A Belief Propagation Approach to Traffic Prediction using Probe Vehicles

Cyril Furtlehner   Jean-Marc Lasgouttes   Arnaud de La Fortelle

Abstract— **This paper deals with real-time prediction of traffic conditions in a setting where the only available information is floating car data (FCD) sent by probe vehicles. Starting from the Ising model of statistical physics, we use a discretized space-time traffic description, on which we define and study an inference method based on the Belief Propagation (BP) algorithm. The idea is to encode into a graph the *a priori* information derived from historical data (marginal probabilities of pairs of variables), and to use BP to estimate the actual state from the latest FCD. The behavior of the algorithm is illustrated by numerical studies on a simple simulated traffic network. The generalization to the superposition of many traffic patterns is discussed.**

I. Introduction

With an estimated 1% GDP cost in the European Union (i.e. more than 100 billions euros), congestion is not only a time waste for drivers and an environmental challenge, but also an economic issue. Today, some urban and inter-urban areas have traffic management and advice systems that collect data from stationary sensors, analyze them, and post notices about road conditions ahead and recommended speed limits on display signs located at various points along specific routes. However, these systems are not available everywhere and they are virtually non-existent on rural areas.

In this context, the EU-funded REACT project developed new traffic prediction models to be used to inform the public and possibly to regulate the traffic, on all roads. The REACT project combines a traditional traffic prediction approach on equipped motorways with an innovative approach on non-equipped roads. The idea is to obtain floating car data from a fleet of probe vehicles and reconstruct the traffic conditions from this partial information. Two types of approaches are usually distinguished for traffic prediction, namely *data driven* (application of statistical models to a large amount of data, for example regression analysis) and *model based* (simulation or mathematical models explaining the traffic patterns). Models (see e.g. [1], [2] for a review) may range from microscopic description encoding drivers behaviors, with many parameters to be calibrated, to macroscopic ones, based on fluid dynamics, mainly adapted to highway traffic and subject to controversy [3], [4]. Intermediate kinetic

description including cellular automata [5] are instrumental for powerful simulation and prediction systems in equipped road networks [6]. On the other hand, the statistical approach mainly focuses on time series analysis on single road links, with various machine learning techniques [7], [8], while global prediction systems on a network combine data analysis and model simulations [6], [9]. For more information about traffic prediction methods, we refer also the reader to [10], [11].

We propose here a hybrid approach, by taking full advantage of the statistical nature of the information, in combination with a stochastic modeling of traffic patterns and a powerful message-passing inference algorithm. The belief-propagation algorithm, originally designed for bayesian inference on tree-like graphs [12], is widely used in a variety of inference problem (e.g. computer vision, coding theory...) but to our knowledge has not yet been applied in the context of traffic prediction. The purpose of this paper is to give the first principles of such an approach, able to exploit both space and time correlation on a traffic network. The main focus is on finding a good way to encode some coarse information (typically whether traffic on a segment is fluid or congested), and to decode it in the form of real-time traffic reconstruction and prediction. In order to reconstruct the traffic and make predictions, we propose to use the so-called Bethe approximation of an underlying disordered Ising model (see e.g. [13]), to encode the statistical fluctuations and stochastic evolution of the traffic and the belief propagation (BP) algorithm, to decode the information. Those concepts are familiar to the computer science and statistical physics communities since it was shown [14] that the output of BP is in general the Bethe approximation.

The paper is organized as follows: Section II describes the model and its relationship to the Ising model and the Bethe approximation. The inference problem and our strategy to tackle it using the belief propagation approach are stated in Section III. Section IV is devoted to a more practical description of the algorithm, and to numerical results illustrating the method. Finally, some new research directions are outlined in Section V.

II. Traffic description and statistical physics

We consider a road network for which we want to both *reconstruct* the current traffic conditions and *predict* future evolutions. To this end, all we have is partial information in the form of floating car data arriving every minute or so. The difficulty is of course that the up-to-date data only covers part of the network and that the rest has to be inferred from this.

C. Furtlehner is with Project Team TAO at INRIA Futurs, Université de Paris-Sud 11 - Bâtiment 490, 91405 Orsay Cedex, France. (*cyril.furtlehner@inria.fr*)

J.-M. Lasgouttes is with Project Team IMARA at INRIA Paris-Rocquencourt, Domaine de Voluceau - BP 105, 78153 Rocquencourt Cedex, France. (*jean-marc.lasgouttes@inria.fr*)

