Identification of new patterns in urban traffic flows

Yingqi Gu[♯], Emanuele Crisostomi[♭], Mingming Liu[♯], Robert Shorten [♯]

***Abstract*— Traffic flow pattern identification, as well as anomaly detection, is known to be an important component for traffic operations and control. Alongside classical applications, mainly, to improve the safety and the comfort of drivers, more recently there is a growing interest in gathering personalised route information to provide customised services. With this latter application in mind, in this paper we investigate the ability of simple macroscopic information (i.e., time varying junction turning probabilities) to identify changes in nominal urban traffic flows, most likely due to the occurrence of external events (e.g., road works or traffic congestions). Some preliminary results obtained with the use of a realistic mobility simulator are also illustrated and discussed, and some candidate applications are briefly outlined.**

I. Introduction

The ability to know vehicular flows in urban contexts has a number of benefits, among others (i) it gives the possibility to predict traffic or pollution build-ups and take pre-emptive countermeasures in advance; (ii) it gives the possibility to plan the construction of new infrastructure to improve the driving experience (e.g., roundabouts, parking or charging stations, new roads, improved traffic lights); (iii) it gives the possibility to deliver customised advertisement.

In particular, some of such facilities can be particularly convenient when the knowledge of a specific vehicle driving along a given road is associated with the knowledge of the specific characteristics of the owner of the vehicle (e.g., to deliver customised services). Note that for this purpose simple detector loops, or other sensors that only count the number of vehicles that take a given road, are ineffective as they do not keep track of what specific vehicles are actually driving in the area.

An alternative approach to keep track of the history of trips of single vehicles is to use on-board electronic equipment to register the sequence of roads that is taken by a vehicle. Assuming that the drivers are then willing to share their own historical information with a central infrastructure, then it is possible to centrally aggregate all the data, that can be used to provide the aforementioned services.

The authors gratefully acknowledge funding for this research provided by Science Foundation Ireland under grant 11/PI/1177.

[♯]School of Electrical, Electronic and Communications Engineering, University College Dublin, Ireland {yingqi.gu@ucdconnect.ie}, {mingming.liu, robert.shorten}@ucd.ie

[♭]Department of Energy, Systems, Territory and Constructions Engineering, University of Pisa, Italy {emanuele.crisostomi}@unipi.it

The possibility to aggregate historical data of single drivers to create a database of trips (i.e., not only Origin-Destination (OD) matrices, but actually all the sequences of roads taken to get to the destination as well) poses a number of interesting problems to the transportation community, namely, (i) the ability to handle big data; (ii) privacy concerns; (iii) how classic transportation optimisation problems (e.g., routing problems) can be improved by the knowledge of such data; (iv) development of new customised services to fully exploit the available data.

In this paper we are interested in addressing one specific problem: how the available data should be cleaned to avoid biased information being included in the aggregated data. As an explanatory example, let us assume that one road is closed due to road works, or due to a car accident. Then vehicles are duly re-routed along other alternative roads. Such trips do not correspond to nominal patterns (i.e., under nominal regular circumstances the drivers would have taken different trips to reach their destination) and thus should not be considered and aggregated as they would introduce wrong information in the cumulative historical database.

This paper is organised as follows: the next section briefly overviews the state of the art in the subject of interest. Section III explains how we formulate the problem in order to identify when non-nominal patterns occur. Trips associated with non-nominal patterns can then be disregarded and not considered in the historical aggregated database. In Section IV we show some preliminary results that we have obtained in realistic simulations performed with the mobility simulator SUMO [1]. Some preliminary applications are described in Section V. Finally, in Section VI we conclude the paper and briefly outline how we plan to extend the work presented here.

