Estimating Latent Population Flows from Aggregated Data via Inversing Multi-Marginal Optimal Transport

Sikun Yang*† Hongyuan Zha‡

Abstract

We study the problem of estimating latent population flows from aggregated count data. This problem arises when individual trajectories are not available due to privacy issues or measurement fidelity. Instead, the aggregated observations are measured over discrete-time points, for estimating the transition flows among states. Most related studies tackle the problems by learning the transition parameters of a time-homogeneous Markov process. Nonetheless, most real-world population flows can be influenced by various uncertainties such as traffic jam and weather conditions. Thus, in many cases, a time-homogeneous Markov model is a poor approximation of the much more complex population flows. To circumvent this difficulty, we resort to a multi-marginal optimal transport (MOT) formulation that can naturally represent aggregated observations by constrained marginals, and encode transition matrices by the cost functions. In particular, we propose to learn the time-varying transition matrices by learning the cost matrices of the MOT formulation, and to estimate latent transition flows simultaneously. The experiments on both synthetic and real data, demonstrate the improved accuracy of the proposed algorithms in estimating transition flows, compared against the related methods.

1 Introduction

This work focuses on the problems where data about individuals are not readily available because of various reasons such as privacy issues and measurement fidelity. Instead, we only have access to the population-level aggregate data that could be *incomplete* and *noisy*. For instance, when studying the infectious disease spreading [5], it is too expensive or even impossible to track the trajectory of each individual. Nevertheless, the number of individuals in some regions over discrete time points, can be measured using sensing devices. Statistical analysis of these aggregate data is challenging, and has received amounts of attention in diverse fields including estimating ensemble flows [8, 15], steering opinion dynamics among humans [28], epidemic forecasting [20] among others.

Over last decade, many efforts, such as collective graphical models (CGMs) [21, 22, 25, 14], have been dedicated to the problem of inference and learning with aggregated data. These methods often assume that the individuals behind aggregated data, behave according to a time-homogeneous Markov chain. However, in many cases, the individual movement behaviors are significantly affected by various factors including weather conditions, traffic situations, and so on. Hence, estimating latent transition flows only with a time-homogeneous Markov model, may lead to a poor approximation in many cases. In addition, some recent studies [8, 9, 23] try to estimate the latent transition flow with marginally aggregated observations, using a multi-marginal optimal transport formulation. By representing aggregated observations by constrained marginals, the transition flows can be estimated by inferring the corresponding transport plan of the MOT problem. In particular, the MOT formulation enables to readily apply Sinkhorn algorithm to perform efficient inference in collective graphical models with guaranteed convergence. Following this success, Singh *et al.* [24] developed an approximate expectation-maximization (EM) algorithm to learn the parameters of a time-homogeneous Markov process, from the marginally aggregate observations of an ensemble flow. Despite being simple and tractable, this method still strictly assumes that each individual behaves according to a time-homogeneous Markov model, which thus may lead to a poor approximation of complicated real-world flows.

In this work, we propose to estimate the latent transition flows from aggregated data, by learning the cost functions of the tree-structured multi-marginal optimal transport framework. In particular, our method allows the estimated transition parameters to be time-varying, and thus demonstrated improved accuracy in analyzing real-world population flows, compared with existing time-homogeneous Markov models. In particular, the main contributions of this paper are:

- We propose an expectation-maximization (EM)-type algorithm to simultaneously learn the time-dependent transition kernels of the MOT formulation, and to estimate the transition flows using Sinkhorn belief propagation algorithm (Sec.4). The uniqueness of the recovered transition parameters can be ensured under some mild

*Corresponding author.

†Shenzhen Institute of Artificial Intelligence and Robotics for Society; Great Bay University, Dongguan 523808, China. yangsikun@cuhk.edu.cn

‡Shenzhen Institute of Artificial Intelligence and Robotics for Society; School of Data Science, The Chinese University of Hong Kong, Shenzhen. zhahy@cuhk.edu.cn

conditions, as proved by the recent studies in inverse optimal transport [16].

- We also investigate regularized convex optimization algorithms [4] to construct cost functions as sparse linear combinations of some basis distance functions, which allow to learn more complicated cost functions than symmetric ones.
- Experiments are conducted on both synthetic and real-world flow data, to demonstrate the improved performance of the proposed methods in estimating transition flows, compared with existing methods with time homogeneous transition kernels.

2 Related Work

Collective graphical models (CGMs) is proposed by [21] as a formalism to perform inference in aggregate noisy data including ensemble flows. Sheldon *et al.* [22] studied the intractability of the exact marginal inference in CGMs, and proposed an approximate maximum a posteriori (MAP) estimation as a substitute. Following this success, Sun *et al.* [25] developed the non-linear belief propagation algorithm to perform approximate MAP inference in CGMs. Bethe-RDA is another algorithm dedicated to aggregate inference in CGMs via regularized dual averaging (RDA) with guaranteed convergence. Recently, Bernstein and Sheldon [3] developed an approach of moments estimator to learn the parameters of the Markov model from aggregate noisy flows. Haasler *et al.* [11, 8] recently investigated the problems of estimating ensemble flows from a graphical-structured multi-marginal optimal transport perspective. In particular, Haasler *et al.* [9] studied a graphical-structured multi-marginal optimal transport, which allows to consider various related problems such as information fusion under a unified MOT framework. Singh *et al.* [23] first studied the inference (filtering) problems in CGMs based upon the graphical-structured MOT framework. Singh *et al.* [24] derived an approximate EM algorithm to conduct learning and inference in time-homogeneous collective graphical models.

To the best of our knowledge, most collective graphical models try to capture transition flows using time-homogeneous Markov models. In contrast, we aim to learn the time-dependent transition matrices indirectly by learning the corresponding cost functions. Our methods are based upon a tree-structured multi-marginal optimal transport (MOT) formalism. Using the MOT framework, the existing works [11, 9] only estimate the transition flows with a predetermined transition kernel. In this paper, we try to learn time-varying transition parameters and also to estimate the transition flows, from noisy aggregate data. In addition, the proposed methods are closely related to inverse optimal transport [13, 16, 4], where they only aim to learn the cost functions from the observed matchings. In constrast, this work needs to recover transition flows from noisy observations, and to learn the time dependent cost functions, iteratively. Other related studies include collective flow diffusion models (CFDM) [26, 1], which can incorporate people's travel duration between locations for estimating transition flows. Neural collective graphical models (CGMs) can estimate population flows by incorporating additional spatiotemporal information into transition kernel parameterized by neural nets [12]. The CFDM and Neural CGMs need to explicitly model observation noise, while the proposed methods can compactly capture noisy observations via constrained marginals.