A. de La Fortelle is with INRIA Paris-Rocquencourt and École des Mines de Paris, CAOR Research Centre, 60 boulevard Saint-Michel, 75272 Paris Cedex 06, France. (*arnaud.de_la_fortelle@ensmp.fr*)
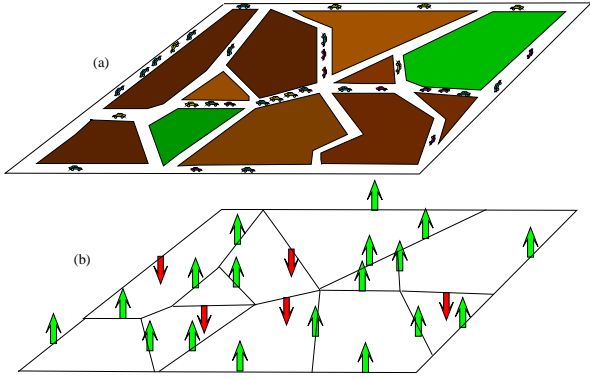
Fig. 1. Traffic network (a) and Ising model (b) on a random graph. Up (resp. down) arrows correspond to fluid (resp. congested) traffic.

In order to take into account both spatial and temporal relationships, the graph on which our model is defined is made of space-time vertices that encode both a location (road link) and a time (discretized on a few minutes scale). More precisely, the set of vertices is $\mathcal{V} = \mathcal{L} \otimes \mathbb{Z}^+$, where $\mathcal{L}$ corresponds to the links of the network and $\mathbb{Z}^+$ to the time discretization. To each point $\alpha = (\ell, t) \in \mathcal{V}$, we attach an information $\tau_\alpha \in \{0, 1\}$ indicating the state of the traffic (1 if congested, 0 otherwise). On such a model, the problems of prediction and reconstruction are equivalent, since they both amount to estimating the value of a subset of the nodes of the graph. The difference, which is obvious for the practitioner, lies mainly in the nature (space or time) of the correlations which are most exploited to perform the two tasks.

Each vertex is correlated to its neighbors (in time and space) and the evaluation of this local correlation determines the model. In other words, we assume that the joint probability distribution of $\tau_\mathcal{V} \stackrel{\text{def}}{=} \{\tau_\alpha, \ \alpha \in \mathcal{V}\} \in \{0, 1\}^\mathcal{V}$ is of the form

$$p(\{\tau_\alpha, \alpha \in \mathcal{V}\}) = \prod_{\alpha \in \mathcal{V}} \phi_\alpha(\tau_\alpha) \prod_{(\alpha, \beta) \in \mathcal{E}} \psi_{\alpha\beta}(\tau_\alpha, \tau_\beta) \quad (1)$$

where $\mathcal{E} \subset \mathcal{V}^2$ is the set of edges, and the local correlations are encoded in the functions ψ and ϕ. $\mathcal{V}$ together with $\mathcal{E}$ describe the space-time graph $\mathcal{G}$ and $\mathcal{V}(\alpha) \subset \mathcal{V}$ denotes the set of neighbors of vertex α.

The model described by (1) is actually equivalent to an Ising model on $\mathcal{G}$, with arbitrary coupling between adjacent spins $s_\alpha = 2\tau_\alpha - 1 \in \{-1, 1\}$, the up or down orientation of each spin indicating the status of the corresponding link (Fig. 1). The homogeneous Ising model (uniform coupling constants) is a well-studied model of ferro (positive coupling) or anti-ferro (negative coupling) material in statistical physics. It displays a phase transition phenomenon with respect to the value of the coupling. At weak coupling, only one disordered state occurs, where spins are randomly distributed around a mean-zero value. Conversely, when the coupling is strong, there are two equally probable states that correspond to the onset of a macroscopic magnetization either in the up or down direction: each spin has a larger probability to be oriented in the privileged direction than in the opposite one. From the point of view of a traffic network, this means that such a model is able to describe three possible traffic regimes: fluid (most of the spins up), congested (most of the spins down) and dense (roughly half of the links are congested). For real situations, we expect other types of congestion patterns, and we seek to associate them to the possible states of an inhomogeneous Ising model with possibly negative coupling parameters, referred to as spin glasses in statistical physics [13].

The practical information of interest which one wishes to extract from (1) is in the form of local marginal distributions $p_\alpha(\tau_\alpha)$, once a certain number of variables have been fixed by probe vehicles observations. They give the probability for a given node to be saturated at a given time and in turn can be the basis of a travel time estimation. From a computational viewpoint, the extraction cost of such an information from an Ising model on a multiply connected graph is known to scale exponentially with the size of the graph, so one has to resort to some approximate procedure. As we explain now such an approximation exists for dilute graphs (graphs with a tree-like local structure).

