Socially-Equitable Interactive Graph Information Fusion-based Prediction for Urban Dockless E-Scooter Sharing

Suining He
University of Connecticut
suining.he@uconn.edu

Kang G. Shin
University of Michigan–Ann Arbor
kgshin@umich.edu

ABSTRACT

Urban dockless e-scooter sharing (DES) has become a popular Web-of-Things (WoT) service and widely adopted globally. Despite its early commercial success, conventional mobility demand and supply prediction based on machine learning and subsequent redistribution may favor advantaged socio-economic communities and tourist regions, at the expense of reducing mobility accessibility and resource allocation for historically disadvantaged communities. To address this unfairness, we propose a socially-Equitable Interactive Graph information fusion-based mobility flow prediction system for Dockless E-scooter Sharing (**EIGDES**). By considering city regions as nodes connected by trips, EIGDES learns and captures the complex interactions across spatial and temporal graph features through a novel interactive graph information dissemination and fusion structure. We further design a novel model learning objective with metrics that capture both the mobility distributions and the socio-economic factors, ensuring spatial fairness in the communities' resource accessibility and their experienced DES prediction accuracy. Through its integration with the optimization regularizer, EIGDES *jointly* learns the DES flow patterns and socio-economic factors, and returns socially-equitable flow predictions. Our in-depth experimental study upon more than 2,122,270 DES trips from three metropolitan cities in North America has demonstrated EIGDES's effectiveness in accurate prediction of DES flow patterns with substantial reduction of mobility unfairness.

CCS CONCEPTS

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

KEYWORDS

Dockless escooter sharing, interactive graph information fusion, socially-equitable flow prediction

ACM Reference Format:
Suining He and Kang G. Shin. 2022. Socially-Equitable Interactive Graph Information Fusion-based Prediction for Urban Dockless E-Scooter Sharing . In *Proceedings of the ACM Web Conference 2022 (WWW '22), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3485447.3512145

WWW '22, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9096-5/22/04…$15.00
https://doi.org/10.1145/3485447.3512145

1 INTRODUCTION

Driven by its on-demand mobility convenience and last-mile connectivity, the dockless electric (e)-scooter sharing (DES) has emerged as a ubiquitous Web-of-Things (WoT) service [22, 40], and proliferated within and across neighborhoods of many smart cities. Due to dockless parking and easy rental using smartphone apps, the distribution of e-scooters can change significantly over time. Consequently, the operation utility and service quality of DES systems largely rely on the balance between demands and supplies, and proper allocation of such mobility resources. Besides, additional infrastructures — such as side-walk geo-fences, e-scooter charging docks, and other convenience support — may be allocated to regions with high drop-offs or arrivals of e-scooters.

These mobility distribution and coordination operations are often empowered by accurate pick-up/drop-off *flow prediction* based on advanced machine learning and data analytics [19, 20]. In practice, the fulfilled historical demand likely originates from tourism regions with already robust mobility options or communities with financial affordability. Hence, the existing algorithms that redistribute the mobility resources may favor such demand with bias, and discriminate the communities with low-income or historically underrepresented attributes [19, 46]. This is mainly due to the nature that conventional learning algorithms seek to *fit* the fulfilled historical demands as close as possible for the single-minded "accuracy" measures.

Nevertheless, the low fulfilled ridership in disadvantaged communities does not necessarily indicate the low mobility demand [19]. Such biased/overfitted predictions and subsequent profit-driven matching of supplies towards demands could trade off the *accessibility* of all citizens, particularly those historically disadvantaged groups, for mobility services, which accounts for urban transportation *fairness* or *equity*. The resultant spatial mobility resource imbalance could jeopardize the long-term sustainability of urban community development [55]. Since the recent COVID-19 pandemics have made micro-mobility options [20] like DES the limited and affordable alternatives to public transportation (particularly for families without private cars), such a spatial mismatch between on-demand mobility options and communities has exacerbated with the disproportionate effect of COVID-19 cases in underrepresented, disadvantaged, and low-income communities as well as long-lasting social inequality and exclusion [31].

To address the above civic transportation disparity concerns, we focus on the emerging DES systems to design a novel *socially-equitable* flow prediction framework with mobility analysis and modeling. In particular, the framework is to predict the fairness-aware DES demands (pick-ups) and supplies (drop-offs) to enable *equal* spatial mobility resource distribution, accessibility, and opportunity across diverse social groups and communities. Despite

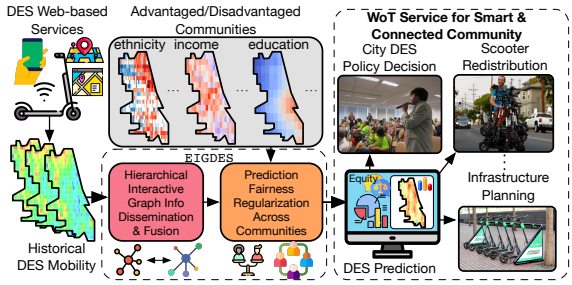

Fig. 1. Illustration of socially-equitable DES mobility modeling.

the pioneering studies in [14, 66], there remain two major issues that need to be addressed for socially-equitable DES deployment.

(1) While various spatio-temporal mobility studies for DES systems have been conducted recently [14, 22, 66], few of them have comprehensively studied the mobility *fairness* of DES deployment in metropolitan cities. The dockless, last-mile, and easy-manuevering features of DES largely result in a broad spectrum (near/distant ranges and short/long terms) of spatial and temporal connectivities across city regions where different communities are located. Such dynamic DES distributions make the accommodation of different communities' mobility demand and accessibility highly challenging. There exist pressing needs for integrating the potential disparities in allocated scooter resources and accommodating differences in model performance across communities, which, however, remain largely unexplored.

(2) The complex DES mobility patterns result from the complex *interactions* between the scooter riders and the urban environments. Riders' mobility preferences and routines largely interact with the spatial adjacency, temporal closeness, and other heterogeneous and multi-level correlations of the city regions. Characterizing such an interaction with complex dimensions, heterogeneous representations, and varying degrees is essential but challenging for the DES mobility modeling.

To realize a socially-equitable WoT-based DES service for *smart and connected community*, we propose a socially-**E**quitable **I**nteractive **G**raph fusion framework for **D**ockless **E**-scooter **S**haring (EIGDES) systems. As illustrated in Fig. 1, via modeling the spatio-temporal mobility connectivities across city regions as graphs, we design the spatio-temporal DES flow (pick-ups/drop-offs) prediction framework based on *interactive graph information dissemination and fusion*, and incorporate *mobility fairness metrics* across communities as regularization upon the core model learning. EIGDES accounts for essential socioeconomic attributes like social ethnicity, income, and education, and regularizes the DES flow predictions for fair distributions across the advantaged and disadvantaged groups.

In summary, this paper makes the following contributions:

(a) **Hierarchical Interactive Graph Information Dissemination and Fusion Designs for DES Flow Prediction**: Considering the city regions as nodes connected by graph edges that represent DES trips and multiple spatio-temporal correlations, EIGDES disseminates the embedded information of the graphs to differentiate their importance. Then, the hierarchical structures augmented with interactive attentions actively capture the dynamic *interactions* across these graphs, and accurately forecast the DES flows (pick-ups/drop-offs) at different city regions.

(b) **Socially-Equitable DES Flow Prediction Studies**: To mitigate the mobility fairness gap, we have designed and integrated a novel mobility fairness regularizer within the learning process of EIGDES, accommodating various social equity perspectives in DES mobility modeling. In particular, we have studied mobility fairness metrics regarding resource non-parity and prediction performance difference. These metrics will help adapt the DES predictions under different mobility fairness evaluation and deployment scenarios.

(c) **Extensive Experimental Evaluations**: We have conducted extensive experimental studies based on real-world datasets from three DES pilot programs, *i.e.*, Chicago IL and Minneapolis MN in US, and Edmonton in Canada, which consist of total 2,122,270 trips, as well as related socioeconomic and urban data sources. The results demonstrate effectiveness, accuracy, and fairness of EIGDES's model, which is important for the trending DES deployment in benefiting a broader spectrum of urban communities. Our approach will be the key enabler for the policy decision, scooter redistribution, infrastructure planning, and many other services as motivated in Fig. 1.

• **Relevance to Web**: Our accurate prediction and socially-equitable regularization in this paper will enhance the understanding of proactiveness and inclusiveness of emerging WoT services for the essential use-cases targeted by the World Wide Web Consortium (W3C) WoT [10] that include *Transportation & Logistics* and *Smart Cities*. The interactive graph fusion we developed will be broadly applicable for web-based pervasive computing in transport industry [22, 43], urban planning [63], and digital governance [42], as well as extensible to the web data integration and mining [28].

2 RELATED WORK

We briefly review the related work in the following two categories.

• **Spatio-Temporal Mobility Data Mining**: Spatio-temporal mobility modeling has paved new ways for many urban and WoT computing applications [23, 35, 45]. Hulot *et al.* [27] studied the station-level demand prediction for system rebalancing. Ma *et al.* [39] studied the convolutional neural network to process and predict the traffic speed. Zhang *et al.* [65] proposed a residual neural network for predictions of urban flows. Liang *et al.* [35] developed a spatio-temporal relation network for fine-grained flow prediction. Besides predictive modeling, researchers have started to derive insights for societal impacts introduced by the new mobility platforms [24, 36]. Unlike these prior studies, we focus on predictive flow modeling for socially-equitable transportation planning, particularly upon the emerging dockless e-scooter sharing (DES) deployment. Furthermore, instead of aggregating the graphs [15, 22, 23, 34], we have designed a novel graph information fusion differentiating and learning the hierarchical *interactions* between various spatial and temporal graphs across the urban regions, hence achieving accurate predictions upon complex DES flows.

• **Socially-Equitable Modeling**: How to identify and remove biases against sensitive groups has become a key issue for machine learning and web communities, especially for many essential but increasingly intrusive urban applications. Yao *et al.* studied fairness for collaborative filtering [60]. Zemel *et al.* proposed fair representation encoding for group and individual fairness [61]. Hossain *et al.* studied fair classifiers via the economic fairness notions [26].

While fairness-aware *classification* has been studied extensively, fewer studies explored the *regression* designs. A pioneering study conducted by Kamishima *et al.* [29] regularized and controlled the trade-off between accuracy and fairness in logistic regression. Berk *et al.* [11] studied a convex family of fairness regularizers. Inspired by the aforementioned studies, we design fairness regularization mechanisms for DES systems, and conduct comprehensive studies to enhance the fairness of its regressed spatio-temporal predictions.

