Routing Choice Information Maximising Robust Optimal Sensor Placement Against Variations of Traffic Demand Based on Importance of Nodes

Jordan Ivanchev[1], Heiko Aydt[2], Alois Knoll[3]

***Abstract*—This paper defines the measure of importance for a node (intersection) in a transportation network, based on its topology and traffic demand. Consequentially the measure is used in order to define and solve the sensor placement problem that maximises the information gain regarding agents' routing choices by sensing the most uncertain areas in the system. It is demonstrated that utilising the strategy for placing the sensors described in this work makes the performance robust against short and long term variations of traffic patterns. Finally, a method for finding the optimal number of sensors to be installed in a city is proposed. It models and maximises the trade-off between cost, performance, robustness and reliability of the sensor placement problem solution.**

I. Introduction

IDENTIFYING the most important modules or elements of a complex system is a problem that is of great interest to engineers and researchers. The entities or sub-systems that turn out to be of higher significance are, depending on the system, either cautiously monitored, robustly controlled, or studied in order to gain deeper understanding of the system's dynamics.

In the case of transportation systems, important links or nodes are usually sensed in order to get information about the overall traffic state. Engineers go even further by trying to change and control traffic parameters at such locations by planning new infrastructure developments [1], control strategies [2], novel policies [3] etc.

The aim of sensing traffic has been mostly in order to determine the flows in a city. The problem of optimal placement of counting sensors in order to estimate an OD matrix has been around for more the four decades [4]. Knowing the OD matrix, the flows can be extracted and knowing the flows, the delays on each link can also be evaluated, thus gathering some aggregated information about the traffic situation.

Given the increased pace of introduction of new technologies to the market and growing availability of computing power, traffic sensing and city planning are getting more interdependent and strongly connected. There are methods developed that even use sensed data in real time in order to apply changes to the traffic system [5].

Therefore, sensors may not be placed with the sole reason to observe traffic. Smart cities use their sensors' data streams in order to optimize their performance. With the increased number of sensor types such as plate scanning, velocity measuring, emissions etc, and their reduced error rate, it is now a matter of great importance to shift the sensor placement problem toward a more active goal. The information stream coming out from the sensors should be utilised by control algorithms or long term planning strategies in order to "actively" sense the traffic by controlling it at the same time.

Fundamentally, sensors are put in such positions so that they maximize the information gain. In other words, the chosen locations to be sensed are usually the ones that we are most uncertain about. Depending on the definition of information the placement problem takes different forms. Usually, information is considered to be the number of agents that pass through a link or the link's flow velocity. If, however, we want to find the intrinsically important locations, we need to look for the places where the choices, that lead to those characteristics, are made. The locations of activity should be sensed in order to actively sense the traffic. Consequentially, instead of examining links, a more topological approach should be taken and nodes should be examined instead. The nodes are actually the places where agents make choices, and what is sensed at the links are just the consequences of those choices.

The uncertainty of the agents choices at every node should be weighted by the number of agents that utilize it in order to be able to evaluate their importance. Naturally, the locations that need to be sensed in order to maximize the information gain are the ones of higher importance. It is important to note that, when examining nodes instead of links the importance values and the dynamics of each node are expected to be weakly correlated to their spatial position. This excludes the need for a mutual information optimisation approach in order to find the optimal sensor placement, because there is basically no redundant information in the sensing network.

Another pressing matter that has been ignored in the past is the robustness of the sensors. Usually, robustness is understood as the error rate or redundancy of a particular sensor placement. In this work, however, we look at robustness from another angle. Due to the fact that sensors are quite expensive and their installation is consuming both time and resources, we want to minimize the need of moving the sensors around (if at all possible) after they are once installed. In this sense, robustness of a sensor placement can be defined as the characteristic of the set of locations that they will stay important when we change the conditions in the system. Such changes may include short term changes in the OD matrix such

[1]TUM CREATE, 1 CREATE Way, 10-02 CREATE Tower, Singapore 138602 jordan.ivanchev@tum-create.edu.sg

[2]TUM CREATE, 1 CREATE Way, 10-02 CREATE Tower, Singapore 138602 heiko.aydt@tum-create.edu.sg

[3]Technische Universität München (TUM), Institute for Informatics, Robotics and Embedded Systems, Germany knoll@in.tum.de

as daily variations of traffic (evening rush hour vs. morning rush hour or weekday against weekend) and also long term changes in the network demand such as people moving around the city and changing living districts and jobs, building of new living complexes or business centers, districts etc. Such type of changes might severely alter the situation for a given sensor placement and thus make the investment obsolete. The robustness of a planned sensor network against variations in the traffic demand is of great importance, especially for megacities.

In a more fundamental aspect, there exist measures governed mostly by the OD matrix (link flow, link average speed) and measures connected to the topology (centrality, heterogeneity, etc.). There have been previous efforts to define entropy of a node or a link but only in a pure topological sense [6]. We strongly believe in the need of employing the information contained in the OD matrix as well in order to come up with a more useful definition of importance of a node. In this study we define the entropy of the node and consequently its importance by using both information about the traffic demand and topological information about the network, which surely gives a better overview. The information from both sources must be entangled since they are actively affecting each other, in order to get one single measure that represents all the information we posses.

The main contributions of this work are:

- Definition of entropy of a network and importance of nodes.
- Study on the robustness of the measure against changes in the OD matrix.
- Design of methods for finding the most robust optimal sensor placement against short and long term variations
- Design of a method for finding the optimal number of sensors to be placed in a given network.

II. Literature Review

Determining the importance of locations in traffic networks is crucial. One of the main areas, where the importance of locations plays a great role is traffic sensing. In most sensor placement problems, the set of locations to be sensed is chosen so that, the resulting synthesis of data is the most informative, which boils down to sensing the important locations. There are many attempts to find optimal sensor placement in order estimate an important traffic characteristic. One of the most comprehensive surveys [7] discusses and summarizes the existing sensor location problems, looking at different types of sensors as well as at the observability and the estimation problems of traffic flows. It also reviews models that deal with describing the optimisation problem. Moreover, it describes different rules for optimisation and analyses methods such as flow intercepting, demand intercepting, independence of traffic counts (mutual information).