3 Background

Notations. By $\exp(\cdot), \ln(\cdot), \odot, /$, we denote the element-wise exponential, logarithm, multiplication, and division of vectors, matrices and tensors, respectively. The outer product is denoted by $\otimes$. Let $\mathbf{p}$ and $\mathbf{q}$ be two nonnegative vectors, matrices or tensors of the same dimension. The normalized Kullback-Leibler (KL) divergence of $\mathbf{p}$ from $\mathbf{q}$ is defined as $H(\mathbf{p}|\mathbf{q}) \equiv \sum_i (p_i \ln(\frac{p_i}{q_i}) - p_i + q_i)$, where $0 \ln 0$ is defined to be 0. Similarly, defined $H(\mathbf{p}) \equiv H(\mathbf{p}|\mathbf{1}) = \sum_i (p_i \ln(p_i) - p_i + 1)$, which is effectively the negative of the entropy of $\mathbf{p}$.

3.1 Optimal transport. Here we only consider the discrete optimal transport problems, and refer to [27] for its continuous counterpart. Let $\boldsymbol{\mu}_1 \in \mathbb{R}^{d_1}_{\geq 0}$ and $\boldsymbol{\mu}_2 \in \mathbb{R}^{d_2}_{\geq 0}$ be two distributions with equal mass. The optimal transport (OT) aims at finding a transport mapping from $\boldsymbol{\mu}_1$ to $\boldsymbol{\mu}_2$, while minimizing the total transport cost. In particular, the transport cost is defined by an underlying cost matrix $C \in \mathbb{R}^{d_1 \times d_2}$, where C_{i_1,i_2} measures the cost of moving an unit mass from location i_1 to i_2. Hence, the Monge-Kantorovich formulation of OT is to find a transport plan by solving the following optimization problem

$$\min_{M \in \Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \langle C, M \rangle,$$

where $\langle C, M \rangle = \sum_{i_1,i_2} C_{i_1,i_2} M_{i_1,i_2}$, and $\Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ denotes the set of nonnegative matrices satisfying maringal constraints specified by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Computing the exact OT problem requires solving a linear program with time complexity $\mathcal{O}(n^3 \ln n)$ [19], which is too expensive for large-scale settings. To avoid excessive computational cost, Cuturi [6] introduces an entropy regularization term $H(M) = \sum_{i_1,i_2} (M_{i_1,i_2} \ln(M_{i_1,i_2}) - M_{i_1,i_2} + 1)$, and thus forms an approximate OT problem as

$$\min_{M \in \Pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)} \Big\{ \langle C, M \rangle + \epsilon H(M) \Big\}, \tag{3.1}$$

where $\epsilon \geq 0$. When ϵ approaches 0, one recovers the canonical OT. For $\epsilon > 0$, taking the dual of the approximation leads

to a strictly convex optimization problem, which enables us to obtain an unique solution up to multiplication/division by a constant [7].

3.2 Multi-marginal optimal transport. Multi-marginal optimal transport (MOT) generalizes bi-marginal OT by considering optimal transport problems involving multiple marginal constraints. More specifically, the MOT problem is to find a transport plan between a set of marginals $\{\boldsymbol{\mu}_j\}_{j=1,2,\dots,J}$, where $\boldsymbol{\mu}_j \in \mathbb{R}_{\geq 0}^{d_j}$. In this setting, the transport cost is encoded as $C = [C_{i_1,i_2,\dots,i_J}] \in \mathbb{R}^{d_1 \times \cdots \times d_J}$, and the transport plan is denoted by $M = [M_{i_1,i_2,\dots,i_J}] \in \mathbb{R}_{\geq 0}^{d_1 \times \cdots \times d_J}$. For a tuple $(i_1, i_2, \dots, i_J)$, $C_{i_1,i_2,\dots,i_J}$ denotes the transport cost of moving an unit mass, and $M_{i_1,i_2,\dots,i_J}$ describes the amount of mass transported for that tuple. Naturally, the Monge-Kantorovich formulation of MOT reads

$$\begin{aligned} (3.2) \qquad \min_{M} \quad & \langle C, M \rangle \\ \text{subject to} \quad & P_j(M) = \boldsymbol{\mu}_j, \quad \text{for } j \in \Gamma, \end{aligned}$$

where $\Gamma \subset \{1, 2, \dots, J\}$ denotes an index set specifying which marginal constraints are given. The projection of the tensor M on its j-th marginal is given by

$$(3.3) \quad P_j(M) = \sum_{i_1,\dots,i_{j-1},i_{j+1},\dots,i_J} M_{i_1,\dots,i_{j-1},i_j,i_{j+1},\dots,i_J}.$$

Note that the original multi-marginal optimal transport formulation [17, 18] specifies all the marginal distributions as its constraints. Here we consider the case where only a subset of marginals are explicitly given, i.e., $\Gamma \subset \{1, 2, \dots, J\}$. This arises in many cases of interests including dynamic network flows [10] and Barycenter problems [2].

The entropy regularized MOT reads

$$\begin{aligned} \min_{M} \quad & \Big\{ \langle C, M \rangle + \epsilon H(M) \Big\} \\ \text{subject to} \quad & P_j(M) = \boldsymbol{\mu}_j, \quad \text{for } j \in \Gamma. \end{aligned}$$

Using the Lagrangian duality theory, it is not hard to see the optimal solution of the entropy regularized MOT is of the form $M = K \odot B$ where $K = \exp(-C/\epsilon)$, and $B = \mathbf{b}_1 \otimes \cdots \otimes \mathbf{b}_J$ with

$$\mathbf{b}_j = \begin{cases} \exp(\boldsymbol{\alpha}_j/\epsilon), & \text{if } j \in \Gamma \\ \mathbf{1}, & \text{otherwise} \end{cases}$$

where $\boldsymbol{\alpha}_j \in \mathbb{R}^n$ denotes the dual variable corresponding to the constraint $P_j(M) = \boldsymbol{\mu}_j$, for $j \in \Gamma$. The generalized Sinkhorn algorithm solves entropy regularized MOT problems by iteratively updating the vectors $\mathbf{b}_j$, for $j \in \Gamma$, as

$$\mathbf{b}_j \leftarrow \mathbf{b}_j \odot \boldsymbol{\mu}_j / P_j(K \odot B).$$

Note that the computational complexity of Sinkhorn algorithm still scales exponentially with J because the number of elements in M is $d_1 \times \cdots \times d_J$.