On a simply connected graph, the knowledge of $p_\alpha(\tau_\alpha)$ the one-vertex and $p_{\alpha\beta}(\tau_\alpha, \tau_\beta)$ the two-vertices marginal probabilities is sufficient [12] to describe the measure (1).

$$\begin{aligned} p(\tau_\mathcal{V}) &= \frac{\prod_{(\alpha,\beta)\in\mathcal{E}} p_{\alpha\beta}(\tau_\alpha, \tau_\beta)}{\prod_{\alpha\in\mathcal{V}} p_\alpha(\tau_\alpha)^{q_\alpha-1}} \\ &= \prod_{\alpha\in\mathcal{V}} p_\alpha(\tau_\alpha) \prod_{(\alpha,\beta)\in\mathcal{E}} \frac{p_{\alpha\beta}(\tau_\alpha, \tau_\beta)}{p_\alpha(\tau_\alpha)p_\beta(\tau_\beta)}, \end{aligned} \quad (2)$$

where q_α denotes the number of neighbors of α. Since our space time graph $\mathcal{G}$ is multi-connected, this relationship between local marginals and the full joint probability measure can only be an approximation, which in the context of statistical physics is referred to as the Bethe approximation. This approximation is provided by the minimum of the so-called Bethe free energy, which, based on the form (2), is an approximate form of the Kullback-Leibler distance,

$$D(b\|p) \stackrel{\text{def}}{=} \sum_{\tau_\mathcal{V}} b(\tau_\mathcal{V}) \ln \frac{b(\tau_\mathcal{V})}{p(\tau_\mathcal{V})},$$

between the reference measure p and an approximate one b. This rewrites in terms of a free energy as

$$D(b\|p) = \mathcal{F}(b) - \mathcal{F}(p),$$

where

$$\mathcal{F}(b) \stackrel{\text{def}}{=} \mathcal{U}(b) - \mathcal{S}(b), \quad (3)$$

with the respective definitions of the energy $\mathcal{U}$ and of the entropy $\mathcal{S}$

$$\begin{aligned} \mathcal{U}(b) &\stackrel{\text{def}}{=} - \sum_{(\alpha,\beta)\in\mathcal{E}} \sum_{\tau_\alpha,\tau_\beta} b_{\alpha\beta}(\tau_\alpha, \tau_\beta) \log \psi_{\alpha\beta}(\tau_\alpha, \tau_\beta) \\ &\quad - \sum_{\alpha\in\mathcal{V}} \sum_{\tau_\alpha} b_\alpha(\tau_\alpha) \log \phi_\alpha(\tau_\alpha), \\ \mathcal{S}(b) &\stackrel{\text{def}}{=} - \sum_{(\alpha,\beta)\in\mathcal{E}} \sum_{\tau_\alpha,\tau_\beta} b_{\alpha\beta}(\tau_\alpha, \tau_\beta) \log b_{\alpha\beta}(\tau_\alpha, \tau_\beta) \end{aligned}$$

$$+ \sum_{\alpha \in \mathcal{V}} \sum_{\tau_\alpha} (q_\alpha - 1) b_\alpha(\tau_\alpha) \log b_\alpha(\tau_\alpha).$$

The set of single vertex b_α and and two-vertices $b_{\alpha\beta}$ marginal probabilities that minimize (3) form the Bethe approximation of the Ising model. For reasons that will become evident in Section III-B, these will also be called *beliefs*. It is known that the quality of the approximation may deteriorate in the presence of short loops. In our case, the fact that the nodes are replicated along the time axis alleviates this problem.

In practice, what we retain from an inhomogeneous Ising description is the possibility to encode a certain number of traffic patterns in a statistical physics model. This property is also shared by its Bethe approximation (BA) and it is actually easier to encode the traffic patterns in this simplified model rather than the original one. Indeed, it will be shown in Section III-C that the computation of the BA from the marginal probabilities is immediate.

The data collected from the probe vehicles is used in two different ways. The most evident one is that the data of the current day directly influences the prediction. In parallel, this data is collected over long periods (weeks or months) in order to estimate the model (1). Typical historical data that is accumulated is

- $\hat{p}_\alpha(\tau_\alpha)$: the probability that vertex α is congested ($\tau_\alpha = 1$) or not ($\tau_\alpha = 0$);
- $\hat{p}_{\alpha\beta}(\tau_\alpha, \tau_\beta)$: the probability that a probe vehicle going from α to $\beta \in \mathcal{V}(\alpha)$ finds α with state τ_α and β with state τ_β.

