Group-Based Dimensionality Reduction and Estimation for Heterogeneous Large-Scale Traffic Networks

Matteo Scandella[1*], Michelangelo Bin[1], and Thomas Parisini[1,2]

Abstract— State estimation for traffic networks is a particularly challenging problem in view of their large dimensionality, and since models are often inaccurate and the interaction patterns unpredictable. In this article, we approach the problem by mixing aggregation-based complexity reduction and nonlinear filtering. We subdivide vehicles into groups and derive a lower-dimensional approximate model where vehicles belonging to the same group are represented by a unique random variable matching their average characteristics. Then, we propose a procedure to estimate the statistical properties of the group variables from partial measurements. Connections to car-following models are discussed, and the developed methodology is illustrated through numerical simulations.

I. INTRODUCTION

According to the fifth IPCC report on mitigation of climate change [1], in 2010 the transport sector was responsible for approximately 23% of total energy-related CO_2 emissions [1, Chapter 8]. Moreover, "Greenhouse gas (GHG) emissions from the transport sector have more than doubled since 1970, and have increased at a faster rate than any other energy end-use sector . . . Around 80% of this increase has come from road vehicles" [1, Section 8.1, p. 605]. The negative environmental impact of road traffic is exacerbated in the presence of congestion, which also has several other consequences of financial and social nature. As road traffic increases every year, regulating traffic flows efficiently is becoming an ever more important control challenge.

Whether control is actuated through traffic lights, pricing mechanisms, ramp meters, or self-driving cars, every control policy needs information on the current traffic conditions. As measurements are generally limited, most of the required information must be estimated. However, state estimation for traffic networks is a particularly challenging problem; the available mathematical models, such as car-following equations [2]–[4], are typically inaccurate and cover only limited situations, and the interaction patterns change unpredictably. As a consequence, approaches based on canonical deterministic observer theory (e.g., [5]–[7]), where a precise knowledge of the vehicles' models and the interaction topology is required, are typically ruled out.

[1]Imperial College London, London, UK.

[2]KIOS Research and Innovation Center of Excellence, University of Cyprus, Cyprus, and Department of Engineering and Architecture, University of Trieste, Trieste, Italy.

*Corresponding author. Email: `m.scandella@imperial.ac.uk`.

Instead, stochastic and filtering-based methods [8]–[10] are structurally more suited to handle the large uncertainty characterizing traffic applications. However, typical numeric implementations, such as those based on Monte Carlo methods and particle filters [8], [9], [11], do not scale well with the dimension of the state space. Therefore, their applicability is anyway limited for traffic networks, which are often of large scale. A possible approach mitigating the effects of a large dimensionality is the Ensemble Kalman Filter [9], [12]–[14]. By mixing an extended Kalman filtering design and Monte Carlo approximations, it avoids storing and computing explicitly the estimate covariance matrix, which for large dimensions may be intractable, and marginals are instead inferred from the particles empirical distribution. Unfortunately, the full dimensionality of the state space is maintained.

In this article, we pursue a different approach mixing nonlinear filtering and a priori model reduction by clustering. We consider a possibly large number of vehicles modeled as Markov processes. We first subdivide the vehicles into a fixed number of groups (for instance, self-driving and human-driven cars, motorbikes or trucks). For each group, we define a representative "group variable" modeling the "average" individual. Then, we construct a reduced model describing the evolution of the group variables. The model delivers, for each group, an approximate distribution of which the vehicles belonging to the group can be seen as samples. Finally, we devise a procedure to estimate the lower-dimensional group model from partial observations.

The idea of reducing complexity by clustering is certainly not new [15]. To name a few related works, complexity reduction based on equivalence relations is used in [16] to simplify control designs in Boolean networks. The problem of clustering and estimation of the aggregated states in networks of linear systems is studied in [17]. In [18] and [19], model similarity-based clustering and mean-field approximations are combined to simplify the policy design in major-minor LQG games. Mean-field approximations are also used in [20], where self and human-driven vehicles are modeled by means of a set of interconnected PDEs. Condensation of the network's connected components is used in [21] to decompose the optimization problem underlying the control policy design into a sequence of lower-complexity problems. Condensation of weakly connected components is also used in [22] to estimate the travel time in large traffic networks. Nevertheless, at the best of authors' knowledge, the integration of similarity-based clustering and filtering of large-scale traffic networks is novel.

Organization. In Section II, we detail the traffic network's model and the main assumptions. In Section III, we develop the group-based lower-dimensional approximation of the original network. In Section IV, we introduce the measurement model and relate it to the group variables. In Section V, we detail the proposed estimation procedure. In Section VI, we make a parallel with microscopic car-following models and, finally, in Section VII we present some numerical simulations.

Notation. We denote by $\mathbb{R}$ and $\mathbb{N}$ the set of real and natural numbers, respectively. If $\leq$ is a preorder on S, for every $x, y \in S$ we let $x : y = S_{\geq x} \cap S_{\leq y}$. Non-strict set inclusion is denoted by $\subseteq$. By $A \setminus B$ we denote the set difference between A and B. If $B = \{b\}$ is a singleton, we write $A \setminus b$ for $A \setminus \{b\}$. If $B = (x_i)_{i=1,\dots,n}$ is an n-tuple, we write $A \setminus B$ for $A \setminus \{x_1, \dots, x_n\}$. If $(x_i)_{i \in \mathcal{I}}$ is a family indexed by $\mathcal{I}$ and $I \subseteq \mathcal{I}$, we let $x_I := (x_i)_{i \in I}$. All probability distributions introduced in the following are defined on an underlying probability space (Ω, Σ, P). Throughout the article, "density" is used as short-hand for "probability density function".

II. THE FRAMEWORK

A. Traffic Networks

We consider a set $\mathcal{V}$ of $n \in \mathbb{N}_{>0}$ vehicles driving on a cyclic road. We denote by $\sigma^1(i)$ the vehicle immediately in front of $i \in \mathcal{V}$, and by $\sigma^{-1}(i)$ the one immediately behind[^1]. We assume each driver is influenced only by the first d vehicles in front and the first d vehicles behind, for some arbitrary $d \in \{1, \dots, n-1\}$. For each $i \in \mathcal{V}$, we define the tuples $B_i := (\sigma^{-d}(i), \dots, \sigma^{-1}(i)) \in \mathcal{V}^d$ and $F_i := (\sigma^1(i), \dots, \sigma^d(i)) \in \mathcal{V}^d$ called, respectively, the *backward* and *forward neighborhood* of i and collecting, respectively, the first d vehicles behind and in front of i, on which the driving choices of i depends. For notation convenience, we let $\Delta_i := (B_i, i, F_i) = (\sigma^{-d}(i), \dots, \sigma^{-1}(i), i, \sigma^1(i), \dots, \sigma^d(i)) \in \mathcal{V}^d$, to which we refer as the *dependency tuple* of i.