Fortunately, the tree-structured cost tensors in many cases of interests, allow us to make the computation of the marginal projections feasible [11]. More specifically, let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be a tree with $\mathcal{V}$ denoting the nodes, and $\mathcal{E}$ the edges. Assume that the cost tensor C can be decomposed according to a tree structure $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with J nodes as

$$(3.4) \qquad C_{i_1,\dots,i_J} = \sum_{(v,u)\in\mathcal{E}} C^{(v,u)}_{i_v,i_u},$$

where $C^{(v,u)}$ denotes the cost matrix between marginals $\boldsymbol{\mu}_v$ and $\boldsymbol{\mu}_u$, for $(v, u) \in \mathcal{E}$. We refer to the problem 3.2 with a cost of the form 3.4 as a tree-structured multi-marginal optimal transport problem. In particular, for an entropy regularized MOT problem with a tree-structured cost, the transport plan is $M = K \odot B$. By letting $K^{(v,u)} \equiv \exp(-C^{(v,u)}/\epsilon)$, the projection of M on the j-th marginal is given by

$$\begin{aligned} (3.5) \quad & P_j(M)_{i_j} \\ & = (\mathbf{b}_j)_{i_j} \sum_{i_1,\dots,i_{j-1},i_{j+1},\dots,i_J} \prod_{(v,u)\in\mathcal{E}} K^{(v,u)}_{i_v,i_u} \prod_{v\in\mathcal{V}\setminus j} (\mathbf{b}_v)_{i_v}. \end{aligned}$$

This sum only involves matrix-vector multiplications, and hence substantially reduces the computational complexity compared with the brute force summation in Eq. 3.3. The full algorithm is introduced as Sinkhorn belief propagation algorithm in [11], for graphically structured MOT problems.

4 Problem Formulation

Consider a population of N individuals (e.g., pedestrians, bikes, cars), each of which independently behaves according to a Markov chain. Let the states of the Markov chain be $X = \{X_1, \dots, X_S\}$ with S being the number of states. In particular, the transition parameters of the Markov chain allows to be time-varying, and specified by $\mathbf{A}^t$, where $A^t_{ij} = p(s_{t+1} = x_j \mid s_t = x_i)$ denotes the transition probability from state x_i at time t, to state x_j at time $t+1$. Let $(\boldsymbol{\mu}_t)_i$ denote the number of individuals appearing in state x_i at time t, and $\mathbf{M}^t = [M^t_{ij}]$, where M^t_{ij} denote the number of individuals moving from state x_i at time t, to state x_j at time $t + 1$. Hence, $\boldsymbol{\mu}_t$ represents the marginal aggregate observation at time t, and $\mathbf{M}^t$ captures the transition flows from time t to $t + 1$. We also denote the noisy observation of $\boldsymbol{\mu}_t$ by $\tilde{\boldsymbol{\mu}}_t$. Fig. 1 illustrates an example of the studied problem. The probability of the transition flow observed during the time interval $[t, t + 1]$ is given by

$$p(\mathbf{M}^t) = \prod_{i=1}^{S} \frac{(\boldsymbol{\mu}_t)_i}{\prod_{j=1}^{S} M^t_{ij}} \prod_{j=1}^{S} (A^t_{ij})^{M^t_{ij}}.$$

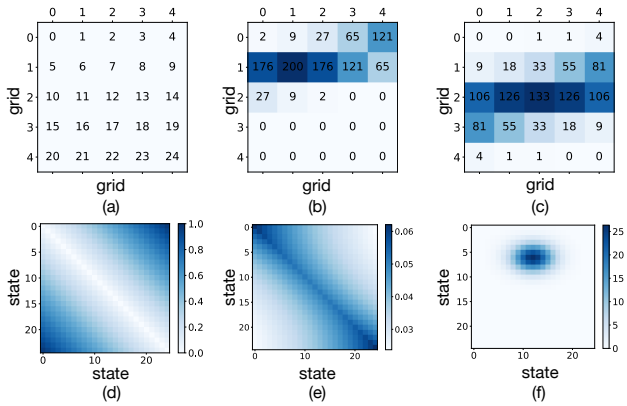

Figure 1: An illustration of the studied problem. The 5×5 grid cells form 25 states (a). There are $1,000$ individuals moving among these states. The aggregated observations (the number in each cell, indicates the observed count of individuals in that state) are measured at two consecutive time steps, as shown in (b) and (c), respectively. This work aims to estimate the latent transition flow (f) (the value of (i, j)-th entry denotes the number of individuals moving from state i to state j at the target time step) by learning the cost matrices (d). The transition matrix is displayed in (e).

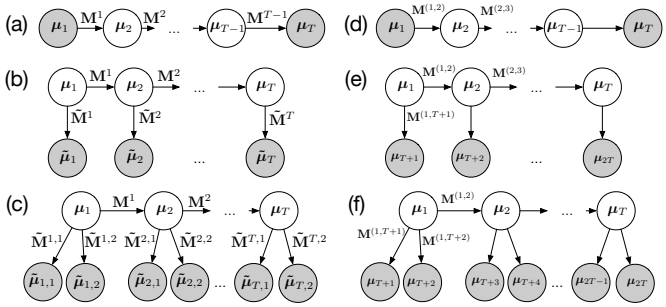

Figure 2: An illustration of the original ensemble flow estimation problem in [9, 11], where the transition parameter $\mathbf{A}$ and two marginals $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_T$ are known, and the goal is to estimate transition flows and intermediate marginals (a). Latent flow estimation problems (b) and (c), provide multiple noisy marginals, for which we aim to estimate the underlying transition flows, marginals, and to learn transition parameters. In (c), $\tilde{\boldsymbol{\mu}}_{t,s}$ denotes the s-th noisy measurements at time t, and $\tilde{\mathbf{M}}^{t,s}$ denotes the transition flows between $\tilde{\boldsymbol{\mu}}_{t,s}$ and $\boldsymbol{\mu}_t$. The graphs in (d)(e)(f) refer to the tree-structured representation of the problems in (a)(b)(c), respectively. In particular, we use $\mathbf{M}^{(u,v)}$ to denote the transition flow between nodes u and v for the ease of notation.