The computation of $\hat{p}_\alpha$ and $\hat{p}_{\alpha\beta}$ requires a proper congestion state indicator τ_α that we assume to be the result of the practitioner's pretreatment of the FCD. The definition of this indicator is a problem of its own and is outside of the scope of this article. A relevant FCD variable is instantaneous speed. An empirical threshold may be attached to each link in order to discriminate (in a binary or in a continuous manner) between a fluid and a congested situation. Another approach is to convert the instantaneous speed in a probability distribution of the local car density, when an empirical fundamental diagram is known for a given link. Aggregation over a long period of these estimators yields then the desired historical data. The edges (α, β) of the space time graph $\mathcal{G}$ are constructed based on the presence of a measured mutual information between α and β, which is the case when $\hat{p}_{\alpha\beta}(\tau_\alpha, \tau_\beta) \neq \hat{p}_\alpha(\tau_\alpha)\hat{p}_\beta(\tau_\beta)$.

III. THE RECONSTRUCTION AND PREDICTION ALGORITHM

A. Statement of the inference problem

We turn now to our present work concerning an inference problem, which we set in general terms as follows: a set of observables $\tau_\mathcal{V} = \{\tau_\alpha, \ \alpha \in \mathcal{V}\}$, which are stochastic variables are attached to the set $\mathcal{V}$ of vertices of a graph. For each edge $(\alpha, \beta) \in \mathcal{E}$ of the graph, an accumulation of repetitive observations allows to build the empirical marginal probabilities $\{\hat{p}_{\alpha\beta}\}$. The question is then: given the values

of a subset $\tau_{\mathcal{V}^*} = \{\tau_\alpha, \ \alpha \in \mathcal{V}^*\}$, what prediction can be made concerning $\overline{\mathcal{V}^*}$, the complementary set of $\mathcal{V}^*$ in $\mathcal{V}$?

There are two main issues:

- how to encode the historical observations (inverse problem) in an Ising model, such that its marginal probabilities on the edges coincide with the $\hat{p}_{\alpha\beta}$?
- how to decode in the most efficient manner—typically in real time—this information, in terms of conditional probabilities $P(\tau_\alpha | \tau_{\mathcal{V}^*})$?

The answer to the second question will somehow give a hint to the first one.

B. The belief propagation algorithm

BP is a message passing procedure [12], which output is a set of estimated marginal probabilities (the beliefs $b_{\alpha\beta}$ and b_α) for the measure (1). The name "belief" reflects the artificial intelligence roots of the algorithm. The idea is to factor the marginal probability at a given site in a product of contributions coming from neighboring sites, which are the messages. The messages sent by a vertex α to $\beta \in \mathcal{V}(\alpha)$ depends on the messages it received previously from other vertices:

$$m_{\alpha \to \beta}(\tau_\beta) \leftarrow \sum_{\tau_\alpha \in \{0,1\}} n_{\alpha \to \beta}(\tau_\alpha) \phi_\alpha(\tau_\alpha) \psi_{\alpha\beta}(\tau_\alpha, \tau_\beta), \quad (4)$$

where

$$n_{\alpha \to \beta}(\tau_\alpha) \stackrel{\text{def}}{=} \prod_{\gamma \in \mathcal{V}(\alpha) \setminus \{\beta\}} m_{\gamma \to \alpha}(\tau_\alpha). \quad (5)$$

The messages are iteratvely propagated into the network with a parallel, sequential or random policy. If they converge to a fixed point, the beliefs b_α are then reconstructed according to

$$b_\alpha(\tau_\alpha) \propto \phi_\alpha(\tau_\alpha) \prod_{\beta \in \mathcal{V}(\alpha)} m_{\beta \to \alpha}(\tau_\alpha), \quad (6)$$

and, similarly, the belief $b_{\alpha\beta}$ of the joint probability of $(\tau_\alpha, \tau_\beta)$ is given by

$$\begin{aligned} b_{\alpha\beta}(\tau_\alpha, \tau_\beta) &\propto n_{\alpha \to \beta}(\tau_\alpha) n_{\beta \to \alpha}(\tau_\beta) \\ &\times \phi_\alpha(\tau_\alpha) \phi_\beta(\tau_\beta) \psi_{\alpha\beta}(\tau_\alpha, \tau_\beta). \end{aligned} \quad (7)$$

In the formulas above and in the remainder of this paper, the symbol $\propto$ indicates that one must normalize the beliefs so that they sum to 1. A simple computation shows that equations (6) and (7) are compatible, since (4)–(5) imply that

$$\sum_{\tau_\alpha \in \{0,1\}} b_{\alpha\beta}(\tau_\alpha, \tau_\beta) = b_\beta(\tau_\beta).$$

In most studies, it is assumed that the messages are normalized so that

$$\sum_{\tau_\beta \in \{0,1\}} m_{\alpha \to \beta}(\tau_\beta) = 1.$$

holds. The update rule (4) indeed indicates that there is an important risk to see the messages converge to 0 or diverge to infinity. It is however not immediate to check that the normalized version of the algorithm has the same fixed points

as the original one (and therefore the Bethe approximation). This point has been analyzed in [15] and the conclusion is that the fixed points of both version of the algorithm coincide, except possibly when the graph has a unique cycle. We can safely expect that it is not the case in practical situations.

