, we extend the idea of Mutual Information (MI) maximization (Oord, Li, and Vinyals 2018) to the COVID-19 mobility prediction, which has been proven to be effective in learning high-quality representations with skew distributions (Wang et al. 2021a). Unlike previous graph MI maximization methods (Velickovic et al. 2019) which only have single input (i.e., graph) and output (i.e., location embedding), our formulation involves multiple variables including two inputs (i.e., the geometric graph and diffusions) and two representations (i.e., location and diffusion representations). It is difficult to enforce consistency among these four variables within the framework of MI maximization. In this work, we present **Deep Graph Diffusion Infomax (DGDI)**, to tackle the above challenges. DGDI is defined between two joint distributions: the joint distribution of the graph input and location representations and that of diffusions and diffusion representations. Moreover, our theoretical derivations show that DGDI can be lower bounded by a linear combination of two univariate MI: the univariate MI between the graph input and diffusion representations and the univariate MI between diffusion representations and location representations. Upon this decomposition, the univariate MI can be easily computed via InfoNCE (Oord, Li, and Vinyals 2018) or Deep Graph Infomax (DGI) (Velickovic et al. 2019). We then derive the visit probability by a similarity function of diffusion and location representations.

Social impact Given the existence of immune escape (i.e., vaccines may fail to protect people) (Zhang et al. 2022), especially for some COVID-19 variants (e.g., Omicron), and the specific drug for COVID-19 is still in an early stage, non-pharmaceutical interventions are still practical and effective to combat COVID-19. However, since these interventions may cause substantial economic and societal loss while hurt individuals' mental health and social security, effectively quantifying human mobility during the COVID-19 pandemic to assess NPI effectiveness is necessary to balance costs. Besides, human mobility models can benefit the control of other epidemics as well, such as influenza (Venkatramanan et al. 2021) and ebola (Peak et al. 2018).

Related Work

In this section, we review human mobility and graph contrastive learning papers that are related to our work.

Human Mobility Modeling

Non-pharmaceutical interventions have been one of the most effective tools to defeat COVID-19 transmission and most NPIs aim to change human mobility to reduce the contact rate. Outside the Artificial Intelligence (AI) community, plenty of works (Gibbs et al. 2020; Hu et al. 2021; Levin et al. 2021; Gibbs et al. 2021; Lai et al. 2020; Chang et al. 2021a) have studied the relations between human mobility and NPIs or COVID-19 transmission, which could provide valuable insights for future public health efforts. Most of them focus on computing mobility metrics via aggregating mobility data, e.g., daily average fraction of residents staying at home (Levin et al. 2021), daily average number of trips (Hu et al. 2021) and movement flow matrix (Gibbs et al. 2020). Nevertheless, these macro-view researches are too simple to model real-world mobility, as they treat individuals without differentiation. Note that in the AI domain, there are analogous tasks such as mobility prediction and diffusion prediction which can be utilized to model individual trajectories. Thus in this work, we fill this gap and derive a mobility prediction method for COVID-19 cases. We believe such an approach can help discover more valuable insights during NPI designs.

Existing mobility prediction methods in AI community can be broadly categorized into: Matrix Factorization (MF) (Liu et al. 2013), Markov Chain (MC) (Zhang, Chow, and Li 2014; Chen, Liu, and Yu 2014), and deep learning models which consist of recurrent models (Yao et al. 2017; Sun et al. 2020b; Li et al. 2021), attention mechanism (Yu et al. 2020), and graph neural networks (Lim et al. 2020). Diffusion prediction models follow a similar research line, from early simple linear models (Kempe, Kleinberg, and Tardos 2003; Granovetter 1978; Barbieri, Bonchi, and Manco 2012; Saito et al. 2009; Rong, Zhu, and Cheng 2016; Rong, Cheng, and Mo 2015) to recent deep learning based models (Feng et al. 2018b; Wang et al. 2017a; Sankar et al. 2020; Wang et al. 2021b, 2017b; Wang, Chen, and Li 2018; Islam et al. 2018; Yang et al. 2021) including recurrent neural networks (RNNs) and variational autoencoders (VAEs).

Compared to diffusion prediction methods, mobility prediction models usually have an additional user representation model learned from users' history trajectories (users generally have more than one trajectory) and user features. Nevertheless, in our task, users have no such rich features but a single trajectory due to privacy protection. In this regard, we follow the settings of diffusion prediction which only rely on the location visiting history and a graph to make predictions.

Graph Contrastive Learning

The recent advance of contrastive learning in computer vision motivates the studies of graph contrastive learning (Velickovic et al. 2019; Jing, Park, and Tong 2021; Peng et al. 2020; You et al. 2020; Sun et al. 2020a; Qiu et al. 2020). DGI (Velickovic et al. 2019) maximizes the mutual information between node representations and global summary of the graph. HDMI (Jing, Park, and Tong 2021) leverages two signals to train the model: the mutual information between location embedding and global summary, and the mutual dependence between location embedding and location attributes. HGMI (Li et al. 2022) maximizes the mutual information among three variables. To the best of our knowledge, existing work (Xue and Salim 2021) only explores the advantage of self-supervised learning on COVID-19 Cough Classification, aiming at learning robust representations of respiratory sounds. Our work is the first study that models

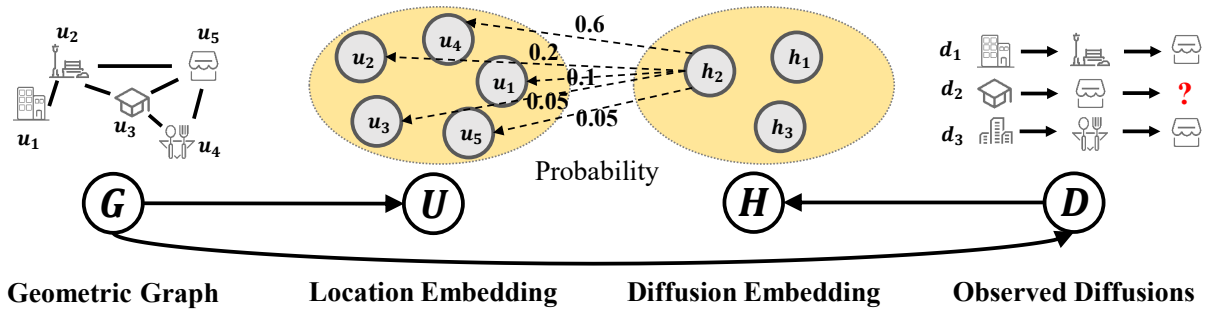

Figure 1: An example of the COVID-19 mobility prediction. The geometric graph records the distance information (two points are connected if they are within a Euclidean distance). The diffusion illustrates a location history of the COVID-19 case.

the mutual information among the graph, nodes and diffusions in COVID-19 diffusion prediction domain.

