

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Artificial Intelligence

www.elsevier.com/locate/artint

Time-delayed collective flow diffusion models for inferring latent people flow from aggregated data at limited locations [☆]



Yusuke Tanaka ^{a,c,*}, Tomoharu Iwata ^b, Takeshi Kurashima ^a, Hiroyuki Toda ^a, Naonori Ueda ^{b,d}, Toshiyuki Tanaka ^c

^a NTT Service Evolution Laboratories, NTT Corporation, 1-1 Hikari-no-oka, Yokosuka-Shi, Kanagawa, 239-0847, Japan

^b NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai, Seika-Cho, Soraku-gun, Kyoto, 619-0237, Japan

^c Graduate School of Informatics, Kyoto University, Yoshida Hon-machi, Sakyo-ku, Kyoto, 606-8501, Japan

^d RIKEN Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan

ARTICLE INFO

Article history:

Received 22 January 2020

Received in revised form 19 October 2020

Accepted 15 November 2020

Available online 20 November 2020

Keywords:

Collective graphical models

Travel duration

Aggregated population data

ABSTRACT

The rapid adoption of wireless sensor devices has made it easier to record location information of people in a variety of spaces (e.g., exhibition halls). Location information is often aggregated due to privacy and/or cost concerns. The aggregated data we use as input consist of the numbers of incoming and outgoing people at each location and at each time step. Since the aggregated data lack tracking information of individuals, determining the flow of people between locations is not straightforward. In this article, we address the problem of inferring latent people flows, that is, transition populations between locations, from just aggregated population data gathered from observed locations. Existing models assume that everyone is always in one of the observed locations at every time step; this, however, is an unrealistic assumption, because we do not always have a large enough number of sensor devices to cover the large-scale spaces targeted. To overcome this drawback, we propose a probabilistic model with flow conservation constraints that incorporate travel duration distributions between observed locations. To handle noisy settings, we adopt noisy observation models for the numbers of incoming and outgoing people, where the noise is regarded as a factor that may disturb flow conservation, e.g., people may appear in or disappear from the predefined space of interest. We develop an approximate expectation-maximization (EM) algorithm that simultaneously estimates transition populations and model parameters. Our experiments demonstrate the effectiveness of the proposed model on real-world datasets of pedestrian data in exhibition halls, bike trip data and taxi trip data in New York City.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Location information of people is now being gathered in various spaces such as exhibition halls, shopping malls, amusement parks, and urban cities. Understanding people's mobility patterns from location data is of critical importance in many fields including navigation systems [11], travel route recommendation [18], location-based mobile advertising [7], urban

[☆] A preliminary version of this work appeared in the Proceedings of IJCAI'18.

* Corresponding author at: NTT Service Evolution Laboratories, NTT Corporation, 1-1 Hikari-no-oka, Yokosuka-Shi, Kanagawa, 239-0847, Japan.
E-mail address: yusuke.tanaka.rh@hco.ntt.co.jp (Y. Tanaka).

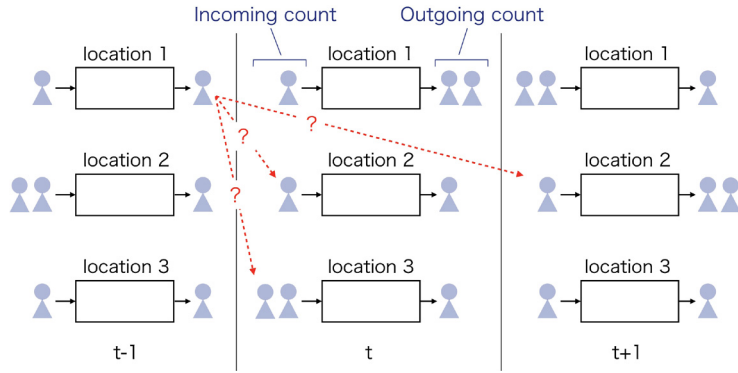


Fig. 1. Aggregated population data. Observed the numbers of incoming and outgoing people at each observed location and at each time step (blue). The next location of each individual is unobserved (red). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

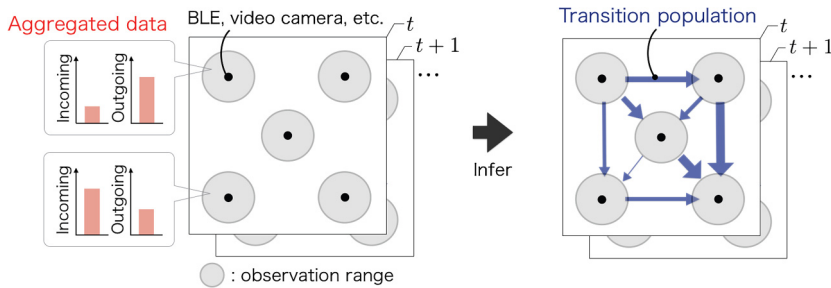


Fig. 2. The problem we focus on: Inferring latent people flows between locations from just aggregated population data. The left part represents the number of incoming and outgoing people at each location and at each time step, which we use as input data. The right part is the output to be estimated: Unobserved transition populations between locations.

planning [39], and disaster management [29]. For example, finding popular routes of people in exhibition halls yields better route recommendations. Knowing mobility patterns in shopping malls is useful for providing advertising information according to their current locations.

There are often cases where location tracking of individuals is difficult; aggregated population data are only available instead, whose two major reasons are as follows: Firstly, privacy concerns are recently increasing, location information is thus not only anonymized but also often aggregated [6]. Secondly, tracking data collection is too costly because one requires to prepare a software system for tracking users and to obtain the user's agreement for tracking user location. The typical response is to use people counting systems for directly obtaining aggregated population data. These systems are constructed by various sensors such as video cameras [5] and inductive-loop traffic detectors [16]; these are cost-effective and easy to install.

The aggregated population data considered in this study consist of the numbers of incoming and outgoing people at each observed location and at each time step. Fig. 1 shows an example of the aggregated data collected from three observed locations. For example, in an exhibition hall, we may be able to obtain the number of visitors entering or leaving each event booth over time via wireless technologies, e.g., Bluetooth Low Energy (BLE). Other examples include pedestrian data for each attraction in an amusement park [8], each store in a shopping mall [25], and traffic data from intersections [17]. In this article, given just aggregated data, we address the problem of inferring latent people flows, i.e., transition populations between locations, shown in Fig. 2. Solving this problem allows us to analyze mobility patterns while preserving privacy and/or saving data collection costs.

Recently, many methods based on deep neural networks have been developed for analyzing aggregated population data [10,37,38,40]. The problem addressed in these studies is related but essentially different from ours. The aim of these studies is to predict the future numbers of incoming and outgoing people at each observed location; thus, they cannot estimate the transition populations between locations from just the numbers of incoming and outgoing people. A DNN-based method has been developed for predicting transition populations [41]; however, the prediction function of the method is learned in the supervised setting, so that one has to collect the transition populations between locations for training the model. Our study, on the other hand, considers the problem of estimating the transition populations from just the numbers of incoming and outgoing people, where the estimation process can be achieved in the unsupervised setting; namely it does not require the training data for transition populations. The unsupervised setting we focus on is practically reasonable, because most of the aggregated datasets gathered by cost-effective ways (e.g., people counting systems) do not have transition information.

A similar problem has been tackled by collective flow diffusion models (CFDM), in which people flows are assumed to be probabilistic diffusion processes on a graph; nodes are locations and edges are paths between locations [17]. Several extensions of CFDM have recently been developed for estimating transition populations between locations from aggregated population data [1,8,13,14]. The basic idea of these models is to infer transition populations by maximizing a likelihood subject to constraints that represent people flow conservation. The constraints assume that any person who leaves a location at one time step always arrives at another location at the same time step. In other words, these models assume that everyone is always in one of the observed locations at every time step. This assumption, however, is too restrictive for many real-world settings, because we do not always have a large enough number of sensor devices to cover a large-scale space of interest: Each sensor device has an observation range in which people can be counted (depicted by the gray circles in Fig. 2), meaning that people might be moving through passages that are outside the ranges at an observation time step. In that case, the flow conservation constraints might not be satisfied, and thus CFDM fails to estimate transition populations between locations accurately.

We propose a new probabilistic model, called *Time-delayed Collective Flow Diffusion Models (T-CFDM)* hereinafter, that robustly infers transition populations between locations in more practical settings where the observation range of sensor devices is limited and some people are not observed in any location in some time periods. In T-CFDM, people are assumed to move from location to location according to transition probabilities that depend on their current locations. Since the transition populations are not observed, we treat them as latent variables. Our key idea is to design flow conservation constraints that incorporate travel duration distributions between locations. They aim to model the people in transit, that is, those who leave one location at one time step and arrive at another location after some delay; moreover, our modeling allows us to capture the heterogeneity in travel duration among individuals as it treats travel duration as a random variable that follows a probability distribution, namely it is not a point estimate. This is a critical advance in modeling the temporal dynamics of people flows. For example, in an exhibition hall, some people might promptly move to their next location, but others might take a rest before arriving at another location; in urban areas, moving speeds might depend on the means of transport used (e.g., walking, bike, car).

