Transportation Research Part B 91 (2016) 250–269

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Clustering of heterogeneous networks with directional flows based on “Snake” similarities

Mohammadreza Saeedmanesh, Nikolas Geroliminis*

Urban Transport Systems Laboratory (LUTS), School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne (EPFL), GC C2 390, Station 18, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 23 December 2015
Revised 27 April 2016
Accepted 17 May 2016
Available online 4 June 2016

Keywords:
Graph partitioning
Spatiotemporal correlation
Macroscopic fundamental diagram
Non-negative matrix factorization
Missing data

ABSTRACT

Aggregated network level modeling and control of traffic in urban networks have recently gained a lot of interest due to unpredictability of travel behaviors and high complexity of physical modeling in microscopic level. Recent research has shown the existence of well-defined Macroscopic Fundamental Diagrams (MFDs) relating average flow and density in homogeneous networks. The concept of MFD allows to design real-time traffic control schemes specifically hierarchical perimeter control approaches to alleviate or postpone congestion. Considering the fact that congestion is spatially correlated in adjacent roads and it propagates spatiotemporaly with finite speed, describing the main pockets of congestion in a heterogeneous city with small number of clusters is conceivable. In this paper, we propose a three-step clustering algorithm to partition heterogeneous networks into connected homogeneous regions, which makes the application of perimeter control feasible. The advantages of the proposed method compared to the existing ones are the ability of finding directional congestion within a cluster, robustness with respect to parameters calibration, and its good performance for networks with low connectivity and missing data. Firstly, we start to find a connected homogeneous area around each road of the network in an iterative way (i.e. it forms a sequence of roads). Each sequence of roads, defined as ‘snake’, is built by starting from a single road and iteratively adding one adjacent road based on its similarity to join previously added roads in that sequence. Secondly, based on the obtained sequences from the first step, a similarity measure is defined between each pair of the roads in the network. The similarities are computed in a way that put more weight on neighboring roads and facilitate connectivity of the clusters. Finally, Symmetric Non-negative Matrix Factorization (SNMF) framework is utilized to assign roads to proper clusters with high intra-similarity and low inter-similarity. SNMF partitions the data by providing a lower rank approximation of the similarity matrix. The proposed clustering framework is applied in medium and large-size networks based on micro-simulation and empirical data from probe vehicles. In addition, the extension of the algorithm is proposed to deal with the networks with sparse measurements where information of some links is missing. The results show the effectiveness and robustness of the extended algorithm applied to simulated network under different penetration rates (percentage of links with data).

---

* Corresponding author. Tel.: +41 21 69 32481; fax: +41 21 69 35060.
E-mail addresses: mohammadreza.saeedmanesh@epfl.ch (M. Saeedmanesh), nikolas.geroliminis@epfl.ch (N. Geroliminis).

http://dx.doi.org/10.1016/j.trb.2016.05.008
0191-2615/

1. Introduction

Traffic congestion appears with different shapes and patterns and might propagate in particular directions varying from day to day. The network infrastructure in smart cities, allows for sensing and collecting massive spatiotemporal data, such as the trajectories of many vehicles from navigation devices, which represent proxies for human mobility patterns. This 'big mobility data' provides a unique observatory that can help to understand how congestion develops and evolves, discover hidden patterns and identify models that can contribute in efficient traffic management techniques to improve cities' mobility and accessibility. However, unpredictability of travel behavior and high complexity of accurate physical modeling have challenged researchers to develop realistic microscopic level models that are able to replicate congestion spreading and growth in large traffic networks.

There is a strong understanding and vast literature of congestion dynamics and spreading in one-dimensional traffic systems with a single mode of traffic, e.g. a highway section with cars. Besides traffic scientists, mathematicians and physicists have also contributed to the field of traffic flow modeling. Because of the numerous publications, we refer the reader to Helbing (2001) for an overview. Briefly speaking, the main modeling approaches can be classified as car-following models (e.g. Gazis et al., 1959; Gipps, 1981); cellular automata (e.g. Nagel and Schreckenberg, 1992); gas-kinetic models (e.g. Herman et al., 1972); first-order and higher order flow models such as Lighthill and Whitham (1955), Payne (1971) and Whitham (1974).