Preliminaries

Problem Definition

A set of nodes $\boldsymbol{V} = \{v_1, v_2, \ldots, v_N\}$ is used to represent real-world locations or areas in COVID-19 transmission. In this work, locations and nodes are used interchangeably. An $N \times N$ adjacency matrix $\boldsymbol{A}$ is used to describe a **geometric graph** $\boldsymbol{G} = \{\boldsymbol{V}, \boldsymbol{A}\}$ for nodes in $\boldsymbol{V}$. $\boldsymbol{A}_{ij} \in \{0, 1\}$ represents whether there is an edge between nodes v_i and v_j or not, e.g., nodes connected by nearest neighbors in COVID-19 transmission. Let a set of **diffusions** $\boldsymbol{D} = \{d_1, d_2, \ldots, d_M\}$ represent the COVID-19 cases, in which a diffusion d_i is an ordered sequence of location visit histories in ascending order of time denoted by $d_i = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$. The k^{th} visited location of d_i is recorded as v_{i_k}.

Given the pair $\{\boldsymbol{G}, \boldsymbol{D}\}$, our problem is defined as learning a prediction model $f(\boldsymbol{G}, \boldsymbol{D})$ to estimate the probability of visiting an unvisited location v: $p(v|d_i)\ \exists\ v \in \boldsymbol{V} - \{v_i\}_{k=1}^{K}$. Due to a large number of potential unvisited locations, we formulate our prediction as an information retrieval problem. For example, for d_1 in Figure 1, given the observed visited locations (i.e., hotel, park, and mall), the model ranks the probability of visiting unvisited locations (i.e., school, canteen, flat).

Framework

In our problem setting, we have two information sources: a geometric graph $\boldsymbol{G}$ and a set of diffusions $\boldsymbol{D}$. Our purpose is to learn diffusion representations $\boldsymbol{H} = \{h_i\}_{i=1}^{M}$ and location representations $\boldsymbol{U} = \{u_i\}_{i=1}^{N}$, where h_i denotes the derived diffusion representation for d_i and u_i denotes the derived location representation for location v_i. These representations can be retrieved and used to compute a visit likelihood $p(v|d_i)$.

We seek to maximize mutual information for two subsystems: the graph joint distribution $(\boldsymbol{G}, \boldsymbol{U})$ to model the geometric information and the diffusion joint distribution $(\boldsymbol{D}, \boldsymbol{H})$ to model the COVID-19 cases, namely,

$$\arg\max I((\boldsymbol{G}, \boldsymbol{U}); (\boldsymbol{D}, \boldsymbol{H})), \tag{1}$$

Intuitively, this target enforce consistency between the geometric information and the COVID-19 cases, which is named **Deep Graph Diffusion Infomax (DGDI)**.

We consider the geometric graph $\boldsymbol{G}$, the location representations $\boldsymbol{U}$, the diffusions $\boldsymbol{D}$ and the diffusion representations $\boldsymbol{H}$ as four random variables and forming a probabilistic graphical model, as illustrated in Figure 1. As the figure makes clear, the diffusion d_i is assumed to be sampled once from the geometric graph $\boldsymbol{G}$; the location representation u_i directly depends on the graph $\boldsymbol{G}$; the diffusion representation $\boldsymbol{H}$ directly depends on the diffusion $\boldsymbol{D}$. As the joint distribution in Eq. 1 makes the data appearances more sparse, we seek to find a marginal distribution lower bound to relieve the data sparsity issue.

Based on the above assumptions, we have the following decomposition theorem to make the computation of DGDI feasible.

Theorem 1 The graph diffusion mutual information $I((\boldsymbol{G}, \boldsymbol{U}); (\boldsymbol{D}, \boldsymbol{H}))$ can be lower bounded by a sum of univariate mutual information, namely,

$$I((\boldsymbol{G}, \boldsymbol{U}); (\boldsymbol{D}, \boldsymbol{H})) \geq \frac{1}{4} I(\boldsymbol{G}; \boldsymbol{H}) + \frac{1}{4} I(\boldsymbol{H}; \boldsymbol{U}) \tag{2}$$

here $I(\boldsymbol{G}; \boldsymbol{H})$ is the mutual information between graph input and diffusion representations, $I(\boldsymbol{H}; \boldsymbol{U})$ is the mutual information between diffusion representations and location representations.

To prove Theorem 1, we first introduce one lemma.

Lemma 1 For an arbitrary set of three random variables R_1, R_2, R_3, we have

$$I(R_1; R_2, R_3) \geq \frac{1}{2}(I(R_1; R_2) + I(R_1; R_3)) \tag{3}$$

here $I(R_1; R_2, R_3)$ is the mutual information between variable R_1 and the joint distribution of R_2 and R_3 .

To prove Lemma 1, we make use of the chain rule for mutual information.

$$\begin{aligned} I(R_1; R_2, R_3) &= I(R_1; R_3) + I(R_1; R_2|R_3) \\ &\geq I(R_1; R_3) \end{aligned} \tag{4}$$

The last inequality holds as mutual information is non-negative. Accordingly, we have

$$I(R_1; R_2, R_3) \geq I(R_1; R_2) \tag{5}$$

Based on Eq. 4 and Eq. 5, we complete the proof of Lemma 1.