One of the most standard traffic characteristic to be observed is the OD matrix. Estimating it from sensor data has become a central problem. In [8] the sensor location problem for OD matrix estimation is defined and a solution is suggested. This study deals with counting sensors, while other studies also include the possibility of using AVI (Automated Vehicle Identification) readers, which are more informative since they also collect information about the identity of the car, which allows for easier tracking and therefore path estimation [9]. In [10] both types of sensors are used in a method that places counting sensors and AVI readers to maximize the expected information gain for an OD demand estimation problem. It also takes into consideration uncertainty in historical demand information. A technique for calculating the optimal number and locations of plate scanning sensors for a given OD matrix is also presented in [11]. Those approaches are centred around the goal of estimating the OD matrix. In most of the cases they are applied on artificial networks as a proof of concept, however their high complexity might turn into a disadvantage if one tries to apply such a strategy for a real life megacity. Therefore there is a need for sensor placement method that is less computationally intensive so that it is practically applicable.

Once the locations of the sensors are fixed one might use a linear approximation technique is used to estimate offline the OD pairs using traffic counts such as the one described in [12]. In case plate scanning sensors are used a method for path reconstruction from such type of data can be used as in [9]. In [13] methods for extracting information from sensors data in order to estimate travel times are discussed, while also looking at sensor failure probabilities.

There are more universal approaches for choosing the most important locations to be sensed, which are based on maximizing information gain. There are information theoretic techniques such as [14],where a non-myopic strategy is used to find the most informative locations for sensors, [15] where a Kalman filtering structure is employed in order to solve a traverse time prediction problem via optimally placing sensors, and [16] where a method for target localization and tracking is presented, which computes the posterior target location distribution minimizing its entropy.

Information theoretic approaches, however, may vary among each other. In [17] traffic phenomena are modelled as Gaussian processes and the sensor locations are based on different information theory approaches. They discuss maximizing entropy for sensor locations and also mutual information between the locations and demonstrate that the mutual information approach performs better for certain type of scenarios. Moreover the method is extended to find robust placement against failures of sensors and uncertainties in the model and uses real world data sets. This is a generic method that can be applied to many different types of sensors. It locates the most representative links in the network that reduce the uncertainty about the unobserved links. There is not method that is able to determine the most important links in the sense of locations where agents makes the choices that are later observed at the representative links.

With the advancement of technology some type of sensors now can be mobile instead of static, while granting better coverage. In [18] a mobile traffic surveillance method is presented. A routing problem is defined such that it computes the optimal paths for the mobile sensors and show that in most cases it performs better than the static network. Link importance is defined as a combination of importance on three levels: link level, which looks at V/C ratio, path level, examining estimation of path travel time, and on a network level, looking at the percentage of OD pairs that use this link. Those components encompass the finding in [19], where the authors demonstrate using traffic indicators that importance of road segments is mainly determined by the network structure and the flows. Even though, this statement is clearly known there is still no indicator of importance of road segments that fully utilizes the flow information and the topological properties of the network. As it can be seen most methods to determine important locations for sensor placement are based mostly on the flows, while in a separate part of literature people look at purely topological properties of transportation graphs.

Important locations can be determined based solely on the topology of a network. Some efforts deal with identifying critical links using a network robustness index based on link flows, link capacity and network topology as in [20]. In [21] the most vital links or nodes are defined as the first n links or nodes whose removal will lead to the biggest increase in average shortest path distance. While in [22] the importance of roads is simply defined to be proportional to the traffic load on them, in [23] three measures of centrality for a street are suggested: closeness, betweenness and straightness and their correlation to various economic activities in the respective areas are examined.

Moreover, the network itself can have some properties that are usually based on the structure of the system and not on local properties of its elements. In [24] the development of the Swiss road and railway network during the second half of the 20th century is investigated. It is observed that the spatial structure of transportation networks is very specific, which makes it hard to analyse using methods developed for complex networks. In [25] Existing measures of heterogeneity, connectivity, accessibility, and interconnectivity are reviewed and three supplemental measures are proposed, including measures of entropy, connection patterns, and continuity. Entropy is used in order to determine the heterogeneity of the network regarding a chosen parameter.

The topology of a network holds an enormous amount of information. Using it we can get insights into the structure of the roads (transportation networks are organized hierarchically as shown in [26]). In [27] they measure the efficiency and accessibility in Paris and London based on the network connectedness. Moreover, this information can be utilized in order to reconstruct agent's trajectories from GPS signals as in [28]. There is also a family of graph measures based on entropy that are very well summarized in the survey [6]. It includes some measures from chemical structural analysis and social network analysis. The survey examines the overall connectedness of graphs such as the topological information content and the entropy of the weights of the edges. However measure of local features' such as entropy of nodes is defined as well, based on length of links connected to it. The centrality measure of links is also defined. Most of the measures deal with evaluating the information content in the graph itself. Those measures are highly uncorrelated, which means that they capture different aspects of graphs, so the proper measure should be chosen.

Once a measure of importance is defined and the most informative locations are chosen, there is one more aspect that needs to be examined. The robustness of those choices depends on the evolution of both the topology of the network and on the evolution of the OD matrix as well, and those two are also interdependent, of course. In [29] the evolution of the topology of networks is observed that demonstrates high degree of self-organization and spontaneous organization of hierarchies in the city of Indiana. Also variations in the relative importance of parts of the network are observed. In [30] the evolution over 200 years of a North Milan road network is observed. Two main processes can explain the processes that occur. Densification of the road network around the main roads and emergence of new roads as a results of urbanisation. An evaluation of the robustness against long term network evolution for any type of sensor placement is lacking at the moment.

In order to analyse traffic and plan for its surveillance one needs a model. Dynamic traffic assignment models such as the one described in [31] need a dynamic network load model and routing choices of agents model, which basically means that they need the OD matrix combined with a routing model such as in [32] based on stochastic conditions. Although patterns seem not to vary excessively as observed in [33]. It is shown that daily traffic is highly predictable and that there exist regular patterns that can be exploited. This stability of choices made by traffic participants together with network topology also leads to traffic concentration on mainly a few links of the network as shown in [34].

