Approximative network partitioning for MFDs from stationary sensor data

Lukas Ambühl

Allister Loder

Nan Zheng

Monica Menenedez

Kay W. Axhausen

STRC

18th Swiss Transport Research Conference
Monte Verità / Ascona, May 16 – 18, 2018

Institute for Transport Planning and Systems (IVT)

Approximative network partitioning for MFDs from stationary sensor data

Lukas Ambühl, Allister Loder,Kay W.
Axhausen
IVT
ETH Zurich
Stefano-Franscini-Platz
phone: +41-44-633 32 51
fax:
ambuehll@ethz.ch

Nan Zheng
School of Transportation Science and
Engineering
Beihang University
China
phone:
fax:
nan_zheng@buaa.edu.cn

Monica Menenedez
Division of Civil Engineering
NYU
Abu Dhabi
phone:
fax:
monica.menendez@nyu.edu

May 2018

Abstract

Macroscopic fundamental diagrams (MFD) are typically estimated for sub-regions in cities, where the zoning conventionally follows the directive of homogeneity in the road network and traffic. Existing network partitioning algorithms fail to work for MFD estimation with stationary sensor data as their link speed and link density measurement are biased by the location of the sensor. However, in this paper, we present an approach for approximative network partitioning for MFDs from stationary sensors.

Keywords

MFD; Network Partitioning; Heterogeneity; Stationary Detectors

1 Introduction

More and more field experiments show that when aggregating single road measurements of flow and density in an urban network, a smooth and well-defined relationship between average network flow, density and speed emerges (Geroliminis and Daganzo, 2008, Buisson and Ladier, 2009, Ambühl *et al.,* 2017b, 2018, 2017a, Loder *et al.,* 2017). These relationships define the macroscopic fundamental diagram (MFD) that is considered to be a characteristic feature of urban transportation networks (Daganzo, 2007, Daganzo and Geroliminis, 2008). A well-defined MFD can be expected when the analyzed sub-region of the city is homogeneous in terms of the road network as well as network topology (Geroliminis and Sun, 2011, Ji and Geroliminis, 2012). Knowing the MFDs for different regions in a city allows to advance and optimize network-wide traffic control in order to reduce congestion (Kouvelas *et al.,* 2017, Haddad and Geroliminis, 2012, Ramezani *et al.,* 2015). However, means to identify homogeneous urban regions only exist for unbiased link speed, flow and density measurements (Ji and Geroliminis, 2012, Saeedmanesh and Geroliminis, 2016, Ji *et al.,* 2014). The widely available traffic data from stationary traffic sensors (Leclercq *et al.,* 2014), however, cannot be easily used with such algorithms.

In this paper, we combine recent advances in estimating MFDs (Ambühl *et al.,* 2017b, 2018) to propose a Monte-Carlo simulation-like partitioning approach that approximates homogeneous regions. The algorithm generates a set of several possible partitioning outcomes, estimates for each outcome the MFDs, evaluates each possible partitioning outcomes' homogeneity, and then suggests the partitioning outcome with the lowest heterogeneity as the algorithm's solution. The algorithm is only approximative because the shape of the homogeneous neighborhoods depends on the spatial coverage of the stationary sensors and their distribution within the length of the links considered. Thus, by design, the solution is always inferior to existing link-based approaches (e.g., Ji and Geroliminis, 2012, Saeedmanesh and Geroliminis, 2016, Ji *et al.,* 2014) if these methods are applicable. Otherwise our proposed algorithms provides a reasonable answer to the partitioning problem.

This paper is organized as follows. The next section describes the partitioning algorithm in detail. Then, we apply the proposed algorithm to large scale data sets of London and Zurich. This paper then ends with a discussion of the contribution of this methodology as well as its limitations.

2 The Partitioning Algorithm

This partitioning algorithm is, in particular, designed for stationary traffic sensors, but is also applicable to floating car data and link based measurements as well. We discuss the MFD estimation in Section 2.1 and then introduce the partitioning approach in Section 2.2.