Interestingly, a large deviation interpretation [9] has shown that as the number of individuals N tends to infinity, if $\frac{1}{N}\boldsymbol{\mu}_t \to \bar{\boldsymbol{\mu}}_t$, and $\frac{1}{N}\mathbf{M}^t \to \bar{\mathbf{M}}^t$, the log-likelihood of the transition flow $\mathbf{M}^t$ can be well approximated as

$$\frac{1}{N}\log p(\mathbf{M}^t) \to -H(\bar{\mathbf{M}}^t \mid \mathrm{diag}(\bar{\boldsymbol{\mu}}_t)\mathbf{A}^t).$$

Fig. 2(a) illustrates a scenario for which given two marginal observations $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_T$, the previous studies[9, 11] aim to estimate the transition flows $\{\mathbf{M}^t\}_{t=1}^{T-1}$ using the predetermined transition matrix $\mathbf{A}$, for time-homogeneous Markov models. Fig. 2(b) describes the problem setting, where we only have access to the noisy observations $\{\tilde{\boldsymbol{\mu}}_t\}_{t=1}^{T}$, of the true marginals $\{\boldsymbol{\mu}_t\}_{t=1}^{T}$. Let $\tilde{\mathbf{M}}^t$ denote the flows between the true aggregate $\boldsymbol{\mu}_t$ and its noisy observation $\tilde{\boldsymbol{\mu}}_t$. Hence, given the noisy aggregates $\{\tilde{\boldsymbol{\mu}}_t\}_{t=1}^{T}$ [1], we can estimate latent transition flows by solving a convex optimization problem given by

$$\begin{aligned}(4.6)\qquad \min_{\substack{\mathbf{M}^{[1:(T-1)]},\\ \tilde{\mathbf{M}}^{[1:(T-1)]},\boldsymbol{\mu}_{[1:T]}}} \quad & \left\{\sum_{t=1}^{T-1} H(\mathbf{M}^t \mid \mathrm{diag}(\boldsymbol{\mu}_t)\mathbf{A}^t) \right. \\ & \left. + \sum_{t=1}^{T-1} H(\tilde{\mathbf{M}}^t \mid \mathrm{diag}(\tilde{\boldsymbol{\mu}}_t)\tilde{\mathbf{A}})\right\} \\ \text{subject to } & \tilde{\mathbf{M}}^t\mathbf{1} = \boldsymbol{\mu}_t, \quad (\tilde{\mathbf{M}}^t)^{\mathrm{T}}\mathbf{1} = \tilde{\boldsymbol{\mu}}_t, \\ & \mathbf{M}^t\mathbf{1} = \boldsymbol{\mu}_t, \quad (\mathbf{M}^t)^{\mathrm{T}}\mathbf{1} = \boldsymbol{\mu}_{t+1}, \\ & \text{for } \ t = 1, \ldots, T-1.\end{aligned}$$

where $\tilde{\mathbf{A}}$ denotes the emission parameter, which determines the conditional distribution of the noisy observation $\tilde{\boldsymbol{\mu}}_t$ given the true marginal $\boldsymbol{\mu}_t$.

Remark 1. Fig. 2(e) depicts an equivalent tree structure[2] of Fig. 2(b). If we define the cost matrix $\mathbf{C}^t = -\epsilon\log(\mathbf{A}^t)$ and $\tilde{\mathbf{C}} = -\epsilon\log(\tilde{\mathbf{A}})$, the convex optimization problem in Eq. 4.6 is equivalent to a multi-marginal optimal transport problem specified by

$$\begin{aligned}(4.7)\qquad \min_{\mathbf{M}} \quad & \left\{\langle \mathbf{C}, \mathbf{M}\rangle + \epsilon H(\mathbf{M} \mid \mathbf{1}_{S\times\cdots\times S})\right\} \\ \text{subject to} \quad & P_j(\mathbf{M}) = \boldsymbol{\mu}_j, \quad \text{for } j \in \Gamma,\end{aligned}$$

[1] Hereafter, we use $\boldsymbol{\mu}_t$, $\mathbf{M}^t$ to denote the normalized observations $\bar{\boldsymbol{\mu}}_t, \bar{\mathbf{M}}^t$, respectively, for ease of notation.

[2] We use $\mathbf{C}^{(u,v)}$ in the tree structure representation instead of $\mathbf{C}^t$, for ease of notation. The meaning of $\mathbf{C}^{(u,v)}$ can be understood from the context.

Algorithm 1 Sinkhorn Belief Propagation Algorithm

Input: Tree-structured graph $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V}$ the node set, $\mathcal{E}$ the edge set, and the indices of the constrained marginals Γ, marginal observations $\{\boldsymbol{\mu}_j\}_{j\in\Gamma}$ and edge potentials $\boldsymbol{\Phi}^{(t,v)} = \exp(-\mathbf{C}^{(t,v)}/\epsilon)$ for $(t, v) \in |\mathcal{E}|$

Output: the transition flows $\mathbf{M}^{(t,v)}(x_t, x_v) \propto \boldsymbol{\phi}^{(t,v)}(x_t, x_v) \prod_{k\in N(t)} \breve{\mathbf{M}}_{k\to t}(x_t) \prod_{k\in N(v)} \breve{\mathbf{M}}_{k\to v}(x_v)$

1: Initialize the messages $\breve{\mathbf{M}}_{v\to u}(x_u)$, $\forall (v, u) \in \mathcal{E}$
2: **repeat**
3: **for** $v \in \Gamma$ **do**
4: Update $\breve{\mathbf{M}}_{v\to u}(x_u) \propto \sum_{x_v} \phi^v(x_v, x_u) \frac{\boldsymbol{\mu}_v(x_v)}{\breve{\mathbf{M}}_{u\to v}(x_v)}$, $\forall u \in N(v)$
5: Update all the messages on the path from v to v_{next} $\breve{\mathbf{M}}_{v\to u}(x_u) \propto \sum_{x_v} \phi^v(x_v, x_u) \prod_{k\in N(v)\setminus u} \breve{\mathbf{M}}_{k\to v}(x_v)$
6: **end for**
7: **until** convergence

where the cost tensor $\mathbf{C} \in \mathrm{R}^{S\times\cdots\times S}$ decomposes as $\mathbf{C}_{i_1,\ldots,i_J} = \sum_{(u,v)\in\mathcal{E}} \mathbf{C}^{(u,v)}_{i_u,i_v}$, according to the tree structure. The marginal observation $\boldsymbol{\mu}_j$ equals to the projection of the tensor-valued transport plan $\mathbf{M} \in \mathrm{R}^{S\times\cdots\times S}$ on its j-th mode. Similarly, the transition flow $\mathbf{M}^{(u,v)}$ can be obtained by the projection of the tensor-valued transport plan $\mathbf{M}$ on its (u, v)-th modes, i.e., $\mathbf{M}^{(u,v)} = P_{(u,v)}(\mathbf{M})$. We refer to Sec.4.2 in [9] for the detailed proof of the equivalence between Eq. 4.6 and 4.7.

In particular, our goal is to develop an EM type algorithm to iteratively recover the transition flows $\{\mathbf{M}^{(u,v)}\}_{(u,v)\in\mathcal{E}}$ and to learn the unknown cost matrices $\{\mathbf{C}^{(u,v)}\}_{(u,v)\in\mathcal{E}}$, given the noisy marginal observations $\{\tilde{\boldsymbol{\mu}}_t\}_{t=1}^T$. More specifically, in the M-step, we consider learning the cost matrices $\mathbf{C}^{(u,v)(\ell)}$ of ℓ-th iteration given the two marginal observations $\boldsymbol{\mu}_u$ and $\boldsymbol{\mu}_v$, and the estimated transition flow $\mathbf{M}^{(u,v)(\ell)}$ of the previous iteration. In the E-step, the expectation of transition flow $\mathbf{M}^{(u,v)(\ell+1)}$ is updated based upon the estimated $\mathbf{C}^{(u,v)(\ell)}$. Moreover, via the tree-structured MOT framework, the proposed method can be well extended to more complicated scenarios where multiple noisy aggregates are available for each time step (Fig. 2(c)).

