Detecting Dynamic Critical Links within Large Scale Network for Traffic State Prediction

Pierre-Antoine Laharotte, Romain Billot, Nour-Eddin El Faouzi

Detecting Dynamic Critical Links
within Large Scale Network for Traffic State
Prediction

Pierre-Antoine LAHAROTTE
Univ. Gustave Eiffel, Univ. Lyon, ENTPE
UMR LICIT
F-69675 Lyon, France
pierre-antoine.laharotte@univ-eiffel.fr

Romain BILLOT
IMT Atlantique, LUSSI
UMR CNRS 6285 Lab-STICC (DECIDE)
CS 83818, Brest, France
romain.billot@imt-atlantique.fr

Nour-Eddin EL FAOUZI
Univ. Gustave Eiffel, Univ. Lyon, ENTPE
UMR LICIT
F-69675 Lyon, France
nour-eddin.elfaouzi@univ-eiffel.fr

Abstract—Can we expose the relationship between the physical dynamics of a network and its predictability? To contribute to this point, we propose a dimensionality reduction method for network states prediction based on spatiotemporal data. The method is intended to deal with large scale networks, where only a subset of critical links can be relevant for accurate multidimensional prediction (MIMO) performances. The algorithm is based on Latent Dirichlet Allocation (LDA) to highlight relevant topics in terms of networks dynamics. The feature selection trick relies on the assumption that the most representative links of the most dominant topics are critical links for short term prediction. The method is fully implemented to an original application field: short term road traffic prediction on large scale urban networks based on GPS data. Results highlight significant reductions in dimensionality and execution time, a global improvement of prediction performances as well as a better resilience to non recurrent traffic flow conditions.

Index Terms—feature selection, network, spatiotemporal dynamic, traffic, Multi-Input Multi-Output prediction

I. INTRODUCTION

A. A general issue

We study short term state prediction, the task of predicting the picture of a network at time $t + h$, based on historical data. The directed network is composed of P links, each link being characterized by a state at time t. For many large scale networks such as energy, transport, or internet, a current challenge relies on the ability to scale up prediction performances based on complex spatiotemporal data. It is reached thanks to Multi-Input Multi-Output (MIMO) supervised regressions from K-NN [1] to NN [2] and M-SVR [3]. The **main challenge** of these approaches lies in the high-dimensionnality and dynamic of the system. From a supervised learning standpoint, this implies to either propose a **dynamic compact representations of the individuals** or **to dynamically reduce inputs dimensions** [4], [5] , coping with relevance and redundancy [6]. This common objective is shared by many application fields [7] with various methods [8] and the interest in the ML community is still strong due to increasing dimensionality [9], constrained budget [10] or even security [11] constraints. A second challenge lies in the

resilience to non recurrent network conditions such as traffic congestion. Hence, the research issue can be summarized as the selection of the best subset of critical links (inputs of the system) in spatiotemporal networks with following properties:

- Best trade-off between dimensionality reduction and generalization performances,
- Resilience to recurrent or non recurrent congested phases,
- Understanding of the relationship between dynamics of the physical system and network's preditability.

We formulated the problem as a feature selection process where the P links at time t are the initial variables of the network. The introduced algorithm (see Figure 1) is based on the 4 following assumptions and steps operating off-line:

1) The link criticality is related to its physical dynamic, i.e. the trend a link has to propagate or catalyze its traffic states to its parents or children, with some time lags to be estimated. Hence, each link is characterized, at any time step, by a **state configuration** that includes itself and its closer spatiotemporal neighbors : its immediate parents (downstream) and children (upstream).
2) Used as a non-supervised process, a Latent Dirichlet Allocation is performed to highlight **time-independent underlying state configuration topics** and **to dynamically sort links** (input) for any state configuration topic.
3) The most contributive links to the most efficient topic cluster (in terms of traffic prediction) are assumed to be the most **critical links**. As the contribution of links to topics vary with time and network dynamics, the set of critical links vary accordingly.
4) The **optimal number of critical links** is selected in terms of trade-off between global prediction error and dimensionality reduction. The prediction method is considered as a black box. The critical links are assumed to be independent of the prediction method and make sense for any other application.

The study proposes 2 independent contributions:

- **classification of the network links** according to some underlying traffic dynamic features and identification of
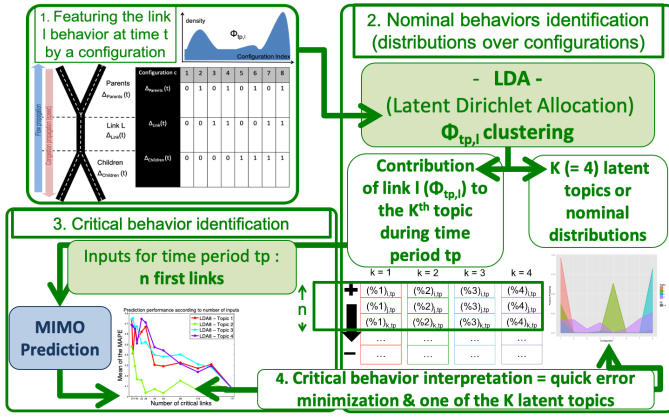
Fig. 1: Comprehensive process applied to a link l.

the critical cluster thanks to a Multi-Input Multi-Output (MIMO) forecasting method;

- **proposition of a subset feature selection process** based on a ranking indicator oriented by the critical category, as one potential application.

A constant effort has been made in this study to maintain the relationship between physical approach and statistical aspects, what deep learning methods do not explicitly allow. It aims to provide a better understanding of their networks to managers.

B. A domain-specific application

Although the proposed methodology is designed to be generalized to various kinds of dynamic networks, we assume that the definition of criticality strongly depends on the application. As a matter of fact, the direction of the information through the network is application-dependent. We design and apply our methodology to short-term road traffic prediction, assuming flow conservation at the nodes of the oriented network. In this context, traffic states are represented by mean speeds (based on GPS [**?**]) on road links during a given time period.