2.1 MFD estimation

We estimate MFDs from stationary traffic sensors by averaging link flows and densities weighted by the link length (Leclercq *et al.,* 2014). In case of S links monitored by stationary sensors, the network average flows and densities are calculated by Eqn. 1 and density in Eqn. 2.

$$q_{MFD}(t) = \frac{\sum_{i=1}^{S} q_i(t) l_i}{\sum_{i=1}^{S} l_i} \tag{1}$$

$$k_{MFD}(t) = \frac{\sum_{i=1}^{S} k_i(t) l_i}{\sum_{i=1}^{S} l_i} \tag{2}$$

An important issue with stationary traffic sensors is that $k_{MFD}(t)$ is sensitive to the sensor location and $k_{MFD}(t)$ is only unbiased if detectors are uniformly distributed within the length of links (Leclercq *et al.,* 2014, Buisson and Ladier, 2009). This condition is rarely satisfied in real urban networks. Therefore, we use information on the sensor location within the link, in order to correct for the biased density estimate (Ambühl *et al.,* 2017b). The idea is to account for the spatial differences introduced by any sensor distribution, and re-weight the density values in a way that a *uniform* distribution can be achieved therefore satisfying the condition by Courbon and Leclercq (2011), Leclercq *et al.* (2014) for accurate network-level density estimations. For this, we require the exact location of each sensor i with its location on the monitored lane l_i measured by the distance to the downstream traffic signal p_i and calculate each sensor's relative

position to the traffic signal r_i as simple as in Eqn. 3.

$$r_i = \frac{p_i}{l_i} \tag{3}$$

To then obtain an unbiased estimate of $k_{MFD}(t)$, the distribution of r_i across all sensors must be analyzed in order to divide M_j sensors according to their relative position into J groups. Each group should at least contain a couple of sensors. Then Eqn. 4 and 5 provide the mathematical formulation for the corrected and unbiased MFD estimation.

$$q_{MFD}(t) = \frac{1}{J} \sum_{j=1}^{J} \sum_{i \in M_j} \frac{q_i(t) \, l_i}{\sum l_i} \tag{4}$$

$$k_{MFD}(t) = \frac{1}{J} \sum_{j=1}^{J} \sum_{i \in M_j} \frac{k_i(t) \, l_i}{\sum l_i} \tag{5}$$

$$with \quad M = \bigcup_{j \in J} M_j \qquad\qquad and \quad M_j = \{i \in M \mid \frac{j-1}{J} < r_i < \frac{j}{J}\}$$

Last, the space-mean speed in the MFD $v_{MFD}(t)$ can then be calculated using the fundamental equation of traffic flow, i.e. $v_{MFD}(t) = q_{MFD}(t) / k_{MFD}(t)$.

2.2 Partitioning

The proposed algorithm consists of three steps. In the first step, a set of Φ possible partitioning outcomes is randomly generated. Based on the number of sensors S (links) in the network as well as the expected number of up to H homogeneous regions, the required size of Φ leading to a acceptable partitioning solution has to be estimated. In principle, the Stirling numbers of the second kind provides the upper most estimation of Φ, but as the regions should be compact in size and contiguous, Φ can be magnitudes smaller than this upper bound. Before the MFD estimation, all randomly generated possible partitioning outcomes should be assessed to exclude

all sets with unrealistic, unfeasible partitioning outcomes or where regions with too few sensors do not allow MFD estimation.

In the second step, for all sets of partitioning outcomes taken into account, we now estimate the MFDs according to Eqns. 4 and 5. In the MFD estimation, we then calculate the level of heterogeneity α for each region based on a re-sampling method (Ambühl *et al.*, 2018). In this method, we estimate MFDs not only with a full sample of sensors, but also estimate MFDs with a re sampling without replacement at 50-90 % of the entire number of sensors in the region and estimate by how much the capacity on average increases with lower sample size as a mean of heterogeneity in the network. Arguably, the more heterogeneous a region is, the more likely it is to have a small number of sensors provide a large deviation from the mean of all measurements which is revealed from this method.

In the third step, we evaluate for each partitioning outcome the overall level of heterogeneity in all regions together. We can determine a curve defined by the re-sampled MFD capacity as a function of the sample size. For simplicity, our objective function for the partitioning is the area below such curve of all sub-regions' heterogeneity α. The solution of the algorithm is then the outcome with the smallest area of all regions' α.

