INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Belief-Propagation Algorithm for a Traffic Prediction System based on Probe Vehicles

Cyril Furtlehner — Arnaud de La Fortelle — Jean-Marc Lasgouttes

N° 5807

Janvier 2006

Thème BIO

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--5807--FR+ENG

Belief-Propagation Algorithm for a Traffic Prediction System based on Probe Vehicles

Cyril Furtlehner* †, Arnaud de La Fortelle*, Jean-Marc Lasgouttes*

Thème BIO — Systèmes biologiques
Projet Preval

Rapport de recherche n° 5807 — Janvier 2006 — 23 pages

Abstract: A traffic reconstruction and prediction algorithm based on probe vehicles is discussed in the present paper. Traffic information is provided by a set of probe vehicles circulating randomly on the network, in the form of average traffic intensity and correlations local in time and spatial position. The road network and the traffic are modeled as a queueing system on a planar graph with local interactions. Using statistical physics methods, a reconstruction algorithm is built and evaluated on a traffic toy model, where the queues have a finite capacity and specific state-dependent transitions rates are used to mimic typical situations of traffic-jams. The reconstruction algorithm consists of a message-passing procedure between sites corresponding to roads-segments at given day-time, which propagates both backward and forward in time, to reconstruct the past traffic and to make predictions.

This work is partly funded by the European project REACT and is being implemented on a server for real tests with a fleet of probe cars.

Key-words: traffic reconstruction, prediction, transport, probe vehicles, Ising model

* INRIA Rocquencourt – Domaine de Voluceau B.P. 105 – 78 153 Le Chesnay Cedex (France)
† École des Mines de Paris – CAOR research centre – 60, boulevard Saint-Michel – 75272 Paris cedex 06 (France)

Un algorithme de propagation de croyances pour un système de prédiction de traffic utilisant des véhicules traceurs

Résumé : Un algorithme de reconstruction et de prédiction de trafic par des véhicules traceurs est présenté dans ce rapport. Dans notre approche, les informations concernant l'intensité du trafic en temps réel sont collectées par des vehicules traceurs qui circulent aléatoirement sur le réseaux. Ces informations permettent d'obtenir des données moyennes et des corrélations entre liens proches à dates consécutives. Le réseau routier et le trafic sont modélisés par des files d'attente sur un graphe planaire, couplées par des interactions locales. Grâce à des méthodes issues de la physique statistique, un algorithme de reconstruction et de prédiction est proposé, puis évalué par des simulations sur un modèle dynamique simplifié de trafic, où les files ont une capacité finie et des taux de transition dépendants de l'état sont choisis de façon à reproduire des situations typiques d'encombrement. L'algorithme de reconstruction, à partir des données temps réel et des données historiques, consiste en une procédure de passage de messages entre sites d'un nouveau graphe. Les sites de ce graphe sont les liens du réseau de base, pris à des instants discrets et connectés entre eux en fonctions des corrélations obtenues aux niveau des carrefours. Ces messages permettent de propager positivement et négativement dans le temps l'information et ainsi de reconstruire le trafic passé et de donner des prédictions sur le trafic à venir.

Ce travail est en partie financé par le projet européen REACT. Les algorithmes sont en cours d'implémentation sur un serveur pour des expérimentations sur une flotte de véhicules traceurs.

Mots-clés : reconstruction de trafic, prédiction, transport, véhicules traceurs, modèle d'Ising

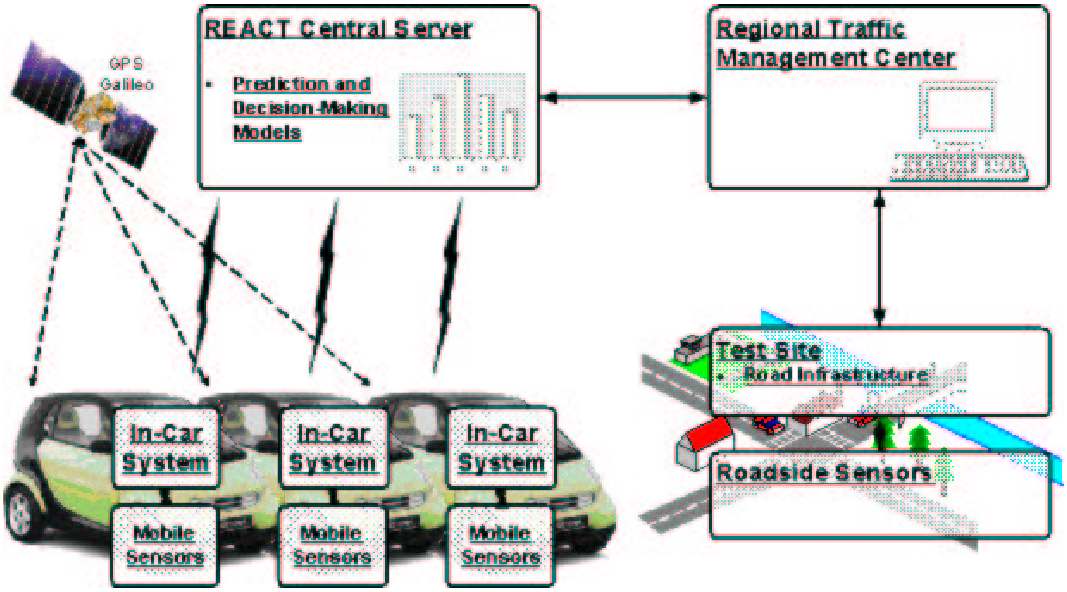

Fig. 1.1: REACT system

1 Introduction

1.1 Probe vehicles and traffic reconstruction

The developments detailed in this report are driven by the application for which they are needed: reconstruct and predict the traffic on roads where information is in general not available. Indeed, traffic data is currently available only on equipped motorways, i.e. on a very small proportion of the road network. The cost of the static traffic sensors (e.g. magnetic loops) and the cost of variable messages panels make it impossible to be a global solution.

However, accidents (more than 40,000 deaths per year in the European Union) and congestion (over 1% of GDP) lead the European Commission to a major effort to reduce these. As a part of this effort, the European project REACT (Realizing Enhanced Safety and Efficiency in European Road Transport) aims at developing a system (see Figure 1.1) that will work towards the Community's objectives of reducing road transport deaths and increasing road infrastructure capacity by means of probe vehicles equipped with a traffic sensor (sensing traffic speed, density and flow).

In this REACT system, INRIA is in charge of developing a module for reconstruction and prediction of the traffic in rural roads, more precisely in road not equipped with static sensors. This report presents the model we use and the algorithms implemented in this module.

1.2 Modeling and results

The main problem of traffic reconstruction on non-equipped roads is not really the lack of fixed detectors, it is the deep difference in the models that can be used. Equipped roads are generally motorways and flows (density and velocity) are measured regularly along the road and on all entrances and exits. Then a traffic flow model *at equilibrium* is used and predictions are incrementally calculated (see [8, 6] and references therein for more models).