Traditionally, forecasting objectives are achieved through univariate time series prediction based on various techniques, from statistical models (ARIMA family [12]) to the non parametric regressive approach with artficial neural networks ([2], [13]), support vector or random forest regression ([14]). Vlahogianni et al. [15] provide a wide overview of state of the art. They highlight that urban road traffic networks are a representative example of current challenges in short term prediction in large scale networks. The urban networks are characterized by high variability over time, making prediction challenging for multiple time horizons [16]. They are also highly sensitive to non recurrent events (congestion, accidents, road works), as shown in [17]. Finally, the spatio-temporal structure is of paramount importance for coming up with the best possible prediction method ([18], [19]). Indeed the immediate upstream and downstream neighbors of a link may impact its current traffic state and such contribution may vary in the time.

The usual approach to captur spatio-temporal structure lies on the identification of spatial or temporal autocorrelation between one link and its neighboorhood [20], but are applied to non-MIMO approaches. The applied strategies can broadly be separated into 2 categories: The spatial analysis aiming to fix the optimal neighboorhood of the link to consider ([21]) and the temporal analysis targeting the correlation weight within a specific and predefined neighboorhood of a specific link [22]. None are able to captur both [23].

Motivated by a clear need for high-scale performance, Multi-Input Multi-Output (MIMO)) approaches have recently emerged ([1], [24]–[26]). In contrary to local methods, which predict links states individually ([19], [27]), MIMO approaches predict pictures of the network, each individual (row) being a snapshot of the network at time t with potentially short memory ($t - 1$, etc). Originally its main benefit lies in the opportunity to preserve the full spatio-temporal structure between the network links, hence to make a more accurate prediction [1], [24] with a shorter computation time and to leave the selection issue to the forecasting method. Nevertheless some noisy or unsignificant variables may alter the forecasting performance by hidding the latent spatio-temporal structure [2]. A MIMO prediction approach will be used in this paper and the goal is then to perform a dimensionality reduction by selecting the optimal subset of critical links in order to ensure fast, accurate and physically understandable traffic prediction.

This paper is organized as follows. In section II, the prediction problem is defined as well as the design of dynamic state configurations (step 1 in Figure 1). In section III, the LDA process is used to highlight some time-independent underlying structur among state configuration distributions (step 2). The algorithm for feature selection (step 3) is then applied. It enables a selection of an optimal subset of critical links for on-line prediction purpose (step 4). Experiments are carried out in section IV.

II. PROBLEM SETUP AND ASSUMPTIONS

A. The prediction problem

The network is composed of P links. Let $X_{t,l}$ be the state of a link l at time t, $\forall t \in [1, T]$. The state vector is a snapshot of the network states (e.g. speed) at time t, *i.e.* the simultaneous values on the P links, written

$$\boldsymbol{X_t} = \left(X_{t,1}, ..., X_{t,P} \right)^T.$$

The learning algorithm makes a prediction about the global network state given a time horizon h,

$$\widehat{\boldsymbol{X_{t+h}}} = g(X_{t,1}, ..., X_{t,P})^T.$$

Depending on the targeted states, the inferred function g can be either a classification or a regression function [28]. Various strategies exist for the learner g and the goal of this paper is not to propose a new prediction method. It will be rather considered as a black box. The only prerequisite is to use a MIMO prediction method, e.g. Multi-Output K-NN. To deal with the curse of dimensionality affecting these MIMO approaches, it is assumed that only the information sheltered by a subset of critical links is relevant to ensure both accurate prediction and smaller execution time. The main issue tackled in this paper is the selection of n critical links that will

ensure an optimal trade-off between global prediction error and computation time. In order to identify what is behind the concept of criticality, we start by focusing the analysis on the links as components of the network.
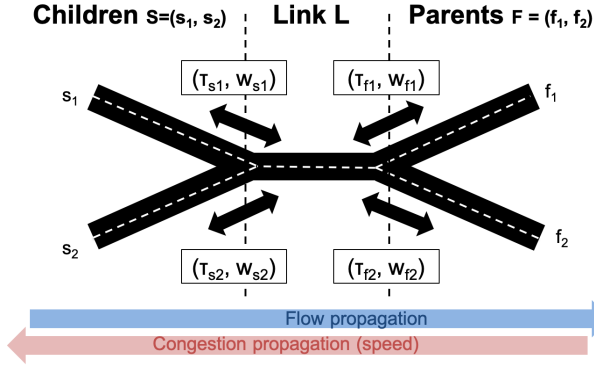

Fig. 2: Application to road traffic state configuration : flow and speed (congestion) directions.

B. From current link state to local configuration

In an oriented network, we consider that the current state of a link l is only dependent on (i) the inbounding throughput of its children (upstream) and on (ii) the outbounding capacity of its parents (downstream), what affects the flow speed by upstreaming. Hence, our assumption is that critical links result from a specific configuration, *i.e.* relationships with its direct neighborhood, or distribution over configurations. We introduce the concept of **local state configuration** $\delta_l(t)$ for a link l at time step t as an encoding of a set composed of: the current state of a link l, written $X_{t,l}$, and the state of its immediate parents and children with appropriate time lag. Parents and children share with the current link a node of the oriented network. We base our process on the cross-correlation to expose this specific relationship between a link, its parents and its children. To depict the signal (speed) propagation (Figure 2), we assume that link $X_{t,l}$ is impacted by its parent f ($f \in [1,F]$, where F is the number of parents for a link l) with a (negative) time lag τ_f and a contribution w_f. The link l impacts its children s ($s \in [1,S]$ with S the number of children) with a (positive) time lag τ_s and a significance w_s. In order to estimate the two components τ and w we define

