

# Emergence of Global Speed Patterns in a Traffic Scenario<sup>\*</sup>

Richard Holzer<sup>1</sup>, Hermann de Meer<sup>1</sup>, and Cristina Beltran Ruiz<sup>2</sup>

<sup>1</sup> Faculty of Informatics and Mathematics, University of Passau,  
Innstraße 43, 94032 Passau, Germany  
{holzer,demeer}@fim.uni-passau.de

<sup>2</sup> SICE, Sepulveda 6, E-28108 Alcobendas, Madrid, Spain  
cbeltran@sice.com

**Abstract.** We investigate different analysis methods for traffic data. The measure for emergence can be used to identify global dependencies in data sets. The measure for target orientation can be used to identify dangerous situations in traffic. We apply these measures in a use case on a data set of the M30 highway in Madrid. The evaluation shows that the measures can be used to predict or to identify abnormal events like accidents in traffic by an evaluation of velocity data or flow data measured by detectors at the road. Such events leads to a decrease of the measures of emergence and target orientation.

**Keywords:** Quantitative Measures, Emergence, Target Orientation, Traffic Safety.

## 1 Introduction

In traffic, abnormal events like accidents on highways may cause a sudden drop in traffic flow, which might lead to dangerous situations, that may result in follow-on accidents [1]. To increase safety in traffic, Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication technologies could be used to provide other drivers with the necessary information or to give a direct advice (e.g. “slow down”) to avoid further accidents. In [2] some V2V systems have been proposed to achieve this goal. To be able to advice drivers to a specified behavior, an evaluation of the current situation is necessary: How can a dangerous situation be predicted or at least identified by using measurements of data? Measurements of the parameters of the vehicles (e.g. position or speed) can either be done by using a system inside of the vehicles or by a system in the infrastructure. Although a system in the infrastructure has no access to the whole trajectory of a vehicle, a set of sensors at the road can measure the parameters of the vehicles passing the sensors. If we consider a set of such detectors at different positions,

---

<sup>\*</sup> This research is partially supported by the SOCIONICAL project (FP7, ICT-2007-3-231288), by the Network of Excellence EINS (FP7, ICT-2011-7-288021) and by the Network of Excellence EuroNF (FP7, ICT-2007-1-216366).

the data of all detectors might be sufficient to identify abnormal events (e.g. an accident) in the traffic flow. For this purpose, measures are needed to quantify the safety relevant properties of the current state.

In the recent years, some quantitative measures for properties in self-organizing systems have been developed [3], [4], [5]:

- Autonomy
- Emergence
- Global state awareness
- Target orientation
- Resilience
- Adaptivity

In a traffic scenario, some of these measures can be used for the analysis of safety relevant properties in the system: By specifying the overall goal of the system in form of a fitness function, the measures for target orientation, resilience and adaptivity can be used in different contexts to evaluate the average fitness of the system and to identify abnormal events like accidents. Also the measure of emergence might be useful for the interpretation of the measured data: A high emergence of velocities indicates high dependencies between the velocities of different vehicles, while a low emergence of velocities indicates few dependencies, which can be interpreted as a disharmony in the traffic flow.

After the identification of the presence of an abnormal event, the harmonization of the traffic flow (see [6] and [2]) can be increased either by overhead speed signs or direct signals between the vehicles (in case of V2V communication).

In this paper we analyze a set of 349920 data collected and received from detectors at the highway M30 in Madrid. M30 highway is controlled by an Intelligent Transport System (ITS) implemented as a four-level control architecture that comprises two control centers (main and backup) at the top level, the intermediate level of communications, the distributed intelligence level and field equipment at the bottom level; the latter is responsible of continuously measuring different traffic values and of producing output values at different time intervals, normally minutes.

For this study, access to a specific set of registered data measured by the ITS system implemented in M30 has been granted between SICE as co-author of the paper and Madrid City Council as owner of the data. The collected dataset used in this study include output values of velocity, intensity and occupancy measured by minute at each lane of a road segment of around 12 km of the whole highway. Those data are subject to IPR protection and will not be reproduced in this paper.

The measure of emergence [3] applied to the velocity data can be used to measure dependencies between the velocities of different cars. Analogously the emergence measure can also be applied to flow data (temporal density) to detect dependencies in flow data. The measure for target orientation [4] shows, whether the overall goal (traffic safety) is satisfied. The goal is formally specified by using the variance coefficient of velocities (see [7] and [2]). Section 2 specifies the micro-level model of systems based on [3]. Sections 3 and 4 specify the methodology for the evaluation

of emergence and target orientation in systems. In Section 5 these measures are applied to a data set of the M30 highway in Madrid. In Section 6 the results of the evaluation are discussed and Section 7 concludes the paper.