This assumption is not valid anymore for most of the non equipped roads, either rural roads or streets: the topology is much richer and the traffic on a segment between two intersection is really dynamic, e.g. due to traffic lights or simply because the segments are very short (with respect to the flow) and the number of vehicles per segment is low. Stochastic models are natural here, and the most popular ones consist in queueing networks. Since the technology of probe vehicles (both in-car and on the central server) is not mature enough to provide modules working in this context, INRIA proposed a model based on statistical physics to solve the problem of traffic reconstruction. Note that reconstruction and prediction are two facets of the same problem. In reconstruction one has measurements on only a subset of the network and the values on the other roads have to be deduced. In prediction one has values for the present and the past, and the values for the future are to be determined.

The REACT project considers an alternative system for non equipped roads. It adopts the point of view that cars inside the traffic should be able to provide a large amount of information, sufficient to reconstruct a global view of the traffic state in a given network. The purpose of the present paper is to propose an algorithm, which would be able to perform such a traffic reconstruction, using the information collected by so-called probe vehicles. Section 2 shows how Ising models from statistical physics can be used as models of traffic state. The algorithm, which is a message-passing procedure well adapted to inference problems on graphs, will be exposed in Section 3, where we show also how Ising models appear naturally. Since we need to be able to benchmark the algorithm, we describe in Section 4 a toy traffic model implemented using a simulator (Figure 1.2). It is used to perform the numerical experiments that are the subject of Section 5.

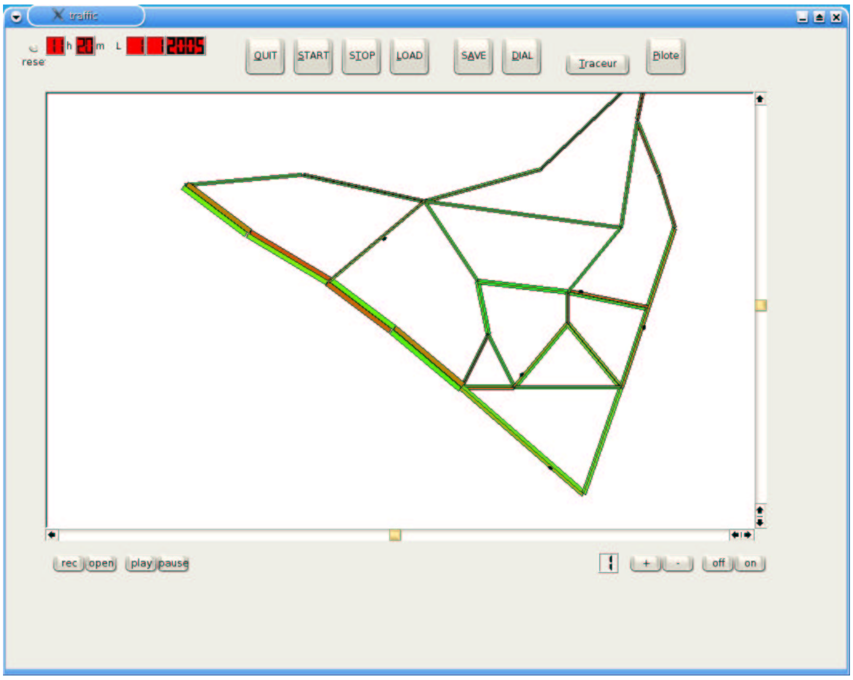

Fig. 1.2: The traffic simulator for performance tests, the road network has 21 nodes and 60 links, black dots are the probe vehicles, road-links are colored from green to red to indicate the local intensity of the traffic.

2 Traffic description and statistical physics

Continuous traffic flow models have flaws when it comes to the modeling of cities or rural roads traffic. Indeed, the velocity flow field is subject to much greater fluctuations induced by the nature of the network (presence of intersections and short distance between two intersections) than by the traffic itself. These fluctuations can be both spatial and temporal (a traffic light at a cross-road, a road-work, etc). There is no local stationary regime for the velocity and the dynamics is dominated by the fluctuations. This requires a stochastic modeling. The model we propose has two components: a queueing network and time discretization.

2.1 Queueing network

The first component is a standard queueing network, where a queue represents a link between two intersections. Instead of the number of customers in the queue (i.e. the number of vehicles in the given road-link), which is difficult to measure, the observable we are interested in is the escape-time, that is the time interval between the entrance and the escape of the link. This is the information needed for each link of the network (in practice since it is a fluctuating quantity we consider probability measure of escape-time) to reconstruct the expected travel-time between two arbitrary points of the network. This information is easily obtained by probe vehicles. Combined with the velocity information, it may serve to determine locally the intensity of the traffic, in particular whether it is saturated or not. However, this information will always be sparse and we need a way to infer the distribution of escape-times over the network at any time from the information collected by the probe vehicles. This is the purpose of the second component of our model.

2.2 Time discretization

In addition to the spatial discretization (set of links) we have a time discretization (typically a few minutes), which gives us a graph $\mathcal{G} = \mathcal{N} \otimes \mathbb{Z}^+$, where $\mathcal{N}$ corresponds to the network and $\mathbb{Z}^+$ to the time discretization. To each point $\alpha = (\ell, t) \in \mathcal{G}$, we attach an information $\tau_\alpha \in \{0, 1\}$ indicating the state of the traffic (1 if congested, 0 otherwise). Each cell is correlated to its neighbor (in time and space) and the evaluation of this local correlation determines the model. In other words, we assume that the joint probability distribution over the set is of the form

$$p(\{\tau_\alpha; \alpha \in \mathcal{G}\}) = \prod_{\alpha\in\mathcal{G}} \prod_{v\in V(\alpha)} \psi(\alpha; v)$$

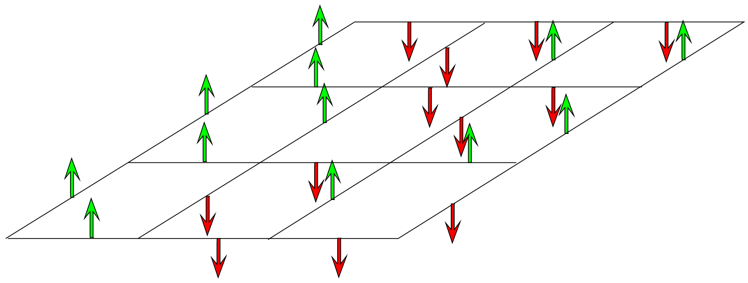

Fig. 2.1: Ising model above T_c

where $V(\alpha) \subset \mathcal{G}$ is the set of neighbors of α and the local correlation is encoded in the function ψ. In turn, message-passing algorithms (for example belief-propagation) can be used from the known information and the knowledge of ψ to extract information from the joint probability, to give for example typical samples of τ_α. The simplest model that takes into account this correlation is an Ising model [2] on $\mathcal{G}$. The partition function would read

$$Z(\beta) = \sum_{\{s_\alpha \in \{-1,1\}\}} \exp\Big[-\beta \sum_{\alpha \in \mathcal{G}} \sum_{v \in V(\alpha)} J_{\alpha;v} s_\alpha s_v + H_\alpha s_\alpha\Big] \tag{2.1}$$

where $s_\alpha \in \{-1, 1\}$ represents the traffic congestion ($s_\alpha = 2\tau_\alpha - 1$), H_α is a constraint field associated to historical data, $J_{\alpha;v}$ is the coupling between neighbor sites with respect to position and time and T is temperature which measures the randomnes of the system. The seemingly strange convention for s_α values comes from the fact that Ising models represent the positive or negative magnetization of particles.

