Real-data modelling of transportation networks

Mahsa Faizrahnemoon

Hamilton Institute

National University of Ireland Maynooth

This dissertation is submitted for the degree of

Doctor of Philosophy

April 2016

I would like to dedicate this thesis to my parents and my sister

Acknowledgements

Abstract

In this thesis, after introducing the basics of Markov chains and mathematically analysing and proving the clustering properties of the eigenvector corresponding to a complex eigenvalue close to $1 + 0i$, we develop several applications of Markov chains in transportation networks. We model bike sharing systems, bus networks, and multi-modal public transportation networks using Markov chains. The validation of the models is done by using real data from Boston and London for the bike sharing systems and the multi-modal public transportation networks respectively. We validate the Markov chain models that we developed for the bus network by using some data that we extracted from SUMO [69] (Simulator of Urban MObility). After successfully validating the models, we extract some important quantities of the Markov chains. These quantities provide useful information about the networks that help to improve the network for different purposes. Since some real data is used to validate our models, we need to know how trustworthy our result is. Therefore, we then define a set of indicators to extract the quality of data. The output of each indicator is a value between zero and one. If the value is close to one, it means that the quality of the data is high and we can trust the data and the result. The indicators are tested on some data from London highways. At the end a framework for the real time trading of budgeted emission rights between a fleet of participating vehicles is presented. The trading problem is formulated as an optimization problem and is solved by different algorithms. The results of some simulations are represented to compare the speed of convergence of the algorithms.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Traffic modelling

Traffic modeling is an area that has recently been growing in fields such as applied mathematics, networking and transportation engineering. Most of the traffic models have been based on postulated mathematical descriptions. Unfortunately it is hard to verify such models experimentally. These days, huge amounts of data can be extracted from mobile phones, GPS devices, and from connected cars. Therefore, an alternative description of mobility patterns has been provided. We shall see examples of this in the main body of the thesis. The main objective of traditional traffic theory is to understand the traffic speed, flow, and density. These quantities should be known for an efficient dimensioning of the transport infrastructure. This information can be helpful to resolve traffic problems. These models are mainly used to develop simulation tools in oder to help traffic designers. Different aspects of traffic modeling include (i) Trip modeling from Origin-Destination matrices; (ii) modeling of individual vehicle paths; (iii) driver modeling, and (iv) macroscopic flow modeling and abstractions. In the context of ITS applications – many other aspects have to be studied. This includes modeling the interaction between individual vehicles, and with infrastructure, and the ability to model closed loop behavioural aspects. There are many surveys on this topic – we shall not repeat these here; the interested reader may refer to [39], [12], [38], and [45]. While we shall use traditional models and simulators to validate our hypothesis as we proceed, we shall take a very different approach to these traditional methods. Namely, we shall assume the existence of large scale GPS/SCATS/Mobile phone data sets, and represent these data sets at a macroscopic level using stochastic, in particular, Markovian, techniques. As we shall see, a large variety of ITS systems and networks can be modelled in this manner with a minimum of mathematical and structural assumptions.

1.2   Markov chain models

Most of the previous work involves a microscopic description of mobility. This is good from the point of view of an individual, but often, cities require a macroscopic description. Much of the original work in this thesis takes this viewpoint. In order to develop our macroscopic approach, we wish to take advantage of all the information gathered from instrumenting vehices for V2V and V2I communication. One key feature in modelling is that the models should not only model traffic flows and road dynamics but also that they should be constructed from real data obtained from the road network which is gained in real time. The ability to use vehicles as sensors to harvest information in real time suggests the possibility to deploy new tools to model road networks. One useful tool is that of Markov Chains. Markov chains can easily be built from real data and they are fast and effective simulation tools [32].

1.3   Collaborative mobility

The whole area of traffic modelling, which is the main focus of this thesis, benefits enormously from the vast amounts of data becoming available as a result of advances in the field of collaborative mobility. In particular, this body of data allows for more data-driven models to be constructed and validated than was possible before. Collaborative mobility is rapidly becoming one of the main growth areas in fields such as applied mathematics, operations research, networking and transportation engineering. Due to advances both in technology and in applications based on this technology, collaborative mobility is currently not only driving new research, but also new propositions for cities.

Roughly speaking, collaborative mobility has motivated research into technologies that underpin this new field, and into new services that make use of the capabilities arising from inter-vehicular communications. In this latter category, much focus has been on safety applications, with some more recent attention on environmental applications. The benefits from connected cars and from collaborative mobility are clear to see. For example, active safety applications such as accident prevention applications will clearly benefit from communication between the vehicles. Further, by collecting traffic status data from a wider area, not only the traffic flow could be improved for road users, but also travel times and emissions from vehicles could be reduced [45].

The concept of vehicle to vehicle (V2V) or vehicle to infrastructure (V2I) communication seems to be straightforward, but designing such systems is technically and economically challenging. Some of the challenges are well known and are described in [45]. These include:

- the fact that V2V and V2I communications present scenarios with unfavourable characteristics for developing wireless communications;

- that the fair and efficient use of the available bandwidth of the wireless channel is a difficult task in a totally decentralized network;

- that the challenges of a decentralized self organizing network are particularly stressed by moving vehicles;

- and that applications require a balance between security and privacy needs.

From an economic point of view, the main challenges have been quantifying the cost benefit relationship of V2V and V2I communication, and developing and installing the infrastructure required to support Intelligent Transportation Systems (ITS). Despite these challenges, ITS is now fast becoming a reality and is spawning a host of new services and innovation. We now briefly review some of the historical progress in this area.

1.3.1   History and projects

The history of the use of radio for vehicle to roadside and vehicle to vehicle communication is strongly connected to the evolution of ITS. Surprisingly, even though there is much current interest in this topic, work has been going on in this general area for many decades, and many excellent surveys on this topic now exist [45, 109]. We shall not repeat this here, but rather give a very brief summary on some of the most high profile works (following closely [45]). Surprisingly, as was described in [45], the basic concepts of roadway automation which is communication and control techniques to make road traffic safe, efficient, and environmentally friendly, were exhibited by General Motors in 1939 [116] [72]. Since then dedicated ITS systems have been the subject of research and development efforts around the world for some time. Specific examples in the USA include an Electronic Route Guidance System (ERGS) which was proposed in [107] in 1970; in Japan, the Comprehensive Automobile Traffic Control System (CACS) project was carried out by the Agency of Industrial Science and Technology of the Ministry of International Trade and Industry (MITI) from 1973 to 1979. More recently, in Europe, the topic of ITS has been heavily promoted by the EU with specific programmes including the PROMETHEUS (Program for European Traffic with Highest Efficiency and Unprecedented Safety) programme, and the Green Car

programme [18]. The former of these programmes focussed on vehicle safety, whereas the Green Car programme is concerned principally with ECO mobility - particularly in the context of electric and plug-in-hybrid vehicles. Some other programmes of note include the California Partners for Advanced Transit and Highways (PATH) programme in which platooning was a primary focus. Another recent driver of interest in ITS has come from the area of cooperative driver assistance systems [46]. Some recent specific projects in Europe on such systems include the Cartalk and FleetNet projects [35], both of which investigated technologies and applications for cooperative driver assistance. Examples of similar projects in the US include Cooperative Intersection Collision Avoidance System (CICAS) project; see [81] [78].

In Europe, many recent projects have considered ITS technologies and applications. Examples of other large historical projects include, COOPERS, CVIS, PReVENT, SAFESPOT, DRIVE C2X and TEAM. For example, the project COOPERS (2006-2010) focused on telematics applications for cooperative traffic management; CVIS (2006-2010) mainly focused on development and testing of vehicle to infrastructure communication; SAFESPOT aims to design cooperative systems for road safety based on vehicle to vehicle and vehicle to infrastructure communication; and PReVENT (2006-2008) was about development of preventive safety applications and technologies. See [45] for a description of these projects. Two more recent projects that have been very high profile in Europe include DRIVE C2X and TEAM. DRIVE C2X (http://www.drive-c2x.eu/project) is very much concerned with the technologies, testing and technologies underpinning ITS, whereas TEAM is very much applications focussed. Of particular relevance to the work in this thesis is the project TEAM and the work presented here is part of the TEAM project. The goal of this project is to optimize facets of traffic networks. TEAM is a large European project. FIAT, BMW, NOKIA, VOLVO, INTEL, and NEC are some of the industrial companies involved in TEAM. Some of the research in TEAM has been done by National University of Ireland Maynooth, TU Berlin, and Fraunhofer FOKUS.

As the interested reader is no doubt aware, many papers and surveys have been completed on this topic; see [109] for example. We shall not repeat this here. However, we shall mention some of the main application areas to benefit from ITS developments, and innovations in collaborative mobility.

1.3.2   Applications

To-date applications of V2V technologies can be classified as follows [61]:

(1) Road safety applications;

(2) Traffic efficiency and management applications;

(3) Infotainment applications.

(4) Pollution control and environment

   As the reader is no doubt aware, road safety applications are intended to reduce the rate of traffic accidents. In spite of the fact that the vehicles are more advanced in terms of safety devices like anti-lock brakes and airbags, the number of fatalities on the roads is still too large. It seems that the combination of human behaviour (drivers) and the increased functionality of vehicles brings about risky driving leading to many accidents worldwide. A good overview of such applications can be found in [117], [10], [29], [60], [108], [22]. Most of the accidents in road vehicles occur at intersections and result in head-on, rear-end and lateral vehicle collisions. The basic idea in road safety applications is to transfer information between the vehicles and the road side units to alert drivers of problems (such as ice, potholes, congestion, or other accidents) so that evasive action can be undertaken either by the driver or by an active safety system. Currently, active safety systems rely on real-time feedback from sensors in vehicles or located in dedicated roadside infrastructure. By increasing the accuracy positioning and the technology of communication between the vehicles, cooperative driving applications can introduce anticipatory behaviour within safety systems; see [37], [98] and [95]. In particular, anticipatory systems can discourage the drivers from entering dangerous situations. There are many other advantages in communication-enabled cooperative applications. A non-comprehensive list includes the enforcement of (possibly varying) speed limits, sharing travel times to improve the accuracy of forecasts in car navigators, safer implementation of some car manoeuvers to mitigate the risk of collisions, such as communication aided intersection approaching or lane changing. The interested reader is referred to [61], [10], [108], [29], [22], [125], [126] to see projects/programmes that develop such ideas. Examples of systems that already exist include the following.

(1) *Intersection collision warning systems*

(2) *Lane change assistance systems*

(3) *Overtaking vehicle warning systems*

(4) *Head on collision warning systems.*

(5) *Emergency vehicle warning:*

(6) *Emergency electronic brake lights*

7) *Collision risk warning*

There are many others. Another area to benefit greatly from ITS applications is the area of traffic management. Traffic efficiency and management applications mainly focus on improving the vehicle traffic flow. Often, such applications use ITS technologies to provide updated local information and maps to vehicle fleets. According to [29] and [61] two main groups of applications in this space are speed management and cooperative navigation. Other novel applications of the traffic framework suggested in [19] are discussed in [18], [109], [111], [32], [31], [110], and [20]. In these references, they apply the framework of [19] to model traffic induced pollution and to model the consumption of battery power in electric vehicles. Both TEAM and the project GREEN TRANSPORTATION NETWORKS (www.smarttransport.ucd.ie) are focussed on such issues and show how pollution may be managed through emissions sharing between vehicles based on V2V. Finally, infotainment applications have also benefitted greatly from ITS technologies. These include global internet services. Applications developed using in-car internet services and connectivity include in-car TV, internet, email and fleet management and insurance services [10], [29], [22], [21]

1.4 Markov chains applications

During last decade Markov chains are becoming more appealing in many applications because of their capabilities. In many applications the statistical properties of Markov chains are important and in some applications the randomness is useful. A theoretical property which is useful in some applications is the existence of stationary distribution, which will be later explained in this thesis. Markov chains are particularly suited to big-data (large data sets that traditional data processing applications are not enough to deal with) applications, as found in transport, for several reasons:

- *(i)* Microscopic behaviour is embedded into the chain through aggregation; namely, in the form of probabilities. These probabilities are easily measured or calculated (turning probabilities, bus occupancies, bike pickup and delivery data) without need for large data processing or storage capabilities;

- *(ii)* Many of the key properties (e.g., Perron eigenvectors and Mean First Passage Time matrices) of a Markov chain can be calculated in a recursive fashion using simple update formula [70]. The suitability of Markov chains for big-data application is

discussed, for example, in the context of Google's PageRank algorithm [71]. Well-established and robust algorithms are available to handle data-sets of the size of thousands, if not millions, of web-pages that might contain the relevant information pertaining the user's query. Some examples at this regard are given in the remainder of the section;

- *(iii)* Many of the properties of the chain correspond to real quantities of interest to network designers. We shall have more to say on this in the next sections;

- *(iv)* The suitability of Markov chains for capturing and modelling complex dynamics is discussed and justified, among others, in [109] and [36].