II. State of the Art

Traffic flow pattern identification, as well as anomaly detection, is known to be an important component for traffic operations and control [2]. In paper [3] the authors identify and classify different traffic flow patterns at single intersections using a fuzzy c-means clustering algorithm to identify the different peaks of traffic during the same day. In [2] the authors used massive data-sets collected by traffic loop detectors in the urban network of Northern Virginia to identify recurrent traffic flow patterns. In this work, the authors also proved that there exists almost no correlation between spatial and temporal anomaly degrees, and that changes of land use and travel modes, due to major municipal constructions, do cause shifts in traffic flow patterns. Microscopic traffic variables, such as the relative speed, inter-vehicle time gap and lane changing, were used by [4] to infer, and classify, traffic anomalies; here the authors were interested in identifying the traffic anomalies that can lead to traffic incidents, also known as incident precursors. Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications may be also used to obtain the classification of anomalies in a distributed way [5]. A large database of truck road data, collected in several European countries by Volvo Group Trucks Technology was used in [6] with the ultimate goal again of preventing car accidents from occurring. Finally in [7] an interactive visual platform was developed to explore historical data and predict the future traffic. The platform was also equipped with the ability to visualise and cluster road behaviours, detect anomalies, predict traffic, explore the behavioral similarities between different roads, test hypotheses, and predict traffic flows after hypothetical incidents imposed by the human operator.

None of the papers previously mentioned addresses exactly the problem of our interest, i.e., identify when the sequence of roads taken to reach the destination of interest is a *nomimal* sequence, or is the consequence of some other event (e.g., road works, car accidents, usage of a car navigator, etc). Similar problems are however encountered in other contexts as well, e.g., control theory [8] or economics [9]. In this context a closely related paper is [10] where the authors developed mathematical tools to separate data-sets between the nominal data and the data affected by the exogenous actions of operators/controllers. More in detail in [10] the authors assumed that the observed time series was produced by a *regular* Markov chain, that corresponded to the nominal behaviour of the driver, which in turn was modulated by another *latent* random variable, produced itself by another Markov chain, to take into account the probability of external events coming into play (e.g., the use of a recommender system). In the end, the authors of [10] proposed a modified Baum-Welch algorithm to estimate the parameters of the Markov Modulated Markov Chain (MMMC) with a closed loop effect that seems to work well in practice, although formal justifications of the algorithm were not given. Also, the methodology was not tested for a transportation application. Our objective in this paper is to evaluate the ability of a very simple approach, namely, the monitoring of junction turning probabilities, to correctly identify the occurrence and the impact of some typical traffic anomalies.

III. Problem Formulation

In this paper we shall use macroscopic traffic variables, i.e. junction turning probabilities, to infer when anomalous patterns occur. If one considers a roundabout (but the same philosophy may be used for junctions regulated with signals or traffic lights as well) connected with four two-way roads, then junction turning probabilities may be expressed through a 4×4 matrix $\mathbf{M}$ whose rows correspond to the incoming roads and columns correspond to the outgoing roads. Here we briefly note that this approach can also be used to model other types of roundabouts, e.g., a roundabout connected with three two-way roads (T-intersection), which leads to a 3×3 matrix $\mathbf{M}$. Each entry $\mathbf{M}_{ij}$ of the matrix $\mathbf{M}$ denotes the probability to choose the j'th exit of the roundabout after having entered from the i'th entrance. With this in mind, it is not difficult to observe that matrix $\mathbf{M}$ is essentially a row-stochastic matrix (i.e. sum of each row of the matrix $\mathbf{M}$ equals 1). In practice, matrix $\mathbf{M}$ can be either estimated by using detector loops appropriately placed in the roundabout, or can be estimated by aggregating data of trips of several vehicles. However, once a matrix $\mathbf{M}$ is constructed, it can not be disaggregated to retrieve information of single vehicles. Also, note that so-built matrices $\mathbf{M}$ correspond to sub-matrices of the transition matrix in Markov-chain based models of urban traffic, as described in [11] and [12].