The joint probability distribution then reads

$$p(\{s_\alpha; \alpha \in \mathcal{G}\}) = \frac{1}{Z(\beta)} \exp\Big[-\frac{1}{T} \sum_{\alpha \in \mathcal{G}} \sum_{v \in V(\alpha)} J_{\alpha;v} s_\alpha s_v + H_\alpha s_\alpha\Big] \tag{2.2}$$

The homogeneous Ising Model (coupling constants $J_{\alpha;v} = J$ and $H_\alpha = H$) is a well-studied model of ferro ($J > 0$) or anti-ferro ($J < 0$) material subject to a uniform magnetic-field H in statistical physics. It displays a phase transition when the space-dimension of the network is bigger than one, with respect to a critical temperature T_c. When $T > T_c$ only a disordered state occurs, spins are randomly distributed around a mean-zero value (Figure 2.1). Instead, when $T < T_c$, in absence of the external magnetic field, two states are equally probable and correspond to the

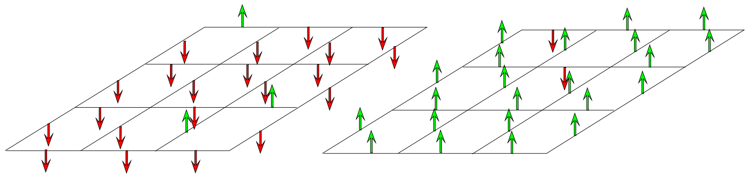

Fig. 2.2: Ising model below T_c with a broken symmetry

onset of a macroscopic magnetization either in the up or down direction, which means that each spin has a larger probability to be oriented in the privileged direction than in the opposite one (Figure 2.2).

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 a random-Ising model (i.e. with general parameters), which has also been well-studied in statistical physics, in the context of spin-glasses [4]. Indeed in some cases, when the system is frustrated, because some of the couplings are negative, leading to a certain number of contradictions, a proliferation of metastable states occurs, which eventually scales exponentially with the size of the system (Figure 2.3).

3 The reconstruction and prediction algorithm

3.1 Historical data

The purpose of collecting historical data is to estimate the parameters $J_{\alpha;v}$ and H_α of the model (2.1)–(2.2). Once these coefficients are known, the information provided by the probe vehicles will allow to freeze a certain number of the spins, and if this information is sufficient, this will drive the system to one of his possible states, which corresponds to a given congestion pattern.

The problem of reconstructing the traffic from a sparse information is an inference problem which can be simply stated as follows. First although the procedure could be implemented for real variables, such as traffic density or velocity flow, we aim at inferring over the network a binary type information, namely whether on a given link the traffic is congested or not. Let us call $\tau_\ell(t) \in \{0, 1\}$ this variable for the link ℓ

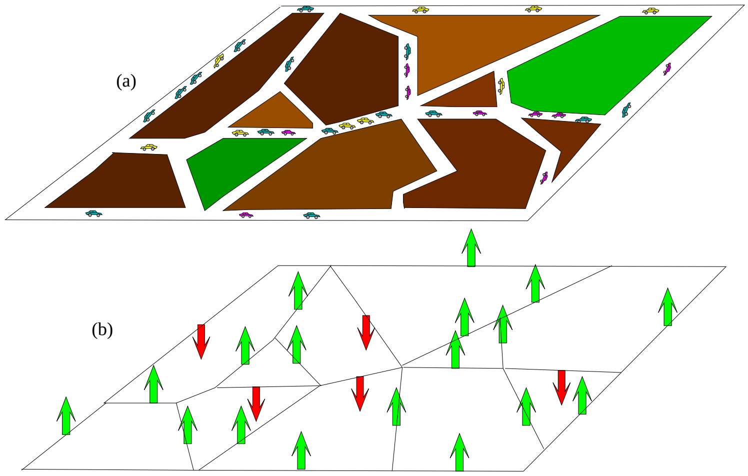

Fig. 2.3: Traffic network (a) and Ising model (b) on a random graph

and at the discrete time-step t. This information can be communicated by the probe vehicle either as

- a firm statement: the traffic *is* ($\tau_\ell(t) = 1$) or *is not*, ($\tau_\ell(t) = 0$) congested;
- a probability: given the current speed-velocity v, there is a certain probability $\hat{p}_{\ell,t}(v)$ that the traffic is congested.

In any case, for each link it is possible to accumulate such data in order to construct an historical value. Let $\{\tau_\ell^i(t), i = 1 \dots n_{\ell,t}\}$ a set of traffic informations collected by the probe vehicles, $n_{\ell,t}$ represents the number of times a probe vehicles was in the link ℓ at day-time t since the beginning of data collection. Then, the historical value

$$h_{\ell,t} \stackrel{\text{def}}{=} \frac{1}{n_{\ell,t}} \sum_{i=1}^{n_{\ell,t}} \tau_\ell^i(t),$$

which represents the probability of congestion of link ℓ at time-step t. If the historical information is given by a velocity set $\{v_\ell^i(t), i = 1 \dots n_{\ell,t}\}$, then instead we define

$$h_{\ell,t} \stackrel{\text{def}}{=} \frac{1}{n_{\ell,t}} \sum_{i=1}^{n_{\ell,t}} p_{\ell,t}\big(v_\ell^i(t)\big).$$

In addition, let us assume that two links ℓ and ℓ' are directly connected through some node. If a probe vehicle goes from ℓ to ℓ', it will be able to send an information concerning $\tau_\ell \tau_{\ell'}$. From this one can compute the correlation between $\alpha = (\ell, t)$ and $\beta = (\ell', t+1)$

$$\sigma^h_{\alpha\beta} \stackrel{\text{def}}{=} \frac{\varepsilon_{\alpha\beta} - h_\alpha h_\beta}{\sqrt{h_\alpha(1-h_\alpha)h_\beta(1-h_\beta)}}$$

according to the historical averaging procedure

$$\varepsilon_{\alpha\beta} \stackrel{\text{def}}{=} \frac{1}{n_\alpha \beta} \sum_{i=1}^{n_{\alpha\beta}} \tau^i_\alpha \tau^i_\beta,$$

where $\{(\tau^i_\alpha, \tau^i_\beta), i = 1 \ldots n_{\alpha\beta}\}$ is the set of information collected by the probe cars.