An innovative data-driven framework for transport which was based on the mathematical theory underlying the well-known Google's PageRank algorithm has recently been proposed in [19] to describe urban vehicular traffic. In that reference, graph theory and Markov chain ideas were used to reveal non-trivial patterns of urban mobility and to support transportation experts with practical tools to solve a number of mobility applications. The contribution in [19] is to propose a new paradigm for modelling road network dynamics. The authors in [19] use Markov chains to explore properties of road networks. These include: identifying the sensitive links in the network; measuring how connected the network actually is (graph connectivity, sub-communities); the design of networks that are in some sense maximally mixing; and the ability to predict the consequences of a situation like if a link fails due to road works or any unexpected event like an accident. Such information cannot be extracted easily from conventional simulators. The work of [19] was well received by the transport community giving rise to further research along the same lines, see [18],[111], [130], [1], [4] and [89].

In [4] it has been shown that the approach used in [19] still works when the scale of the network is enlarged and more variables are added to the roads. Different lengths, speed limits, amount of lanes and priority regulations all play a part in everyday road networks and they are taken into account in [4]. The network considered in [4] to validate the approach of [19] is 'De Uithof' which is a district in the city of Utrecht in Netherlands, see Figure 1.1. The result shows that the idea of [19] is useful for a more complicated and bigger network.

The fact that huge amount of taxi cabs are equipped with GPS can be considered as a great potential to investigate the dynamic of the city through the mobile sensors, but the bias or sufficiency of this data for city wide traffic modeling could be an issue which should be taken into account. However, a recent research [105] shows that GPS traces of taxis can be
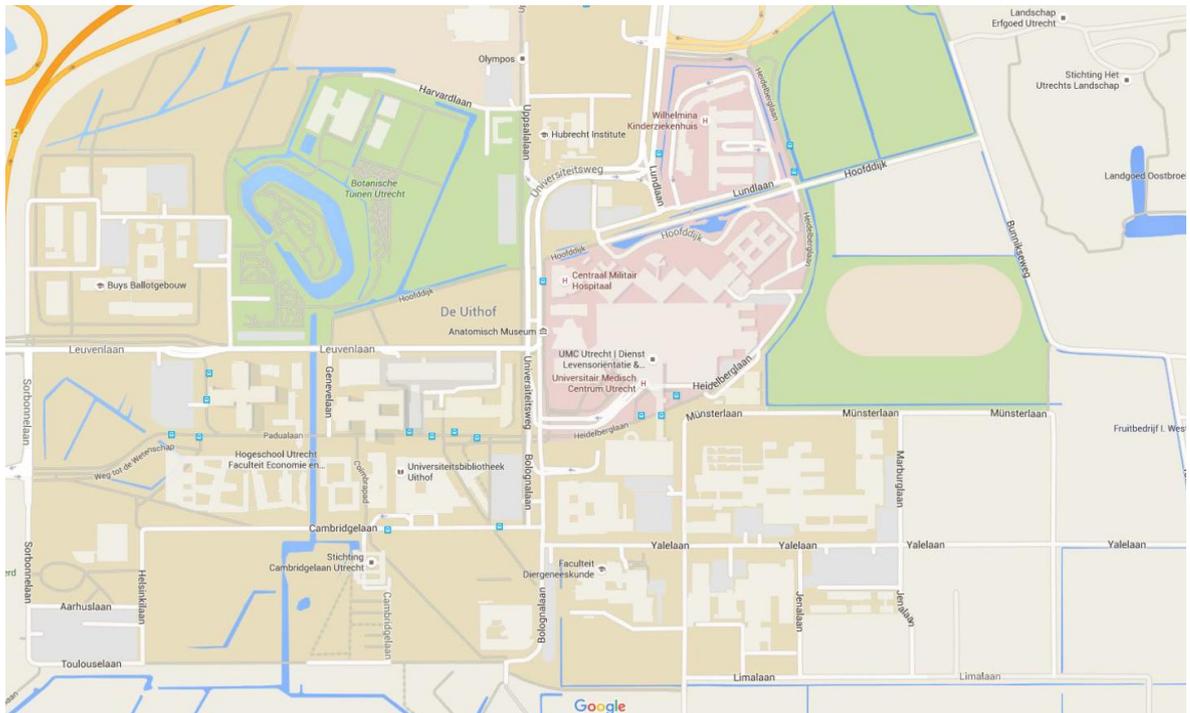
Fig. 1.1 De Utihoft in Google maps

used to approximate traffic patterns in a city-scale road network in a proper way. Several researches, for example [106] [133] [77] [132], have been done by using GPS traces. Some people used Markov chains as their tools to use GPS traces. For example in [54] it is shown that it is possible to predict the path between two points in the road network accurately by using a Markov chain approach. Then they conclude that this prediction can be used for traffic simulation. In some research such as [129] and [122] vehicle trajectories in urban road networks have been predicted. According to [19], Markov chains are strong mathematical tools that have some other properties that can be used to model citywide traffic dynamics. In [88] they propose a method to construct a time-dependent and scalable Markov chain from GPS traces. They implemented the framework based on a released data set of taxi cabs in Beijing from Microsoft Research Asia [132]. The proposed conceptual framework of [88] is represented in Fugure 1.2. In their approach, a Markov chain is constructed periodically in coexistence with a continuous data stream, emitted from taxi cabs as a part of the everyday life in a city. It can also be seen that a Markov chain is used for different segments of road network, different time periods, and different time resolutions depending on the frequency rate of GPS traces. Then, several properties of each constructed Markov chain like community detection, time-dependent expected travel times, real time path planning and road network engineering can be calculated. This kind of information is not easy to achieve directly from data. These result can be helpful in planning interventions in the urban network. They extract the expected car density in the network, they identify the critical road segments, they detect spatiotemporal communities and they measure some other important information.

1.4.1   Open questions

Although a lot of work has already been done in the area of using Markov chain techniques in transportation networks, there are still plenty of problems to consider in this field.

Markov chains have already been used to model road networks, they have also been applied in emission control. One of the problems to solve is to use the Markov chains to model public transport networks and bike sharing networks. In Chapters 2 and 3 of this thesis, we explain how to build transition matrices of Markov chains for bus networks and bike sharing networks.

The bigger question is to use Markov chains to model multi-modal networks. It can be a network of buses, tube, people, etc. The Markov chain model should be able to handle different types of transport at the same time because the reality is that people may take different types of public transportation in their journey. In Chapter 3 we propose a multi-modal transportation model which is validated by real data.
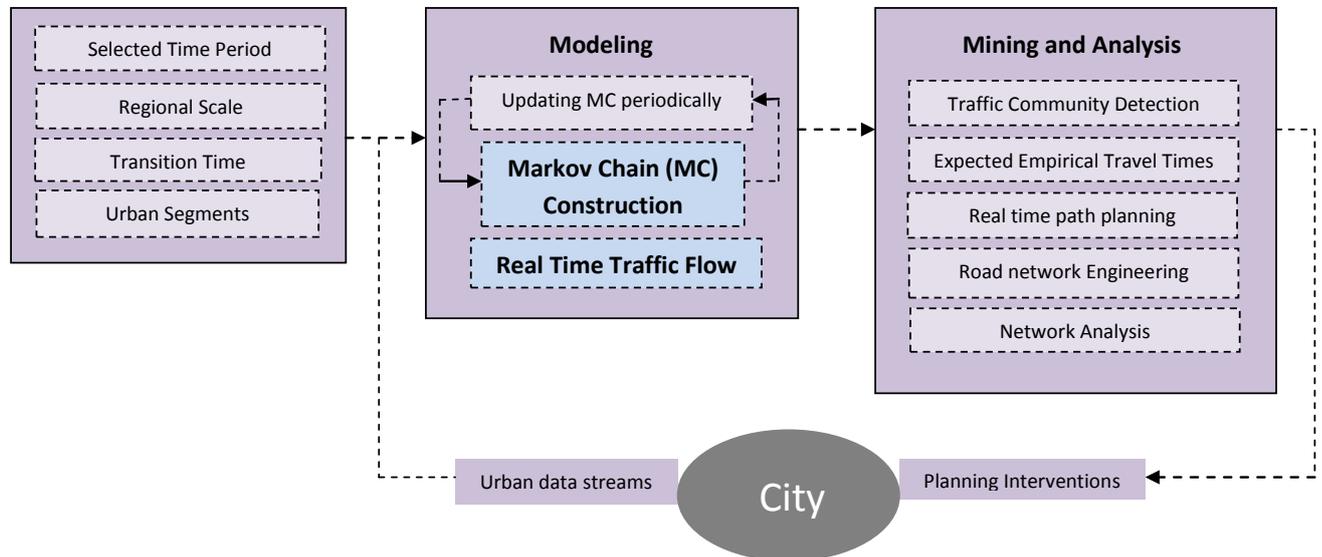
Fig. 1.2 A conceptual framework for Markovian modelling of traffic flow in coexistence with urban data streams. This figure is taken from [88]

Finally, tools for assesing data quality and new tools to analyze Markov chains, arising from Markov chains in transportation, are urgently needed.

1.4.2 A final comment

In our approach we use SUMO (Simulator of Urban MObility) which is a free open source traffic simulator and has been available since 2001 [69]. SUMO is able to simulate road vehicles, public transport and even pedestrians. It has strong supporting tools that makes it possible to handle tasks like route finding, visualization, emission calculation, extracting data, etc.

1.5 Contribution and publications

Apart from chapters 1 and 6 which are the introduction and the conclusion of the thesis my main contributions throughout the rest of the thesis are listed in the following paragraph.

In chapter 2 using Markov chain techniques to model bike network was my idea. In chapter 3, my main contribution was using the model for a multi-modal transportation

network but I did some simulations in the other parts like the SUMO simulation to extract the data for the bus network. In chapter 4, my main contribution was to use the bootstrap method to extract the quality of traffic data and in chapter 5, speeding up the rate of convergence by getting more information from the center was my idea.

I list below the conference and journal papers arising from the work of this thesis and Table 1.1 lists some of the major notational and terminological conventions adopted throughout the thesis.

1. Conference papers

    (a) M. Faizrahnemoon, F. Alesiani, L. Matias, "A Scenario-Oriented approach for Noise detection on Traffic Flow data", 18th IEEE International Conference on Intelligent Transportation Systems (ITSC), 2015.

    (b) M. Faizrahnemoon, A. Schlote, E. Crisostomi, and R. Shorten,"A Markov Chain based model for a bike sharing system", International Conference on Connected Vehicles and Expo(ICCVE), 2015.

    (c) M. Faizrahnemoon, A. Schlote, E. Crisostomi, and R. Shorten, "A Google-like model for public transport", International Conference on Connected Vehicles and Expo(ICCVE), 2013.

    (d) F. Häusler, M. Faizrahnemoon, E. Crisostomi, A. Schlote, I. Radusch, and R. Shorten, "A framework for real-time emissions trading in large-scale vehicle fleets", 9th ITS European Congress, 2013 (winner of the best scientific paper award).

2. Journal papers

    (a) M. Faizrahnemoon, A. Schlote, L. Maggi, E. Crisostomi, and R. Shorten, "A big-data model for multi-modal public transportation with application to macroscopic control and optimisation", International Journal of Control, 2015.

    (b) F. Häusler, M. Faizrahnemoon, E. Crisostomi, A. Schlote, I. Radusch, and R. Shorten, "A framework for real-time emissions trading in large-scale vehicle fleets", IET Intelligent Transport Systems,vol. 9, Issue. 3, pp. 275-284, 2015.

