

Explainable Origin-Destination Crowd Flow Interpolation via Variational Multi-Modal Recurrent Graph Auto-Encoder

Qiang Zhou^{1,*}, Xinjiang Lu^{2,†}, Jingjing Gu¹, Zhe Zheng¹, Bo Jin³, Jingbo Zhou²

¹Nanjing University of Aeronautics and Astronautics, Nanjing, China

²Baidu Research, Beijing, China

³Dalian University of Technology, Dalian, China

{zhouqnuuacs, gujingjing, zhengzhe}@nuaa.edu.cn, {luxinjiang, zhoujingbo}@baidu.com, jinbo@dlut.edu.cn

Abstract

Origin-destination (OD) crowd flow, if more accurately inferred at a fine-grained level, can potentially enhance the efficacy of various urban applications. In practice, for mining OD crowd flow with effect, the problem of spatially interpolating OD crowd flow occurs because of the ineluctable missing values. This problem is further complicated by the inherent scarcity and noise nature of OD crowd flow data. In this paper, we propose an uncertainty-aware interpolative and explainable framework, namely UApex, for realizing reliable and trustworthy OD crowd flow interpolation. Specifically, we first design a Variational Multi-modal Recurrent Graph Auto-Encoder (VMR-GAE) for uncertainty-aware OD crowd flow interpolation. A key idea here is to formulate the problem as semi-supervised learning on directed graphs. Next, to mitigate the data scarcity, we incorporate a distribution alignment mechanism to introduce supplementary modals into variational inference. Then, a dedicated decoder with a Poisson prior is proposed for the task. Moreover, to make VMR-GAE more trustworthy, we develop an efficient and uncertainty-aware explainer that can explain spatiotemporal topology via the Shapley value. Extensive experiments on two real-world datasets validate that VMR-GAE outperforms the state-of-the-art baselines. Also, an exploratory empirical study shows that the proposed explainer can generate meaningful spatiotemporal explanations.

Introduction

Inferring human mobility at a more fine-grained level, i.e., origin-destination (OD) crowd flow, can benefit various new applications in the urban computing area. For example, with the cross-region OD crowd flow rather than the incoming/out-going flow in a region, the decision-making for alleviating traffic jams can be more efficient by focusing on the problematic OD pairs without excessive traffic restrictions. Therefore, it is the importance of OD crowd flow that has attracted attention from both academia and industry.

For instance, (Xu et al. 2023; Lin et al. 2023; Huang et al. 2022; Zhang et al. 2021b) were devoted to overcoming the sparsity in OD crowd flow prediction. (Wang et al. 2019c)

*This work was done when the first author was an intern at Baidu Research.

†Corresponding author.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

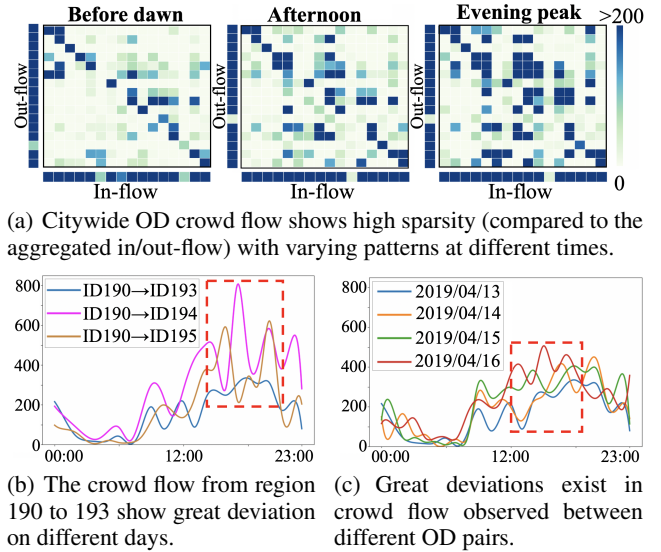


Figure 1: The citywide OD crowd flow usually depicts not only sparsity but also uncertainty.

proposed a sparsity-aware multi-task framework for OD matrix prediction. To predict OD-based ride-hailing demands, DiDi (Ke et al. 2021) evades the sparsity in OD matrix by directly modeling correlations between OD pairs. However, these works neglect the noise and uncertainty in OD crowd flow data. Some researchers have paid attention to the uncertainty-aware OD prediction. For example, (Li et al. 2022) leveraged the Bayesian approximation to quantify the uncertainty in OD demand, whereas, the proposed approach can hardly extrapolate unseen OD demands.

As the crowd flow data becomes more fine-grained, more data quality issues come to light. As shown in Fig. 1(a), the citywide OD crowd flow is sparse most of the time compared to the aggregated in/out-flow observed within regions. Hence, the innate data scarcity makes it challenging to learn effective representations without enough context. Meanwhile, Fig. 1(b) and 1(c) show that the crowd flow is highly uncertain temporally and spatially. Thus, it is critical to quantify the uncertainty of OD crowd flow.

Since the advancement in spatiotemporal data mining

techniques (Zhang et al. 2019; Miao et al. 2023), we could have drawn more insights from SOTA models to enhance the OD crowd flow prediction/analytics. However, it is precisely due to the data issues of OD crowd flow that the progress in relevant research has been limited. Hence, we suggest a two-stage approach to the research. In the first stage, our focus lies in resolving the data issues through OD crowd flow interpolation. In the second stage, we can subsequently employ the latest methods for accurate OD crowd flow analytics. In this thought, the key is how to realize reliable and trustworthy OD crowd flow interpolation regarding the inherent data scarcity and noise nature. Moreover, although we can adopt uncertainty quantification to reduce the risk of interpolative errors (Zhou et al. 2021), the black-box model can be brittle and unfair (Li et al. 2023; Ali et al. 2023), and can harm the critical downstream tasks like signal control and autonomous driving.

To this end, we propose an uncertainty-aware interpolative and explainable framework (UApeX), which includes an interpolator and its corresponding explainer. For the interpolator, we propose a Variational Multi-modal Recurrent Graph Auto-Encoder (VMR-GAE) by incorporating contextual information into Bayesian deep learning with uncertainty quantification. Specifically, we first formulate OD crowd flow as a dynamic directed graph where the edges encode the flow values. Then, we build the diffusion GCN (Atwood and Towsley 2016) with a recurrent structure to study the spatiotemporal dependency at each time step. To enrich the representation of OD crowd flow, we exploit additional contexts into our VMR-GAE by proposing a prior distribution alignment method. Next, we reconstruct the graph of OD crowd flow at the target time step through the deep decoder with a Poisson prior. Moreover, we derive an uncertainty-aware spatiotemporal explainer to equip VMR-GAE with explainability based on the Shapley value, which discovers the most important subgraph structures from the spatiotemporal topology perspective. Given the explainer, we further propose an efficient explanation generation algorithm encouraged by the characteristics of urban structure. To sum up, our contributions are as follows:

- We advocate investigating OD crowd flow interpolation to address the data issues, thereby enabling the insights of SOTA spatiotemporal models and emphasizing intrinsic OD crowd flow characteristics.
- We propose a Variational Multi-modal Recurrent Graph Auto-Encoder, which can incorporate multi-modals and uncertainty quantification for OD crowd flow interpolation to overcome the sparsity and noise issues.
- We derive an uncertainty-aware spatiotemporal explainer and its efficient exploration process to enhance interpolation outcomes' credibility further.

