CONTIGUITY-CONSTRAINED PARTITIONING OF HETEROGENEOUS URBAN NETWORKS

Mohammadreza Saeedmanesh

Nikolas Geroliminis

May 2016

STRC | 16th Swiss Transport Research Conference

Monte Verità / Ascona, May 18 – 20, 2016

CONTIGUITY-CONSTRAINED PARTITIONING OF HETEROGENEOUS URBAN NETWORKS

Mohammadreza Saeedmanesh
Urban Transport Systems Laboratory
Ecole Polytechnique Fédérale de Lausanne
Lausanne GC C2 390, Station 18,
Lausanne 1015, Switzerland
phone: +41-21-69-35397
fax: +41-21-69-35060
mohammadreza.saeedmanesh@epfl.ch

Nikolas Geroliminis
Urban Transport Systems Laboratory
Ecole Polytechnique Fédérale de Lausanne
Lausanne GC C2 389, Station 18,
Lausanne 1015, Switzerland
phone: +41-21-69-32481
fax: +41-21-69-35060
nikolas.geroliminis@epfl.ch

May 2016

Abstract

The spatial correlation of urban traffic congestion allows the development of clustering algorithms to divide heterogeneous networks into relatively homogeneous and spatially connected sub-regions. Specifically, by knowing the distribution of congestion over the network and combining monitoring techniques with control theory, recent developments are capable of designing real-time hierarchical traffic control schemes to protect regions with high level of congestion. The modeling tool to advance this work is the network macroscopic fundamental diagram (MFD). The concept of MFD specifies the aggregated traffic state in a region (i.e. linking the average space-mean flow with accumulation) and it has been incorporated as a model in different control approaches, mainly perimeter or boundary control. Since these approaches control the transfer flows on the boundaries, homogeneous and compact areas with smooth boundaries ease the applicability and efficiency of control. In this paper, we model the clustering problem as a mixed-integer linear optimization, taking into account contiguity and size constraints for the clusters. The objective function comprises a weighted average of two terms representing heterogeneity and non-compactness of the clusters. The weights are defined by the user according to the desired relative importance of the homogeneity to the compactness. The homogeneity is calculated as the sum of all squared differences of the link densities within a cluster from the approximate mean of that cluster. Compactness is attained by minimizing the sum of all spatial distances for the links in each cluster to the center of that cluster. The preliminary results of the proposed clustering framework in a simulated network show the flexibility to make a trade-off between compactness and homogeneity in different clusters.

Keywords

Graph partitioning, Conegstion propagation, Mixed integer linear programming, MFD

1 Introduction

Traffic congestion in urban networks appear in different regions with different shapes and might propagate in particular directions varying from day to day. There is a strong effort in the last decades for traffic flow models in one-dimensional traffic systems (see (Helbing, 2001) for an overview). Nevertheless, literature in network level dynamics and congestion propagation is limited and has been mainly built on micro-simulations of link-level traffic dynamics. Due to unpredictability of travel behaviours, accurate physical modeling remains still challenging and simulation results may not be realistic for dynamic systems with stochastic characteristics. Recently, it has been observed with simulated and empirical data in (Geroliminis and Daganzo, 2008), (Buisson and Ladier, 2009), and (Gayah and Daganzo, 2011) that a low scatter unimodal relationship exists between space-mean flow and density in networks with homogeneous traffic conditions. Spatial heterogeneity in the distribution of congestion can significantly influence the shape and the scatter in MFD curves (Geroliminis and Sun, 2011) and (Saberi *et al.,* 2014). The spatial correlations of urban traffic congestion allows the development of clustering algorithms to divide heterogeneous urban networks into homogeneous regions (see (Ji and Geroliminis, 2012), (Saeedmanesh and Geroliminis, 2016a)). Specifically, by knowing the distribution of congestion over the network, we are capable of designing real-time hierarchical traffic control schemes (e.g. perimeter control, gating, etc.) to improve the mobility and avoid gridlock conditions in regions with high level of congestion (Kouvelas *et al.,* 2015) and (Keyvan-Ekbatani *et al.,* 2012). Note that, MFD has been utilized as a key tool for modeling in these studies. A detailed literature review of network modeling and control can be found for example in (Haddad *et al.,* 2013) and (Mahmassani *et al.,* 2013).