$$\boldsymbol{X_l} \star \boldsymbol{X_{t-\tau,f}} = \int \boldsymbol{X_l} \boldsymbol{X^*_{t-\tau,f}} \mathrm{d}t.$$

as the cross-correlation between the state vectors of link l and parent f with a (negative) time lag τ, and, respectively,

$$\boldsymbol{X_l} \star \boldsymbol{X_{t+\tau,s}} = \int \boldsymbol{X_l} \boldsymbol{X^*_{t+\tau,s}} \mathrm{d}t.$$

as the cross-correlation between the state vectors of links l and child s with a (positive) time lag τ. Then,

$$\forall f \in [1,F],$$
$$\begin{cases} \tau_f &= \arg\max_\tau (\boldsymbol{X_l} \star \boldsymbol{X_{t+\tau,f}}), \\ w_f &= \frac{|\boldsymbol{X_l} \star \boldsymbol{X_{t+\tau_f,f}}|}{\sum_{i=1}^{F} |\boldsymbol{X_l} \star \boldsymbol{X_{f,t+\tau_i}}|}. \end{cases} \quad (1)$$

and

$$\forall s \in [1,S],$$
$$\begin{cases} \tau_s &= \arg\max_\tau (\boldsymbol{X_l} \star \boldsymbol{X_{t+\tau,s}}), \\ w_s &= \frac{|\boldsymbol{X_l} \star \boldsymbol{X_{t+\tau_f,s}}|}{\sum_{i=1}^{S} |\boldsymbol{X_l} \star \boldsymbol{X_{t+\tau_i,s}}|}. \end{cases} \quad (2)$$

In this study, a stationarity assumption is made for the time lags τ (positive or negative), and the contribution weights w. In order to retrieve comparable configurations for any link, a further assumption is made by grouping together the parents (resp. children). This assumption is non-binding, because only the comprehensive impact of its parents (resp. children) matters for the current link (only the constraint matters, not the origin of the constraint). It enables to define the variable $\boldsymbol{\delta_l} = (\boldsymbol{\delta_{Parents}}, \boldsymbol{\delta_{Link}}, \boldsymbol{\delta_{Children}})$ which is related to the weighted variations of links states according to previously defined time lags:

$$\begin{cases} \delta_{Parents}(t) &= \sum_{f=1}^{F} (w_f \times (X_{t-\tau_f,f} - X_{t-\tau_f-1,f})), \\ \delta_{Link}(t) &= (X_{t,l} - X_{t-1,l}), \\ \delta_{Children}(t) &= \sum_{s=1}^{S} (w_s \times (X_{t+\tau_s,s} - X_{t+\tau_s-1,s})). \end{cases} \quad (3)$$

The $\boldsymbol{\delta_l}$ variable depicts the state (speed) variation between time t and time $t-1$, with an adjusted shift for parents and children of link l. The goal is to determine whether a link l has been impacted by a variation coming from its parents or/and whether the link l has propagated this variation to its children. Then, coding a local configuration consists in discretizing in order to interpret the $\boldsymbol{\delta_l}$ values according to the real world application (see Figure 3). For sake of simplicity (and power of interpretation), we encode $\boldsymbol{\delta_l}$ into binary variables $\Delta_l(t) = (\Delta_{Parents}(t), \Delta_{Link}(t), \Delta_{Children}(t))$ that take on 0 or 1 according to the sign of $\boldsymbol{\delta_l}$ ones. Assuming that the propagation of a degradation is of interest (*i.e.* a decrease in the state value), then

$$\begin{cases} \Delta_{Parents}(t) &= 1 \text{ if } \delta_{Parents}(t) < 0, \ 0 \text{ otherwise}, \\ \Delta_{Link}(t) &= 1 \text{ if } \delta_{Link}(t) < 0, \ 0 \text{ otherwise}, \\ \Delta_{Children}(t) &= 1 \text{ if } \delta_{Children}(t) < 0, \ 0 \text{ otherwise}. \end{cases} \quad (4)$$

A local configuration at time t, $C_{t,l}$, is then fully defined by the triplet $\Delta_l(t) = (\Delta_{Parents}(t), \Delta_{Link}(t), \Delta_{Children}(t))$. Let $c (c \in [1,N_c])$ be one configuration. As illustrated by Figure 3 for this simplified case, eight ($N_c = 2^3$) potential local configurations result from this triplet discretization process.

C. Adaptation to the case of road traffic prediction

The method is applied to road traffic prediction involving physics concepts illustrated by figure 2. As a signal, a throughput of vehicles travels from upstream to downstream link, from children to parents. As a reflected signal, congestion appears when the demand is higher than the offer. Congestion, featured by speed, propagates upstream, from parents to children. In this study, links are valued by mean speeds and a degradation of the state (congestion) is related to a negative variation of speed.
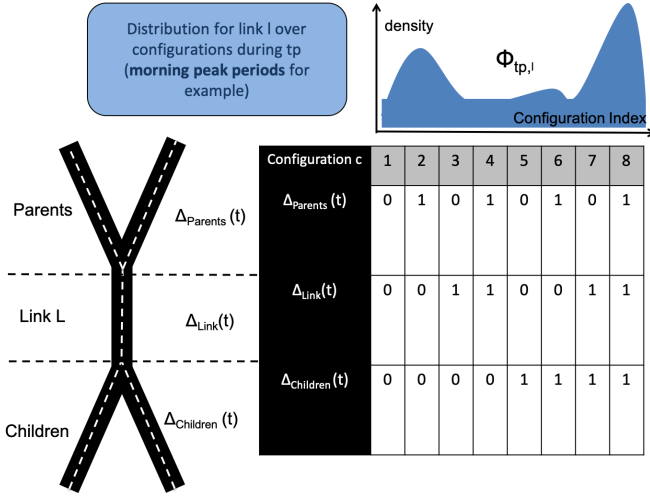
Fig. 3: Local configurations for a link l.