Literature in network level dynamics and congestion propagation is limited especially in large urban networks. Previous works have mainly built on micro-simulations of link-level traffic dynamics or graphical visualizations of congestion without any metrics and dynamic models. However, both the unpredictability of travel behaviors and high complexity of accurate physical modeling remain challenging and simulation results may be time consuming and not realistic for dynamic systems with stochastic characteristics. The main characteristic of traffic in urban networks is that congestion is spatially correlated in adjacent roads and it can propagate with some finite speed in time and space. These correlations allow describing the main pockets of congestion in a city with a small number of clusters without the need for detailed information in every link of the network (Ji et al., 2014). However transportation networks have unique dynamic features and a direct application of an arbitrary clustering algorithm may not produce a desired solution.

With respect to network level, it has been observed with empirical and simulated data in Geroliminis and Daganzo (2008), Buisson and Ladier (2009), and Gayah and Daganzo (2011) that by spatially aggregating the highly scattered plots of flow vs. density from individual links (e.g. 1 min data), the scatter almost disappears and a well-defined curve exists between space-mean flow and density. The idea of an MFD with an optimum accumulation belongs Godfrey (1969) and later re-introduced by Daganzo (2007), Herman and Prigogine (1979) and Mahmassani et al. (1987). Geroliminis and Daganzo (2008) showed that the shape of MFD is a property of the network infrastructure and control and not very sensitive to the demand. This is important for modeling purposes, as details in individual links are not necessary to describe congestion in cities. It can also be utilized to introduce simple control strategies to improve mobility in multi-region city centers building on the concept of an MFD, like in Ramezani et al. (2015), Keyvan-Ekbatani et al. (2015), Haddad and Shraiber (2014), and others. A detailed literature review of network modeling and control can be found for instance in Haddad et al. (2013) or Mahmassani et al. (2013). Furthermore, latest works extend the single-mode MFD to a bi-modal where cars and buses share the same infrastructure and look at passenger flow dynamics in addition to vehicular dynamics (Chiabaut et al., 2014; Zheng and Geroliminis, 2013), and (Chiabaut, 2015). Estimating MFDs with different type of data (loop detectors, probe vehicles or a combination) are investigated in Leclercq et al. (2014), Du et al. (2015), Ji et al. (2014) and others.

Recent findings from empirical and simulated data Geroliminis and Sun (2011), Mazloumian et al. (2010), Daganzo et al. (2011), and Knoop and Hoogendoorn (2013) have identified the spatial distribution of vehicle density in the network as one of the key components that influence the shape and the scatter of an MFD. These findings are of great importance because the concept of an MFD can be applied for heterogeneously loaded cities with multiple centers of congestion, if these cities can be partitioned into a number of homogeneous clusters. The objectives of partitioning are to obtain (i) small variance of road densities within a cluster, which increases the network flow for the same average density and (ii) connectivity and spatial compactness[1] of each cluster which makes the application of perimeter control strategies feasible. The proposed mechanism in this paper can produce a partitioning with a desired number of clusters that contain connected roads with small density variance. Furthermore, this mechanism demonstrated superiority of both effectiveness and robustness in comparison with standard clustering algorithms.

There is a vast literature on studying clustering algorithms in several fields such as community detection (Lancichinetti and Fortunato, 2009), data-mining (Jiawei Han, 2000), and image segmentation (Shi and Malik, 2000). Depending on the type of application, problems related to clustering generally fall into two main categories: unconstrained and constrained clustering. The former does not have any constraints on objects to be in the same cluster (i.e. all the set partitions are feasible) while the latter is limited in search space meaning that some of the clustering assignments are not feasible. In our urban partitioning problem, we look for connected and compact shaped clusters with homogeneous traffic conditions to have low-scatter MFDs, which allows us to utilize perimeter control strategies that work based on the concept of MFD.

---

[1] A measure of compactness can be defined as a summation of distances (shortest path) of all the roads in the cluster from the center of that cluster. The center of a cluster is defined as a road or set of roads with minimum average distance to all the roads in that cluster.