E-step. With the cost matrices $\{\mathbf{C}^{(u,v)}\}_{(u,v)\in\mathcal{E}}$ updated in the M-step, the optimization problem in Eq.4.6 can be equivalently solved via an entropy regularized multi-marginal optimal transport formulation in Eq. 4.7. Hence, the tree-structure induced by the latent flow estimation, enables us to readily utilize Sinkhorn belief propagation (SBP) algorithm to recover the latent transition flows $\{\mathbf{M}^{(u,v)}\}_{(u,v)\in\mathcal{E}}$. The SBP algorithm for the E-step is detailed in Algorithm 1.

M-step. Given the marginal observations $\{\boldsymbol{\mu}_j\}_{j\in\Gamma}$, and the estimated transition flows $\{\mathbf{M}^{(u,v)}\}_{(u,v)\in\mathcal{E}}$, the parameter learning of the collective graphical models in Eq. 4.6, becomes an inverse multi-marginal optimal transport problem given by

$$\min_{\mathbf{C},\boldsymbol{\alpha}} \Big\{F(\boldsymbol{\alpha}, \mathbf{C}) + R(\mathbf{C})\Big\}, \tag{4.8}$$

where $F(\boldsymbol{\alpha}, \mathbf{C}) \equiv \langle \mathbf{M}^{(\ell)}, \mathbf{C}\rangle - \sum_{j\in\Gamma}\langle \boldsymbol{\alpha}_j, \boldsymbol{\mu}_j\rangle + \epsilon\langle \mathbf{K}, \mathbf{B}\rangle$ is a convex function, $\boldsymbol{\alpha} \equiv [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{|\Gamma|}]$ denote the dual variables corresponding to the marginal constraints $\{P_j(\mathbf{M}) = \boldsymbol{\mu}_j\}_{j\in\Gamma}$, $\mathbf{K} \equiv \exp(-\frac{\mathbf{C}}{\epsilon})$, $\mathbf{B} \equiv \mathbf{b}_1 \otimes \cdots \otimes \mathbf{b}_{|\Gamma|}$ where $\mathbf{b}_j \equiv \exp(\frac{\boldsymbol{\alpha}_j}{\epsilon})$, and $R(\mathbf{C})$ is the regularization imposed on the cost tensor $\mathbf{C}$. The detailed derivation of Eq. 4.8 can be found in the appendix of [29]. Note that the optimization problem in Eq.4.8 admits infinitely many solutions without additional regularization imposing on cost tensor $\mathbf{C}$.

Some recent advancements [13, 16] in solving the problem of inverse optimal transport(IOT), has proved that the IOT problem admits an unique solution if the cost function is restricted to belong to a set of symmetric matrices with zero diagonal elements, and thus the proximal operator can be specified by

$$\mathbf{C}^{(u,v)} = \mathrm{prox}_{\gamma R}(\hat{\mathbf{C}}^{(u,v)}) = (\hat{\mathbf{C}}^{(u,v)} + (\hat{\mathbf{C}}^{(u,v)})^{\mathrm{T}})/2,$$

and followed by enforcing the diagonal entries of $\mathbf{C}^{(u,v)}$ to be 0. In our case, the cost tensor of the inverse multi-marginal optimal transport problem, naturally decouples, according to the tree structure as $\mathbf{C}_{i_1,\ldots,i_J} = \sum_{(u,v)\in\mathcal{E}} \mathbf{C}^{(u,v)}_{i_u,i_v}$. Thus, we impose symmetric and zero-diagonal constraints straightforwardly on each of the cost matrices $\mathbf{C}^{(u,v)}$, instead of restricting a symmetric cost tensor. One instance of symmetric cost matrices is $\mathbf{C}^{(u,v)} = [C^{(u,v)}_{ij}]$ with $C^{(u,v)}_{ij} = |x_i - x_j|^2$ where x_i and x_j denote i-th and j-th locations, respectively. In particular, $\boldsymbol{\alpha}$ can be updated using Sinkhorn belief propagation algorithm for entropy regularized MOT. More specifically, to solve the convex optimization problem in Eq.4.8, a block coordinate descent scheme detailed in Algorithm 2 can be considered to alternatively update $\mathbf{C}$ and $\boldsymbol{\alpha}$.

Although the symmetric and zero-diagonal constraints ensure the unique solution, the cost matrices between the states might be more complex. For instance, in many urban population data [24], most individuals are transitioning from suburb towards downtown areas in the early morning, while they are moving back in the opposite direction, in the late evening. Inspired by recent advances in the optimal matching studies [4], we consider constructing the time-dependent cost matrices as a sparse, linear combination of basis distance matrices. More specifically, $\mathbf{C}^{(u,v)} = \sum_{q=1}^{Q} \beta^{(u,v)}_q \mathbf{D}^q$ where $D^q_{ij} = |x_i - x_j|^q$ denotes the (i, j)-th element of the q-th basis distance matrix, x_i is the i-th location, $\boldsymbol{\beta}^{(u,v)} =$

Algorithm 2 Iterative Scaling Algorithm for Learning Cost Functions

Input: The expected transition flows $\mathbf{M}^{(u,v)(\ell)}$, and the marginal observations $\boldsymbol{\mu}_u$ and $\boldsymbol{\mu}_v$

Output: the cost matrix $\mathbf{C}^{(u,v)}$

Initialize $\epsilon, \mathbf{C}^{(u,v)}, \boldsymbol{\alpha}^u, \boldsymbol{\alpha}^v$, and set $\mathbf{b}_j = \exp(\boldsymbol{\alpha}^j/\epsilon)$ for $j = u, v$

repeat

$\Sigma \leftarrow \exp\Big(-\frac{\mathbf{C}^{(u,v)}}{\epsilon}\Big)$

$\mathbf{b}_u \leftarrow \boldsymbol{\mu}_u/(\Sigma \mathbf{b}_v)$

$\mathbf{b}_v \leftarrow \boldsymbol{\mu}_v/(\Sigma^{\mathrm{T}} \mathbf{b}_u)$

$\Sigma \leftarrow \mathbf{M}^{(u,v)(\ell)}/(\mathbf{b}_u \mathbf{b}_v^{\mathrm{T}})$