3 Empirical Application

3.1 General

We will now discuss the application of our algorithm in an empirical context. First, the spatial information of all sensor locations is obtained. Second, we generate several sets of possible partitioning outcomes by randomly creating contiguous regions or clusters. Third, we then estimate MFDs for each contiguous region in each of the possible partitioning outcomes. In this estimation procedure we then correct for the placement bias of loop detectors (Ambühl *et al.*, 2017b). Fourth, we estimate the level of heterogeneity in each MFD (Ambühl *et al.*, 2018). Last, the partitioning outcome with the lowest total level of heterogeneity is then the most homogeneous partitioning outcome.

The algorithm results in a partitioning solution where the network is divided into Ξ neighborhoods. This partitioning solution is identified by having the lowest total heterogeneity over

all neighborhoods. In contrast to existing partitioning algorithms, this algorithm does not find the solution by iteratively changing the shape of neighborhoods, but by a Monte Carlo method simulating Δ possible partitioning outcomes and then choosing the optimal partitioning outcome.

Consider an urban road network where the streets are monitored with in total L loop detectors. Each detector i is identified by its coordinates x_i and y_i. Each lane per street is covered only by one detector. For each detector, the length l_i of the monitored lane and the position with respect to the downstream traffic signal p_i is known. To then account for the placement bias of loop detectors we introduced the measure of relative position to the traffic signal r_i as simple as in Eqn. 3.

Each partitioning outcome is randomly generated using a simplified *k-means* clustering approach. We start by considering the locations of all sensors and selecting randomly $B \in [B_{min}; B_{max}]$ sensors in the network from which we start to cluster. Allowing B to vary creates more variation in the simulation of partitioning outcomes. All random partition outcomes are then compared to one another and only unique outcomes are kept.

As each neighborhood requires a certain minimum size, certain shape and a certain minimum number of sensors to estimate the MFD, we suggest to consider only those randomly generated partitioning outcomes that satisfy the three mentioned conditions.

3.2 Data

Before applying in the next Section our proposed methodology, we briefly discuss the preparation of the loop detector data sets from London (UK) and Zurich (CH). Loop detectors report flow $q_i(t)$ and occupancy $o_i(t)$. Where flow carries information directly useful for the MFD estimation, the measured occupancy is only a proxy for the link density and speed. The occupancy measurements can be transformed into density using the detector and (average) car length (Coifman, 2001), but these parameters are usually unknown. If further the detector length has an unknown distribution, the transformation gets even more complicated. Therefore, we use a two-step routine to calibrate the MFDs. In the first step, we calibrate each detector to an expected free flow speed at its location in order to reduce the bias from different detector and car lengths. The expected free flow speed at the upstream intersection ($r = 1$) is $v_{max,us}$ and the expected free flow speed at the downstream intersection ($r = 0$) is $v_{max,ds}$. For simplicity, we

assume a linear relationship, where we obtain Eqn. 6 which must be solved for the calibration scalar c_i.

$$\max_t \left(q_i\left(t\right)/o_i\left(t\right)\right) = c_i\left(v_{max,ds} + r_i\left(v_{max,us} - v_{max,ds}\right)\right) \tag{6}$$

In empirical contexts, one might prefer not the maximum of the speed distribution, but the 90th or 95th-percentile to avoid outliers. Here, we set $v_{max,ds} = 30km/h$ and $v_{max,us} = 40km/h$. In the second step, we then estimate the MFD in the usual approach with Eqns. 4 and 5 and than calibrate the free-flow speed in the MFD $\bar{v}_{MFD}$ with the free-flow speed obtained from the same time period from the Google directions API $\bar{v}_{API}$. The resulting calibration equation for $k_{MFD}\left(t\right)$ then follows from Eqn. 7.

$$k_i\left(t\right) = \frac{o_i\left(t\right)}{c_i}\frac{\bar{v}_{API}}{\bar{v}_{MFD}} \tag{7}$$