III. Measuring Importance of Nodes

In this section we introduce the measure of importance of nodes. We define a node as important if many agents pass through it and we are uncertain about the choices they make. In order to get the uncertainty we need an entropy measure at the node. After that we simply weight it by the throughput of agents. In this way we measure how much this node adds to the overall uncertainty of the road network given an OD matrix. Let us introduce some notation that will be used throughout the paper first:

N_{ij} - number of cars that pass sequentially through node i and j
P_l - the path of the l-th agent
f^l_{ij} - function that is one if the sequence of nodes ij is the that path of agent l
A - a set containing all the agents
p_{ij} - probability that an agent that is at node i will continue on to node j
S_i - set of nodes that are successors to node i
H_i - entropy of node i
I_i - importance of node i
N^r_{ij} - number of cars that pass sequentially through node i and j during time period r
H^t_i - entropy of node i during time period r
I^t_i - importance of node i during time period r
$\bar{I}_i$ - overall daily importance of node i

T - number of regions the day is split into
R - total reduced entropy
$\bar{I}_i^d$ - the overall importance of node i for a degree of perturbation d
L - a set of sensor locations
L^d - locally optimal sensor placement for a degree of perturbation d
L_o - globally optimal robust sensor placement
$Var_d[I_i^d]$ - the variance of I_i across all possible values of the d coefficient
$E_d[I_i^d]$ - the expectation of I_i across all possible values of the d coefficient
g^d - a function that takes as argument a set of sensor locations L and return the total reduced entropy for a given degree of perturbation d
V_{L_o} - variation level of the importance values of sensor placement L_o
$M_{L_o}^d$ - percentage of mismatched sensors between locally optimal placement L^d and globally optimal robust placement L_o
M_{L_o} - the overall percentage of mismatched sensors for all degrees of perturbation
Q_{L_o} - performance measure of robust optimal solution L_o
K_{L_o} - cost of installing solution L_o
U_{L_o} - utility function value of solution L_o

As we know Shannon's entropy is calculated using the transition probabilities between the states of the system. Let us assume that the state of an agent is its current link. Then the set of possible transitions from this state represents the agent turning on any of the links that are successors of the current link. The entropy of the node connecting those links is calculated using this information.

Here are the steps taken in order to calculate the importance of a node:

1) Calculate turning probabilities:
Let N_{ij} be the number of cars that pass through the i-th node and after that through the j-th node and let P_l be the path of the l-th agent. Then let the function f_{ij}^l:

$$f_{ij} = \begin{cases} 1 & \text{if nodes } ij \text{ are in } P_l \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Then:

$$N_{ij} = \sum_{l=1}^{|A|} f_{ij}^l(P_l) \tag{2}$$

, where $|A|$ is the number of agents.
Let p_{ij} be the probability that an agent at node i continues to node j
Let S_i be the set of nodes that are successors of node i. Then we can define the turning probability as the ratio between the number of cars that pass through node i and then proceed to node j and the number of cars that pass through node i altogether :

$$p_{ij} = \frac{N_{ij}}{\sum_{k \in S_i} N_{ik}} \tag{3}$$

2) Calculate the entropy at every node:
The entropy of a node i, H_i, is calculated using Shannon's entropy definition. A state is represented as the current link an agent is on and the transition probabilities are the turning probabilities from this node to its successors. Then the entropy becomes:

$$H_i = -\sum_{j \in S_i} p_{ij} \log p_{ij} \tag{4}$$

3) Weight the entropy of every node with the number of agents that pass through it
In order to differentiate between nodes that have a high entropy value and respectively, high and low traffic throughput, we weight the entropy of every node by the number of agents utilising it. The importance of node i is defined as:

$$I_i = H_i \sum_{j \in S_i} N_{ij} \tag{5}$$

Due to the fact that traffic demands change throughout the day so does the importance value of the nodes. First of all, agents make different routing choices in depending on the time of day and second the traffic volumes naturally also vary. Some nodes may experience high importance values during morning rush hour while having lower values during the evening. In case sensors are placed at nodes, whose importance value varies significantly throughout the day, they cannot be moved if some other nodes become more important. This is the reason why we need to find out the nodes that overall, have the biggest importance values across the day.

Since it is also important to study the daily variation of importance let us examine the notation describing splitting the day into time regions:
N_{ij}^t - the number of agents that go from node i to node j in period t
H_i^t - the entropy of node i during period t
I_i^t - the importance of node i during period t

Next step is to come up with an importance value representative for the whole day. Some regions of the day are of less interest than others simply because the amount of information that can be extracted is smaller. Typically, the factor that plays the largest role in this case is the traffic amount. Therefore, we compute the total importance of a node for the whole day, using a weighted average of importance values of the node for different regions of the day. The weight function is governed by the number of agents that pass through the node during the respective time region. Then, we can define the overall daily importance of a node as:

$$\bar{I}_i = \sum_{t=1}^{T} I_i^t \frac{\sum_{k \in S_i} N_{ik}^t}{\sum_{k \in S_i} N_{ik}} \tag{6}$$

The second term in the sum is simply the number of cars that pass through the node throughout time region t over the total number of cars that pass throughout the whole day and T is the number of regions the day is split into. This definition of overall importance puts an emphasis on the nodes that are interesting during the important parts of the day.

IV. Achieving Robustness against changes in the OD matrix

In reality apart from the changes of the OD matrix throughout the day, there is another process that alters the traffic demand in a less intense but more gradual way. This process is a result of long term changes to both the agent population and the city structure. In order demonstrate that our method is robust against such type of variations we implement a generic way to "alter" or "perturb" the traffic demand.

A. Methodology for altering the OD matrix

Let every agent have a list of itineraries, which is composed of separate trips. Every trip has an origin, destination and start time. It is usually the case that an agent takes two trips per day: from home to work in the morning and from work to home in the evening. Let us take two agents. We assume that the first origin and the last destination in the itinerary of those agents is their place of residence. Then we exchange those places for the agents. We do this for a predetermined percentage of all the agents that we call degree of disturbance. By executing this strategy the number of people starting from or arriving at all the regions is not changed. This means that the intensity of people starting from the various regions in not changed as is the intensity of people arriving at those regions. The factor that is perturbed is exactly the OD matrix, since the connections between origins and destinations are altered.