Notation	meaning
MFPT	Mean First Passage Time
m_{ij}	entries of the means first passage time matrix
MC	Markov Chain
$\mathbb{P}$	transition probablity matrix
$\mathbb{P}_{i,j}$	blocks of matrix $\mathbb{P}$
$\mathbb{P}_{ij}$	entries of matrix $\mathbb{P}$
π	Perron eigenvector
K	Kemeny constant
ITS	Intelligent Transportation System
ρ	spectral radius
$\alpha + i\beta$	complex eigenvalue
$X + iY$	complex eigenvector

Table 1.1 Table of the main abbrevations and notations in the thesis

Chapter 2

Bike-sharing: An example of Markov modelling with constant density

Abstract

The first objective idea of this Chapter is to provide the basic notions of Markov chains. The basic definitions can be found in classic books such as [84] and [96] and their properties can be found in classic references like [71] and [62]. The second objective is to prove clustering properties of the second eigenvector of the transition matrix. We extend the proof in [19] to the case in which there exists a complex eigenvalue. This part of the work was done when I was visiting Professor Stephen Kirkland at the University of Manitoba. It is a joint work with him, Jane Breen, Emanuele Crisostomi, and Robert Shorten. The third objective of this Chapter is to explain how Markov chains can be used in practice to model a transportation network. The main references of this part are [19] [88] [4]. Finally, the last part of the chapter describes a practical application of the previously described approach. In particular, we propose a Markov chain based metodology to model a bike-sharing system. The approach is particularly attractive for this specific application, given that, differently from the prior applications, the total number of vehicles (i.e., bikes) is now constant in the whole mobility network. The proposed methodology is validated on real data from the bike-sharing system is Boston, USA, and an application of the model is illustrated. This work is a joint work with Emanuele Crisostomi, Arieh Schlote and Robert Shorten and has been presented at the ICCVE Conference 2015.

2.1   Basics of Markov chains

Throughout this thesis only discrete time, finite state, homogeneous Markov chains are considered. In this situation, a Markov chain is a discrete time stochastic process $x_k, \quad k \in \mathbb{N}$ and characterized by the equation

$$
\begin{aligned}
p(x_{k+1} = S_{i_{k+1}} | x_k = S_{i_k}, ..., x_0 = S_{i_0}) = \\
p(x_{k+1} = S_{i_{k+1}} | x_k = S_{i_k}) \qquad \forall k \geq 0,
\end{aligned}
\tag{2.1}
$$

A Markov chain with n states is completely described by the $n \times n$ transition probability matrix $\mathbb{P}$, whose entry $\mathbb{P}_{ij}$ denotes the probability of passing from state S_i to state S_j in exactly one step. $\mathbb{P}$ is a row-stochastic non-negative matrix, as the elements in each row are probabilities and they sum up to 1[84].

2.1.1 Markov chains and graph theory

A graph is a set of nodes (corresponding to the states of the Markov chain) that can be connected by edges or arcs. We will only consider directed arcs in our approach. To be more accurate let us consider a finite set of n nodes, then for each pair $i, j = 1, ..., n$ an edge from node i to node j indicates that it is possible to make a direct transition from state i to state j in the Markov chain. The case $i = j$ is called a self-loop arc of the graph.

It is possible to give a weight to each arc that corresponds for example to the cost of using that edge. If the aggregate cost of all outgoing arcs of each node is normed to 1 then the costs can be interpreted as the probabilities of using the corresponding arcs to go from the node which is the origin of the arc to the node which is the destination of the arc.

There is a strong link between Markov chains with finite state space and graphs that makes the interpretation of the Markov chain concept easier. States of the chain can be associated with nodes of the graph and non-zero probabilities of transition between the two states S_i and S_j in the chain can be associated with the directed edge (i, j) which is the edge between the corresponding nodes with the given probability as a weight, see Figure 2.1 for example. Therefore, graphs can be analysed using methods for Markov chains.

$$A = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$
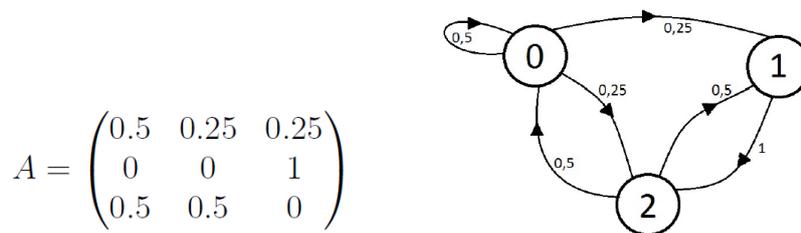
Fig. 2.1 An example of a transition matrix and its corresponding graph.

An important property of a graph is connectivity. A directed graph is called strongly connected if starting from any node of the graph it is possible to reach any other node of the graph by following the arcs. Strong connectivity holds if and only if the transition matrix of the Markov chain corresponding to the graph is irreducible. Throughout this part of the

thesis we shall assume that all considered Markov chain transition matrices are irreducible. This allows us to apply the Perron-Frobenius theorem, see for example [53] and [71]:

- The spectral radius of $\mathbb{P}$ is 1;

- 1 also belongs to the spectrum of $\mathbb{P}$, and it is called the Perron root;

- The Perron root has an algebraic multiplicity of 1;

- The left-hand Perron eigenvector π is the unique vector defined by $\pi^T \mathbb{P} = \pi^T$, such that every single entry of π is strictly positive and $\|\pi\|_1 = 1$. Except for positive multiples of π there are no other non-negative left eigenvectors for $\mathbb{P}$.

One of the main properties of irreducible Markov chains, that will be used and explained later in chapters 3 and 4 of this thesis, is that the i'th component π_i of the vector π represents the long-run fraction of time that the chain will be in state S_i. The row vector π^T is also called the stationary distribution vector of the Markov chain.

2.1.2 Mean first passage times and the Kemeny constant

The mean first passage time and the Kemeny constant are two quantities of Markov chains which are related to random walks on the graph corresponding to the transition matrix of the Markov chain. A random walk on a graph is a discrete time stochastic process $S(k)$, $k \geq 0$. The state space of the random walk is the nodes of the graph. Starting in an initial node, the random walk in every time step follows one of the edges leading out of the node it is currently in, where each arc is chosen randomly according to the probability of that edge. If $S(0) = i$, it means that the walk starts from node i, then the first time that the walk reaches node j is called the first passage time from i to j and it can be computed as

$$\min_{k \geq 0}\{k : S(k) = j\}, \tag{2.2}$$

The expected value of the first passage time is called the mean first passage time. The mean first passage time from node i to node j will be shown as m_{ij} in this thesis and can be computed as

$$m_{ij} = P_{ij} + \sum_{k \neq j} P_{ik}(1 + m_{kj}) = 1 + \sum_{k \neq j} P_{ik} m_{kj}. \tag{2.3}$$

The formula in (2.3) indicates that with probability P_{ij} it takes exactly one step to go from i to j and for $k \neq j$ the first step takes the random walk from node i to node k from which it takes an expected time of m_{kj} to reach node j.

One way of computing the mean first passage time arises from the study of the group inverse of a matrix. A transition matrix $\mathbb{P}$ with 1 as a simple eigenvalue gives rise to a singular matrix $\boldsymbol{I} - \mathbb{P}$ (where the identity matrix $\boldsymbol{I}$ has appropriate dimensions) which is known to have a group inverse $(\boldsymbol{I} - \mathbb{P})^{\#}$. The group inverse is the unique matrix such that $(\boldsymbol{I} - \mathbb{P})(\boldsymbol{I} - \mathbb{P})^{\#} = (\boldsymbol{I} - \mathbb{P})^{\#}(\boldsymbol{I} - \mathbb{P})$, $(\boldsymbol{I} - \mathbb{P})(\boldsymbol{I} - \mathbb{P})^{\#}(\boldsymbol{I} - \mathbb{P}) = (\boldsymbol{I} - \mathbb{P})$, and $(\boldsymbol{I} - \mathbb{P})^{\#}(\boldsymbol{I} - \mathbb{P})(\boldsymbol{I} - \mathbb{P})^{\#} = (\boldsymbol{I} - \mathbb{P})^{\#}$. More properties of group inverses and their applications to Markov chains can be found in [82]. The group inverse $(\boldsymbol{I} - \mathbb{P})^{\#}$ contains important information on the Markov chain and it will be often used in this thesis. For this reason, it is convenient to denote this matrix as $Q^{\#}$. The mean first passage time (MFPT) m_{ij} from the state S_i to the state S_j denotes the expected number of steps to arrive at destination S_j when the origin is S_i, and the expectation is averaged over all possible paths following a random walk from S_i to S_j. If we denote by $q_{ij}^{\#}$ the ij entry of the matrix $Q^{\#}$, then the mean first passage time can be computed as

$$m_{ij} = \frac{q_{jj}^{\#} - q_{ij}^{\#}}{\pi_j} \qquad i,j = 1,...,n, i \neq j. \tag{2.4}$$

The diagonal entries of the mean first passage time matrix are $m_{ii} = \frac{1}{\pi_i}$.

If the random walk starts from node i to destination j which is chosen randomly according to the stationary distribution of the Markov chain, π, we can compute the expected number of steps to reach the destination, Kemeny constant, according to the following formula

$$K = \sum_{j=1}^{n} m_{ij}\pi_j \tag{2.5}$$

where the right-hand side is independent of the choice of the origin state S_i. This is a very important well known result from [62]. An interpretation of this result is that the expected time to get from an initial state S_i to a destination state S_j (selected randomly according to the stationary distribution π) does not depend on the starting point S_i [27]. Therefore, the Kemeny constant is an intrinsic measure of a Markov chain and it can be interpreted as the expected duration of an average trip. The Kemeny constant can be viewed as an efficiency indicator for a network.

If the transition matrix $\mathbb{P}$ has eigenvalues $\lambda_1 = 1, \lambda_2, ..., \lambda_n$, then another way of computing K is [74]

$$K = \sum_{j=2}^{n} \frac{1}{1 - \lambda_j}. \tag{2.6}$$

As can be seen from Equation (2.6), K is only related to the particular matrix $\mathbb{P}$. The value of K is not too large if the eigenvalues $\lambda_2,...,\lambda_n$ are weel-separated from 1 and it increases if at least one eigenvalue is close to 1.

2.2 Clustering properties of the second eigenvector

In [19], the authors give broader meaning than just the spectrum of the matrix to the term spectral analysis for directed graphs and use the group inverse of the transition matrix of the Markov chain in their approach. The group inverse captures important information on the Markov chain, see [83]. For example, the mean first passage time, m_{ij}, which is the expected number of steps to arrive at destination state S_j from state S_i, is computed by using the group inverse according to the formula (2.4) which is also represented in [17]. In [19] the authors use the formulation of the Kemeny constant, which is an intrinsic measure of a Markov chain, in [74] to show that when the eigenvalues of the transition matrix approach to 1, the Kemeny constant increases. It should be noticed that in [19], no assumption is put on the transition matrix to ensure real eigenvalues (like what they did in [75] or in [16]). The following theorem and its proof are represented in [19]:

Theorem 1. *Let $\mathbb{P}$ be an irreducible stochastic matrix and suppose that $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbb{P}$. Let $v = \left[v_1^T \mid -v_2^T \mid \mathbf{0}^T \right]^T$ be a corresponding λ eigenvector (with $v_1 > 0$ and $v_2 > 0$) and let us partition the matrix $\mathbb{P}$ conformally as* $\begin{bmatrix} \mathbb{P}_{1,1} & \mathbb{P}_{1,2} & \mathbb{P}_{1,3} \\ \mathbb{P}_{2,1} & \mathbb{P}_{2,2} & \mathbb{P}_{2,3} \\ \mathbb{P}_{3,1} & \mathbb{P}_{3,2} & \mathbb{P}_{3,3} \end{bmatrix}$ *and label the subsets of the partition as S_1, S_2 and S_0 respectively. Then:*

1. *$\rho\left(\mathbb{P}_{1,1}\right)$, $\rho\left(\mathbb{P}_{2,2}\right) \geq \lambda$.*

2. *There are subsets $\tilde{S}_1 \subseteq S_1$, $\tilde{S}_2 \subseteq S_2$, and positive vectors $\tilde{w}_1^T$, $\tilde{w}_2^T$ with supports on $\tilde{S}_1$, $\tilde{S}_2$ respectively such that $\tilde{w}_1^T \mathbf{1} = \tilde{w}_2^T \mathbf{1} = 1$ and $\sum_{i \in \tilde{S}_1} \tilde{w}_1(i) \sum_{j \notin \tilde{S}_1} \mathbb{P}_{i,j} = 1 - \rho\left(\mathbb{P}_{1,1}\right) \leq 1 - \lambda$ and $\sum_{i \in \tilde{S}_2} \tilde{w}_2(i) \sum_{j \notin \tilde{S}_2} \mathbb{P}_{i,j} = 1 - \rho\left(\mathbb{P}_{2,2}\right) \leq 1 - \lambda$.*

3. *For any $j \in \tilde{S}_2$, $\sum_{i \in \tilde{S}_1} \tilde{w}_1(i) m_{ij} \geq \dfrac{1}{1 - \rho\left(\mathbb{P}_{1,1}\right)} \geq \dfrac{1}{1 - \lambda}$ and for any $j \in \tilde{S}_1$, $\sum_{i \in \tilde{S}_2} \tilde{w}_2(i) m_{ij} \geq \dfrac{1}{1 - \rho\left(\mathbb{P}_{2,2}\right)} \geq \dfrac{1}{1 - \lambda}$, where m_{ij} are elements of the mean first passage matrix.*

In the previous theorem the third partition can be empty without affecting the validity of the theorem; $\mathbf{0}$ and $\mathbf{1}$ are column vectors of zeros and ones of appropriate dimensions; $\rho(A)$ indicates the spectral radius of matrix A; the support of a vector is the set of coordinates on which the vector is nonzero.