In practical situations, the flow conservation constraints might not strictly hold, because the observations are noisy. For example, people may enter or exit from the predefined space, implying that the total number of people is not constant. Another example is that sensing errors might be contained in the observations. To handle the noisy settings that may disturb flow conservation, we adopt noisy observation models for the numbers of incoming and outgoing people, as in [28]. In other words, we treat the flow conservation constraints as soft constraints. This allows determination of noise variance at respective locations via maximum likelihood estimation. We develop an approximate expectation-maximization (EM) algorithm that simultaneously estimates transition populations and model parameters.

The major contributions of this article include the following:

- We propose T-CFDM, a first model that incorporates travel duration distributions for robustly inferring transition populations from noisy aggregated population data gathered at limited locations. This contribution is significant in that it broadens the applicability of CFDM-based methods to many practical cases.
- We develop an efficient approximate EM algorithm to estimate transition probabilities, transition populations, travel duration distributions, and noise variances, simultaneously.
- We evaluate the effectiveness of the proposed model by using real-world datasets, pedestrian location logs from large-scale exhibition halls, and bike trip data and taxi trip data from New York City.

This article is organized as follows: In Section 2, we outline related work. Section 3 describes our problem of inferring latent people flows from just aggregated population data. In Section 4, we formulate the proposed model that incorporates travel duration distributions between locations. In Section 5, we present the approximate EM algorithm for learning model parameters. Section 6 demonstrates the effectiveness of the proposed model by using multiple real-world datasets. Finally, we present concluding remarks and a discussion of future work in Section 7. This article is an extended version of [32]. The main differences from [32] are described in Appendix A.

2. Related work

2.1. Collective graphical models

Collective graphical models (CGMs) [26] have been recently developed as a general framework for analyzing aggregated data. They have been applied to modeling contingency tables [26], and bird migration [27]. Prior works provide efficient inference techniques for CGMs based on the maximum a posteriori [28], MCMC sampler [26], message passing [31], variational Bayesian inference [14], and minimum convex cost flow algorithm [2]. Noisy observation models for aggregated data are introduced in [28,21,26]; they allow for treatments with noisy observation settings. Collective flow diffusion model (CFDM) [17] is the first application of the efficient inference techniques developed for CGMs to the transportation domain. By maximizing a likelihood subject to constraints that represent people flow conservation, this model can estimate transition populations between locations from just aggregated data [17,8]. Several recent works have attempted to estimate

people flows from spatiotemporal population data [14,1,13]. These models do not, however, consider people's travel duration between locations, and thus they might fail to estimate the transition populations in the case where the number of sensor devices is not sufficient to cover a large-scale space of interest and some people in transit are outside the observation range.

The proposed model is the extension of CFDM, and incorporates travel duration distributions into the constraints that represent flow conservation. This allows us to more accurately estimate the transition populations by considering the travel durations between observed locations even if we have just aggregated population data gathered at limited locations.

2.2. Information diffusion models

Information diffusion models are related to our proposal, because their main goal is to estimate latent flows of a piece of information over social networks [15,20]. Recent progress was to estimate the hidden network structure and the model parameters that govern the diffusion process of information [12,19,24,33]. These methods incorporate a continuous-time distribution to model the temporal dynamics of information flows, with the aim of describing the spread of information from node to node with delays over time. In the analysis of information flows, however, it is not necessary to consider flow conservation, a critical factor in modeling the temporal dynamics of people flows.

On the basis of the idea of information diffusion modeling, we have designed travel duration distributions between locations. By maximizing the likelihood with flow conservation constraints that incorporate the travel duration distributions, our model can infer latent people flows between locations from aggregated population data.

2.3. Urban computing

Many statistical methods have been developed for analyzing people's mobility patterns from their trajectory data gathered by mobile devices [9,22,18,39]. The common assumption is that we can obtain the histories of locations visited by each user; that is, they can be tracked individually over time. The location information of individuals is, however, often aggregated to protect privacy, and the resulting aggregated data do not contain the tracking information of individuals, making these existing methods inapplicable. Many works have been published recently that use the latest technologies, i.e., deep neural networks (DNNs), for analyzing aggregated population data such as the numbers of incoming and outgoing people at each location [10,37,38,40]. The problem of these works is, while considering various external factors (e.g., weather), to predict the future numbers of incoming and outgoing people, which is essentially different from the problem addressed in this article. In other words, these DNN-based methods cannot solve the problem of estimating the transition populations between locations from the numbers of incoming and outgoing people. Although a DNN-based method has been developed for predicting transition populations, it requires the transition populations for training in the supervised setting [41]. Our study aims to estimate the transition populations between locations in the unsupervised setting, namely it does not require the training data for transition populations.

The method of [36] tried to recover user trajectories given aggregated data and transition probabilities. Its drawback is that the transition probabilities must be manually set by using other information such as distances between locations: The method cannot estimate the transition populations between locations from just aggregated population data.

Different from the prior works in the field of urban computing, our model can estimate, from just the numbers of incoming and outgoing people, not only the transition populations but also transition probabilities. The estimation process can be achieved in the unsupervised setting, meaning that it does not require the training data for transition populations or the auxiliary information (e.g., distances). Moreover, our model enables us to estimate the travel duration distributions; this yields robust estimations even if the number of observed locations is limited.

3. Problem setting

In this section, we describe the aggregated population data considered here, i.e., the numbers of incoming and outgoing people at each location and at each time step, and define our problem of inferring latent people flows. The notations used in this article are listed in Table 1.

Aggregated population data. Let $t \in \{1, \dots, T\}$ be a discrete time step. Let \mathbf{V} denote a set of locations and $i \in \mathbf{V}$ denote a location. Let $Y_{ti}^{\text{out}} \in \mathbb{N}^0$ be the number of outgoing people of location i at time step t , where $\mathbb{N}^0 = \{0, 1, 2, \dots\}$ denotes the set of nonnegative integers, and $Y_{ti}^{\text{in}} \in \mathbb{N}^0$ be the number of incoming people of location i at time step t , where we assume that these flows are aggregated by each time interval $[t-1, t)$, namely people who left or entered a location during the interval $[t-1, t)$ are counted in the flows at time step t . Aggregated population data refer to the paired sets of the outgoing people counts $\mathbf{Y}^{\text{out}} = \{Y_{ti}^{\text{out}} \mid t = 1, \dots, T; i \in \mathbf{V}\}$ and of the incoming people counts $\mathbf{Y}^{\text{in}} = \{Y_{ti}^{\text{in}} \mid t = 1, \dots, T; i \in \mathbf{V}\}$.

Problem. Our problem is to infer latent people flows, i.e., transition populations between locations, from aggregated population data; this problem is illustrated schematically in Fig. 2. Let $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ be the undirected graph that represents accessibility information in such a way that the neighbor $\mathbf{E}_i = \{j \in \mathbf{V} \mid (ij) \in \mathbf{E}\}$ of location $i \in \mathbf{V}$ on \mathbf{G} represents the set of locations that are directly accessible from i . Let $M_{tij} \in \{0, \dots, Y_{ti}^{\text{out}}\}$ be the transition population, that is, the number of people who left location $i \in \mathbf{V}$ at time step t and whose next location is $j \in \mathbf{E}_i$. We define the set of transition populations as $\mathbf{M} = \{M_{tij} \mid t = 1, \dots, T; i \in \mathbf{V}; j \in \mathbf{E}_i\}$. Given $\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, T$ and \mathbf{G} , our goal is to infer transition populations \mathbf{M} .

Table 1
Notation.

Symbol	Description
T	number of time steps
t	time step, $t \in \{1, \dots, T\}$
\mathbf{V}	set of locations
\mathbf{G}	undirected graph, $\mathbf{G} = (\mathbf{V}, \mathbf{E})$
i, j	locations, $i, j \in \mathbf{V}$
\mathbf{E}_i	set of locations directly accessible from location i , $\mathbf{E}_i = \{j \in \mathbf{V} \mid (ij) \in \mathbf{E}\}$
Y_{ti}^{out}	number of outgoing people at location i and at time step t , $Y_{ti}^{\text{out}} \in \mathbb{N}^0$
Y_{ti}^{in}	number of incoming people at location i and at time step t , $Y_{ti}^{\text{in}} \in \mathbb{N}^0$
N_{ti}^{out}	noise-free variable for Y_{ti}^{out}
N_{ti}^{in}	noise-free variable for Y_{ti}^{in}
M_{tij}	transition population leaving location i at time step t and whose next location is j , $M_{tij} \in \{0, \dots, Y_{ti}^{\text{out}}\}$
θ_{ij}	transition probability that people move from location i to j , $\theta_{ij} \geq 0$, $\sum_{j \in \mathbf{E}_i} \theta_{ij} = 1$
$\boldsymbol{\gamma}_{ij}$	set of parameters of the travel duration distribution from location i to j
σ_i^2	noise variance for the number of outgoing people at location i
λ_i^2	noise variance for the number of incoming people at location i

4. Model

We propose T-CFDM (Time-delayed Collective Flow Diffusion Models), a probabilistic model for inferring latent people flows, i.e., transition populations between locations, from aggregated population data gathered at limited locations. Note that we use just incoming and outgoing people counts as inputs, meaning that we attempt to infer transition populations in the unsupervised setting. We first model the temporal dynamics of people flows in the noise-free setting, that is, the numbers of incoming and outgoing people are strictly preserved. Travel duration distributions between locations are incorporated into the flow conservation constraints. This modeling is especially advantageous in such a situation that the sensor devices have limited observation range and some people are not observed in any location in some time periods. We then introduce the noisy observation models for the numbers of incoming and outgoing people; they allow us to treat the flow conservation constraints as soft constraints and thus handle noisy settings.