## 2 Discrete Micro-level Model of Systems

We consider a discrete system  $\mathcal{S}$  with many entities and interactions between the entities. If the topology of such a system is static then it can be represented [3] by a directed graph  $G = (V, E)$ , where the nodes of the graph represent the entities and the edges represent the communication channels. For dynamic topologies a time dependent directed graph can be used. Each entity has an internal state, which might change during the time. The communication channels between the entities can be used for interaction between the entities, i.e. an entity can send some data to some other entities, which may have some influence on their internal states. For the set of all points in time, when a change of the system occurs (e.g. a change in the state of an entity or an interaction between entities), we use  $T = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ .

Let  $Conf$  be the set of all global configurations of  $\mathcal{S}$ , i.e. each  $c = (c_V, c_E) \in Conf$  is a pair, where  $c_V = (c_v)_{v \in V}$  contains the current internal state  $c_v$  for each entity  $v \in V$  and  $c_E = (c_e)_{e \in E}$  contains the current value  $c_e$  on the communication channel for each  $e \in E$ . The behavior of the entities (change of the state and interaction) might be deterministic or stochastic. Also the initialization of the system at time  $t = 0$  might be deterministic or stochastic. For  $t \in T$  let  $Conf_t$  be the random variable for the configuration at time  $t$  taking values in  $Conf$ . A trajectory of the system is a realization  $(c_0, c_1, c_2, \dots) \in Conf^T$  of the stochastic process  $(Conf_t)_{t \in T}$ . The trajectory describes the sequence of configurations during a run of the system: At each time  $t \in T$  the system is in a configuration  $c_t \in Conf$ , and in the next step the changes in the entities and in the communication channels lead to the next configuration  $c_{t+1} \in Conf$  at time  $t + 1$ .

## 3 Emergence

In [3] the measure of emergence at each point in time is defined by using the entropies of the communication channels between the entities. Here we generalize this concept to an arbitrary family of measures defined on the set of global states.

For a discrete random variable  $X : \Omega \rightarrow W$  on a probability space  $\Omega$  into a set  $W$  of values the entropy [8] is defined by

$$H(X) = - \sum_{w \in W} P(X = w) \log_2 P(X = w)$$

with  $0 \cdot \log_2 0 := 0$ . This entropy measures the average number of bits needed to encode the outcome of the random variable  $X$  in an optimal way. If  $X$  has a uniform distribution on a finite set  $W$ , then we get  $H(X) = \log_2(|W|)$ . For

other distributions of  $X$  this value is an upper bound, i.e. we always have  $H(X) \leq \log_2(|W|)$ . The entropy measure can also be applied on tuples of random variables:

$$H(X, Y) = - \sum_{w \in W, w' \in W'} P(X = w, Y = w') \log_2 P(X = w, Y = w')$$

The entropy of a pair  $(X, Y)$  of random variables is not greater than the sum of the entropies of the random variables:

$$H(X, Y) \leq H(X) + H(Y)$$

To formally define the level of emergence of the system, we have to specify the properties of interest. These properties of interest can be specified by a family of maps. Let  $J$  be an at most countable set and  $m = (m_j)_{j \in J}$  be a family of maps

$$m_j : Conf \rightarrow W_j$$

for a value set  $W_j$  for  $j \in J$ . For  $c \in Conf$  let

$$m(c) = (m_j(c))_{j \in J} \in \prod_{j \in J} W_j$$

The level of emergence at time  $t \in T$  is defined by

$$\varepsilon_t = 1 - \frac{H(m(Conf_t))}{\sum_{j \in J} H(m_j(Conf_t))}$$

Note that we have

$$\varepsilon_t \in [0, 1]$$

because of

$$H(m(Conf_t)) \leq \sum_{j \in J} H(m_j(Conf_t))$$

Therefore we get a map

$$\varepsilon : T \rightarrow [0, 1] \quad \text{with} \quad t \mapsto \varepsilon_t$$

Note that the definition of level of emergence in [3] is a special case of this definition: We can choose  $J = E$  as the set of all communication channels and  $m_e(c_V, c_E) = c_e$  as the value on the channel  $e \in E$ . While [3] only considers emergence in communication, the definition above can be used in a much broader sense. First, the measures  $m = (m_j)_{j \in J}$  have to be specified to describe, in which type of data we would like to identify emergent patterns. Then the global emerging patterns are measured by comparing the entropy of the specified measures  $m$  in the current configuration with the sum of the entropies

$H(m_j(Conf_t))$  for  $j \in J$ . If the values of the measures  $m_j$  are independent, then the whole entropy is identical to the sum of the entropies:

$$H(m(Conf_t)) = \sum_{j \in J} H(m_j(Conf_t))$$

In this case we get  $\varepsilon_t = 0$ . The more dependencies are between the measures  $m_j$ , the smaller is the global entropy  $H(m(Conf_t))$ , i.e. the quotient

$$\frac{H(m(Conf_t))}{\sum_{j \in J} H(m_j(Conf_t))}$$

decreases with higher dependencies. Therefore the value  $\varepsilon_t$  indicates the existence of dependencies (which can be seen as global patterns) in the values of the measures  $m_j$ : If the emergence  $\varepsilon_t$  is near 0, then all measures  $m_j$  are nearly independent at the current point in time  $t \in T$ . If the emergence  $\varepsilon_t$  is near 1, then there are many dependencies between the measures at the current point in time.