3.3 Preliminary Results

We create for every city 10000 randomly generated clusters. For the city of Zurich, B_min and B_max range from 5 to 12, respectively. The minimum number of loops required per cluster are 40. In order to evaluate the algorithm, we show the best four partitioning outcomes. Figure 1 shows the partitioning outcomes, where one colour corresponds to one cluster. We can see that the partition outcomes found all have the four clusters, which are very similar to each other. Note, that this also true for the next best partition outcomes (not shown here for readability. This indicates that our proposed partitioning is indeed robust. Another key thing to note is that all partitions identify topographical boundaries, such as the lake of Zurich or the hill range to the north west of the city. Here the algorithm finds the correct topographical borders, without any additional information. More interestingly the four clusters found more or less follow the boundaries of the control regions, that the city has defined for its traffic management strategy. We can see this as a further confirmation of our algorithm.
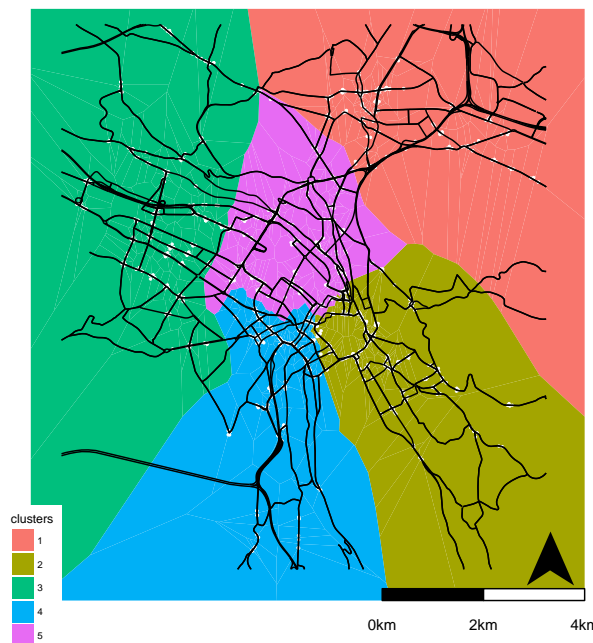
The same analysis is performed on the data of the inner city of London (limits are chosen to
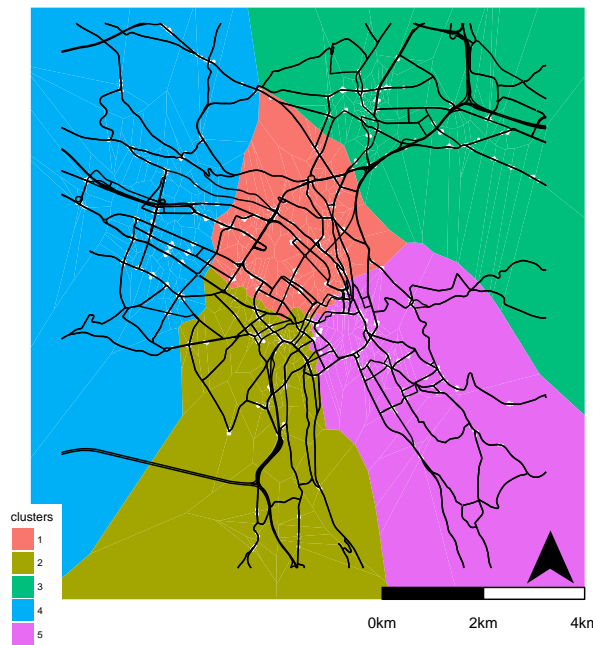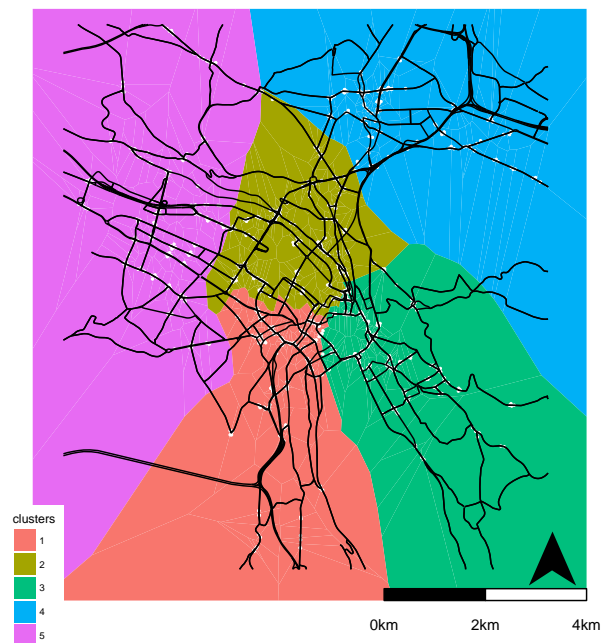
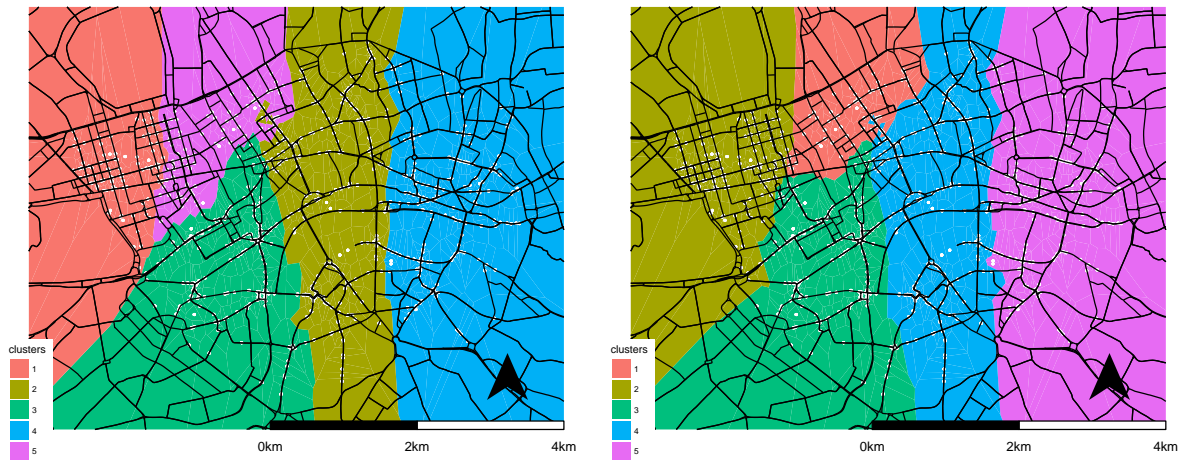(a) Best partitioning for Zurich.
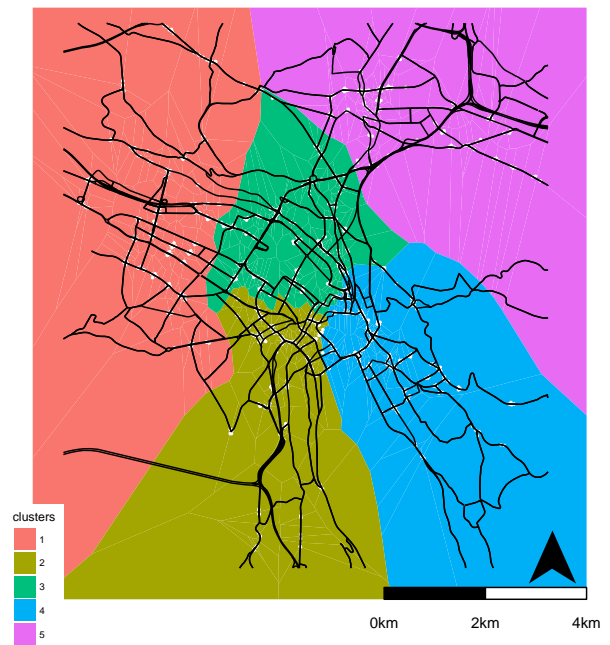
(b) 2nd best partitioning for Zurich.