We suppose that each vehicle belongs to exactly one among m pre-specified *groups* clustering vehicles according to an arbitrary criterion. For instance, groups may distinguish self-driving or human-driven cars, motorcycles, and trucks. Formally, a *group assignment* $G = (G_k)_{k=1,\dots,m}$ is a sequence of m non-empty disjoint sets $G_1, \dots, G_m \subseteq \mathcal{V}$ covering $\mathcal{V}$. We denote by $\mathcal{G}$ the set of all possible group assignments and, for notation convenience, we let $\mathcal{K} := \{1, \dots, m\}$. In the following, we identify groups G_k with their indices k and call the elements of $\mathcal{K}$ groups as well.

We associate with each vehicle $i \in \mathcal{V}$ a parameter h_i ranging in a given subset $\mathcal{H} \subseteq \mathbb{R}^{n_h}$, $n_h \in \mathbb{N}_{>0}$, and representing the vehicle's unique characteristics in terms of way of driving and interacting with the other vehicles. For instance, h_i may coincide with the parameters defining a car-following model [2], [4]. We denote by $\mathcal{H}^{\mathcal{V}}$ the set of all possible parameter assignments $i \mapsto h_i$. With these definitions in

mind, we define a *traffic network* as a tuple $(\mathcal{V}, \sigma, G, h)$ where $G : \Omega \to \mathcal{G}$ is a random group assignment, and $h : \Omega \to \mathcal{H}^{\mathcal{V}}$ a random parameter assignment. For notation convenience, for each $i \in \mathcal{V}$ we define the variable g_i in such a way that, for each $\omega \in \Omega$, $g_i(\omega)$ equals the (unique) $k \in \mathcal{K}$ such that $i \in G_k(\omega)$. We let $g := (g_i)_{i \in \mathcal{V}} \in \mathcal{K}^n$.

Throughout the paper, we focus on traffic networks satisfying the following group-homogeneity assumption.

Assumption 1: There exists a (known) $\rho = (\rho_i)_{i \in \mathcal{K}} \in (0, 1)^m$ such that, for every $\bar{g} = (\bar{g}_i)_{i \in \mathcal{V}} \in \mathcal{K}^n$,

$$P(g = \bar{g}) = \prod_{i \in \mathcal{V}} \rho_{\bar{g}_i} \qquad \triangleleft$$

This assumption states that the probability of the vehicle $i \in \mathcal{V}$ to be in the group G_k is equal to ρ_k and that this event is independent of the groups of all the other vehicles.

The interaction network previously defined by σ is supposed static, since σ does not depend on time. Nevertheless, this is done only for ease of exposition and we underline that, at least for what concerns the estimation procedure presented in Section V, time invariance of the interaction topology is not actually required, provided that the homogeneity of group distributions asked by Assumption 1 holds at all times $t \in \mathbb{N}$.

B. Vehicles Dynamics

Each vehicle $i \in \mathcal{V}$ is modeled by means of a (discrete-time) stochastic process $(\omega, t) \in \Omega \times \mathbb{N} \mapsto x_i^t(\omega) = (d_i^t(\omega), v_i^t(\omega), \theta_i^t(\omega)) \in \mathcal{X} := \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \Theta$, in which $d_i^t(\omega) \in \mathbb{R}_{\geq 0}$ is the distance of vehicle i to its successor $\sigma(i)$, $v_i^t(\omega) \in \mathbb{R}_{\geq 0}$ is the vehicle's cruise speed, and $\theta_i^t(\omega) \in \Theta \subseteq \mathbb{R}^{n_\theta}$, $n_\theta \in \mathbb{N}$, gathers additional vehicle-specific state variables that can be used to accommodate different advanced car-following models. In the following, we let $x^t := (x_i^t)_{i \in \mathcal{V}} \in \mathcal{X}^n$, and we omit the argument ω when not strictly necessary.

As the vehicles are arranged in a cycle, the length of the road is given by $\ell := \sum_{i \in \mathcal{V}} d_i^0$. Notice that this definition implies that, in addition to the subdivision into groups and the individual parameters, in our setting also the road length ℓ is a random variable. Instead of fixing ℓ beforehand, proceeding in this way allows us to handle cyclic roads while preserving independence of the initial conditions x_i^0, as implied by the forthcoming Assumption 2.

Assumption 2: There exist (known) $\lambda : \mathcal{H} \times \mathcal{K} \to [0, \infty]$ and $\pi^0 : \mathcal{X} \times \mathcal{H} \to [0, \infty]$ satisfying

$$\int_{\mathcal{H}} \lambda(\eta, k) \, d\eta = 1, \qquad \forall k \in \mathcal{K},$$

$$\int_{\mathcal{X}} \pi^0(\xi, h) \, d\xi = 1, \qquad \forall h \in \mathcal{H},$$

such that, for all $i \in \mathcal{V}$, the variables x_i^0, h_i, and g_i have joint density

$$\psi^0(\xi, \eta, k) = \rho_k \pi^0(\xi, h) \lambda(\eta, k). \qquad (1)$$

$$\triangleleft$$

In Assumption 2, λ plays the role of a density function for h_i given that $i \in G_k$, and models the fact that the

[^1]: Formally, $\sigma : \mathcal{V} \to \mathcal{V}$ is a bijection with the property that for any two $i, j \in \mathcal{V}$ there exists $k \in \mathbb{N}$ such that $\sigma^k(i) = j$.
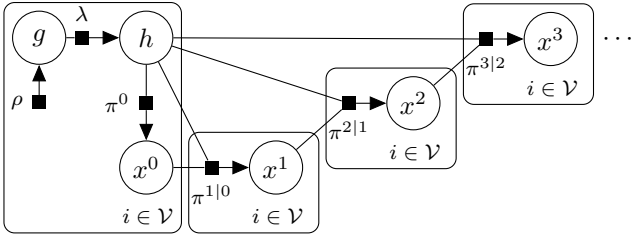
Fig. 1. Factor graph [23] of the stochastic model (2).

parameters corresponding to vehicles in the same group are identically distributed (although, we stress, they are not equal). Likewise, π^0 plays the role of a density function for the initial condition x_i^0 given the value of the parameter h_i. In the following, we adopt the more suggestive notation $\lambda(\cdot|k) := \lambda(\cdot, k)$ and $\pi^0(\cdot|h) := \pi^0(\cdot, h)$, and we shall do the same for all functions having a similar interpretation.