We then prove Theorem 1,

$$\begin{aligned} &I((\boldsymbol{G}, \boldsymbol{U}); (\boldsymbol{D}, \boldsymbol{H})) \geq \frac{1}{2} I((\boldsymbol{G}, \boldsymbol{U}); \boldsymbol{H}) + \frac{1}{2} I((\boldsymbol{G}, \boldsymbol{U}); \boldsymbol{D}) \\ &\geq \frac{1}{4} I(\boldsymbol{U}; \boldsymbol{H}) + \frac{1}{4} I(\boldsymbol{G}; \boldsymbol{H}) + \frac{1}{2} I((\boldsymbol{G}, \boldsymbol{U}); \boldsymbol{D}) \\ &= \frac{1}{4} I(\boldsymbol{U}; \boldsymbol{H}) + \frac{1}{4} I(\boldsymbol{G}; \boldsymbol{H}) + \frac{1}{2} I(\boldsymbol{G}; \boldsymbol{D}) + \frac{1}{2} I(\boldsymbol{U}; \boldsymbol{D}|\boldsymbol{G}) \\ &= \frac{1}{4} I(\boldsymbol{U}; \boldsymbol{H}) + \frac{1}{4} I(\boldsymbol{G}; \boldsymbol{H}) + \frac{1}{2} I(\boldsymbol{G}; \boldsymbol{D}) + 0 \\ &\geq \frac{1}{4} I(\boldsymbol{U}; \boldsymbol{H}) + \frac{1}{4} I(\boldsymbol{G}; \boldsymbol{H}) \end{aligned} \tag{6}$$

The first and second inequalities hold according to Lemma 1; the third equality holds because of the chain rule for mutual information; $I(U; D|G) = 0$ holds as U and D are conditional independent given G (under the proposed graphical model); the last inequality holds as $I(\boldsymbol{G}; \boldsymbol{D})$ is positive and irrelevant with our optimization variables. Thus, we complete the proof of Theorem 1.

Model

In this section, we first introduce our method to compute the location, diffusion and graph representations, then describe how to maximize DGDI in details.

Representation Model

The goal of our representation model is to produce fixed-length embedding vectors of locations, diffusions and graph. We first obtain location embeddings $\boldsymbol{U}$ from random initializing and then propose a simple yet effective way to compute graph and diffusion representations.

Diffusion representation To capture the graph characteristics and the diffusion temporal influence, the diffusion representation part takes the geometric graph G and a set of diffusions D as inputs. Accordingly, our model consists of two major components: (1) a Graph Neural Network (GNN) that smooths each location representation according to the graph topology; and (2) a self-attention layer that quantifies the varying effect of previous visited locations. Our objective is to encode the graph information into location embeddings in GNN and then feed them into the temporal self-attention layer.

We use GNN model to encourage locations that are close in the geometric graph G to share similar latent representations, which may benefit the predicted task in the absence of explicit location attributes. For example, in the COVID-19 spread, locations that are close in geometric space may be visited by the same cases. We empirically test Graph Convolutional Network (GCN) (Kipf and Welling 2017), Graph Attention Network (GAT) (Velickovic et al. 2018), and Graph Isomorphism Network (GIN), and find that GCN performs best. Its update rule is:

$$\boldsymbol{Z}^{(l)} = \text{ReLU}(\hat{\boldsymbol{A}} \boldsymbol{Z}^{(l-1)} \boldsymbol{W}^{(l-1)}), \tag{7}$$

where $\hat{\boldsymbol{A}}$ is a normalized adjacency matrix and $\boldsymbol{Z}^{(l)}$ ($\boldsymbol{Z}^{(0)} = \boldsymbol{U}$) denotes the output representations of l-th layer GCN.

The latent embedding $\boldsymbol{Z}$ is then utilized to compute the diffusion embedding. In the COVID-19 predictions, one observation is that the true spread processes may not strictly follow the sequential assumption, e.g., a location can be visited depending on any of the previous visited locations. To capture such dependency, a self-attention layer followed by a Multi-Layer Perceptron (MLP) is adopted. Specifically, given a diffusion $d_i = \{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\}$, the k-th location is represented by z'_k (we omit the subscript i for simplicity):

$$z'_k = z_k + \text{PE}(k) \tag{8}$$

where $\text{PE}(k)$ is the positional-encoding (Vaswani et al. 2017) that only depends on the position k. The even and odd elements are $\sin(k/10000^{i/L})$ and $\cos(k/10000^{i-1/L})$ respectively (L is the dimension of encoding). Then the diffusion representation h_i for the diffusion d_i can be computed by:

$$h_i = \text{MLP}\left(\sum_{j=1}^{k} \alpha_{jk} f_v(z'_j)\right), \tag{9}$$

$$\alpha_{jk} = \frac{\exp(< f_q(z'_k), f_k(z'_j) >)}{\sum_{j=1}^{k} \exp(< f_q(z'_k), f_k(z'_j) >)}, \tag{10}$$

where $< \cdot, \cdot >$ denotes the inner product of two vectors. z'_k denotes the last location of d_i and attention weight $\alpha_{jk} \in \mathbb{R}$ denotes the j-th node's contribution to the final diffusion representation. Following the previous work (Vaswani et al. 2017), we apply three linear transform functions f_k, f_q, and f_v (i.e., $f(z) = z\boldsymbol{W}$) on latent embeddings.

Graph representation The representation of the entire graph can be obtained by aggregating the output of GNN, which is termed as graph pooling (Ying et al. 2018). The pooling function transforms arbitrary-sized location embeddings to a fixed-length representation, which can be a simple mean or a more sophisticated graph-level pooling function such as Attention Pooling (Li et al. 2019; Lee, Lee, and Kang 2019). Here we use mean pooling for simplicity:

$$g = \frac{1}{N} \sum_{i=1}^{N} z_i, \tag{11}$$

where g denotes the graph representation.