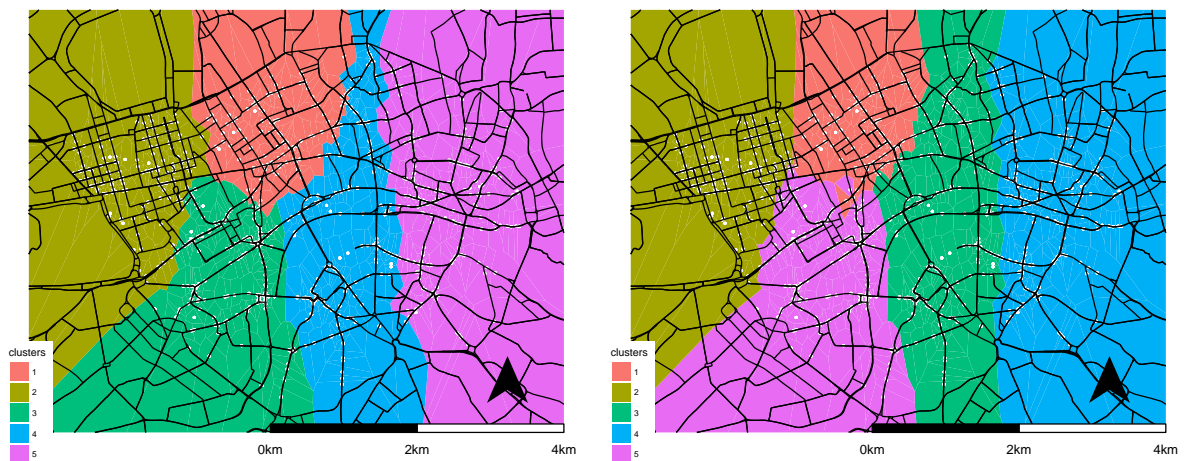(c) 3rd best partitioning for Zurich.