Regarding the distribution of the vehicles' variables at times $t > 0$, we make the following Markov assumption.

Assumption 3: There exists a family $\left(\pi^{t|t-1}\right)_{t\in\mathbb{N}_{>0}}$ of (known) functions $\pi^{t|t-1} : \mathcal{X} \times \mathcal{X}^{2d+1} \times \mathcal{H} \to [0, \infty]$ satisfying

$$\int_{\mathcal{X}} \pi^{t|t-1}\left(\xi|\chi, \eta\right) d\xi = 1, \quad \forall (\chi, \eta) \in \mathcal{X}^{2d+1} \times \mathcal{H},$$

such that, for each $t \in \mathbb{N}$, the variables $x^{0:t}$, h and g have joint density

$$\Psi^{0:t}\left(x^{0:t}, h, g\right) = \Psi^0\left(x^0, h, g\right) \prod_{i\in\mathcal{V}} \prod_{\tau=1}^{t} \pi^{\tau|\tau-1}\left(x_i^\tau|x_{\Delta_i}^{\tau-1}, h_i\right)$$

(2)

with $\Psi^0\left(x^0, h, g\right) := \prod_{i\in\mathcal{V}} \psi^0\left(x_i^0, h_i, g_i\right).$ ◁

The quantity $\pi^{t|t-1}\left(x_i^\tau|x_{\Delta_i}^{\tau-1}, h_i\right)$ is interpreted as the density of x_i^t given the value of the parameter h_i of i and that of the state variables x_j^{t-1} at time $t-1$ for all vehicles j in the dependency tuple Δ_i of i. Later in Section VI, we discuss a method to construct $\pi^{t|t-1}$ from the knowledge of a difference equation satisfied by the state variables x_i^t. We underline that Assumption 3 implies independence of x_i^0 from $\left(x_j^0, h_j, g_j\right)$ for all $j \in \mathcal{V} \setminus i$.

Under Assumption 3, we can write

$$\Psi^{0:t}\left(x^{0:t}, h, g\right) =$$
$$= \prod_{i\in\mathcal{V}} \rho_{g_i} \pi^0\left(x_i^0|h_i\right) \lambda\left(h_i|g_i\right) \prod_{\tau=1}^{t} \pi^{\tau|\tau-1}\left(x_i^\tau|x_{\Delta_i}^{\tau-1}, h_i\right).$$

Additionally, the joint density (2) can be described graphically using the factor graph [23] in Figure 1. Finally, we define the conditional distributions relating the variables at time t and those at time $t-1$ as

$$\Psi^{t|t-1}\left(x^t, h, g|x^{t-1}\right) = \prod_{i\in\mathcal{V}} \rho_{g_i} \lambda\left(h_i|g_i\right) \pi^{t|t-1}\left(x_i^t|x_{\Delta_i}^{t-1}, h_i\right).$$
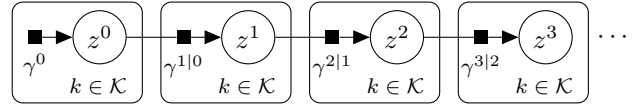
(3)

Fig. 2. Factor graph [23] of the stochastic model (4).

III. GROUP-BASED DIMENSIONALITY REDUCTION

As the cardinality n of $\mathcal{V}$ increases, working with the joint distribution $\Psi^{0:t}$ may easily become impractical due to the rapid increase with n of the dimension of the vehicles' state space. In this section, we develop a low-dimensional approximation of $\Psi^{0:t}$ by clustering vehicles belonging to the same group and representing them with a single "group variable" modeling the "average" individual of the group. The methodology developed in this section is a fundamental part of the estimation procedure presented in Section V.

A. Main Idea

The basic idea is to substitute the state $x_{G_k}^t$ of all vehicles in a group G_k with a single representative *group variable* z_k^t distributed in a way that matches the average group characteristics. Then, we substitute the model joint distribution $\Psi^{0:t}$ with a function of the form (cf. (2))

$$\widehat{\Psi}^{0:t}\left(z^{0:t}\right) := \prod_{k\in\mathcal{K}} \gamma_k^0\left(z_k^0\right) \prod_{\tau=1}^{t} \gamma_k^{\tau|\tau-1}\left(z_k^\tau|z^{\tau-1}\right),$$

(4)

in which $z^t := \left(z_k^t\right)_{k\in\mathcal{K}}$, $z^{0:t} = \left(z^\tau\right)_{\tau=0,\dots,t}$, and γ_k^0 and $\gamma_k^{t|t-1}$ are functions playing the same role, respectively, of π^0 and $\pi^{t|t-1}$ in (1) and (2). Specifically, γ_k^0 models the density of the group variable z_k^0, which is representative of the initial condition x_i^0 of all vehicles i belonging to the same group G_k. The initial conditions of such vehicles, indeed, may be thought as being distributed according to γ_k^0. Instead, $\gamma_k^{t|t-1}\left(\cdot|z^{t-1}\right)$ represents the density of the group variable z_k^t given the value of z^{t-1} at previous time $t-1$. The joint density (4) can be described graphically using the factor graph [23] in Figure 2.

We underline that, since we only keep m group variables, one for each group, the resulting model concerns probability distributions over a state space of a dimension that scales in m but not in n as long as m is fixed independently on n. For instance, in the example of Section VII, we only use $m = 2$ groups (human-driven and self-driving vehicles) independently on n. Nevertheless, we also remark that the approximation involves averages over all group characteristics (the parameters h_i), so as clustering together vehicles with high variability may result in an inaccurate approximation.

In the remainder of this section, we construct and characterize the functions γ_k^0 and $\gamma_k^{t|t-1}$.