Related Work

Origin-Destination Crowd Flow. Human mobility modeling and crowd flow prediction have been paid much attention to. Recently, the interest in OD crowd flow has been gaining momentum. Some traditional methods, like Matrix Factorization (MF) and Collaborative Filtering (CF), have been

explored for OD crowd/traffic flow prediction (Deng et al. 2016; Gu et al. 2020; Ros-Roca et al. 2022). With the powerful expressive capability, graph neural networks (GNNs) have been applied to incorporate spatial dependencies in OD representation (Wang et al. 2019c; Rong et al. 2021; Xu et al. 2023; Huang et al. 2023; Shi et al. 2020; Feng et al. 2021). However, none focuses on noise issues or the uncertainty of prediction results.

Trustworthy Graph Neural Network. Uncertainty quantification has become a common practice to support critical decision-making in risk-aware applications (Wang et al. 2019a; Kong, Sun, and Zhang 2020; Gal and Ghahramani 2016; Lakshminarayanan, Pritzel, and Blundell 2017). In urban computing, some traffic prediction methods (Zhou et al. 2020, 2021; Wang et al. 2023a), including OD crowd flow inference (Pitombeira-Neto, Loureiro, and Carvalho 2020; Jeong and Park 2021; Wang et al. 2023b), have been encouraged to help in understanding the model behavior through uncertainty quantification. Still, these works can hardly be applied to the OD matrix with high sparsity.

Because of the significance of explainability (Xie et al. 2020), there have been studies on this topic (Li et al. 2016; Lundberg and Lee 2017; Cheng et al. 2020). To investigate GNN explainability, several explainers (Ying et al. 2019; Luo et al. 2020; Yuan et al. 2021, 2020) that focus on identifying important components related to a graph have achieved success. Especially, (Zhang et al. 2021a) leverage the Shapley value from game theory to capture the subgraph importance, which inspired us to deliberate the contributions of structural information in the urban environment. Despite the above pioneer studies on explainability, how to develop an efficient algorithm to make the explainer more practical in real-world applications is still an open question.

Preliminaries

Definitions and Problem Statements

Definition 1 (OD Crowd Flow) *OD crowd flow involves a user set \mathcal{U} and a region pair $\langle r_i, r_j \rangle$ ($i \neq j$) that \mathcal{U} move from r_i to r_j in a period of time $[t_0, t]$ ($t_0 < t$), which is denoted by $\tau_{i,j}^{(t_0:t)} : r_i, t_0 \xrightarrow{\mathcal{U}} r_j, t$. $\tau_{i,j}^{(t_0:t)}$ can be simplified as $\tau_{i,j}^{(t)}$ to denote the crowd flow observed at time step t .*

Definition 2 (OD Matrix) *Given a region set $\mathcal{R} = \{r_1, \dots, r_{n_t}\}$, an OD matrix $\mathbf{A}^{(t)} \in \mathbb{R}^{n_t \times n_t}$ records the OD crowd flow between each pair of regions in \mathcal{R} , i.e. $A_{i,j}^{(t)} = \tau_{i,j}^{(t)}$, $1 \leq i, j \leq n_t$.*

Definition 3 (OD Graph) *An OD graph is a dynamic directed graph that evolves over time, which is denoted by $\mathcal{G} = \{G^{(1)}, G^{(2)}, \dots, G^{(T)}\}$, where $G^{(t)} = (\mathcal{V}, \mathcal{E}^{(t)})$ is an OD graph at time step t with \mathcal{V} and $\mathcal{E}^{(t)}$ being the corresponding node and directed edge sets, respectively. The set $\mathcal{E}^{(T)}$ reflects the paired OD crowd flow, and $\mathbf{A}^{(t)} \in \mathbb{R}^{n_t \times n_t}$ is the adjacency matrix of $G^{(t)}$. The region/station features of nodes in \mathcal{V} are stored in $\mathbf{X}^{(t)} \in \mathbb{R}^{n_t \times m}$.*

Then, we introduce the context modal that differs from the target OD crowd flow as supplementary knowledge.

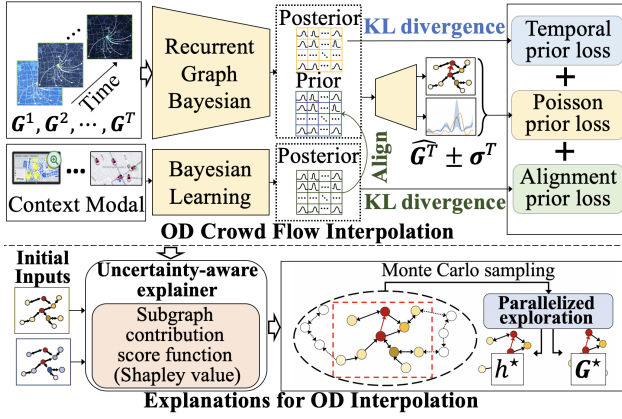


Figure 2: Framework overview of UApex.

Definition 4 (Context Modal) The context modal \mathcal{C} is another set of city data correlated to the target OD crowd flow. Via a specific deep learning architecture, the input of \mathcal{C} can be mapped to a regional representation of time step t , denoted as $\tilde{\mathbf{Z}}^{(t)} \in \mathbb{R}^{n \times d}$.

Due to the data from urban sensors often having missing values, we take the OD graph \mathcal{G} as well as the context modal \mathcal{C} (if available) as input to interpolate the missing values in $\mathbf{A}^{(T)}$. Formally, we have the following problem statement:

Problem 1 (OD Crowd Flow Interpolation) Given a dynamic OD graph \mathcal{G} , learn a function $\mathcal{F}(\cdot)$ that takes $\mathbf{A}^{(\leq T)}$ and $\mathbf{X}^{(\leq T)}$ as input. If available, $\mathcal{F}(\cdot)$ can take additional contexts \mathcal{C} as input. Then, we infer a spatial interpolated OD matrix $\hat{\mathbf{A}}^{(T)}$ of the last time step T :

$$\{\mathbf{A}^{(\leq T)}, \mathbf{X}^{(\leq T)}, \mathcal{C}\} \xrightarrow{\mathcal{F}(\cdot)} \hat{\mathbf{A}}^{(T)}. \quad (1)$$