It has been realized a few years ago [16] that the fixed points of the BP algorithm coincide with stable points of the Bethe free energy (3), and that moreover stable fixed points correspond to local minima of (3) [17]. BP is therefore a simple and efficient way to compute the Bethe approximation of our inhomogeneous Ising model. We propose to use the BP algorithm for two purposes: estimation of the model parameters (the functions $\psi_{\alpha\beta}$ and ϕ_α) from historical data and reconstruction of traffic from current data.

C. Setting the model with belief propagation

The fixed points of the BP algorithm (and therefore the Bethe approximation) allow to approximate the joint marginal probability $p_{\alpha\beta}$ when the functions $\psi_{\alpha\beta}$ and ϕ_α are known. Conversely, it can provide good candidates for $\psi_{\alpha\beta}$ and ϕ_α from the historical values $\hat{p}_{\alpha\beta}$ and $\hat{p}_\alpha$.

To set up our model, we are looking for a fixed point of the BP algorithm satisfying (4)–(5) and such that $b_{\alpha\beta}(\tau_\alpha, \tau_\beta) = \hat{p}_{\alpha\beta}(\tau_\alpha, \tau_\beta)$ and therefore $b_\alpha(\tau_\alpha) = \hat{p}_\alpha(\tau_\alpha)$.

It is easy to check that the following choice of ϕ and ψ,

$$\psi_{\alpha\beta}(\tau_\alpha, \tau_\beta) \;=\; \frac{\hat{p}_{\alpha\beta}(\tau_\alpha, \tau_\beta)}{\hat{p}_\alpha(\tau_\alpha)\hat{p}_\beta(\tau_\beta)}, \tag{8}$$

$$\phi_\alpha(\tau_\alpha) \;=\; \hat{p}_\alpha(\tau_\alpha), \tag{9}$$

leads (1) to coincide with (2). They correspond to a normalized BP fixed point for which all messages are equal to $1/2$. It has been shown in [15] that this form of ϕ and ψ is in some sense canonical: any other set of functions yielding the same beliefs is equal to (8)–(9), up to a change of variable. This equivalence holds for the beliefs at the other fixed points of the algorithm and their stability properties.

This crucial result means that it is not needed to *learn* the parameters of the Ising model, but that they can be readily recovered from the Bethe approximation. The message update scheme (4) of previous section can therefore be recast as

$$m_{\alpha\to\beta}(\tau_\beta) \leftarrow \sum_{\tau_\alpha \in \{0,1\}} n_{\alpha\to\beta}(\tau_\alpha)\hat{p}_{\alpha\beta}(\tau_\alpha | \tau_\beta) \tag{10}$$

and the beliefs are now expressed as

$$b_\alpha(\tau_\alpha) \propto \hat{p}_\alpha(\tau_\alpha) \prod_{\gamma \in \mathcal{V}(\alpha)} m_{\gamma\to\alpha}(\tau_\alpha), \tag{11}$$

$$b_{\alpha\beta}(\tau_\alpha, \tau_\beta) \propto \hat{p}_{\alpha\beta}(\tau_\alpha, \tau_\beta) n_{\alpha\to\beta}(\tau_\alpha) n_{\beta\to\alpha}(\tau_\beta). \tag{12}$$

There is no guarantee that the trivial constant fixed point is stable. However, the following theorem, proved in [15], shows that this can be decided from the mere knowledge of the marginal which we want to model.

Theorem 1: The fixed point $\{\hat{p}\}$ is stable if, and only if, the matrix defined, for any pair of oriented edges $(\alpha, \beta) \in \mathcal{E}$,

$(\alpha', \beta') \in \mathcal{E}$, by the elements

$$J_{\alpha\beta}^{\alpha'\beta'} = \big(\hat{p}_{\alpha\beta}(1|1) - \hat{p}_{\alpha\beta}(1|0)\big) \mathbb{1}_{\{\alpha' \in \mathcal{V}(\alpha)\setminus\{\beta\},\ \beta'=\alpha\}},$$

has a spectral radius (largest eigenvalue in norm) smaller than 1.

A sufficient condition for this stability is therefore

$$\big|\hat{p}_{\alpha\beta}(1|1) - \hat{p}_{\alpha\beta}(1|0)\big| < \frac{1}{q_\alpha - 1}, \quad \text{for all } \alpha \in \mathcal{V}, \beta \in \mathcal{V}(\alpha).$$

In addition, on a dilute graph, the knowledge of the Jacobian coefficient distribution and the connectivity distribution of the graph is enough to determine the stability property by a mean field argument [15].