The theorem shows how an eigenvector corresponding to an eigenvalue close to 1 can be used to detect nearly disconnected groups of states in a Markov chain. The clustering idea is

formalised through parts 2 and 3 of the theorem in terms of small probabilities of going from one part of the graph to the other, and with high mean first passage times.

Proof. 1. We have $\mathbb{P}_{1,1}v_1 = \lambda v_1 + \mathbb{P}_{12}v_2$. Let z^T be a Perron vector for $\mathbb{P}_{1,1}$. Then $\rho(\mathbb{P}_{1,1})z^Tv_1 = z^T\mathbb{P}_{1,1}v_1 = \lambda z^Tv_1 + z^T\mathbb{P}_{1,2}v_2 \geq \lambda z^Tv_1$. The inequality follows because all terms are positive. Therefore, comparing the first and last term of the chain of inequalities, we obtain $\rho(\mathbb{P}_{1,1}) \geq \lambda$. An analogous argument applies for $\rho(\mathbb{P}_{2,2})$.

2. Let w_1^T be a left Perron vector for $\mathbb{P}_{1,1}$, normalised so that $w_1^T\mathbf{1} = 1$. Partition S_1 as $\tilde{S}_1 \cup \bar{S}_1$, where $\tilde{S}_1$ is the support of w_1^T, and denote the corresponding subvector of w_1^T by $\tilde{w}_1^T$. Let w_2^T, $\tilde{w}_2^T$, $\tilde{S}_2$ and $\bar{S}_2$ denote the analogous quantities for $\mathbb{P}_{2,2}$. Let us write $\mathbb{P}$ in partitioned form as

$$
\left[
\begin{array}{c|c|c|c|c}
\mathbb{P}_{\tilde{S}_1\tilde{S}_1} & \mathbb{P}_{\tilde{S}_1\bar{S}_1} & \mathbb{P}_{\tilde{S}_1\tilde{S}_2} & \mathbb{P}_{\tilde{S}_1\bar{S}_2} & \mathbb{P}_{\tilde{S}_1S_0} \\
\hline
\mathbb{P}_{\bar{S}_1\tilde{S}_1} & \mathbb{P}_{\bar{S}_1\bar{S}_1} & \mathbb{P}_{\bar{S}_1\tilde{S}_2} & \mathbb{P}_{\bar{S}_1\bar{S}_2} & \mathbb{P}_{\bar{S}_1S_0} \\
\hline
\mathbb{P}_{\tilde{S}_2\tilde{S}_1} & \mathbb{P}_{\tilde{S}_2\bar{S}_1} & \mathbb{P}_{\tilde{S}_2\tilde{S}_2} & \mathbb{P}_{\tilde{S}_2\bar{S}_2} & \mathbb{P}_{\tilde{S}_2S_0} \\
\hline
\mathbb{P}_{\bar{S}_2\tilde{S}_1} & \mathbb{P}_{\bar{S}_2\bar{S}_1} & \mathbb{P}_{\bar{S}_2\tilde{S}_2} & \mathbb{P}_{\bar{S}_2\bar{S}_2} & \mathbb{P}_{\bar{S}_2S_0} \\
\hline
\mathbb{P}_{S_0\tilde{S}_1} & \mathbb{P}_{S_0\bar{S}_1} & \mathbb{P}_{S_0\tilde{S}_2} & \mathbb{P}_{S_0\bar{S}_2} & \mathbb{P}_{S_0S_0}
\end{array}
\right] . \tag{2.7}
$$

We have

$$
1 = \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\tilde{S}_1}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\bar{S}_1}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\tilde{S}_2}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\bar{S}_2}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1S_0}\mathbf{1} =
$$

$$
= \rho(\mathbb{P}_{11}) + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\bar{S}_1}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\tilde{S}_2}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\bar{S}_2}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1S_0}\mathbf{1}
$$

so that $1 - \lambda \geq 1 - \rho(\mathbb{P}_{11}) = \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\bar{S}_1}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\tilde{S}_2}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1\bar{S}_2}\mathbf{1} + \tilde{w}_1^T\mathbb{P}_{\tilde{S}_1S_0}\mathbf{1}$ and the desired inequality follows. An analogous argument applies to $(\mathbb{P}_{2,2})$.

3. Fix $j \in \tilde{S}_2$ and let $\mathbb{P}_{(j)}$ be formed from $\mathbb{P}$ by deleting its jth row and column. Then for any $i \in \tilde{S}_1$ we have

$$
m_{ij} = e_i^T\left(I - \mathbb{P}_{(j)}\right)^{-1}\mathbf{1} \geq e_i^T\left(I - \mathbb{P}_{\tilde{S}_1\tilde{S}_1}\right)^{-1}\mathbf{1}.
$$

The last inequality follows as $\mathbb{P}_{\tilde{S}_1\tilde{S}_1}$ is a submatrix of $\mathbb{P}_{(j)}$. Hence, $\sum_{i\in\tilde{S}_1}\tilde{w}_1(i)m_{ij} \geq \tilde{w}_1^T\left(I - \mathbb{P}_{\tilde{S}_1\tilde{S}_1}\right)^{-1}\mathbf{1} = \frac{1}{1 - \rho(\mathbb{P}_{1,1})} \geq \frac{1}{1 - \lambda}$. A similar argument establishes the desired inequality for $j \in \tilde{S}_1$.

$\square$

The argument given in [15] is valid only if the second largest eigenvalue is real. However, in some circumstances the second eigenvalue is complex. In Chapter 3 of this thesis we face this situation ,see section 3.2.1. Figure 3.5 represents the mean first passage time and the second largest modulus eigenvector of the transition matrix of the graph represented in Figure 3.2. It is obvious that the clusters are identifed by both the mean first passage time and the second largest eigenvector. From Figure 3.5b we can see that the values of the entries of the second eigenvector are complex and they are capable of representing the clusters of the graph. Based on our knowledge clustering properties of the eigenvectors corresponding to the complex eigenvalues close to one have never been investigated before. Our motivation now is to study this case.

We have a theorem which finds bounds for the spectral radius of the diagonal blocks of the transition matrix whose eigenvalue is complex. Then according to part 3 of theorem (1), it is possible to identify the clusters. This means that we prove that: if $\mathbb{P}$ is a row stochastic matrix, and there exist eigenvalues with magnitude close to 1, and which have small imaginary parts, then the eigenvectors associated with these eigenvalues reveal clusters in the network. loosely understand a cluster to be a set of nodes from which it is unlikely to go to other sets of nodes of the graph.

We found it more understandable for the reader to start by giving a summary proof of this result, and then conclude with a formal theorem stating the above result.

Outline proof of Theorem 2 (following page 23) :

Suppose $\mathbb{P}$ is a stochastic irreducible matrix with an eigenvalue $\lambda = \alpha + i\beta$ whose corresponding eigenvector is $X + iY$. Therefore, We have

$$\mathbb{P}(X + iY) = (\alpha + i\beta)(X + iY) \tag{2.8}$$

We separate the real part and imaginary part and multiply them by scalars s and t as follows:

$$s(\mathbb{P}X) = s(\alpha X - \beta Y) \tag{2.9}$$

$$t(\mathbb{P}Y) = t(\beta X + \alpha Y) \tag{2.10}$$

We sum the equations (2.9) and (2.10) to get the linear combination as:

$$\mathbb{P}(sX + tY) = (\alpha s + t\beta)X + (\alpha t - \beta s)Y \tag{2.11}$$

Different cases of s and t will be considered through the proof.

$$s = 1, \qquad t = 0 \tag{2.12a}$$

$$s = 1, \qquad t > 0 \tag{2.12b}$$

$$s = 1, \qquad t < 0 \tag{2.12c}$$

$$s = 0, \qquad t = 1 \tag{2.12d}$$

$$s > 0, \qquad t = 1 \tag{2.12e}$$

$$s < 0, \qquad t = 1 \tag{2.12f}$$

1) In (2.12a) we assume that $s = 1$ and $t = 0$ then equation (2.11) is:

$$\mathbb{P}X = \alpha X - \beta Y \tag{2.13}$$

We partition equation (2.13) as (assuming that S_1, S_2 and S_3 denote the index sets of the partition):

$$\begin{bmatrix} \mathbb{P}_{1,1} & \mathbb{P}_{1,2} & \mathbb{P}_{1,3} \\ \mathbb{P}_{2,1} & \mathbb{P}_{2,2} & \mathbb{P}_{2,3} \\ \mathbb{P}_{3,1} & \mathbb{P}_{3,2} & \mathbb{P}_{3,3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \alpha \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} - \beta \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \tag{2.14}$$

In which $X_1 > 0$, $X_2 < 0$, and $X_3 = 0$, First row of the matrix equation (2.14) is:

$$\mathbb{P}_{1,1}X_1 + \mathbb{P}_{1,2}X_2 + \mathbb{P}_{1,3}X_3 = \alpha X_1 - \beta Y_1 \tag{2.15}$$

since $\mathbb{P}_{1,2}X_2 < 0$ and $\mathbb{P}_{1,3}X_3 = 0$ we have

$$\mathbb{P}_{1,1}X_1 \geq \alpha X_1 - \beta Y_1 > 0 \tag{2.16}$$

Here we should note that it is vital to further discussion on these clusters that we assume that

$$\alpha X_1(j) - \beta Y_1(j) > 0 \tag{2.17}$$

We consider (2.17) as a hypotheses to be satisfied in order to state the result of the lower bounds. Although (2.19) technically still holds in case this hypotheses do not, it is worthless to us, since the spectral radius is always nonnegative.

Inequality (2.16) leads us to

$$\mathbb{P}_{1,1}\mathbf{1} \geq \min_{j} \frac{\alpha X_1(j) - \beta Y_1(j)}{X_1(j)}\mathbf{1} \tag{2.18}$$

where $\mathbf{1}$ is a vector of ones. According to Theorem 8.3.2 in [53], from (2.18), we conclude that

$$\rho(\mathbb{P}_{1,1}) \geq \min_{j} \frac{\alpha X_1(j) - \beta Y_1(j)}{X_1(j)} \tag{2.19}$$

Similarly we can show that

$$\rho(\mathbb{P}_{2,2}) \geq \min_{j} \frac{\alpha X_2(j) - \beta Y_2(j)}{X_2(j)} \tag{2.20}$$

Equations (2.19) and (2.20) give us bounds for $\rho(\mathbb{P}_{1,1})$ and $\rho(\mathbb{P}_{2,2})$.

If α is close to 1 and β is close to 0, then the lower bounds in (2.19) and (2.20) are close to 1, indicating that the vertices indexed by S_1 and S_2 display some clustering behaviour in the Markov chain represented by $\mathbb{P}$ because according to the theorem 1 when the lower bounds for $\rho(\mathbb{P}_{1,1})$ and $\rho(\mathbb{P}_{2,2})$ are close to 1, the probability of traveling between digonal block $\mathbb{P}_{1,1}$ and $\mathbb{P}_{2,2}$ is very low which means that the vertices indexed by S_1 and S_2 are clusters.