D. The comprehensive methodology

The comprehensive 4-steps methodology is exposed in Figure 1. It illustrates the off-line process applied to a training dataset used to define the criticality and to associate to any timeperiod the appropriate subset of links, *i.e.* inputs of the prediction. It is assumed that the iteration of a specific underlying (but unknown) link configuration, called critical, (step 1-2) during a timeperiod involves the necessity to keep the link to perform prediction. The critical link configuration (step 2) is supposed to be highlighted (step 3) by comparing the prediction performances of various link configurations. For a specific link configuration, the selection of links is operated by ranking the links in terms of iteration during the timeperiod. The link confguration associated to the best prediction performance with the less of inputs (links) is assumed as the critical configuration. Finally the on-line stage is implemented, the link selection is operated according to the timeperiod on a new dataset and its performance is assessed.

One could assume that one of the 8 configurations (see section II-B) matches with the expected critical configuration and leverage it to discriminate the links used as input of the prediction process. Skipping the section III, this assumption is evaluated in section IV through the M3 method. It is pointed out that isolated configurations does not perform accurately and further investigations is of interest, what is exposed in section III and in algorithm 1.

III. A LDA-BASED ALGORITHM FOR FEATURE SELECTION: MINING CRITICAL LINKS

In this section, the comprehensive algorithm 1 is exposed and then each step are described. In step 0, a further refinement of the link configuration is introduced by featuring any link l by its distribution $\boldsymbol{\Phi_{tp,l}}$ over the N_c configurations during a time window TP (see function ConfigFrequencyInTP). Here, the full time period T of the training set is split into TP disjointed homogeneous time-periods (*e.g.* days periods). Then, the LDA makes use of the local configurations distributions $\boldsymbol{\Phi_{tp,l}}$ to highlight a set of K candidates configuration to

the criticality (latent topics) (step 1). Any latent topic is a leverage candidate to discriminate the links and is assessed according to the prediction performance (step 2). The most representative links of the best topic (in terms of prediction error) are considered as critical links (step 2 and 3).

Algorithm 1 LDA based Feature Selection Process

INPUTS: training dataset X (size $T \times P$,) new dataset newX (size $T \times P$), N_Δ {number of discrete categories}, $N_c = N_\Delta^3$, K

PRESET FUNCTIONS: CrossCorrelation {eq 1}, TripletBuilder {eq 3}, Discretizer {eq 4}, IdxConfig, ConfigFrequencyInTP {assess iteration of configurations for timeperiod tp}, LDA, DescendingSort, AssessTimePeriod, AssessPerf, MIMO, ElbowMethod

  STEP 0: *Configuration encoding* {see section II}
  for $l = 1$ **to** P **do**
    $\tau[l], w[l]$ = CrossCorrelation(X, DistMat, l);
    δ_l = TripletBuilder(X, DistMat, l, $\tau[l], w[l]$);
    $\boldsymbol{\Delta_l}$ = Discretizer(δ_l, N_Δ) ; C[., l] = IdxConfig($\boldsymbol{\Delta_l}$);
    {C is a matrix of size $T \times P$}
    for $tp = 1$ **to** TP **do**
      $\Phi(tp, l, 1 : N_c)$ = ConfigFrequencyInTP(C, tp, N_c);
    end for
    {Φ is a tensor of size $TP \times P \times N_c$}
  end for
  STEP 1: *Latent Topics determination* {see section III-A}
  Λ, π = LDA(K, Φ);
  {Λ (size $K \times N_c$) & π (size $(P \times TP \times K)$}
  STEP 2: *Topic selection* {see section III-B}
  {Build ranking indicator}
  for $k = 1$ **to** K **do**
    for $tp = 1$ **to** TP **do**
      $SortedLinks(1 : P, tp, k)$ = DescendingSort ($\pi(1 : P, tp, k)$, $[1 : P]$);
    end for
  end for
  {Apply 'Best First' strategy}
  for $nb_{input} = 1$ **to** P **do**
    for t = 1 **to** T **do**
      tp = AssessTimePeriod(t);
      input = newX(t, $SortedLinks(1 : nb_{input}, tp, k)$);
      Prediction (t, nb_{input}, k) = MIMO(input);
    end for
    MAPE(k, nb_{input}) = AssessPerf(Prediction);
  end for
$\hat{k}, \dot{n}$ = find(min(MAPE))
  STEP 3: *Optimal number of variables selection* {see section III-C}
$\hat{n}$ = ElbowMethod(MAPE($\hat{k}$, .))

A. Step 1: Determination of latent topics (LDA)

A Latent Dirichlet Allocation (LDA) process has been applied to all $\boldsymbol{\Phi_{tp,l}}$, *i.e.* probability distributions of all links over the $N_c = 8$ configurations during all homogeneous time

periods tp, $tp \in [1, TP]$. As a generative probabilistic model, the LDA provides two specific outputs (Λ, π) allowing to characterize and build topics over the configurations. These topics are representative of the networks dynamics in terms of local behaviors. Probabilistic Topic Models such as LDA [29] have been already applied to other issues: from mobile phone data [30], to Bicycle Sharing Systems (BSS) in the transportation domain [31]. The main principle of topic models such as LDA is that each link l contributes, during a specific time period tp, to the network mechanism given its associated behavior $\boldsymbol{\Phi_{tp,l}}$. The LDA aims to highlight the latent pattern within a corpus of links behaviors leading to a better understanding of the relationship between link and network dynamics. In contrary to the previous application in the transportation domain [31], the applied process returns to LDA basics with a slight variation: words configurations replace words and the document is assumed to be a bag of words configurations. The analogy between the original LDA and this application context is described as follows:

- **Corpus of documents**: set of $\boldsymbol{\Phi_{tp,l}}$, $\forall tp \in [1, TP], \forall l \in [1, P]$,
- **Document (bag of words)**: $\boldsymbol{\Phi_{tp,l}}$, distribution over the possible configurations for the link l during the time period tp,
- **Word**: local configuration $c \in [1, N_c]$, based on the triplet $\Delta_l(t) = \big(\Delta_{Parents}(t), \Delta_{Link}(t), \Delta_{Children}(t)\big)$.