3.2 A Bayesian network

Assuming that the traffic flow evolution has a natural causal orientation, we consider that the state of the traffic in a given link of the network at time t is conditioned by the state of the network at time $t-1$. As a result, the probability that an outgoing link of a given cross-roads a is saturated at time t is conditioned to the states of incoming roads to a at time $t-1$.

Based on the data which may be realistically extracted from the traffic network by the probe vehicles, we assume that a one-day sample of traffic is distributed according to a probability function, which we detail now.

On the top of the traffic network, the correlations we measure lead to define the graph $\mathcal{G}$ from

- the nodes (ℓ, t) corresponding to a link ℓ of the traffic network with a time-stamp t;
- the correlation links, which simply join the nodes (ℓ, t) and $(\ell', t+1)$ if the corresponding links ℓ and ℓ' are connected by a non-vanishing $\sigma_{\ell\ell'}$. These are local correlations, since ℓ and ℓ' have a node in common.

The graph is oriented, and we say that $\alpha = (\ell, t)$ is a parent node of $\beta = (\ell', t')$—and β is a child of α—if they are linked by a finite correlation $\sigma_{\ell\ell'} \neq 0$ and if $t' = t - 1$ (see Figure 3.2). For a given node α we call $a(\alpha)$ the subset of nodes which are parents of α and $c(\alpha)$ the subset of nodes which are child of α. To each node α

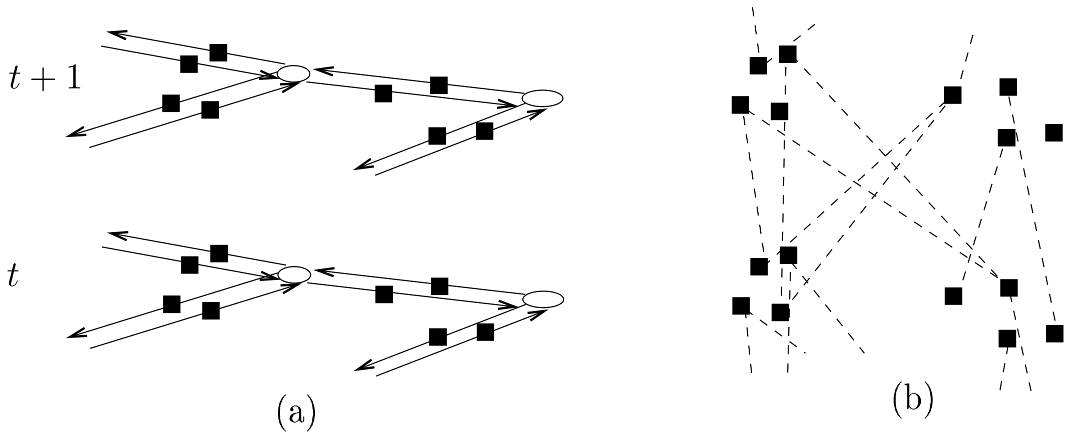

Fig. 3.1: (a) the road network duplicated for two time-stamps t and $t+1$ with the cross-roads represented by circles; (b) the corresponding correlation graph where nodes are time-stamped roads represented by black squares.

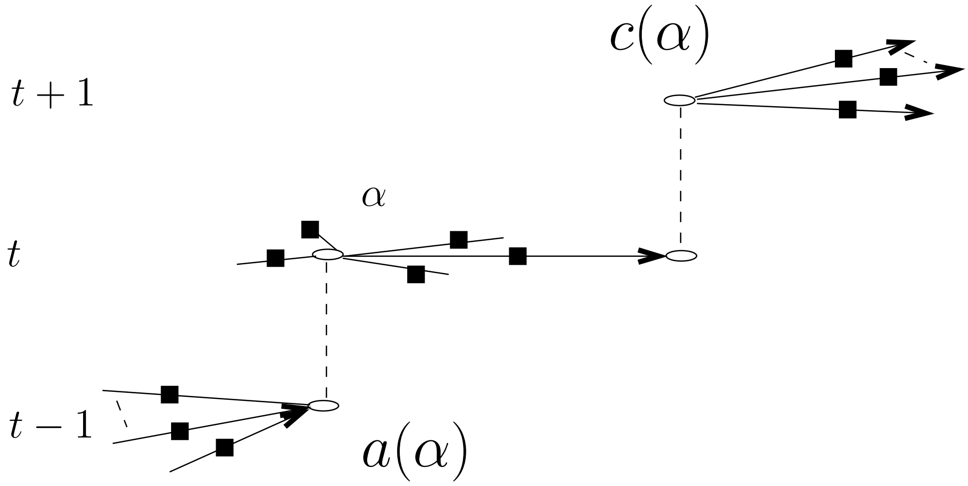

Fig. 3.2: A node α, its set of parents $a(\alpha)$ and its set of children $c(\alpha)$.

we attach the traffic variable $\tau_\alpha \in \{0,1\}$. The joint probability distribution of the unknown variables τ_α is assumed to be of the form

$$p_B(\{\tau_\alpha, \alpha \in \mathcal{G}\}) = \prod_{\alpha\in\mathcal{G}} p(\tau_\alpha|\{\tau_\beta, \beta \in a(\alpha)\})$$

Let $\mathcal{U}$ be the subset of nodes α^* for which the probe vehicles have provided an information, which in general will be the probability p_{α^*} of saturation ($\tau_{\alpha^*} = 1$). By construction, the joint probability measured for this subset has a product form

$$q(\{\tau_{\alpha^*}, \alpha^* \in \mathcal{U}\}) \stackrel{\text{def}}{=} \prod_{\alpha^*\in\mathcal{U}} \bigl(\tau_{\alpha^*} p_{\alpha^*} + \bar{\tau}_{\alpha^*}(1 - p_{\alpha^*})\bigr),$$

using the boolean notation $\bar{\tau} = 1 - \tau$. The purpose of the algorithm is to determine the complete set of marginal functions

$$p_\alpha(\tau_\alpha) = \sum_{\{\tau_\beta,\beta\neq\alpha\}} \frac{p(\tau_\alpha, \{\tau_\beta, \beta \in \mathcal{G}\})}{p(\{\tau_{\alpha^*}, \alpha^* \in \mathcal{U}\})} q(\{\tau_{\alpha^*}, \alpha^* \in \mathcal{U}\}).$$