2) Now according to (2.12b) we assume $s = 1$ and $t > 0$, according to equation (2.11) we have

$$\mathbb{P}(X + tY) = (\alpha + t\beta)X + (\alpha t - \beta)Y \tag{2.21}$$

Now we use the same partitiong of the matrix $\mathbb{P}$ that we used in (2.14)

$$\begin{bmatrix} \mathbb{P}_{1,1} & \mathbb{P}_{1,2} & \mathbb{P}_{1,3} \\ \mathbb{P}_{2,1} & \mathbb{P}_{2,2} & \mathbb{P}_{2,3} \\ \mathbb{P}_{3,1} & \mathbb{P}_{3,2} & \mathbb{P}_{3,3} \end{bmatrix} \begin{bmatrix} X_1 + tY_1 \\ X_2 + tY_2 \\ X_3 + tY_3 \end{bmatrix} = (\alpha + t\beta) \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} - (\alpha t - \beta) \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \tag{2.22}$$

Here, we want to partition the vector $X + tY$ to positive, negative and zero parts (like what we did in (2.14))

If $Y_1 > 0$, (since $X_1 > 0$) then $X_1 + tY_1 > 0$.

If $Y_1 < 0$, then t should be $0 < t < \min_{j:Y_1(j)<0} -\frac{X_1(j)}{Y_1(j)}$ to satisfy $X_1 + tY_1 > 0$.

we partition X_3 and Y_3 as follows:

$$X_{3,1} = 0 \quad Y_{3,1} > 0 \tag{2.23}$$

$$X_{3,2} = 0 \quad Y_{3,2} < 0 \tag{2.24}$$

$$X_{3,3} = 0 \quad Y_{3,3} = 0 \tag{2.25}$$

We update the partitioning as :

$$\begin{bmatrix} \tilde{\mathbb{P}}_{1,1} & \tilde{\mathbb{P}}_{1,2} & \tilde{\mathbb{P}}_{1,3} \\ \tilde{\mathbb{P}}_{2,1} & \tilde{\mathbb{P}}_{2,2} & \tilde{\mathbb{P}}_{2,3} \\ \tilde{\mathbb{P}}_{3,1} & \tilde{\mathbb{P}}_{3,2} & \tilde{\mathbb{P}}_{3,3} \end{bmatrix} \begin{bmatrix} \tilde{X}_1 + t\tilde{Y}_1 \\ \tilde{X}_2 + t\tilde{Y}_2 \\ \tilde{X}_3 + t\tilde{Y}_3 \end{bmatrix} = (\alpha + t\beta) \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \tilde{X}_3 \end{bmatrix} - (\alpha t - \beta) \begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \tilde{Y}_3 \end{bmatrix} \tag{2.26}$$

in which $\tilde{X}_1 + t\tilde{Y}_1 > 0$, $\tilde{X}_2 + t\tilde{Y}_2 < 0$, and $\tilde{X}_3 + t\tilde{Y}_3 = 0$ and

$$\begin{bmatrix} X_1 \\ X_{3,1} \end{bmatrix} = \tilde{X}_1 \qquad \begin{bmatrix} Y_1 \\ Y_{3,1} \end{bmatrix} = \tilde{Y}_1 \tag{2.27}$$

where $X_1 > 0$, $X_{3,1} = 0$, and $Y_{3,1} > 0$

$$\begin{bmatrix} X_2 \\ X_{3,2} \end{bmatrix} = \tilde{X}_2 \qquad \begin{bmatrix} Y_2 \\ Y_{3,2} \end{bmatrix} = \tilde{Y}_2 \tag{2.28}$$

$$X_{3,3} = \tilde{X}_3 \qquad Y_{3,3} = \tilde{Y}_3 \tag{2.29}$$

where $X_2 < 0$, $X_{3,2} = 0$, and $Y_{3,2} > 0$.

We should note that this repartitioning is simply allowing the option of including some extra vertices in the clusters S_1 and S_2 by including indices corresponding to positive entries of Y_3 to S_1, and indices corresponding to negative entries of Y_3 to S_2. In fact $S_1 \subseteq \tilde{S}_1$ and $S_2 \subseteq \tilde{S}_2$ (assuming that $\tilde{S}_1, \tilde{S}_2, \tilde{S}_3$ are the index sets of the partitioning in (2.26)). Interpreting in terms of potential clustering behaviour, we are simply allowing the possible addition of more nodes into our existing set to potentially get our hand on a more accurate cluster.

Proceeding as before, the first row of the matrix equation (2.26) leads us to

$$\tilde{\mathbb{P}}_{1,1}(\tilde{X}_1 + t\tilde{Y}_1) \geq \alpha(\tilde{X}_1 + t\tilde{Y}_1) + \beta(t\tilde{X}_1 - \tilde{Y}_1) > 0. \tag{2.30}$$

As we already explained for (2.17), we need to ensure that the right hand side of (2.30) is positive in order for our conclusions to be worthwhile. If $S_1 = \tilde{S}_1$, then the positivity of

$\alpha X_1 - \beta Y_1$ is enough to ensure the positivity of the right hand side of (2.30), but if S_1 is a proper subset of $\tilde{S}_1$, the case that $\tilde{X}_1 = 0$ and $\tilde{Y}_1 > 0$ is considered, we must have $t > \frac{\beta}{\alpha}$.

From equation (2.30) we have

$$\tilde{\mathbb{P}}_{1,1}\mathbf{1} \geq \min_j(\alpha + \beta(\frac{\tilde{X}_1(j) + t\tilde{Y}_1(j)}{\tilde{X}_1(j) + t\tilde{Y}_1(j)})).\mathbf{1} \qquad (2.31)$$

According to the Theorem 8.3.2 in [53]

$$\rho(\tilde{\mathbb{P}}_{1,1}) \geq \min_j(\alpha + \beta(\frac{t\tilde{X}_1(j) + \tilde{Y}_1(j)}{\tilde{X}_1(j) + t\tilde{Y}_1(j)})). \qquad (2.32)$$

In order to optimize the lower bound we should maximize the function $\frac{\tilde{X}_1(j) + t\tilde{Y}_1(j)}{\tilde{X}_1(j) + t\tilde{Y}_1(j)}$ over t. Since it is an increasing function of t, we reach the maximum by taking the limit as t approaches its higher value, which can be either $t_M = \min_{j:Y_1(j)<0} -\frac{X_1(j)}{Y_1(j)}$ (if $Y_1 < 0$) or infinity (in case that $Y_1 > 0$).

Two cases may happen when t approches t_M:

- Case 1

$$\exists j \quad s.t. \quad \tilde{X}_1(j) + t_M\tilde{Y}_1(j) \neq 0. \qquad (2.33)$$

In this case since the function is increasing the maximum happens at t_M so

$$\rho(\tilde{\mathbb{P}}_{1,1}) \geq \min_j(\alpha + \beta(\frac{t_M\tilde{X}_1(j) - \tilde{Y}_1(j)}{\tilde{X}_1(j) + t_M\tilde{Y}_1(j)})) \qquad (2.34)$$

so we have a bound for ρ.

- Case 2

$$\forall j \quad \tilde{X}_1(j) + t_M\tilde{Y}_1(j) = 0. \qquad (2.35)$$

If equation (2.35) happens, it means that no node is moved from S_3 to S_1 because $X_{3,1} + t_M Y_{3,1}$ cannot be zero according to (2.23). Therefore, the partitioning and the bound are the same as (2.22) and (2.19).

From (2.35), we have

$$\forall j \quad \tilde{Y}_1(j) = \frac{-1}{t_M}\tilde{X}_1(j) \qquad (2.36)$$

therefore

$$\tilde{Y}_1 = \frac{-1}{t_M}\tilde{X}_1 \tag{2.37}$$

which is representing the linear dependence between the vectors $\tilde{X}_1$ and $\tilde{Y}_1$. We can easily add another condition that the vectors $\tilde{X}_1$ and $\tilde{Y}_1$ should be linearly independent to avoid this situation.

If t goes to infinity then

$$\rho(\tilde{\mathbb{P}}_{1,1}) \geq \min_j (\alpha + \beta \frac{\tilde{X}_1(j)}{\tilde{Y}_1(j)}) \tag{2.38}$$

With the same procedure for the second diagonal block we have

$$\rho(\tilde{\mathbb{P}}_{2,2}) \geq \min_j (\alpha + \beta (\frac{t_M\tilde{X}_2(j) - \tilde{Y}_2(j)}{\tilde{X}_2(j) + t_M\tilde{Y}_2(j)})) \tag{2.39}$$

$$\rho(\tilde{\mathbb{P}}_{2,2}) \geq \min_j (\alpha + \beta \frac{\tilde{X}_2(j)}{\tilde{Y}_2(j)}) \tag{2.40}$$

We analysed the first two cases (2.12a) and (2.12b) in detail. By taking the same procedure for (2.12c), (2.12d), (2.12e), and (2.12f), we reach the bounds that we mention in the following theorem:

Theorem 2. *Let $\mathbb{P}$ be an irreducible stochastic matrix. Suppose that $\lambda = \alpha + i\beta$ is an eigenvalue of $\mathbb{P}$ and let $X + iY$ be the corresponding right eigenvector where $X, Y \in \mathbb{R}^n$. we partition the matrix $\mathbb{P}$ in six different ways, the ways that we partition was explained in the proof. For $i = 1, 2, 3$, let S_i, $\tilde{S}_i$, $\bar{S}_i$, $\ddot{S}_i$, $\tilde{\tilde{S}}_i$, and $\bar{\bar{S}}_i$ be the index sets described in the proof. Let X_i, Y_i, $\tilde{X}_i$, $\tilde{Y}_i$, $\bar{X}_i$, $\bar{Y}_i$, $\ddot{X}_i$, $\ddot{Y}_i$, $\tilde{\tilde{X}}_i$, $\tilde{\tilde{Y}}_i$, $\bar{\bar{X}}_i$, and $\bar{\bar{Y}}_i$ be the subvectors of X and Y corresponding to the index sets, and let $\mathbb{P}_{i,i}$, $\tilde{\mathbb{P}}_{i,i}$, $\bar{\mathbb{P}}_{i,i}$, $\ddot{\mathbb{P}}_{i,i}$, $\tilde{\tilde{\mathbb{P}}}_{i,i}$, and $\bar{\bar{\mathbb{P}}}_{i,i}$ be the principal submatrices of $\mathbb{P}$ corresponding to the index sets. Assume that X_i and Y_i, $\tilde{X}_i$ and $\tilde{Y}_i$, $\bar{X}_i$ and $\bar{Y}_i$, $\ddot{X}_i$ and $\ddot{Y}_i$, $\tilde{\tilde{X}}_i$ and $\tilde{\tilde{Y}}_i$, $\bar{\bar{X}}_i$ and $\bar{\bar{Y}}_i$ are linearly independent for $i = 1, 2$. Then:*

1. *If* $\quad \alpha X_1 - \beta Y_1 > 0 \quad$ *then* $\quad \rho(A_{1,1}) \geq \alpha - \beta \min_j \left\{ \frac{Y_1(j)}{X_1(j)} \right\}.$

2. *If* $\quad \alpha X_2 - \beta Y_2 < 0 \quad$ *then* $\quad \rho(A_{2,2}) \geq \alpha - \beta \min_j \left\{ \frac{Y_2(j)}{X_2(j)} \right\}.$

3. *If* $\quad \alpha \tilde{X}_1 - \beta \tilde{Y}_1 > 0 \quad$ *then* $\quad \rho(\tilde{A}_{1,1}) \geq \alpha - \beta \min_j \left\{ \frac{t\tilde{X}_1(j) - \tilde{Y}_1(j)}{\tilde{X}_1(j) + t\tilde{Y}_1(j)} \right\}.$

   where $t > 0$ and is bounded above by

$$\min_{\tilde{Y}_1(j)<0}\left\{\frac{-\tilde{X}_1(j)}{\tilde{Y}_1(j)}\right\} \qquad and \qquad \min_{\tilde{Y}_2(j)>0}\left\{\frac{-\tilde{X}_2(j)}{\tilde{Y}_2(j)}\right\}$$

$$if \quad \tilde{Y}_1 > 0 \ and \ \tilde{Y}_2 < 0 \qquad then \qquad \rho(\tilde{A}_{1,1}) \geq \alpha + \beta \min_j\left\{\frac{\tilde{X}_1(j)}{\tilde{Y}_1(j)}\right\}.$$