In T-CFDM, the temporal dynamics of people flows are assumed to be probabilistic diffusion processes on a graph, where the nodes are locations and the edges are paths between locations. Even if the graph structure such as a road network or a set of neighbor information is unknown, our model is still applicable by assuming a complete graph among locations. We assume that people move from location to location in accordance with location-dependent and time-independent transition probabilities. Let $\theta_{ij} \geq 0$ be the transition probability that people move from location i to j , where $\sum_{j \in \mathbf{E}_i} \theta_{ij} = 1$. Let N_{ti}^{out} and N_{ti}^{in} denote noise-free variables for Y_{ti}^{out} and Y_{ti}^{in} , respectively. We assume that $\mathbf{M}_{ti} = \{M_{tij} \mid j \in \mathbf{E}_i\}$ follows a multinomial distribution,

$$p(\mathbf{M}_{ti} \mid N_{ti}^{\text{out}}, \boldsymbol{\theta}_i) = \frac{N_{ti}^{\text{out}}!}{\prod_{j \in \mathbf{E}_i} M_{tij}!} \prod_{j \in \mathbf{E}_i} \theta_{ij}^{M_{tij}}, \quad (1)$$

where $\boldsymbol{\theta}_i = \{\theta_{ij} \mid j \in \mathbf{E}_i\}$. The transition population M_{tij} satisfies the following two relations to ensure flow conservation:

$$N_{ti}^{\text{out}} = \sum_{j \in \mathbf{E}_i} M_{tij}, \quad (2)$$

$$N_{ti}^{\text{in}} = \sum_{j \in \mathbf{E}_i} \sum_{t'=1}^t F(\Delta_{tt'}; \boldsymbol{\gamma}_{ji}) M_{t'ji}. \quad (3)$$

The conservation for outgoing counts (2) indicates that the sum of people leaving location i at time step t equals the outgoing count at the same time step. The conservation for incoming counts (3) indicates that the weighted sum of people leaving location j before time step t toward location i equals the incoming count of location i at time step t ; the idea behind (3) is that people who leave one location in one time step arrive at another location after some delay. Here, $\Delta_{tt'} = t - t'$ is travel duration, where t and t' are the arriving and leaving time steps, respectively. Weight function $F(\Delta; \boldsymbol{\gamma}_{ji})$ is a travel duration probability where $\Delta \in \{0, \dots, T-1\}$, which is the probability that the travel duration is Δ ; $\boldsymbol{\gamma}_{ji}$ denotes its parameters for transitions from location j to i , which are assumed time-invariant, that is, the travel duration distributions do not vary in time. Details of $F(\Delta; \boldsymbol{\gamma}_{ji})$ are given in the following paragraphs. In our modeling, the travel duration is treated as a random variable that follows the probability distribution; it is not a point estimate. This allows us to handle the heterogeneity in travel duration among individuals.

Table 2
Travel duration probabilities.

	Travel duration distribution $f(\tau; \boldsymbol{\gamma}_{ji})$	Travel duration probability $F(\Delta; \boldsymbol{\gamma}_{ji})$
Exponential	$\begin{cases} \alpha_{ji} e^{-\alpha_{ji} \tau} & \text{if } 0 \leq \tau \\ 0 & \text{otherwise} \end{cases}$	$e^{-\alpha_{ji} \Delta} - e^{-\alpha_{ji} (\Delta+1)}$
Rayleigh	$\begin{cases} \alpha_{ji} \tau e^{-\frac{1}{2} \alpha_{ji} \tau^2} & \text{if } 0 \leq \tau \\ 0 & \text{otherwise} \end{cases}$	$e^{-\frac{1}{2} \alpha_{ji} (\Delta)^2} - e^{-\frac{1}{2} \alpha_{ji} (\Delta+1)^2}$
Weibull	$\begin{cases} \alpha_{ji} \beta_{ji} (\alpha_{ji} \tau)^{\beta_{ji}-1} e^{-(\alpha_{ji} \tau)^{\beta_{ji}}} & \text{if } 0 \leq \tau \\ 0 & \text{otherwise} \end{cases}$	$e^{-(\alpha_{ji} \Delta)^{\beta_{ji}}} - e^{-(\alpha_{ji} (\Delta+1))^{\beta_{ji}}}$

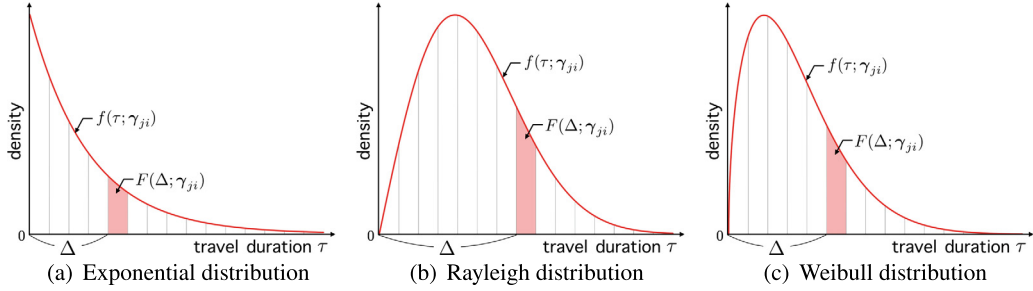


Fig. 3. Visualization of travel duration probabilities. The red lines are the travel duration distributions $f(\tau; \boldsymbol{\gamma}_{ji})$, and the red areas are the travel duration probabilities $F(\Delta; \boldsymbol{\gamma}_{ji})$. The gray vertical lines are spaced by unit time steps.

To derive the travel duration probability $F(\Delta; \boldsymbol{\gamma}_{ji})$, we first introduce continuous travel duration distribution $f(\tau; \boldsymbol{\gamma}_{ji})$ as the probability density function of continuous travel duration $\tau \geq 0$. The travel duration probability $F(\Delta; \boldsymbol{\gamma}_{ji})$ is then calculated by the following integral of $f(\tau; \boldsymbol{\gamma}_{ji})$ over the interval from Δ to $\Delta + 1$:

$$F(\Delta; \boldsymbol{\gamma}_{ji}) = \int_{\Delta}^{\Delta+1} f(\tau; \boldsymbol{\gamma}_{ji}) d\tau. \tag{4}$$

Notice that our model does not depend on the particular choice of the travel duration distribution, so that one can use any distribution as $f(\tau; \boldsymbol{\gamma}_{ji})$. Table 2 summarizes the travel duration distributions and the travel duration probabilities when using exponential, Rayleigh, and Weibull distributions, which are widely used for assessing duration in various fields such as user modeling [34] and diffusion modeling [24,19]. The exponential and Rayleigh distributions are one-parameter distributions, where we set $\boldsymbol{\gamma}_{ji} = \alpha_{ji}$ and $\alpha_{ji} > 0$. The Weibull distribution is a more flexible distribution that has two parameters, where we set $\boldsymbol{\gamma}_{ji} = \{\alpha_{ji}, \beta_{ji}\}$. Here, $\alpha_{ji} > 0$ is the scale parameter and $\beta_{ji} > 0$ is the shape parameter of the distribution. The exponential distribution is a special case of the Weibull distribution with $\beta_{ji} = 1$; the Weibull distribution with $\beta_{ji} = 2$ is equivalent to the Rayleigh distribution. Fig. 3 illustrates the examples of using the different distributions.

In practical situations, the flow conservation constraints (2) and (3) might not strictly hold, because the observations are noisy. The noise can be regarded as factors that may disturb flow conservation, for example, people may enter or exit from the predefined space, e.g., an exhibition hall, that is being targeted. Another example is that sensing errors might be contained in the observations. To handle noisy settings, we adopt noisy observation models for the numbers of incoming and outgoing people in a manner similar to that described in [28,21,26]; namely we assume that the observed outgoing count Y_{ti}^{out} and the observed incoming count Y_{ti}^{in} are random variables dependent on their noiseless counterparts N_{ti}^{out} and N_{ti}^{in} , respectively. This allows us to treat the flow conservation constraints as soft constraints. Although the use of Poisson distributions with means of N_{ti}^{out} and N_{ti}^{in} would be the most natural option since Y_{ti}^{out} and Y_{ti}^{in} are count data, it makes it difficult to adjust the strength of the soft constraints, for the variances of the Poisson distributions are determined by their means. Another option is to use the following Gaussian distributions by regarding Y_{ti}^{out} and Y_{ti}^{in} as real numbers,

$$p(Y_{ti}^{\text{out}} | N_{ti}^{\text{out}}, \sigma_i^2) = \frac{1}{\sqrt{2\pi \sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2} (Y_{ti}^{\text{out}} - N_{ti}^{\text{out}})^2\right), \tag{5}$$

$$p(Y_{ti}^{\text{in}} | N_{ti}^{\text{in}}, \lambda_i^2) = \frac{1}{\sqrt{2\pi \lambda_i^2}} \exp\left(-\frac{1}{2\lambda_i^2} (Y_{ti}^{\text{in}} - N_{ti}^{\text{in}})^2\right), \tag{6}$$

Algorithm 1: Inference procedure for the proposed model.