The problem of partitioning the heterogeneous network into spatially connected and homogeneous regions is considered as a particular case of clustering which is known as contiguity-constrained clustering. There are several methods introduced to deal with this problem which can be divided into two main groups according to their strategies to maintain connectivity for obtained clusters. In one hand there are conventional clustering methods which implicitly fulfill the connectivity by incorporating spatial information into the classification data. The method presented in (Ji and Geroliminis, 2012) indirectly imposes connectivity by assuming similarity only between neighboring roads. However, in non-grid networks, this method tends to partition from the locations where network has low connectivity regardless of the level of congestion. Moreover, the method requires a connected graph of the network and missing values or malfunctioning detectors might create difficulties in application of the method. The proposed 'Snake' method in (Saeedmanesh and Geroliminis, 2016a) tries to overcome the aforementioned difficulty and develops a clustering methodology that is able to find directional congestion within a cluster and has good performance for networks with low connectivity. The proposed method in (Saeedmanesh and Geroliminis, 2016a) utilizes the symmetric non-negative Matrix Factorization to find clusters from similarity matrix obtained by running different snakes in the network. In the other hand, there are two categories of methods that explicitly impose the connectivity: (i) heuristic approaches; (ii) exact optimization models. The heuristic approaches are effectively utilized in cases where we have to aggregate high number of roads in big urban networks. Different types of agglomerative hierarchical clustering methods proposed in (Guo, 2008) are examples of heuristic methods. At the beginning of these algorithms, each road is considered as one cluster and a tree is derived by merging two clusters that contains adjacent roads. Then, clusters are obtained by cutting one edge from the tree until reaching predefined number of regions. A main challenge in heuristic algorithms is to avoid local optimal solutions, which motivated researches to develop exact models for contiguity constrained clustering problem. We refer the reader to (Carlos *et al.,* 2007) for more detailed overview about existing methods.

In this study, we formulate the clustering problem as a mixed-integer linear programming (MILP), taking into account contiguity and size constraints for the clusters. The objective of the optimization is to maximize the weighted average of the homogeneity and compactness of the clusters. In fact, compact areas with smooth boundaries ease the applicability and efficiency of control as these approaches control the transfer flows on the boundaries. The weights in the objective function are defined by the user according to the desired relative importance of the homogeneity to the compactness. The proposed clustering framework can be applied in heterogeneous large-scale real networks and has the flexibility to make a trade-off between compactness and homogeneity in different clusters. Note that, the current formulation has some advantages compared to the formulation presented in (Saeedmanesh and Geroliminis, 2016b) by the same authors: (i) contiguity is modeled in such a way that needs much less decision variables which reduces the computational complexity; (ii) the compactness is defined in a more precise way using the new formulation (i.e. the shortest path inside each cluster is considered rather than the shortest path in the entire network); (iii) compactness is directly written as a linear function of decision variables while in Saeedmanesh and Geroliminis (2016b) the initial quadratic compactness function is approximated by linear functions using upper and lower bounds.

The remainder of this paper is organized as follows: In Section 2, we introduce the proposed optimization model for partitioning problem and derive a MILP formulation. Moreover, modeling different objectives (compactness, homogeneity) and constraints (contiguity, size, etc.) are explained in details. The Section 3 presents the partitioning results with different number of partitions. Different objective weights are presented and compared for a grid-type medium size network (San Francisco with about 400 links). The paper concludes with a discussion about the results and ideas for further research.