An analytical evaluation of the level of emergence in a complex system is usually impossible because of the huge state space  $Conf$ . Therefore we need approximation methods [9] to be able to evaluate at least an approximation of the emergence of a system.

To transfer the definition of the entropy  $H(X)$  of a random variable  $X$  to a sequence  $(X(\omega_1), \dots, X(\omega_n))$  of realizations of the random variable  $X$ , we use relative frequencies of the realizations for the probabilities. This leads to the following definition: For a tuple  $x = (x_1, \dots, x_n) \in W^n$  the relative frequency of a value  $w \in W$  in  $x$  is defined by

$$rel_w(x) := \frac{|\{i \leq n : x_i = w\}|}{n}$$

The entropy  $H(x)$  of  $x$  is defined by

$$H(x) = - \sum_{w \in W} rel_w(x) \log_2 rel_w(x)$$

This definition can be extended to matrices<sup>1</sup>  $A = (A_{ij})_{i \leq n, j \leq r} \in W^{n \times r}$  by identifying  $A$  with the sequence of rows of  $A$ . For  $i \leq n$  and  $j \leq r$  let  $A_i. \in W^r$  be the  $i$ -th row of  $A$  and let  $A_.j \in W^n$  be the  $j$ -th column of  $A$ . The relative frequency of a tuple  $w = (w_1, \dots, w_r) \in W^r$  in  $A$  is

$$rel_w(A) := \frac{|\{i \leq n : A_i. = w\}|}{n}$$

---

<sup>1</sup> It is also possible to use “inhomogeneous” matrices, where the entries of different columns come from different value sets  $W_j$ . In this case, the matrix  $A$  can be seen either as an element of  $\prod_{j \leq r} W_j^n$  (sequence of columns) or as an element of  $\left( \prod_{j \leq r} W_j \right)^n$  (sequence of rows).

The entropy of  $A$  is defined by

$$H(A) = - \sum_{w \in W^r} rel_w(A) \log_2 rel_w(A)$$

The emergence of  $A$  is defined by

$$\varepsilon(A) = 1 - \frac{H(A)}{\sum_{j \leq r} H(A_{.j})}$$

Now we can use these concepts to approximate the level of emergence  $\varepsilon_t$  of a large system at time  $t \in T$  with respect to a given family  $m = (m_j)_{j \in J}$  of measures with a finite set  $J$ . For this purpose we need some realizations of the stochastic process  $(Conf_t)_{t \in T}$ . Such realizations can either be received from simulations or from measured real data. Let  $n$  be the number of realizations and for  $1 \leq i \leq n$  let  $Conf_{t,i} \in Conf$  be the global state at time  $t \in T$  for the  $i$ -th realization. For each  $t \in T$  we can define the  $n \times |J|$  matrix  $A = (A_{ij})_{i \leq n, j \in J}$  by  $A_{ij} = m_j(Conf_{t,i})$ . Then we can observe:

- For  $i \leq n$  the  $i$ -th row of this matrix is the  $i$ -th realization of  $m(Conf_t)$ :

$$A_{i.} = m(Conf_{t,i})$$

- For  $j \in J$  the  $j$ -th column of  $A$  consists of all realizations of  $m_j(Conf_t)$ :

$$A_{.j} = (m_j(Conf_{t,i}))_{i \leq n}$$

Then we can use  $\varepsilon(A)$  as an approximation of  $\varepsilon_t$ : The relative frequencies of the values  $w = m_j(Conf_{t,i})$  correspond to the probabilities  $P(m_j(Conf_t) = w)$  and the relative frequencies of the tuples

$$w = (w_1, \dots, w_{|J|}) = m(Conf_{t,i}) = A_{i.}$$

correspond to the probabilities  $P(m(Conf_t) = w)$ . In Section 5 this approximation  $\varepsilon_t \approx \varepsilon(A)$  will be used for the evaluation of an approximation of the emergence in a traffic scenario by using real data.

## 4 Target Orientation

The goal of a system [4] can be specified by a fitness function  $f : Conf \rightarrow [0, 1]$  on the global state space. The fitness function describes, which configurations are good ( $f(c) \approx 1$ ) and which configurations are bad ( $f(c) \approx 0$ ). At each point in time  $t \in T$  this fitness function can be applied to  $Conf_t$  to measure the fitness of the system at time  $t$ . Since  $Conf_t$  is a random variable,  $f(Conf_t)$  is also a random variable. The level of target orientation of the system at time  $t \in T$  is the mean value of this random variable:

$$TO_t = E(f(Conf_t))$$

This induces a map

$$TO : T \rightarrow [0, 1] \quad \text{with} \quad t \mapsto TO_t$$

As for the measure for the emergence, the measure of target orientation usually can not be calculated analytically in large systems, so we need again realizations (either by simulations or real data) for the stochastic process  $(Conf_t)_{t \in T}$ . Let  $n$  be the number of realizations,  $t \in T$  and for  $1 \leq i \leq n$  let  $Conf_{t,i}$  be the  $i$ -th realization of the random variable  $Conf_t$ . Then the mean value  $E(f(Conf_t))$  of the fitness function can be approximated by the average fitness of the realizations:

$$TO_t \approx \frac{1}{n} \sum_{i=1}^n f(Conf_{t,i})$$

The accuracy of this approximation increases with the number  $n$  of realizations, but in some cases it is not possible to produce many realizations of the random variable  $Conf_t$ . For example, if the measure of target orientation is applied to real data measured in a given system, then we have only one realization of  $Conf_t$  for each point in time  $t$ . While for the measure of emergence a single realization is not sufficient to calculate an approximation of  $\varepsilon_t$ , this is not so critical for the level of target orientation: The target orientation  $TO : T \rightarrow [0, 1]$  can be approximated by applying the fitness function to the realization:

$$TO_t \approx f(Conf_{t,1})$$

For the emergence we will see in the next section, that aggregation techniques can help to overcome this problem.

## 5 Evaluation of Highway Data

In this section we evaluate data measured at the highway M30 in Madrid. We use 81 detectors of the whole M30 road network, each of them located in one road lane, to collect minute by minute measures of the velocity of cars passing the detectors during one working day. Table 1 shows the positions in *km* (relative to a reference point) of detectors at 24 different positions at the M30 highway. Let  $P$  be the set of these 24 positions. For each position  $p \in P$  the second row of Table 1 shows the number of lanes  $L_p$ . For each lane  $l \in \{1, 2, \dots, L_p\}$  a detector at position  $p$  measures some data for lane  $l$ . Once per minute, each detector produces some output values. The output of the detector is (among others)

- the number of vehicles, which have passed the detector at position  $p$  in this lane  $l$  in a time interval of length *1min*, and
- the average value of the measured velocities of these vehicles, when they passed the detector in the time interval.

**Table 1.** Positions of detectors in *km* (relative to a reference point) and number of lanes

Position (km)	0.0	0.2	0.9	1.4	2.1	2.4	3.4	4.0	4.4	4.7	5.0	5.6	6.1	6.6	7.1	8.3	8.6	9.0	9.4	9.8	10.7	11.0	11.6	12.0
Lanes	3	3	3	3	3	3	3	4	3	3	3	3	4	3	4	5	5	3	3	3	5	2	3	4

We analyze the emergence and the target orientation of the measured data of one day (=1440 minutes), where an accident has occurred on the highway at time  $t = 1170min$  (which is 7:30pm on that day) near position  $p = 6.1$ . For each detector position  $p \in P$  and each lane  $l \leq L_p$  we have some output values of the detectors for the time intervals  $I_t := [t, t + 1[$  of length  $1min$  for  $t \in T' := \{0, 1, \dots, 1439\}$ . Let  $N_{p,l,t}$  the number of vehicles passing the detector in the time interval  $I_t$  and let  $v_{p,l,t}$  be the averaged velocity values of these vehicles measured at the detector.

For our micro-level model we can use the set of all detectors

$$D = \{d_{p,l} \mid p \in P, l \leq L_p\}$$

as a subset of all entities:  $D \subseteq V$ . Other entities are the vehicles on the road. The interactions between the vehicles and the detectors cause a change of the internal states of the detectors  $d \in D$ . In a configuration  $c \in Conf$  each detector  $d_{p,l} \in D$  contains two values in its internal state:  $N_{p,l,c}$  is the number of vehicles detected and  $v_{p,l,c}$  is the average measured velocity of these vehicles. We assume a finite accuracy for the detectors, so the set of all possible outputs of a detector  $d_{p,l} \in D$  is finite:

$$|\{(N_{p,l,c}, v_{p,l,c}) : c \in Conf\}| < \infty \quad \text{for } p \in P, \quad l \leq L_p$$