Input : $\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \mathbf{V}, \{\mathbf{E}_i \mid i \in \mathbf{V}\}, T$
Output: $\mathbf{M}, \Theta, \Gamma, \Sigma, \Lambda$
1: Initialize $\mathbf{M}, \Theta, \Gamma, \Sigma, \Lambda$
2: **repeat**
3: /* E-step */
4: Update \mathbf{M} by solving (11)
5: /* M-step */
6: Update θ_{ij} by (13) for $i \in \mathbf{V}$; for $j \in \mathbf{E}_i$
7: Update Γ, Σ, Λ by solving (14)
8: **until** Convergence

where $\sigma_i^2 > 0$ and $\lambda_i^2 > 0$ are noise variances for outgoing and incoming count, respectively, at location i . These variances play a role in controlling the penalty incurred in violating the soft constraints. This option is also advantageous in that we can use continuous optimization methods to estimate the transition populations. Details are given in Section 5. In the following, we use the Gaussian distributions as the noise models.

Another way of handling noisy observations was introduced in [14,11,32]; it consists of the following procedures: (a) assume the noise-free model as in (1); (b) construct the objective function with penalty terms that are the squared difference between the left- and right-hand sides in each of (2) and (3); (c) determine a hyperparameter for controlling the penalty terms by a validation procedure. A drawback of this approach is that the validation procedure in (c) is time-consuming; to reduce the computation time the hyperparameter should be shared among all the locations even though the noise variances might be different among locations. Different from this approach, the noisy observation models considered here are cost-effective and allow us to determine the noise variances at respective locations via maximum likelihood estimation.

We summarize the parameters of T-CFDM as follows: Transition probabilities $\Theta = \{\theta_i \mid i \in \mathbf{V}\}$; parameters of travel duration distributions $\Gamma = \{\gamma_{ij} \mid i \in \mathbf{V}; j \in \mathbf{E}_i\}$; noise variances for outgoing counts $\Sigma = \{\sigma_i \mid i \in \mathbf{V}\}$; noise variances for incoming counts $\Lambda = \{\lambda_i \mid i \in \mathbf{V}\}$.

5. Inference

Overview. We develop an approximate expectation-maximization (EM) algorithm for simultaneously estimating the transition populations \mathbf{M} between locations and the set of parameters $\Phi = \{\Theta, \Gamma, \Sigma, \Lambda\}$. Following the standard prescription of the EM algorithm [3], we define the Q-function

$$Q(\hat{\Phi}, \Phi) = \sum_{\mathbf{M}} p(\mathbf{M} \mid \mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \hat{\Phi}) \log p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \mathbf{M} \mid \Phi), \quad (7)$$

which is the logarithm of the complete-data likelihood of Φ , averaged with respect to the posterior distribution of \mathbf{M} conditional on data $\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}$ and evaluated using the current parameter estimate $\hat{\Phi}$. Summing up over all possible values of \mathbf{M} in (7) is infeasible due to computation costs, and thus we replace the posterior expectation with a *plug-in* procedure of the maximum a posteriori (MAP) estimate: The approximate Q-function is obtained by plugging the MAP estimate $\hat{\mathbf{M}}$ of \mathbf{M} given by

$$\hat{\mathbf{M}} = \arg \max_{\mathbf{M}} \log p(\mathbf{M} \mid \mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \hat{\Phi}) \quad (8)$$

into the complete-data log likelihood, as

$$Q^{\text{approx}}(\hat{\Phi}, \Phi) = \log p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \hat{\mathbf{M}} \mid \Phi). \quad (9)$$

Note that, although we adopt the shorthand notation $\hat{\mathbf{M}}$ for $\hat{\mathbf{M}}(\hat{\Phi})$ here and hereafter, $\hat{\mathbf{M}}$ actually depends on $\hat{\Phi}$ via (8). A similar approximate EM algorithm was introduced in [28], and their computational experiments demonstrated that it provides an efficient and effective way of estimating the transition populations between locations and the model parameters. We obtain the MAP estimates of \mathbf{M} in E-step, and obtain the maximum likelihood estimates of Φ by maximizing $Q^{\text{approx}}(\hat{\Phi}, \Phi)$ in M-step. Our inference procedure is shown in Algorithm 1. Discussions of the approximate EM algorithm and its validity are included in Appendix B. Given a complete graph among locations, the computational complexity of the inference algorithm is $O(T^2|\mathbf{V}|^2)$ for each iteration, where $|\mathbf{V}|$ is the number of locations. Although the complexity drastically increases when T and/or $|\mathbf{V}|$ is large, in practice, one can deal with the large-scale problems by using the following procedures: (a) ignore transition with long durations; (b) incorporate neighbor information from a road network. Then, its complexity can be reduced to $O(T\bar{T}|\mathbf{V}||\bar{\mathbf{E}}|)$, where $\bar{T} (\ll T)$ is a period long enough to complete the transition, and where $|\bar{\mathbf{E}}| (\ll |\mathbf{V}|)$ is the average number of neighbor locations. Details of E- and M-steps are described in the following paragraphs.

E-step. Suppose that the current estimate $\hat{\Phi}$ of the parameters Φ is given. Using Bayes' rule, the log-posterior distribution of \mathbf{M} is given by

$$\begin{aligned}
 & \log p(\mathbf{M} \mid \mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \hat{\Phi}) \\
 & \propto \log p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \mathbf{M} \mid \hat{\Phi}) \\
 & \approx \sum_{t=1}^T \sum_{i \in \mathbf{V}} \left[-\frac{1}{2\hat{\sigma}_i^2} \left(Y_{ti}^{\text{out}} - \sum_{j \in \mathbf{E}_i} M_{tij} \right)^2 - \frac{1}{2\hat{\lambda}_i^2} \left(Y_{ti}^{\text{in}} - \sum_{j \in \mathbf{E}_i} \sum_{t'=1}^t F(\Delta_{tt'}; \hat{\gamma}_{ji}) M_{t'ji} \right)^2 \right. \\
 & \quad \left. + \left(\sum_{j \in \mathbf{E}_i} M_{tij} \right) \log \left(\sum_{j \in \mathbf{E}_i} M_{tij} \right) + \sum_{j \in \mathbf{E}_i} \left(M_{tij} \log \hat{\theta}_{ij} - M_{tij} \log M_{tij} \right) \right] \\
 & =: \mathcal{L}(\mathbf{M}),
 \end{aligned} \tag{10}$$

where we use the relations (2) and (3); we employ Stirling’s approximation, $\log n! \approx n \log n - n$, in order to calculate $\log M_{tij}!$ and $(\sum_{j \in \mathbf{E}_i} M_{tij})!$ efficiently, as in [28]. Here, we define the objective function for \mathbf{M} as $\mathcal{L}(\mathbf{M})$. The optimization problem to be solved to determine $\hat{\mathbf{M}}$ is as follows:

$$\begin{aligned}
 & \underset{\mathbf{M}}{\text{maximize}} \quad \mathcal{L}(\mathbf{M}) \\
 & \text{subject to} \quad M_{tij} \geq 0, \quad t = 1, \dots, T; \quad i \in \mathbf{V}; \quad j \in \mathbf{E}_i,
 \end{aligned} \tag{11}$$

where we use a continuous relaxation from $M_{tij} \in \{0, 1, \dots, Y_{ti}^{\text{out}}\}$ to $M_{tij} \geq 0$, which enables us to use various continuous optimization methods. This article solves the optimization problem by using the L-BFGS-B method [4]. The derivative of the objective function $\mathcal{L}(\mathbf{M})$ (10) with respect to M_{tij} is described in Appendix C.

M-step. Given the current MAP estimate $\hat{\mathbf{M}}$ of \mathbf{M} , the approximate Q-function (9) is as follows:

$$\begin{aligned}
 Q^{\text{approx}}(\hat{\Phi}, \Phi) & = \log p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \hat{\mathbf{M}} \mid \Phi) \\
 & \propto \sum_{t=1}^T \sum_{i \in \mathbf{V}} \left[-\frac{1}{2} \log \sigma_i^2 - \frac{1}{2} \sigma_i^2 \left(Y_{ti}^{\text{out}} - \sum_{j \in \mathbf{E}_i} \hat{M}_{tij} \right)^2 - \frac{1}{2} \log \lambda_i^2 \right. \\
 & \quad \left. - \frac{1}{2\lambda_i^2} \left(Y_{ti}^{\text{in}} - \sum_{j \in \mathbf{E}_i} \sum_{t'=1}^t F(\Delta_{tt'}; \gamma_{ji}) \hat{M}_{t'ji} \right)^2 + \sum_{j \in \mathbf{E}_i} \hat{M}_{tij} \log \theta_{ij} \right] \\
 & =: \mathcal{J}(\Theta, \Gamma, \Sigma, \Lambda),
 \end{aligned} \tag{12}$$

where the objective function for the model parameters is defined as $\mathcal{J}(\Theta, \Gamma, \Sigma, \Lambda)$. The maximum likelihood estimate of θ_{ij} is explicitly given by the following closed-form solution via the Lagrangian multiplier method,

$$\theta_{ij} = \frac{\sum_{t=1}^T \hat{M}_{tij}}{\sum_{t=1}^T \sum_{j \in \mathbf{E}_i} \hat{M}_{tij}}. \tag{13}$$

When we use the Weibull distribution as the travel duration distribution, which has two parameters, i.e., $\gamma_{ji} = \{\alpha_{ji}, \beta_{ji}\}$, for each pair of locations, the optimization problem for parameters Γ , Σ , and Λ is as follows:

$$\begin{aligned}
 & \underset{\Gamma, \Sigma, \Lambda}{\text{maximize}} \quad \mathcal{J}(\Theta, \Gamma, \Sigma, \Lambda) \\
 & \text{subject to} \quad \alpha_{ji} > 0, \quad j \in \mathbf{V}; \quad i \in \mathbf{E}_j, \\
 & \quad \quad \quad \beta_{ji} > 0, \quad j \in \mathbf{V}; \quad i \in \mathbf{E}_j,
 \end{aligned} \tag{14}$$

which we solve by using the L-BFGS-B method [4].