We further study the explanation problem for OD crowd flow interpolation. In common practice, we derive the hidden state representation $\mathbf{h}^{(T-1)}$ given the input $G^{(<T)}$. Therefore, we suggest to consider $G^{(T)}$ at time T , and the hidden state $\mathbf{h}^{(T-1)}$ at time $T-1$ for yielding concise explanations. Formally, we have another problem statement:

Problem 2 (Explainer of OD Crowd Flow Interpolation) Given a trained OD crowd flow interpolation model \mathcal{F} , the explainer generates G^* and \mathbf{h}^* for outcomes $\hat{A}_{i,j}^{(T)} \in \hat{\mathbf{A}}^{(T)}$, where G^* is a subgraph of $G^{(T)}$ and \mathbf{h}^* is a subset of $\mathbf{h}^{(T-1)}$. More precisely, the generated explanations are the subgraph of $G^{(T)}$ and the subset of $\mathbf{h}^{(T-1)}$ (denoted by $G' \subseteq G^{(T)}$ and $\mathbf{h}' \subseteq \mathbf{h}^{(T-1)}$, respectively) that maximize the contributions to the interpolation:

$$G^*, \mathbf{h}^* = \underset{G' \subseteq G^{(T)}, \mathbf{h}' \subseteq \mathbf{h}^{(T-1)}}{\operatorname{argmax}} \mathcal{S}(\mathcal{F}, G', \mathbf{h}'), \quad (2)$$

where $\mathcal{S}(\cdot)$ is the score function to measure the contribution.

Framework Overview

Fig. 2 depicts the overview of UApex that consists of two stages, i.e., OD crowd flow interpolation and explanation

exploration. In the first stage, UApex takes the OD graph as input, then estimates the prior and posterior distributions of latent variables via Bayesian graph networks. If available, context modalities can be adopted to enrich the original representations through prior distribution alignment. Afterward, a deep decoder inputs the enriched representation and implements the directed graph reconstruction. In the second stage, we develop an uncertainty-aware explainer for producing explanations of VMR-GAE, in which a score function based on the Shapley value is proposed.

Methodology

OD Crowd Flow Interpolation

VMR-GAE employs the encoder-decoder architecture comprising three components that require detailed exposition.

Variational Recurrent Graph Encoder The variational graph recurrent neural network (VGRNN) proposed by (Hajiramezanali et al. 2019) learns the prior distribution parameters based on the historical data to capture dependencies within the graph evolution process. To further modify this uncertainty-aware spatiotemporal model for the directed OD graph, we incorporate diffusion convolution gate recurrent units (Li et al. 2018) to capture hidden states $\mathbf{h}^{1 \sim (T-1)}$ from historical OD crowd flow $(\mathbf{A}^{<t}, \mathbf{X}^{<t})$. Also, we stack diffusion graph convolution layers as our encoder, denoted as $\varphi(\cdot)$, for directed OD graph representation learning. Formally, the prior and posterior distributions of encoded representation $\mathbf{Z}^{(t)}$ at time step t as follows,

$$p(\mathbf{Z}^t | \mathbf{A}^{<t}, \mathbf{X}^{<t}) = \prod_{i=1}^n \mathcal{N}(\mu_{i,pr}^t, \operatorname{diag}((\sigma_{i,pr}^t)^2)), \quad (3)$$

$$q(\mathbf{Z}^t | \mathbf{A}^t, \mathbf{X}^t, \mathbf{h}^{t-1}) = \prod_{i=1}^n \mathcal{N}(\mu_{i,en}^t, \operatorname{diag}((\sigma_{i,en}^t)^2)),$$

where μ_{pr}^t and σ_{pr}^t denote the parameters of the conditional prior distribution, derived from hidden states \mathbf{h}^{t-1} after passing through a set of fully connected layers. μ_{en}^t and σ_{en}^t denote the parameters of the approximated posterior, derived by the encoder as follows,

$$\mu_{en}^t = \varphi_{\mu}(\mathbf{A}^t, [\mathbf{X}^t, \mathbf{h}^{t-1}]), \quad (4)$$

$$\sigma_{en}^t = \varphi_{\sigma}(\mathbf{A}^t, [\mathbf{X}^t, \mathbf{h}^{t-1}]),$$

where $[\cdot, \cdot]$ means concatenation in the feature dimension, φ_{μ} and φ_{σ} are encoders, respectively. Thereafter, we can capture the dependencies of the OD graph evolution by minimizing the difference between the prior and posterior.

Multi-modal Distribution Alignment It has been proven that incorporating information from other modals can optimize the empirical risk and improve the latent representation quality (Huang et al. 2021). Thus, multi-modal data is exactly what we need for enhancing the robustness of OD crowd flow interpolation. Since early fusion could hardly work due to the data gaps, e.g., the different lengths of time steps, we suggest to incorporate context modal after representation learning. While, this is challenging because of

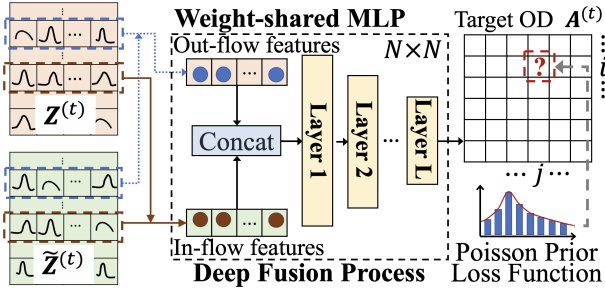


Figure 3: The decoder architecture of VMR-GAE.

the nature of our variational encoder, which implies that the representations of a context modal $\tilde{\mathbf{Z}}^{(t)}$ should follow specific distributions. In practice, the prior of $\tilde{\mathbf{Z}}^{(t)}$ can be arbitrary, while most practitioners adopt the standard Gaussian distribution (Kipf and Welling 2016). However, this might be deficient when the cross-domain gap is considerable. In transfer learning, domain adaptation (Long et al. 2016; Kamath, Liu, and Whitaker 2019) is often used to minimize the difference between the target and source domains via maximum mean discrepancy (MMD) loss. Inspired by this, we propose a simple but effective distribution alignment mechanism to align the representation distributions of the context modal with the prior distributions in Eq. (3) as follows.

$$p(\tilde{\mathbf{Z}}^t) = p(\mathbf{Z}^t | \mathbf{A}^{<t}, \mathbf{X}^{<t}). \quad (5)$$

Note we discard the gradient of Eq. (5) when using back-propagation for training to avoid cross-modal data corruption. Accordingly, we can concatenate the representations of multi-modals along the feature dimension without worrying about the distribution differences in the latent space.

Decoder for Graph Reconstruction Inner production as a common method (Hajiramezanali et al. 2019; Bonner et al. 2019) for relation reconstruction proves insufficient in capturing the dynamics of OD crowd flow. We need more powerful operators to enhance the graph reconstruction performance (Kipf and Welling 2016; Hajiramezanali et al. 2019). As illustrated in Fig. 3, we stack the multi-layer perceptron (MLP) for OD graph reconstruction.