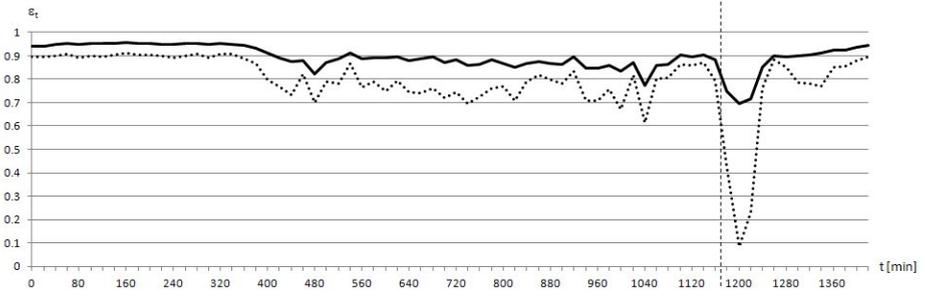
## 5.1 Emergence of the Data Set

To be able to evaluate the emergence of the data in the traffic scenario, we now have the problem, that we do not have many realizations of the random variable  $Conf_t$ , so the methodology of section 3 can not be applied directly. But an aggregation of data can help to solve this problem: When we consider the average velocity  $v_{p,l,t}$  measured in a time interval  $I_t$  at one detector at position  $p \in P$  on lane  $l \leq L_p$ , then there will be usually only a small change to the value in the next time interval  $I_{t+1}$ , i.e. the difference  $|v_{p,l,t} - v_{p,l,t+1}|$  is usually not very large. Therefore we can aggregate a number  $q \geq 1$  of time intervals  $I_t, I_{t+1}, \dots, I_{t+q-1}$  into one large time interval

$$I_{t,q} := I_t \cup I_{t+1} \cup \dots \cup I_{t+q-1}$$

and consider the corresponding velocity values

$$v_{p,l,t}, v_{p,l,t+1}, \dots, v_{p,l,t+q-1}$$



**Fig. 1.** Emergence  $\varepsilon$  of the velocity in dependency of the time [min]. For the solid line the data of all detectors were used, while for the dotted line only the data of detectors with position  $p \in [0km, 6.1km]$  were used. An accident happened near position  $6.1km$  at time  $t = 1170$ .

as different realizations of the random variable  $V_{p,l,t,q}$  for the average velocity in  $I_{t,q}$  at position  $p$  on lane  $l$ . In the following we use  $q = 20$ . For each  $t \in \{0, 20, 40, \dots, 1420\}$  we have now  $q = 20$  realizations of the random variable  $V_{p,l,t,q}$ . Now we have to fix the measures  $m = (m_j)_{j \in J}$  for the emergence of the system. We could use the the measure

$$m_{p,l}(c) = v_{p,l,c} \quad \text{for } p \in P, \quad l \leq L_p$$

but this would cause very low values for the relative frequencies, because two realizations of the velocity usually have a different value. We use again the concept of aggregation, but this time on the values for the velocity: We do not consider single velocity values, but intervals  $K_0, K_1, \dots$  of velocity values. For this purpose we fix the interval length  $k = 30km/h$  and get the intervals

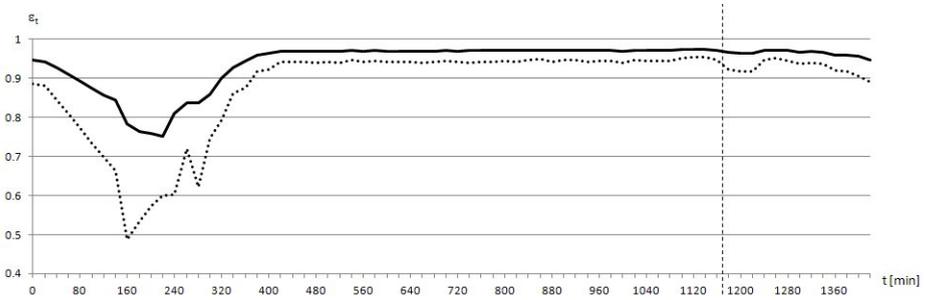
$$K_s = [s \cdot k, (s + 1) \cdot k[ \quad \text{for } s = 0, 1, 2, \dots$$

The corresponding measure is

$$m_{p,l}(c) = \lfloor v_{p,l,c}/k \rfloor$$

This measure yields the value  $s$  such that the realization of the velocity is in the interval  $K_s$ . As mentioned in Section 3 the values of the measures of the realizations form a matrix  $A$ , where the  $i$ -th row of the matrix is  $m(Conf_{t,i})$ , i.e. the  $i$ -th row consists of the values  $m_{p,l}(Conf_{t,i})$  for all positions  $p \in P$  and all lanes  $l \leq L_p$ . The approximated emergence  $\varepsilon_t \approx \varepsilon(A)$  in dependency of the time  $t \in \{0, 20, 40, \dots, 1420\}$  is shown in Figure 1 (solid line).

At  $t = 1170$  there is an accident on the highway near position  $p = 6.1$ , which causes a blocked lane. After this point in time we see a decrease of the emergence for a duration of about 80 minutes. Then the emergence gets back to a normal value. In a normal traffic flow, the speed of vehicles is much more harmonized than for situations with a blocked lane (caused by an accident or some other abnormal event), so a normal traffic flow shows more dependencies



**Fig. 2.** Emergence  $\varepsilon$  of the vehicle flow in dependency of the time [min]. For the solid line the data of all detectors were used, while for the dotted line only the data of detectors with position  $p \in [0km, 6.1km]$  were used. An accident happened near position  $6.1km$  at time  $t = 1170$ .

between the velocities of the vehicles, which is indicated by a higher value for the emergence  $\varepsilon_t$ . When we restrict our evaluation on the area  $[0km, 6.1km]$  (before the accident), then the decrease of the emergence at time  $t = 1170$  is much stronger (dotted line in Figure 1). This results from the fact that in the area behind the accident  $[6.1km, 12km]$  the traffic flow is much more harmonized than before the accident, so the restriction of the area of interest to the problem zone leads to a stronger irregularity in the graph of  $\varepsilon$  compared to the normal traffic flow.