6. Experiments

6.1. Data

We evaluated the proposed model using real-world datasets: Pedestrian data from exhibition halls, and bike trip data and taxi trip data from New York City. Details of the datasets are shown in subsequent paragraphs.

Pedestrian data from exhibition halls. The data consist of pedestrian location logs, which were collected at an event that attracted large crowds, Niconico Chokaigi 2016,¹ held at Makuhari Messe located near Tokyo, Japan, from 10:00 a.m. to 6:00 p.m. on April 29th, 2016. The event was spread over four exhibition halls, Hall 1, Hall 2, Hall 3, and Hall 4 with sizes

¹ <http://www.chokaigi.jp/2016/en/>.

Table 3
The totals of incoming and outgoing people count in the real-world datasets.

Data		Pedestrian data							
Time-of-day	10:00 a.m.–2:00 p.m.				2:00 p.m.–6:00 p.m.				
Area	Hall 1	Hall 2	Hall 3	Hall 4	Hall 1	Hall 2	Hall 3	Hall 4	
# outgoing	19,667	20,957	8,605	6,531	11,139	14,069	6,965	4,533	
# incoming	19,840	21,198	8,770	6,630	10,974	13,829	6,806	4,437	

Data		Bike trip data				Taxi trip data			
Time-of-day	8:00 a.m.–4:00 p.m.		4:00 p.m.–12:00 p.m.		8:00 a.m.–4:00 p.m.		4:00 p.m.–12:00 p.m.		
Date	Mar. 1	Jun. 1	Mar. 1	Jun. 1	Mar. 1	Jun. 1	Mar. 1	Jun. 1	
# outgoing	7,275	13,277	7,091	14,931	93,514	91,911	107,923	104,188	
# incoming	7,361	13,398	7,279	15,372	93,839	91,523	100,632	107,701	

of $186.3 \text{ m} \times 124.8 \text{ m}$, $183.2 \text{ m} \times 124.8 \text{ m}$, $127.6 \text{ m} \times 124.8 \text{ m}$, and $190.9 \text{ m} \times 108.3 \text{ m}$, respectively. The number of event booths, $|\mathbf{V}|$, in Hall 1, Hall 2, Hall 3, and Hall 4 were 38, 27, 10, and 9, respectively. We analyze data for these four halls separately, so that we do not consider edges across halls but those within each hall only. \mathbf{E}_i is a set of edges assuming a complete graph for each of the exhibition halls. We gathered pedestrian location logs by placing Bluetooth beacons at each booth. The technology allows us to observe the times at which each user entered or left the observation range (at most 10–15 meters from each beacon). The data consist of 3,727 mobile users who agreed to provide detailed information of location over time. The original data contained time stamps of arrival and departure at booths for each user, allowing users to be tracked over time. The users enter and leave each exhibition hall at any time step, making aggregated population data noisy. In our experiments, we created aggregated incoming and outgoing count data at each booth, where the time interval was set to 3 minutes. Since the tendency of people flows, i.e., transition probabilities, could depend on time-of-day, the data were divided into two subsets, one from 10:00 a.m. to 2:00 p.m., and the other from 2:00 p.m. to 6:00 p.m.; the number of observation time steps T was thus 80. Note that the user tracking information was used only for evaluating the estimation performance for transition populations and travel duration probabilities; we did not use the tracking information in the inference process. The totals of incoming and outgoing pedestrians at all booths in each hall/time-of-day are shown in Table 3.

Traffic data in urban areas. To validate the performance of our model, we used two open datasets, bike trip data² and taxi trip data³ in New York City. These datasets consist of trip records holding Trip id, pickup location, dropoff location, pickup date and time, and dropoff date and time. Note that location information was available only when people started and finished their trips: The trajectories during the trips were not recorded. In our experiments, we used the data from 8:00 a.m. to 4:00 p.m. and from 4:00 p.m. to 12:00 p.m. on March 1st and June 1st, and aggregated the data into incoming and outgoing people counts at grid cells, the sizes of which in the bike trip data and the taxi trip data were $2 \text{ km} \times 2 \text{ km}$ (12×12 grid cells) and $3 \text{ km} \times 3 \text{ km}$ (18×18 grid cells), respectively. Here, we omitted grid cells if their incoming and outgoing counts were lower than a threshold; the resulting set \mathbf{V} of locations for the bike trip data and the taxi trip data consisted of 11 and 14 grid cells, respectively. \mathbf{E}_i was a set of edges assuming a complete graph for all datasets. The time interval in both datasets was set to 10 minutes; the number of observation time steps T was thus 48. The totals of incoming and outgoing bikes/taxis at overall grid cells on each date/time-of-day are shown in Table 3. In the original data, a pair of pickup and dropoff is completely given; however, either pickup or dropoff may be missing near the edges of observation periods as the data are divided into the subsets by time-of-day. Then, the resulting aggregated data are noisy.

6.2. Transition population estimation

We evaluated the proposed model, T-CFDM, in terms of the performance in estimating the transition populations \mathbf{M} . We compared T-CFDM with the collective flow diffusion model (CFDM) [17]. Unlike our model, CFDM does not consider travel duration between locations. In addition, we compared our model with the following two baselines: *Popularity* and *Uniform*. Popularity assumes that people move to other locations in proportion to location popularity regardless of current locations; then the estimated transition population \hat{M}_{tij} at time step t is given by

$$\hat{M}_{tij} = Y_{ti}^{\text{out}} \times \frac{\sum_{t=1}^T Y_{tj}^{\text{in}}}{\sum_{t=1}^T \sum_{j \in \mathbf{E}_i} Y_{tj}^{\text{in}}}. \quad (15)$$

Uniform uses a discrete uniform distribution to estimate transition populations; it assumes that people move to neighbor locations with equal probability $1/|\mathbf{E}_i|$, where $|\mathbf{E}_i|$ is the number of neighbors of location i . In T-CFDM, we used the Weibull

² <https://www.citibikenyc.com/system-data>.

³ http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.

Table 4

Mean normalized absolute errors and standard errors of the estimates of transition populations. The double star (**) indicates that the differences between T-CFDM and the other methods are statistically significant (Student's t-test) at the level of $P < 0.05$. The single star (*) indicates that the differences between T-CFDM and the others, except for T-CFDM (validation), are statistically significant.

(a) Pedestrian data (10:00 a.m.–2:00 p.m.)				
	Hall 1	Hall 2	Hall 3	Hall 4
T-CFDM	1.174 ± 0.016*	0.995 ± 0.013**	0.689 ± 0.014**	0.600 ± 0.022*
T-CFDM (validation)	1.190 ± 0.015	1.056 ± 0.013	0.798 ± 0.013	0.624 ± 0.021
CFDM	1.261 ± 0.009	1.127 ± 0.010	0.864 ± 0.013	0.672 ± 0.022
Popularity	1.674 ± 0.008	1.500 ± 0.008	1.106 ± 0.016	0.779 ± 0.020
Uniform	1.767 ± 0.007	1.596 ± 0.007	1.130 ± 0.015	1.165 ± 0.016
(b) Pedestrian data (2:00 p.m.–6:00 p.m.)				
	Hall 1	Hall 2	Hall 3	Hall 4
T-CFDM	1.439 ± 0.026*	1.436 ± 0.010*	0.908 ± 0.022**	0.743 ± 0.043*
T-CFDM (validation)	1.445 ± 0.022	1.476 ± 0.011	0.922 ± 0.029	0.759 ± 0.043
CFDM	1.506 ± 0.025	1.515 ± 0.010	0.995 ± 0.026	0.780 ± 0.045
Popularity	1.825 ± 0.017	1.789 ± 0.010	1.253 ± 0.029	0.924 ± 0.038
Uniform	1.904 ± 0.016	1.857 ± 0.010	1.289 ± 0.029	1.282 ± 0.038
(c) Bike trip data				
	8:00 a.m.–4:00 p.m.		4:00 p.m.–12:00 p.m.	
	Mar. 1	Jun. 1	Mar. 1	Jun. 1
T-CFDM	0.561 ± 0.013**	0.630 ± 0.010**	0.628 ± 0.031*	0.590 ± 0.027*
T-CFDM (validation)	0.612 ± 0.012	0.677 ± 0.012	0.655 ± 0.029	0.637 ± 0.024
CFDM	0.655 ± 0.012	0.737 ± 0.012	0.706 ± 0.029	0.651 ± 0.027
Popularity	0.691 ± 0.013	0.751 ± 0.012	0.739 ± 0.028	0.671 ± 0.027
Uniform	1.025 ± 0.009	0.953 ± 0.013	1.036 ± 0.025	0.943 ± 0.025
(d) Taxi trip data				
	8:00 a.m.–4:00 p.m.		4:00 p.m.–12:00 p.m.	
	Mar. 1	Jun. 1	Mar. 1	Jun. 1
T-CFDM	0.363 ± 0.004**	0.433 ± 0.004**	0.429 ± 0.022*	0.366 ± 0.012**
T-CFDM (validation)	0.413 ± 0.004	0.450 ± 0.004	0.431 ± 0.016	0.402 ± 0.012
CFDM	0.472 ± 0.004	0.496 ± 0.004	0.469 ± 0.059	0.448 ± 0.012
Popularity	0.505 ± 0.004	0.520 ± 0.004	0.477 ± 0.016	0.493 ± 0.012
Uniform	1.043 ± 0.005	1.043 ± 0.005	0.923 ± 0.014	0.929 ± 0.010