When external events occur for some reasons, local choices of drivers at roundabouts may change correspondingly as well. We shall evaluate such changes at every roundabout i in the area of interest by evaluating the local quantity

$$\left\|\mathbf{M}^i(k+\Delta T)-\mathbf{M}^i(k)\right\|_F, \quad \forall i=1,...,R \qquad (1)$$

at each time step k, where $\mathbf{M}^i(k)$ denotes the matrix of junction turning probabilities at the i'th roundabout, and R is the number of observed roundabouts in the area of interest. $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, and we have

$$\left\|\mathbf{M}^i(k)\right\|_F=\sqrt{\sum_{r=1}^{d_i}\sum_{c=1}^{d_i}\mathbf{M}^i_{rc}(k)^2}, \qquad (2)$$

where d_i denotes the number of incoming or outgoing roads connected to roundabout i. The choice of ΔT represents the interval of time used to identify changes in the patterns. Small values of ΔT allow one to quickly detect abrupt changes in mobility (e.g., the city centre may be closed at a specific time), but events occurring at a slower time scale may be regarded as noise. On the other hand, higher values of ΔT allow one to better appreciate events occurring at a slower time scale, while instantaneous changes may be noticed with some delay. Moreover, it may be convenient in some cases to consider the metric (i.e. equation (1)) in the whole area by summing the contributions of each local roundabout i as follows:

$$\sum_{i=1}^{R}\left\|\mathbf{M}^i(k+\Delta T)-\mathbf{M}^i(k)\right\|_F. \qquad (3)$$

The new indicator provides a single curve (i.e., as a function of the time step k) that allows one to detect anomalies in a wide area by simple visual inspection. However, anomalies in this pattern do not allow one to infer if all roundabouts, or only a subset of them, have seen different junction turning probabilities occurring. In this case, the local indicator (1) may be used instead, as will be illustrated in the following section.

IV. SUMO Simulations

A. Simulation Set-up

In this section, we evaluate the performance of the proposed method in a realistic traffic scenario. The vehicular flows are simulated using the popular mobility simulator SUMO [1]. All the simulations are performed over the road network of a small area in Dublin city, Ireland, depicted in Figure 1, imported from OpenStreetMap [13].

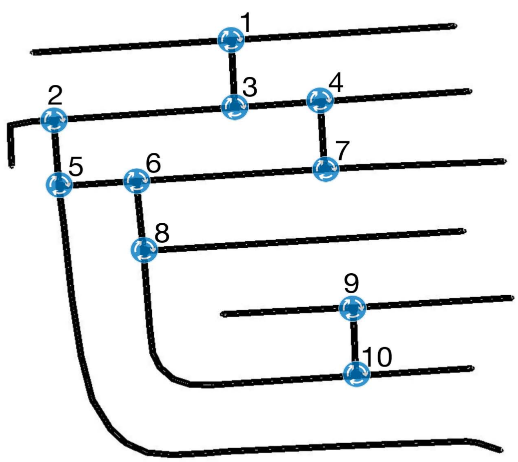

Fig. 1. Road network of a small area in Dublin city, Ireland. The network used in the SUMO simulation is imported from OpenStreetMap. All roundabouts in the network are marked with blue roundabout symbols and have been denoted with different ID numbers reported in the map.

We assume that there are 160 vehicles with personalised origin and destination pairs travelling in the network during a period 24 hours. We assume that vehicles may choose different routing strategies to get to their destination, e.g., shortest path [14], [15] or minimum expected travelling time [16], [17]. In our simulations, we always considered one time step equal to one second, and chose ΔT equal to 100 seconds. In the following sections we investigate the ability of the proposed methodology to detect the arising of different anomalies in the nominal driving patterns of the vehicles.

B. Road access control

In some circumstances the number of vehicles entering a given road, or area, may be controlled. A classic example is to control the number of vehicles entering the city centre to mitigate polluting emissions. We simulate the occurrence of such a situation by limiting the number of vehicles that are allowed to take a given exit at a roundabout (i.e., this corresponds to assuming that only a fraction of vehicles, for instance low-polluting vehicles, have the permission to enter a given area at a given time). Note also that access control may be used in other circumstances as well, for instance due to road maintenance or as a consequence of a car accident. Accordingly, in the first simulation we assume that all roundabouts operate normally in the first twelve hours. During this interval of time the system learns nominal patterns of vehicles. Then we assume that 20% of the vehicles (randomly chosen) is forced not to take a given exit road at Roundabout 3 between hour 12 and hour 15. Similarly, an outgoing road at Roundabout 6 is closed to 60 % of the vehicles between hour 16 and hour 18.