Realizing the socially-equitable transportation planning has become essential with the city expansion and growing social disparities. Hosford *et al.* [25] investigated the equity of access to bike sharing at multiple cities in Canada. Ge *et al.* [18] studied the racial and gender discrimination in the expanding transportation network companies. Yan *et al.* [57] explored the region-based and individual-based fairness for car-sharing and bike-sharing. Despite the aforementioned studies, few of them have considered fairness modeling within the emerging *e-scooter deployment*, and provided comprehensive data-driven and fairness-aware mobility analysis upon the prediction framework [17, 20]. Different from [29, 57, 58], EIGDES fills this essential gap, and provides important deployment insights for the DES planning related to resource non-parity and prediction performance difference.

3 DATA, CONCEPTS, & PROBLEM

3.1 Datasets Studied

We have collected the following datasets for our data analysis and social-equitable mobility modeling:

- *DES Trips*: We have collected the DES trips from three metropolitan cities in North America: Chicago IL and Minneapolis MN in US, and Edmonton, Canada. Bird, Lime, and Spin are the major DES service providers in these pilot programs. Specifically, the DES dataset from Chicago [3] consists of 604,146 trips which span from June 15, 2019 to October 15, 2019, while the one from Minneapolis [6] contains a total of 1,004,920 trips collected during two periods: July 10 to November 30, 2018 and May 1 to November 30, 2019. The DES trip dataset from Edmonton [4] is comprised of a total 513,204 trips retrieved from the E-Scooter Share API from August 28 to October 30, 2020. Each trip record contains the pick-up/drop-off GPS coordinates and timestamps. For the Chicago, Minneapolis, and Edmonton datasets, trips which last for over 7 hours, less than 0 miles, or greater than 24 miles have been removed before further processing.
- *Socioeconomic Datasets*: We have collected the socioeconomic data (census tracts) related to sensitive attributes, including population, social ethnicity, education, and income (poverty) level, for Chicago and Minneapolis from the US Census Database [8], and Edmonton from the Canada Federal Census Tract in Alberta [2]. We note that all socioeconomic datasets used in our studies are from open data portals and hence no institutional review board (IRB) review is required.

For each city, we have collected the meteorological and weather conditions for the external data analysis. Specifically, we collect the weather data consisting of 11 categories, *i.e.*, temperature, precipitation, wind speed, and other 8 weather conditions from the Weather Archive website [9]. Furthermore, we have collected the *city map* and *points-of-interests* (POIs) from the OpenStreetMap (OSM) [7]. In summary, from 13 POI categories, we have collected 15,026 POIs from Chicago, 2,204 POIs from Minneapolis, and 4,102 POIs from Edmonton. We also collected the *city events* to characterize the urban momentum from the Eventful website [5]. In particular, we have collected 767, 2,686, and 3,250 urban events in 12 categories from Chicago, Minneapolis, and Edmonton within the studied time scope. Detailed categories of weather conditions, POIs, and city events are listed in Table 3 (see Appendix A).

3.2 Important Concepts & Problem Definition

We further define the important concepts in EIGDES as follows.

• **Spatio-Temporal DES Mobility Processing**: We first briefly discuss how we process the map and time information in EIGDES.

For ease of spatial modeling and computational efficiency (similar to [59, 65]), we discretize the entire city map into $N = H \times W$ rectangular grids or regions, where H (W) is the number of grids/regions in the latitudinal (longitudinal) direction. Given the above discretization, we denote the set of all city regions as $\mathbf{G} = \{\mathbf{g}_i,\ i \in \{1, \ldots, N\}\}$. Similarly, we discretize the time domain into fixed intervals (say, 30 or 60min), where each time interval is indexed by k.

Then, by considering each city region as a node, we can represent the entire DES network as a network graph $\mathcal{G} = (\mathbf{V}, \{\mathbf{A}\})$, where $\{\mathbf{A}\}$ is a set of spatial and temporal graph adjacency matrices representing the correlations among the regions. Note that each of the DES trips contains the pick-up/drop-off GPS coordinates, from which we could infer the start/destination regions. Based on the spatial discretization, the DES flows refer to the aggregate numbers of *outgoing* and *incoming* DES trips at each grid, where we characterize the flows via the numbers of *pick-ups* (out flows) and *drop-offs* (in flows). Let $I_i^{(k)}$ and $O_i^{(k)}$ be the in and out flows in interval k at the city region $\mathbf{g}_i$. We then have the flows at time interval k at all the N city regions as $\mathbf{Y}^{(k)} \in \mathbb{R}^{N \times 2}$, whose i-th row is $[I_i^{(k)}, O_i^{(k)}]$.

• **Socioeconomic Notations**: Based on the census (*e.g.*, US Census) and previous social fairness studies [41, 51, 55], for each of the sensitive socioeconomic attributes α, including social ethnicity (Caucasian or not), education (college-educated or not), and income level (above poverty line or not), we consider a binary group feature setting, and denote the resultant *advantaged* and *disadvantaged* communities, respectively, as $\mathbf{c}_\alpha^+$ and $\mathbf{c}_\alpha^-$.

In practice, as the census tracts are different from the map discretization, we find the socioeconomic attributes for each city grid by allocating those in the census tracts proportionally based on the distance between centers of the grids and the census tracts. For each city region $\mathbf{g}_i$ ($i \in \{1, \ldots, N\}$), we let P_i be the percentage of population there, *i.e.*, the population of $\mathbf{g}_i$ divided by the entire city's population. For fine-grained characterization (Sec. 5.2), we further let $\omega_{i,\alpha}^+$ ($\omega_{i,\alpha}^-$) be the respective percentage of populations within $\mathbf{g}_i$ belonging to the advantaged (disadvantaged) groups of attribute α, *i.e.*, the populations of advantaged (disadvantaged) groups divided by the total populations P_i in $\mathbf{g}_i$.

• **Problem Definition**: Given the historical DES flows $\{\mathbf{Y}^{(k')}\}$ ($k' \in \{k - W, \ldots, k - 1\}$) of the recent W intervals at the N city regions, the DES network graph $\mathcal{G}$ (including POI and map information), and external factors $\mathbf{E}$ (say, meteorological factors and urban events), we aim to design a function $f(\{\mathbf{Y}^{(k')}\}, \mathcal{G}, \mathbf{E}, \{\mathbf{c}_\alpha^+, \mathbf{c}_\alpha^-\})$,

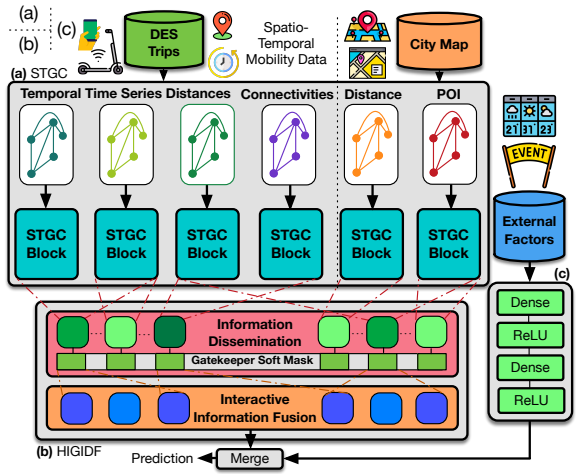

Fig. 2. Prediction framework overview of EIGDES.

which jointly predicts the DES flow distributions (pick-ups/drop-offs or demands/supplies) of all city regions at the next time interval k, *i.e.*, $\widehat{\mathbf{Y}}^{(k)} \in \mathbb{R}^{N\times 2}$, and enables socially-equitable accessibility and resource allocation for communities of both $\mathbf{c}_\alpha^+$ and $\mathbf{c}_\alpha^-$.

4 INTERACTIVE GRAPH INFORMATION FUSION FOR PREDICTION

4.1 Prediction Framework Overview

We first overview the core prediction model framework in Fig. 2, which consists of the following three major designs.

(a) *Spatio-Temporal Graph Construction* (STGC): We first process the adjacencies of the city regions in terms of spatial distances, POI similarities, and multi-level temporal time-series correlations in order to construct multiple spatial and temporal graphs for EIGDES learning. Given above graphs, we further construct spatio-temporal graph convolution (STGC) blocks to capture the spatio-temporal correlations and obtain the graph embeddings.

(b) *Hierarchical Interactive Graph Information Dissemination & Fusion* (HIGIDF): Given the graph embeddings, EIGDES interactively learns the graph information through a novel hierarchical dissemination and fusion mechanism. Specifically, we have designed a gatekeeper soft mask upon the graph embeddings from different spatio-temporal graphs to first control and differentiate the information dissemination. Through the importance differentiation, EIGDES then correlates the graph embeddings that are more important for the subsequent information fusion, yielding accurate predictions of DES flows.

(c) *Integrating External Factors*: Besides STGC and HIGIDF, we further incorporate the external factors (weather conditions and events), and the result will be merged with the output from HIGIDF for the final DES flow predictions.

4.2 Spatio-Temporal Graph Construction

We construct the spatio-temporal graphs as follows.

• **Spatial Proximity Φ**: According to the first law of geography [49], we consider the grids which are closer to each other are more likely correlated in the DES flows. Let $\text{dist}(i,j)$ be the geographic distance (in km) between the city regions i and j, and we define the spatial proximity between them as

$$\Phi_{(i,j)} \triangleq (1+\text{dist}(i,j))^{-1}, \tag{1}$$

which implies higher correlations across the neighboring regions. Then, we obtain the proximity graph adjacency matrix $\Phi \in \mathbb{R}^{N\times N}$.

• **Spatial POI Correlations Ω**: For each region $\mathbf{g}_i$, we find the POI vector $\mathbf{v}_i$, each dimension of which represents the number of POIs belonging to a certain category (say, "residential" in Table 3 in Appendix A). We further define the POI similarities between two city regions, $\mathbf{g}_i$ and $\mathbf{g}_j$, based on the cosine similarity, *i.e.*,

$$\Omega_{(i,j)} \triangleq \left(\mathbf{v}_i \cdot \mathbf{v}_j\right) / \left(|\mathbf{v}_i| \cdot |\mathbf{v}_j|\right). \tag{2}$$

Then, we obtain the POI adjacency matrix $\Omega \in \mathbb{R}^{N\times N}$ characterzing the POI similarity across city regions.