Next to the emergence of the velocity data, we can also investigate the emergence of the flow of vehicles (temporal density). The value  $N_{p,l,t}$  measures the flow of cars at position  $p \in P$  for lane  $l \leq L_p$  with the unit  $min^{-1}$ . For the evaluation of the emergence of the flow data, we use the same methodology as for the velocity data. The only difference is the aggregation size on the value set: For the velocity data we aggregated the data into intervals of length  $k = 30km/h$ , while for the flow data we use the interval length  $k = 4min^{-1}$ . The aggregation on the time axis is for the flow data the same as for the velocity data:  $q = 20$ . The results are shown in Figure 2. At the time of accident, the evaluation of the whole data set (solid line in Figure 2) shows only a very small decrease. When we only evaluate the data of the detectors before the accident (i.e. with position  $p \leq 6.1km$ ) there is a larger decrease at the time of accident (dotted line in Figure 2). The graph shows that for  $t \in [80min, 320min]$  the emergence of the flow is below the normal value. This might result from the low traffic during the night: From 1:20am (which is  $t = 80min$ ) to 5:20am (which is  $t = 320min$ ) there is much less traffic on the road, so a disharmony in the traffic flow is more likely than on a full road. In the time interval  $[80, 320[$  the average number of detected cars per minute is

$$N_{avg} = \frac{1}{240} \sum_{t=80}^{319} N_t \approx 102$$

where

$$N_t = \sum_{p \in P, l \leq L_p} N_{p,l,t}$$

is the total number of cars detected in the interval  $I_t$  (which is usually much higher than the number of cars on the road segment, because each car is counted by more than one detector in  $1min$ ). In the time interval  $[320, 1440[$  the average number of detected cars is  $N_{avg} \approx 1357$ . While for the emergence of velocities the difference of the amount of traffic had only a small effect on the results (see Fig. 1), the emergence measure for the flow should only be interpreted in connection with the data for the amount of traffic.

## 5.2 Target Orientation of the Data Set

To be able to evaluate the highway system with the measure for target orientation, we have to specify the goal of the system. The main goal is the safety, so we have to define, which configurations are safe and which configurations are dangerous. In [2] three different measures for safety in traffic have been proposed:

- based on the variance of the velocities,
- based on the variance of accelerations,
- based on the mean time to collision (TTC)

From the data of the detectors at the M30 highway we are not able to calculate the accelerations or the TTC values for the vehicles, but we have some aggregated velocity values. Each detector yields the number of vehicles  $N_{p,l,t}$  which have passed the detector at position  $p \in P$  in lane  $l \leq L_p$  in the interval  $I_t$ . The average measured velocity of these vehicles can be used to specify the fitness function  $f : Conf \rightarrow [0, 1]$  on the set of all global configurations. We consider a configuration as dangerous, if the detectors yield velocity data with a high variance coefficient, because in this case an accident is more likely than for a homogeneous traffic flow, where the velocities of the vehicles are very similar. We only consider configurations, where at least one vehicle with positive velocity was detected (otherwise the variance coefficient of velocities does not exist). Let  $c \in Conf$ . Let

$$N_c = \sum_{p \in P, l \leq L_p} N_{p,l,c}$$

be the total number of detected vehicles for the configuration  $c$ . The average measured velocity of the detected vehicles in the configuration  $c$  is

$$\mu_c = \frac{1}{N_c} \sum_{p \in P, l \leq L_p} N_{p,l,c} \cdot v_{p,l,c}$$

For the empirical variance of the velocities, the detectors do not provide enough information, because some velocities are already aggregated into a single value  $v_{p,l,c}$ . But we can use  $N_{p,l,c}$  as a weight in the formula of the empirical variance:

$$\sigma_c^2 = \frac{1}{N_c - 1} \sum_{p \in P, l \leq L_p} N_{p,l,c} \cdot (v_{p,l,c} - \mu_c)^2$$

Then

$$\frac{\sigma_c}{\mu_c}$$

is the empirical variance coefficient of the velocity for the configuration  $c$ . Let

$$K \geq \max_{c \in Conf} \frac{\sigma_c}{\mu_c}$$

be a normalization constant.<sup>2</sup> The fitness  $f(c)$  of the configuration  $c$  is defined by

$$f(c) := 1 - \frac{1}{K} \cdot \frac{\sigma_c}{\mu_c}$$

A low value for the fitness implies a high variance coefficient for the velocities, i.e. the “bad states” are those configurations, where the velocities of different vehicles have a relative high variance.