4. If $\quad \alpha\tilde{X}_2 - \beta\tilde{Y}_2 < 0 \qquad then \qquad \rho(\tilde{A}_{2,2}) \geq \alpha + \beta \min_j\left\{\frac{t\tilde{X}_2(j)-\tilde{Y}_2(j)}{\tilde{X}_2(j)+t\tilde{Y}_2(j)}\right\}.$

   where $t > 0$ and is bounded above by

$$\min_{\tilde{Y}_1(j)<0}\left\{\frac{-\tilde{X}_1(j)}{\tilde{Y}_1(j)}\right\} \qquad and \qquad \min_{\tilde{Y}_2(j)>0}\left\{\frac{-\tilde{X}_2(j)}{\tilde{Y}_2(j)}\right\}$$

$$if \quad \tilde{Y}_1 > 0 \ and \ \tilde{Y}_2 < 0 \qquad then \qquad \rho(\tilde{A}_{2,2}) \geq \alpha + \beta \min_j\left\{\frac{\tilde{X}_2(j)}{\tilde{Y}_2(j)}\right\}.$$

5. If $\quad \alpha\bar{X}_1 - \beta\bar{Y}_1 > 0 \qquad then \qquad \rho(\bar{A}_{1,1}) \geq \alpha - \beta \min_j\left\{\frac{\bar{Y}_1(j)}{\bar{X}_1(j)}\right\}.$

6. If $\quad \alpha\bar{X}_2 - \beta\bar{Y}_2 > 0 \qquad then \qquad \rho(\bar{A}_{2,2}) \geq \alpha - \beta \min_j\left\{\frac{\bar{Y}_2(j)}{\bar{X}_2(j)}\right\}.$

7. If $\quad \alpha\ddot{Y}_1 + \beta\ddot{X}_1 > 0 \qquad then \qquad \rho(\ddot{A}_{1,1}) \geq \alpha + \beta \min_j\left\{\frac{\ddot{X}_1(j)}{\ddot{Y}_1(j)}\right\}.$

8. If $\quad \alpha\ddot{Y}_2 + \beta\ddot{X}_2 < 0 \qquad then \qquad \rho(\ddot{A}_{2,2}) \geq \alpha + \beta \min_j\left\{\frac{\ddot{X}_2(j)}{\ddot{Y}_2(j)}\right\}.$

9. If $\quad \alpha\tilde{\tilde{Y}}_1 + \beta\tilde{\tilde{X}}_1 > 0 \qquad then \qquad \rho(\tilde{\tilde{A}}_{1,1}) \geq \alpha + \beta \min_j\left\{\frac{\tilde{\tilde{X}}_1(j)}{\tilde{\tilde{Y}}_1(j)}\right\}.$

10. If $\quad \alpha\tilde{\tilde{Y}}_2 + \beta\tilde{\tilde{X}}_2 > 0 \qquad then \qquad \rho(\tilde{\tilde{A}}_{2,2}) \geq \alpha + \beta \min_j\left\{\frac{\tilde{\tilde{X}}_2(j)}{\tilde{\tilde{Y}}_2(j)}\right\}.$

11. If $\quad \alpha\bar{\bar{Y}}_1 + \beta\bar{\bar{X}}_1 > 0 \qquad then \qquad \rho(\bar{\bar{A}}_{1,1}) \geq \alpha + \beta \min_j\left\{\frac{\bar{\bar{X}}_1(j)-s\bar{\bar{Y}}_1(j)}{s\bar{\bar{X}}_1(j)+\bar{\bar{Y}}_1(j)}\right\}.$

   where $s < 0$ and is bounded below by

$$\min_{\bar{\bar{X}}_1(j)>0}\left\{\frac{-\bar{\bar{Y}}_1(j)}{\bar{\bar{X}}_1(j)}\right\} \qquad and \qquad \min_{\bar{\bar{X}}_2(j)<0}\left\{\frac{-\bar{\bar{Y}}_2(j)}{\bar{\bar{X}}_2(j)}\right\}$$

$$if \quad \bar{\bar{X}}_1 < 0 \ and \ \bar{\bar{X}}_2 > 0 \qquad then \qquad \rho(\bar{\bar{A}}_{1,1}) \geq \alpha - \beta \min_j\left\{\frac{\bar{\bar{Y}}_1(j)}{\bar{\bar{X}}_1(j)}\right\}.$$

12. If $\quad \alpha\bar{\bar{Y}}_2 + \beta\bar{\bar{X}}_2 < 0 \qquad then \qquad \rho(\bar{\bar{A}}_{2,2}) \geq \alpha + \beta \min_j\left\{\frac{\bar{\bar{X}}_2(j)-s\bar{\bar{Y}}_2(j)}{s\bar{\bar{X}}_2(j)+\bar{\bar{Y}}_2(j)}\right\}.$

   where $s < 0$ and is bounded below by

$$\min_{\bar{\bar{X}}_1(j)>0}\left\{\frac{-\bar{\bar{Y}}_1(j)}{\bar{\bar{X}}_1(j)}\right\} \qquad and \qquad \min_{\bar{\bar{X}}_2(j)>0}\left\{\frac{-\bar{\bar{Y}}_2(j)}{\bar{\bar{Y}}_2(j)}\right\}$$

$$if \quad \bar{\bar{X}}_1 > 0 \; and \; \bar{\bar{X}}_2 > 0 \qquad then \qquad \rho(\bar{\bar{A}}_{2,2}) \geq \alpha - \beta \min_j \left\{ \frac{\bar{\bar{Y}}_2(j)}{\bar{\bar{X}}_2(j)} \right\}.$$

Example: In the next chapter we have an example of a transition matrix with complex eigenvalues and eigenvectors, see Figure 3.5. The complex eigenvalues is $0.9989 + 0.0005i$. It is obvious that its real part is close to 1 and its imaginary part is close to zero. In Figure 3.5 the entries of the corresponding eigenvector are represented and it can be seen that they are revealing clusters of the network. The index sets are $S_1 = \{1, 2, ..., 11\}$, $S_2 = \{12, 13, ..., 17\}$ and S_3 is empty. It should be mentioned that all the entries of the vector $\alpha X_1 - \beta Y_1$ are greater than zero in this case.

If we assume that α is very close to 1 and β is very close to 0, then these lower bounds of spectral radius lead us to the clustering properties in an analogous manner as in Theorem 1 [19].

2.3   Markov chains in transportation

2.3.1   Markov chains and big data

To gain insight into fast recalculation of Markov chain quantities for changed data, regard the following theorem from [70] addressing row updates to Markov chain transition matrices and their effect on the stationary distributions.

Theorem 3. *Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be irreducible $n \times n$ Markov chain transition matrices that satisfy the relationship $\tilde{\mathbb{P}} - \mathbb{P} = -e_i \delta^\top$, where e_i is a vector of zeros of length n with a 1 in the i'th position and $\delta \in \mathbb{R}^n$. Let π and $\tilde{\pi}$ be the respective left Perron eigenvectors of $\mathbb{P}$ and $\tilde{\mathbb{P}}$. Let $Q = (I - \mathbb{P})$ and let $Q^\#$ be its group inverse. Then*

$$\tilde{\pi}^\top = \pi^\top - \varepsilon^\top, \tag{2.41}$$

where $\varepsilon^\top = \frac{\pi_i}{1 + \delta^\top Q^\# e_i} \delta^\top Q^\#$. Further, with $\mathbf{1} \in \mathbb{R}^n$ being the vector of all ones,

$$(I - \tilde{\mathbb{P}})^\# = Q^\# + \mathbf{1}\varepsilon^\top (Q^\# - \frac{\varepsilon^\top Q^\# e_i}{\pi_i} I) - \frac{Q^\# e_i \varepsilon^\top}{\pi_i}. \tag{2.42}$$

The above theorem allows us to explicitly compute the stationary distribution of a Markov chain after updating a single row using the original stationary distribution and the group inverse of $(I - \mathbb{P})$. It also allows us to directly compute the group inverse of the updated Markov chain and this theorem can thus be used iteratively to obtain updated stationary distributions for arbitrary changes in the transition matrix by means of describing them as

consecutive row updates. In some situations even simpler formulas can be obtained. For example, the following theorem was proved in [111].

Theorem 4. *Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be irreducible $n \times n$ Markov chain transition matrices such that $\tilde{\mathbb{P}}$ is obtained from $\mathbb{P}$ by multiplying the i'th diagonal entry with a factor $w_i > 0$ for each $i = 1 \dots, n$ and scaling the off diagonal entries in each row so that their ratios remain constant. Let π and $\tilde{\pi}$ be the respective left Perron eigenvectors of $\mathbb{P}$ and $\tilde{\mathbb{P}}$. Then*

$$\tilde{\pi} = \kappa W \pi, \tag{2.43}$$

where $W = diag(w_1, \dots, w_n)$ and $\kappa = \frac{1}{\|W\pi\|_1}$ is a scaling factor that ensures that the entries of $\tilde{\pi}$ sum to 1.

This property is particularly useful for our model. It will be shown later in the next chapter that changing the diagonal entries corresponds to changing the public transport service frequency.

2.3.2   Markov chains and road networks

It is easy and straightforward to build a connection between a road network and a Markov chain if a city map is interpreted as a directed graph, where nodes and edges of the graph correspond to junctions of the road and connections between the roads respectively. In the literature related to urban networks, this representation is sometimes called primal, see [100]. On the other hand, in the dual representation streets correspond to nodes and junctions to edges. In Figures 2.2 and 2.3, the examples of the primal and dual graphs used in [19] are represented. It is easy to see that the network of Figure 2.2 is designed to represent a city with two main communities connected through a junction, which is called junction D in Figure 2.2.

In the work of [51] which was published in the late eighties, the use of graph theory to analyse urban networks was proposed. After that the idea was further developed in the later works [49, 50]. An important achievement of applying graph theory in urban networks was the establishment of correlation between the topological accessibility of streets and urban properties like pedestrian and vehicular flows, human way-finding, microeconomic vitality and social liveability, see [49] for example. The topology of the urban network can mathematically be analysed by finding the degree of the nodes of the corresponding graph, the characteristic path lengths and clustering coefficients. More recently, algorithms like Google's PageRank, have also been used to analyse the topology of urban networks, see [57], [59], and [58]. All approaches exploit well-established mathematical tools borrowed from
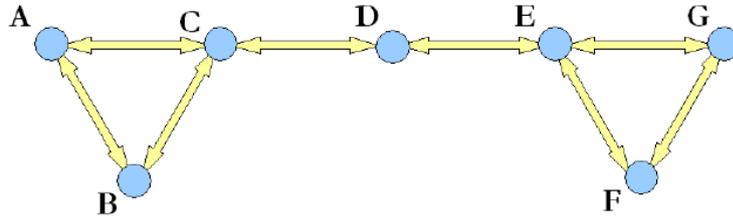
Fig. 2.2 An example of Primal network used in [19]
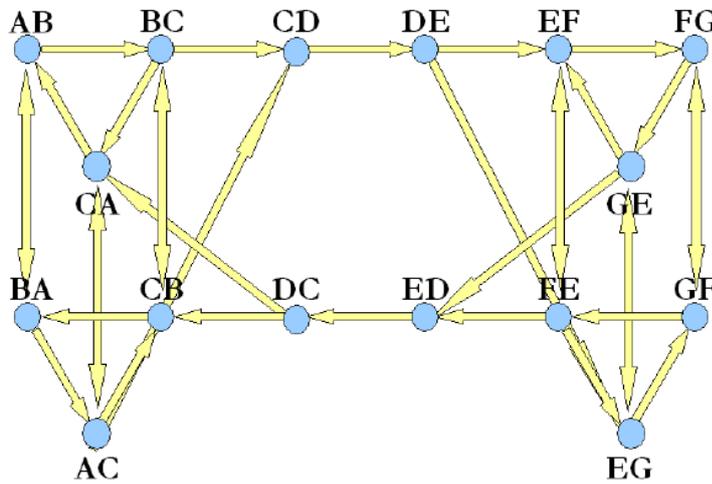
Fig. 2.3 The dual network of Figure 2.2 which is used in [19]

graph theory, but the starting point is still a simple plain urban map (or its dual representation) which usually does not have any quantitative data that are important and required to evaluate traffic. This means that some important variables like speed limits, street lengths, junction turning probabilities, numbers of lanes, presence of traffic lights and priority rules are neglected in their approaches. In [19] they propose a data-driven model with the strong mathematical background of Markov chain theory, that also takes into account all the previous quantitative parameters, which are important and clearly affect traffic flows.