2 Methodological framework

Consider an urban network with uneven and inconsistent distribution of congestion, which occurs frequently over a day in real networks with high travel demand. The main objective is to partition such a network with a large number of roads into connected, homogeneous and spatially compact sub-regions. An ideal sub-region is a relatively homogeneous area characterized by a set of following properties: (i) spatial connectivity (i.e. roads inside each cluster are connected to each other) which facilitates effective traffic management strategies such as perimeter control and gating; (ii) spatial compactness is an important aspect of clustering as very weird shapes might make perimeter control inefficient. Given that drivers tend to choose routes without too many turns, a non-smooth boundary where perimeter control is applied might create shortest paths with a large number of turns and change the behavior of drivers in non-predictable ways; (iii) certain minimum number of links that each cluster should contain. Regions with a few number of roads (e.g. a single road in an extreme case) have high scatter MFDs which creates difficulties for multi-region control strategies (like perimeter control ad gating). These control strategies assume a well-defined low-scatter relation between the production and accumulation for each region.

Indeed, homogeneity and compactness criteria are two conflicting objectives that need to be taken into account at the same time. Hence, the optimization problem is formulated with an objective of maximizing a weighted sum of homogeneity and compactness for different clusters. Note that, these weights represent the importance of each two objectives and can be defined by the user. The connectivity in each cluster is explicitly imposed by constraints. This type of partitioning is known as 'Contiguity-constrained clustering' in the literature. In the following sections (Section 2.1, Section 2.2), we explain in details the metrics used to quantify homogeneity and compactness.

2.1 Homogeneity metric

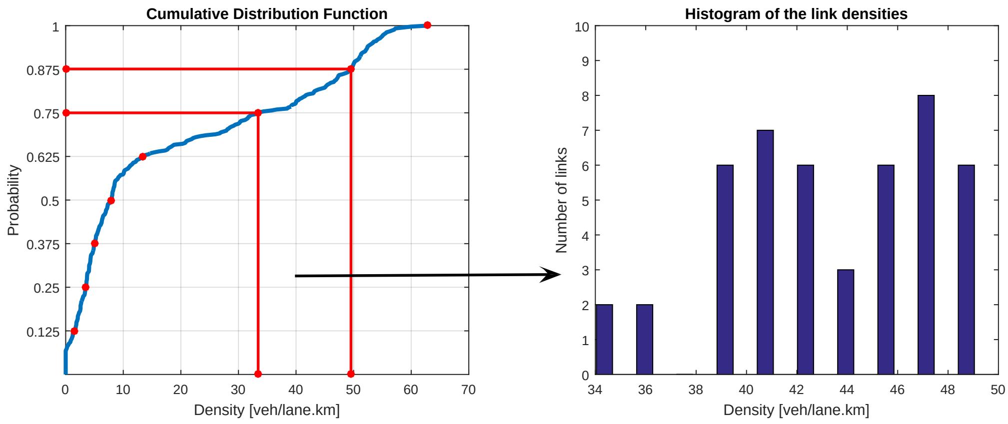
A well-established index to quantify the heterogeneity is the normalized total variance (TV_n) of different clusters presented in Eq. (1), which considers both cluster size and data variation within the clusters. This formula is mainly utilized to make an a posteriori evaluation on the partitioning results once the clusters are determined. However, it is not straightforward to use it as an objective function in the optimization process since the average values of the clusters are not known in advance and vary by any changes in clusters. This creates non-linearity and non-convexity that unease classical approaches for solving linear or convex problems.

$$TV_n = \frac{\sum_{i=1}^{N_c} \sum_{j=1}^{N} x_{ij} \times (d_j - \frac{\sum_{k=1}^{N} x_{ik} \times d_k}{\sum_{k=1}^{N} x_{ik}})^2}{\sum_{j=1}^{N} (d_j - \frac{\sum_{k=1}^{N} d_k}{N})^2} \tag{1}$$

where x_{ij} is a binary variable indicating if link j belongs to cluster i or not and d_j is the data representing the traffic condition in link j. Note that, N and N_c are numbers of links and desired clusters respectively. However, this metric is a nonlinear non-convex function of decision variables x_{ij}s that unease classical convex optimization formulations.