• **Region-to-Region Connectivities Λ**: The DES riders usually exhibit frequent travel patterns among certain pairs of city regions due to the riders' commute routes and ride preferences. For instance, we show in Fig. 11 (see Appendix A) the transitions across the selected top 5% regions with the largest total in and out flows in the cities of Chicago and Minneapolis, where the warmer colors indicate the stronger region-to-region connectivities. Inspired by these, we further take into account the region-to-region connectivities in characterizing the correlations of city regions.

Specifically, let $T_{(i,j)}^{(k)}$ be the number of transitions from $\mathbf{g}_i$ to $\mathbf{g}_j$ in the time interval k, and we obtain the vector of transitions from a region i to all N regions as $\vec{\mathbf{U}}_i^{(k)} = [T_{(i,1)}^{(k)}, \ldots, T_{(i,N)}^{(k)}]$. Based on this, we leverage the dot product between $\vec{\mathbf{U}}_i^{(k)}$ and $\vec{\mathbf{U}}_j^{(k)}$ to characterize their similarities in region-to-region connectivities, and transform it to the range of $[0,1]$ based on a sigmoid function, *i.e.*,

$$\Lambda_{i,j} = \left(1+\exp\left(-\vec{\mathbf{U}}_i^{(k)} \cdot \vec{\mathbf{U}}_j^{(k)}\right)\right)^{-1}. \tag{3}$$

In other words, the more similar transition trends from two regions i and j to other regions, the higher $\Lambda_{i,j}$ in region-to-region connectivities. Then, we obtain the adjacency matrix $\mathbf{\Lambda} \in \mathbb{R}^{N\times N}$ representing region connectivities.

• **Multi-Level Time-series Correlations Ψ**: Besides the spatial proximity, POI correlations, and connectivities, we further take into account the multi-level temporal dependency across the city regions. Since the DES flows at grids are time series, one might consider the typical method based on dynamic time warping (DTW) [12, 34] to measure their mutual similarity. However, the conventional DTW for time series similarity calculation can be computationally expensive in practice. Therefore, we propose a novel temporal representation of the graph based on the LB_Keogh bound (*LBK*) [30], which can efficiently return a *lower-bound* of the DTW distance between two temporal sequences in linear time (illustrated in Alg. 1). This way, we can facilitate the temporal correlation calculation.

Specifically, for each region i, let $\mathbf{O}_i^{(k)}[W'] = \left[O_i^{(k-W')}, \ldots, O_i^{(k)}\right]$ and $\mathbf{I}_i^{(k)}[W'] = \left[I_i^{(k-W')}, \ldots, I_i^{(k)}\right]$ be the time series of out and in flows in the time window of W'. When comparing two out flows $\mathbf{O}_i^{(k)}[W']$ and $\mathbf{O}_j^{(k)}[W']$ (similarly for in flows $\mathbf{I}_i^{(k)}[W']$ and $\mathbf{I}_j^{(k)}[W']$) of grids i and j, *LBK* takes in the global path constraints,

$$\epsilon_j - Z \le \epsilon_i \le \epsilon_j + Z, \tag{4}$$

where ϵ_i and ϵ_j are the indices of the warping path *w.r.t.* the two time series, and $Z > 0$ is a predefined path constraint. Then, we find the upper and lower values, uv and lv, along the warping path (Lines 3 to 4). Afterwards, we obtain a lower-bounding measure, denoted as lb_diff, of the accumulated difference (Lines 5 to 9).

Algorithm 1: Calculation of $LBK(\mathbf{O}_i^{(k)}[W'], \mathbf{O}_j^{(k)}[W'])$.

Input: $\mathbf{O}_i^{(k)}[W']$, $\mathbf{O}_j^{(k)}[W']$: two time series (similarly for $\mathbf{I}_i^{(k)}[W']$, $\mathbf{I}_j^{(k)}[W']$); Z: path constraint.
Output: *LBK* difference of the two time series.

```
lb_diff←0;  O_i←O_i^(k)[W'];  O_j←O_j^(k)[W'];
for ϵ, o in enumerate(O_i) do
    uv←max(O_j[(ϵ-Z if ϵ-Z ≥ 0 else 0): (ϵ+Z)]); /* upper val */
    lv←min(O_j[(ϵ-Z if ϵ-Z ≥ 0 else 0): (ϵ+Z)]); /* lower val */
    if o > uv then
        lb_diff←lb_diff + (o-uv)^2;
        else if o < lv then
            lb_diff←lb_diff + (o-lv)^2;
        end
    end
end
return lb_diff;
```

We then form the temporal correlations across grids i and j, denoted as $\Psi(i,j)$ ($\Psi(i,j) \in (0,1)$), accounting for the similarities respectively between out and in flows of grids i and j, *i.e.*,

$$\Psi_{(i,j)}^{(k)}[W'] \triangleq \left(1 + LBK_{(i,j)}^{(k)}[W']\right)^{-1}, \tag{5}$$

where $LBK_{(i,j)}^{(k)}[W']$ is given by sum of *LBK* for in and out flows, *i.e.*, $LBK(\mathbf{O}_i^{(k)}[W'], \mathbf{O}_j^{(k)}[W']) + LBK(\mathbf{I}_i^{(k)}[W'], \mathbf{I}_j^{(k)}[W'])$. In other words, the smaller the bounds, the more similar two grids are in terms of temporal flows. Based on Eq. (5), we obtain the temporal correlations $\mathbf{\Psi}^{(k)}[W'] \in \mathbb{R}^{N \times N}$.

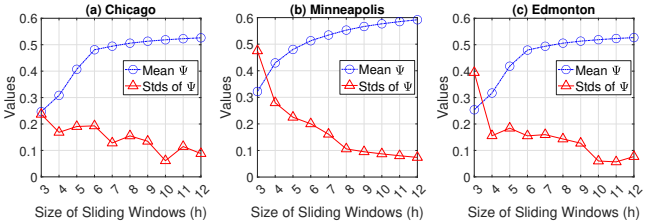

Fig. 3. Multi-level Ψ in (a) Chicago, (b) Minneapolis, & (c) Edmonton.

We illustrate in Fig. 3 the means and standard deviations (Stds) of $\Psi^{(k)}[W']$ given different values of W' in (a) Chicago, (b) Minneapolis, and (c) Edmonton. We can observe the lower and more dynamic short-term DES flow correlations given smaller W', as well as the higher and stabler long-term correlations with larger W'. In order to accommodate multiple different levels of temporal closeness between the city regions, we adopt the multi-level *LBK* bounds given *short*, *medium*, and *long* historical time windows (denoted as $W^{(\mathrm{S})}$, $W^{(\mathrm{M})}$, and $W^{(\mathrm{L})}$). Based on Fig. 3, we set the numbers of historical time intervals ($W^{(\mathrm{S})}, W^{(\mathrm{M})}, W^{(\mathrm{L})}$) as (3, 6, 12) for Chicago, (4, 8, 12) for Minneapolis, and (4, 6, 10) for Edmonton. Our experimental studies also corroborate the comparable accuracy of *LBK* with DTW ((vi) in Fig. 7).

4.3 Spatio-Temporal Graph Convolution

We adopt the graph convolutions in order to capture the DES correlations and support the message passing across different nodes of city regions. We illustrate the structures of spatio-temporal graph convolution (STGC) blocks in Fig. 4(a).

Specifically, each adjacency matrix representing a spatial or temporal graph, $\mathbf{A}_n \in \left\{\mathbf{\Phi}, \mathbf{\Omega}, \mathbf{\Lambda}, \mathbf{\Psi}^{(k)}[W^{(\mathrm{S})}], \mathbf{\Psi}^{(k)}[W^{(\mathrm{M})}], \mathbf{\Psi}^{(k)}[W^{(\mathrm{L})}]\right\}$ is fed to the n-th STGC block. $n \in \{1, \ldots, B\}$ and $B = 6$ are used in our studies given all the graphs formed in Sec. 4.2. Let $\sigma(\cdot)$ be the `sigmoid` activation function. Given the output signals from the preceding $(l-1)$-th layer, denoted as $\mathbf{H}_n^{(l-1)} \in \mathbb{R}^{N \times d'_G}$, we derive the l-th graph convolution layer ($l \in \{1, \ldots, L\}$) based on the input spatial or temporal graph, *i.e.*,

$$\mathbf{H}_n^{(l)} \triangleq \sigma\left(\widetilde{\mathbf{A}}_n \mathbf{H}_n^{(l-1)} \Theta_n^{(l)}\right), \tag{6}$$

where $\Theta^{(l)} \in \mathbb{R}^{d_G \times d'_G}$ is the parameter matrix to be learned, and we adopt the symmetric normalized Laplacian $\mathcal{L}(\cdot)$ and have $\widetilde{\mathbf{A}}_n = \mathcal{L}(\mathbf{A}_n) = \widetilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{A}_n \widetilde{\mathbf{D}}^{-\frac{1}{2}}$, with the degree matrix $\widetilde{\mathbf{D}}(i,i) = \sum_j \mathbf{A}_n(i,j)$. At the 1-st layer of STGC, `EIGDES` takes in the historical mobility flows $\mathbf{Y}^{(k-1)}$ in the last time interval $(k-1)$ as the graph signals and have $\mathbf{H}_n^{(1)} \triangleq \sigma\left(\widetilde{\mathbf{A}}_n \mathbf{Y}^{(k-1)} \mathbf{\Theta}_n^{(1)}\right)$. In this paper, `EIGDES` will take in a total of B spatial and temporal graphs and process them respectively through B STGC blocks. The output from the L-th graph convolution layer of the n-th STGC blocks, denoted as $\mathbf{H}_n^{(L)} \in \mathbb{R}^{N \times d_G}$, will be fed to the module IGIDF for interaction learning.

4.4 Hierarchical Interactive Graph Information Dissemination & Fusion Design

• **Interactive Graph Information Dissemination**: After obtaining the graph embeddings from the STGC blocks regarding different spatial and temporal graphs, we *interactively* extract and propagate the feature information across these graphs in different city regions. For this, we have designed an interactive graph information dissemination mechanism (illustrated in Fig. 4(b)) to *hierarchically* propagate more feature information across the heterogeneous spatial and temporal graph embeddings derived from Sec. 4.2 that are more relevant to the predicted flows.