Using this analogy, the generative process shall be rewritten in the following form. First, the K latent nominal distributions or topics are drawn using a Dirichlet distribution over the set of configurations:

$$\boldsymbol{\Lambda^{(k)}} \sim \mathcal{D}(\beta), \forall k \in \{1, \dots, K\}. \tag{5}$$

The optimal number of K is determined using a perplexity measure [32]. Then, each bag of local configurations $\Phi_{tp,l}$ in the set of bags $\{1, \dots, P \times TP\}$ is supposed to be generated according to the following two-step process.

1) Draw the proportions of the topics in the bag:

$$\pi^{(tp,l)} \sim \mathcal{D}(\alpha).$$

2) For each configuration of the bag (tp, l)
   a) Draw its topic k:

   $$k \sim \mathcal{M}(1, \pi^{(tp,l)}).$$

   b) Draw a configuration c using topic k:

   $$c \sim \mathcal{M}(1, \boldsymbol{\Lambda^{(k)}}).$$

The probabilistic model describing this generative process is still the classic LDA model. Main outputs from LDA are the $\boldsymbol{\Lambda^{(k)}}$ and the $\pi^{(tp,l)}$ which can be interpreted in the context of link behavior analysis as follows:

- $\Lambda_{k,c} \equiv P\big(c \mid k\big)$: Discrete distribution over configurations. They can be interpreted as typical distributions over the configurations, *i.e.* as nominal behaviors for links. $\boldsymbol{\Lambda^{(k)}}$ enables the interpretation of topic k.
- $\pi^{(l,tp)} \equiv P\big(k \mid l \text{ and } tp\big)$: Contribution of link l during time period tp to build latent topic k.

B. Step 2: Topic selection

For any topic k and time-period tp, a ranking indicator $SortedLinks_{tp,k}$ (see DescendingSort) is built: links l are sorted by descending $\pi^{(l,tp)}(k)$. Indeed, the higher $\pi^{(l,tp)}(k)$ is, the higher the link l contributes to topic k during time-period tp. Then, for any topic and appropriate time-period, the links ranking is used to select the nb_{input} first features (links) to use as inputs of a MIMO algorithm (black box), where nb_{input} is ranging from 1 to P. The prediction error is assessed by computing the average $\boldsymbol{MAPE}$ and is used as a performance proof to discriminate ranking indicators $SortedLinks_{tp,k}$ (see AssessPerf).

$$MAPE_{k,n} = mean_{l \in [1,P]}\big(\frac{1}{T}\sum_{t=1}^{T}|\frac{\widehat{X_{t+h,l}} - X_{t+h,l}}{X_{t+h,l}}| \times 100\big)$$

Then, the topic minimizing the prediction error with the smaller nb_{input} is considered as the optimal topic $\hat{k}$.

C. Step 3: Optimal number of variable selection

The performance target of this feature selection process consists in finding a trade-off between the minimization of the global error and the reduction of the number of input variables. The Elbow method (see ElbowMethod) applied to the performance $MAPE_{\hat{k}}$ of the critical topic enables the identification of the optimal number $\hat{n}$ of variables. The Elbow method is a visual process assuming that the optimal value is defined by a visible curve inflexion (Elbow effect).

IV. EXPERIMENTS ON REAL DATA

A. Traffic data

The algorithm is applied to real GPS data collected at a city-wide scale (Nantes, France) during 3 months ($T = 4368$, from September to November 2013) on a urban road network composed of 1286 links. A link state $X_{t,l}$ represents the mean speed of all GPS-equipped vehicles that have been recorded on the link l during the last 30 minutes-long time step. The learner g, assumed as a black box, is a multi-dimensionnal K-NN regression algorithm [1]. The dataset is divided into a training subset composed of September and October for LDA topic determination (offline) and a validation subset (November) for performance assessment. Some periods of November experimented unusual traffic conditions (incidents). It has been labelled with "unusual conditions" and compared to other periods labelled with "usual conditions".

B. Results of the feature selection algorithm applied to traffic data

Considering a binary encoding of the triplet $\Delta_l(t)$ (see section II) 8 local traffic configurations are obtained. According to traffic experts, the time is split into $TP = 4$ disjoint homogeneous time periods: morning peak period, noon, evening peak period, night.

The **step 1** of algorithm 1 consists in applying the LDA to local traffic configurations (bags of $\Phi_{tp,l}$). It leads to $K = 4$ topics, as depicted by Figure 4. Each of these topics comes
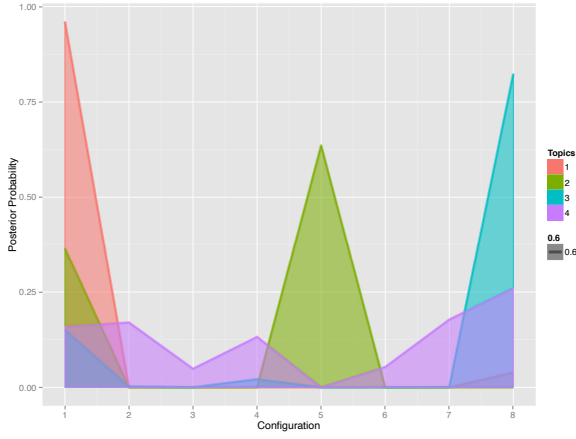
Fig. 4: Distribution of configurations, featuring $\mathbf{\Lambda}^{(k)}$ for each of the 4 topics.

generated by configurations 1 *(0,0,0)* and 5 *(0, 0, 1)* as seen on Figure 4. It means that critical links are mainly representing 2 trends: one minor trend with free flow traffic and one major trend degrading traffic states on its children. It supports our hypothesis concerning the relationship between link criticality and prediction performance.