Hence, we try to find an approximation for TV$_n$ metric that facilitates the optimization process. To do so, we give the algorithm a bunch of candidate values for clusters' mean values and let the algorithm chooses proper values among them during the optimization process. In fact, we try to replace the second term in the numerator of Eq. 1 by some predefined candidate values. It is necessary to find a smart way to determine limited number of values (in the range of one dozen) representing approximate cluster mean values. Indeed, the values depend on the traffic data (density, speed, occupancy, etc.), number of desired clusters, and the spatial distribution of congestion along the network; however, we can have a proper approximation by only considering the data distribution as congestion is spatially correlated in the network. In this study, given the number of candidate values N_s, we divide the data into N_s quantile classes and compute the mean value of the data between two consecutive quantiles and use it as a candidate value. Fig. 1 depicts the case where the data of a simulated case study (described later in details) is divided into eight quantiles and histogram of the data is schematically presented for one of the quantiles. Mean of the data in each quantile is calculated and utilized as a candidate value for the mean of the cluster. Therefore, we have as many candidate points as the number of quantile classes. It is clear that with this method, more candidate points will be picked from parts of the domain with high probability density values which is consistent with the two following facts: (i) more precision is needed for the parts of the domain with large amount of data; (ii) multiple clusters might exist with close average values in different locations as we have too many links within the same range.

Figure 1: Empirical Cumulative Distribution Function of the link densities divided into eight quantiles



2.2 Compactness metric

To introduce a measure of non-compactness for clusters, we utilize the graph representation of the network. Based on the network structure, a graph $G = (V, E)$ is built in which V and E are sets of nodes and edges respectively. Each road in the network is represented by a node $u \in V$ in a graph and a value of traffic parameter is assigned to that node (i.e. density, speed, etc.). There is an edge between roads, which are connected to each other. Two links are spatially connected if they have a common intersection. Since essentially traffic congestion propagates from the adjacent roads in the network, we opt for the shortest path distance between the roads rather than the Euclidean distance. It should be noted that there is a clear difference between compactness and connectivity of the clusters in the terminology of graph theory. A sub-graph is called connected when there exists a path between each pair of nodes through the existing edges. To best our knowledge, there is no unique definition for compactness. A good measure of compactness can be defined as the total distance (TD) of all the roads to the center of their own clusters. Note that center of a cluster is a node or set of nodes with minimum average distance to all the nodes in that cluster. Centers of the clusters are specified by the algorithm within the optimization framework from the whole set of nodes.

2.3 Problem formulation

We first define the sets and indices used to describe the model as well as the variables and parameters; then, detailed mathematical optimization formulations are presented.

Table 1: Decision variables and parameters of MILP

Decision variables	
y_j	Distance of node 'j' to the center of its cluster
x_{ijk}	Binary variable indicating if node k is connected to the center of cluster 'i' via its neighboring node j or not
x_{ioj}	Binary variable indicating if node j is selected as a center of cluster 'i' or not
Parameters and sets	
$\mathcal{L}$	Set of nodes in the graph (roads in the network)
$\bar{\mu}_i$	i^{th} candidate cluster mean value
$\mathbf{N}(j)$	Set of neighboring links of link j
N, N_s	Number of links and candidate mean values respectively
N_c	Desired number of clusters
N_{min}	Minimum number of links in selected clusters(minimum size)

$$\min_{x_{ijk}, y_j} \alpha \times \left(\sum_{i=1}^{N_s} \sum_{j=1}^{N} \sum_{k \in \mathbf{N}(j)} x_{ikj} \times (d_j - \bar{\mu}_i)^2 \right) + (1 - \alpha) \times \sum_{j=1}^{N} y_j \tag{2}$$

$$\sum_{i=1}^{N_s} \sum_{m \in \mathbf{N}(j)} x_{imj} = 1 \quad \forall j \in \mathcal{L} \tag{3}$$

$$x_{ijk} \leq x_{ioj} + \sum_{m \in \mathbf{N}(j), m \neq k} x_{imj} \quad \forall i = \{1, \ldots, N_s\}, \forall j \in \mathcal{L}, \forall k \in \mathbf{N}(j) \tag{4}$$