Maximization of DGDI

Training objective According to Theorem 1, DGDI is lower bounded by two univariate mutual information: mu-

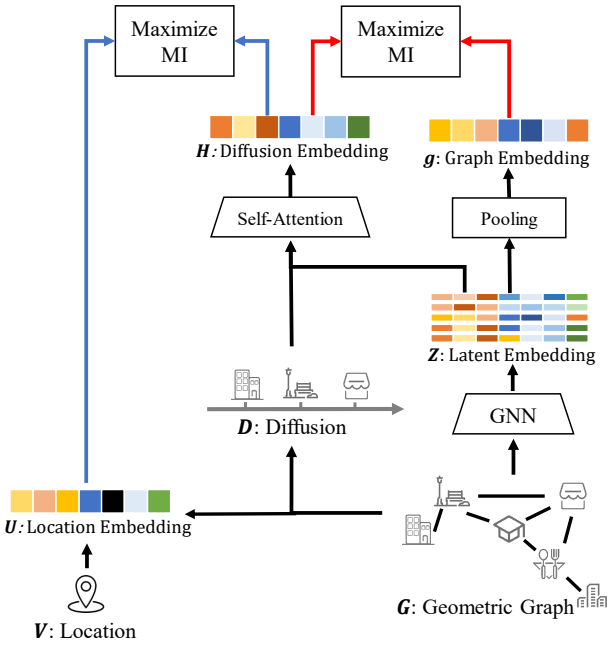

Figure 2: Overview of the proposed DGDI. The graph embedding is obtained via pooling the latent embeddings, and the diffusion embedding is calculated via self-attention with the input of location embeddings. Finally, we compute the MI between the location embedding and diffusion embedding, and the MI between the diffusion and graph embedding.

tual information between graph input and diffusion representations and that between diffusion representations and location representations. Accordingly, our training objective becomes:

$$\mathcal{L} = \lambda_1 \mathcal{L}_G + \lambda_2 \mathcal{L}_U \tag{12}$$

Following existing work (Peng et al. 2020), λ_1 and λ_2 are set to tunable parameters for better performance. $\mathcal{L}_G$ and $\mathcal{L}_U$ are the training losses for $I(\boldsymbol{G};\boldsymbol{H})$ and $I(\boldsymbol{H};\boldsymbol{U})$ maximization respectively. Figure 2 illustrates the overall framework of DGDI. During our experiments, we found that deep graph infomax (DGI) (Velickovic et al. 2019) and InfoNCE (Oord, Li, and Vinyals 2018) are simple but effective methods to estimate $I(\boldsymbol{G};\boldsymbol{H})$ and $I(\boldsymbol{H};\boldsymbol{U})$. They share a similar idea which encourages the consistency between the representations of positive pairs and the divergence between that of negative pairs. Although there exist other estimators (e.g., GMI (Peng et al. 2020) or InfoGraph (Sun et al. 2020a)), we empirically show that DGI and InfoNCE have already achieve state-of-the-art performance in Section . Thus, we leave the testing of other choices for future work.

$I(\boldsymbol{G};\boldsymbol{H})$ computation MI between global graph embedding g and representations of local parts are maximized under the framework of DGI (Velickovic et al. 2019). Since diffusions are assumed to be sampled from the graph, we treat diffusions as the local parts and maximize MI between global graph embedding g and diffusion embedding h. A negative sampling strategy is then leveraged to generate negative graph $\hat{g}$. More specifically, we randomly shuffle the location embedding matrix and get $\hat{\boldsymbol{U}}$; we let $\hat{\boldsymbol{U}}$ go through the GNN module and the graph representation module and get $\hat{g}$. A discriminator is then utilized to distinguish the diffusion representation and graph representations. Formally, the loss is defined as follows:

$$\mathcal{L}_G = \sum_{h \in \boldsymbol{H}} \mathbb{E}[\log f_d(h,g)] + \mathbb{E}[\log(1 - f_d(h,\hat{g}))], \tag{13}$$

where f_d denotes the inner-product discriminator:

$$f_d(h,g) = \sigma(< h, g >), \tag{14}$$

where σ denotes the sigmoid activation function and $f_d(h,\hat{g})$ is calculated in the same way.

$I(\boldsymbol{H};\boldsymbol{U})$ computation $I(\boldsymbol{H};\boldsymbol{U})$ is maximized between diffusion representation H and location representation U. We use InfoNCE (Oord, Li, and Vinyals 2018) to maximize the MI of these two representations. In terms of sampling strategies, a diffusion and its next visited location are considered as a positive pair while a diffusion and uninfected locations are as negative pairs. To summarize, $\mathcal{L}_U$ is computed by:

$$\mathcal{L}_U = -\sum_{i=1}^{M}\sum_{k=1}^{K} \log \frac{\exp(< h_{i_k}, u > /\tau)}{\sum_{u' \in \boldsymbol{U}} \exp(< h_{i_k}, u' > /\tau)}, \tag{15}$$

where h_{i_k} denotes the diffusion representation of d_i's sub-diffusion ending at the position k and τ is the hyper-parameter, known as the temperature. Inner-product is employed as the similarity function of InfoNCE.

For model inference, we formulate the predicted task as an information retrieval problem. Given an input diffusion, our model ranks a location list according to the possibility of being visited. And the possibility equals to the inner product of the diffusion representation and the location representations.

Time Complexity

The cost per training iteration of DGDI contains the computation of GCN (i.e., Eq. 7) and the self-attention layer (i.e., Eq. 9 and Eq. 10). The graph convolution operations take $O(|\boldsymbol{E}|F^2)$ (Kipf and Welling 2017) where $|\boldsymbol{E}|$ denotes the number of edges and F is dimension of latent embeddings. The time complexity of the self-attention layer is $O(|\boldsymbol{D}|K^2F)$ where $|\boldsymbol{D}|$ denote the size of observed cascades and K is the length of the longest cascade. Thus, the overall complexity of DGDI is $O(|\boldsymbol{E}|F^2 + |\boldsymbol{D}|K^2F)$. We compare DGDI with Inf-VAE (Sankar et al. 2020) and one of our baselines SNIDSA (Wang, Chen, and Li 2018). The core operations of Inf-VAE are GCN with inner-product and a co-attention layer which take $O(|\boldsymbol{E}|F^2 + |\boldsymbol{V}|^2F + |\boldsymbol{D}|K^2F)$. SNIDSA is a recurrent model with a structural attention network whose complexity is $O(|\boldsymbol{D}|KF^2 + |\boldsymbol{V}|F^2 + |\boldsymbol{E}|F)$. Since Inf-VAE requires the computation of each node pair, its time complexity is highest among these three models.