This procedure is visualised in Fig. 1

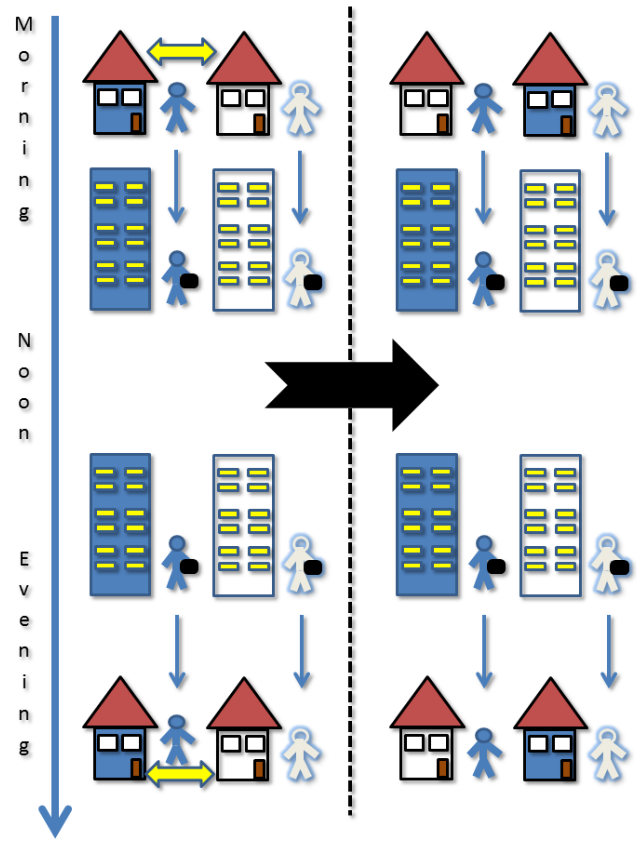


Fig. 1: Diagram illustrating the exchange of origins or living locations of two agents. Both agents still have the same work locations however they switch their homes. In this way the OD pairs intensity is changed.

B. Strategy for robust placement

In order to find locations that are optimal for performance and robust against variations in the OD matrix we have to find a measure that represents the importance of a set of nodes for different degrees of perturbation. This is the overall importance of the chosen locations. Every node has an importance measure $\bar{I}_i$. We assume that we can take a given number of sensors out of all possible locations and then we calculate the total reduced entropy in the network which is:

$$R = \sum_{i=1}^{|L|} \bar{I}_i \tag{7}$$

For every different degree of perturbation every node has a calculated importance value $\bar{I}_i^d$ where d is the degree of perturbation.

Let the resulting reduced entropy from a set of locations L for different degrees of perturbation be calculated by the function g^d:

$g^d(L) = R$, let the optimal placement for a given degree of perturbation d be L^d and $g^d(L^d) = R^d$

We are looking for an optimal placement L_o that maximizes the reduced entropy relative to the local maximum across the various perturbations:

$$\max_{L_o} \sum\nolimits_d \frac{g^d(L_o)}{g^d(L^d)} \tag{8}$$

C. Strategy for finding the optimal number of sensors

There are four aspects that should be taken into account when designing a utility function that should be maximised in order to find the optimal number of sensors.

1) The variation of the importance value across the perturbations.
   Every node has a different importance value across the perturbations $\bar{I}_i^d$. We want to evaluate the degree of variation so that we can locate globally important nodes rather than nodes that have just one high importance value. In order to do that, we calculate the variance for every node i across different degrees of perturbations d: $Var_d[\bar{I}_i^d]$. We scale it by the average value across the perturbations so that this measure is comparable to others:

$$\frac{Var_d[\bar{I}_i^d]}{E_d[\bar{I}_i^d]} \tag{9}$$

In order to evaluate the total variation level of the importance for a sensor placement we calculate the average of the scaled variances for all chosen locations:

$$V_{L_o} = E_i \left[\frac{Var_d[\bar{I_i^d}]}{E_d[\bar{I_i^d}]} \right] \tag{10}$$

We can vary the number of nodes to be included in the set of optimal locations L_o that can also be referred to as its cardinality: $|L_o|$. The goal is to minimise the variation of importance of the same node across the degrees of perturbation in order to ensure robustness of the placement.

2) The percentage of mismatched sensors.
We define the function $M^d_{L_o}$ as the percentage of sensors that are mismatched between the optimal sensor placement for a certain degree of perturbation L^d for a given number of sensors, and the robust optimal solution L_o, which is the cardinality of the difference between the two sets divided by the cardinality of the set:

$$M^d_{L_o} = \frac{\left|L_o \setminus L^d\right|}{|L_o|} \tag{11}$$

Then the overall percentage of mismatched sensors is just the average of this measure across all degrees of perturbation:

$$M_{L_o} = E_d[M^d_{L_o}] \tag{12}$$

This is a measure of distance between the optimal solution for d and the robust optimal solution for all degrees of perturbation. It can also be understood as a value signifying the percentage of sensors that need to be moved in order to reach the local optimal solution.
This measure should be minimised if we want to ensure robustness of the placement. Meaning that the sensor locations should be as universal as possible.

3) Performance measure of the robust optimal solution compared to the local optimal solutions
This measure is used to describe how close is the robust optimal solution to perfectly match the locally optimal solutions.

$$Q_{L_o} = \sum\nolimits_d \frac{g^d(L_o)}{g^d(L^d)} \tag{13}$$

This measure should be maximised since we aim for maximum performance.

4) Cost of sensors
We also include a function that punishes high number of sensors. For simplicity we just use a linear function that grows with the increase of number of sensors:

$$K_{L_o} = \alpha\,|L_o| \tag{14}$$

The utility function that needs to be maximised subject to the number of sensors or the cardinality of the set L_o then becomes:

$$\max_{|L_o|} U_{L_o} = w_1 Q_{L_o} - w_2 V_{L_o} - w_3 K_{L_o} - w_4 M_{L_o}\ , \quad \text{where } \sum_{i=1}^{4} w_i = 1 \tag{15}$$

All the separate functions are scaled to assume values between 0 and 1, however depending on the designers choice some measures can be given more weight by varying w_{1-4}. On the graph we can see all the separate functions and the utility function that determines the optimal sensor number.

V. Case Study: Singapore

In this section we are going to demonstrate the functioning of the described methodologies in a case study about Singapore. We start with calculating the entropies of every node of the network within each of the time regions they day is split into. A video can be found of the evolution of the importance measures of the nodes at VIDEO.

Following this we apply the robustness against daily variations technique. This allows us to get the overall daily importances of the nodes and use them to find the optimal sensors placement as described in section III. A picture of the Singaporean road network with the importances of nodes can be found on Fig. 2. It can be observed that the sensors cover the city completely with accents on the central business district, the highway intersections and intersection of highways with other large roads. More than that, there are also plenty of sensors in the living areas, which however have lower importance values due to the smaller number of cars that go through those intersections.

Next, we try to simulate change in traffic demand as explained section IV-A. Fig 3 visualises the results of applying the change. Since it is not practical to visualise all OD pairs, in our visualisation we are showing only the intensities of OD pairs that have as origin the university area around the Nanyang Technical University in the western part of the city. We can observe the change in the destinations intensities as people increase their trips to the east part of the city while reducing the trips that stay within the western part.