3.3 A pairwise random Markov field

The Bayesian representation of the inference graph, although exact, appears difficult to handle with the message passing procedure which will be explained later. When it comes to make explicit the set of conditional probabilities $p(\tau_\alpha|\{\tau_\beta, \beta \in a(\alpha)\})$, it is tempting to approximate this as

$$p(\tau_\alpha|\{\tau_\beta, \beta \in a(\alpha)\}) = \frac{\prod_{\beta\in a(\alpha)} p(\tau_\alpha|\tau_\beta)}{Z_\alpha(\{\tau_\beta, \beta \in a(\alpha)\})},$$

with Z_α a normalizing constant

$$Z_\alpha(\{\tau_\beta, \beta \in a(\alpha)\}) \stackrel{\text{def}}{=} \sum_{\tau_\alpha\in\{0,1\}} \prod_{\beta\in a(\alpha)} p(\tau_\alpha|\tau_\beta),$$

ensuring that $p(\tau_\alpha|\{\tau_\beta, \beta \in a(\alpha)\})$ is a probability measure for τ_α. Because of this normalization constant, it will be necessary to pass messages among parents, having a child in common, in addition to the messages between parents and children. Such a complication is difficult to handle, and we call for an additional approximation by redefining the joint probability measure in terms of a pairwise random Markov field

$$p_{MF}(\{\tau_\alpha, \alpha \in \mathcal{G}\}) = \frac{1}{Z} \prod_{\alpha\in\mathcal{G}} \prod_{\beta\in a(\alpha)} \psi(\tau_\alpha, \tau_\beta).$$

where the pairwise interaction functions are simply given by

$$\psi(\tau_\alpha, \tau_\beta) = p(\tau_\alpha|\tau_\beta).$$

From the historical data we may have an heuristic estimation of the conditional probabilities $p(\alpha|\beta)$ between to nodes $\alpha = (\ell, t)$ and $\beta = (\ell', t+1)$ of the graph such that ℓ and ℓ', the corresponding links, are connected by a correlation link. In the limit where the historical data is exact, i.e. if the accumulation time is sufficiently long, the joint probability $p(\tau_\alpha, \tau_\beta)$ tends to

$$p(\tau_\alpha, \tau_\beta) = \varepsilon_{\alpha\beta}\tau_\alpha\tau_\beta + (h_\beta - \varepsilon_{\alpha\beta})\bar{\tau}_\alpha\tau_\beta + (h_\alpha - \varepsilon_{\alpha\beta})\tau_\alpha\bar{\tau}_\beta + (1 - h_\alpha - h_\beta + \varepsilon_{\alpha\beta})\bar{\tau}_\alpha\bar{\tau}_\beta.$$

The conditional probability therefore reads

$$\begin{aligned} p(\tau_\beta|\tau_\alpha) &= \frac{p(\tau_\alpha, \tau_\beta)}{p(\tau_\alpha, \tau_\beta) + p(\tau_\alpha, \bar{\tau}_\beta)} \\ &= \tau_\beta\Big[\frac{\varepsilon_{\alpha\beta}}{h_\alpha}\tau_\alpha + \frac{h_\beta - \varepsilon_{\alpha\beta}}{1 - h_\alpha}\bar{\tau}_\alpha\Big] \\ &\quad + \bar{\tau}_\beta\Big[\frac{h_\alpha - \varepsilon_{\alpha\beta}}{h_\alpha}\tau_\alpha + \frac{1 - h_\alpha - h_\beta + \varepsilon_{\alpha\beta}}{1 - h_\alpha}\bar{\tau}_\alpha\Big]. \end{aligned} \tag{3.1}$$

At this point it is straightforward to make the connection with the random Ising model formulation of (2.1)–(2.2). Indeed, setting again for $\alpha \in \mathcal{G}$

$$s_\alpha \stackrel{\text{def}}{=} 2\tau_\alpha - 1,$$

and using the properties of binary variables, we have

$$\psi(s_\alpha, s_\beta) \stackrel{\text{def}}{=} \exp\Big(\log p(\tau_\alpha|\tau_\beta)\Big) = \exp\Big(J_{\alpha\beta}s_\alpha s_\beta + H_{\alpha\beta}s_\alpha + H_{\beta\alpha}s_\beta + C_{\alpha\beta}\Big),$$

with

$$\begin{cases} J_{\alpha\beta} \stackrel{\text{def}}{=} \dfrac{1}{4}\log\dfrac{\varepsilon_{\alpha\beta}(1 - h_\alpha - h_\beta + \varepsilon_{\alpha\beta})}{(h_\alpha - \varepsilon_{\alpha\beta})(h_\beta - \varepsilon_{\alpha\beta})}, \\ H_{\alpha\beta} \stackrel{\text{def}}{=} \dfrac{1}{4}\log\dfrac{\varepsilon_{\alpha\beta}(1 - h_\alpha)^2(h_\alpha - \varepsilon_{\alpha\beta})}{(1 - h_\alpha - h_\beta + \varepsilon_{\alpha\beta})h_\alpha^2(h_\beta - \varepsilon_{\alpha\beta})}, \\ H_{\beta\alpha} \stackrel{\text{def}}{=} \dfrac{1}{4}\log\dfrac{\varepsilon_{\alpha\beta}(h_\beta - \varepsilon_{\alpha\beta})}{(1 - h_\alpha - h_\beta + \varepsilon_{\alpha\beta})(h_\alpha - \varepsilon_{\alpha\beta})}, \\ C_{\alpha\beta} \stackrel{\text{def}}{=} \dfrac{1}{4}\log\dfrac{\varepsilon_{\alpha\beta}(h_\alpha - \varepsilon_{\alpha\beta})(h_\beta - \varepsilon_{\alpha\beta})}{(1 - h_\alpha - h_\beta + \varepsilon_{\alpha\beta})h_\alpha^2(1 - h_\alpha)^2}. \end{cases}$$

Finally

$$H_\alpha \stackrel{\text{def}}{=} \sum_{\beta\in a(\alpha)} (H_{\alpha\beta} + H_{\beta\alpha}) + \sum_{\beta\in c(\alpha)} (H_{\alpha\beta} + H_{\beta\alpha}),$$

completes the connection with the model of Section 2.2.

3.4 A mean-field iteration scheme

In order to reconstruct the traffic we use a message-passing procedure among the nodes of the graph, such that each node sends to its neighbors its last updated evaluation of the probability p_α to be saturated. In turn the node evaluate the probability of being saturated from the received messages, accordingly to the mean-field rules we detail now.

Assume that α is collecting the information from its parents in $a(\alpha)$, these neighbors are not connected and a loop passing through several other nodes is needed in general before a connection is established. Accordingly we neglect the correlations between these nodes and write the joint probability $p\big(\tau_\alpha, \{\tau_\beta\}_{\beta\in a(\alpha)}\big)$ as

$$\begin{aligned} p\big(\tau_\alpha, \{\tau_\beta\}_{\beta\in a(\alpha)}\big) &= p(\tau_\alpha)\, p\big(\{\tau_\beta\}_{\beta\in a(\alpha)}|\tau_\alpha\big) \\ &= \big(\tau_\alpha h_\alpha + \bar\tau_\alpha(1-h_\alpha)\big) \prod_{\beta\in a(\alpha)} p(\tau_\beta|\tau_\alpha) \end{aligned}$$

with $p(\tau_\beta|\tau_\alpha)$ given by (3.1).