Dataset	# Nodes	# Links	# Diffusions
COVID-HK	3,274	603,667	2,536
COVID-MLC	4,091	58,079	1,887

Table 1: Dataset statistics.

Experiments

Experimental Settings

Datasets Two public available datasets are employed to evaluate our proposed model in COVID-19 transmission scenarios: COVID-HK and COVID-MLC. To better serve our problem, imported cases are excluded and we retain cases who have at least 2 locations. In terms of the geometric graph, two locations are connected if their distance is less than 3 kilometers. **Note that all data are published by the government and do not contain sensitive information (i.e., anonymous).**

- **COVID-HK.** The COVID-HK dataset is released by the Hong Kong government [4], which records locations visited by the COVID-19 cases from Jan 28, 2020 to Sep 1, 2021.
- **COVID-MLC.** The COVID-MLC dataset is provided by Beijing Advanced Innovation Center for Big Data and Brain Computing[5], which records locations from Jan 1, 2020 to March 22, 2020 and involves twelve provinces of Mainland, China.

Baselines We compared our proposed model with several methods. The first group focuses on diffusion modeling and contains FMC (Zhang, Chow, and Li 2014), LSTM (Hochreiter and Schmidhuber 1997), GRU (Cho et al. 2014) and DeepMove (Feng et al. 2018a). The second group focuses on geometric graph modeling and includes GCN (Kipf and Welling 2017) and GIN (Xu et al. 2019). It makes use of geometric graph information and represents diffusions by their last location representations. The third group uses both geometric graph and diffusion information and includes Topo-LSTM (Wang et al. 2017a), SNIDSA (Wang, Chen, and Li 2018), FOREST (Yang et al. 2019) and Inf-VAE (Sankar et al. 2020). These methods are the state-of-the-art in diffusion prediction tasks.

Implementation details For Topo-LSTM[6], DeepMove[7], SNIDSA[8] and Inf-VAE[9], we use the code provided by authors. We implement FOREST and our model by PyTorch.

Evaluation metrics According to the timestamp, the dataset is split to 70%, 10% and 20% for training, validation and testing. As stated in Section , we consider our prediction task as a ranking task. Specifically, the unvisited locations are ranked based on their predicted visit probabilities. Two widely used ranking metrics are adopted: Recall@K which denotes the fraction of infected locations among the top-k predicted locations, and MAP@K which jointly measures the existence and position of the target location in the rank list.

Setup For fair comparison, all models are trained by Adam optimizer (Kingma and Ba 2015) with a learning rate of $\{0.001, 0.0005, 0.0001\}$ and mini-batch size of 16. The dimension for representations and hidden states in all models are 32. For DGDI, we set λ_2 to 1 and tune λ_1 and τ within the ranges of $\{0.1, 0.2, ..., 0.5\}$ and $\{0.5, 0.6, ..., 1\}$. We use GCN as our graph model and its layer number is searched within 2. For FOREST, we use recommended settings in their papers (Yang et al. 2019). We run each method 5 times and report its average accuracy.

Overall Performance Comparison

Table 2 lists the performance for all methods where K is set to 3, 5, and 10. We can observe:

- Sequence Models that only utilize diffusions (i.e., FMC, LSTM, and DeepMove) perform worst in most cases, indicating that simply modeling the diffusion information is not sufficient. Although graph models (i.e., GCN and GIN) ignore previous visited locations to make predictions, they outperforms LSTM and DeepMove under all circumstances. One possibility is that most people only have a limited circle of activities (e.g., near their home or workplace) . Thus, integrating the geometric information allow the model to provide a better predictions.
- The integration of sequence models and the graph structure results in performance improvements. Topo-LSTM, SNIDSA, FOREST, Inf-VAE, and DGDI beat models based on pure sequence (i.e., FMC, LSTM, and DeepMove) in most cases, indicating the importance of jointly modeling the diffusion and graph information.
- DGDI significantly outperforms all the baselines on both MAP and Recall, which indicates that more to-be-visited locations are found and their ranks are higher in the predicted list. For example, our model beats the second best model Inf-VAE by 1.18 w.r.t. MAP@3 on COVID-HK and 0.93 w.r.t. MAP@10 on COVID-MLC. These improvements are owed to both the model architecture and the contrastive loss.

Comparison w.r.t. data sparsity To investigate the model performance w.r.t. data sparsity of location appearance, we divide the frequency of location appearance into three groups which are $\{< 7(547), 7-13(30), > 13(10)\}$ on COVID-HK and $\{< 8(912), 8-17(44), > 17(17)\}$ on COVID-MLC, e.g., $< 7(547)$ indicates that there are 547 locations appears less than 7 times in training diffusions. We can see that most locations only have limited appearances and only a small number of locations appear frequently. The recall@10 of DGDI and the two strongest baselines (i.e., Inf-VAE and FOREST) are displayed in Figure 3. Specifically, the results of each group indicate how much they contribute to the overall results of the entire test set (i.e., the **sum** of

[4]https://chp-dashboard.geodata.gov.hk/covid-19/en.html

[5]https://github.com/BDBC-KG-NLP/COVID-19-tracker

[6]https://github.com/vwz/topolstm

[7]https://github.com/vonfeng/DeepMove

[8]https://github.com/zhitao-wang/Sequential-Neural-Information-Diffusion-Model-with-Structure-Attention