For  $t \in T' := \{0, 1, \dots, 1439\}$  we have  $n = 1$  realization  $Conf_{t,1}$  of the random variable  $Conf_t$  for the configuration at time  $t$  containing all necessary information ( $v_{p,l,t}$  and  $N_{p,l,t}$ ) for the calculation of the fitness of the configuration  $Conf_{t,1}$ . The level of target orientation at time  $t \in T'$  is approximated by the fitness of the realization:

$$TO_t \approx f(Conf_{t,1}) = 1 - \frac{1}{K} \cdot \frac{\sigma_t}{\mu_t}$$

where  $\frac{\sigma_t}{\mu_t}$  is the empirical variance coefficient of measured velocities in the interval  $I_t$  for the realization  $Conf_{t,1}$ . For our data set we can set  $K := 1$  because all variance coefficients are below 1. The approximation of the target orientation  $TO : T' \rightarrow [0, 1]$  is shown in Figure 3.

At time  $t = 1170$  an accident has occurred. In Figure 3 we see that the level of target orientation is low at  $t = 1170$ , we have  $TO_t = 0.68$ . This means that we have a high variance coefficient for the velocities. After the accident, the level of target orientation still decreases to its minimum  $TO_t = 0.56$  at time  $t = 1247$  and then it goes back to a normal level  $\geq 0.9$  at time  $t = 1291$ .

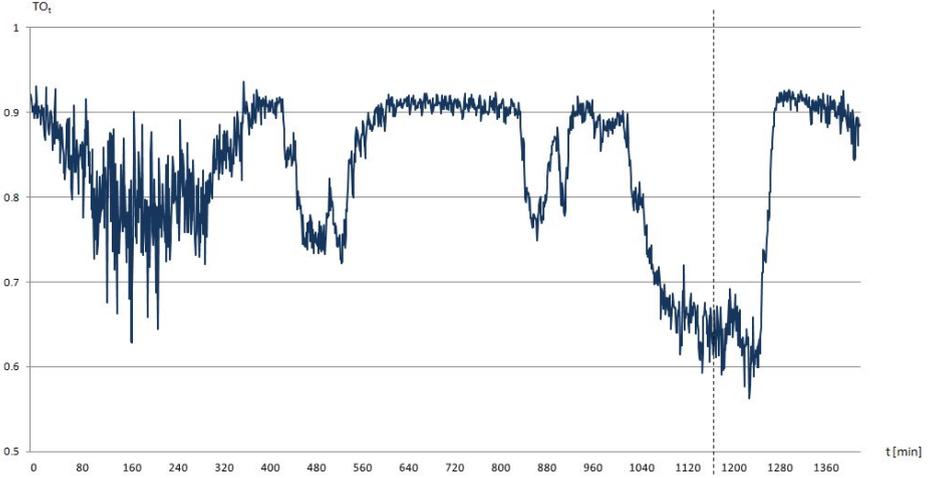
## 6 Discussion of the Results

In the previous section we applied measures on the M30 highway data set:

- emergence of velocity,
- emergence of flow,
- target orientation with respect to the variance coefficient of velocities.

For these measures, an accident on the highway causes a significant decrease of the values of the measures. For the level of emergence in velocities this can be explained by the fact that a homogeneous traffic flow shows more dependencies in the velocity data than a traffic flow that is disturbed by an event like an

<sup>2</sup> Such  $K$  exists because of the finite accuracy of the detectors.



**Fig. 3.** Target orientation  $TO$  of the system in dependency of the time [ $min$ ]

accident. Analogously this argument holds for the level of emergence in flow, but the decrease is not so strong like for the emergence of velocities. For the level of target orientation we already have a low value at the point in time, when the accident occurred, so the low fitness value of the global configuration indicates a dangerous situation, where accidents are more likely than for harmonic traffic flows with a low variance coefficient of velocities. Therefore an evaluation of the measure for target orientation can be used as an indicator for the prediction of accidents: A high value for  $TO_t$  indicates a low probability for accidents, while a low value for  $TO_t$  indicates a higher probability. For the level of emergence of velocities and densities, the values of the measures were high at the point in time, when the accident occurred, so these measure do not predict the accident, but since they decrease after the accident, all three measures can be used as indicators for the identification of accidents: If they have low values, then this might be caused by an abnormal event like an accident.

Also in the situation after the accident the level of target orientation and the levels of emergence stay at low values for more than one hour. Due to the accident, one lane of the highway was blocked, so a traffic jam occurred near the place of accident. The vehicles near the accident have a low velocity, while the vehicles far away from the accident have a high velocity, so the variance of the velocities is high, which is indicated by a low value of target orientation. Concerning the emergence of velocities, this situation shows less dependencies between the velocities of different vehicles than a harmonic traffic flow without a blocked lane, so we get a low value for the emergence for the velocity data.