Simulation results are shown in Figure 2 and Figure 3. Figure 2 shows a single curve that aggregates the effects of all roundabouts. Both anomalies occurring at Roundabout 3 and 6 are clearly detected, but the second one is obviously more evident as more vehicles are involved in the detours. Figure 3 distinguishes the effects of single roundabouts, so one can clearly identify which specific roundabout has been affected by a road access control.

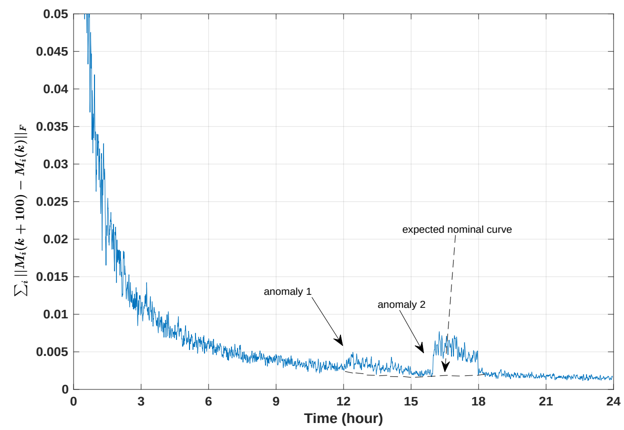

Fig. 2. Aggregated effect when anomalies occur. Effects are more evident when anomaly 2 occurs as more vehicles are detoured from their nominal patterns.

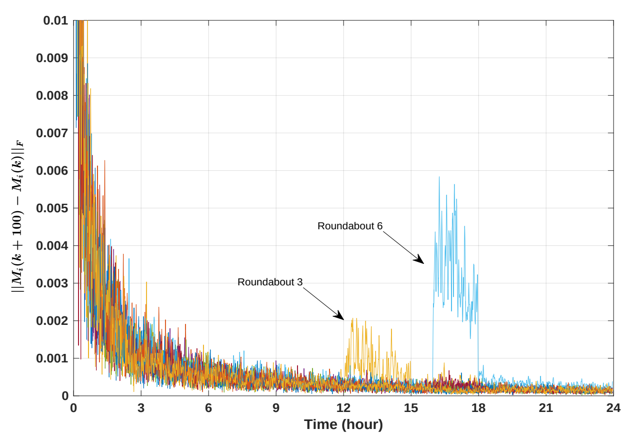

Fig. 3. Aggregated detection effect of all roundabouts in the first scenario.

C. Road access control and minimum time routing

In the new simulation, similarly from before, we assume that one exit road of Roundabout 3 is closed between hour 12 and hour 15, this time to 40 % of the vehicles. Between hour 16 and hour 18, we assume that all vehicles change their routing strategy to the minimum time routing path (i.e., this is automatically estimated in SUMO by using the travel time information along each edge). Note that this does imply that all vehicles change their path, as in practice the minimum time path may coincide with their nominal choice anyway. This event simulates the possibility that vehicles are recommended to follow a car navigator (e.g., to respond in real-time to external events, like traffic build-ups). Here we consider a "minimum time path" routing strategy, but in principle any other routing choice may be considered. The important aspect is that we wish to detect the fact that after hour 15 some vehicles will modify their usual paths. Results are now given in Figure 4, and both events are correctly captured in the figure. Furthermore, it is possible to note that the first anomaly involves a single roundabout (correctly, Roundabout 3), while the single anomaly involves all roundabouts.