The following step is finding the optimal sensor placement for Singapore that is robust against such type of variations in the OD matrix as described in section IV-B. In order to evaluate the performance of the robust placement we run the following experiment:

1) For each degree of perturbation run a set of 10 simulations in order to get an averaged value for all the required parameters.
2) Using the simulation outputs, calculate the turning probabilities, entropies, and importance of all nodes
3) Find the optimal placement of sensors for every degree of perturbation
4) Using the optimisation strategy described above, calculate a robust sensor placement.
5) Compare the performance of the robust sensor placement to the performance of the locally optimal (in the sense of perturbation degree) sensor placements. The performance in this case is the ratio between the total reduced

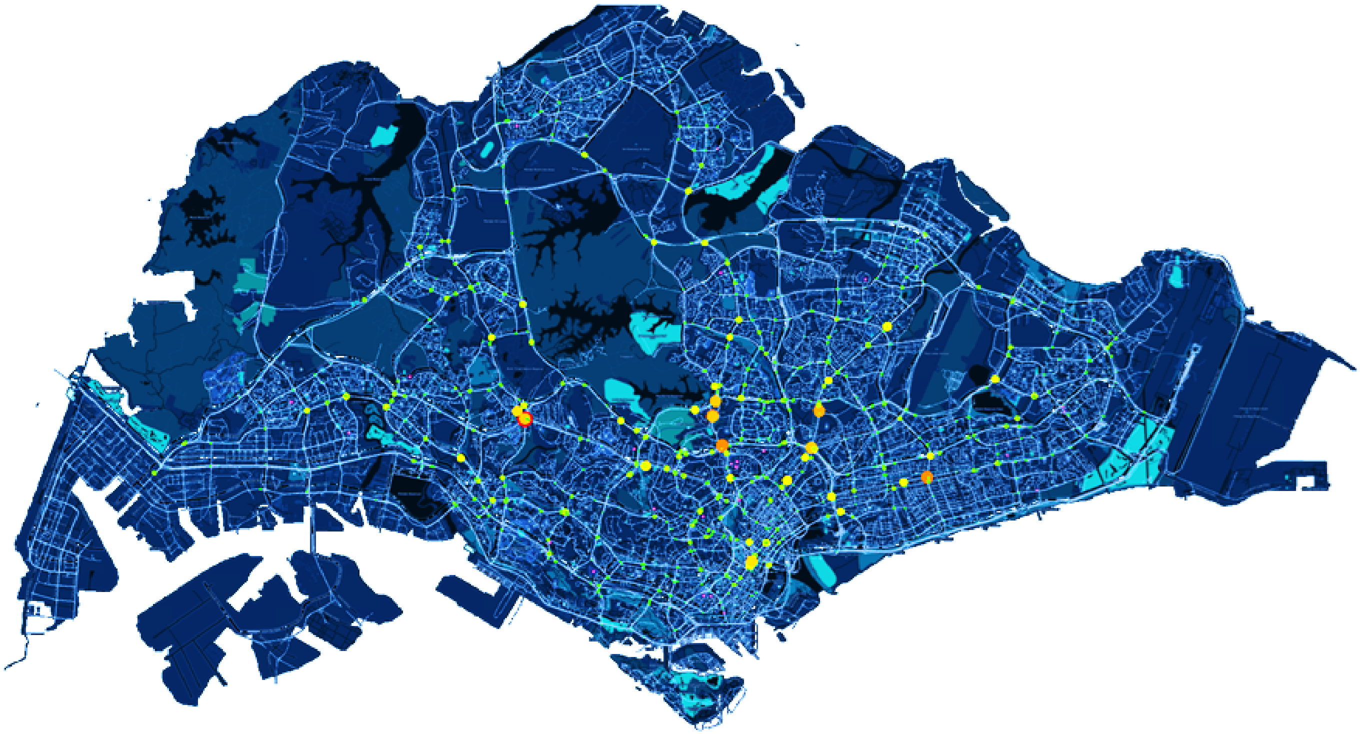


Fig. 2: Averaged importance value of nodes for a full day. The sensor placement resulting from those values is optimally robust against daily traffic variations.

entropy R of the robust placement to the total reduced entropy of the locally optimal placements.

In Fig. 4 we can see a comparison of the performance of the optimally robust method versus the locally optimal solutions for sensor placement.

After that we calculate the performance of those sets of locations for other degrees of perturbations. For example the blue line on Fig. 4 represents the optimal sensor placement for the original OD matrix. Naturally, since we are examining the optimal placement for 0 degrees perturbation the performance at 0 is 100%. We can then see that the performance of this sensor placement if the traffic is governed by the OD matrix perturbed by 5% decreases. The more we perturb the traffic demand, the more the performance of the optimal sensor placement calculated from the original OD matrix, decreases. The goal of the method is to achieve robustness in the sense that the performance stays high as we vary the OD matrix. We have also plotted the performance of our robust sensor placement solution. It can be observed that the robust solution does not vary that much when the OD matrix is perturbed and is performing better than the rest.

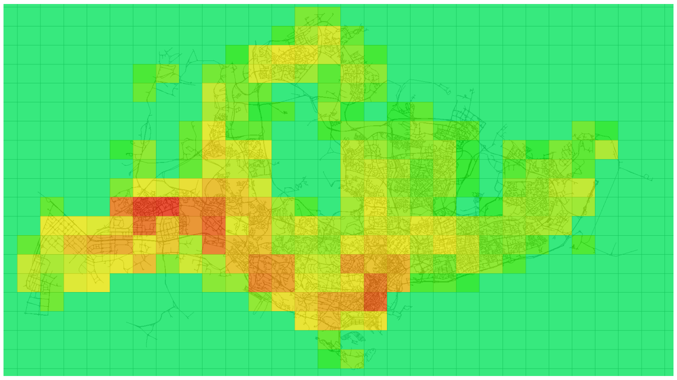
Finally, we have to compute the optimal number of sensors to be placed in Singapore as described in section IV-C. For the sake of simplicity let all factors be equally important. In Fig. 5 the functions related to the process of finding the sensor count are plotted. On the last sub-graph we can see the utility function whose maximum corresponds to the optimal number of sensors to be installed. We can see that in the case of Singapore this number is 582. Surely, if some factors are more important than others, they can be weighted differently and this will affect the optimal number of sensors.