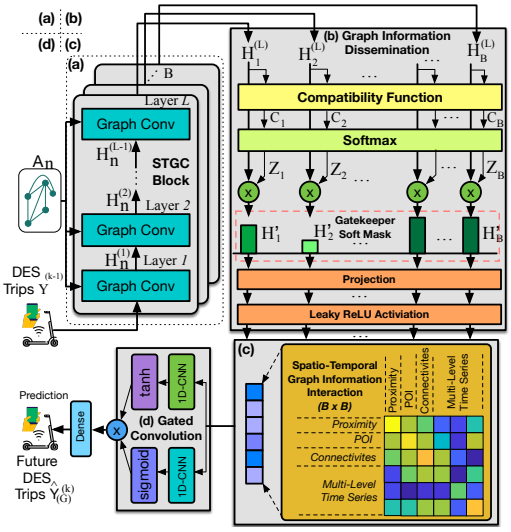

Fig. 4. Core prediction architecture of `EIGDES`.

Specifically, we have designed an interactive graph information attention mechanism between the outputs from STGC blocks to first generate the soft mask for *information control*. The soft mask serves as an information *gatekeeper* to control the information dissemination across the graphs of the city regions, and ensure that they are the most relevant for DES flow prediction.

Let $\mathbf{W}_n^Q, \mathbf{W}_n^K \in \mathbb{R}^{N_F \times d_K}$, and $\mathbf{W}_n^V \in \mathbb{R}^{N_F \times d_V}$ be the matrices of learnable hyperparameters (N_F is number of filters; d_K and d_V

are predefined dimensions). Given input graph embedding $\mathbf{H}_n^{(L)}$ ($n \in \{1, \ldots, B\}$), we first map the graph embeddings from STGC blocks towards $\mathbf{Z}_n$, *i.e.*, $\mathbf{Z}_n =$

$$\texttt{softmax}\left(C_n\right) \cdot \left(\mathbf{H}_n^{(L)} \mathbf{W}_n^V\right) = \frac{\exp\left(C_n\right)}{\sum_{n=1}^{B} \exp\left(C_n\right)} \cdot \left(\mathbf{H}_n^{(L)} \mathbf{W}_n^V\right) \quad (7)$$

where the function C_n ($n \in \{1, \ldots, B\}$) represents the *compatibility* between mappings $\mathbf{Q}_n = \mathbf{H}_n^{(L)} \mathbf{W}_n^Q$ and $\mathbf{K}_n = \mathbf{H}_n^{(L)} \mathbf{W}_n^K$, *i.e.*,

$$C_n = \left(\mathbf{Q}_n \cdot \mathbf{K}_n^{\intercal}\right) / \kappa = \left(\left(\mathbf{H}_n^{(L)} \mathbf{W}_n^Q\right) \cdot \left(\mathbf{H}_n^{(L)} \mathbf{W}_n^K\right)^{\intercal}\right) / \kappa, \quad (8)$$

where we scale the output of the dot product via $\kappa = \sqrt{d_K}$ to mitigate vanishing gradients within the subsequent softmax activation function (particularly given large dot products [50]). This way, we can determine the relative strength that each input graph should be kept in the gatekeeper.

With $\sigma(\cdot)$ as the `sigmoid` function, we form the soft mask $\mathbf{Z}$ as gatekeeper, *i.e.*,

$$\mathbf{Z} = \sigma\left(\left[\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n, \ldots, \mathbf{Z}_B\right]\right). \quad (9)$$

Then, we generate the updated graph embeddings $\mathbf{H}'$ which aim to differentiate the heterogeneous mobility information, *i.e.*,

$$\mathbf{H}' = \left[\mathbf{Z}_1^{\intercal} \cdot \mathbf{H}_1^{(L)}, \quad \ldots, \quad \mathbf{Z}_n^{\intercal} \cdot \mathbf{H}_n^{(L)}, \quad \ldots, \quad \mathbf{Z}_B^{\intercal} \cdot \mathbf{H}_B^{(L)}\right]. \quad (10)$$

By multiplying the gatekeeper soft mask and the input embeddings, we can encourage the information dissemination of the spatial and temporal graphs with greater importance and penalize others that are less important. This way, we can restrict the information propagation within the subset of the input graphs.

• **Interactive Graph Information Fusion**: Given the disseminated information, we conduct the interactive graph information fusion. Let $\mathbf{U}_n^P \in \mathbb{R}^{N_P \times d_V}$ and $\mathbf{W}_n^P \in \mathbb{R}^{1 \times 2N_P}$ be the linear projection parameters, and b_n is the offset parameter. Then, we disseminate the information across the graphs to form their mutual *interactions* (Fig. 4(c)), *i.e.*,

$$\mathbf{G}_n = \mathbf{U}_n^P \mathbf{H}_n' + b_n, \quad \beta_{m,n} = \texttt{LReLU}\left((\mathbf{W}_n^P)^{\intercal}\left[\mathbf{G}_m, \mathbf{G}_n\right]\right), \quad (11)$$

where $m, n \in \{1, \ldots, B\}$, and $\texttt{LReLU}(\cdot)$ is the leaky rectified linear activation function. We further leverage above $\beta_{m,n}$ to compute the interaction weight $\Gamma_{m,n}$ which represents the proportion of information disseminated from the STGC blocks m to n, *i.e.*,

$$\Gamma_{m,n} = \frac{\exp(\beta_{m,n})}{\sum_{m=1}^{B} \exp(\beta_{m,n})}. \quad (12)$$

This way, we actively characterize interactions between each pair of spatial or temporal graphs, instead of aggregating the multiple graphs [22, 38] that might amortize the essential mobility features. The final output given the information interaction becomes the total information weighted and aggregated from others, *i.e.*,

$$\widetilde{\mathbf{H}}_n = \texttt{ReLU}\Big(\sum_{m=1}^{B} \Gamma_{m,n} \mathbf{G}_n\Big), \quad \text{and} \quad \widehat{\mathbf{H}} = \mathbf{W}_{\mathrm{I}} \widetilde{\mathbf{H}}_n, \quad (13)$$

where $\mathbf{W}_{\mathrm{I}} \in \mathbb{R}^{N \times d_V}$ is the trainable parameter after the interaction.

Through the hierarchical structures of dissemination and fusion, `EIGDES` can actively capture the correlations between the graphs from coarse-grained (differentiated by gatekeeper soft mask) to fine-grained (mutual interactions) levels, hence adapting to the broad spectrum of connectivities in DES mobility.

• **Gated Convolution**: Based on our extensive analysis, there exist both local (short-range) and global (long-range) spatio-temporal dependencies across different city regions. For instance, we observe in Chicago that 25% of the scooter trips last for more than 0.24h and reach neighborhood beyond 1.755km. In order to capture the long-range spatial and temporal correlations with respect to the distant neighbors, we further design the gated convolution module within `EIGDES` to enhance the prediction. Specifically, as illustrated in Fig. 4(d), given the fused features from Eq. (13), we have

$$\widetilde{\mathbf{H}} = \tanh\left(\mathbf{\Theta}_1 * \widetilde{\mathbf{H}} + b_1\right) \odot \sigma\left(\mathbf{\Theta}_2 * \widetilde{\mathbf{H}} + b_2\right), \quad (14)$$

where $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ are two independent 1D convolution operations upon the input $\widetilde{\mathbf{H}}$, $\odot$ represents Hadamard product, and b_1, b_2 are the bias terms. $\widehat{\mathbf{H}}$ is further reshaped into $\widehat{\mathbf{H}}'' \in \mathbb{R}^{N \times 2}$ and fed to dense neural network for the final prediction, returning $\mathbf{Y}_{(\mathrm{G})}$, *i.e.*,

$$\mathbf{Y}_{(\mathrm{G})} = \texttt{ReLU}(\texttt{Dense}(\widehat{\mathbf{H}}'')), \quad (15)$$

which is finally reshaped into $\widehat{\mathbf{Y}}_{(\mathrm{G})} \in \mathbb{R}^{N \times 2}$.

• **External Factor Integration**: In addition to the spatial distributions and temporal trends, we integrate the external factors of time, holidays, and weather conditions within `EIGDES` (Fig. 2(c)). We use one-hot encoding for the categorical weather data such as shower or not (one means the condition is observed and zero means otherwise). We count the number of each category of the events recorded within each time interval. We conduct min-max normalization upon the numerical weather data (*e.g.*, temperature) and the event numbers with respect to each dimension. By concatenating the weather and event features, we then form a 23-D external vector $\mathbf{F}_{(\mathrm{E})}$ as the input for a multi-layer dense neural network, with the `ReLU` activation functions interleaving the two `Dense` layers, *i.e.*,

$$\begin{aligned} \mathbf{X}_0 &= \texttt{Dense}\left(\mathbf{F}_{(\mathrm{E})}\right), \qquad \mathbf{X}_1 = \texttt{ReLU}\left(\mathbf{X}_0\right), \\ \mathbf{X}_2 &= \texttt{Dense}\left(\mathbf{X}_1\right), \qquad \mathbf{X}_3 = \texttt{ReLU}\left(\mathbf{X}_2\right), \end{aligned} \quad (16)$$

which is finally reshaped into the DES distributions $\widehat{\mathbf{Y}}_{(\mathrm{E})} \in \mathbb{R}^{N \times 2}$. Due to the high dimensions of external factors (say, weather conditions and events), the first `Dense` layer serves as *embedding* to further incorporate the features in input $\mathbf{F}_{(\mathrm{E})}$.

Finally, the final DES flow prediction $\widehat{\mathbf{Y}}^{(k)}$ is given by

$$\widehat{\mathbf{Y}}^{(k)} = \mathbf{W}_{(\mathrm{G})} \widehat{\mathbf{Y}}_{(\mathrm{G})} + \mathbf{W}_{(\mathrm{E})} \widehat{\mathbf{Y}}_{(\mathrm{E})}, \quad (17)$$

where $\mathbf{W}_{(\mathrm{G})}$ and $\mathbf{W}_{(\mathrm{E})}$ are trainable parameters.

5 SOCIALLY-EQUITABLE ADAPTATION

5.1 Design Motivations

Based on the equality or horizontal equity [33, 48, 55], considering *demands* (pick-ups) and *supplies* (drop-offs) as indicator of mobility resources and opportunities, we target at: *(1) equal distribution and allocation of such DES service resources*, and *(2) consistent prediction performance (accuracy)* across the advantaged and disadvantaged communities. Towards these *fairness-aware* goals, we *regularize* the predictions of `EIGDES` during mobility learning. For fine-grained characterization, our (un)fairness design particularly takes into account *individual* or *per capita* share of the resources, where for each community we consider their demographic sub-population distribution (characterized by $\omega_{i,\alpha}^{+}$ and $\omega_{i,\alpha}^{-}$ in Sec. 3.2) in the grids.