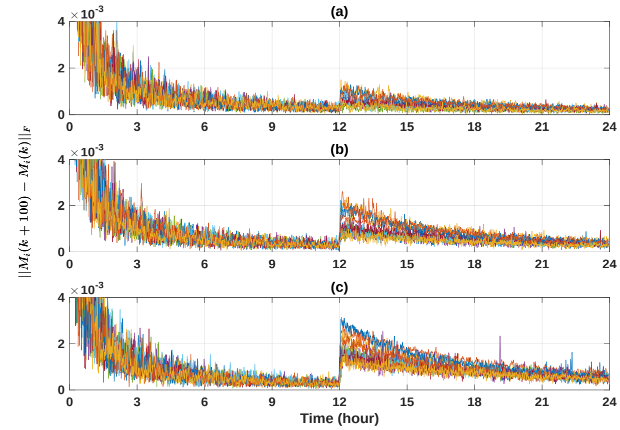

Fig. 5. All roundabouts detect that some changes in the nominal patterns occur after 12 hours. These are more evident when a larger percentage of vehicles follow the new recommendation (figure below).

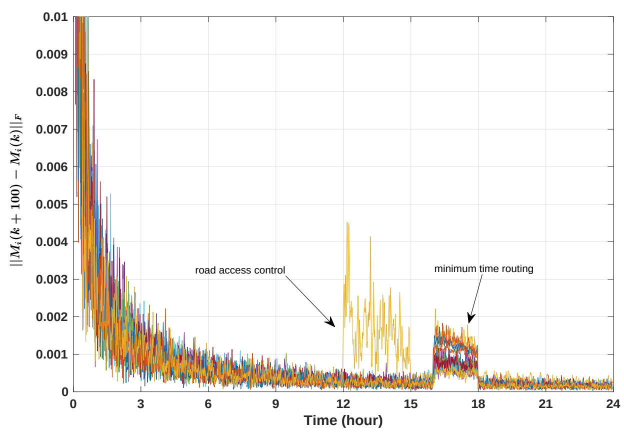

Fig. 4. In the second simulation, only Roundabout 3 correctly detects a local effect, while all roundabouts recognise a global effect (i.e., the use of a possibly different routing strategy).

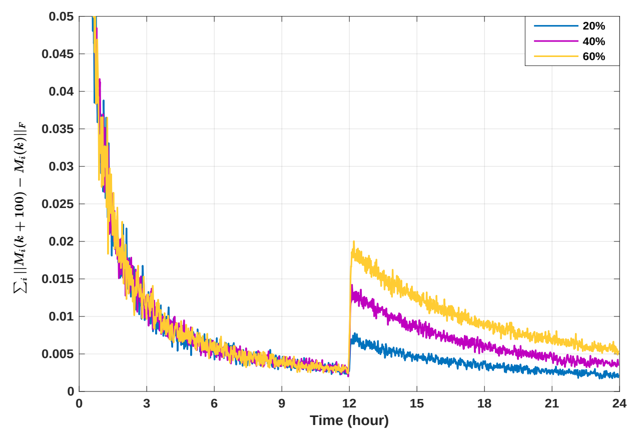

Fig. 6. Aggregated detection effect of all roundabouts when different percentages of vehicles change their routing path from the nominal one.

D. Minimum time routing

In our last simulation, we assume that at hour 12 a different percentage of vehicles (i.e., 20%, 40% and 60%) start following a minimum time routing strategy. Again, this corresponds to assuming that vehicles are recommended to follow a car navigator to respond to some external events, but only a fraction of vehicles follows the recommendation while the others neglect the recommendation. Sub-plots (a), (b) and (c) in Figure 5 show that all roundabouts detect the new behaviour of the vehicles in the three scenarios respectively. Obviously, the anomalous behaviour is more recognisable when more vehicles follow the recommendation as seen in sub-plot (c). Figure 6 shows the aggregate effect in the three scenarios. In both figures it is possible to see that at steady-state the effect of the anomaly tends to disappear.

V. Applications

In this section, we give examples and outline several directions where our method can be useful in real applications.