D. Traffic reconstruction and prediction

Let $\mathcal{V}^*$ be the set of vertices that have been visited by probe vehicles. Reconstructing traffic from the data gathered by those vehicles is equivalent to evaluating the conditional probability

$$p_\alpha(\tau_\alpha | \tau_{\mathcal{V}^*}) = \frac{p_{\alpha,\mathcal{V}^*}(\tau_\alpha, \tau_{\mathcal{V}^*})}{p_{\mathcal{V}^*}(\tau_{\mathcal{V}^*})}, \tag{13}$$

where $\tau_{\mathcal{V}^*}$ is a shorthand notation for the set $\{\tau_\alpha\}_{\alpha \in \mathcal{V}^*}$.

The BP algorithm applies to this case if a specific rule is defined for vertices $\alpha \in \mathcal{V}^*$: since the value of τ_α is known, there is no need to sum over possible values and (10) becomes

$$m_{\alpha\to\beta}(\tau_\beta) \leftarrow n_{\alpha\to\beta}(\tau_\alpha)\hat{p}_{\alpha\beta}(\tau_\alpha | \tau_\beta). \tag{14}$$

IV. PRACTICAL CONSIDERATIONS AND SIMULATION

The algorithm outlined in Section III can be summarized by the flowchart of Fig. 2. It is supposed to be run in real time, over a graph which corresponds to a time window (typically a few hours) centered around present time, with probe vehicle data added as it is available. In this perspective, the reconstruction and prediction operations are done simultaneously on an equal footing, the distinction being simply the time-stamp (past for reconstruction or future for prediction) of a given computed belief. The output of the previous run can be used as initial messages for a new run, in order to speedup convergence. Full re-initialization (typically a random set of initial messages) has to be performed within a time interval of the order but smaller than the time-scale of typical traffic fluctuations. We have tested the algorithm on the artificial traffic network shown on the program's screenshot of Fig. 3. To this end, we used a simulated traffic system which has the advantage to yield exact empirical data correlations. For real data, problems may arise because of noise in the historical information used to build the model; this additional difficulty will be treated in a separate work. The simulator implements a queueing network, where each queue represents a link of the traffic network (a single-way lane) and has a finite capacity. To each link, we attach a variable $\rho \in [0, 1]$, the car density, which is represented by a color code on the user interface snapshot.

As already stated in Section II, the physical traffic network is replicated, to form a space time graph, in which each
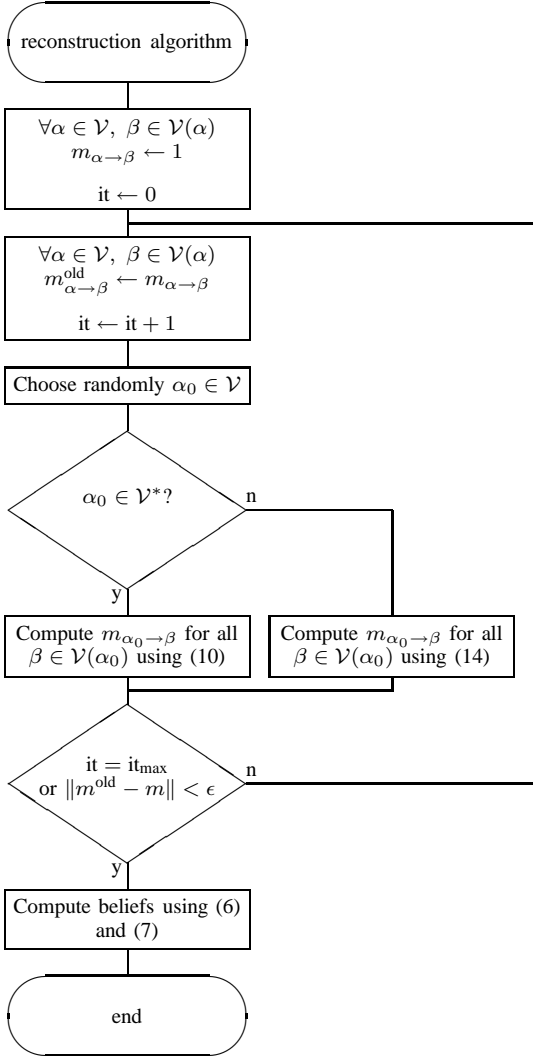
Fig. 2. The traffic reconstruction algorithm. The parameters, besides those that have already been defined elsewhere, are the total number of iterations it_{max}, the maximal error $\epsilon > 0$ and a norm $\|\cdot\|$ on the set of messages, which choice is up to the implementer. Likewise, the update policy described here is "random", but parallel or sequential updates can also be used.