$$y_k \geq y_j - \left((1 - \sum_{i=1}^{N_s} x_{ijk}) \times N \right) + 1 \quad \forall k \in \mathcal{L}, j \in \mathbf{N}(k) \tag{5}$$

$$\sum_{i=1}^{N_s} \sum_{j=1}^{N} x_{ioj} = N_c \tag{6}$$

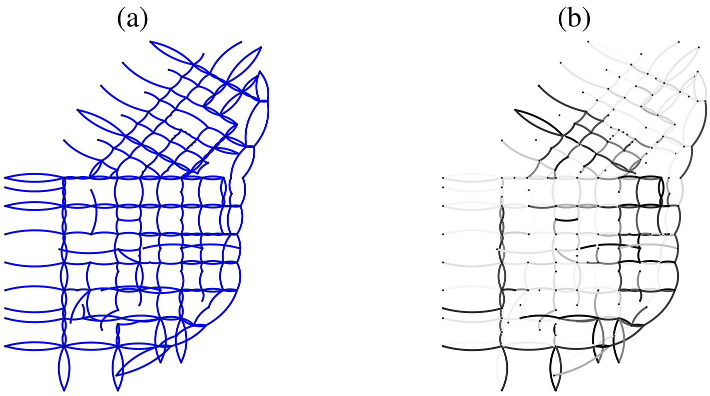
$$\sum_{j=1}^{N} x_{ioj} \leq 1 \tag{7}$$

$$N_{min} \times (\sum_{j=1}^{N} x_{ioj}) \leq \sum_{j=1}^{N} \sum_{m \in \mathbf{N}(j)} x_{imj} \tag{8}$$

$$x_{ijk} \in \{0,1\} \qquad 0 \leq y_j \leq N \tag{9}$$

The objective function expressed in Eq. (2), includes both measures of heterogeneity and non-compactness with weights α and $1-\alpha$ respectively. As you can see, the objective function is linear in terms of decision variables x_{ikj} and y_j. The most challenging part is to model connectivity in graphs and enforce clusters to be connected. Given a connected sub-graph with n nodes and an arbitrary node j in that set, we can always find at least one ordered tree connecting all the nodes in that sub-graph with j as its root node. All the nodes in the graph are connected to the center of cluster (root node) through a sequence of neighboring nodes. Variable x_{imj} indicates whether link j is connected to the center of cluster i through its neighboring link m or not. Note that, $x_{imk} = 1$ also implies that nodes m and j both belong to cluster i. In this case, node m is considered as a parent node for child node j. Constraint (3) enforces that each node only belongs to one cluster and it has one and only one parent node. Constraint (4) establishes that node k can be connected through link j to the center only if link j is already connected to the center. Decision variable y_j can be interpreted as the level of node j which represents how far that node is from the center. Apparently, for each node, this value depends on its path to the center. Constraints (5) insures that the level of a child node should be at least one unit bigger than its parent. Therefore, y_j represent the graph distance (i.e. number of edges in the obtained ordered tree) between node j and the center of its cluster. Since number of candidate points N_s exceeds number of desired clusters N_c, we enforce to have as many center nodes as N_c in constraints (6)- (7). Constraints (8) enforce the size of the clusters to be bigger than or equal to a predefined value N_{min}.

Figure 2: (a) The test site of San-Francisco (b) Grey-scale representation of link densities in the test site of San-Francisco.