A. Identification of Driver's Travelling Intentions

As we have briefly mentioned, the ultimate objective of our approach is to gain the ability to distinguish nominal and induced (i.e., by external events) patterns of drivers in urban mobility. In this subsection, we shall show that the nominal patterns of the users may be significantly different from those that would have been observed without pre-filtering the data to remove trips taken in anomalous situations. For this purpose we simulate a vehicle driving in the same urban network depicted in Figure 1 for two hours per day, on average, for 70 days. We further assume that one road accessible through Roundabout 3 gets actually closed to traffic (for instance, to simulate that due to pollution build-up some roads in the city centre are closed to conventional traffic). This happens for some hours during the simulation, up to 40 % of the overall time of the simulation (in all cases, connectivity of the graph is preserved, as it is still possible to reach each node in the graph taking different

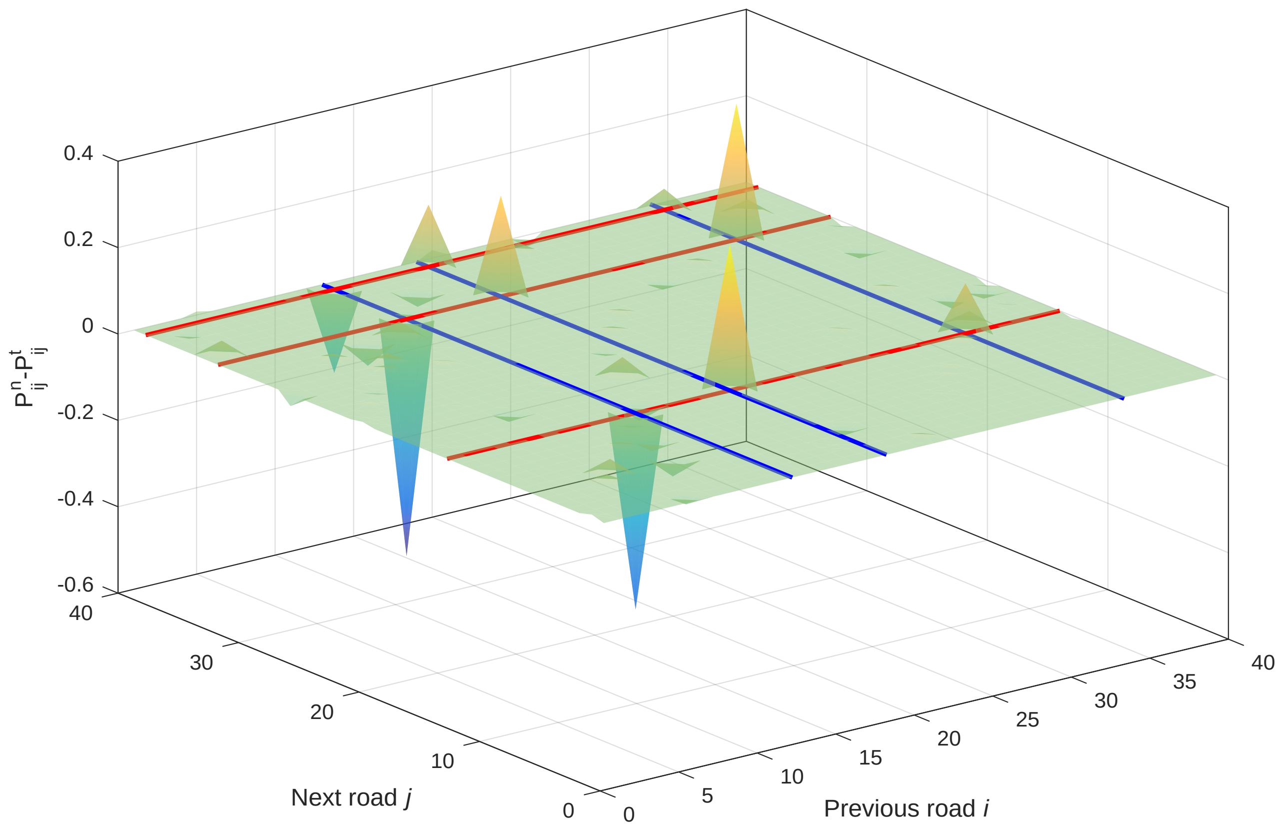

Fig. 7. The transition matrices of the nominal behaviours of single drivers and the one affected by outer events may be quite different.

paths). In the cases that such a road was indeed included in the nominal route, then the vehicle has to re-route along a different path. Figure 7 shows that the nominal behaviour of the driver would be rather different from the one that is inferred by averaging the choices of the driver during the whole observed time.