B. Construction of γ_k^0

For each $i \in \mathcal{V}$, the joint density of $\left(x_i^0, h, g\right)$ can be obtained as

$$\alpha_i^0\left(x_i^0, h, g\right) = \int_{\mathcal{X}^{n-1}} \Psi^0\left(x^0, h, g\right) dx_{\mathcal{V}\setminus i}^0.$$

With $\delta_{i,j}$ denoting the Kronecker delta function, we define $\gamma_k^0 : \mathcal{X} \to [0,\infty]$ as the marginal

$$\gamma_k^0 (\cdot) := \frac{1}{c_k} \sum_{g \in \mathcal{K}^n} \int_{\mathcal{H}^{\mathcal{V}}} \sum_{i \in \mathcal{V}} \delta_{g_i,k}\, \alpha_i^0 (\cdot, h, g)\; dh,$$

in which c_k is the normalization constant

$$c_k := \int_{\mathcal{X}} \sum_{g \in \mathcal{K}^n} \int_{\mathcal{H}^{\mathcal{V}}} \sum_{i \in \mathcal{V}} \delta_{g_i,k}\, \alpha_i^0 (\xi, h, g)\; dh\, d\xi.$$

The following lemma expresses γ_k^0 in terms of the known quantities π^0 and η.

Lemma 1: Under Assumptions 1, 2, and 3,

$$\gamma_k^0 \left(z_k^0 \right) = \int_{\mathcal{H}} \pi^0 \left(z_k^0 | \eta \right) \lambda \left(\eta | k \right)\; d\eta, \qquad \forall k \in \mathcal{K}.$$

$\square$

It is worth noticing that, according to Lemma 1, γ_k^0 equals the average of the initial-state densities $\pi^0 (\cdot | h_i)$ over all parameters h_i associated with group G_k.

C. Construction of $\gamma_k^{t|t-1}$

For each $t \in \mathbb{N}_{>0}$, we define

$$\overline{\Psi}^{t|t-1} \left(x^t, h, g \,\middle|\, z^{t-1} \right) := \Psi^{t|t-1} \left(x^t, h, g \,\middle|\, (z_{g_i}^t)_{i \in \mathcal{V}} \right),$$

which is obtained by substituting each x_i^{t-1} with its group variable $z_{g_i}^{t-1}$ in (3). Then, along the lines of the previous section, for each $i \in \mathcal{V}$, we define

$$\alpha_i^{t|t-1} \left(x_i^t, h, g \,\middle|\, z^{t-1} \right) := \int_{\mathcal{X}^{n-1}} \overline{\Psi}^{t|t-1} \left(x^t, h, g \,\middle|\, z^{t-1} \right)\; dx_{\mathcal{V} \setminus i}^t.$$

Finally, we let

$$\gamma_k^{t|t-1}(\cdot | z^{t-1}) := \frac{1}{b_k} \sum_{g \in \mathcal{K}^n} \sum_{i \in \mathcal{V}} \delta_{g_i,k} \int_{\mathcal{H}^{\mathcal{V}}} \alpha_i^{t|t-1}(\cdot, h, g | z^{t-1})\, dh$$

in which $b_k := \rho_k n$ is a normalization constant. The following lemma expresses $\gamma_k^{t|t-1}$ in terms of λ and $\pi^{t|t-1}$. To simplify the notation, for $\nu \in \mathcal{V}^{2d}$ and $t \in \mathbb{N}$, we let $z_{k \cup \nu}^t := (z_{\nu_1}^t, \ldots, z_{\nu_d}^t, z_k^t, z_{\nu_{d+1}}^t, \ldots, z_{\nu_{2d}}^t)$. The variable $z_{k \cup \nu}^t$ approximates the dependency tuple of the representative variable z_k^t when the backward neighbors belong to the groups $(\nu_q)_{q=1,\ldots,d}$ and the forward neighbors to the groups $(\nu_q)_{q=d+1,\ldots,2d}$.

Lemma 2: Under Assumptions 1, 2, and 3,

$$\gamma_k^{t|t-1}(\cdot | z^{t-1}) = \sum_{\nu \in \mathcal{K}^{2d}} \prod_{q=1}^{2d} \rho_{\nu_q} \int_{\mathcal{H}} \pi^{t|t-1}(\cdot | z_{k \cup \nu}^{t-1}, \eta) \lambda(\eta|k)\, d\eta.$$

$\square$

It is worth observing that, according to Lemma 2, $\gamma_k^{t|t-1}(\cdot | z^{t-1})$ equals the average over all possible parameters associated with group G_k and all possible neighboring groups of the transition probabilities $\pi^{t|t-1}$ (cf. Lemma 1).

IV. The Measurement Model

In this section we consider the problem of transferring the information about individual vehicles coming from measurements to information about group variables. In particular, we suppose that, at each time $t \in \mathbb{N}_{>0}$, we are given a sequence of measurements $y^t := (y_i^t)_{i \in U^t}$ regarding a subset $U^t \subseteq \mathcal{V}$ of vehicles with cardinality u_t, where each y_i^t is a random variable taking values in a (common) subset $\mathcal{Y} \subseteq \mathbb{R}^{n_y}$. Moreover, we suppose we know the likelihood of such outcomes according to the following assumption.

Assumption 4: For each $t \in \mathbb{N}_{>0}$, there exists a (known) function $\mu^t : \mathcal{Y}^{u_t} \times \mathcal{X}^{u_t} \times \mathcal{K}^{u_t} \to [0,\infty]$, satisfying

$$\int_{\mathcal{Y}^{u_t}} \mu^t (\zeta | \chi, k)\; d\zeta = 1, \qquad \forall (\chi, k) \in \mathcal{X}^{u_t} \times \mathcal{K}^{u_t},$$

such that $y^{0:t}$, $x^{0:t}$, h and g have joint density

$$\Upsilon^{0:t} \left(y^{0:t}, x^{0:t}, h, g \right) = \Psi \left(x^{0:t}, h, g \right) \prod_{\tau=1}^{t} \mu^\tau \left(y^\tau \,|\, x_{U^\tau}^\tau, g_{U^\tau} \right).$$

$\triangleleft$

For each $t \in \mathbb{N}_{>0}$, $\mu^\tau (\cdot | x_{U^\tau}^\tau, g_{U^\tau})$ represents the density of y^t given the value of the groups g_{U^τ} and all state variables x_i^t with $i \in U^t$. A method to construct μ from an individual measurement model is discussed in Section VI.