First, we reparameterize the representation distributions and concatenate the multi-modal representations. The outcomes are fed into $f_{\mathbf{z}_{in}}$ and $f_{\mathbf{z}_{out}}$, respectively.

$$\mathbf{Z}_{in}^t = f_{\mathbf{z}_{in}}([\mathbf{Z}^t, \tilde{\mathbf{Z}}^t]), \quad \mathbf{Z}_{out}^t = f_{\mathbf{z}_{out}}([\mathbf{Z}^t, \tilde{\mathbf{Z}}^t]). \quad (6)$$

In Eq. (6), the weights of $f_{\mathbf{z}_{in}}$ and $f_{\mathbf{z}_{out}}$ are shared across all nodes and time steps. Afterward, we concatenate each pair of node-wise in-flow and out-flow features, i.e., $\mathbf{Z}_{OD,i,j}^t = [\mathbf{Z}_{out,i}^t, \mathbf{Z}_{in,j}^t]$. Then, \mathbf{Z}_{OD}^t is fed into the MLP where the weights are shared among OD pairs, such that VMR-GAE can generalize to the unseen OD pairs for OD crowd flow interpolation.

Finally, the loss function of VMR-GAE is derived from the negative logarithm of the evidence lower bound.

$$\begin{aligned} \mathcal{L} = - \sum_{t=1}^T & \left(\mathbb{E}_{\mathbf{Z}^t \sim q(\mathbf{Z}^t | \mathbf{A}^t, \mathbf{X}^t, \mathbf{h}^{t-1})} \right. \\ & \mathbb{E}_{\tilde{\mathbf{Z}}^t \sim q(\tilde{\mathbf{Z}}^t | \tilde{\mathbf{A}}^t)} \log p(\mathbf{A}^t | \mathbf{Z}^t, \tilde{\mathbf{Z}}^t) \\ & + \text{KL}(q(\tilde{\mathbf{Z}}^t | \tilde{\mathbf{A}}^t) || p(\mathbf{Z}^t | \mathbf{A}^{<t}, \mathbf{X}^{<t})) \\ & \left. + \text{KL}(q(\mathbf{Z}^t | \mathbf{A}^t, \mathbf{X}^t, \mathbf{h}^{t-1}) || p(\mathbf{Z}^t | \mathbf{A}^{<t}, \mathbf{X}^{<t})) \right). \end{aligned} \quad (7)$$

Since we can regard a regional trip as an event that occurred, the OD crowd flow can be approximated with a Poisson distribution (Gu et al. 2020). In practice, we pre-define the posterior distribution $p(\mathbf{A}^t | \mathbf{Z}^t, \tilde{\mathbf{Z}}^t)$ as Poisson distribution.

Explanations for OD Interpolation

Spatiotemporal Explainer Post hoc explanation techniques can help us understand black box models without in-process interpolation overhead. Traditional GNN explanation methods (Ying et al. 2019; Luo et al. 2020) mainly examine the individual importance of nodes/edges while ignoring the subset importance, such that the complex entirety of urban crowd flow, such as the living circle effect, cannot be well studied. Inspired by SubgraphX (Yuan et al. 2021) which addresses subgraph importance, we propose to apply Shapley value (Kuhn and Tucker 1953) for the urban topology importance evaluation. However, this approach encounters a challenge when applied to spatiotemporal models.

To realize it, we derive a spatiotemporal explainer and design a score function \mathcal{S} with the Shapley value in Eq. (2) for the set (G', \mathbf{h}') . Specifically, given a set combining \mathcal{V}^T and \mathbf{h}^{T-1} , i.e., $\{v_1, \dots, v_n, \mathbf{h}_1, \dots, \mathbf{h}_n\}$, we have $(G', \mathbf{h}') = \{v_1, \dots, v_{n_1}, \mathbf{h}_1, \dots, \mathbf{h}_{n_2}\}$ ($0 < n_1, n_2 < n$). Then, the set of all players is denoted by $\mathbf{P} = \{(G', \mathbf{h}'), v_{n_1+1}, \dots, v_n, \mathbf{h}_{n_2+1}, \dots, \mathbf{h}_n\}$, and the Shapley value $\psi(\mathcal{F}, G', \mathbf{h}')$ omitting model \mathcal{F} can be calculated as,

$$\begin{aligned} \psi(G', \mathbf{h}') = \sum_{\mathbf{S} \subseteq \mathbf{P} \setminus (G', \mathbf{h}')} & \left\{ \frac{(|\mathbf{P}| - |\mathbf{S}| - 1)! |\mathbf{S}|!}{|\mathbf{P}|!} \right. \\ & \left. \times [\mathcal{C}(\mathbf{S} \cup (G', \mathbf{h}')) - \mathcal{C}(\mathbf{S})] \right\}, \end{aligned} \quad (8)$$

where $\mathbf{S} = (G^S, \mathbf{h}^S)$ is an arbitrary coalition set of players, and $\mathcal{C}(\cdot)$ is defined to derive the contribution given \mathcal{F} . The lack of important information can result in uncertain interpolation. In other words, the contribution is negatively correlated with the outcome deviation. As the variational model \mathcal{F} quantifies the uncertainty along with the interpolation, we adopt the outcome deviation plus the confidence error to measure the contribution accordingly. Formally, the contribution function \mathcal{C} is as follows,

$$\mathcal{C}(\mathcal{F}, \mathbf{S} \cup (G', \mathbf{h}')) = e^{-\beta(\hat{A}_{i,j}^{(T)} - A_{i,j}^{(T)})^2} - \hat{\sigma}_{i,j}, \quad (9)$$

where β is the controlling weight, $\hat{\sigma}$ is the interpolation deviation, or regarded as the uncertainty.

Explanations Exploration Due to the high complexity of Eq. (8), we aim to propose an efficient exploration method. Regarding properties of the Shapley value to our problem, we have the variant of Linearity axiom (Zhang et al. 2020):

Theorem 1 *If two independent games (u and v) can be merged into one game (w), then the Shapley value of the new game can also be merged, i.e., $\forall (G', \mathbf{h}') \subseteq (G^{(T)}, \mathbf{h}^{(T-1)})$, $\psi_w(G', \mathbf{h}') = \psi_u(G', \mathbf{h}') + \psi_v(G', \mathbf{h}')$.*

Accordingly, we derive the split spatiotemporal explainer from Eq. (2) by jointly using Eq. (8) and Theorem 1. The Problem 2 can be rewritten as follows.

$$\begin{aligned} G^*, \mathbf{h}^* &= \operatorname{argmax}_{G' \subseteq G^{(T)}, \mathbf{h}' \subseteq \mathbf{h}^{(T-1)}} \psi(G', \mathbf{h}') \\ &= \operatorname{argmax}_{G' \subseteq G^{(T)}, \mathbf{h}' \subseteq \mathbf{h}^{(T-1)}} (\psi_u(G', \mathbf{h}') + \psi_v(G', \mathbf{h}')) \quad (10) \\ &= \operatorname{argmax}_{G' \subseteq G^{(T)}} \psi_u(G') + \operatorname{argmax}_{\mathbf{h}' \subseteq \mathbf{h}^{(T-1)}} \psi_v(\mathbf{h}'). \end{aligned}$$

Therefore, we can reduce the time overhead by half with parallel exploration. While, the time complexity is still up to $\mathcal{O}(n^2 \cdot 2^n)$, n is the number of regions.