For the level of emergence we have the problem that we are not able to calculate the exact value  $\varepsilon_t$  because of the huge state space of the system, so we

used the method of aggregation to approximate the value  $\varepsilon_t$ . It is unclear how good or how bad this approximation is. If we consider the enumerator

$$H(m(\text{Conf}_t))$$

in the definition of emergence  $\varepsilon_t$ , then we can observe that this value stays constant in our approximation method: Because of the large state space two realizations of  $\text{Conf}_t$  received from the data set will never yield the same value for  $m$ , i.e. we have

$$m(\text{Conf}_{t,i}) \neq m(\text{Conf}_{t,j})$$

for different realizations. Therefore, the relative frequencies, which are used for the approximations of the probabilities will either be 0 or 1. The result of the approximation  $\varepsilon_t \approx \varepsilon(A)$  used in the previous section only depends on the approximation of the denominator

$$\sum_{j \in J} H(m_j(\text{Conf}_t))$$

in the definition of the emergence. The evaluation shows that the changes in the denominator are sufficient to identify abnormal events like an accident. Therefore we can conclude that the value of the approximation  $\varepsilon(A)$  might differ very much from the real value  $\varepsilon_t$ , but changes in the function  $\varepsilon : T \rightarrow [0, 1]$  can also be found in the approximation, so the approximation contains enough information to be able to identify some properties of interest.

## 7 Conclusion and Future Work

In this paper we have investigated different analysis methods for traffic data: Emergence and target orientation. The measure for emergence applied to a data set of velocities or flow data can be used to identify global dependencies in the data set. The measure for target orientation can be used to identify dangerous situations in traffic. We applied both measures in a use case on a data set of the M30 highway in Madrid. The evaluation shows that the measures can be used to identify and predict abnormal events like accidents in traffic by an evaluation of velocity data: A decreasing level emergence of velocities indicates a reduction of dependencies between the velocities of different vehicles, which implies a disharmony in the traffic flow. A decreasing level of target orientation with respect to the specified fitness function indicates a higher variance of velocities, so we again have an indication for the disharmony in the traffic flow, which implies a more dangerous situation.

Note that these measures can not only be applied for the identification of accident but for arbitrary events disturbing the harmony of the traffic flow: Whenever the measures indicate a disharmonious traffic flow, a change of the behavior of the drivers might be desirable to harmonize the flow again and to increase safety in traffic. By using detectors at roads, which measures the velocities of passing vehicles, the evaluations of the measures can be used to control

overhead signs for speed limits to increase the safety of the traffic scenario, either by increasing the harmonization of the flow to prevent accidents or — in the case, that already an accident (or any other abnormal event disturbing the traffic flow) has already happened — to prevent follow-on accidents.

For future work it might be interesting to evaluate other (non-highway) data with some other abnormal events (e.g. traffic jams without accidents) disturbing the harmonization of the traffic flow to see, whether the proposed measures can also be used in other scenarios.

## References

1. Masinick, J., Teng, H.: An analysis on the impact of rubbernecking on urban freeway traffic. Tech. Rep., Centre for Transportation Studies of the University of Virginia (2004)
2. Fullerton, M., Holzer, R., De Meer, H., Beltran, C.: Novel assessment of a peer-peer road accident survival system. In: IEEE Self-adaptive and Self-organizing Systems Workshop Eval4SASO 2012 (2012)
3. Holzer, R., de Meer, H., Bettstetter, C.: On autonomy and emergence in self-organizing systems. In: Hummel, K.A., Sterbenz, J.P.G. (eds.) IWSOS 2008. LNCS, vol. 5343, pp. 157–169. Springer, Heidelberg (2008)
4. Holzer, R., de Meer, H.: Quantitative Modeling of Self-organizing Properties. In: Spyropoulos, T., Hummel, K.A. (eds.) IWSOS 2009. LNCS, vol. 5918, pp. 149–161. Springer, Heidelberg (2009)
5. Auer, C., Wüchner, P., de Meer, H.: The degree of global-state awareness in self-organizing systems. In: Spyropoulos, T., Hummel, K.A. (eds.) IWSOS 2009. LNCS, vol. 5918, pp. 125–136. Springer, Heidelberg (2009)
6. Assenmacher, S., Killat, M., Schmidt-Eisenlohr, F., Vortisch, P.: A simulative approach for the identification of possibilities and impacts of v2xcommunication. In: 15th World Congress on ITS, pp. 29–38 (2008)
7. Klein, A.: Untersuchung der Harmonisierungswirkung von Streckenbeeinflussungsanlagen. In: 4. Aachener Simulations Symposium (2011)
8. Cover, T.M., Thomas, J.A.: Elements of Information Theory. 2nd edn. Wiley (2006)
9. Holzer, R., De Meer, H.: Methods for approximations of quantitative measures in self-organizing systems. In: Bettstetter, C., Gershenson, C. (eds.) IWSOS 2011. LNCS, vol. 6557, pp. 1–15. Springer, Heidelberg (2011), The original publication is available at [www.springerlink.com](http://www.springerlink.com)