distribution for modeling travel duration distribution as shown in Section 4. We consider two ways of handling noisy observations. The first one is the validation approach described in the fifth paragraph of Section 4, which we call *T-CFDM (validation)*. Since the validation procedure is time-consuming, the hyperparameter was shared among all the locations. The second one is to use the noisy observation models, which we call *T-CFDM*. This one is beneficial in that the parameter (i.e., the noise variance) for each location can be efficiently obtained by maximum likelihood estimation. The evaluation metric is the mean normalized absolute error (MNAE) in transition populations given by:

$$\frac{1}{T} \sum_{t=1}^T \frac{\sum_{i \in \mathbf{V}} \sum_{j \in \mathbf{E}_i} |\hat{M}_{tij} - M_{tij}^*|}{\sum_{i \in \mathbf{V}} \sum_{j \in \mathbf{E}_i} M_{tij}^*}, \quad (16)$$

where M_{tij}^* is the true transition population. As described in Section 6.1, the original data actually have the detailed tracking information of each user, and that is why we can use the noise-free data M_{tij}^* in the evaluation process.

Table 4 shows MNAE and the standard error for T-CFDM, T-CFDM (validation), CFDM, Popularity and Uniform. For all datasets, T-CFDM performed better than the other methods. We found similar results using other evaluation metrics (e.g., MAE, RMSE). Since the proposed model considers the travel duration distribution between locations, it can more accurately estimate the transition populations than CFDM. We can also see that T-CFDM matched or bettered the estimation performance of T-CFDM (validation). These results indicate that the formulation based on the noisy observation models is effective for accurately estimating the transition populations. In the following, we present the results of further assessment of T-CFDM.

Table 5
Mean Kullback-Leibler divergence for the estimation of travel duration probability.

(a) Pedestrian data (10:00 a.m.–2:00 p.m.)				
	Hall 1	Hall 2	Hall 3	Hall 4
T-CFDM	1.567 ± 0.050	1.547 ± 0.050	1.627 ± 0.087	1.669 ± 0.231
CFDM	1.800 ± 0.091	1.695 ± 0.082	1.675 ± 0.122	1.748 ± 0.127
(b) Pedestrian data (2:00 p.m.–6:00 p.m.)				
	Hall 1	Hall 2	Hall 3	Hall 4
T-CFDM	1.583 ± 0.051	1.560 ± 0.053	1.613 ± 0.071	1.560 ± 0.125
CFDM	1.758 ± 0.078	1.737 ± 0.103	1.670 ± 0.087	1.854 ± 0.082
(c) Bike trip data				
	8:00 a.m.–4:00 p.m.		4:00 p.m.–12:00 p.m.	
	Mar. 1	Jun. 1	Mar. 1	Jun. 1
T-CFDM	1.489 ± 0.194	1.353 ± 0.172	1.498 ± 0.218	1.182 ± 3.129
CFDM	2.936 ± 0.048	3.052 ± 0.047	2.942 ± 0.056	3.129 ± 0.043
(d) Taxi trip data				
	8:00 a.m.–4:00 p.m.		4:00 p.m.–12:00 p.m.	
	Mar. 1	Jun. 1	Mar. 1	Jun. 1
T-CFDM	1.432 ± 0.142	2.444 ± 0.135	2.060 ± 0.161	2.130 ± 0.102
CFDM	3.123 ± 0.058	3.069 ± 0.049	3.044 ± 0.072	3.062 ± 0.058

6.3. Travel duration probability estimation

We evaluated performance of estimating travel duration probabilities. We used mean Kullback-Leibler (KL) divergence between the true and the estimated travel duration probabilities over respective pairs of locations as the evaluation metric:

$$\frac{1}{|\mathbf{V}|} \sum_{j \in \mathbf{V}} \frac{1}{|\mathbf{E}_j|} \sum_{i \in \mathbf{E}_j} \sum_{\Delta=0}^{T-1} P_{ji}^*(\Delta) \log \frac{P_{ji}^*(\Delta)}{F(\Delta; \hat{\mathbf{y}}_{ji})}, \tag{17}$$

where $P_{ji}^*(\Delta)$ and $F(\Delta; \hat{\mathbf{y}}_{ji})$ are the true travel duration probability and its estimate, respectively, for transition from location j to location i . Table 5 shows the mean KL divergence and the standard error for T-CFDM and CFDM. In CFDM, the travel duration probability equals 1 if $\Delta = 0$ and 0 otherwise; CFDM assumes that all people who leave a location at a time step have always arrived at another location at the same time step. As shown, T-CFDM achieved lower mean KL divergence values for all datasets. The results show that T-CFDM accurately estimated the travel duration probabilities. The performance improvements of the bike trip data and the taxi trip data were larger than those of the pedestrian data. The results are reasonable because the distances between observed locations are relatively large in the urban traffic data compared with the pedestrian data in the exhibition halls; in such cases, introduction of the travel duration probability is more helpful. The results of Tables 4 and 5 indicate that the incorporation of the people’s travel durations into the model is important for estimating the transition populations accurately.

Fig. 4 illustrates visualization examples of transition matrices and travel duration probabilities estimated by T-CFDM for the respective times-of-day in the bike trip data and the taxi trip data. The transition matrix is the total number of bikes that moved between each pair of origin location i and destination location j ; its elements were calculated as follows: $M_{ij} = \sum_{t=1}^T M_{rij}$. As shown in Fig. 4, T-CFDM accurately estimated transition matrix from just aggregated population data. T-CFDM also flexibly fitted to the various travel duration distributions. The results show that T-CFDM allows us to capture the heterogeneity in travel duration among individuals.

6.4. Visualization of the estimated people flows

In this section, we present qualitative comparisons of the estimated transition populations in the pedestrian data. Fig. 5 visualizes the pedestrian flows in Hall 1 and Hall 2. In Fig. 5, we illustrate the true pedestrian flow on the left, and the estimates of T-CFDM, CFDM, Popularity and Uniform on the right. As shown in Fig. 5, T-CFDM better discerned the pedestrian flows than the other methods. CFDM tends to output some false flows. This is mainly because CFDM is based on the unrealistic assumption that all pedestrians who left a location at one time step should arrive at another location at the same time step; and thus CFDM misestimates the transition populations between locations. T-CFDM, on the other hand, could more accurately estimate the transition populations because it considers travel durations between locations. The visualization results are useful for optimizing navigation systems. For example, discovering popular routes of pedestrians yield better

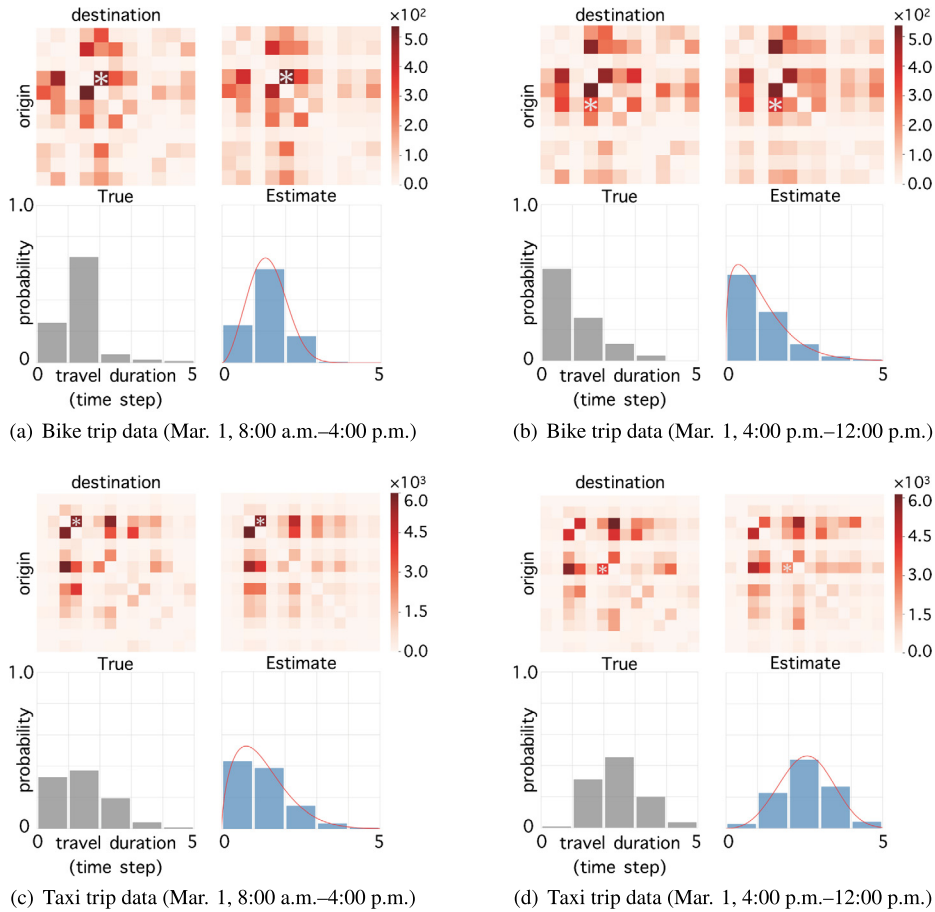


Fig. 4. Visualization of transition matrices and travel duration probabilities. (Upper part of each figure) Heatmap visualization of the transition matrix. The true matrix is shown on the left and the estimate of our model is shown on the right. (Lower part of each figure) Travel duration probability $F(\Delta; \hat{\gamma}_{ji})$ estimated by T-CFDM for the origin-destination pair specified by the white asterisk (*) in the corresponding upper heatmap; the left one is the true probability and the right one is the estimate of T-CFDM. Red line represents the travel duration distribution $f(\tau; \hat{\gamma}_{ji})$ estimated by T-CFDM.

route recommendations; those that are likely to be chosen by visitors. The results are also useful to marketers when they want to optimize the strategies of location-based advertising. For example, analyzing the transition relation between booths provide better understanding of the visitors’ interests. This makes it easier for marketers to determine which advertisements to serve to the visitors according to their interests and current locations.