To explain the OD interpolation efficiently, we suggest to carefully examine the urban environment as shown in the Fig. 2 bottom half. First, because human mobility adheres to certain constraints, such as geographical distance, it is wise to initialize G' with the nodes only close to the target origin and destination. Besides, we can consider the connectivity as the criteria to filter out trivial nodes. Moreover, we leverage the Monte Carlo sampling to approximate the calculation of Eq. (8). With the above steps, we develop an approximated exploration process with complexity $\mathcal{O}(k_1 k_2 M_C)$, where k_1 and k_2 are two integers to be set for initial exploration space ($0 < k_1, k_2 \ll n$), and M_C is the number of steps in Monte Carlo sampling. Finally, we apply the same procedure to historical explanations exploration. Thanks to parallel exploration, the overall time complexity remains $\mathcal{O}(k_1 k_2 M_C)$.

Experiments

Experimental Setups

Datasets. 1) The Beijing dataset was collected via an anonymized navigation application. 2) NYC dataset was generated from NYC taxicab data published by NYC TLC¹. The basic statistics of Beijing and NYC datasets can be found in Table 3. Besides, the rush hours were defined as 8:00 – 10:00 and 16:00 – 18:00 for both Beijing and NYC; the off-peak hours were 12:00 – 13:00 and 22:00 – 23:00 for Beijing, and 12:00 – 14:00 and 22:00 – 00:00 for NYC.

Baselines. We compare VMR-GAE with SOTAs for flow interpolation, i.e., a Matrix Factorization (MF) model (i.e., SVD (Ma, Sun, and Qin 2017)), deep graph models (i.e., GAE (Kipf and Welling 2016), 2D-GCN (Shi et al. 2020), NGCF (Wang et al. 2019b), and T-NGCF), and uncertainty-aware deep models (i.e., VGAE (Kipf and Welling 2016), VDGNN (Zhou et al. 2020), DropOut BNN (Gal and Ghahramani 2016), and DeepEnsembles (Lakshminarayanan, Pritzel, and Blundell 2017)).

Implementation details. The unit time step is 1h in Beijing and 2h in NYC datasets. We consider all OD pairs with non-zeros across $\leq T$ time slices to form a set of (unique)

OD pairs. Then, we perform a random 7:1:2 division to create the training, validation, and test sets. During training, we mask the flow values of OD pairs in the validation/test sets to zero. During validation/testing, we interpolate the values of masked OD pairs and compare them with ground truths. Each method’s performance is reported via the averaged scores of 10 runs in terms of MAE, RMSE, and MAPE.

Performance Evaluation

We report the overall performance comparison for OD crowd flow interpolation on Beijing and NYC datasets in Tables 1 and 2, respectively. We can generally observe that VMR-GAE outperforms baselines consistently, especially in rush hours and holidays. To explain this, we find OD flow often exhibit larger volumes in rush-hours and higher uncertainty in holidays, which leads to the most remarkable performance improvements. On the other hand, uncertainty quantification would like to produce conservative outcomes, which may sacrifice the improvement w.r.t. RMSE.

To be specific, we then elaborate the comparison results as follows: 1) Traditional MF model (SVD) cannot fit the nonlinear region relationships in OD crowd flow sufficiently. 2) 2D-GCN is limited by the fixed network structure, thus can only achieve better performances than some traditional methods. 3) NGCF that combines the deep architecture and GNNs into collaborative filtering performs better than 2D-GCN. However, T-NGCF (a variant of NGCF) that captures temporal dependency has little improvement. One possible reason is that the temporal dependency is hard to capture due to the instability of OD crowd flow. 4) VGAE performs similarly to GAE since their representation power is insufficient. 5) VDGNN outperforms VGAE with modeling temporal dependency, while it can hardly beat other uncertainty-aware methods due to the deficiency of inner product decoder. 6) The results of the other two uncertainty-aware methods (which use NGCF or T-NGCF as basis) are similar, DeepEnsembles is the second-best in both two datasets, validating the effectiveness of uncertainty quantification. 7) VMR-GAE outperforms DropOut BNN and DeepEnsembles, which validates the effectiveness of supplementary information for OD crowd flow interpolation.

Effectiveness of uncertainty quantification We randomly selected two OD pairs from the Beijing testing set and examined the outcomes with uncertainty quantification. As depicted in Fig. 4, VMR-GAE can give the corresponding confidence interval to cover the ground truths, such as 12:00-13:00 in the left figure and 18:00-19:00 in the right figure. Also, we can see that the OD flows in these two cases are highly deviated, but VMR-GAE still makes stable interpolation due to the stochastic factors modeling. Specifically, there are many educational institutions in region 132, such that obvious uncertainties can be found after school. In contrast, more commercial, residential and entertainment facilities exist in region 104, which result in high uncertainties at midnight and rush hours. These results verify VMR-GAE can discover the correct uncertainty in OD pairs.

¹<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

Method	Weekday Rush Hour	Weekday Off-Peak	Holiday Rush Hour	Holiday Off-Peak
SVD	53.229/68.947/1.613	49.633/63.378/1.474	54.791/71.092/1.724	51.533/66.121/1.531
GAE*	38.303/52.934/1.546	36.488/50.811/1.426	39.328/53.905/1.582	36.829/50.748/1.444
2D-GCN	42.084/57.740/1.052	42.450/57.622/1.029	43.835/59.663/1.112	43.550/58.830/1.075
NGCF	30.998/42.152/0.838	30.021/41.220/0.766	32.494/44.017/0.863	30.991/42.197/0.807
T-NGCF	35.531/48.205/0.799	34.549/46.580/0.748	36.924/48.959/0.876	35.260/48.091/0.770
VGAE*	43.911/59.595/1.059	42.320/56.818/1.060	44.016/60.635/1.012	42.008/56.976/1.048
VDGNN*	42.327/57.771/1.079	38.588/52.820/0.896	40.481/55.838/1.081	39.980/54.630/0.996
DropOut BNN	31.065/42.305/0.830	30.031/41.213/0.773	32.485/44.035/0.866	31.244/42.479/0.803
DeepEnsembles	29.450/ 40.482/ 0.782	28.537/ 39.726/ 0.715	30.910/ 42.328/ 0.807	29.571/ 40.644/ 0.763
VMR-GAE	27.400/40.455/0.674 -6.96%/-0.01%/-13.8%	26.644/38.673/0.696 -6.63%/-2.65%/-2.66%	25.535/38.501/0.661 -17.4%/-9.04%/-18.1%	26.637/40.115/0.728 -9.92%/-1.30%/-4.59%

Table 1: Performance comparisons in terms of MAE, RMSE, and MAPE on Beijing dataset.