(d) 4th best partitioning for Zurich.

Figure 1: The best four partitioning outcomes for the city of Zurich. The white dots represent the location of the loop detectors.

(a) Best partitioning for London.

(b) 2nd best partitioning for London.

(c) 3rd best partitioning for London.

(d) 4th best partitioning for London.

Figure 2: The best four partitioning outcomes for the city of London. The white dots represent the location of the loop detectors.

correspond to the congestion charge zone). Again, we see that the four best clusters are similar to each other in Figure 2. The detailed interpretation of the result is on-going work.

In order to evaluate how different the randomly generated partitioning outcomes are from each other, let us consider Figure 3. Here, we show, for every loop detectors its relation with every
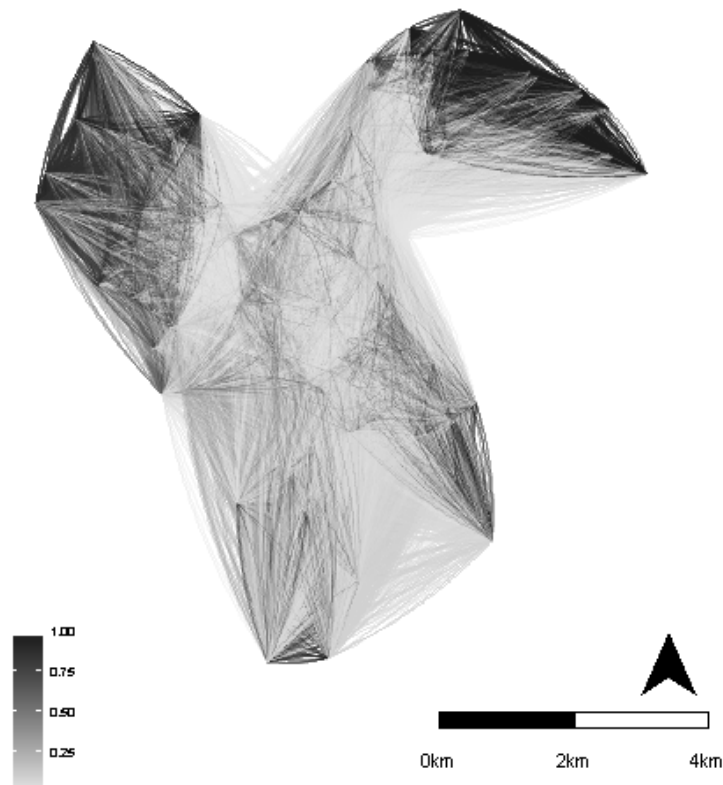
Figure 3: The connectivity of different partitioning outcomes: Each detector is connected to another one, if they are both in the same cluster. The more likely it is for them to appear in the same cluster the stronger their connection is shown.

other loop detector with respect to all considered partitioning outcome for the city of Zurich. In other words, the strength of a line connecting two loop detectors indicates for how many partitioning outcomes these detectors belong to the same cluster. From this graph, we see that the partitions generated cover the network well. This means that the partitions are very different from one another, thereby exploring the space well, finding many different possible partitioning outcomes.

4 Discussion

In this paper, we propose an approximative network partitioning for the MFD estimation. We propose to partition urban road networks into homogeneous sub networks by a Monte-Carlo-like partitioning algorithm. The algorithm is in particular useful for MFDs estimated from stationary traffic sensors when existing partitioning approaches are not applicable. For such sensors, a

strong bias on the density (and speed) measurements is expected. Based on a heterogeneity index, which is independent of the sensors' positions, we detect the most suitable network partition. As this approach draws the solution from a finite number randomly generated outcomes, the solution might not be optimal and is thus only approximate and inferior to existing approaches when they are applicable.