7. Conclusion

We have proposed the Time-delayed Collective Flow Diffusion Models (T-CFDM) for inferring latent people flows, i.e., transition populations between observed locations, from just aggregated population data. An important characteristic of the T-CFDM is the incorporation of the travel duration probabilities into the people flow conservation constraints; the advantage of which is that it can accurately infer transition populations in more practical settings where the observation range of sensor devices is limited and some people are not observed in any location in some time periods. Since the T-CFDM adopts the noisy observation model for the numbers of incoming and outgoing people, it remains applicable even if the flow conservation constraints do not strictly hold. The approximate expectation-maximization (EM) algorithm that we have developed allows us to estimate transition populations and model parameters simultaneously. Experiments on real-world datasets confirmed that the T-CFDM can accurately infer transition populations between observed locations.

Although our results are encouraging, our model can be further improved in a number of research directions. First, we can extend the model to capture time-varying people flows. One of the approaches to address this issue is to consider a mixture of multiple diffusion processes, each of which is shared among the time-of-day on which tendency of people flows is similar, as in [14]. On the basis of that idea, we will refine T-CFDM and develop its inference algorithm so that it can take account of flows changing over time. Second, a mixture modeling could also be helpful for handling aggregated data that mix multiple types of people (e.g., age and gender). In that case, each process corresponds to the flows of the different type of people. This approach could distinguish different types of flow patterns without labels of their types; however, it is not

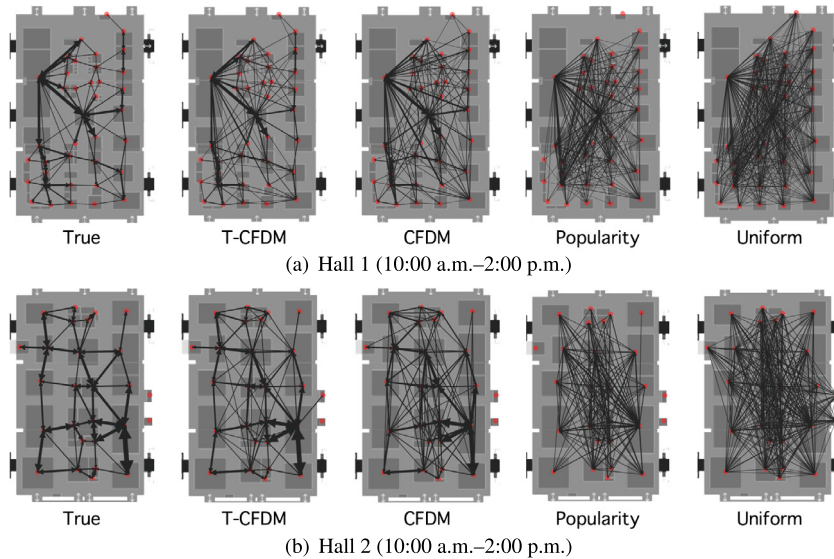


Fig. 5. Visualization comparison of pedestrian flows between booths in Hall 1 and Hall 2. The red dots represent the locations of booths, and the directed edges represent the pedestrian flows between locations. The edge widths are proportional to the transition populations between the respective booth pairs. Note that we omitted those edges whose transition populations were lower than a threshold, and bidirectional edge widths are proportional to the average of the transition populations between the pair of locations.

straightforward to determine the mixture coefficients. This problem is one of future works. Third, we plan to incorporate the Bayesian approach for estimating transition populations and model parameters; it can be expected to provide better results while increasing computations involved (e.g., [26]). Fourth, we will extend the model to utilize external information such as the weather condition and the program of events by integrating it with deep neural networks, as in [13]. Fifth, we will explore advanced algorithms for MAP estimation of transition populations \mathbf{M} , which may be the computational bottleneck in large-scale problems. It would be promising to develop an inference method for T-CFDM by utilizing the recently published techniques based on message-passing algorithms [31] or minimum convex cost flow algorithms [2]. Sixth, our study focuses on the case where aggregated population data are only available, but in practice one might be able to use the detailed information of a limited number of people who agreed with location tracking. In that case, we think that it would be important to consider a hybrid approach for utilizing both types of data, that is, aggregated data and tracking data, to improve the estimation performance. Lastly, we will explore the applicability of our model to other domains. For example, our model could be helpful in the biological sciences: It could be used for modeling animal migration on the basis of the counts of animals in different locations, as in [27,28]. One of our future works is to use various datasets (e.g., eBird data [30]) for further evaluation of the T-CFDM.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Difference from our conference version

In this appendix, we summarize the main differences from [32]. We extend the model in [32] by adopting the noisy observation models, as in [28], for handling noisy settings, where people flow conservation does not strictly hold (Section 4), and develop an approximate EM algorithm for learning the parameters of the extended model (Section 5). We also conduct the extensive experiments: (a) we compare the extended model with the preliminary conference version in terms of the performance in estimating the transition populations (Section 6.2); (b) we mention three types of travel duration distributions (Table 2 and Fig. 3), and conduct the experiments using a more flexible two-parameter distribution, i.e., a Weibull distribution (Section 6); (c) we add the experiments using the data that are divided into the subsets by time-of-day since the tendency of people flows could vary with time-of-day (Section 6).

Appendix B. Discussion of approximate EM algorithm

In this appendix, we elaborate on the approximate EM algorithm and its validity. Let us start discussion of the exact EM algorithm. The aim of the EM algorithm is to maximize the following incomplete-data log-likelihood,

$$\log p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}} | \Phi) = \log \sum_{\mathbf{M}} p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \mathbf{M} | \Phi), \quad (18)$$

in which latent variables (i.e., \mathbf{M}) are marginalized. However, it is often difficult to directly evaluate the log-likelihood (18); then one can consider the following lower bound,

$$\log p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}} | \Phi) \geq \sum_{\mathbf{M}} q(\mathbf{M}) \log p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \mathbf{M} | \Phi) - \sum_{\mathbf{M}} q(\mathbf{M}) \log q(\mathbf{M}), \quad (19)$$

which is derived from Jensen's inequality, where $q(M_{tij}) \geq 0$ and $\sum_{j \in \mathbf{E}_i} q(M_{tij}) = 1$. If we let $q(\mathbf{M}) = p(\mathbf{M} | \mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \hat{\Phi})$ (the posterior of the latent variables \mathbf{M} evaluated with the current estimate $\hat{\Phi}$ of the parameter Φ), one can obtain the Q-function as the lower bound of the log-likelihood (18) as follows:

$$\log p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}} | \Phi) \geq Q(\hat{\Phi}, \Phi) + \text{const.}, \quad (20)$$

where $Q(\hat{\Phi}, \Phi)$ is given by (18). By maximizing $Q(\hat{\Phi}, \Phi)$ iteratively, one can always obtain the local optima of parameters that maximize the incomplete-data log-likelihood (18). Details are described in [23, Section 11.4.7.2].

In the approximate EM algorithm we developed, we replace the posterior expectation in (7) with a plug-in procedure of the MAP estimate $\hat{\mathbf{M}}$ represented by (8). This procedure, substituting $\mathbf{M} = \hat{\mathbf{M}}$ to $p(\mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \mathbf{M} | \Phi)$, corresponds to the bounding procedure in (19) with the probability distribution $q(\mathbf{M}) = \delta_{\hat{\mathbf{M}}, \mathbf{M}}$, obtaining the approximate Q-function $Q^{\text{approx}}(\hat{\Phi}, \Phi)$ (8). It should be noted that $Q^{\text{approx}}(\hat{\Phi}, \Phi)$ is still a lower bound of the log-likelihood (18). $Q^{\text{approx}}(\hat{\Phi}, \Phi)$ becomes equal to the exact Q-function on the condition that the posterior distribution of \mathbf{M} in (7) can be written as $p(\mathbf{M} | \mathbf{Y}^{\text{out}}, \mathbf{Y}^{\text{in}}, \hat{\Phi}) = \delta_{\hat{\mathbf{M}}, \mathbf{M}}$. In that case, it is guaranteed to converge to local optima that maximize the incomplete-data log-likelihood like the exact EM algorithm. If the above condition is not satisfied, one cannot always obtain local optima that are the same estimated by the exact EM algorithm. Nevertheless, a plug-in procedure of the MAP estimate is expected to be a good approximation in situations where the posterior distribution of \mathbf{M} is concentrated around the MAP estimate $\hat{\mathbf{M}}$. The approximate EM algorithm is also advantageous in that it makes parameter learning efficient by avoiding a posterior expectation in (7). Moreover, according to the previous work [28], the effectiveness of the approximation procedure, i.e., a plug-in procedure of the MAP estimate $\hat{\mathbf{M}}$, has been shown experimentally.