Method	Weekday Rush Hour	Weekday Off-Peak	Holiday Rush Hour	Holiday Off-Peak
SVD	12.056/21.097/1.062	7.723/13.286/0.751	10.220/16.967/0.923	8.887/15.107/0.772
GAE*	10.046/20.683/0.897	7.523/14.629/0.899	9.116/17.194/0.894	8.854/16.756/0.899
2D-GCN	10.009/19.163/0.724	7.490/13.472/0.696	8.507/14.760/0.678	8.787/15.256/0.762
NGCF	9.030/17.665/0.694	6.211/11.572/0.561	7.851/13.709/0.683	7.310/12.927/0.598
T-NGCF	9.677/19.193/0.683	7.627/14.289/0.590	8.467/15.189/0.652	8.211/14.822/0.625
VGAE*	10.786/18.167/0.936	8.176/13.052/0.816	9.503/14.860/0.886	9.215/14.779/0.832
VDGNN*	10.382/18.114/0.690	7.657/12.982/0.619	10.042/17.240/0.657	9.057/15.262/0.627
DropOut BNN	9.649/19.175/0.676	6.544/12.432/0.533	8.484/15.352/0.646	7.739/13.981/0.595
DeepEnsembles	8.358/ 16.281/ 0.624	6.162/ 11.401/ 0.529	7.382/ 13.290/ 0.625	7.116/ 12.872/ 0.579
VMR-GAE	8.152/16.211/0.570 -2.46%/-0.43%/-8.65%	6.092/11.344/0.504 -1.14%/-0.50%/-4.91%	6.932/12.472/0.548 -6.10%/-6.16%/-12.3%	7.091/12.572/0.576 -0.35%/-2.33%/-0.52%

Table 2: Performance comparisons in terms of MAE, RMSE, and MAPE on NYC dataset.

Dataset	Beijing	NYC
Time interval	1 hour	2 hours
#Nodes (#Regions)	213	263
Avg. #edges in OD graph	19949.50	1625.86
Avg. edge weight in OD graph	201.33	8.78
Avg. #edges in context graph	202.79	951.50
Avg. edge weight in context graph	1255.37	1.89

Table 3: Dataset description.

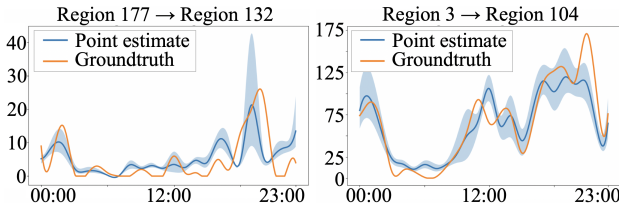


Figure 4: Interpolation results with uncertainty quantification for two OD pairs in Beijing on April 12, 2019.

Ablation Study

Effectiveness of deep decoder To investigate the effect of deep decoder, we construct a simplified model that employs the inner product decoder instead. We show the results w.r.t. RMSE in Table 4. As can be seen, the simplified model performs worse than VMR-GAE, since the deep decoder can effectively model the complex patterns of OD crowd flow.

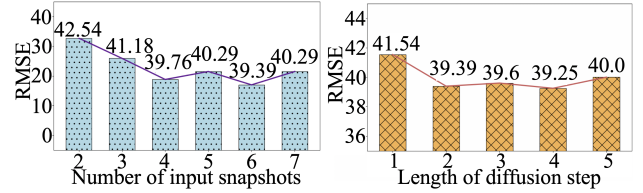


Figure 5: Parameter sensitivity analysis.

Method	Beijing	NYC
w/o deep decoder	45.425	12.556
w/o multi-modal inputs	47.449	12.987
w/o alignment	42.308	11.211
VMR-GAE	39.437	11.137

Table 4: Ablation study given by RMSE. 'w/o deep decoder' is the variant that uses inner product decoder instead. 'w/o multi-modal' does not use context graph. 'w/o alignment' omits distribution alignment in data fusion.

Effectiveness of multi-modal To verify the effect of context information, we compare the variants without contextual inputs and without distribution alignment in Table 4. The RMSE of the variant VMR-GAE w/o context graph decreases a lot, especially on the Beijing dataset. Since the supplementary information used in the NYC dataset (i.e. Green Taxi) is highly sparse, the error increase is smaller. Besides, the lack of distribution alignment also indicates the effectiveness of our assumption in the latent space.

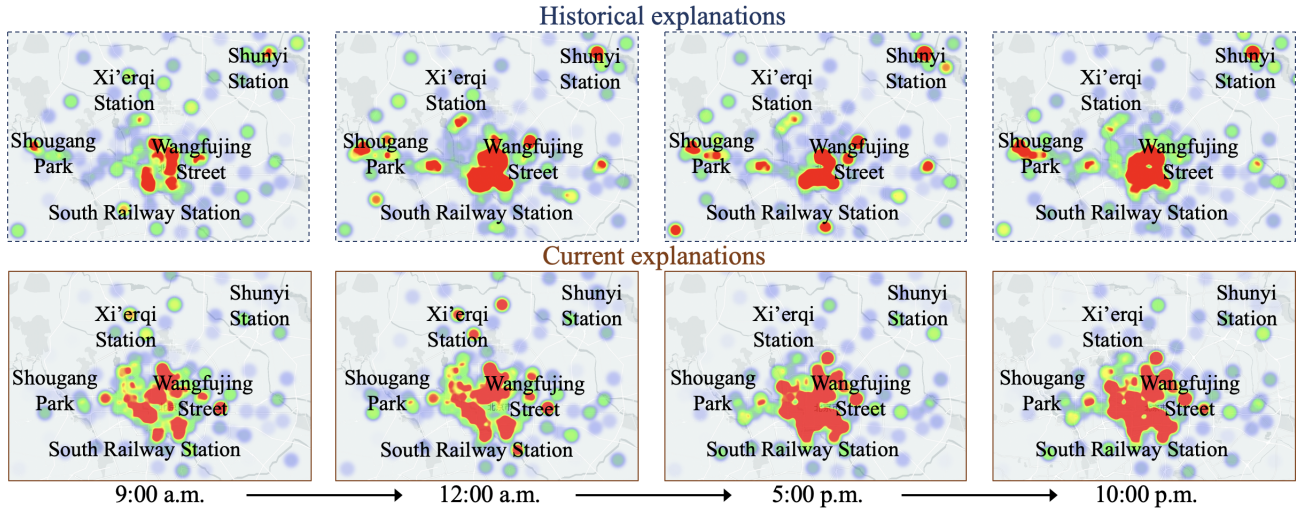


Figure 6: The spatial importance at different times derived from the overall spatiotemporal explanations on April 17, 2019.