Using empirical data from Zurich and London, we show that the method is capable of finding homogeneous regions in a large cities. The results are promising as they show how the clusters with the lowest heterogeneity indices are very similar to each other. The proposed method is flexible, does not require much input and yields geographically contiguous clusters. Furthermore, the number of partitions is not pre-defined, while a reasonable range thereof is an input of the algorithm.

It is clear that this approximative partitioning algorithm bears certain limitations. The algorithm is based on a Monte-Carlo approach, involving a large number of randomly generated possible clusters. This is not a problem from a computational perspective; the generation of such clusters is relatively fast. Nonetheless, future research should investigate ways that directly partition a network in a single shot, similar to the methods available for FCD based partitioning.

Acknowledgments

References

5 References

Ambühl, L., A. Loder, M. Bliemer, M. Menendez and K. W. Axhausen (2017a) A functional form for the macroscopic fundamental diagram with a physical meaning, *Arbeitsberichte*

Verkehrs- und Raumplanung, **1306**.

Ambühl, L., A. Loder, M. C. J. Bliemer, M. Menendez and K. W. Axhausen (2018) Introducing a re-sampling methodology for the estimation of empirical macroscopic fundamental diagrams, *Transportation Research Record: Journal of the Transportation Research Board*, **(accepted)**.

Ambühl, L., A. Loder, M. Menendez and K. W. Axhausen (2017b) Empirical macroscopic fundamental diagrams: New insights from loop detector and floating car data, paper presented at the *96th Annual Meeting of the Transportation Research Board*.

Buisson, C. and C. Ladier (2009) Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams, *Transportation Research Record*, 127–136.

Coifman, B. (2001) Improved velocity estimation using single loop detectors, *Transportation Research Part A: Policy and Practice*, **35**, 863–880.

Courbon, T. and L. Leclercq (2011) Cross-comparison of macroscopic fundamental diagram estimation methods, *Procedia - Social and Behavioral Sciences*, **20**, 417–426.

Daganzo, C. F. (2007) Urban gridlock: Macroscopic modeling and mitigation approaches, *Transportation Research Part B: Methodological*, **41**, 49–62.

Daganzo, C. F. and N. Geroliminis (2008) An analytical approximation for the macroscopic fundamental diagram of urban traffic, *Transportation Research Part B: Methodological*, **42**, 771–781.

Geroliminis, N. and C. F. Daganzo (2008) Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, *Transportation Research Part B: Methodological*, **42**, 759–770.

Geroliminis, N. and J. Sun (2011) Properties of a well-defined macroscopic fundamental diagram for urban traffic, *Transportation Research Part B: Methodological*, **45** (3) 605–617.

Haddad, J. and N. Geroliminis (2012) On the stability of traffic perimeter control in two-region urban cities, *Transportation Research Part B: Methodological*, **46**, 1159–1176.

Ji, Y. and N. Geroliminis (2012) On the spatial partitioning of urban transportation networks, *Transportation Research Part B: Methodological*, **46** (10) 1639–1656.

Ji, Y., J. Luo and N. Geroliminis (2014) Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems, *Transportation Research Record: Journal of the Transportation Research Board*, **2422**, 1–11.

Kouvelas, A., M. Saeedmanesh and N. Geroliminis (2017) Enhancing model-based feedback perimeter control with data-driven online adaptive optimization, *Transportation Research Part B: Methodological*, **96**, 26–45.

Leclercq, L., N. Chiabaut and B. Trinquier (2014) Macroscopic fundamental diagrams: A cross-comparison of estimation methods, *Transportation Research Part B: Methodological*, **62**, 1–12.

Loder, A., L. Ambühl, M. Menendez and K. W. Axhausen (2017) Empirics of multi-modal traffic networks – using the 3d macroscopic fundamental diagram, *Transportation Research Part C: Emerging Technologies*, **82**, 88–101.

Ramezani, M., J. Haddad and N. Geroliminis (2015) Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control, *Transportation Research Part B: Methodological*, **74**, 1–19.

Saeedmanesh, M. and N. Geroliminis (2016) Clustering of heterogeneous networks with directional flows based on "snake" similarities, *Transportation Research Part B: Methodological*, **91**, 250–269.