Next, we define the function $\widehat{\mu}^t : \mathcal{Y}^{u_t} \times \mathcal{X}^m \to [0,\infty]$ as

$$\widehat{\mu}^t \left(y^t | z^t \right) := \mu^t \left(y^t \,\middle|\, (z_{g_i}^t)_{i \in U^t}, g_{U^\tau} \right), \tag{5}$$

which is obtained from μ^t by substituting each x_i^t with the corresponding group variable $z_{g_i}^t$. Finally, according to (4) and Assumption 4, in the terms given in (5) we approximate the joint density of $y^{0:t}$ and $z^{0:t}$ as

$$\widehat{\Upsilon}^{0:t}(y^{0:t}, z^{0:t}) = \prod_{k \in \mathcal{K}} \gamma^0 (z_k^0) \prod_{\tau=1}^{t} \gamma_k^{\tau|\tau-1}(z_k^\tau | z^{\tau-1}) \widehat{\mu}^\tau (y^\tau | z^\tau). \tag{6}$$

V. The Filtering Procedure

In this section, we propose a filtering procedure employing the low-dimensionality model developed in Section III and the measurement model of Section IV to estimate, for each $k \in \mathcal{V}$ and $t \in \mathcal{N}$, statistics about the group variables. In particular, we are interested in approximating integrals of the form

$$I^t \left(\Phi^t \right) = \int_{\mathcal{X}^{m(t+1)}} \Phi^t (\xi)\, \zeta^{0:t} \left(\xi | y^{0:t} \right)\; d\xi$$

where, for all $t \in \mathbb{N}$, $\Phi^t : \mathcal{X}^{m(t+1)} \to \mathbb{R}$ is a "test function" describing the properties of the vehicles of interest (e.g., the projection $\Phi^t((d_k^\tau, v_k^\tau, \theta_k^\tau)_{k \in \mathcal{K}, \tau \in 0:t}) = v_j^t$ gives the velocity of a vehicle in the j-th group at time t), and

$$\zeta^t \left(\cdot | y \right) := \frac{\widehat{\Upsilon} \left(y^{0:t}, \cdot \right)}{\int_{\mathcal{X}^{m(t+1)}} \widehat{\Upsilon} \left(y^{0:t}, \xi \right)\; d\xi}$$

can be interpreted as the posterior distribution of the group variables given the available measurements.

The main idea behind the proposed procedure is to use a Sequential Monte Carlo (SMC) approach. This methodology mixes *prediction* steps, where the functions $\pi^{t|t-1}$ (Lemma 2) are employed, and *correction* steps, where the measurement model $\widehat{\mu}^t$ (equation (5)) is used. The functions λ and π^0 are instead used in the initialization phase. In particular, at each $t \in \mathbb{N}$, the methodology extracts S samples z_s^t, associates with each sample a weight $w_s^t \in [0,1]$, and then approximates I^t as

$$I^t\left(\Phi^t\right) \approx \widehat{I}^t\left(\Phi^t\right) := \sum_{s=1}^{S} w_s^t \Phi^t\left(z_s^t\right).$$

Since (6) can be seen as a classical Hidden Markov Model (HMM), we propose to use a bootstrap particle filter with multinomial resampling [8], [11]. However, in certain conditions, it is possible to use more advanced versions such as the auxiliary particle filter [24], or use other resampling techniques in case of high path degeneracy. The overall approximation procedure is detailed below:

Algorithm: Estimation of $I^t\left(\Phi^t\right)$

A. Initialization:

1. **Propagation:** For each group $k \in \mathcal{K}$:

   (a) Extract S samples $(h_{k,s}^0)_{s\in\mathcal{S}}$ from $\lambda(\cdot|k)$.

   (b) Extract S samples $(z_{k,s}^0)_{s\in\mathcal{S}}$ in such a way that, for each $s = 1,\ldots,S$, $z_{k,s}^0$ is sampled from $\pi^0(\cdot|h_{k,s}^0)$.

2. **Concatenation:** Set $z_s^0 = (z_{k,s}^0)_{k\in\mathcal{K}}$

3. **Weighting:** Define the weights $w_s^0 := 1/S$, $\forall s \in \mathcal{S}$.

B. Iteration: For each $t \in \mathbb{N}_{>0}$:

1. **Resampling:** Extract S samples $(a_s^t)_{s\in\mathcal{S}} \in \mathcal{S}^S$ randomly in such a way that, for each $s,q \in \mathcal{S}$, $a_s^t = q$ with probability w_q^{t-1}.

2. **Propagation:** for each group $k \in \mathcal{K}$:

   (a) Extract S samples $(h_{k,s}^t)_{s\in\mathcal{S}}$ from $\lambda(\cdot|k)$.

   (b) for each possible position $q \in \{1,\ldots,2d\}$ in the neighborhood, extract S samples $(g_{k,q,s}^t)_{s\in\mathcal{S}}$ randomly in such a way that, for every $q \in \{1,\ldots,2d\}$, every $s \in \mathcal{S}$, and every $r \in \mathcal{K}$, $g_{k,q,s}^t = r$ with probability ρ_r.

   (c) Extract S samples $(z_{k,s}^t)_{s\in\mathcal{S}}$ in such a way that, for every $s \in \mathcal{S}$, $z_{k,s}^t$ is sampled from $\pi^{t|t-1}(\cdot|\chi_{k,s}^t, h_{k,s}^t)$, with $\chi_{k,s}^t$ the $(2d+1)$-ple

   $$\chi_{k,s}^t := \left(z_{g_{k,1,s}^t, a_s^t}^{t-1}, \ldots, z_{g_{k,d,s}^t, a_s^t}^{t-1}, z_{k,a_s^t}^{t-1}, \right.$$
   $$\left. z_{g_{k,d+1,s}^t, a_s^t}^{t-1}, \ldots, z_{g_{k,2d,s}^t, a_s^t}^{t-1}\right).$$

3. **Concatenation:** Set $z_s^t := \left(z_s^{t-1}, (z_{k,s}^t)_{k\in\mathcal{K}}\right)$

4. **Weighting:** After receiving the measurement outcomes y^t, define the weights

   $$w_s^t := \frac{\overline{w}_s^t}{\sum_{q\in\mathcal{S}} \overline{w}_q^t}, \qquad \forall s \in \mathcal{S},$$

   in which $\overline{w}_s^t := \widehat{\mu}^t\left(y^t\big|(z_{k,s}^t)_{k\in\mathcal{K}}\right)$, $\forall s \in \mathcal{S}$

VI. Distributions from Difference Equations

In this section, we discuss how the functions $\pi^{t|t-1}$ and μ, appearing in Assumptions 2, 3 and 4 and supposed known, can be computed when the state variable x_i^t of each vehicle satisfies a (stochastic) difference equation of a certain kind. The class of models considered includes the relevant case of car-following models, which are widespread in the modeling of human-driven and self-driven vehicles [2], [4].