VI. Conclusion

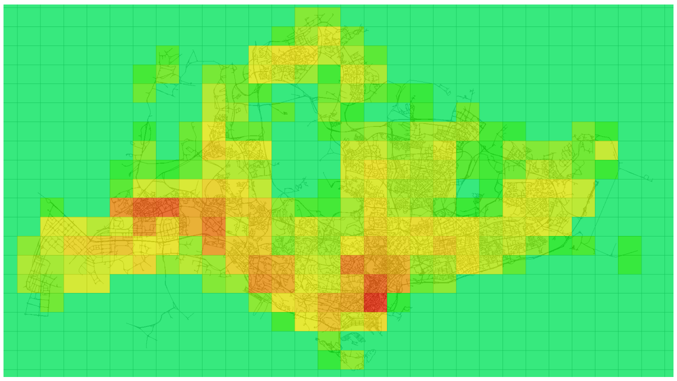
In this paper we have pointed out the need for an importance measure that is able to combine and refine the information about the traffic demand contained in the OD matrix and the information about the topology of the network. In this way one can gain the ability to point out to the locations in the network that hold the biggest amount of uncertainty related to agents' routing choices, which we believe is a crucial underlying factor that determines traffic conditions in a network. We have discussed that nodes should be examined instead of links since the intersections are the places were decisions are made and the roads are the locations were the results of those choices are observed.

We have defined a measure of importance that satisfies the aforementioned conditions using information theory. More precisely, the measure is a combination between the flow through a node and the entropy of the node itself. The novel definition of entropy of a node is dictated by the routing choices agent's made instead of by purely topological factors.

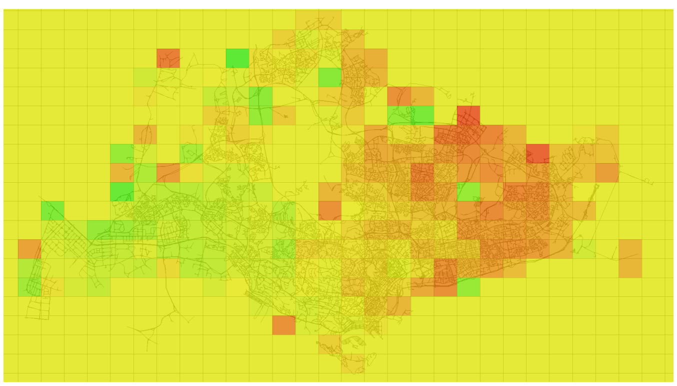
We have observed that the importance of nodes can vary throughout the day due to changes in traffic patterns, moreover



(a) Original OD



(b) OD with 30 % Perturbation



(c) Difference Map

Fig. 3: Visualisation of effects of perturbing the OD matrix. On 3a can be seen the original intensity OD pairs with destination NTU, while on 3b the intensities of the same pairs are shown after the OD matrix has been 30 % perturbed. On 3c we can see a heat map showing the differences between the two, where yellow means the same intensity, red means more and green means less.

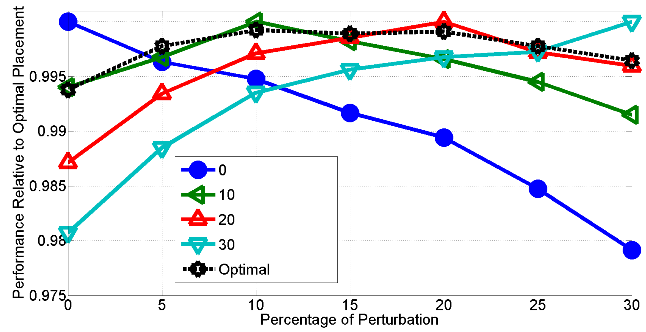


Fig. 4: Comparison of the performance of the optimally robust method across the various degrees of perturbation for the OD matrix to the performance of the locally optimal sensor placements. It can be observed that the robust placement is performing better so it is more invariant to changes in the OD matrix.

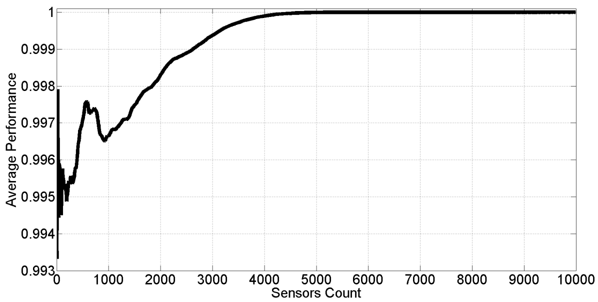
we have designed a method that finds the most robust sensor placement against such type of changes, which we call short term traffic demand variations. Long term changes are also being addressed by our work. We have defined a method to simulate long term city dynamics and their effect on the traffic demand in the city. Moreover, a method is described in order to find a robust sensor placement against such type of changes, providing certainty that sensors will not have to be moved once they are installed.

Finally, we have designed a method that allows designers to weight various factors connected to their preferences regarding the sensor network and its functionality, in order to determine the optimal number of sensors that need to be placed. The utility function consists of the variation factor of the sensor readings, the average percentage of mismatched sensors under varying traffic demand, the performance and the sensor instalment and sustaining cost.

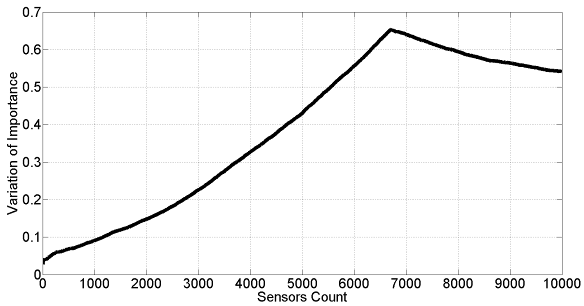
Future work on this topic would require the development of a more comprehensive tool for modelling long term changes in city dynamics, such as building new living or business areas, building new road segments etc. Implementation of such types of changes will bring qualitatively different type of OD matrix variations, since this processes will create completely new origins and destinations. Moreover the population growth should be modelled as well.

Another interesting research aspect would be to use the measure of importance in order to reconstruct agents' trajectories. Due to the fact, that by design the sensed locations maximise the information about agents' routing choices, the sensed data can be very useful in order to determine the paths agents take and consequentially the OD matrix. However, attention should be given to the issue that regions where agents actually begin or end their trips are usually of low importance value.