5.2 Resource & Performance Fairness Metrics

For ease of description, we let $\widehat{y}_i^{(k)} \in \left\{\widehat{\mathbf{I}}_i^{(k)}, \widehat{\mathbf{O}}_i^{(k)}\right\}$ be the estimated demand or supply at $\mathbf{g}_i$ at interval k. Then for each interval k, we find each (un)fairness measure, denoted as $U_x^{(k)}(\alpha)$, for a fairness category x given a socioeconomic attribute α (*e.g.*, social ethnicity).

• **Resource Non-Parity Unfairness**: In practice, the DES flow prediction will determine the allocation parity in scooters and other

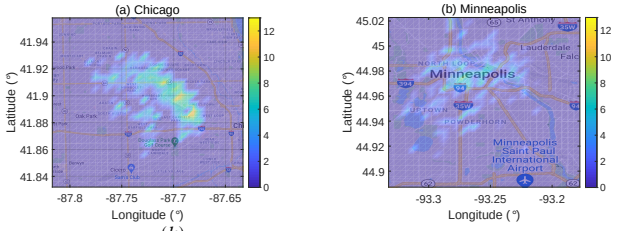

Fig. 5. $U_{\text{Par}}^{(k)}$ in (a) Chicago and (b) Minneapolis.

support infrastructures. Thus, we measure the non-parity unfairness [29, 60] based on absolute difference between the flow predictions for advantaged and disadvantaged groups for all city regions at time interval k given a certain socioeconomic attribute α, *i.e.*,

$$U_{\text{Par}}^{(k)}(\alpha) \triangleq \frac{1}{\sum_i^{2N} \widehat{y}_i^{(k)}} \sum_i^{2N} \left| \mathbb{Q}\left[\widehat{y}_i^{(k)}\right]_\alpha^+ - \mathbb{Q}\left[\widehat{y}_i^{(k)}\right]_\alpha^- \right|, \quad (18)$$

where $\mathbb{Q}\left[\widehat{y}_i^{(k)}\right]_\alpha^+ = \frac{\widehat{y}_i^{(k)} \omega_{i,\alpha}^+}{\sum_i^N P_i \omega_{i,\alpha}^+}$ and $\mathbb{Q}\left[\widehat{y}_i^{(k)}\right]_\alpha^- = \frac{\widehat{y}_i^{(k)} \omega_{i,\alpha}^-}{\sum_i^N P_i \omega_{i,\alpha}^-}$ represent the estimated per capita (for each person) predictions in groups $\mathbf{c}^+$ and $\mathbf{c}^-$, and the total prediction $\sum_i^{2N} \widehat{y}_i^{(k)}$ is used for normalization (we have $2N$ as both pick-ups and drop-offs are counted per city region). $U_{\text{Par}}^{(k)}(\alpha)$ takes into account the population percentage of the different groups, and its value becomes small when the mobility resources allocated (predicted demands and supplies) are consistently proportional to their populations.

By substituting $\widehat{y}_i^{(k)}$'s with the ground-truths $y_i^{(k)}$'s, we further illustrate in Fig. 5 the spatial $U_{\text{Par}}^{(k)}$ of Chicago and Minneapolis, *i.e.*, latent *unfairness* within the DES mobility resource distribution. In particular, we show the aggregated non-parity unfairness (with social ethnicity, income, and education) within a selected week (Chicago: 07/07/2019 – 07/13/2019; Minneapolis: 07/22/2018–07/28/2018). Such spatial mismatch between DES mobility resources and communities renders incorporating (un)fairness metrics within mobility modeling both necessary and imperative.

• **Prediction Performance Unfairness**: Prediction performance difference in terms of accuracy will affect the service experience of different communities. To characterize the fairness in DES flow prediction performance for different communities, we measure the inconsistency in unsigned flow estimation errors across advantaged and disadvantaged groups via the absolute value unfairness given a socioeconomic attribute α, *i.e.*, $U_{\text{Perf}}^{(k)}(\alpha) \triangleq$

$$\frac{1}{\sum_i^{2N} \widehat{y}_i^{(k)}} \sum_i^{2N} \left| \left| \mathbb{Q}\left[\widehat{y}_i^{(k)}\right]_\alpha^+ - \mathbb{Q}\left[y_i^{(k)}\right]_\alpha^+ \right| - \left| \mathbb{Q}\left[\widehat{y}_i^{(k)}\right]_\alpha^- - \mathbb{Q}\left[y_i^{(k)}\right]_\alpha^- \right| \right|, \quad (19)$$

where $\mathbb{Q}\left[y_i^{(k)}\right]_\alpha^+ = \frac{y_i^{(k)} \omega_{i,\alpha}^+}{\sum_i^N P_i \omega_{i,\alpha}^+}$ and $\mathbb{Q}\left[y_i^{(k)}\right]_\alpha^- = \frac{y_i^{(k)} \omega_{i,\alpha}^-}{\sum_i^N P_i \omega_{i,\alpha}^-}$ represent the per capita distributions of DES predictions. It characterizes the quality of predictions for each group or community due to the unsigned comparison, regardless of the direction of the differences. In other words, the Eq. (19) aims at minimizing the differences of the absolute errors for the DES flow predictions at the advantaged and disadvantaged communities.

5.3 Socially-Equitable Objective Regularization

The conventional objective of minimizing the prediction errors in deep learning training likely leads to bias towards the historically advantaged communities and tourism regions [19, 20]. Thus, we leverage the unfairness metrics defined in Sec. 5.2 to regularize the learning objective. Specifically, we form the hybrid training objective for EIGDES, which is to minimize the weighted loss in terms of *prediction errors* $\mathcal{L}_{\text{Pred}}$ as well as *(un)fairness costs* $\mathcal{L}_{\text{Fair}}(x)$ given a mobility fairness metric x in Sec. 5.2, *i.e.*,

$$\mathcal{L}_x = \mathcal{L}_{\text{Pred}} + \lambda \mathcal{L}_{\text{Fair}}(x), \quad (20)$$

where $\lambda > 0$ is a weight parameter for model training.

To mitigate impact from the outliers particularly in the squared loss, we adopt the Huber loss [13] for $\mathcal{L}_P$, *i.e.*, $\texttt{Huber}(\widehat{y}_i, y_i, \delta) \triangleq$

$$\begin{cases} \frac{1}{2}\left(\widehat{y}_i - y_i\right)^2, & \text{if} \quad |\widehat{y}_i - y_i| < \delta; \\ \frac{1}{2}\delta^2 + \delta \cdot \left(|\widehat{y}_i - y_i| - \delta\right), & \text{otherwise.} \end{cases} \quad (21)$$

The hyperparameter δ indicates where the loss function changes from a quadratic one to linear, making the loss function less sensitive to the outliers and more robust towards flow dynamics.

In practice, the city planners analyze multiple sensitive socioeconomic attributes $\mathbf{S}$ (say, social ethnicity, income, and education level in our studies) in $\mathcal{L}_{\text{Fair}}(x)$ for comprehensive model training upon an unfairness metric x, *i.e.*, Eqs. (18) and (19).

Therefore, the final loss function for our socially-equitable model training, given each of the target (un)fairness metrics $x \in \{\text{Par}, \text{Perf}\}$, becomes

$$\mathcal{L}_x = \frac{1}{Z} \sum_{i=1}^{Z} \texttt{Huber}\left(\widehat{y}_i, y_i, \delta\right) + \sum_{\alpha \in \mathbf{S}} \lambda U_x(\alpha), \quad \lambda > 0, \quad (22)$$

where α is a certain sensitive socioeconomic attribute and Z is the total number of prediction samples.

Table 1: Overall prediction performance of all approaches.

Schemes	Chicago		Minneapolis		Edmonton	
	RMSE	PCR	RMSE	PCR	RMSE	PCR
HA	7.44	23.6%	6.31	25.6%	7.23	32.3%
GP	6.12	21.9%	6.33	25.1%	7.19	31.2%
ARIMA	5.32	18.1%	4.33	23.5%	5.16	30.7%
RNN	5.69	20.2%	3.62	22.5%	4.52	28.7%
LSTM	5.72	20.5%	3.15	18.7%	4.21	27.4%
GRU	5.36	19.8%	3.02	15.6%	3.72	25.1%
CNN	4.86	17.5%	2.96	18.7%	2.88	21.9%
STC2D	4.56	15.2%	2.87	16.2%	3.24	32.3%
STC3D	4.02	13.6%	2.45	15.3%	2.62	19.6%
STRN	3.75	16.3%	2.16	14.8%	2.67	20.1%
MGCN	3.66	15.3%	2.36	12.8%	2.32	18.7%
GWN	3.86	17.5%	2.86	17.5%	2.91	24.3%
MSGN	3.72	14.2%	2.36	16.7%	2.55	19.7%
MTGNN	4.02	19.2%	2.62	12.9%	2.63	22.1%
CSTN	3.95	19.2%	2.17	13.8%	3.34	23.5%
TLSTM	3.55	13.2%	2.75	13.1%	2.96	25.1%
FST	4.62	18.1%	2.83	15.9%	3.12	24.1%
FRep	4.51	17.2%	2.71	15.1%	3.51	26.7%
EIGDES	**1.98**	**5.2%**	**1.11**	**5.1%**	**1.97**	**8.9%**

6 EVALUATION

We compare the performance of EIGDES with other baseline approaches using root mean square error (RMSE), poor case rate (PCR), as well as the (un)fairness (see Appendixes B and C for detailed definitions, baseline approaches, and additional results).

• **Overall Prediction Accuracy & Fairness**: We first show in Table 1 the overall flow prediction accuracy and fairness of different approaches in Chicago, Minneapolis, and Edmonton. Thanks to the HIGIDF designs, EIGDES is shown to improve prediction performance in terms of RMSE and PCR (at least by 35.26%), including other graph neural network approaches [15, 38, 54]. We can also observe overall higher errors for all approaches in Chicago and Edmonton than in Minneapolis, mainly due to their larger traffic

volumes and more complex routes. As shown in Table 2, EIGDES *disassociates* the DES flows with the sensitive socioeconomic attributes through our regularization, enabling subsequent fair DES resource allocation decisions and prediction quality experienced across communities. In terms of societal implications, we can infer from the fairness results that EIGDES can enhance the scooter accessibility by at least 54.85%, and reduce the prediction performance difference by at least 44.26%, which is essential for urban planners in coordinating the DES systems across different communities.

Table 2: Resource non-parity & prediction performance unfairness.