We suppose that x satisfies

$$x_i^{t+1} = f\left(x_{\Delta_i}^t, h_i, w_i^t\right), \qquad y_i^t = \varrho\left(x_i, \nu_i^t\right), \quad (7)$$

for all $t \in \mathbb{N}$, in which $f : \mathcal{X}^{2d+1} \times \mathcal{H} \times \mathcal{W} \to \mathcal{X}$, $\varrho : \mathcal{X} \times \mathcal{Q} \to \mathcal{Y}$, and, for each $i \in \mathcal{V}$, $w_i = (w_{i,j}^t)_{j=1,\ldots,n_x}$: $\Omega \times \mathbb{N} \to \mathcal{W} \subseteq \mathbb{R}^{n_x}$, $n_x = 2 + n_\theta$, and $\nu_i = (\nu_{i,j}^t)_{j=1,\ldots,n_y}$: $\Omega \times \mathbb{N} \to \mathcal{Q} \subseteq \mathbb{R}^{n_y}$ are stochastic processes and Δ_i is the dependency tuple of vehicle i. We restrict the class of equations to those satisfying the following assumptions.

Assumption 5: For each $(\chi,\eta) \in \mathcal{X}^{2d+1} \times \mathcal{H}$ (resp. $\chi \in \mathcal{X}$), the function $f(\chi,\eta,\cdot)$ (resp. $\varrho(\chi,\cdot)$) is invertible and continuously differentiable on an open set including $\mathcal{W}$ (resp. $\mathcal{Q}$). ◁

Assumption 6: For each $t \in \mathbb{N}$, all processes $w_{i,j}^t$ and $\nu_{r,q}^t$, $i,r \in \mathcal{V}$, $j = 1,\ldots,n_x$, $q = 1,\ldots,n_y$, are independent to one another. Moreover, $(w_{i,j}^t)_{j=1,\ldots,n_x}$ (resp. $(\nu_{r,q}^t)_{q=1,\ldots,n_y}$) have (known) joint density $\varphi_w : \mathcal{W} \to [0,\infty]$ (resp. $\varphi_\nu : \mathcal{Q} \to [0,\infty]$). ◁

Notice that the invertibility property in Assumption 5 is not too restrictive, as the zero function $(\omega,t) \mapsto 0$ can be seen as a stochastic process with density $\varphi(\cdot) = \delta(\cdot)$, and thus "dummy" inputs may be added if needed. Under Assumptions 5 and 6, for each fixed $(\chi,\eta) \in \mathcal{X}^{2d+1} \times \mathcal{H}$ and $t \in \mathcal{N}$, the random variable $\omega \mapsto f(\chi,\eta,w_i^t(\omega))$ satisfies (denote, for simplicity, $f_{\chi,\eta}(\cdot) = f(\chi,\eta,\cdot)$)

$$P\left(f\left(\chi,\eta,w_i^t\right) \in X\right) = P\left(w_i^t \in f_{\chi,\eta}^{-1}(X)\right)$$
$$= \int_{f_{\chi,\eta}^{-1}(X)} \varphi_w\left(w_i^t\right) dw_i^t$$
$$= \int_X \varphi_w\left(f_{\chi,\eta}^{-1}(\xi)\right) \left|\det \mathsf{D}\left(f_{\chi,\eta}^{-1}\right)(\xi)\right| d\xi,$$

where $\mathsf{D}(f_{\chi,\eta}^{-1})$ denotes the Jacobian matrix of $f_{\chi,\eta}^{-1}$. Hence, $f(\chi,\eta,w_i^t(\cdot))$ has density $\varphi_w\left(f_{\chi,\eta}^{-1}(\cdot)\right) \left|\det \mathsf{D}\left(f_{\chi,\eta}^{-1}\right)(\cdot)\right|$. Moreover, similar arguments show that, for every $\chi \in \mathcal{X}$, $\varrho(\chi,\nu_i^t(\cdot))$ has density $\varphi_\nu\left(\varrho_\chi^{-1}(\cdot)\right) \left|\det \mathsf{D}\left(\varrho_\chi^{-1}\right)(\cdot)\right|$.

The previous computations thus justify taking

$$\pi^{t|t-1}(\xi|\chi,\eta) := \varphi_w\left(f_{\chi,\eta}^{-1}(\xi)\right) \left|\det \mathsf{D}\left(f_{\chi,\eta}^{-1}\right)(\xi)\right|,$$

for all $(\xi,\chi,\eta) \in \mathcal{X} \times \mathcal{X}^{2d+1} \times \mathcal{H}$, and

$$\mu^t(\zeta|\chi) := \prod_{i\in U^t} \varphi_\nu\left(\varrho_\chi^{-1}(\zeta_i)\right) \left|\det \mathsf{D}\left(\varrho_\chi^{-1}\right)(\zeta_i)\right|$$

for all $(\zeta,\chi) \in \mathcal{Y}^{u_t} \times \mathcal{X}^{u_t}$.

Remark 1: It is important to notice that the filtering procedure presented in Section V only requires the computation of the density $\widehat{\mu}^t$ and the sampling from $\pi^{t|t-1}(\cdot|\chi,h)$, $\pi^0(\cdot|h)$ and $\lambda(\cdot|k)$, for each $\chi \in \mathcal{X}^{2d+1}$, $h \in \mathcal{H}$ and $k \in \mathcal{K}$.

Therefore, it is not necessary to sample from $\widehat{\mu}^t$ and to know the density $\pi^{t|t-1}(\cdot|\chi, h)$, instead we only need to be able to evaluate the density μ^t and the sampling function f. Hence, for the procedure, Assumptions 5 and 6 can be relaxed for the function f.

VII. NUMERICAL SIMULATIONS

We consider a traffic network populated by self-driving and human-driven vehicles, which we subdivide into $m = 2$ groups, distinguishing self-driven vehicles, identified by the group index $\mathsf{A} := 1$, and human-driven vehicles, identified by the index $\mathsf{H} := 2$.