Their main step in [19] to pass from a road network to a Markov chain is to build the transition matrix of the Markov chain. The transition matrix can be obtained after gathering the average travel times and junction turning probablitites. When they have the travel times of all the roads while they are normalized such that the smallest travel time is 1, then diagonal entries of the transition matrix can be obtained as

$$P_{ii} = \frac{tt_i - 1}{tt_i} \qquad i = 1, ..., n \tag{2.44}$$

In which tt_i is the average travel time for road i. The proof of (2.44) can be found in [19]. The off-diagonal entries of the transition matrix are

$$P_{ij} = (1 - P_{ii}) \cdot t p_{ij}, \qquad i \neq j \tag{2.45}$$

in which p_{ij} is the turning probablity of going from road i to road j.

One of the important achievements of the work in [19] is that they validated their model after building the transition matrix. After that, they extract some information of the network that is not easy to extract without Markov chains. Markov chain model identifies critical links for example, by using the mean first passage times and the Kemeny constant they quantify the efficiency of the network etc, see Table 2.1.

2.3.3   Markov chains and Emission

The main idea of the work in [18] is to use the same framework as [19] to model pollutants. This goal is achieved by replacing time in the Markov chain described earlier, see (2.44) and (2.45), by a unit of pollutant (such as benzene or NO_x). In the framework used in [18], a car is moving in the same road network, and changes or remains in the same state when a unit of pollutant is released. The main difference between the Markov chain in [19] and the chain in [18] is the way that they obtain the diagonal entries of the transition matrices. In fact their emission model in [18] is totally determined by a transition matrix whose diagonal terms indicate the number of units of pollution a fleet of vehicles releases along a road section. The off-diagonal entries of the transition matrix somehow indicate the following road section

Property	Meaning	Suggested applications
Perron eigenvector	Congested roads in the network	Traffic balancing
Second eigenvector	(a) Rate of convergence to stationary distribution, (b) If real, presence of weakly-connected sub-communities	
Second eigenvalue	In case (b) it associates nodes to sub-communities	Reveal hidden sub-communities
Mean first passage times	Average travel time from origin to destination	(Non conventional) routing algorithms
Kemeny constant	Average travel time in the network	Identification of important (critical) roads
Perron eigenvector (primal)	Congested junctions in the network	Timing of traffic lights

Table 2.1 Interpretation of Markov chain quantities in the road network and candidate uses for traffic mitigation and control, see [19].

chosen by a unit of pollution. The emission factors are computed from some functions of type of vehicle, type of pollution of interest and the average speed of vehicle, the parameters of the formula is taken from [6]. They compute the diagonal entries of the Markov chain by the following formula

$$P_{jj} = \frac{f_j . l_j - 1}{f_j . l_j} \tag{2.46}$$

in which f_j is the emission factor of section j and l_j is the length of section j of the road.

After building the transition matrix, they extract some quantities of Markov chain such as the Perron eigenvector, the mean first passage times and the Kemeny constant and they interpret these quantities in the emission framework, see Table 2.2. Apart from the quantities mentioned in Table 2.2, some other quantities such as density of emissions along each road, emissions along each road and total emissions are obtained. They demonstrate the efficacy of the proposed approach by means of some examples.

2.4 Bike sharing

Bike sharing schemes are now available as a mean of transportation in most cities around the world as described among other in [34, 87, 104, 114]; see for example *Vélib'* in Paris, *Bicing* in Barcelona, *Capital Bikeshare* in Washington, D.C. and *Forever Bicycle* in Shanghai. A good overview can be gained from [24] and references therein. They are not only seen as an integral component in the public transportation networks for alleviating congestion and

Markov chain quantity	Green interpretation
Left-hand Perron eigenvector	This vector has as many entries as the number of road segments. Each entry represents the long run fraction of emissions that a fleet of vehicles will emit along the corresponding road segment. It can be used as an indicator of pollution peaks.
Mean first passage emissions	This is a square matrix with as many rows as the number of the road segments. The entry ij represents the expected quantity of emissions that a vehicle releases to go from i to j. The average is with respect to all possible paths from i to j.
Kemeny constant	This number is the average number of emissions released in a random route. It is an indicator of pollution in the entire network.

Table 2.2 interpretation of Markov chain quantities in the emission framework in [18].

obesity problems, see [33, 73, 86, 90], but also as a fundamental tool for combating urban pollution, see [9, 102, 103]. They are also being suggested as a step towards solving the last mile problem[1] in intelligent transportation systems, see [101]. See also [115] for an overview over IT based bike sharing schemes in North America and [56] on a simulation study on the related field of electric bike sharing.

Given such a recent diffusion, there is also an interest in developing optimal ways of managing this mean of transportation, bike-sharing systems. In fact, as pointed out in [87], most bike-sharing systems have to rely on public fundings from municipalities or other public sources, since customer fees are not enough to match the capital costs plus the annual operational costs. Accordingly, there is a strong interest in devising smart ways of reducing operational costs, e.g., to minimise the costs of operating staffed trucks for manual relocation of bicycles in order to balance the difference between supply and demand at various stations (see [99]); or to maximise the revenues obtained by using the bikes for advertising.

Our interest in this chapter is to develop a model to capture the macroscopic aggregated properties of the bike-sharing system. From this perspective, we are mainly interested in answering high-level questions that pertain the identification of the main flows in the bike network, or how important are some bike stations with respect to others. Our ultimate goal is finally to support bike-sharing managers in implementing planning and control actions aiming

[1] see C. MacKechnie, *The Last Mile Problem* at url http://publictransport.about.com/od/Transit_Planning/a/The-Last-Mile-Problem.htm.

at improving the efficiency of the bike network, in a proactive manner (i.e., anticipating the expected outcome of a strategic change in the bike network). The overall problem is particularly challenging if one intends to embed in the model the big quantity of data that become available in quasi real-time from a bike network. As an example, in this chapter we use data collected from the bike-sharing network in Boston, USA[2].

For this purpose, we have used the previously described Markov chain based framework from [19]. They employed graph theory and Markov chain ideas to reveal some non-trivial patterns of urban mobility. In thier approach the density was not constant as the vehicles are constantly leaving and entering the network while in our bike sharing system, the total number of bikes is constant because the bikes are either moving or parked in a station.

The objectives of this chapter are two-fold: (i) first, we explain how the original Markov chain approach of [19] can be adapted to the particular application of interest here, i.e., the bike-sharing network; (ii) second, we show some preliminary results that have been obtained by adopting the proposed modelling framework.

In this chapter, section 2.5 explains how we build a Markov-chain transition matrix from bike trips data, and motivates why Markov chains are expected to be useful to handle big-data analyses. Section 2.6 shortly describes the used database of data and validates the transition matrix built on such data. Section 2.6.1 shows clustering results obtained using the second eigenvector of the transition matrix, and provides an interpretation of the result in terms of the bike network.

2.5   A Markov-chain model for a bike-sharing system

In our model of a bike-sharing network, we assume that the i'th bike station is associated with two states: state BSi refers to a bike that is currently parked at the i'th Bike Station; state TBi refers to a bike that is Travelling from the i'th Bike station to any other destination. Thus, at every time step (in our simulation, one second), a bike staying in the BSi state can either remain in the same state, or move to state TBi (i.e., if somebody picks the bike to reach his destination). Similarly, a bike staying in the TBi state can either remain in the same state (i.e., if it keeps travelling because the destination has not been reached yet) or can move to another state corresponding to any bike station BSj, where the j'th bike station corresponds to the destination. Note that j can be equal to i (i.e., some users may pick up a bike at the i'th bike station, and then return the bike to the same place after having completed their own business). Also, note that it is not possible to pass directly from state BSi to BSj if $i \neq j$ or

---

[2]http://hubwaydatachallenge.org/

from state TBi to TBj if $i \neq j$. Our model is exemplified in the graph shown in Figure 2.4.
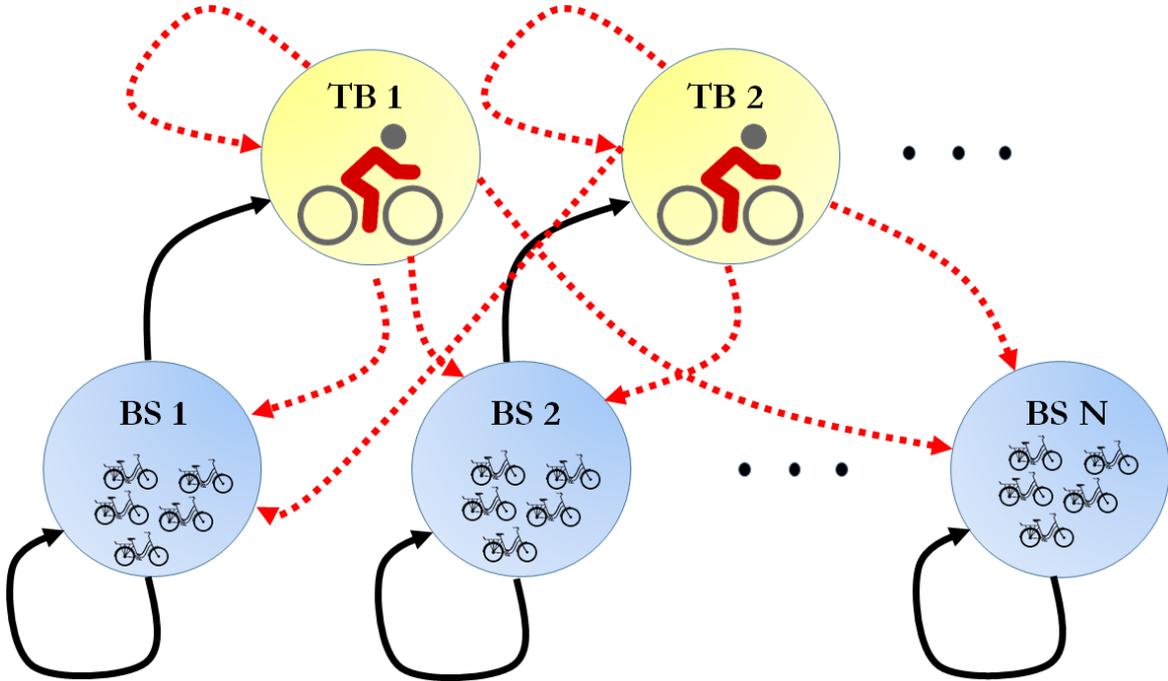
Fig. 2.4 Graph modelling a bike-sharing network. States BS_i's are only connected to themselves or to states TB_i's (black solid connections). States TB_i's are connected to themselves and to every single state BS (red dashed connections).

According to the previous discussion, the transition matrix of the Markov chain has size $2N \times 2N$ where N is the number of bike stations, and it is very sparse (i.e., all but two entries for the states BSi are equal to zero, $\forall i = 1,...,N$, and $n-1$ entries are equal to zero for the states TBi, $\forall i = 1,...,N$). Then, the diagonal entries pertaining the BS states should be built proportional to the average time spent (parked) at a given bike station (i.e., by denoting such time with tp_i, then the entry should be $(tp_i - 1)/tp_i$ as explained in the Appendix 2 in [19]), while the other non-zero entry of the same row is easily determined by enforcing the row to be stochastic (i.e., sum equal to 1). The diagonal entries pertaining the TBi states are similarly computed proportionally to the average length of trips starting from the i'th bike station. The other entries in the same row are chosen proportionally to the frequency with which a given bike station is chosen as a destination for a trip starting from the i'th bike station, and scaled in order to make the row stochastic as well. Thus, the transition matrix can be built only if average parking times, average trip durations and a full list of trips are available. However, note that such a requirement is typically fulfilled, since such data are available anyway for other purposes as well, e.g., typically, GPS-like sensors are mounted on

bikes to prevent people from stealing them; bike stations are monitored to choose optimal relocation strategies to avoid bike stations to remain empty; and finally information about the trips are available for pricing purposes. In particular, such data were publicly available for the city of Boston, USA, and in the next Section we explain how we built the transition matrix for that specific use case. The next subsection collects some known results to motivate the adoption of Markov chains for big-data applications.