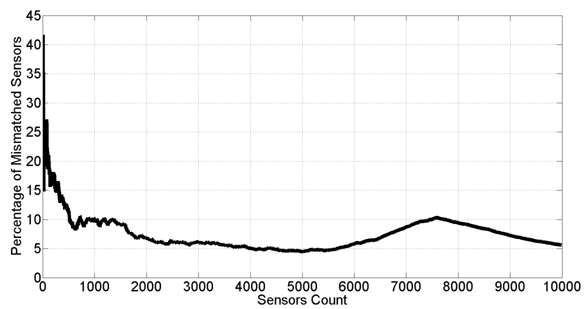
The heterogeneity of the nodes' importance as a network characteristic can be of great importance as well, as it is directly proportional to the utilization factor of the transportation network. In case of homogeneous importance values, there is lack of central points at which congestion is created. Homogeneity of the importance measure also means that agents are evenly spread across the network and utilize fully its infrastructure. Heterogeneity, on the other hand, means that agents' paths are very similar with the exception of several hub points through which everyone passes. This might bring imbalance of traffic on the network as some roads become congested while others stay empty. Following this argument it



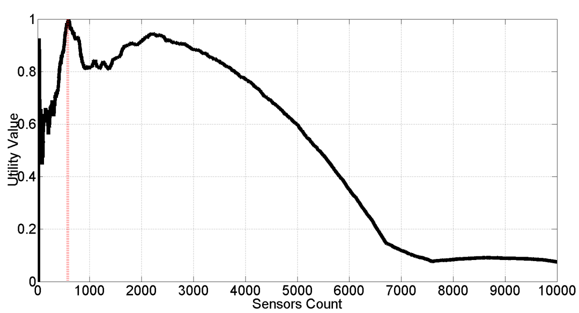
(a) Performance



(b) Reliability



(c) Robustness



(d) Utility Function

Fig. 5: Function that participate in the utility function and utility function to be maximized by the number of sensors: 5a performance of the sensor placement as a function of the sensor number, 5b Overall variation of the importance values of the sensor network as a function of sensor count, 5c Overall Percentage of mismatched sensors for the sensor network as a function of the sensor number, 5d utility function that maximizes 5a and minimizes 5b, 5c and the number of sensors

might be interesting to use the measure of heterogeneity of the importance measure of a network in order to either evaluate the traffic performance or optimise the routing of agents leading to overall reduction of congestion.

Finally, an interesting application of the measure would be to use its daily variation to determine the nodes (cross-sections) that are most dynamical in a city. Those intersections should be given additional attention. For example, this measure might be used in order to find the optimal locations for installing intersection control systems, since it provides us with the information that the flow ratios at those nodes vary significantly throughout the day.

Acknowledgement

This work was financially supported by the Singapore National Research Foundation under its Campus for Research Excellence And Technological Enterprise (CREATE) program.

References

[1] E. Beukes, M. Vanderschuren, and M. Zuidgeest, “Context sensitive multimodal road planning: a case study in cape town, south africa,” *Journal of Transport Geography*, vol. 19, no. 3, pp. 452 – 460, 2011, special Issue : Geographic Information Systems for Transportation. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966692310001341

[2] F.-Y. Wang, “Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications,” *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 3, pp. 630–638, Sept 2010.

[3] C.-L. Fok, M. Hanna, S. Gee, T.-C. Au, P. Stone, C. Julien, and S. Vishwanath, “A platform for evaluating autonomous intersection management policies,” in *Cyber-Physical Systems (ICCPS), 2012 IEEE/ACM Third International Conference on*, April 2012, pp. 87–96.

[4] P. Robillard, “Estimating the o-d matrix from observed link volumes,” *Transportation Research*, vol. 9, no. 2, pp. 123 – 128, 1975. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0041164775900490

[5] K. M. Yousef, M. N. Al-Karaki, and A. M. Shatnawi, “Intelligent traffic light flow control system using wireless sensors networks.” *J. Inf. Sci. Eng.*, vol. 26, no. 3, pp. 753–768, 2010.

[6] M. Dehmer and A. Mowshowitz, “A history of graph entropy measures,” *Information Sciences*, vol. 181, no. 1, pp. 57 – 78, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025510004147

[7] M. Gentili and P. Mirchandani, “Locating sensors on traffic networks: Models, challenges and research opportunities,” *Transportation Research Part C: Emerging Technologies*, vol. 24, no. 0, pp. 227 – 255, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X1200006X

[8] L. Bianco, G. Confessore, and P. Reverberi, “A network based model for traffic sensor location with implications on o/d matrix estimates,” *Transportation Science*, vol. 35, no. 1, pp. 50–60, 2001. [Online]. Available: http://dx.doi.org/10.1287/trsc.35.1.50.10140

[9] E. Castillo, J. M. Menéndez, and P. Jiménez, “Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations,” *Transportation Research Part B: Methodological*, vol. 42, no. 5, pp. 455 – 481, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191261507000975

[10] X. Zhou and G. F. List, “An information-theoretic sensor location model for traffic origin-destination demand estimation applications,” *Transportation Science*, vol. 44, no. 2, pp. 254–273, 2010. [Online]. Available: http://dx.doi.org/10.1287/trsc.1100.0319

[11] R. Mínguez, S. Sánchez-Cambronero, E. Castillo, and P. Jiménez, “Optimal traffic plate scanning location for {OD} trip matrix and route estimation in road networks,” *Transportation Research Part B: Methodological*, vol. 44, no. 2, pp. 282 – 298, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191261509000964

[12] T. Toledo and T. Kolechkina, “Estimation of dynamic origin destination matrices using linear assignment matrix approximations,” *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, no. 2, pp. 618–626, June 2013.