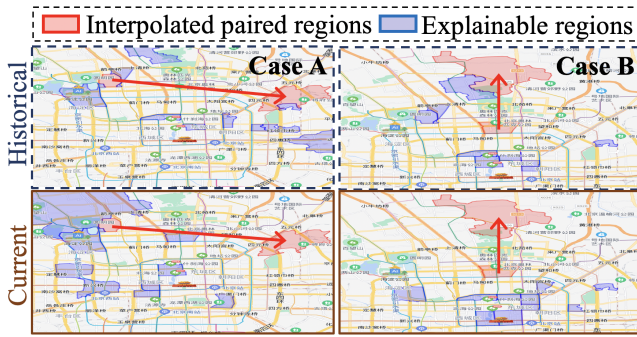


Figure 7: Case study on explanations yielded by UApex. Case A: 9:00 4/13/2019; Case B: 17:00 4/17/2019.

Parameter Sensitivity

In Fig. 5, we report the parameter sensitivity analysis on Beijing dataset w.r.t. two hyperparameters, i.e., the number of input snapshots (controlling the temporal dependency) and the length of diffusion step in the encoder (controlling the spatial dependency). We can observe that both testing errors first decrease along the available information grows up. When the parameters are large enough, performance shows volatility since a larger number of input snapshots or longer diffusion steps will augment the training intricacy. Therefore, we choose the number of input snapshots T as 6 and set the diffusion step to be 2 in this work.

Case Study on OD Explanations

We randomly selected two testing OD pairs in Beijing dataset and visualize explanations accordingly. In case A of Fig. 7, the crowd travels from a college and residential area to a commercial area on the weekend morning. In particular, from historical explanations, we observe that the highlighted regions close to the origin contain many shopping malls, which are similar to the destination. Around the des-

tinuation, some residential areas show importance since they also provide many flow to the destination. In current explanations, the northern neighbor region of the origin is a commercial district that attracts most of the surrounding crowd flow. On the other hand, case B in Fig. 7 shows the crowd flow from a business area to a residential area in the evening rush hour. Many busy downtown regions are highlighted in historical and current explanations, i.e., the leisure places after work can also explain this OD crowd flow.

Exploratory Analysis of Explanations

We summarize the yielded spatiotemporal explanations by counting the occurrence of explainable regions at different times. Basically, the historical explanation reflects the long-term region influence, while the current explanation depicts the instant crowd preference. As shown in Fig. 6, Beijing downtown is always the most popular area, but in the long run, other districts show their importance. Besides, regarding current explanations, the informative regions spread out from morning to sunset, which is in accordance with people's mobility patterns on weekdays. In contrast, the important regions in historical explanations are more scattered, suggesting other developing areas besides the downtown.

Conclusion

In this paper, we proposed UApex for reliable and trustworthy OD crowd flow interpolation, including VMR-GAE and its explainer. Notably, we devised distribution alignment to integrate supplementary modals in VMR-GAE for mitigating the data scarcity. A deep decoder with a Poisson prior was proposed to reveal the complex relationships between OD pairs. Moreover, we developed an uncertainty-aware explainer to effectively explain VMR-GAE from the spatiotemporal topology perspective with the Shapley value. Extensive empirical studies on two real-world datasets validate the superiority and effectiveness of UApex.

Acknowledgments

This research is supported in part by the National Natural Science Foundation of China (Grant No.62072235).

References

- Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J. M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; and Herrera, F. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99: 101805.
- Atwood, J.; and Towsley, D. 2016. Diffusion-convolutional neural networks. In *Advances in neural information processing systems*, 1993–2001.
- Bonner, S.; Atapour-Abarghouei, A.; Jackson, P. T.; Brennan, J.; Kureshi, I.; Theodoropoulos, G.; McGough, A. S.; and Obara, B. 2019. Temporal neighbourhood aggregation: Predicting future links in temporal graphs via recurrent variational graph convolutions. In *2019 IEEE International Conference on Big Data (Big Data)*, 5336–5345. IEEE.
- Cheng, X.; Rao, Z.; Chen, Y.; and Zhang, Q. 2020. Explaining knowledge distillation by quantifying the knowledge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12925–12935.
- Deng, D.; Shahabi, C.; Demiryurek, U.; Zhu, L.; Yu, R.; and Liu, Y. 2016. Latent space model for road networks to predict time-varying traffic. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1525–1534.
- Feng, S.; Ke, J.; Yang, H.; and Ye, J. 2021. A Multi-Task Matrix Factorized Graph Neural Network for Co-Prediction of Zone-Based and OD-Based Ride-Hailing Demand. *IEEE Transactions on Intelligent Transportation Systems*.
- Gal, Y.; and Ghahramani, Z. 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, 1050–1059. PMLR.
- Gu, J.; Zhou, Q.; Yang, J.; Liu, Y.; Zhuang, F.; Zhao, Y.; and Xiong, H. 2020. Exploiting interpretable patterns for flow prediction in dockless bike sharing systems. *IEEE Transactions on Knowledge and Data Engineering*.
- Hajiramezanali, E.; Hasanzadeh, A.; Narayanan, K.; Duffield, N.; Zhou, M.; and Qian, X. 2019. Variational Graph Recurrent Neural Networks. *Advances in Neural Information Processing Systems*, 32: 10701–10711.
- Huang, B.; Ruan, K.; Yu, W.; Xiao, J.; Xie, R.; and Huang, J. 2023. ODformer: Spatial-temporal transformers for long sequence Origin-Destination matrix forecasting against cross application scenario. *Expert Systems with Applications*, 222: 119835.
- Huang, Y.; Du, C.; Xue, Z.; Chen, X.; Zhao, H.; and Huang, L. 2021. What Makes Multimodal Learning Better than Single (Provably). *arXiv preprint arXiv:2106.04538*.
- Huang, Z.; Zhang, W.; Wang, D.; and Yin, Y. 2022. A GAN framework-based dynamic multi-graph convolutional network for origin-destination-based ride-hailing demand prediction. *Information Sciences*, 601: 129–146.
- Jeong, I.-J.; and Park, D. 2021. Stochastic programming approach for static origin-destination matrix reconstruction problem. *Computers & Industrial Engineering*, 157: 107373.
- Kamath, U.; Liu, J.; and Whitaker, J. 2019. Transfer learning: Domain adaptation. In *Deep learning for NLP and speech recognition*, 495–535. Springer.
- Ke, J.; Qin, X.; Yang, H.; Zheng, Z.; Zhu, Z.; and Ye, J. 2021. Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. *Transportation Research Part C: Emerging Technologies*, 122: 102858.
- Kipf, T. N.; and Welling, M. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*.
- Kong, L.; Sun, J.; and Zhang, C. 2020. SDE-Net: Equipping Deep Neural Networks with Uncertainty Estimates. In *International Conference on Machine Learning*, 5405–5415. PMLR.
- Kuhn, H. W.; and Tucker, A. W. 1953. *Contributions to the Theory of Games*, volume 2. Princeton University Press.
- Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *Advances in neural information processing systems*, 30.
- Li, B.; Qi, P.; Liu, B.; Di, S.; Liu, J.; Pei, J.; Yi, J.; and Zhou, B. 2023. Trustworthy AI: From Principles to Practices. *ACM Comput. Surv.*, 55(9).
- Li, C.; Bai, L.; Liu, W.; Yao, L.; and Waller, S. T. 2022. Graph Neural Network for Robust Public Transit Demand Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 23(5): 4086–4098.
- Li, J.; Chen, X.; Hovy, E.; and Jurafsky, D. 2016. Visualizing and Understanding Neural Models in NLP. In *Proceedings of NAACL-HLT*, 681–691.
- Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *International Conference on Learning Representations*.
- Lin, Y.; Wan, H.; Hu, J.; Guo, S.; Yang, B.; Lin, Y.; and Jensen, C. S. 2023. Origin-Destination Travel Time Oracle for Map-based Services. *Proceedings of the ACM on Management of Data*, 1(3): 1–27.
- Long, M.; Wang, J.; Cao, Y.; Sun, J.; and Philip, S. Y. 2016. Deep learning of transferable representation for scalable domain adaptation. *IEEE Transactions on Knowledge and Data Engineering*, 28(8): 2027–2040.
- Lundberg, S. M.; and Lee, S.-I. 2017. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, 4768–4777.
- Luo, D.; Cheng, W.; Xu, D.; Yu, W.; Zong, B.; Chen, H.; and Zhang, X. 2020. Parameterized explainer for graph neural network. *arXiv preprint arXiv:2011.04573*.
- Ma, X.; Sun, P.; and Qin, G. 2017. Nonnegative matrix factorization algorithms for link prediction in temporal networks using graph communicability. *Pattern Recognition*, 71: 361–374.