Schemes	Chicago		Minneapolis		Edmonton	
	U_{Par}	U_{Perf}	U_{Par}	U_{Perf}	U_{Par}	U_{Perf}
HA	171.15	96.33	89.70	59.92	87.66	57.24
GP	115.13	41.21	43.19	31.04	47.10	31.22
ARIMA	91.47	88.21	90.12	41.89	67.21	39.01
RNN	126.60	74.06	58.91	34.24	64.12	48.70
LSTM	112.39	39.40	34.83	29.80	44.31	37.21
GRU	117.36	42.92	51.46	73.46	63.77	51.33
CNN	72.64	59.04	44.78	40.48	52.88	45.96
STC2D	96.33	32.05	40.80	51.46	57.06	43.18
STC3D	84.68	24.55	42.08	52.16	41.65	35.64
STRN	89.92	17.06	37.86	34.24	62.70	33.21
MGCN	63.52	39.08	37.86	32.08	41.70	47.31
GWN	72.64	59.04	61.25	79.04	61.21	45.96
MSGN	96.33	32.05	72.12	37.05	73.23	55.12
MTGNN	72.26	21.52	53.23	27.51	51.21	31.56
CSTN	69.12	21.16	62.92	17.26	62.92	47.13
TLSTM	78.12	27.12	43.92	22.76	73.92	55.91
FST	63.52	39.08	42.64	21.35	49.21	33.10
FRep	57.06	26.31	36.12	24.17	41.55	35.31
EIGDES	**10.26**	**6.52**	**11.72**	**9.62**	**18.76**	**11.95**

• **Fairness with Single Socioeconomic Attribute**: We evaluate in Fig. 6 the resource non-parity fairness and prediction performance fairness of EIGDES given each individual socioeconomic attribute, *i.e.*, social ethnicity (Eth), education (Edu), or income (Inc), in the optimization objective regularizer. In particular, we reduce the number of socioeconomic attributes in Eq. (22) during the model learning. Such studies are suitable for city planning when policy decision-makers are particularly interested in a specific socioeconomic attribute in the city. We can observe the relatively higher importance in mitigating mobility unfairness across ethnic groups for all cities. We can also observe generally lower unfairness in the category of education than income in Chicago, while in Minneapolis and Edmonton overall higher unfairness exists in education than income, implying more college-educated users in the latter cases. Such differences across three metropolitan cities reflect the spatial DES deployment and neighborhood socioeconomic conditions.

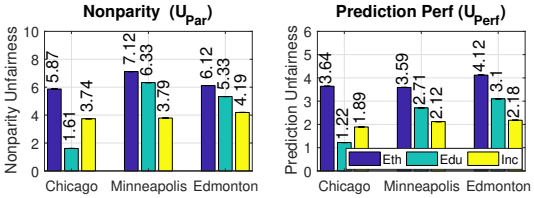

Fig. 6. Fairness performance with single socioeconomic attribute.

• **Model Ablation Studies**: We remove (or replace) different design components of EIGDES, and evaluate the resultant flow prediction RMSE in Fig. 7 (for Chicago). We train the model based on DES trips during June 15–25, 2019, and evaluate based on June 26–30, 2019. We compare the performance of complete settings (denoted as (i) "w/ all") with variations of EIGDES *without*: (ii) external factors (w/o ext), (iii) spatial correlations (w/o $\mathbf{\Phi}$), (iv) POI correlations (w/o $\mathbf{\Omega}$), (v) region-to-region connectivities (w/o $\mathbf{\Lambda}$), (vii) multi-level temporal correlations (w/o $\mathbf{\Psi}$), (viii) information dissemination masking (w/o Mask), (ix) interaction fusion (w/o Fusion), and (x) fairness regularization (w/o Reg). We also have (vi) DTW instead of *LBK* (w/ DTW) for $\mathbf{\Psi}$. We can observe the relative importance of different components in EIGDES. In particular, we can infer from the difference between (i) and (x) that single-minded "accuracy" objective may not necessarily ensure satisfactory prediction performance, while regularization based on more fine-grained socioeconomic information can better identify and characterize the local demand and supply, and enhance model generalization.

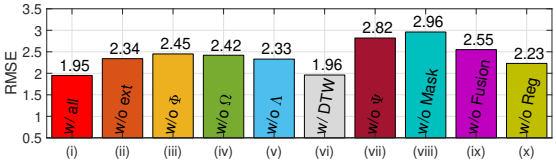

Fig. 7. RMSE of ablation studies on EIGDES.

• **Interaction of Spatial and Temporal Graphs**: Taking Chicago as an example, we visualize in Fig. 8 the $\mathbf{\Gamma} \in \mathbb{R}^{B\times B}$ of two time intervals (a Monday noon and a Sunday noon), where the lighter colors imply more attention weights and more mutual interactions between the graphs (organized as Fig. 4(c)). We can observe the multi-level time-series correlations ($\mathbf{\Psi}$) interact with each other for both cases. For (a) Monday noon, we can observe more interactions between spatial proximity ($\mathbf{\Phi}$), connectivities ($\mathbf{\Lambda}$), and short-term time-series correlations, mainly owing to the transient ride trips during the rush hours. For (b) Sunday noon, besides interactions within each set of spatial and temporal graphs, we can further observe more interactions between POI ($\mathbf{\Omega}$) and long-term time-series correlations. This is mainly due to more recreational rides spanning throughout the weekend around the urban points-of-interest.

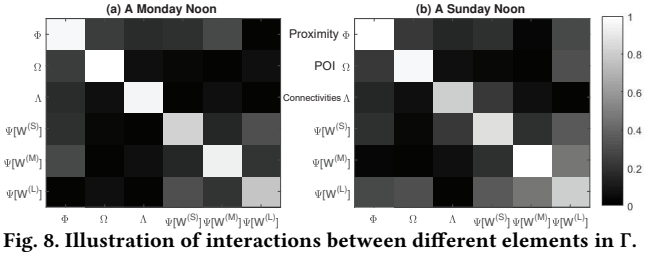

Fig. 8. Illustration of interactions between different elements in $\mathbf{\Gamma}$.

7 CONCLUSION

We have proposed EIGDES, a socially-equitable flow prediction system for dockless e-scooter sharing. EIGDES *hierarchically* incorporates spatial and temporal interactions across different graphs across city regions, via a novel interactive graph information dissemination and fusion framework. By integrating with the optimization regularizer, EIGDES *jointly* learns the complex DES flow patterns and socio-economic factors. EIGDES yields flow distribution predictions which mitigate spatial resource allocation and prediction performance unfairness for disadvantaged communities. Our extensive studies based on over 2.1 million trips in North America have validated EIGDES's effectiveness, fairness, and accuracy in socially-equitable flow prediction.

This project is supported, in part, by the Google Research Scholar Program Award (2021–2022) and the 2021 Nvidia Applied Research Accelerator Program Award.