[13] X. Li and Y. Ouyang, "Reliable sensor deployment for network traffic surveillance," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 218 – 231, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191261510000573
[14] A. Krause and C. Guestrin, "Near-optimal nonmyopic value of information in graphical models," *CoRR*, vol. abs/1207.1394, 2012. [Online]. Available: http://arxiv.org/abs/1207.1394
[15] T. Xing, X. Zhou, and J. Taylor, "Designing heterogeneous sensor networks for estimating and predicting path travel time dynamics: An information-theoretic modeling approach," *Transportation Research Part B: Methodological*, vol. 57, no. 0, pp. 66 – 90, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191261513001677
[16] H. Wang, K. Yao, and D. Estrin, "Information-theoretic approaches for sensor selection and placement in sensor networks for target localization and tracking," *Communications and Networks, Journal of*, vol. 7, no. 4, pp. 438–449, 2005.
[17] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, Jun. 2008. [Online]. Available: http://dl.acm.org/citation.cfm?id=1390681.1390689
[18] N. Zhu, Y. Liu, S. Ma, and Z. He, "Mobile traffic sensor routing in dynamic transportation systems," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 5, pp. 2273–2285, Oct 2014.
[19] E. Jenelius, "Network structure and travel patterns: explaining the geographical disparities of road network vulnerability," *Journal of Transport Geography*, vol. 17, no. 3, pp. 234 – 244, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966692308000550
[20] D. M. Scott, D. C. Novak, L. Aultman-Hall, and F. Guo, "Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks," *Journal of Transport Geography*, vol. 14, no. 3, pp. 215 – 227, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0966692305000694
[21] H. Corley and D. Y. Sha, "Most vital links and nodes in weighted networks," *Operations Research Letters*, vol. 1, no. 4, pp. 157 – 160, 1982. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167637782900207
[22] M. Schreckenberg and R. Selten, *Human behaviour and traffic networks*. Springer, 2004.
[23] S. Porta, V. Latora, F. Wang, S. Rueda, E. Strano, S. Scellato, A. Cardillo, E. Belli, F. Cardenas, B. Cormenzana *et al.*, "Street centrality and the location of economic activities in barcelona," *Urban Studies*, vol. 49, no. 7, pp. 1471–1488, 2012.
[24] A. Erath, M. Löchl, and K. W. Axhausen, "Graph-theoretical analysis of the swiss road and railway networks over time," *Networks and Spatial Economics*, vol. 9, no. 3, pp. 379–400, 2009.
[25] F. Xie and D. Levinson, "Measuring the structure of road networks," *Geographical Analysis*, vol. 39, no. 3, pp. 336–356, 2007. [Online]. Available: http://dx.doi.org/10.1111/j.1538-4632.2007.00707.x
[26] B. M. Yerra and D. M. Levinson, "The emergence of hierarchy in transportation networks," *The Annals of Regional Science*, vol. 39, no. 3, pp. 541–553, 2005.
[27] L. d. F. Costa, B. Travencolo, M. Viana, and E. Strano, "On the efficiency of transportation systems in large cities," *EPL (Europhysics Letters)*, vol. 91, no. 1, p. 18003, 2010.
[28] O. Pink and B. Hummel, "A statistical approach to map matching using road network geometry, topology and vehicular motion constraints," in *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, Oct 2008, pp. 862–867.
[29] F. Xie and D. Levinson, "Topological evolution of surface transportation networks," *Computers, Environment and Urban Systems*, vol. 33, no. 3, pp. 211 – 223, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0198971508000653
[30] E. Strano, V. Nicosia, V. Latora, S. Porta, and M. Barthélemy, "Elementary processes governing the evolution of road networks," *Scientific reports*, vol. 2, 2012.
[31] M. Sanchez-Rico, R. Garcia-Rodenas, and J. Espinosa-Aranda, "A monte carlo approach to simulate the stochastic demand in a continuous dynamic traffic network loading problem," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 3, pp. 1362–1373, June 2014.
[32] T. xing and X. Zhou, "Reformulation and solution algorithms for absolute and percentile robust shortest path problems," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, no. 2, pp. 943–954, June 2013.
[33] Y. Li, D. Jin, P. Hui, Z. Wang, and S. Chen, "Limits of predictability for large-scale urban vehicular mobility," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 6, pp. 2671–2682, Dec 2014.
[34] S. Lämmer, B. Gehlsen, and D. Helbing, "Scaling laws in the spatial structure of urban road networks," *Physica A: Statistical Mechanics and its Applications*, vol. 363, no. 1, pp. 89 – 95, 2006, information and Material Flows in Complex Networks Information and Material Flows in Complex Networks. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437106000938

Jordan Ivanchev is pursuing his Ph.D degree in Computer Science from the Technische Universität München (TUM). He obtained his M.Sc degree from the University College in London (UCL) in Electrical Engineering and Communications and his B.Sc from the Jacobs University Bremen (JUB) in Electrical Engineering and Computer Science. Since January 2011 he held a position as a Research Associate in the MINDS machine learning group at JUB, working on hierarchical architectures for motion control based on recurrent neural networks. He joined TUM CREATE in June 2013 to work on traffic modelling and optimisation at the "Modelling and Optimization of Architectures and Infrastructure" group. His research interests include: automated agent-based simulation calibration, optimal sensor placement placement problems, social optimal routing strategies, traffic modelling, sensing and control.

Heiko Aydt received his Ph.D. degree in Computer Science from Nanyang Technological University (NTU) in Singapore and his M.Sc. degree from the Royal Institute of Technology (KTH) in Stockholm. He also holds a Dipl.-Ing. (FH) in Information Technology from Esslingen University of Applied Sciences. Prior to his doctoral studies, he gained experience as Software Engineer in the automotive industry. In December 2006 he joined the Parallel and Distributed Computing Centre at NTU as Research Associate where he worked on several projects concerned with simulation-based optimization and agent-based crowd simulation. In December 2011, he joined TUM CREATE as Research Fellow where he worked on agent-based traffic simulation and simulation-based optimization topics. As of December 2012, he assumed the role of Principal Investigator for the "Modelling and Optimization of Architectures and Infrastructure" group at TUM CREATE. His research is concerned with the analysis of the potential impact of electromobility on the infrastructure and the environment from a complex systems perspective. His current research interests are agent-based simulation, complex adaptive systems, symbiotic simulation, evolutionary computing and simulation-based optimization.

Alois Knoll received his diploma (M.Sc.) degree in Electrical/Communications Engineering from the University of Stuttgart and his PhD degree in Computer Science from the Technical University of Berlin. He served on the faculty of the computer science department of TU Berlin until 1993, when he qualified for teaching computer science at a university (habilitation). He then joined the Technical Faculty of the University of Bielefeld, where he was a full professor and the director of the research group Technical Informatics until 2001. Between May 2001 and April 2004 he was a member of the board of directors of the Fraunhofer-Institute for Autonomous Intelligent Systems. At AIS he was head of the research group "Robotics Construction Kits", dedicated to research and development in the area of educational robotics. Since autumn 2001 he has been a professor of Computer Science at the Computer Science Department of the Technische Universität München. He is also on the board of directors of the Central Institute of Medical Technology at TUM (IMETUM Garching). Between April 2004 and March 2006 he was Executive Director of the Institute of Computer Science at TUM. His research interests include cognitive, medical and sensor-based robotics, multi-agent systems, data fusion, adaptive systems and multimedia information retrieval.