$\mathbf{C}^{(u,v)} = \mathrm{prox}_{\gamma R}(-\epsilon \log(\Sigma))$

until convergence

Algorithm 3 ISTA Algorithm for Learning Cost Functions

Input: The expected transition flows $\mathbf{M}^{(u,v)(\ell)}$, marginal observations $\boldsymbol{\mu}_u$, $\boldsymbol{\mu}_v$, and basis distance matrices $\{\mathbf{D}^q\}_{q=1}^Q$

Output: the cost matrix $\mathbf{C}^{(u,v)}$

Initialize $\epsilon, \boldsymbol{\beta}^{(u,v)}, \boldsymbol{\alpha}^u, \boldsymbol{\alpha}^v$, and set $\mathbf{C}^{(u,v)} = \sum_{q=1}^Q \beta_q^{(u,v)} \mathbf{D}^q$, and $\mathbf{b}_j = \exp(\boldsymbol{\alpha}^j/\epsilon)$ for $j = u, v$

repeat

Set $\mathbf{C}^{(u,v)} = \sum_{q=1}^Q \beta_q^{(u,v)} \mathbf{D}^q$

$\Sigma \leftarrow \exp\Big(-\frac{\mathbf{C}^{(u,v)}}{\epsilon}\Big)$

$\mathbf{b}_u \leftarrow \boldsymbol{\mu}_u/(\Sigma \mathbf{b}_v)$

$\mathbf{b}_v \leftarrow \boldsymbol{\mu}_v/(\Sigma^{\mathrm{T}} \mathbf{b}_u)$

$\mathbf{M}^{(u,v)(\beta)} \leftarrow \mathbf{b}_u \odot \Sigma \odot \mathbf{b}_v$

$\beta_q^{(u,v)(\ell+1)} = \mathrm{prox}_{\rho\gamma|\cdot|}\Big(\beta_q^{(u,v)(\ell)} - \rho \sum_{i,j} (\mathbf{M}_{ij}^{(u,v)(\ell)} - \mathbf{M}_{ij}^{(u,v)(\beta)}) D_{ij}^q\Big)$

until convergence

$[\beta_1^{(u,v)}, \ldots, \beta_Q^{(u,v)}]$ is a sparse coefficient vector with q-th element determining the usage of $\mathbf{D}^q$ in the construction of $\mathbf{C}^{(u,v)}$. Thus, the learning problem of the cost matrices $\mathbf{C}^{(u,v)}$ reduces to an optimization problem with respect to $\boldsymbol{\beta}^{(u,v)}$ and $\boldsymbol{\alpha}$ as

$$\min_{\boldsymbol{\beta}^{(u,v)}, \boldsymbol{\alpha}^u, \boldsymbol{\alpha}^v} \Big\{ F(\boldsymbol{\alpha}^u, \boldsymbol{\alpha}^v, \boldsymbol{\beta}^{(u,v)}) + \gamma |\boldsymbol{\beta}^{(u,v)}|_1 \Big\},$$

where $F(\boldsymbol{\alpha}^u, \boldsymbol{\alpha}^v, \boldsymbol{\beta}^{(u,v)}) \equiv \sum_{i_u, i_v} e^{[(\boldsymbol{\alpha}^u)_{i_u} + (\boldsymbol{\alpha}^v)_{i_v} - \mathbf{C}_{i_u,i_v}^{(u,v)}]} + \sum_{i_u,i_v} [\mathbf{C}_{i_u,i_v}^{(u,v)} - (\boldsymbol{\alpha}^u)_{i_u} - (\boldsymbol{\alpha}^v)_{i_v}]$, and $\boldsymbol{\alpha}^u, \boldsymbol{\alpha}^v$ denote the dual variables corresponding to $\boldsymbol{\mu}^u, \boldsymbol{\mu}^v$, respectively, and the ℓ_1 penalty term is to enforce a sparse coefficient vector $\boldsymbol{\beta}^{(u,v)}$. As we did in Algorithm 2, $\boldsymbol{\alpha}$ can be updated using Sinkhorn algorithm, and $\boldsymbol{\beta}^{(u,v)}$ is updated using an iterative shrinkage-thresholding algorithm (ISTA) [4], which leads to the second block coordinate descent scheme as detailed in Algorithm 3, for learning cost matrices. The proximal operator is given by the soft-thresholding operator specified by

$$\mathrm{prox}_{\rho\gamma|\cdot|}(x) = \mathrm{sign}(x) \max\{|x| - \rho\gamma, 0\}.$$

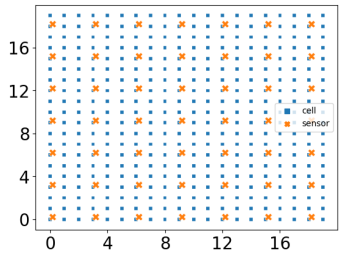

Figure 3: Sensor locations

The proposed EM-type algorithm is to iteratively implement the Sinkhorn belief propagation in the E-step to estimate the expected transition flows, and to learn the cost functions using Algorithm 2 or Algorithm 3 in the M-step. Hereafter, we denote the two developed EM-type algorithms as Sinkhorn belief propagation inverse symmetric transport cost (SBP-ISTC), and Sinkhorn belief propagation iterative shrinkage-thresholding (SBP-ISTA) algorithms.

Computational Cost. For Sinkhorn belief propagation algorithm implemented in the E-step, computing the transition flow matrices $\{\mathbf{M}^{(u,v)}\}_{(u,v)\in\mathcal{E}}$ takes $\mathcal{O}(|\mathcal{E}|S^2)$ time, where S is the number of states, and $|\mathcal{E}|$ is the number of edges in the tree graph. To update the cost function in the M-step, the iterative scaling algorithms enjoy the quadratic computational complexity $\mathcal{O}(S^2)$. For Algorithm 3, the computation cost of ISTA algorithm scales with $\mathcal{O}(Q\mathcal{L})$, where Q denotes the number of basis distance matrices, and $\mathcal{L}$ is the number of inner iterations for the convergence of ISTA algorithm.

5 Experiments

Baselines. The proposed methods were compared with some closely related methods: the collective graphical model [3], constrained norm-product (CNP) algorithm [11], and Sinkhorn belief propagation-Expectation Maximization (SBP-EM) algorithm [24]. The STAY method assumes that all the individuals stay in the same states from time t to time $t+1$, i.e., $\hat{M}_{ii}^t = (\boldsymbol{\mu}_t)_i$ and $\hat{M}_{ij}^t = 0$ for $j \neq i$. The CGM, CNP, and SBP-EM algorithms assume the underlying Markov chains are time-homogeneous, while our proposed methods can estimate time-varying transition probabilities by learning the underlying cost matrices.