The definition of the scheme follows these expressions: suppose that at step s, the individual probabilities are given by p_α^s; the update of p_α at the next step is then given by

$$p_\alpha^{s+1} = \sum_{\{\tau_\beta\}} \frac{p(\tau_\alpha = 1, \{\tau_\beta\})}{\prod_{\beta\in a(\alpha)}\big(\tau_\beta h_\beta + \bar\tau_\beta(1-h_\beta)\big)} \prod_{\beta\in a(\alpha)} \big(\tau_\beta p_\beta^s + \bar\tau_\beta(1-p_\beta^s)\big)$$

and the fixed point is therefore solution of

$$p_\alpha = h_\alpha \prod_{\beta\in a(\alpha)} \Big[1 + \big(\frac{\varepsilon_{\alpha\beta}}{h_\alpha h_\beta} - 1\big)\frac{p_\beta - h_\beta}{1-h_\beta}\Big]. \tag{3.2}$$

As we see, in absence of new day-time data, which would force some of the p_β to be different from their historical value h_β, the scheme naturally converges to the historical expectation values. This mean-field procedure is rather crude and gives unsatisfactory results especially when we try to send messages backward in time, because the approximate independence is lost between neighbors.

Algorithm 1 : Mean Field

Input: A set of space-time nodes $\{(\ell,t), \ell = 1 \ldots N, t = 1 \ldots M\}$ a set of historical data $\{h_{(\ell,t)}, \ell = 1 \ldots N, t = 1 \ldots M\}$ and $\{\varepsilon_{\ell,\ell'}, \ell = 1 \ldots N, \ell' = 1 \ldots N\}$, and the corresponding graph $\mathcal{G}$. A set of visited space-time nodes $\{\alpha^*\}$ and the corresponding weights p_{α^*} send by the probe vehicles.

```
for every node (ℓ, t) of the graph do
    if (ℓ, t) is a visited site then
        initialize the field p_{ℓ,t} = p*_{ℓ,t}
    else
        initialize the field p_{ℓ,t} = h_{ℓ,t}
    end if
end for
for t = 2 to t = M do
    sweep the set of nodes, and update sequentially the weights on all the nodes
    of the graph except the visited ones {α*}, generating the values p*_{ℓ,t},
    using 3.2 with parent neighbors.
end for
return the set of fixed weight p*_{ℓ,t}
```

3.5 A belief-propagation algorithm

The simple mean-field scheme presented above has a major drawback: it does not take full advantage of the information that is provided by the probe cars. Indeed, information propagates only forward in time, although to reconstruct the traffic at a given time t in the past, one can use past as well as future information with respect to time t. Using the Markov random field representation of the day-traffic history we may actually set up a belief-propagation algorithm which will cure this problem. The beliefs are estimations of the marginal probabilities p_α that the algorithm will provide [5]. They would be exact if the graph had a tree-like structure, but in our case the graph has cycles which we expect to be sufficiently large to avoid convergence problems [3]. The complication added to the simple algorithm presented above, is that the iterations will not consists to update the beliefs directly, but a new kind of messages which will help to construct the beliefs and which we define now.

The idea of the belief propagation is to factor the marginal probability at a given site in a product of contributions coming from neighboring sites, which precisely are the messages. The update-rules are the following. Let α a node with parents $a(\alpha)$ and children $c(\alpha)$ (see Figure 3.2). The message sent by α to $\gamma \in V(\alpha) = a(\alpha) \cup c(\alpha)$

is

$$
\begin{cases}
m_{\alpha\to\gamma}(\tau_\gamma) = \sum_{\tau_\alpha} \prod_{\beta\in V(\alpha),\ \beta\neq\gamma} m_{\beta\to\alpha}(\tau_\alpha) p_{\alpha\gamma}(\tau_\alpha|\tau_\gamma), & \text{if } \gamma \in V(\alpha), \\
m_{\alpha\to\gamma}(\tau_\gamma) = \sum_{\tau_\alpha} \prod_{\beta\in a(\alpha),\ \beta\neq\gamma} m_{\beta\to\alpha}(\tau_\alpha) p_{\gamma\alpha}(\tau_\gamma|\tau_\alpha), & \text{if } \gamma \in c(\alpha).
\end{cases}
\tag{3.3}
$$

The beliefs b_α are then reconstructed according to

$$
b_\alpha(\tau_\alpha) \stackrel{\text{def}}{=} \frac{1}{Z_\alpha} \prod_{\beta\in a(\alpha)\cup c(\alpha)} m_{\beta\to\alpha}(\tau_\alpha),
$$

with

$$
Z_\alpha = \sum_{\tau_\alpha\in\{0,1\}} \prod_{\beta\in a(\alpha)\cup c(\alpha)} m_{\beta\to\alpha}(\tau_\alpha).
$$

A specific rule is defined for nodes α^* which have been visited by probe cars,

$$
m_{\alpha^*\to\beta} = \sum_{\tau_{\alpha^*}} p(\gamma|\alpha^*) b_{\alpha^*}(\tau_{\alpha^*})
$$

This procedure allows to go beyond the mean-field approximation and to recover the so-called Bethe approximation for the free-energy [7].

4 A Markovian traffic toy-model

In order to test our algorithm of traffic reconstruction we need a convenient traffic network model, on which it is possible to perform measurements and simulate traffic conditions. The model we use is as follows:

The network is composed of nodes corresponding to crossroads or to virtual cuts if road segments are too long, and of a set of oriented links connecting these nodes. Each link ℓ has a given importance g_ℓ, which by convention is in scale ranging from 1 for little rural roads, to 5 for highways. The length of the link is noted $|\ell|$ and should not exceed a few hundreds of meters. Its capacity is $C_\ell = \text{NCARMAX}\ |\ell| g_\ell$, where NCARMAX is the maximum allowable car density in a traffic jam. Typically, $\text{NCARMAX} = 200$ if $|\ell|$ is measured in kilometers.

To each link ℓ we associate a free escape-time τ_ℓ^0, which corresponds to the time it takes to travel all the way through the link, when the traffic is fluid. To each node a, we associate the mean time τ_a^0 it takes to go through the cross-road, when again

Algorithm 2 : Belief-Propagation

Input: A set of space-time nodes $\{(\ell, t), \ell = 1 \dots N, t = 1 \dots M\}$ a set of historical data $\{h_{(\ell,t)}, \ell = 1 \dots N, t = 1 \dots M\}$ and $\{\varepsilon_{\ell,\ell'}, \ell = 1 \dots N, \ell' = 1 \dots N\}$, and the corresponding graph $\mathcal{G}$. A set of visited space-time node $\{\alpha^*\}$ and the corresponding weights p_{α^*} send by the probe vehicles. a maximal number of iterations $it_{\max}$; a requested precision ϵ

Output: UNCONVERGED if BP has not converged, the set of all messages $m_{\alpha\to\beta}$ if BP has converged

At step $it = 0$:
for every edge $\alpha \to \beta$ of the graph **do**
    if α is a visited site **then**
        initialize the messages $m_{\alpha\to\beta} = \sum_{\tau_\alpha} p(\tau_\beta|\tau_\alpha)p^*_\alpha$
    else
        randomly initialize the messages $m_{\alpha\to\beta}(it = 0) \in [0, 1]$
    end if
end for
for $it = 1$ to $it = it_{\max}$ **do**
    sweep the set of edges in a random order, and update sequentially the messages on all the edgess of the graph except the one originating from a visited node $\{\alpha^*\}$, generating the values $m_{\alpha\to\beta}(it)$, using update rules 3.3
    if $|m_{\alpha\to\beta}(it) - m_{\alpha\to\beta}(it-1)| < \epsilon$ on all the edges **then**
        the iteration has converged and generated $m^*_{\alpha\to\beta} = m_{\alpha\to\beta}(it)$
        goto line 12
    end if
end for
if $it = it_{\max}$ **then**
    return UN-CONVERGED
else $\{it < it_{\max}\}$
    return the set of fixed point $m^*_{\alpha\to\beta} = m_{\alpha\to\beta}(it)$
end if

there is no traffic. Typically if there is a traffic light, this scale represents the average time one waits in front of the light plus the transit time through the cross-road.