REFERENCES

[1] 2019. The Ultimate Guide to being a Lime Scooter Charger/Juicer. https://gridwise.io/the-ultimate-guide-to-being-a-lime-scooter-charger-juicer, Accessed: 2021-10-21. (2019).
[2] 2021. Canada Federal Census Tract in Alberta. https://open.alberta.ca/. (2021).
[3] 2021. Chicago Open Data. https://data.cityofchicago.org/. (2021).
[4] 2021. Edmonton E-Scooter Share. https://data.edmonton.ca/Transportation/E-Scooter-Share-API. (2021).
[5] 2021. Eventful – Local Upcoming Events. www.eventful.com, Accessed: 2021-10-21. (2021).
[6] 2021. Minneapolis Open Data. https://opendata.minneapolismn.gov/. (2021).
[7] 2021. Open Street Map. https://www.openstreetmap.org/, Accessed: 2021-10-21. (2021).
[8] 2021. US Census. https://www.census.gov/, Accessed: 2021-10-21. (2021).
[9] 2021. Weather for 243 countries of the world. https://rp5.ru/Weather_in_the_world, Accessed: 2021-10-21. (2021).
[10] 2022. Web of Things (WoT) Architecture. https://www.w3.org/TR/wot-architecture/#sec-use-cases. (2022).
[11] Richard Berk, Hoda Heidari, Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth. 2017. A convex framework for fair regression. *arXiv preprint arXiv:1706.02409* (2017).
[12] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series.. In *Proc. ACM KDD Workshop*, Vol. 10. Seattle, WA, USA:, 359–370.
[13] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.
[14] Or Caspi, Michael J Smart, and Robert B Noland. 2020. Spatial associations of dockless shared e-scooter usage. *Transportation Research Part D: Transport and Environment* 86 (2020), 102396.
[15] Di Chai, Leye Wang, and Qiang Yang. 2018. Bike flow prediction with multi-graph convolutional networks. In *Proc. ACM SIGSPATIAL*. 397–400.
[16] Zhixuan Fang, Longbo Huang, and Adam Wierman. 2017. Prices and Subsidies in the Sharing Economy. In *Proc. WWW*. 53–62.
[17] Martha Fedorowicz, Emily Bramhall, Mark Treskon, and Richard Ezike. 2020. *New Mobility and Equity: Insights for Medium-Size Cities*. Research Report – Urban Institute.
[18] Yanbo Ge, Christopher R Knittel, Don MacKenzie, and Stephen Zoepf. 2016. *Racial and gender discrimination in transportation network companies*. Technical Report. National Bureau of Economic Research.
[19] Yanbo Ge, Christopher R Knittel, Don MacKenzie, and Stephen Zoepf. 2020. Racial discrimination in transportation network companies. *Journal of Public Economics* 190 (2020), 104205.
[20] Solomon Greene, Graham MacDonald, Olivia Arena, Tanaya Srini, and Ruth Gourevitch. 2019. *Technology and Equity in Cities: Emerging Challenges and Opportunities*. Research Report – Urban Institute.
[21] Suining He and Kang G Shin. 2019. Spatio-temporal capsule-based reinforcement learning for mobility-on-demand network coordination. In *Proc. WWW*. 2806–2813.
[22] Suining He and Kang G Shin. 2020. Dynamic Flow Distribution Prediction for Urban Dockless E-Scooter Sharing Reconfiguration. In *Proc. WWW*. 133–143.
[23] Suining He and Kang G. Shin. 2020. Towards Fine-Grained Flow Forecasting: A Graph Attention Approach for Bike Sharing Systems. In *Proc. WWW*. 88–98.
[24] Tianfu He, Jie Bao, Ruiyuan Li, Sijie Ruan, Yanhua Li, Chao Tian, and Yu Zheng. 2018. Detecting Vehicle Illegal Parking Events using Sharing Bikes' Trajectories.. In *Proc. ACM KDD*. 340–349.
[25] Kate Hosford and Meghan Winters. 2018. Who are public bicycle share programs serving? An evaluation of the equity of spatial access to bicycle share service areas in Canadian cities. *Transportation research record* 2672, 36 (2018), 42–50.
[26] Safwan Hossain, Andjela Mladenovic, and Nisarg Shah. 2020. Designing Fairly Fair Classifiers Via Economic Fairness Notions. In *Proc. WWW*. 1559–1569.
[27] Pierre Hulot, Daniel Aloise, and Sanjay Dominik Jena. 2018. Towards station-level demand prediction for effective rebalancing in bike-sharing systems. In *Proc. ACM KDD*. 378–386.
[28] Di Jin, Cuiying Huo, Chundong Liang, and Liang Yang. 2021. *Heterogeneous Graph Neural Network via Attribute Completion*. 391–400.
[29] Toshihiro Kamishima, Shotaro Akaho, and Jun Sakuma. 2011. Fairness-aware learning through regularization approach. In *Proc. IEEE ICDM Workshops*. 643–650.
[30] Eamonn Keogh and Chotirat Ann Ratanamahatana. 2005. Exact indexing of dynamic time warping. *Knowledge and information systems* 7, 3 (2005), 358–386.
[31] Sage J Kim and Wendy Bostwick. 2020. Social Vulnerability and Racial Inequality in COVID-19 Deaths in Chicago. *Health Education & Behavior* (2020).
[32] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[33] Gerald S Leventhal. 1980. What should be done with equity theory? In *Social Exchange*. Springer, 27–55.
[34] Mengzhang Li and Zhanxing Zhu. 2021. Spatial-Temporal Fusion Graph Neural Networks for Traffic Flow Forecasting. In *Proc. AAAI*, Vol. 35. 4189–4196.
[35] Yuxuan Liang, Kun Ouyang, Junkai Sun, Yiwei Wang, Junbo Zhang, Yu Zheng, David Rosenblum, and Roger Zimmermann. 2021. Fine-Grained Urban Flow Prediction. In *Proc. WWW*. 1833–1845.
[36] Junming Liu, Leilei Sun, Qiao Li, Jingci Ming, Yanchi Liu, and Hui Xiong. 2017. Functional zone based hierarchical demand prediction for bike system expansion. In *Proc. ACM KDD*. 957–966.
[37] Lingbo Liu, Zhilin Qiu, Guanbin Li, Qing Wang, Wanli Ouyang, and Liang Lin. 2019. Contextualized spatial–temporal network for taxi origin-destination demand prediction. *IEEE Transactions on Intelligent Transportation Systems* 20, 10 (2019), 3875–3887.
[38] Man Luo, Bowen Du, Konstantin Klemmer, Hongming Zhu, Hakan Ferhatosmanoglu, and Hongkai Wen. 2020. D3P: Data-Driven Demand Prediction for Fast Expanding Electric Vehicle Sharing Systems. *Proc. ACM IMWUT* 4, 1, Article 21 (March 2020), 21 pages.
[39] Xiaolei Ma, Zhuang Dai, Zhengbing He, Jihui Ma, Yong Wang, and Yunpeng Wang. 2017. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* 17, 4 (2017), 818.
[40] Grant McKenzie. 2019. Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, DC. *Journal of Transport Geography* 78 (2019), 19–28.
[41] Stephen J Mooney, Kate Hosford, Bill Howe, An Yan, Meghan Winters, Alon Bassok, and Jana A Hirsch. 2019. Freedom from the station: Spatial equity in access to dockless bike share. *Journal of Transport Geography* 74 (2019), 91–96.
[42] Nuno Mota, Negar Mohammadi, Palash Dey, Krishna P. Gummadi, and Abhijnan Chakraborty. 2021. *Fair Partitioning of Public Resources: Redrawing District Boundary to Minimize Spatial Inequality in School Funding*. 646–657.
[43] Takuma Oda. 2021. Equilibrium Inverse Reinforcement Learning for Ride-hailing Vehicle Network. In *Proc. WWW*. 2281–2290.
[44] Yan Pan, Ray Chen Zheng, Jiaxi Zhang, and Xin Yao. 2019. Predicting bike sharing demand using recurrent neural networks. *Procedia Computer Science* 147 (2019), 562–566.
[45] Zheyi Pan, Songyu Ke, Xiaodu Yang, Yuxuan Liang, Yong Yu, Junbo Zhang, and Yu Zheng. 2021. AutoSTG: Neural Architecture Search for Predictions of Spatio-Temporal Graph. In *Proc. WWW*. 1846–1855.
[46] Susan Shaheen and Adam Cohen. 2018. Equity and shared mobility. *ITS Berkeley Policy Briefs* 2018, 06 (2018).
[47] Shun-Yao Shih, Fan-Keng Sun, and Hung-yi Lee. 2019. Temporal pattern attention for multivariate time series forecasting. *Machine Learning* 108, 8 (2019), 1421–1441.
[48] Emily Talen. 1998. Visualizing fairness: Equity maps for planners. *Journal of the American Planning Association* 64, 1 (1998), 22–38.
[49] Waldo Tobler. 2004. On the first law of geography: A reply. *Annals of the Association of American Geographers* 94, 2 (2004), 304–310.
[50] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proc. NIPS*. 5998–6008.
[51] Qi Wang, Nolan Edward Phillips, Mario L Small, and Robert J Sampson. 2018. Urban mobility and neighborhood isolation in America's 50 largest cities. *PNAS* 115, 30 (2018), 7735–7740.
[52] Shuai Wang, Tian He, Desheng Zhang, Yunhuai Liu, and Sang H. Son. 2019. Towards efficient sharing: A usage balancing mechanism for bike sharing systems. In *Proc. WWW*. 2011–2021.
[53] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proc. ACM SIGKDD*. 753–763.
[54] Z Wu, S Pan, G Long, J Jiang, and C Zhang. 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In *Proc. IJCAI*.
[55] An Yan and Bill Howe. 2019. Fairness in Practice: A Survey on Equity in Urban Mobility. *IEEE Data Eng. Bull.* 42, 3 (2019), 49–63.
[56] An Yan and Bill Howe. 2019. FairST: Equitable spatial and temporal demand prediction for new mobility systems. In *Proc. ACM SIGSPATIAL*. 552–555.
[57] An Yan and Bill Howe. 2020. Fairness-Aware Demand Prediction for New Mobility. In *Proc. AAAI*, Vol. 34. 1079–1087.
[58] An Yan and Bill Howe. 2021. EquiTensors: Learning Fair Integrations of Heterogeneous Urban Data. In *Proc. ACM SIGMOD*. 2338–2347.
[59] Huaxiu Yao, Yiding Liu, Ying Wei, Xianfeng Tang, and Zhenhui Li. 2019. Learning from multiple cities: A meta-learning approach for spatial-temporal prediction. In *Proc. WWW*. 2181–2191.
[60] Sirui Yao and Bert Huang. 2017. Beyond parity: Fairness objectives for collaborative filtering. In *Proc. NIPS*. 2921–2930.
[61] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. 2013. Learning fair representations. In *Proc. ICML*. 325–333.
[62] Chaoyun Zhang and Paul Patras. 2018. Long-term mobile traffic forecasting using deep spatio-temporal neural networks. In *Proc. ACM MobiHoc*. 231–240.
[63] Chao Zhang, Keyang Zhang, Quan Yuan, Haoruo Peng, Yu Zheng, Tim Hanratty, Shaowen Wang, and Jiawei Han. 2017. Regions, periods, activities: Uncovering

urban dynamics via cross-modal representation learning. In *Proc. WWW*. 361–370.
[64] Cheng Zhang, Linan Zhang, Yangdong Liu, and Xiaoguang Yang. 2018. Short-term prediction of bike-sharing usage considering public transport: A LSTM approach. In *Proc. IEEE ITSC*. 1564–1571.
[65] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proc. AAAI*. 1655–1661.
[66] Zhenpeng Zou, Hannah Younes, Sevgi Erdoğan, and Jiahui Wu. 2020. Exploratory analysis of real-time e-scooter trip data in Washington, DC. *Transportation Research Record* 2674, 8 (2020), 285–299.

APPENDIX

A EXAMPLES OF DATASETS

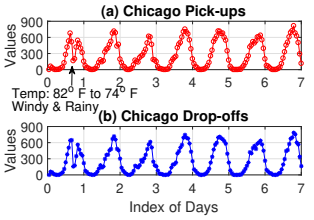

Fig. 9. DES pick-up/drop-off flows in Chicago.

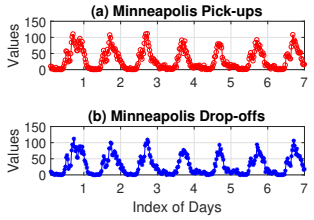

Fig. 10. DES pick-up/drop-offs flows in Minneapolis.

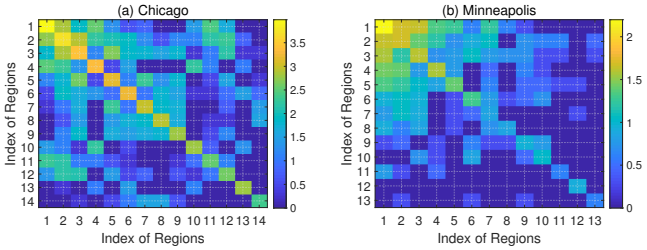

Fig. 11. Transitions in (a) Chicago and (b) Minneapolis ($\log_{10}(\cdot)$).

Taking Chicago and Minneapolis as the examples, we demonstrate the hourly/daily routines and dynamics of total pick-ups/drop-offs within each time interval of a week in Figs. 9 and 10. We can also see in Fig. 9 a sudden drop in pick-ups/drop-offs on Day 1 due to the windy/rainy conditions and temperature drop. Fig. 11 illustrates the region-to-region transitions in Chicago and Minneapolis (Sec. 4.2). Table 3 further summarizes the external factors used in EIGDES.

Category	**POIs** (13-D)	**Weather** (11-D)	**Events** (12-D)
Data	Residential, education, cultural, recreational, social services, transportation, commercial, government, religious, health services, public safety, water, sustenance.	Temperature, precipitation, wind speed, other conditions including misty/drizzle /light rain/shower /snow/freezing/foggy /no significant clouds.	Comedy, conferences, education, food, health, museum, music, networking, outdoors, arts, sports, technology.