The only case where topic 2 is surpassed by topic 4 is for the shortest prediction horizon ($H = 30min$) with perfect traffic conditions. Such modification of the criticality with the prediction horizon makes sense as the topic 4 is relevant with very short-term dynamics of the network horizons : it is influenced by 3 configurations: 2nd (1, 0, 0), 7th (0, 1, 1) & 8th (1, 1, 1), which express the trend of the link to propagate to its children.

with a links ranking indicator, $SortedLinks_{tp,k}$, built from the links contributions to the topic k.

The **step 2** consists in the performance comparison of the emerging topics to identify the critical one as shown on figure 5 and summarized by table I for various prediction horizons. Results are exposed for both recurrent and non recurrent conditions. For sake of simplicity, results are displayed for a network subspace composed of ($P = 177$) links.
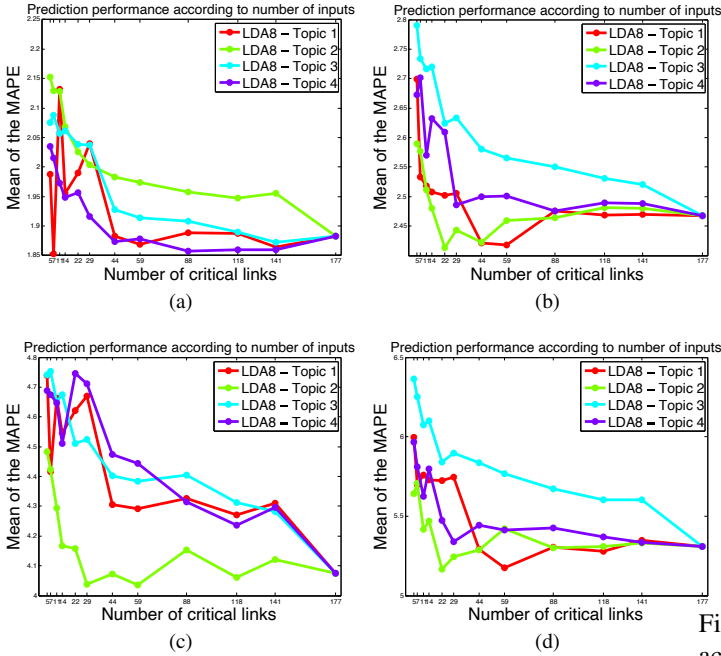

Fig. 5: Performance of the $K = 4$ topic-based ranking according to nb_{input} with prediction horizon h. a) Under recurrent conditions (h=30'). b) Under non recurrent conditions (h=30'). c) Under recurrent conditions (h=90'). d) Under non recurrent conditions (h=90').

The best topic under non recurrent congestion is the 2nd one (green) for most of prediction horizon. This topic is mainly
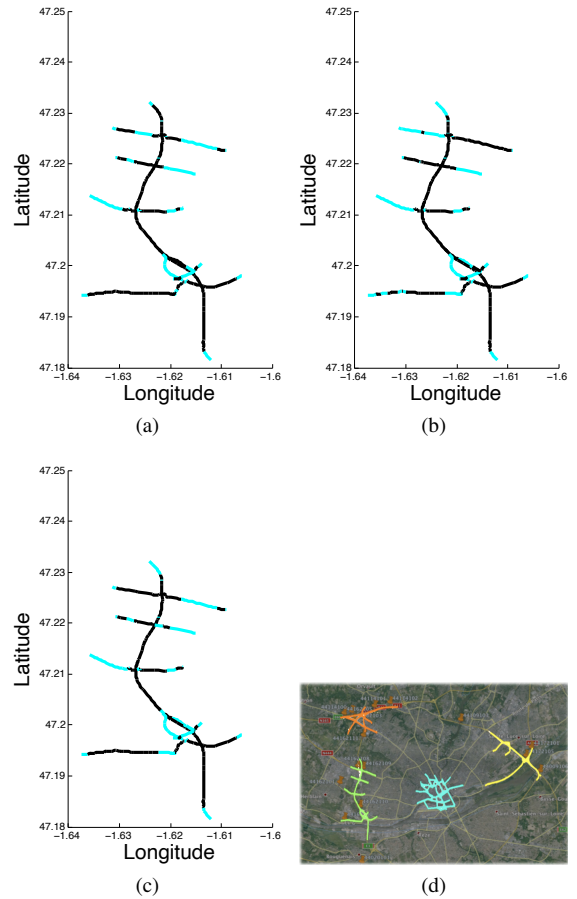

Fig. 6: Identification of the 44 critical links (colored in cyan) according to time period tp and topic 2 of the method LDA8. a) During the night. b) During morning peak period. c) During evening peak period. d) Nantes network.

The **step 3** consists in the determination for any time period of the optimal $\hat{n}$ critical links and is based on the elbow method. Optimal $\hat{n}$ are summarized according to the selected topic $\hat{k}$ by prediction horizon in table I. As expected, optimal topics and hence critical links may also vary with time periods tp, as illustrated by figure 6. The execution time is significantly

Horizon prediction	H = 30 mn	H = 60 mn	H = 90 mn	H = 120 mn	H = 150 mn	H = 180 mn
$\hat{k}$	4 (recurrent) 2 (non recurrent)	2	2	2	2	2
$\hat{n}$	48	30	35	35	32	31
% of time reduction	0.44	0.37	0.34	0.32	0.33	0.34

TABLE I: Time improvement provided by the algorithm.

reduced.