#floating cars: 10
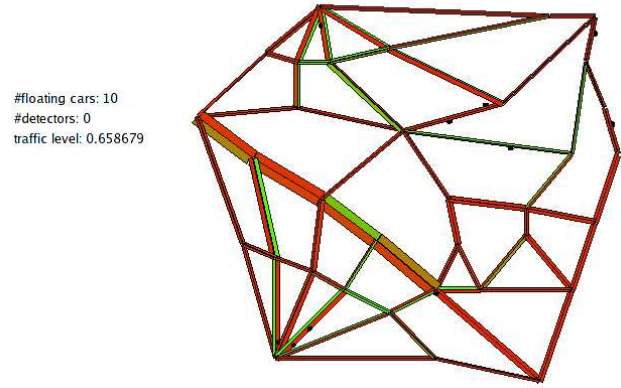#detectors: 0
traffic level: 0.658679

Fig. 3. Traffic network as produced by the simulator. The continuous color code represents the traffic index from 0 (green/light) to 1 (red/dark). There are 35 physical nodes and 122 physical links (road segments), simulated on 40 time steps, which yields a time-space graph $\mathcal{G}$ with 4880 nodes.
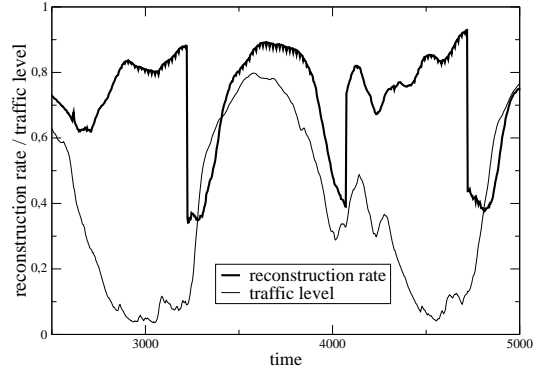

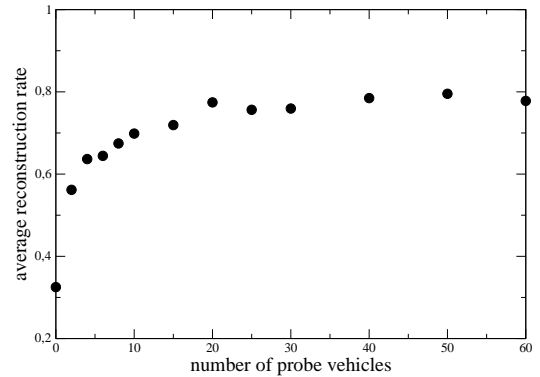Fig. 4. traffic level and reconstruction rate as a function of time (10 probe vehicles).

vertex $\alpha = (\ell, t)$ corresponds to a link ℓ at a given time t of the traffic graph. To any space-time vertex α, we associate a binary congestion variable $\tau_\alpha \in \{0, 1\}$. The statistical physics description amounts to relating the probability of saturation $P(\tau_\alpha = 1)$ to the density ρ_α. For the sake of simplicity, we consider a linear relation and build our historical $\hat{p}$ according to some time averaging procedure. In practice, the car density would not be available from the FCD and a preprocessing of information would be necessary. In our oversimplified setting, the single-vertex beliefs yield directly an estimation of the car density. Nevertheless, more realistic data collection and modeling would be completely transparent w.r.t. the algorithm.

To estimate the quality of the traffic restoration we use the

Fig. 5. average reconstruction rate as a function of the number of probe vehicles.

following estimator:

$$\text{reconstruction rate} \overset{\text{def}}{=} \frac{1}{|\mathcal{V}|} \sum_{\alpha \in \mathcal{V}} \mathbb{1}_{\{|b_\alpha - \rho_\alpha| < 0.2\}},$$

which computes the fraction of space-time nodes α for which the belief b_α does not differ by more than an arbitrary threshold of 0.2 from ρ_α.

A typical prediction time series is shown in Fig. 4. The overall traffic level, characterized by some real number comprised between 0 and 1, oscillates between dense and fluid conditions with a certain amount of noise superimposed. In this setting, we observe first that BP has three fixed points, among which the reference $b = \hat{p}$ (see Section III-C), which is is in fact unstable because it is a superposition of distinct measures. The two additional fixed points represent actually the dense and fluid traffic conditions. These additional states appear spontaneously and some fine tuning is required to control saturation effects [15]. This is reflected in the sudden drops of the reconstruction rate when the algorithm jumps from one state to the other during transient phases. A selection criteria based on free energy measurements may be used to choose the most relevant fixed point [15].

Concerning the dependence of the reconstruction rate with respect to the number of probe vehicles, Fig. 5 indicates that the knowledge on less than 10% of the links (10 vehicles for 122 links) is sufficient in this setting to identify most of the time correctly the traffic regime. However, when the number of probes is increased, the reconstruction rate given by our algorithm saturates around 80%. In addition to the fact that time correlations are not incorporated in our model, the main reason for this saturation is that in our practical implementation [15] correlations between probe vehicles are neglected when imposing (13).