Self-driving vehicles behave according to the IDM (Intelligent Driver Model) car-following model [2]. According to this model, vehicles act rationally, avoid collisions (in most situations), try to maintain a desired velocity, and acquire perfect measurements of their own state and of that of the vehicle ahead. This is a deterministic model, therefore, recalling (7), the function f does not depend on the stochastic process w_i^t and the density $\pi^{t|t-1}$ is a Dirac delta. Since self-driving vehicles act in a predetermined way, for simplicity we assume that they all have the same parameters, reported in Table I. Therefore, the density $\lambda(\cdot|\mathsf{A})$ is a Dirac delta centered at the reported parameters.

Instead, we model human-driven vehicles according to the HDM (Human Driver Model) metamodel applied to the IDM as described in [3]. HDM adds human behavior to the IDM car-following logic by considering a finite reaction time of the driver, imperfect estimation of the state of other vehicles, the ability to react to more than one vehicle ahead, and the implied use of constant velocity assumption when estimating the distance to the other vehicles. This is a stochastic model where the estimation error is modeled using multiple independent Wiener processes. It is important to notice that, for human-driven vehicles, θ_i^t contains the state of the Wiener processes. Since human drivers can have different and unmeasurable characteristic, the parameters of the human-driven vehicles are selected randomly from the ranges reported in Table I. In particular, it is possible to sample from $\lambda(\cdot|\mathsf{H})$ by sampling, independently for each human-driven vehicle, from a uniform distribution defined on the specified range.

The sampling time of all simulations is $0.1\,s$. The initial density $\pi^0(\cdot|h)$ is defined in such a way that $v_i^0 = 0\,{}^m\!/_s$ and $d_i^0 = 40\,m$, $\forall i \in \mathcal{V}$. The Wiener processes for the human-driven vehicles are initialized to 0. Figure 3 shows two possible realizations of the stochastic model (2) with $n = 20$ vehicles, $\rho_\mathsf{A} = 0.25$ and $\rho_\mathsf{H} = 0.75$.

We suppose that some vehicles can communicate periodically some measurements reporting their velocity and distance to the vehicle directly ahead with a small amount of noise. In particular, we will assume that $\mathcal{Y} \in \mathbb{R}_{\geq 0}^2$ and

$$\mu^\tau\left((y_i^\tau)_{i \in U^\tau} | x_{U^\tau}^\tau, g_{U^\tau}\right) =$$
$$= \prod_{i \in U^\tau} \mathcal{N}\left(y_i^\tau \,\middle|\, \begin{bmatrix} d_i^\tau \\ v_i^\tau \end{bmatrix}, \begin{bmatrix} \delta_{g_i} & 0 \\ 0 & \nu_{g_i} \end{bmatrix}\right)$$

Parameter	Self-driven	Human-driven
Algorithm	IDM	HDM
Desired velocity, v_0 $[m/s]$	$33.\overline{3}$	$[26, 34]$
Safe time headway, T $[s]$	1.6	$[1.4, 2]$
Maximum acceleration, a $[m/s^2]$	0.73	$[0.55, 0.75]$
Desired deceleration, b $[m/s^2]$	1.67	$[1.6, 2]$
Acceleration exponent, δ $[\cdot]$	4	$\{4\}$
Jam distance, s_0 $[m]$	2	$[3, 6]$
Reaction time, T' $[s]$	-	$[0.2, 1.2]$
Number of anticipated vehicles, n_a $[\cdot]$	-	$\{1, \ldots, 7\}$
Relative distance error, V_s $[\cdot]$	-	$[0.03, 0.08]$
Inverse TTC error, r_c $[1/s]$	-	$[0.005, 0.015]$
Error correlation time, τ $[s]$	-	$[17, 23]$

TABLE I

PARAMETERS OF THE CAR-FOLLOWING MODELS USED FOR THE TWO GROUP OF DRIVERS. FOR DETAILS ABOUT THE PARAMETERS, THE READER IS REFERRED TO [2] AND [3].

where $\mathcal{N}(\cdot|\mu, \Sigma)$ is the density of a normal distribution with mean μ and covariance matrix Σ, $\delta_\mathsf{A} = 6$, $\delta_\mathsf{H} = 80$, $\nu_\mathsf{A} = \nu_\mathsf{H} = 0.75$. By using such measurements, we then consider the problem of estimating the average states of the vehicles in both groups. Specifically, we aim to approximate $I^t(\Phi^t)$ with $\Phi^t(z^{0:t}) = (\overline{d}_\mathsf{A}^t, \overline{v}_\mathsf{A}^t, \overline{d}_\mathsf{H}^t, \overline{v}_\mathsf{H}^t)$, where $\overline{d}_\mathsf{A}^t$ and $\overline{v}_\mathsf{A}^t$ are, respectively, the distance and velocity components of z_A^t and $\overline{d}_\mathsf{H}^t$ and $\overline{v}_\mathsf{H}^t$ those of z_H^t. To this aim, we compute the estimation $\widehat{I}^t(\Phi^t)$ by using the procedure presented in Section V in three different situations:

A. $u_t = 0$, $\forall t \in \mathbb{N}$

B. $u_t = 20$, at $t = 250$, $t = 500$, $t = 750$, and $t = 1000$, while $u_t = 0$ at all other times. This corresponds to having a new set of measurements every $25\,s$. At each time, Half of the measurements are from self-driven vehicles and half from human-driven.

C. Same as **B**, but all measurements are from self-driven vehicles.

In all cases, we let $\rho_\mathsf{A} = 0.25$ and $\rho_\mathsf{H} = 0.75$. Case **A** is an open loop simulation that employs only the knowledge of the general model. Vice versa, cases **B** and **C** can be seen as two filtering problems with partial information. The evaluation of $\widehat{I}^t(\Phi^t)$ is compared with that of an oracle obtained by simulating model (2) on 50 different traffic networks with $n = 200$ and by computing the samples mean of Φ^t. The measurements of cases **B** and **C** are taken from one of these 50 simulations. The estimation are performed using the procedure explained in Section V with $S = 100$.

Figure 4 compares the estimation error, computed as the normalized difference between the estimation and the oracle, of the three different cases. Here we can observe that the proposed reduced model is able to track the test function Φ with small error (around 1%-3% of the value of the oracle), even with very few number of measurements.