More in detail, we are now comparing the nominal matrix $\mathbf{P}^n$, whose entries $\mathbf{P}^n_{ij}$ represent the true probabilities with which a driver would turn into road j at the end of road i, with the matrix $\mathbf{P}^t$ where probabilities are computed over the total time of observation (i.e., also including when one road was closed to traffic). In the first case, when the system detects a non-nominal situation, then some routing data are filtered out, and not aggregated to estimate the nominal behaviours. The so-built matrices are row-stochastic matrices, and correspond to the Markov chain matrices of single vehicles as defined in [12] after setting to zero the diagonal entries (i.e., neglecting travel times along roads). In Figure 7 the red lines refer to the incoming roads of Roundabout 3, while the blue lines are the outgoing roads of the same Roundabout. As can be seen in the figure, if the two matrices are subtracted entry-wisely, then only the entries involved in the roundabout (i.e., at the intersections between the red and the blue line) are characterised by a value that is clearly different from zero. On the other hand, the other entries of the nominal and total matrices practically coincide.

B. Other applications

Growing concerns regarding a sustainable mobility infrastructure together with an ever-increasing level of connectivity among vehicles are reshaping current mobility paradigms. In this framework, several applications may benefit of the ideas that we described previously, particularly including:

Customised services: Many companies are now starting to provide customised services to drivers. For this purpose the historical travelling data of vehicles are used to infer the characteristics and typical patterns of single vehicles. Little attention is however paid to distinguish through patterns from drivers' choices that are a consequence of external events (e.g., traffic rerouting from municipalities for whatever reasons).

Combating traffic "fake news": As news from single drivers are getting embedded in real-time and off-line routing advisers, there is a growing concern among routing companies that people are starting providing some wrong news (e.g., regarding car accidents or regular traffic build-ups, see [18], [19]) to influence the way routing recommendations are delivered (e.g., to prevent the roads where they live from being included in typical recommended routes).

Identification of the use of car navigators: The plethora of currently existing car navigators, often not installed within the vehicle but used through a smart-phone, makes it challenging to infer what vehicles, and when, are in fact using a car navigator. By aggregating data it would become possible to cross-correlate data to identify when they are used (e.g., to communicate the availability of ancillary services via app).

While the previous applications, and many similar others, are starting getting popular, it is our opinion that little care has been posed to so far in separating the effects of the nominal behaviours of drivers and those induced by recommender systems or municipalities in response to the behaviours of the drivers themselves. The monitoring of some macroscopic quantities, like junction turning probabilities, may be of help as illustrated in this paper.

VI. Conclusions

This paper investigates the ability of simple macroscopic traffic information, namely, junction turning probabilities, to infer the occurrence of anomalies in traffic patterns. Such changes may be due to several external events, e.g., road works, impact of traffic congestion on navigator systems, car accidents, that affect individual routes and in turn introduce a bias in nominal patterns. More in general, while many companies are getting interested in developing customised services to connected drivers, little attention has been posed so far in trying to separate the effects of the nominal behaviours of the drivers, and those induced by recommender systems, or municipalities, in response to the behaviours of the drivers themselves. This paper attempts to take a first step in this direction, and a popular mobility simulator was used to validate some preliminary results in a realistic fashion.