C. Alternative ranking methods

Basically, the method is likely to be sensitive to the ranking strategy and to the type of network. In order to assess this sensitivity, the method proposed above, noted $LDA8$, will be compared to other ranking strategies and tested on Multi-output KNN (as reference method):

Method M1: Criticality is linked to the centered-reduced speed of the link. For each time-period, the median speed is assessed and used as ranking indicator:

$$\pi_{tp,l}^{M1} = median_{tp}(X_{t,l}),$$

$\forall l \in [1, P], \forall tp \in [1, TP]$. The critical links are ranked by increasing speeds values (*i.e* most critical = lowest speed).

Method M1Bis: Criticality is defined by the centered-reduced speed deviation, *i.e.* $\forall l \in [1, P], \delta_l(t) = X_{t,l} - X_{t-1,l}$. For each time-period, the median speed deviation is assessed and used as ranking indicator:

$$\pi_{tp}^{M1Bis}(l) = median_{tp}(|\delta_l(t)|),$$

$\forall l \in [1, P], \forall tp \in [1, TP]$. Critical links have the highest speed variations.

Method M2: Criticality is linked to the probability of negative speed deviation. For each time-period tp, the probability of negative speed deviation is assessed and used as ranking indicator:

$$\pi_{tp}^{M2}(l) = freq_{tp} \, \Delta_{Link}(t),$$

$\forall l \in [1, P], \forall tp \in [1, TP]$. Critical links have the most congested traffic states.

Method M3: Instead of testing LDA topics, this method assumed that any of the 8 configurations ($\forall c \in [1, N_c]$) could be critical, its criticality being defined by $\pi_{tp,l,c}^{M3} = \Phi_{tp,l}(c)$, the probability for link l during time-period tp to be in configuration c. The 5th configuration, which contributes the more to the best topic in method $LDA8$, has been selected as a challenging method. Nevertheless the analysis of all configurations is computationaly expensive.

Method LDA64: This method assumes that the discretization of the traffic states (Discretizer) resulting in $\mathbf{\Delta_l}$ is no more binary, but refined in 4 bins: $\delta > 0, 5$, $0 < \delta \leq 0, 5$, $-0.5 < \delta \leq 0$ or $\delta < -0.5$. 64 (= 4^3) configurations results from this refinement and 7 potential configuration topics are

highlighted by application of the LDA classification. The 6th topic is suitable for featuring the criticality.

Method MLasso: Criticality is assessed through a LASSO method generalized to multidimensionnal outputs [33]. The ranking indicator is built from the contribution of link l to explain the model variance for time-period tp: $\pi_{tp}^{MLasso}(l) = mean(Lasso_{tp}(l))$.
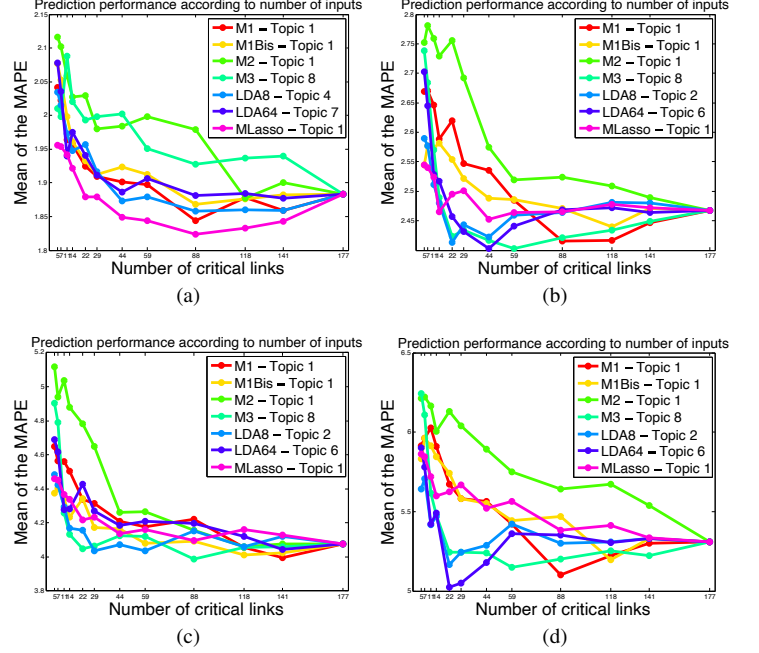
Fig. 7: Prediction performance according to nb_{input} with an horizon h a) under recurrent conditions (h=30'). b) under non-recurrent conditions (h=30'). c) under recurrent conditions (h=90'). d) under non-recurrent conditions (h=90').

D. Comparison of performance and resilience to recurrent & non recurrent conditions

Previous ranking indicators are used to select their n "First Best" links as input variables of the prediction process and the prediction performance is assessed with MAPE in recurrent and non-recurrent conditions. According to figure 7, the proposed method LDA8 outperforms the others under the majority of traffic conditions and prediction horizons, especially when horizon gets higher. Even if some expert methods may reach equivalent performance under usual conditions, the difference becomes more significant under unusual congested traffic. The resilience of the LDA based methods to non-recurrent congestion is clearly highlighted. Furthermore, the previously observed hesitation between 2 topics (section IV-B) may explain the performance of LDA8 at the shorter horizon under usual conditions. Moreover, we can see the interest of refining the bins as the method LDA64 outperforms LDA8 under non recurrent conditions.

V. DISCUSSION

A feature selection process based on variables ranking and fitted for network-wide multidimensional prediction (MIMO)

has been introduced. In this specific case of multidimensional forecast, input variables are link states and the exposed strategy aims to rank these links according to their criticality. The method can be applied to any type of transport network. Moreover, the paper makes three contributions:

1) bringing out a definition of critical links highlighted by statistical concepts (LDA, but) based on physic assumption specifying that critical links impact networks mechanisms by catalyzing or propagating critical events, such as congestion;

2) introducing an innovative process based on LDA to generate ranking indicators for feature selection;

3) applying the method to a real-world application, road traffic prediction, using innovative data collection (GPS).