The network's stochastic evolution is Markovian and continuous w.r.t. time. On each link ℓ a certain number of vehicles $n_\ell(t)$ is circulating at time t and the corresponding load is given by $\rho_\ell(t) = n_\ell(t)/C_\ell(t)$. Let $V^+(a)$ be the set of links which destination is a. To measure the congestion at a, we also define the load $\rho_a(t)$ of the node as

$$\rho_a(t) = \frac{\sum_{\ell\in V^+(a)} g_\ell \rho_\ell(t)}{\sum_{\ell\in V^+(a)} g_\ell}.$$

If a is the destination of link ℓ, and ℓ' starts from there, we associate to the transition

$$n_\ell \longrightarrow n_\ell - 1$$
$$n_{\ell'} \longrightarrow n_{\ell'} + 1.$$

the transition rate

$$\lambda^a_{\ell\ell'}(\rho_\ell, \rho_{\ell'}, \rho_a) \stackrel{\text{def}}{=} \frac{1-\rho_{\ell'}}{\tau_\ell + \tau_a},$$

where

$$\frac{1}{\tau_\ell} \stackrel{\text{def}}{=} \frac{f(\rho_\ell)}{\tau^0_\ell},$$
$$\frac{1}{\tau_a} \stackrel{\text{def}}{=} \frac{(1-\rho_a)}{\tau^0_a},$$

and where f is a positive decreasing function, such that $f(0) = 1$. In order to model a traffic situation with two different regimes, we choose

$$f(\rho) \stackrel{\text{def}}{=} 1 - \eta \frac{\int_0^\rho dx x^n (1-x)^n}{\int_0^1 dx x^n (1-x)^n}, \tag{4.1}$$

with n large. With this choice of function, for $\rho < 0.5$, the traffic is fluid $(f(\rho) \simeq 1)$, whereas for $\rho > 0.5$, it is congested $(f(\rho) \simeq 1 - \eta)$.

Some of the nodes are connected to the outside world and vehicles may enter or leave the system at these nodes. At such a node a, the rate of entrance of vehicles into a link ℓ is

$$\lambda^a_\ell(\rho_\ell) \stackrel{\text{def}}{=} T_{level}\lambda_a(1-\rho_\ell),$$

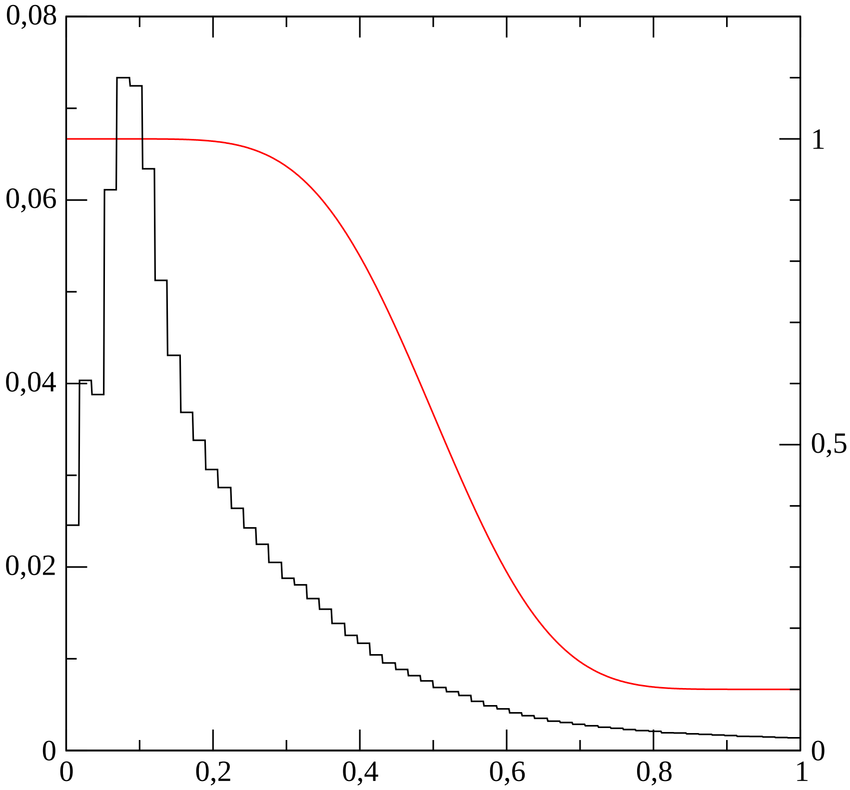

Fig. 5.1: Average distribution of load ρ_ℓ over the network and over day-time (black) and the assumed velocity-load diagram $f(\rho)$ (red).

where $\lambda_a > 0$ is a parameter attached to a and T_{level} is the current global level of traffic. Similarly, for a link ℓ' that ends at a, the exit rate is

$$\mu^a_{\ell'}(\rho_{\ell'}) \stackrel{\text{def}}{=} (1 - T_{level})\frac{\mu_a}{\tau_{\ell'}},$$

with $\mu_a > 0$.

5 Preliminary Simulation Results

We present here results, limited to the mean-field algorithm with causal updates of Section 3.4. A more complete numerical investigation and optimizations of our approach will be the subject of a future paper. This is a preliminary check that the inference principle, based on local correlations, is able to propagate a useful information. The causal updates mean that the mean-field inference rule is applied only forward in time. For the purpose of simulation, we have designed a small network shown on Figure 1.2. The traffic level is a periodic function of period one-day, which has some additional tunable stochastic components in order to have significant fluctuations in the overall traffic intensity (see green curve of Figure 5.3). The velocity-load function f of (4.1) is used with parameters $n = 7$ and $\eta = 0.9$ (see red curve of Figure 5.1). The historical data is extracted directly from the simulation, not from the probe-vehicles, in order to test separately the reconstruction algorithm from the data collection procedure. In addition, this data corresponds to day-time averaging, instead of time stamp average data. The criterion for traffic saturation, is $\tau_\alpha = 1$ if $\rho_\alpha > 0.3$.