[9]https://github.com/aravindsankar28/Inf-VAE

Metric (%)	Recall						MAP					
Model	Covid-HK			Covid-MLC			Covid-HK			Covid-MLC		
	@3	@5	@10	@3	@5	@10	@3	@5	@10	@3	@5	@10
FMC	1.10±0.0	1.10±0.0	1.10±0.0	4.40±0.0	4.63±0.0	5.39±0.0	0.79±0.0	0.79±0.0	0.79±0.0	3.78±0.0	3.83±0.0	3.93±0.0
LSTM	0.73±0.2	0.88±0.3	1.18±0.2	2.73±0.2	3.24±0.2	4.03±0.2	0.56±0.2	0.59±0.2	0.63±0.2	2.17±0.2	2.26±0.2	2.37±0.1
DeepMove	0.94±0.2	1.15±0.2	1.39±0.2	3.29±0.2	3.64±0.2	4.55±0.2	0.77±0.2	0.82±0.1	0.85±0.1	2.82±0.2	2.90±0.2	3.02±0.2
GCN	1.77±0.2	2.17±0.3	3.15±0.3	4.20±0.5	4.83±0.5	6.17±0.6	1.19±0.2	1.28±0.2	1.40±0.2	3.10±0.3	3.25±0.3	3.42±0.2
GIN	1.92±0.2	2.46±0.4	3.42±0.9	4.95±0.2	6.28±0.4	8.15±0.3	1.26±0.2	1.39±0.2	1.51±0.2	4.12±0.2	4.42±0.2	4.67±0.2
Topo-LSTM	1.78±0.2	2.25±0.3	3.30±0.4	4.66±0.1	5.58±0.2	7.17±0.1	1.28±0.1	1.40±0.1	1.54±0.1	3.17±0.2	3.38±0.2	3.58±0.2
SNIDSA	1.44±0.2	2.26±0.1	3.40±0.1	2.11±0.3	5.36±0.2	7.13±0.1	1.10±0.2	1.29±0.2	1.43±0.2	1.53±0.2	2.25±0.1	2.51±0.1
FOREST	1.77±0.3	2.53±0.4	3.50±0.3	4.55±0.2	5.32±0.4	6.48±0.4	1.14±0.1	1.32±0.2	1.45±0.2	3.88±0.3	4.06±0.3	4.21±0.3
Inf-VAE	2.03±0.2	2.54±0.2	3.73±0.2	6.39±0.2	7.33±0.1	8.64±0.3	1.56±0.1	1.66±0.1	1.80±0.1	5.02±0.2	5.20±0.2	5.44±0.2
DGDI	**3.76±0.3**	**4.89±0.4**	**6.95±0.9**	**7.04±0.3**	**8.88±1.1**	**12.19±1.1**	**2.74±0.2**	**2.99±0.2**	**3.25±0.2**	**5.52±0.2**	**5.94±0.3**	**6.37±0.3**

Table 2: MAP@K and Recall@K comparison of different methods on two datasets: our model (denoted by bold) significantly outperforms the strongest baseline.

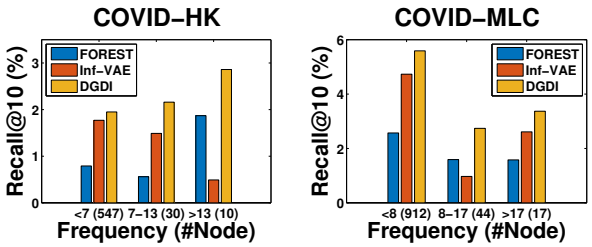

Figure 3: Recall@10 of COVID-HK and COVID-MLC on the frequency of location appearances. DGDI outperforms baselines on all groups.

Method	COVID-HK		COVID-MLC	
	Recall@10	MAP@10	Recall@10	MAP@10
Remove GNN	2.28±0.3	1.18±0.1	8.79±0.4	4.58±0.1
$\lambda_1 = 0$	6.56±0.5	2.75±0.3	11.12±1.1	5.82±0.5
Default	6.95±0.9	3.25±0.2	12.19±1.1	6.37±0.3

Table 3: The MAP@10 and Recall@10 of ablation study. The results are average over 5 runs.

these three results is equal to the results of the entire test set). As the figure shows, the model performs significantly better on head nodes than tail nodes since the performance of only a small number of head nodes is comparable to that of tail nodes. Moreover, we can observe that DGDI outperforms Inf-VAE and FOREST in all groups, demonstrating that DGDI is better than other models even under data sparsity.

Model Ablation Study

To investigate the contribution of our model design choices, the ablation study is conducted including:

1. **Remove GNN**: This method removes the GNN of the model design. Since $I(\boldsymbol{G}; \boldsymbol{H})$ depends on GNN to obtain graph representations. If the GNN is removed, $I(\boldsymbol{G}; \boldsymbol{H})$ is removed as well. Thus, this method only uses self-attention to derive diffusion representations.
2. $\boldsymbol{\lambda_1 = 0}$: Only use $\mathcal{L}_U$ to train the model and GNN is used to derive diffusion representations.

Table 3 shows MAP@10 and Recall@10 of our model and its variants. The last row denotes the performance of our model in default settings.

Due to the lack of graph information, the performances decrease on both datasets by a large margin after removing the GNN, which reflects the importance of geometric information when predicting COVID-19 diffusion. Maximizing $\mathcal{L}_G$ enforces diffusion embeddings to encode the geometric information. When setting $\lambda_1 = 0$, the performance drops in all cases, which demonstrates the effectiveness of $\mathcal{L}_G$. Specifically, the relative improvement w.r.t. MAP@10 on COVID-HK and COVID-MLC is 18.18% and 9.45%, demonstrating the effectiveness of DGDI.

Conclusion

In this work, we argue that human mobility modeling plays an important role during the NPI design and propose to predict case locations in the COVID-19 transmission. We present a novel mutual information maximization framework, named DGDI, to jointly learn the high-quality location, geometric graph, and diffusion representations. A lower bound of DGDI consists of the sum of two univariate mutual information is derived to optimize the model. Compared with other related works, DGDI can better handle data sparsity of diffusion and location appearance. Two COVID-19 datasets are proposed to facilitate the research of COVID-19 mobility modeling. Experimental results on these two real-world datasets show that DGDI outperforms other competitors by a significant margin on both Recall and MAP.

Acknowledgements

The research of Li was supported by NSFC Grant No. 62206067, Tencent AI Lab Rhino-Bird Focused Research Program RBFR2022008 and Guangzhou-HKUST(GZ) Joint Funding Scheme. The research of Tsung was supported by the Hong Kong RGC General Research Funds 16216119 and Foshan HKUST Projects FSUST20-FYTRI03B.