This work only represents a preliminary step towards an automatic procedure to process data from individual vehicles to autonomously split traffic patterns into nominal and anomalous patterns, where the second ones should not be used for aggregation purposes in this context. The ability to tune the parameter ΔT may be used to evaluate different patterns evolving at different time scales. However, the procedure can not be applied in a complete unsupervised fashion as it is, while it may be convenient to further include other macroscopic features, in addition to junction turning probabilities, to infer more complete information; for instance, to infer when an anomalous situation finishes and nominal patterns are restored. In addition, we are interested in including the possibility to recognise the specific anomalies that have occurred, and to correlate (both in time and space) the occurring induced patterns among different vehicles as well. Also correlations among different junctions have not been considered here either. Finally, other applications along the same lines of research may be developed as well, as briefly outlined in V-B.

References

[1] SUMO, Simulation of Urban MObility. Last accessed: January 1, 2018. [Online]. Available: http://www.sumo-sim.org

[2] Z. Zhenhua, Q. He, H. Tong, J. Gou, and X. Li, "Spatial-temporal traffic flow pattern identification and anomaly detection with dictionary-based compression theory in a large-scale urban network," *Transportation Research Part C*, vol. 71, pp. 284–302, 2016.

[3] Y. Zhu, N. Gao, J. Wng, and C. Liu, "Study on traffic flow patterns identification of single intersection intelligent signal control," *Procedia Engineering*, vol. 137, pp. 452–460, 2016.

[4] J. A. Barria and S. Thajchayapong, "Detection and classification of traffic anomalies using microscopic traffic variables," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 695–701, 2011.

[5] S. Thajchayapong, E. S. Garcia-Treviño, and J. A. Barria, "Distributed classification of traffic anomalies using microscopic traffic variables," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 448–458, 2013.

[6] M. Riveiro, M. Lebram, and M. ELmer, "Anomaly detection for road traffic: a visual analytics framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 2260–2270, 2017.

[7] I. Kalamaras, A. Zamichos, A. Salamanis, A. Drosou, D. D. Kehagias, G. Margaritis, S. Papadopoulos, and D. Tzovaras, "An interactive visual analytics platform for smart intelligent transportation systems management," *IEEE Transactions on Intelligent Transportation Systems*, In Press.

[8] L. Ljung, *System Identification: Theory for the User*. Pearson Education, 1998.

[9] H. R. Varian, "Causal inference in economics and marketing," *PNAS*, vol. 113, no. 27, pp. 7310–7315, 2016.

[10] J. Epperlein, R. Shorten, and S. Zhuk, "Learning Markov models from closed loop data-sets," *arXiv*, pp. 1–17, 2017.

[11] E. Crisostomi, S. Kirkland, and R. Shorten, "A Google-like model model of road network dynamics and its application to regulation and control," *International Journal of Control*, vol. 84, no. 3, pp. 633–651, 2011.

[12] A. Schlote, E. Crisostomi, S. Kirkland, and R. Shorten, "Traffic modelling framework for electric vehicles," *International Journal of Control*, vol. 85, no. 7, pp. 880–897, 2012.

[13] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *Pervasive Computing, IEEE*, vol. 7, no. 4, pp. 12–18, 2008.

[14] N. Deo and C.-Y. Pang, "Shortest-path algorithms: Taxonomy and annotation," *Networks*, vol. 14, no. 2, pp. 275–323, 1984.

[15] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "Sumo–simulation of urban mobility: an overview," in *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind, 2011.

[16] W. Dong, "An overview of in-vehicle route guidance system," in *Australasian Transport Research Forum*, vol. 2011. Citeseer, 2011.

[17] V. T. N. Nha, S. Djahel, and J. Murphy, "A comparative study of vehicles' routing algorithms for route planning in smart cities," in *Vehicular Traffic Management for Smart Cities (VTM), 2012 First International Workshop on*. IEEE, 2012, pp. 1–6.

[18] K. D. Atherton, "Israeli students spoof waze app with fake traffic jam, available online at url: https://www.popsci.com/article/gadgets/israeli-students-spoof-waze-app-fake-traffic-jam," 2014.

[19] C. Cimpanu, "Waze drivers can be tracked, network flooded with fake traffic, available online at url: http://news.softpedia.com/news/waze-drivers-can-be-tracked-network-flooded-with-fake-traffic-503473.shtml," 2016.