The proposed methods were evaluated in terms of estimating the transition flows $\{\mathbf{M}^t\}_{t=1}^{T-1}$ only using the marginally aggregated count observations $\{\tilde{\boldsymbol{\mu}}_t\}_{t=1}^T$. The performance in estimating transition flows is evaluated using the

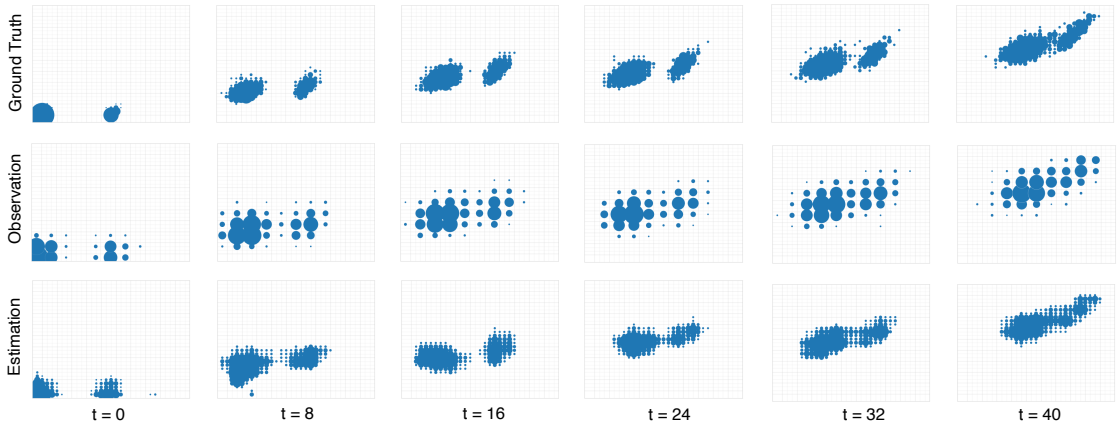

Figure 4: The top plots shows a simulated ensemble flow of $1,000$ particles moving over a 20×20 grid cells over 40 time points; the middle displays the noisy observations of this ensemble flow; the bottom shows the distributions estimated by the proposed method for each corresponding time point. The size of the blue dots is proportional to the number of particles at the corresponding state.

Table 1: Normalized mean absolute error (NMAE) for the estimation of transition flows in the real-world datasets

	Beijing Taxi	San Francisco Cabs	Tokyo Flow	Chukyo Flow	Synthetic	
					time-varying	time homogeneous
STAY	0.378	0.346	0.186	0.187	0.667	0.725
CGM	0.301	0.307	0.181	0.448	0.512	0.634
CNP	0.375	0.296	0.182	0.179	0.428	0.576
SBP-EM	0.344	0.291	0.347	0.375	0.416	0.489
SBP-ISTC	**0.244**	**0.167**	0.166	0.145	**0.389**	**0.473**
SBP-ISTA	0.253	0.187	**0.156**	**0.129**	0.408	0.488

normalized mean absolute error (NMAE) defined by

$$\text{NMAE} = \frac{\sum_{t=1}^{T-1}\sum_{i=1}^{S}\sum_{j\in\mathcal{N}_i}|\hat{M}_{ij}^t - \bar{M}_{ij}^t|}{\sum_{t=1}^{T-1}\sum_{i=1}^{S}\sum_{j\in\mathcal{N}_i}\bar{M}_{ij}^t},$$

where $\mathcal{N}_i$ stands for the set of neighbor states of state i, $\bar{M}_{ij}^t$ is the true number of transitions from state i at time t to state j at time $t+1$, and $\hat{M}_{ij}^t$ denotes the corresponding estimate.

5.1 Synthetic data. We consider simulating an ensemble of $1,000$ individuals moving over a 20×20 grid cells, as shown in Fig. 4. The goal of these individuals is to move from bottom-left and bottom middle corners to top-right corner. In particular, the dynamic behaviors of these individuals are determined by a log-linear distribution, which is modeled by four factors: the physical distance between two states, the angle between moving direction and the direction to the destination, the preference to stay in the original state, and the angle between the moving direction and an external force. The parameters of the log-linear model for the first three factors, are set to be $(3, 5, 10)$, respectively. To simulate time-varying dynamic behaviors of individuals, the angle between the moving direction and an external force is set to be π/t for $t = 1, \ldots, T$. There are 64 sensors placed over the grids as shown in Fig. 3. Instead of collecting the full trajectories of all the particles, these sensors can only measure an aggregated count of individuals currently being observed. The probability of an individual being detected, decreases exponentially as the distance between the individual and the sensor increases. As shown in Fig. 4, although the sensor observations only roughly record the aggregated counts of individuals, the proposed method still well estimate the population flows with a high resolution. Table 1 shows the improved performance of the proposed methods in estimating latent population flow, compared against time-homogeneous models (CGM, CNP, SBP-EM). For a fair comparison, we also consider setting the angle between moving direction and external force to be $\pi/2$ for the whole interval, to simulate an ensemble flow with a time homogeneous transition parameter. Although the proposed model slightly outperforms the other baselines, SBP-EM achieves almost the same accuracy in estimating latent transition flows of a time homogeneous Markov model.

5.2 Real-world data. The performance of the proposed methods in estimating transition flows from aggregated count data, is evaluated using four real-world population flow data. The Beijing Taxi data [30] consists of $10,357$ taxi trajectories collected from February 2, 2008 to February 8,

2008. The grid sizes in this data are 2km×2km(17×17 grid cells), and thus the number of states is 289. The time grid is 15 minutes, and thus the aggregated observations were made for 96 time steps for one day. The second data collected 537 taxi cabs' GPS traces in San Francisco (SF) from to May 18 2008 to May 30 2008. The grid sizes in this data are 2km×2km(13×13 grid cells), and thus the number of states is 169. The time grid is 15 minutes, and thus the aggregated observations were made for 96 time steps for one day. The Tokyo People Flow data[3] consists of 6,432, 9,166, 6,822, 10,134, 6,646, 10,338 individual trajectories on six days in the year of 2013. The grid sizes in this data are 10km×10km(15×15 grid cells), and thus the number of states is 225. The time grid is 30 minutes, and thus the aggregated observations were made for 48 time steps for one day. The Chukyo Flow data cosists of 975, 1,372, 1,195, 1,506, 1,021, 1,615 individuals also on six days in the year of 2013. The data is created in the same ways as Tokyo Flow data, except the grid sizes are 10km×10km(10×10) grid cells.