References

Anderson, C. 2006. *The long tail: Why the future of business is selling less of more*. Hachette Books.

Barbieri, N.; Bonchi, F.; and Manco, G. 2012. Topic-Aware Social Influence Propagation Models. In *ICDM*, 81–90.

Chang, S.; Pierson, E.; Koh, P. W.; Gerardin, J.; Redbird, B.; Grusky, D.; and Leskovec, J. 2021a. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840): 82–87.

Chang, S.; Pierson, E.; Pang, W. K.; Gerardin, J.; and Leskovec, J. 2020. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 1–6.

Chang, S.; Wilson, M. L.; Lewis, B. L.; Mehrab, Z.; Dudakiya, K. K.; Pierson, E.; Koh, P. W.; Gerardin, J.; Redbird, B.; Grusky, D.; Marathe, M.; and Leskovec, J. 2021b. Supporting COVID-19 Policy Response with Large-scale Mobility-based Modeling. In *KDD*, 2632–2642.

Chen, M.; Liu, Y.; and Yu, X. 2014. NLPMM: A Next Location Predictor with Markov Modeling. In *PAKDD*, 186–197.

Cho, K.; van Merrienboer, B.; Gülçehre, Ç.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *EMNLP*, 1724–1734.

Clouston, S. A.; Natale, G.; and Link, B. G. 2021. Socioeconomic inequalities in the spread of coronavirus-19 in the United States: A examination of the emergence of social inequalities. *Social Science & Medicine*, 268: 113554.

Feng, J.; Li, Y.; Zhang, C.; Sun, F.; Meng, F.; Guo, A.; and Jin, D. 2018a. DeepMove: Predicting Human Mobility with Attentional Recurrent Networks. In *WWW*, 1459–1468.

Feng, S.; Cong, G.; Khan, A.; Li, X.; Liu, Y.; and Chee, Y. M. 2018b. Inf2vec: Latent Representation Model for Social Influence Embedding. In *ICDE*, 941–952.

Gibbs, H.; Liu, Y.; Pearson, C. A.; Jarvis, C. I.; Grundy, C.; Quilty, B. J.; Diamond, C.; and Eggo, R. M. 2020. Changing travel patterns in China during the early stages of the COVID-19 pandemic. *Nature communications*, 11(1): 1–9.

Gibbs, H.; Nightingale, E.; Liu, Y.; Cheshire, J.; Danon, L.; Smeeth, L.; Pearson, C. A.; Grundy, C.; working group, L. C. C.-.; Kucharski, A. J.; et al. 2021. Detecting behavioural changes in human movement to inform the spatial scale of interventions against COVID-19. *PLoS computational biology*, 17(7): e1009162.

Granovetter, M. 1978. Threshold models of collective behavior. *American journal of sociology*, 83(6): 1420–1443.

Haug, N.; Geyrhofer, L.; Londei, A.; Dervic, E.; Desvars-Larrive, A.; Loreto, V.; Pinior, B.; Thurner, S.; and Klimek, P. 2020. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature human behaviour*, 4(12): 1303–1312.

Hochreiter, S.; and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural Comput.*, 9(8): 1735–1780.

Hu, S.; Xiong, C.; Yang, M.; Younes, H.; Luo, W.; and Zhang, L. 2021. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. *Transportation Research Part C: Emerging Technologies*, 124: 102955.

Islam, M. R.; Muthiah, S.; Adhikari, B.; Prakash, B. A.; and Ramakrishnan, N. 2018. DeepDiffuse: Predicting the 'Who' and 'When' in Cascades. In *ICDM*, 1055–1060.

Jing, B.; Park, C.; and Tong, H. 2021. HDMI: High-order Deep Multiplex Infomax. In *WWW*, 2414–2424.

Kempe, D.; Kleinberg, J. M.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *KDD*, 137–146.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

Laborde, D.; Martin, W.; Swinnen, J.; and Vos, R. 2020. COVID-19 risks to global food security. *Science*, 369(6503): 500–502.

Lai, S.; Ruktanonchai, N. W.; Zhou, L.; Prosper, O.; Luo, W.; Floyd, J. R.; Wesolowski, A.; Santillana, M.; Zhang, C.; Du, X.; et al. 2020. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *nature*, 585(7825): 410–413.

Lee, J.; Lee, I.; and Kang, J. 2019. Self-Attention Graph Pooling. In *ICML*, volume 97 of *Proceedings of Machine Learning Research*, 3734–3743.

Levin, R.; Chao, D. L.; Wenger, E. A.; and Proctor, J. L. 2021. Insights into population behavior during the COVID-19 pandemic from cell phone mobility data and manifold learning. *Nature Computational Science*, 1(9): 588–597.

Li, J.; Huang, Y.; Chang, H.; and Rong, Y. 2022. Semi-Supervised Hierarchical Graph Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Li, J.; Rong, Y.; Cheng, H.; Meng, H.; Huang, W.; and Huang, J. 2019. Semi-Supervised Graph Classification: A Hierarchical Graph Perspective. In *WWW*, 972–982.

Li, X.; Hu, R.; Wang, Z.; and Yamasaki, T. 2021. Location Predicts You: Location Prediction via Bi-direction Speculation and Dual-level Association. In *IJCAI*, 529–536.

Lim, N.; Hooi, B.; Ng, S.; Wang, X.; Goh, Y. L.; Weng, R.; and Varadarajan, J. 2020. STP-UDGAT: Spatial-Temporal-Preference User Dimensional Graph Attention Network for Next POI Recommendation. In *CIKM*, 845–854.

Liu, B.; Fu, Y.; Yao, Z.; and Xiong, H. 2013. Learning geographical preferences for point-of-interest recommendation. In *KDD*, 1043–1051.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation Learning with Contrastive Predictive Coding. *CoRR*, abs/1807.03748.