V. Conclusion and perspectives

We have presented a novel methodology for reconstruction and prediction of traffic using the belief propagation algorithm on floating car data. We have shown how the underlying Ising model can be determined in a straightforward manner and that it is unique up to some change of variables. The algorithm has been implemented and illustrated using an artificial traffic model. While our main focus is currently on testing on more realistic simulations, several generalizations are considered for future work, using the extension of the results of Section III to a more general factor graph setting done in [18]:

Firstly, the binary description corresponding to the underlying Ising model is arbitrary. Traffic patterns could be represented in terms of s different inference states. A Potts model with s-states variables would leave the belief propagation algorithm and its stability properties structurally unchanged. However, since the number of correlations to evaluate for each link is $s^2 - 1$, this number of states should be subject to an optimization procedure.

Secondly, our way of encoding traffic network information might need to be augmented to cope with real world situations. This would simply amount to use a factor-graph used to propagate this information. In particular it is likely that a great deal of information is contained in the correlations of local congestion with aggregate traffic indexes, corresponding to sub-regions of the traffic network. Taking these correlations into account would result in the introduction of specific variables and function nodes associated to these aggregate traffic indexes. These aggregate variables would naturally lead to a hierarchical representation of the factor graph, which is necessary for inferring the traffic on large scale network. Additionally, time dependent correlations which are needed for the description of traffic, which by essence is an out of equilibrium phenomenon, could be conveniently encoded in these traffic index variables.

Ultimately, for the elaboration of a powerful prediction system, the structure of the information content of a traffic-road network has to be elucidated through a specific statistical analysis. The use of probe vehicles, based on modern communications devices, combined with a belief propagation approach, is in this respect a very promising approach.

References

[1] D. Chowdhury, L. Santen, and A. Schadschneider, "Statistical physics of vehicular traffic and some related systems," *Physics Report*, vol. 329, p. 199, 2000.

[2] A. Klar, R. Kuehne, and R. Wegener, "Mathematical models for vehicular traffic," *Surv. Math. Ind.*, vol. 6, p. 215, 1996.

[3] C. Daganzo, "Requiem for second order fluid approximation of traffic flow," *Transportation Research*, vol. B, 1995.

[4] A. Aw and M. Rascle, "Resurrection of "second order" models of traffic flow," *SIAM Journal on Applied Mathematics*, vol. 60, no. 3, pp. 916–938, 2000.

[5] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *J. Phys. I,2*, pp. 2221–2229, 1992.

[6] R. Chrobok, J. Wahle, and M. Schreckenberg, "Traffic forecast using simulations of large scale networks," in *Proceedings*, ser. Intell. Transport. Sys. Conf. Oakland(CA), USA: IEEE, August 25-29 2001.

[7] T. Guozhen, Y. Wenjiang, and H. Ding, "Traffic flow prediction based on generalized neural network," in *Proceedings*, ser. Intell. Transport. Sys. Conf. Washington, D.C. USA: IEEE, October 3-6 2004.

[8] W. Chun-Hsin, H. Jan-Ming, and D. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transport. Syst.*, vol. 5, no. 4, p. 276, 2004.

[9] H. Kanoh *et al.*, "Short-term traffic prediction using fuzzy c-means and cellular automata in a wide-area road network," in *Proceedings of the 8th International*, ser. Conf. Intell. Transport. Sys. Vienna, Austria: IEEE, September 13-16 2005.

[10] T. Benz *et al.*, "Information supply for intelligent routing services – the INVENT traffic network equalizer approach," in *Proceedings of the ITS World Congress*, 2003.

[11] H. H. Versteegt and C. M. J. Tampère, "PredicTime - state of the art and functional architecture," TNO Inro, Tech. Rep. 2003-07, 2003.

[12] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*. Morgan Kaufmann, 1988.

[13] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987.

[14] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inform. Theory.*, vol. 51, no. 7, pp. 2282–2312, 2005.

[15] C. Furtlehner, J.-M. Lasgouttes, and A. de La Fortelle, "Belief propagation and Bethe approximation for traffic prediction," INRIA, Rapport de Recherche 6144, Mar. 2007.

[16] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," *Advances in Neural Information Processing Systems*, pp. 689–695, 2001.

[17] T. Heskes, "Stable fixed points of loopy belief propagation are minima of the Bethe free energy," *Advances in Neural Information Processing Systems*, vol. 15, 2003.

[18] C. Furtlehner and J.-M. Lasgouttes, "Belief propagation inference with a prescribed fixed point," 2007, submitted to NIPS'07.