Results. The normalized absolute error averaged over all the time steps for each data, is presented in Table 1. For all the datasets, the proposed methods achieved higher accuracy than the other methods. In particular, we found that our proposed methods outperform the closely related SBP-EM algorithm by allowing the underlying cost matrices to be time-varying. In addition, we found that the SBP-ISTA performed better than SBP-ISTC in estimating transition flows in the Tokyo and Chukyo People Flow data. We looked into this data, and found that most individuals were moving from outer suburb regions to inner downtown areas in the morning, while transitioning on the opposite direction in the evening. The transition flows collected in this data, exhibit asymmetric moving patterns at different time steps. Hence, SBP-ISTA achieved higher accuracy by constructing more complicated cost matrices, compared with SBP-ISTC that enforces symmetric structured cost matrices.

6 Conclusion

This paper proposed to estimate latent transition flows from marginally aggregated observations, via a tree-structured multi-marginal optimal transport framework. More specifically, the proposed methods allow the transition kernels behind population flows to be time-varying, by learning the time-dependent cost functions. In particular, the cost matrices can be uniquely recovered with theoretical guarantees, by imposing structural restrictions over those cost matrices. Hence, the transition flows can be estimated with time-varying transitioning parameter, using Sinkhorn belief propagation algorithm. We demonstrate how the proposed method estimate the latent ensemble flows, using both time homogeneous and time-varying synthetic dynamic flows. The experiments on four real-world population flow data, show the improved accuracy of the proposed methods in estimating latent transition flows, compared with the others assuming time-homogeneous transition kernels.

[3]Data sources: SNS-based People Flow Data, http: //nightley.jp/archives/1954

7 Acknowledgements

This work is partially supported by a grant from the Shenzhen Science and Technology Program (JCYJ20210324120011032) and Shenzhen Institute of Artificial Intelligence and Robotics for Society.

References

[1] Yasunori Akagi, Takuya Nishimura, Yusuke Tanaka, Takeshi Kurashima, and Hiroyuki Toda. Exact and efficient inference for collective flow diffusion model via minimum convex cost flow algorithm. In *AAAI*, pages 3163–3170, 2020.

[2] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyre. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.

[3] Garrett Bernstein and Daniel Sheldon. Consistently estimating markov chains with noisy aggregate data. In *AISTATS*, pages 1142–1150, 2016.

[4] Guillaume Carlier, Arnaud Dupuy, Alfred Galichon, and Yifei Sun. Sista: learning optimal transport costs under sparsity constraints. *CoRR*, 2020.

[5] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 589(7840):82–87, January 2021.

[6] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NIPS*, pages 2292–2300, 2013.

[7] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and its applications*, 114:717–735, 1989.

[8] Isabel Haasler, Axel Ringh, Yongxin Chen, and Johan Karlsson. Estimating ensemble flows on a hidden Markov chain. In *CDC*, pages 1331–1338, 2019.

[9] Isabel Haasler, Axel Ringh, Yongxin Chen, and Johan Karlsson. Multimarginal optimal transport with a tree-structured cost and the Schrödinger bridge problem. *SIAM J. Control. Optim.*, 59(4):2428–2453, 2021.

[10] Isabel Haasler, Axel Ringh, Yongxin Chen, and Johan

Karlsson. Scalable computation of dynamic flow problems via multi-marginal graph-structured optimal transport. *CoRR*, 2106.14485, 2021.

[11] Isabel Haasler, Rahul Singh, Qinsheng Zhang, Johan Karlsson, and Yongxin Chen. Multi-marginal optimal transport and probabilistic graphical models. *IEEE Trans. Inf. Theory*, 67(7):4647–4668, 2021.

[12] Tomoharu Iwata and Hitoshi Shimizu. Neural collective graphical models for estimating spatio-temporal population flow from aggregated data. In *AAAI*, pages 3935–3942, 2019.

[13] Ruilin Li, Xiaojing Ye, Haomin Zhou, and Hongyuan Zha. Learning to match via inverse optimal transport. *J. Mach. Learn. Res.*, 20:80:1–80:37, 2019.

[14] Dixin Luo, Hongteng Xu, Yi Zhen, Bistra Dilkina, Hongyuan Zha, Xiaokang Yang, and Wenjun Zhang. Learning mixtures of Markov chains from aggregate data with structural constraints. *IEEE Transactions on Knowledge and Data Engineering*, 28(6):1518–1531, 2016.

[15] Shaojun Ma, Shu Liu, Hongyuan Zha, and Haomin Zhou. Learning stochastic behaviour from aggregate data. In *ICML*, pages 7258–7267, 2021.

[16] Shaojun Ma, Haodong Sun, Xiaojing Ye, Hongyuan Zha, and Haomin Zhou. Learning cost functions for optimal transport. *CoRR*, 2021.

[17] B. Pass. Uniqueness and Monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.

[18] B. Pass. On the local structure of optimal measures in the multi-marginal optimal transportation problem. *Calc. Var. Partial Differential Equations*, 43(3-4):529–536, 2012.

[19] Ofir Pele and Michael Werman. Fast and robust earth mover's distances. *2009 IEEE 12th International Conference on Computer Vision*, pages 460–467, 2009.

[20] Evan L Ray. Ensemble forecasts of Coronavirus disease 2019 (COVID-19) in the U.S. *medRxiv*, 2020.

[21] Daniel R Sheldon and Thomas Dietterich. Collective graphical models. In *NIPS*, pages 1161–1169, 2011.

[22] Daniel Sheldon, Tao Sun, Akshat Kumar, and Thomas G. Dietterich. Approximate inference in collective graphical models. In *ICML*, pages 1004–1012, 2013.

[23] Rahul Singh, Isabel Haasler, Qinsheng Zhang, Johan Karlsson, and Yongxin Chen. Inference with aggregate data: An optimal transport approach. *CoRR*, abs/2003.13933, 2020.

[24] Rahul Singh, Qinsheng Zhang, and Yongxin Chen. Learning hidden Markov models from aggregate observations. *CoRR*, abs/2011.11236, 2020.

[25] Tao Sun, Daniel Sheldon, and Akshat Kumar. Message passing for collective graphical models. In *ICML*, pages 853–861, 2015.

[26] Yusuke Tanaka, Tomoharu Iwata, Takeshi Kurashima, Hiroyuki Toda, and Naonori Ueda. Estimating latent people flow without tracking individuals. In *IJCAI*, pages 3556–3563, 2018.

[27] C. Villani. Topics in optimal transportation theory. 01 2003.

[28] Yichen Wang, Evangelos A. Theodorou, Apurv Verma, and Le Song. Steering opinion dynamics in information diffusion networks. *CoRR*, 2016.

[29] Sikun Yang and Hongyuan Zha. Estimating latent population flows from aggregated data via inversing multi-marginal optimal transport. *CoRR*, abs/2212.14527, 2022.

[30] Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. T-drive: Enhancing driving directions with taxi drivers' intelligence. *IEEE Trans. on Knowl. and Data Eng.*, 25(1):220–232, jan 2013.