The fraction of the nodes which are saturated can then be deduced from Figure 5.1. To assess the fidelity of our traffic reconstitution we use two indicators, which are time dependent,

$$I_1(t) \stackrel{\text{def}}{=} \sum_{\alpha\in\mathcal{G}_t} p_\alpha\tau_\alpha + \bar{\tau}_\alpha(1-p_\alpha) \tag{5.1}$$

$$I_2(t) \stackrel{\text{def}}{=} \sum_{\alpha\in\mathcal{G}_t} p_\alpha\tau_\alpha \tag{5.2}$$

where the p_α are the predictions and $\tau_\alpha \in \{0,1\}$ come directly from the simulation. $\mathcal{G}_t$ is the subgraph of $\mathcal{G}$ corresponding to nodes which have a day time-stamp smaller than t. An instance of a reconstructed network is given in Figure 5.2. From the load distributionof Figure 5.1, given the value of the threshold, we see that the great majority of the links are not saturated. Therefore, the second estimator is much more relevant than the first one, since it takes only the saturated links in account. In this respect we see an improvement in Figure 5.3 in predicting the presence of traffic jams when probe-car are used to infer the traffic network. At the end of each day, the fraction of detected jams, which from historical data is around 20% in Figure 5.3(a), rises to above 50% with only one probe vehicle (Figure 5.3(b)) and reaches around 70% with 5 probe cars (Figure 5.3(c)). Instead the first estimator which in our case is more representative of the population of unsaturated nodes, does not discriminate between historical data, and probe vehicles inference data.

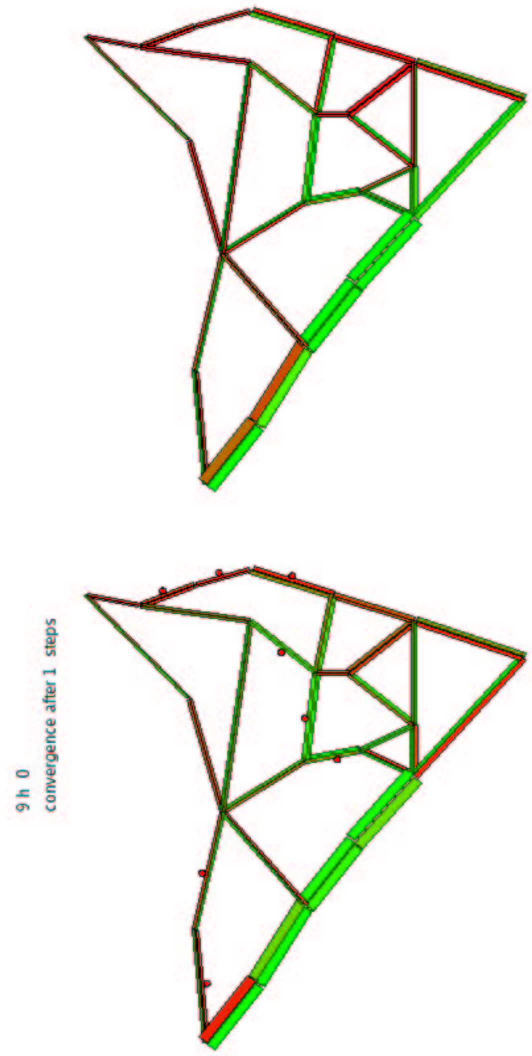

Fig. 5.2: Comparison of the real traffic network situation (left) with the inferred one (right) with five probe-vehicles, using the mean-field algorithm. Red dots represent the presence of probe vehicles during the time interval (10 min). Color varies from green (fluid) to red (saturated), proportionally to the load $\rho_\ell(t)$ (left) or to p_α (right).

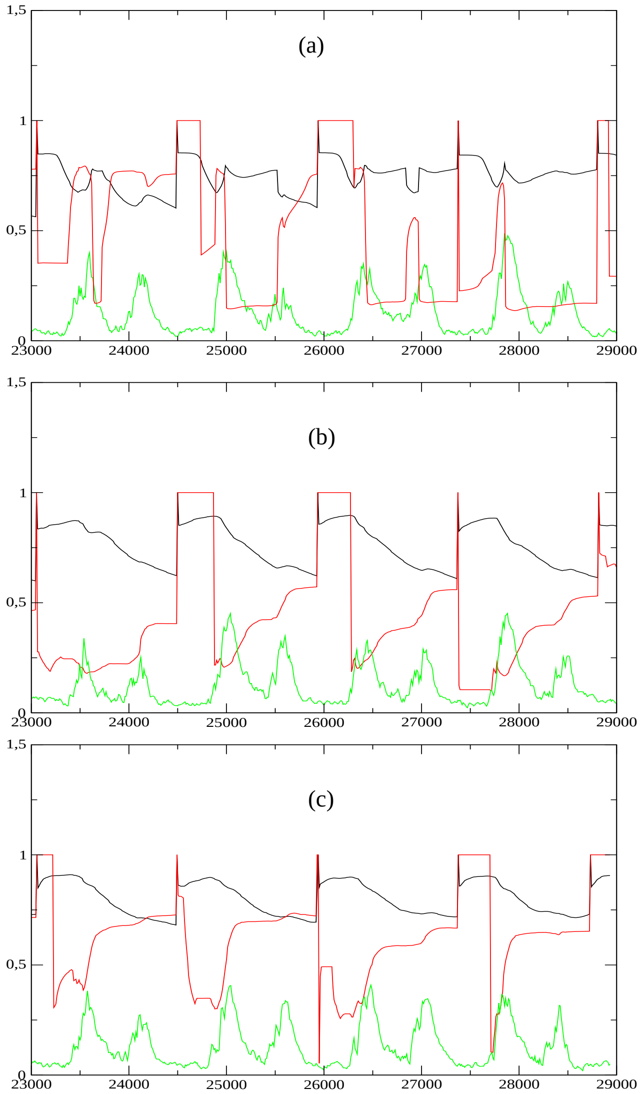

Fig. 5.3: Performance evaluation of the mean-field algorithm over four day-time cycles, by comparing traffic reconstruction with historical data (a), with 1 probe vehicle traffic reconstruction (b) and 5 probe vehicles (c). Curves in green are the traffic intensity measured over the network, in black the estimator 5.1, in red the estimator 5.2.

References

[1] REACT, Realizing Enhanced Safety and Efficiency in European Road Transport. `http://www.react-project.org/`.

[2] Ising, E. Beitrag zur theory des ferromagnetismus. *Zeitschr. f. Phys. 31* (1925).

[3] Kschischang, F., Frey, B., and Loeliger, H. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory 47* (2001), 498–519.

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

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

[6] Smith, B., Williams, B., and Oswald, R. Comparison of parametric and nonparametric models for traffic flow forecasting. In *Transportation Research Part C* (2002), vol. 10.

[7] Yedidia, J., W.T., F., and Weiss, Y. Generalized belief propagation. *Advances in Neural Information Processing Systems* (2001), 689–695.

[8] You, J., and Kim, T. J. Towards developing an expert GIS-based travel time forecasting model with congestion pattern analysis. In *proceedings of ITS congress* (1998).

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399