VIII. CONCLUSIONS

We have proposed an estimation method for large-scale traffic networks based on a priori complexity reduction and nonlinear filtering. Complexity reduction is obtained by clustering vehicles into groups and maintaining a single representative variable for each group. Proceeding in this way
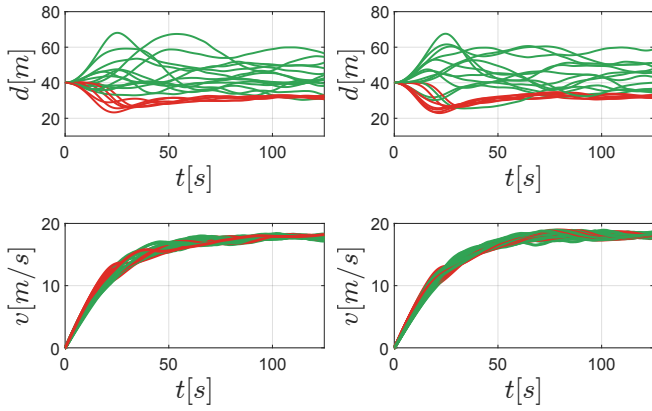
Fig. 3. Example showing two possible trajectories of the velocities and inter-vehicle distances obtained by using the model (2) with the settings explained in Section VII. Green lines refer to human-driven vehicles. Red lines to self-driving vehicles.
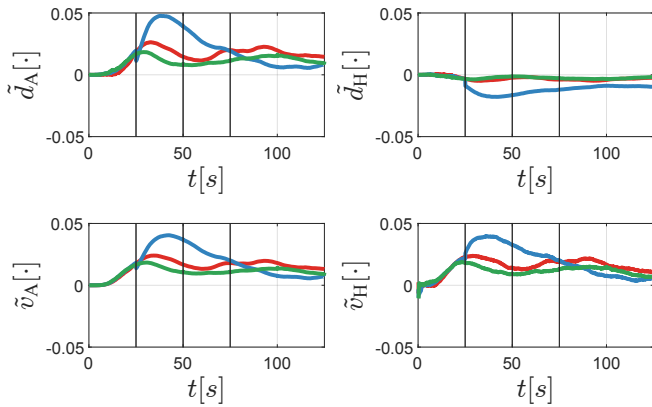

Fig. 4. Relative estimation error in case **A** (red lines), case **B** (blue lines) and **C** (green lines) with respect to the oracle for all the 4 components of $\hat{I}^t\left(\Phi^t\right)$. Left plots refer to self-driven vehicles. Right plots refer to human-driven vehicles. Top plots refer to the estimation of the distances. Bottom plots refer to the estimation of the velocities.

allows using efficient, albeit approximate, estimation techniques, such as that presented in Section V. The numerical simulations reported in Section VII show that the proposed methodology delivers a good estimate of the groups' mean velocity and inter-vehicles distance in the case in which vehicles are modeled using a mixture of IDM and HDM car-following equations.

REFERENCES

[1] O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickeland, and J. C. Minx, "IPCC, 2014: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," Cambridge University Press, Tech. Rep., 2014.
[2] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E*, vol. 62, pp. 1805–1824, 2000.
[3] M. Treiber, A. Kesting, and D. Helbing, "Delays, inaccuracies and anticipation in microscopic traffic models," *Phys. A: Stat. Mech. Appl.*, vol. 360, no. 1, pp. 71–88, 2006.
[4] A. Ferrara, S. Sacone, and S. Siri, *Freeway Traffic Modelling and Control*. Springer, 2018.
[5] G. Besançon, *Nonlinear Observers and Applications*. Springer, 2007, vol. 363.
[6] V. Andrieu and L. Praly, "On the existence of a Kazantzis-Kravaris/Luenberger observer," *SIAM J. Control Optim.*, vol. 45, no. 2, pp. 432–456, 2006.
[7] P. Bernard, *Observer design for nonlinear systems*. Springer, 2019.
[8] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, 2002.
[9] H. Fang, N. Tian, Y. Wang, M. Zhou, and M. A. Haile, "Nonlinear Bayesian estimation: from Kalman filtering to a broader horizon," *IEEE/CAA J. Autom. Sin.*, vol. 5, no. 2, pp. 401–417, 2018.
[10] A. H. Jazwinski, *Stochastic processes and filtering theory*. Academic Press, 1970.
[11] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Stat. Comput.*, vol. 10, no. 3, pp. 197–208, 2000.
[12] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *J. Geophys. Res. Oceans*, vol. 99, no. C5, pp. 10 143–10 162, 1994.
[13] G. Burgers, P. J. van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble Kalman filter," *Mon. Weather Rev.*, vol. 126, no. 6, pp. 1719–1724, 1998.
[14] J. Mandel, L. Cobb, and J. D. Beezley, "On the convergence of the ensemble Kalman filter," *Appl. Math.*, vol. 56, no. 6, pp. 533–541, 2011.
[15] M. Aoki, "Control of large-scale dynamic systems by aggregation," *IEEE Trans. Autom. Control*, vol. 13, no. 3, pp. 246–253, 1968.
[16] R. Li, Q. Zhang, and T. Chu, "On quotients of Boolean control networks," *Automatica*, vol. 125, p. 109401, 2021.
[17] M. U. B. Niazi, C. C. de Wit, and A. Y. Kibangou, "Average state estimation in large-scale clustered network systems," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 4, pp. 1736–1745, 2020.
[18] P. E. Caines and A. C. Kizilkale, "Mean field estimation for partially observed LQG systems with major and minor agents," *IFAC Proc. Vol.*, vol. 47, no. 3, pp. 8705–8709, 2014.
[19] D. Firoozi and P. E. Caines, "ϵ-Nash equilibria for major–minor LQG mean field games with partial observations of all agents," *IEEE Trans. Autom. Control*, vol. 66, no. 6, pp. 2778–2786, 2021.
[20] K. Huang, X. Di, Q. Du, and X. Chen, "Scalable traffic stability analysis in mixed-autonomy using continuum models," *Transp. Res. C: Emerg. Technol.*, vol. 111, pp. 616–630, 2020.
[21] G. Casadei and C. Canudas De-Wit, "Network games: Condensation of the graph as a hierarchical interpretation of the game," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 9655–9660, 2017.
[22] G. Casadei, V. Bertrand, B. Gouin, and C. Canudas-de Wit, "Aggregation and travel time calculation over large scale traffic networks: An empiric study on the Grenoble City," *Transp. Res. C: Emerg. Technol.*, vol. 95, pp. 713–730, 2018.
[23] B. J. Frey, "Extending factor graphs so as to unify directed and undirected graphical models," in *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, 2002, pp. 257–264.
[24] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. Am. Stat. Assoc.*, vol. 94, no. 446, pp. 590–599, 1999.