In addition, we mention another aspect of the approximate EM algorithm, whose procedure is represented in Algorithm 1. In Algorithm 1, the latent variables \mathbf{M} and the model parameters Φ are updated alternately. This procedure is equivalent to a kind of coordinate ascent algorithm [35] using $Q^{\text{approx}}(\hat{\Phi}, \Phi)$ as the objective function. Accordingly, Algorithm 1 monotonically increases the objective function $Q^{\text{approx}}(\hat{\Phi}, \Phi)$ until it reaches a local optimum as long as the objective function is differentiable.

Appendix C. Derivative of the objective function $\mathcal{L}(\mathbf{M})$ (10) with respect to M_{tij}

This appendix describes the first derivative of the objective function $\mathcal{L}(\mathbf{M})$ (10) with respect to M_{tij} , which is required for estimating the transition populations \mathbf{M} based on the L-BFGS-B method. The derivative is given by

$$\begin{aligned} \frac{\partial \mathcal{L}(\mathbf{M})}{\partial M_{tij}} &= \frac{1}{\hat{\sigma}_i^2} \left(Y_{ti}^{\text{out}} - \sum_{j \in \mathbf{E}_i} M_{tij} \right) + \frac{1}{\hat{\lambda}_j^2} \sum_{s=1}^T \left(Y_{sj}^{\text{in}} - \sum_{i \in \mathbf{E}_j} \sum_{t=1}^s F(\Delta_{st}; \hat{\mathbf{y}}_{ij}) M_{tij} \right) F(\Delta_{st}; \hat{\mathbf{y}}_{ij}) \\ &\quad + \log \left(\sum_{j \in \mathbf{E}_i} M_{tij} \right) + \log \theta_{ij} - \log M_{tij}, \end{aligned} \quad (21)$$

where s is an auxiliary time variable.

References

- [1] Y. Akagi, T. Nishimura, T. Kurashima, H. Toda, A fast and accurate method for estimating people flow from spatiotemporal population data, in: IJCAI'18, 2018, pp. 3293–3300.
- [2] Y. Akagi, T. Nishimura, Y. Tanaka, T. Kurashima, H. Toda, Exact and efficient inference for collective flow diffusion model via minimum convex cost flow algorithm, in: AAAI'20, 2020, pp. 3163–3170.
- [3] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
- [4] R.H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. 16 (1995) 1190–1208.
- [5] A.B. Chan, Z.S.J. Liang, N. Vasconcelos, Privacy preserving crowd monitoring: counting people without people models or tracking, in: CVPR'08, 2008, pp. 1–7.
- [6] C.Y. Chow, M.F. Mokbel, Trajectory privacy in location-based services and data publication, ACM SIGKDD Explor. Newsl. 13 (1) (2011) 19–29.
- [7] S. Dhar, U. Varshney, Challenges and business models for mobile location-based services and advertising, Commun. ACM 54 (5) (2011) 121–129.
- [8] J. Du, A. Kumar, P. Varakantham, On understanding diffusion dynamics of patrons at a theme park, in: AAMAS'14, 2014, pp. 1501–1502.
- [9] F. Giannotti, M. Nanni, F. Pinelli, D. Pedreschi, Trajectory pattern mining, in: KDD'07, 2007, pp. 330–339.
- [10] M. Hoang, Y. Zheng, A. Singh, FCCF: forecasting citywide crowd flows based on big data, in: SIGSPATIAL'16, 2016, pp. 1–10.

- [11] H. Huang, G. Gartner, A survey of mobile indoor navigation systems, in: G. Gartner, F. Ortig (Eds.), *Cartography in Central and Eastern Europe*, Springer, Berlin Heidelberg, 2010, pp. 305–319.
- [12] T. Iwata, A. Shah, Z. Ghahramani, Discovering latent influence in online social activities via shared cascade Poisson processes, in: *KDD'13*, 2013, pp. 266–274.
- [13] T. Iwata, H. Shimizu, Neural collective graphical models for estimating spatio-temporal population flow from aggregated data, in: *AAAI'19*, 2019, pp. 3935–3942.
- [14] T. Iwata, H. Shimizu, F. Naya, N. Ueda, Estimating people flow from spatio-temporal population data via collective graphical mixture models, *ACM Trans. Spat. Algorithms Syst.* 3 (1) (2017) 1–18.
- [15] D. Kempe, J. Kleinberg, Éva Tardos, Maximizing the spread of influence through a social network, in: *KDD'03*, 2003, pp. 137–146.
- [16] L.A. Klein, *Sensor Technologies and Data Requirements for ITS*, Artech House, 2001.
- [17] A. Kumar, D. Sheldon, B. Srivastava, Collective diffusion over networks: models and inference, in: *UAI'13*, 2013, pp. 351–360.
- [18] T. Kurashima, T. Iwata, G. Irie, K. Fujimura, Travel route recommendation using geotags in photo sharing sites, in: *CIKM'10*, 2010, pp. 579–588.
- [19] T. Kurashima, T. Iwata, N. Takaya, H. Sawada, Probabilistic latent network visualization: inferring and embedding diffusion networks, in: *KDD'14*, 2014, pp. 1236–1245.
- [20] J. Leskovec, L.A. Adamic, B.A. Huberman, The dynamics of viral marketing, *ACM Trans. Web* 1 (1) (2007).
- [21] L.P. Liu, D. Sheldon, T.G. Dietterich, Gaussian approximation of collective graphical models, in: *ICML'14*, 2014, pp. 1602–1610.
- [22] A. Monreale, F. Pinelli, R. Trasarti, F. Giannotti, WhereNext: a location predictor on trajectory pattern mining, in: *KDD'09*, 2009, pp. 637–646.
- [23] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, The MIT Press, 2012.
- [24] M.G. Rodriguez, D. Balduzzi, B. Schölkopf, Uncovering the temporal dynamics of diffusion networks, in: *ICML'11*, 2011, pp. 561–568.
- [25] A.W. Senior, L.M. Brown, A. Hampapur, C.F. Shu, Y. Zhai, R.S. Feris, Y. li Tian, S. Borger, C.R. Carlson, Video analytics for retail, in: *AVSS'07*, 2007, pp. 423–428.
- [26] D. Sheldon, T.G. Dietterich, Collective graphical models, in: *NIPS'11*, 2011, pp. 1161–1169.
- [27] D. Sheldon, M.A.S. Elmohamed, D. Kozen, Collective inference on Markov models for modeling bird migration, in: *NIPS'08*, 2008, pp. 1321–1328.
- [28] D. Sheldon, T. Sun, A. Kumar, T.G. Dietterich, Approximate inference in collective graphical models, in: *ICML'13*, 2013, pp. 1004–1012.
- [29] X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki, Prediction of human emergency behavior and their mobility following large-scale disaster, in: *KDD'14*, 2014, pp. 5–14.
- [30] B.L. Sullivan, C.L. Wood, M.J. Iliff, R.E.B.D. Fink, S. Kelling, eBird: a citizen-based bird observation network in the biological sciences, *Biol. Conserv.* 142 (10) (2009) 2282–2292.
- [31] T. Sun, D. Sheldon, A. Kumar, Message passing for collective graphical models, in: *ICML'15*, 2015, pp. 853–861.
- [32] Y. Tanaka, T. Iwata, T. Kurashima, H. Toda, N. Ueda, Estimating latent people flow without tracking individuals, in: *IJCAI'18*, 2018, pp. 3556–3563.
- [33] Y. Tanaka, T. Kurashima, Y. Fujiwara, T. Iwata, H. Sawada, Inferring latent triggers of purchases with consideration of social effects and media advertisements, in: *WSDM'16*, 2016, pp. 543–552.
- [34] L. Wang, Y. Zheng, X. Xie, W.Y. Ma, A flexible spatio-temporal indexing scheme for large-scale GPS track retrieval, in: *MDM'08*, 2008, pp. 1–8.
- [35] S.J. Wright, Coordinate descent algorithms, *Math. Program.* 151 (1) (2015) 3–34.
- [36] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, D. Jin, Trajectory recovery from ash: user privacy is not preserved in aggregated mobility data, in: *WWW'17*, 2017, pp. 1241–1250.
- [37] H. Yao, X. Tang, H. Wei, G. Zheng, Z. Li, Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction, in: *AAAI'19*, 2019, pp. 5668–5675.
- [38] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, Z. Li, Deep multi-view spatial-temporal network for taxi demand prediction, in: *AAAI'18*, 2018, pp. 2588–2595.
- [39] J. Yuan, Y. Zheng, X. Xie, Discovering regions of different functions in a city using human mobility and POIs, in: *KDD'12*, 2012, pp. 186–194.
- [40] J. Zhang, Y. Zheng, D. Qi, Deep spatio-temporal residual networks for citywide crowd flows prediction, in: *AAAI'17*, 2017, pp. 1655–1661.
- [41] J. Zhang, Y. Zheng, J. Sun, D. Qi, Flow prediction in spatio-temporal networks based on multitask deep learning, *IEEE Trans. Knowl. Data Eng.* (2019), <https://doi.org/10.1109/TKDE.2019.2891537>.