3 Case study and priliminary results

3.1 Network Description

The case study is a 2.5 square mile area of Downtown San Francisco (Financial District and South of Market Area), including about 100 intersections with link lengths varying from 400 to 1300 ft. The number of lanes for through traffic varies from 2 to 5 lanes and the free flow speed is 30 miles per hour. Traffic signals are all multiphase fixed-time operating on a common cycle length of 100 s for the west boundary of the area (The Embarcadero) and 60 s for the rest. A 4hr time-dependent traffic demand (120 time intervals of 2 min) is applied to this network, which produces different spatial and temporal levels of congestion. We select a time at which there is a high range of average density values and sub-regions with different level of congestions could be easily seen in different locations. Furthermore, directional congestion could be observed in some parts of network with bidirectional roads so we are able to investigate the ability of dealing with directional congestion. Fig. 2a and 2b illustrate the network and level of congestion in a certain time in a gray scale format. Note that arc represents different direction in 2 way roads. In Fig. 2b, it could be easily seen that the network is heterogeneous with different levels of congestion. Moreover, there are some bi-directional roads with only one congested direction, which will facilitate to test the performance of model for detecting directional congestion.

Table 2: Heterogeneity and non-compactness measure for different clustering results (2-4 clusters)

Number of clusters/weight		TV_n	TD
Two clusters	$\alpha = 1$	**0.1700**	3129
	$\alpha = 0$	0.9803	**1671**
Three clusters	$\alpha = 1$	**0.1452**	3817
	$\alpha = 0$	0.9519	**1331**
Four clusters	$\alpha = 1$	**0.1440**	2601
	$\alpha = 0$	0.9455	**1187**