Peak, C. M.; Wesolowski, A.; zu Erbach-Schoenberg, E.; Tatem, A. J.; Wetter, E.; Lu, X.; Power, D.; Weidman-Grunewald, E.; Ramos, S.; Moritz, S.; et al. 2018. Population mobility reductions associated with travel restrictions during the Ebola epidemic in Sierra Leone: use of mobile phone data. *International journal of epidemiology*, 47(5): 1562–1570.

Peng, Z.; Huang, W.; Luo, M.; Zheng, Q.; Rong, Y.; Xu, T.; and Huang, J. 2020. Graph Representation Learning via Graphical Mutual Information Maximization. In *WWW*, 259–270.

Qiu, J.; Chen, Q.; Dong, Y.; Zhang, J.; Yang, H.; Ding, M.; Wang, K.; and Tang, J. 2020. GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training. In *KDD*, 1150–1160.

Rong, Y.; Cheng, H.; and Mo, Z. 2015. Why it happened: Identifying and modeling the reasons of the happening of social events. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1015–1024. ACM.

Rong, Y.; Zhu, Q.; and Cheng, H. 2016. A model-free approach to infer the diffusion network from event cascade. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, 1653–1662. ACM.

Saito, K.; Kimura, M.; Ohara, K.; and Motoda, H. 2009. Learning Continuous-Time Information Diffusion Model for Social Behavioral Data Analysis. In *ACML*, volume 5828, 322–337.

Sankar, A.; Zhang, X.; Krishnan, A.; and Han, J. 2020. Inf-VAE: A Variational Autoencoder Framework to Integrate Homophily and Influence in Diffusion Prediction. In *WSDM*, 510–518.

Schwabe, A.; Persson, J.; and Feuerriegel, S. 2021. Predicting COVID-19 Spread from Large-Scale Mobility Data. In *KDD*, 3531–3539.

Sun, F.; Hoffmann, J.; Verma, V.; and Tang, J. 2020a. InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization. In *ICLR*.

Sun, K.; Qian, T.; Chen, T.; Liang, Y.; Nguyen, Q. V. H.; and Yin, H. 2020b. Where to Go Next: Modeling Long- and Short-Term User Preferences for Point-of-Interest Recommendation. In *AAAI*, 214–221.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In *NIPS*, 5998–6008.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *ICLR*.

Velickovic, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep Graph Infomax. In *ICLR*.

Venkatramanan, S.; Sadilek, A.; Fadikar, A.; Barrett, C. L.; Biggerstaff, M.; Chen, J.; Dotiwalla, X.; Eastham, P.; Gipson, B.; Higdon, D.; et al. 2021. Forecasting influenza activity using machine-learned mobility map. *Nature communications*, 12(1): 1–12.

Wang, J.; Zheng, V. W.; Liu, Z.; and Chang, K. C. 2017a. Topological Recurrent Neural Network for Diffusion Prediction. In *ICDM*, 475–484.

Wang, M. L.; Behrman, P.; Dulin, A.; Baskin, M. L.; Buscemi, J.; Alcaraz, K. I.; Goldstein, C. M.; Carson, T. L.; Shen, M.; and Fitzgibbon, M. 2020. Addressing inequities in COVID-19 morbidity and mortality: research and policy recommendations. *Translational Behavioral Medicine*, 10(3): 516–519.

Wang, P.; Han, K.; Wei, X.-S.; Zhang, L.; and Wang, L. 2021a. Contrastive Learning based Hybrid Networks for Long-Tailed Image Classification. In *CVPR*, 943–952.

Wang, R.; Huang, Z.; Liu, S.; Shao, H.; Liu, D.; Li, J.; Wang, T.; Sun, D.; Yao, S.; and Abdelzaher, T. F. 2021b. DyDiff-VAE: A Dynamic Variational Framework for Information Diffusion Prediction. In *SIGIR*, 163–172.

Wang, Y.; Shen, H.; Liu, S.; Gao, J.; and Cheng, X. 2017b. Cascade Dynamics Modeling with Attention-based Recurrent Neural Network. In *IJCAI*, 2985–2991.

Wang, Z.; Chen, C.; and Li, W. 2018. A Sequential Neural Information Diffusion Model with Structure Attention. In *CIKM*, 1795–1798.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2019. How Powerful are Graph Neural Networks? In *ICLR*.

Xue, H.; and Salim, F. D. 2021. Exploring Self-Supervised Representation Ensembles for COVID-19 Cough Classification. In *KDD*, 1944–1952.

Yang, C.; Sun, M.; Liu, H.; Han, S.; Liu, Z.; and Luan, H. 2021. Neural Diffusion Model for Microscopic Cascade Study. *IEEE Trans. Knowl. Data Eng.*, 33(3): 1128–1139.

Yang, C.; Tang, J.; Sun, M.; Cui, G.; and Liu, Z. 2019. Multi-scale Information Diffusion Prediction with Reinforced Recurrent Networks. In *IJCAI*, 4033–4039.

Yao, D.; Zhang, C.; Huang, J.; and Bi, J. 2017. SERM: A Recurrent Model for Next Location Prediction in Semantic Trajectories. In *CIKM*, 2411–2414.

Ying, Z.; You, J.; Morris, C.; Ren, X.; Hamilton, W. L.; and Leskovec, J. 2018. Hierarchical Graph Representation Learning with Differentiable Pooling. In *NeurIPS*, 4805–4815.

You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; and Shen, Y. 2020. Graph Contrastive Learning with Augmentations. In *NeurIPS*.

Yu, F.; Cui, L.; Guo, W.; Lu, X.; Li, Q.; and Lu, H. 2020. A Category-Aware Deep Model for Successive POI Recommendation on Sparse Check-in Data. In *WWW*, 1264–1274.

Zhang, J.; Chow, C.; and Li, Y. 2014. LORE: exploiting sequential influence for location recommendations. In *SIGSPATIAL*, 103–112.

Zhang, L.; Li, Q.; Liang, Z.; Li, T.; Liu, S.; Cui, Q.; Nie, J.; Wu, Q.; Qu, X.; Huang, W.; et al. 2022. The significant immune escape of pseudotyped SARS-CoV-2 Variant Omicron. *Emerging microbes & infections*, 11(1): 1–5.