- Miao, H.; Shen, J.; Cao, J.; Xia, J.; and Wang, S. 2023. MBA-STNet: Bayes-Enhanced Discriminative Multi-Task Learning for Flow Prediction. *IEEE Transactions on Knowledge and Data Engineering*, 35(7): 7164–7177.
- Pitombeira-Neto, A. R.; Loureiro, C. F. G.; and Carvalho, L. E. 2020. A dynamic hierarchical bayesian model for the estimation of day-to-day origin-destination flows in transportation networks. *Networks and Spatial Economics*, 20(2): 499–527.
- Rong, C.; Li, T.; Feng, J.; Yang, H.; Geng, L.; and Li, Y. 2021. Inferring Origin-Destination Flows from Population Distribution. *IEEE Transactions on Knowledge and Data Engineering*.
- Ros-Roca, X.; Montero, L.; Barceló, J.; Nökel, K.; and Gentile, G. 2022. A practical approach to assignment-free Dynamic Origin–Destination Matrix Estimation problem. *Transportation Research Part C: Emerging Technologies*, 134: 103477.
- Shi, H.; Yao, Q.; Guo, Q.; Li, Y.; Zhang, L.; Ye, J.; Li, Y.; and Liu, Y. 2020. Predicting origin-destination flow via multi-perspective graph convolutional network. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, 1818–1821. IEEE.
- Wang, B.; Lu, J.; Yan, Z.; Luo, H.; Li, T.; Zheng, Y.; and Zhang, G. 2019a. Deep uncertainty quantification: A machine learning approach for weather forecasting. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2087–2095.
- Wang, Q.; Wang, S.; Zhuang, D.; Koutsopoulos, H.; and Zhao, J. 2023a. Uncertainty Quantification of Spatiotemporal Travel Demand with Probabilistic Graph Neural Networks. *arXiv:2303.04040*.
- Wang, X.; He, X.; Wang, M.; Feng, F.; and Chua, T.-S. 2019b. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, 165–174.
- Wang, Y.; Yin, H.; Chen, H.; Wo, T.; Xu, J.; and Zheng, K. 2019c. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 1227–1235.
- Wang, Z.; Zhuang, D.; Li, Y.; Zhao, J.; and Sun, P. 2023b. ST-GIN: An Uncertainty Quantification Approach in Traffic Data Imputation with Spatio-temporal Graph Attention and Bidirectional Recurrent United Neural Networks. *arXiv:2305.06480*.
- Xie, N.; Ras, G.; van Gerven, M.; and Doran, D. 2020. Explainable deep learning: A field guide for the uninitiated. *arXiv preprint arXiv:2004.14545*.
- Xu, Y.; Lyu, Y.; Xiong, G.; Wang, S.; Wu, W.; Cui, H.; and Luo, J. 2023. Adaptive Feature Fusion Networks for Origin-Destination Passenger Flow Prediction in Metro Systems. *IEEE Transactions on Intelligent Transportation Systems*, 24(5): 5296–5312.
- Ying, R.; Bourgeois, D.; You, J.; Zitnik, M.; and Leskovec, J. 2019. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32: 9240.
- Yuan, H.; Tang, J.; Hu, X.; and Ji, S. 2020. Xgnn: Towards model-level explanations of graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 430–438.
- Yuan, H.; Yu, H.; Wang, J.; Li, K.; and Ji, S. 2021. On explainability of graph neural networks via subgraph explorations. In *international conference on machine learning*.
- Zhang, H.; Cheng, X.; Chen, Y.; and Zhang, Q. 2020. Game-theoretic interactions of different orders. *arXiv preprint arXiv:2010.14978*.
- Zhang, H.; Xie, Y.; Zheng, L.; Zhang, D.; and Zhang, Q. 2021a. Interpreting Multivariate Shapley Interactions in DNNs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 10877–10886.
- Zhang, J.; Che, H.; Chen, F.; Ma, W.; and He, Z. 2021b. Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method. *Transportation Research Part C: Emerging Technologies*, 124: 102928.
- Zhang, J.; Zheng, Y.; Sun, J.; and Qi, D. 2019. Flow prediction in spatio-temporal networks based on multitask deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 32(3): 468–478.
- Zhou, F.; Yang, Q.; Zhong, T.; Chen, D.; and Zhang, N. 2020. Variational graph neural networks for road traffic prediction in intelligent transportation systems. *IEEE Transactions on Industrial Informatics*, 17(4): 2802–2812.
- Zhou, Z.; Wang, Y.; Xie, X.; Qiao, L.; and Li, Y. 2021. STU-aNet: Understanding uncertainty in spatiotemporal collective human mobility. In *Proceedings of the Web Conference 2021*, 1868–1879.