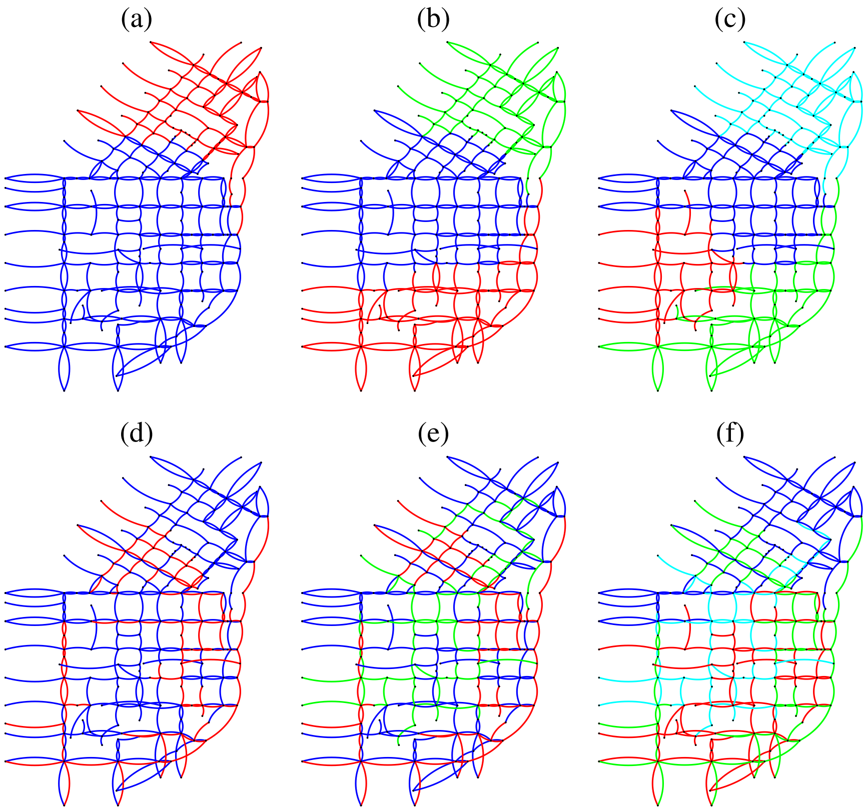
3.2 Partitioning results

In this part, the proposed clustering framework is applied in the part of San-Francisco network. To obtain candidate mean values, the data is divided into eight quantiles using the approach explained above. Efficiency of the partitioning algorithm has been tested for the cases with 2, 3 and 4 clusters. For each of these two cases, the algorithm runs two times considering different weights ($\alpha = 0, 1$) in the objective function. The results of the clustering for different cases are depicted in Fig. 3 and measures of both heterogeneity and non-compactness are calculated and presented in Table 2 for different cases. Figures(3.a, 3.b, 3.c) are corresponding to the case where weight $\alpha = 0$ and we only care about maximizing compactness while Fig.(3.d, 3.e, 3.f) depict the other extreme case where we minimize the heterogeneity in the clusters. By comparing theses results, we can easily conclude that enforcing connectivity is not a sufficient condition for having clusters with compact shape and smooth boundaries. Hence it is necessary to incorporate compactness index in the objective function. As you could see from the obtained values in Table 2, by increasing the weight α from 0 to 1, homogeneity increases while compactness decreases; bold numbers show the best achieved homogeneity and compactness (i.e. minimum TV_n and TD respectively) for different cases with fixed desired number of clusters.

4 Conclusion and future works

This paper presents a method to partition urban traffic networks using link information and network structure. The significance of the proposed method is that, it allows the user to make a

Figure 3: Clustering results for 2, 3 and 4 clusters with $\alpha = 0$ (a-c) and $\alpha = 1$ (d-f).



trade-off between compactness and homogeneity of the clusters which are usually conflicting objectives. Moreover, it explicitly imposes the connectivity, which makes the application of control strategies feasible. The proposed method has been applied to the part of the network of San Francisco with two different weight values defining the importance ratio of compactness to homogeneity. The results indicate the big difference between connectivity and compactness and the necessity of considering compactness in the optimization framework to have clusters with smoother boundaries. The proposed method can be utilized in perimeter control which works based on the concept of MFDs to improve network performance since homogeneous clusters have low scatter MFDs. Future investigation is needed in the following directions: (i) the way of defining weights which can be done by running optimization framework for more combination of weights and obtaining Pareto frontier; (ii) the way to have a more accurate estimation of candidate average values for the clusters. As a future work, it would be challenging to extend the framework from static to dynamic partitioning by integrating time in the optimization framework.

5 References

Buisson, C. and C. Ladier (2009) Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams, *Transportation Research Record*, **2124**, 127–136.

Carlos, J. D., R. Ramos and J. Surinach (2007) Supervised regionalization methods: A survey, *International Regional Science Review*, **30** (3) 195–220.

Gayah, V. V. and C. F. Daganzo (2011) Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability, *Transportation Research Part B*, **45** (4) 643–655.

Geroliminis, N. and C. F. Daganzo (2008) Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, *Transportation Research Part B: Methodological*, **42** (9) 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.

Guo, D. (2008) Regionalization with dynamically constrained agglomerative clustering and partitioning (redcap), *International Journal of Geographical Information Science*, **22** (7) 801–823.

Haddad, J., M. Ramezani and N. Geroliminis (2013) Cooperative traffic control of a mixed network with two urban regions and a freeway, *Transportation Research Part B: Methodological*, **54**, 17 – 36.

Helbing, D. (2001) Traffic and related self-driven many-particle systems, *Rev. Mod. Phys.*, **73**, 1067–1141.

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

Keyvan-Ekbatani, M., A. Kouvelas, I. Papamichail and M. Papageorgiou (2012) Exploiting the fundamental diagram of urban networks for feedback-based gating, *Transportation Research Part B: Methodological*, **46** (10) 1393 –1403.

Kouvelas, A., M. Saeedmanesh and N. Geroliminis (2015) Feedback perimeter control for heterogeneous urban networks using adaptive optimization, paper presented at the *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 882–887.

Mahmassani, H. S., M. Saberi and A. Zockaie (2013) Urban network gridlock: Theory, characteristics, and dynamics, *Transportation Research Part C: Emerging Technologies*, **36**, 480–497.

Saberi, M., H. Mahmassani, T. Hou and A. Zockaie (2014) Estimating network fundamental diagram using three-dimensional vehicle trajectories, *Transportation Research Record: Journal of the Transportation Research Board*, **2422**, 12–20.

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

Saeedmanesh, M. and N. Geroliminis (2016b) An MILP formulation for contiguity-constrained partitioning of heterogeneous urban networks, paper presented at the *Transportation Research Board 95th Annual Meeting*, no. 16-4071.