Table 3: Details of external weather factors and urban features.

B EXPERIMENTAL SETTINGS

• **Comparison Schemes**: We evaluate the following conventional and state-of-the-art mobility prediction algorithms:

- HA: which predicts the flows of a target period based on the historical average (HA) of the same periods in the past. Specifically, we use the historical average during the same period in the historical records; for example, we average all the DES flows during 8:00–8:30 of all recorded Mondays to predict that of 8:00–8:30 of a specific Monday.
- GP/ARIMA: which predict the DES flows with Gaussian process (GP) and auto-regressive integrated moving average (ARIMA), respectively.
- RNN/LSTM/GRU: which predict the DES flows based on the recurrent neural network (RNN) [44], long short-term memory (LSTM) [64], and gated recurrent unit (GRU), respectively. For RNN, LSTM, and GRU, we predict the DES flows based on the K most recent time intervals, and we empirically set $K = 72$ and the units (dimensionality of the output space) as 128.
- CNN: which takes in the historical heatmaps of DES flows and predicts the future flows via conventional convolutional neural network (CNN). We adopt 12 Conv2D layers, each of which uses 32 filters of kernel size 5×5.
- STC2D: which spatio-temporally processes the historical flows via 2D CNN [39], and then further extracts the sequential dependency based on LSTM. We adopt 12 Conv2D layers, each of which uses 32 filters of kernel size 5×5.
- STC3D: which spatio-temporally processes the historical flows via 3D CNN [62], and then extracts the sequential dependency based on LSTM. We adopt 12 Conv3D layers, each of which uses 32 filters of kernel size 5×5.
- STRN: which forecasts the DES flows via the spatio-temporal residual neural network (STRN) [65]. We adopt 12 residual layers, each of which has 32 filters of kernel size 3×3.
- MGCN: which leverages the multi-graph convolutional neural network to predict the DES flows [15].
- GWN: which leverages the graph wave neural network for the traffic flow prediction [54].
- MSGN: which adaptively learns and forecasts the DES traffic flows with the multi-scale graph neural network [38].
- MTGNN: which estimates the future DES flows based on multiple time-series graph neural network [53].
- CSTN: which predicts the DES flows with the convolution embedded LSTM-based method with a contextualized spatio-temporal network [37].
- TLSTM: which predicts the DES flows based on temporal pattern attention long short-term memory [47].
- FST: which takes in three-stream network for equitable mobility resource demand prediction [56, 57]. We respectively adopt 3 Conv3D layers ($3 \times 3 \times 3$ filters), Conv2D (3×3 filters) and Conv1D (3 filters) layers to take in historical demands, urban map road network and city-wide temperature.
- FRep: which learns the fair representation integration to align the heterogeneous data [58].

• **Evaluation Settings**: Unless otherwise stated, we set the following default parameters in evaluations. Based on empirical studies and the DES coverage, we discretize the city map into 32×32 (N = 1,024), and the time into 1-hour and 30-min intervals for Chicago and Minneapolis due to the granularity of their original datasets, respectively. For Edmonton, we partition the city map into 28×28 (N = 784), and time into 30-min intervals. We note that our model is general enough to be adapted to discretization based on Census tracts or zip codes by formulating them into the graph. We set the learning rate to 0.0001, and 500 epochs for all the model evaluated. To emulate the real-world deployment, we divide the dataset on a *monthly* basis and respectively conduct the offline model training as well as online model prediction. Specifically, for each month, we use the first two weeks, *i.e.*, 336 time intervals in Chicago and 672 time intervals in Minneapolis and Edmonton, for the model training, and the rest of that month for prediction (testing).

We have implemented all these schemes based on Python/Tensorflow upon a GPU server with AMD Ryzen Threadripper 3960X Processor (24 Core), 128Gb RAM, and four Nvidia RTX3090 (24Gb DDR6). We adopt the `Adam` optimizer [32], which has been shown to yield faster convergence than the conventional stochastic gradient descent.

• **Model Parameter Settings**: For the EIGDES model designs, we empirically set the important parameters as follows. $Z = 5$ for the multi-level time-series correlations in Eq. (4); $d_G = d'_G = 8$; number of STGC blocks $B = 6$; number of graph convolution layers in each STGC block $L = 8$; $N_F = 8$, $d_V = 8$, $d_K = 2$, and $N_P = 32$ in HIGIDF. We set the output dimensions of $\mathbf{X}_0$ and $\mathbf{X}_2$ in Eq. (16) as 32 and $2N$. For the schemes other than EIGDES, we adopt the Huber loss function only. We set $\delta = 2.5$ in Eq. (21) for all approaches evaluated. We use $\lambda = 0.6$ within the fairness regularizer in Eq. (22). For each (un)fairness metric, we use all three socioeconomic attributes in Eq. (22) by default, *i.e.*, social ethnicity, education, and income. Model training generally takes 0.68s per epoch, and prediction takes 17ms per input of a time interval. In our experiment, we also observe adopting *LBK* substantially reduces EIGDES's prediction time by 74.9% and 46.9% compared with DTW and Pearson correlation.

• **Performance Metrics**: We comprehensively evaluate all the schemes and designs based on the following three metrics:

(1) *Root Mean Square Error* (*RMSE*): which evaluates the spread of the residuals, *i.e.*, RMSE $= \sqrt{\frac{1}{Z}\sum_{i=1}^{Z}(\widehat{y}_i - y_i)^2}$.

(2) *Poor Case Rate* (*PCR*): Furthermore, we define the poor case rate (PCR) as the percentage of predictions which have excessive over- or under-estimations than ground-truths, *i.e.*, $\frac{1}{y_i}|\widehat{y}_i - y_i| > \lambda\%$, where we set $\lambda = 50$. A lower PCR indicates a more robust DES flow prediction performance.

(3) *(Un)Fairness*: We take into account the non-parity, value, absolute value, underestimation, and overestimation unfairness metrics in Sec. 5.2 for the final fairness evaluation (on sensitive socioeconomic attributes of social ethnicity, education and income). In particular, we find the average of all time intervals $k \in \{1, \ldots, K\}$ within the test data, *i.e.*, $U_x = \frac{1}{K}\sum_{k=1}^{K} U_x^{(k)}$, where the metric type $x \in \{\text{Par}, \text{Perf}\}$.

C OTHER RESULTS & DISCUSSION

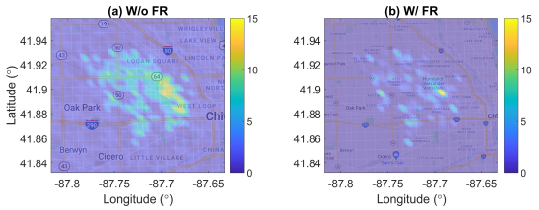

Fig. 12. Spatial non-parity unfairness without (a) and with (b) EIGDES's Regularization (FR).

• **Spatial Fairness Variations**: Taking Chicago as a typical example, we first compare in Fig. 12 the spatial heat-map distributions of non-parity (un)fairness (Eq. (18)) for a selected time interval without (a) and with (b) EIGDES's fairness regularization (FR). The warmer colors imply larger unfairness values. From the difference of the two heatmaps, we can observe that EIGDES substantially reduces the spatial unfairness, particularly around the regions of West Side (including Humboldt Park and West Town) for the disadvantaged communities in Chicago. Such reduction and the socially-equitable prediction results will serve as the basis for the city planning decision-makers.

• **Visualization of Mobility Fairness**: Besides the spatial unfairness, we show in Fig. 13 the temporal non-parity (un)fairness variation without and with EIGDES's regularization in Chicago. We can observe that when there exist large DES traffic volumes during daytime, the non-parity unfairness is large, demonstrating the inherent disparities within the DES distributions. EIGDES substantially reduces the unfairness values of the DES mobility prediction (by 87.97% on average), which can benefit the fairness-aware allocation of DES resource decisions in practice.

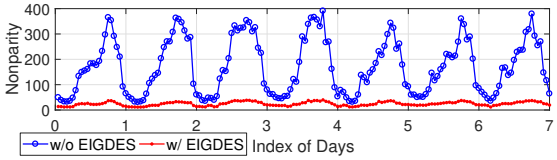

Fig. 13. Temporal unfairness without and with EIGDES.

• **Effect of λ**: We evaluate the RMSE, U_{Par}, and U_{Perf} of EIGDES versus λ (Eq. (22)) in Fig. 14. We can see that with small λ (say, $\lambda \leq 0.3$) EIGDES can achieve better accuracy and mobility fairness than $\lambda = 0$, thanks to the model regularization. With further increase of λ, the errors of EIGDES slightly enlarge with improvement in mobility fairness. Overall, EIGDES can maintain high prediction accuracy and mobility fairness, and we empirically set $\lambda = 0.6$.

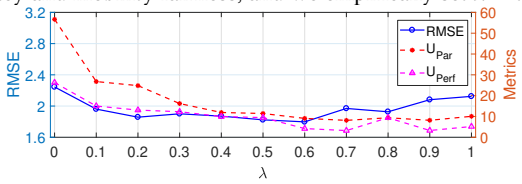

Fig. 14. Performance of EIGDES (RMSE and fairness) vs. λ.

• **Societal Implications**: As illustrated in Fig. 1, according to the EIGDES's predictions integrated with the WoT designs, the city planners as well as the DES service providers can interactively adjust the initial e-scooter deployment (re)distribution, particularly during the city mobility planning and pilot program studies. Furthermore, the e-scooter relocation/redistribution operation can be more fairly adapted towards the demand of diverse communities in the city. Support infrastructures, including geo-fences for parking, e-scooter lanes, and charging docks, can also be planned accordingly. We can extend fairness studies and model insights towards the vulnerable populations in US under the COVID-19 pandemic, enhancing the algorithm designs for the mobility-on-demand companies to augment these communities' access to pharmacy, grocery stores, hospitals, and many other life-essential urban services.

• **Redistribution of Transportation Resources**: Although we focused on DES flow prediction, our EIGDES can serve as an information input for mobility or transportation resource redistribution mechanisms, including but not limited to DES. For example, Lime can coordinate the Lime Juicers [1] to redistribute the scooters towards disadvantaged communities for commute purposes. Such mechanisms can also be extended to the operation of other web-based and WoT platforms [16, 21, 43, 52]. Redistribution may change the subsequent spatial fairness conditions, thus requiring continuous optimization and coordination as well as additional planning information, and we leave such studies as our future work.