This work opens future research directions. The robustness of the process could be improved through

- an assessment of the sensitivity to other MIMO algorithms (Multi-SVR, ...) with shorter time-step (< 30 minutes) or

- an adaptation to real-time estimation of critical section, instead of operating by preset time-periods.

Furthermore, the definition of local configurations could be improved by substituting to the off-line cross-correlation some non-stationary statistical alternatives (mutual information, graphical LASSO [27]).

REFERENCES

[1] Pierre-Antoine Laharotte, Romain Billot, El Faouzi, Hesham A Rakha, et al. Network-wide traffic state prediction using bluetooth data. In *Transportation Research Board 94th Annual Meeting*, number 15-3022, 2015.

[2] Chi-Sen Li and Mu-Chen Chen. Identifying important variables for predicting travel time of freeway with non-recurrent congestion with neural networks. *Neural Computing and Applications*, 23(6):1611–1629, 2013.

[3] Pierre-Antoine Laharotte. *Contributions à la prévision court-terme, multi-échelle et multi-variée, par apprentissage statistique du trafic routier*. PhD thesis, 2016.

[4] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.

[5] Ron Kohavi. Wrappers for feature subset selection. 1996.

[6] Lei Yu and Huan Liu. Efficient feature selection via analysis of relevance and redundancy. *The Journal of Machine Learning Research*, 5:1205–1224, 2004.

[7] Fadi Dornaika, Elena Lazkano, and Basilio Sierra. Improving dynamic facial expression recognition with feature subset selection. *Pattern Recognition Letters*, 32(5):740–748, 2011.

[8] Pratik Jawanpuria, Manik Varma, and Saketha Nath. On p-norm path following in multiple kernel learning for non-linear feature selection. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 118–126, 2014.

[9] Lei Yu and Huan Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *ICML*, volume 3, pages 856–863, 2003.

[10] Feng Nan, Joseph Wang, and Venkatesh Saligrama. Feature-budgeted random forest. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 1983–1991, 2015.

[11] Huang Xiao, Battista Biggio, Gavin Brown, Giorgio Fumera, Claudia Eckert, and Fabio Roli. Is feature selection secure against training data poisoning? In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 1689–1698, 2015.

[12] Billy M Williams and Lester A Hoel. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of transportation engineering*, 129(6):664–672, 2003.

[13] Eleni I Vlahogianni, Matthew G Karlaftis, and John C Golias. Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. *Computer-Aided Civil and Infrastructure Engineering*, 22(5):317–325, 2007.

[14] Benjamin Hamner. Predicting travel times with context-dependent random forests by modeling local and aggregate traffic flow. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 1357–1359. IEEE, 2010.

[15] Eleni I Vlahogianni, Matthew G Karlaftis, and John C Golias. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43:3–19, 2014.

[16] Justin Dauwels, Aamer Aslam, Muhammad Tayyab Asif, Xinyue Zhao, Nikola Mitro Vie, Andrzej Cichocki, and Patrick Jaillet. Predicting traffic speed in urban transportation subnetworks for multiple horizons. In *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, pages 547–552. IEEE, 2014.

[17] Bei Pan, Ugur Demiryurek, and Cyrus Shahabi. Utilizing real-world transportation data for accurate traffic prediction. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 595–604. IEEE, 2012.

[18] Pinlong Cai, Yunpeng Wang, Guangquan Lu, Peng Chen, Chuan Ding, and Jianping Sun. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies*, 62:21–34, 2016.

[19] Aude Hofleitner, Ryan Herring, Pieter Abbeel, and Alexandre Bayen. Learning the dynamics of arterial traffic from probe data using a dynamic bayesian network. *Intelligent Transportation Systems, IEEE Transactions on*, 13(4):1679–1693, 2012.

[20] Andrew D Cliff and J Keith Ord. *The problem of spatial autocorrelation*. University, 1968.

[21] Wanli Min and Laura Wynter. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19(4):606–616, 2011.

[22] Xinyu Min, Jianming Hu, and Zuo Zhang. Urban traffic network modeling and short-term traffic flow forecasting based on gstarima model. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 1535–1540. IEEE, 2010.

[23] Tao Cheng, James Haworth, and Jiaqiu Wang. Spatio-temporal auto-correlation of road network data. *Journal of Geographical Systems*, 14(4):389–413, 2012.

[24] Fernando Pérez-Cruz, Gustavo Camps-Valls, Emilio Soria-Olivas, Juan José Pérez-Ruixo, Aníbal R Figueiras-Vidal, and Antonio Artés-Rodríguez. Multi-dimensional function approximation and regression estimation. In *Artificial Neural Networks—ICANN 2002*, pages 757–762. Springer, 2002.

[25] B. Ghosh, B. Basu, and M. O'Mahony. Multivariate short-term traffic flow forecasting using time-series analysis. 10(2):246–254.

[26] Srinivasa Ravi Chandra and Haitham Al-Deek. Predictions of freeway traffic speeds and volumes using vector autoregressive models. 13(2):53–72.

[27] James Haworth. *Spatio-temporal forecasting of network data*. PhD thesis, UCL (University College London), 2014.

[28] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, 2nd edition, 2000.

[29] D. Blei, M. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[30] Farrahi K. and D. Gatica-Perez. Discovering routines from large-scale human locations using probabilistic topic models. *ACM TIST*, 2(1):3–5, 2011.

[31] E. Come and L. Oukhellou. Model-based count series clustering for bike sharing system usage mining: A case study with the vélib'system of paris. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):39, 2014.

[32] Peter F Brown, Vincent J Della Pietra, Robert L Mercer, Stephen A Della Pietra, and Jennifer C Lai. An estimate of an upper bound for the entropy of english. *Computational Linguistics*, 18(1):31–40, 1992.

[33] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.