2.6   Boston data: validation

The data used in this work involve about 70000 bike trips occurring between July 28 and September 21 in the year 2011. Overall, 53 bike stations and 569 bikes were effectively used in the period under exam. The transition matrix was built upon such available data, and thus corresponds to average values in the examined time range. The transition matrix has size 106×106 and, consistently with the discussion in Section 2.5, not more than $2N + (N+1) \cdot N$ non-zero entries (actually, there are even less non-zero entries, namely, 2654, due to the fact that in the period under exam some particular pairs of origin-destination were never chosen by the users).

All the information required to build the Markov chain transition matrix can be extracted from the list of trips: in particular, every trip stores the ID of the bike, the time at which the bike is taken, the bike station of origin, the time at which the bike is delivered back and the bike station of destination. Other information was also available, such as whether the user was a registered or a casual customer (i.e., without a membership to the bike-sharing service); the zip code of the user; the birth year of the user and the gender. Note that all of such information can be used to deliver particular services (e.g., advertising), but have not been used in this work and their analysis will be left for a future work.

Remark 1 : One might expect that the destination of one trip should coincide with the origin of the next trip (of the same bike). This is true unless the bike gets relocated to another bike station, some time during the two successive trips (e.g., because there is a shortage of bikes at one bike station). Accordingly, in the list of trips from the available data, most of the times the two stations coincide, but in some cases there is a discrepancy. We assumed that every discrepancy was due to relocations, and did not consider such inconsistent trips when computing parking times (because the time of relocation was not available, and thus it could not be established how long the bike had been parked before relocation, or after relocation before being used again).

Remark 2 : Note that the available data only reveal a part of the true story. For instance, some users might have wanted to use a bike to take a specific trip but did not find any available bike at the origin bike station; or some users might have wanted to choose a specific destination but found that it was already full of bikes. We do not have information of such (somewhat) unsatisfied customers, but note that they might introduce some noise in the data (i.e., due to unavailability of bikes, they might choose a different starting point than the desired one; or due to full bike stations, they might choose a different destination point than the desired one).

Figure 2.5 shows in blue the Perron eigenvector of the Markov chain transition matrix built according to the procedure described in Section 2.5. The entries of the Perron eigenvector correspond to the long-run fraction of time that a bike spends in the corresponding state, i.e., either parked at the i'th bike station, or on a trip starting from the i'th bike station. Figure 2.5 also compares the Perron eigenvector with the same information extracted from the same data (and shown in red). As can be noticed from the figure, the two densities are very close, and discrepancies are due to the fact that the real data correspond to a possible realisation of the Markov chain transition matrix given a specific initial distribution of bikes. Another interesting information that can be extracted from the figure is that it is more likely for a bike to be parked at a station than being busy and travelling between the stations.

2.6.1   Clustering

As illustrated in reference [19] in a different context, the advantage of aggregating data in a Markov chain, is that some quantities that can be easily computed in the context of Markov chains have an insightful interpretation in terms of the original data that in some cases can not be easily extracted in other ways. As an example, the "second eigenvector" (i.e., the eigenvector associated with the second eigenvalue of largest module, if real) is known to have clustering properties (see Appendix of the reference [19]). To obtain a meaningful interpretation of clusters in the bike network, we only considered the flows of bikes neglecting the times spent parked in the station, and the time spent travelling along a trip. Thus, we considered the submatrix of the original one whose states would correspond to TBi, and renormalised its rows to have a unitary sum. Then, the second eigenvector of the matrix is shown in Figure 2.6.

The interpretation of clusters is that people most likely travel within a single cluster, and more rarely travel from one cluster to the other cluster. The possible membership to one cluster rather than to another one is given by the sign of the corresponding entry of
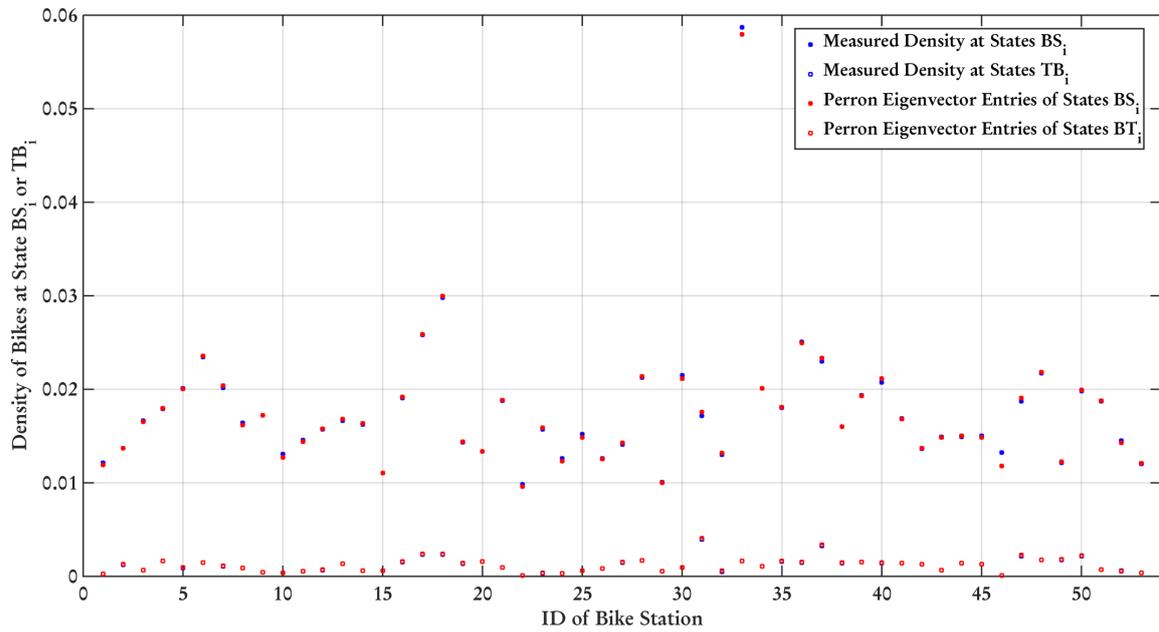
Fig. 2.5 Comparison between the Perron eigenvector of the Markov chain transition matrix (in red) and the same information extracted from the trip data (in blue).
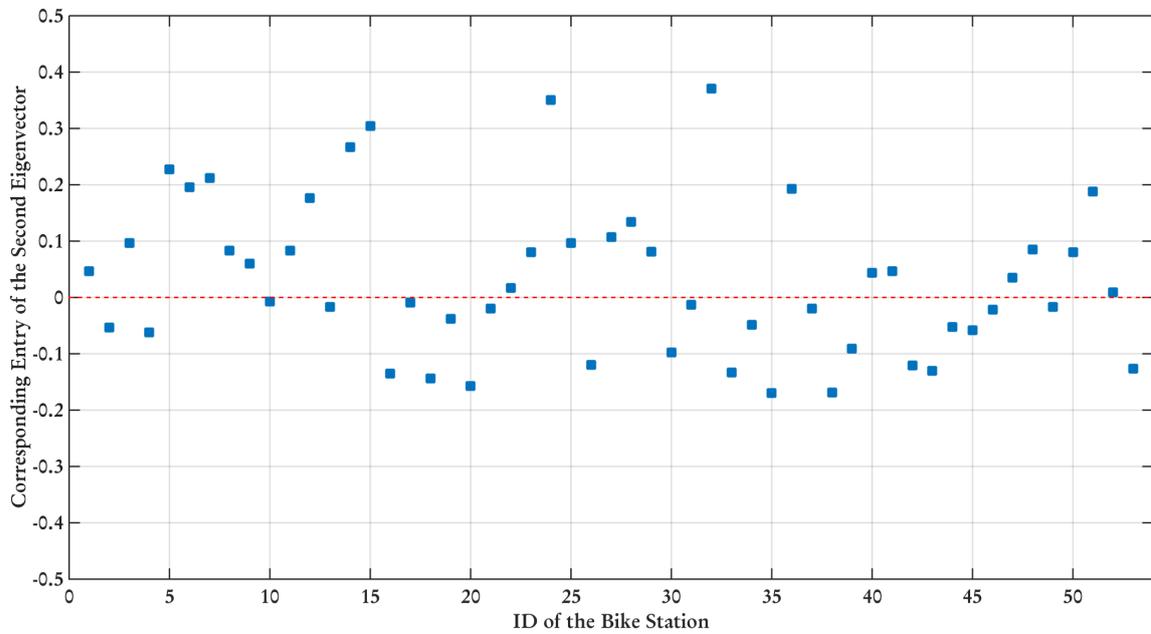
Fig. 2.6 Second eigenvector of the submatrix corresponding to the trips of vehicles.

the second eigenvector, and Figure 2.7 shows the positive and the negative entries of the second eigenvector with a different colour (the figure is taken from Google Map, using the information of the latitude and longitude of each single bike station, available from the data-set). The interesting property of the so-obtained clusters is that they do not exploit any geographical information about the position of the bike stations, but they only exploit the knowledge of the typical trips made by the customers, and thus, of the typical choice of origin-destination pairs.
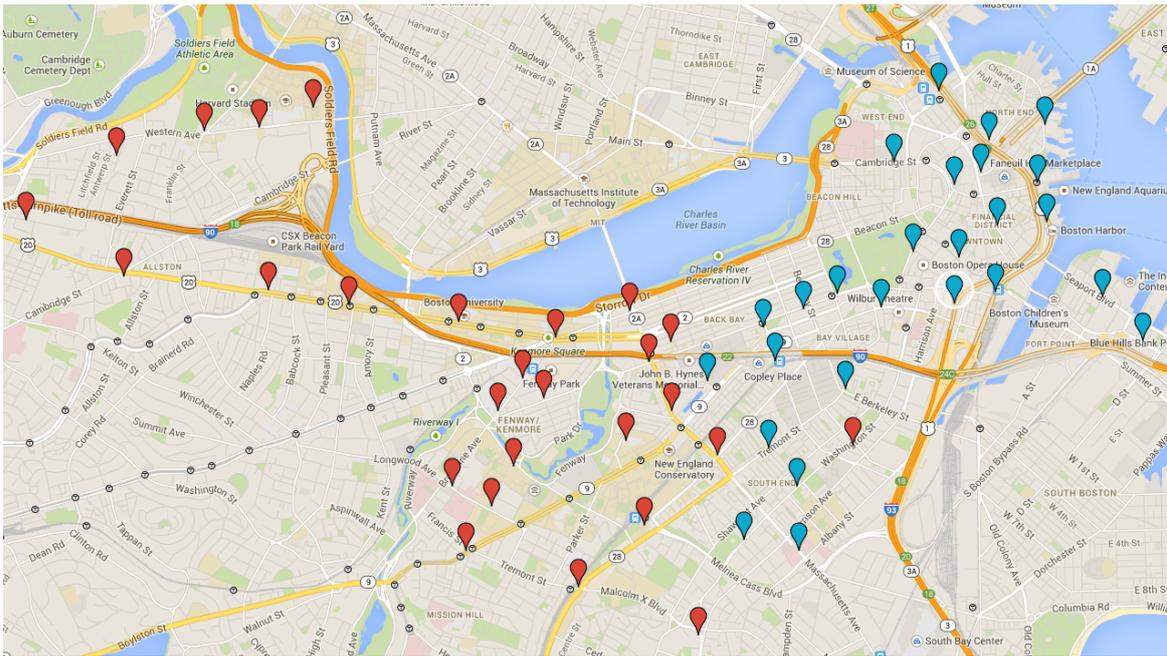
Fig. 2.7 Positive and negative entries of the second eigenvector do seem to provide a natural geographical characterisation of the bike stations in Boston.

As shown in the figure, the clusters do seem to provide a correct information, as the locations are in fact consistently situated in two geographically separated regions.

Remark 3 : In principle, it could occur that different clusters arise at different times of the day. In this case, we further computed the second eigenvector taking into account only the trips occurring in the morning, those in the afternoon, and still we obtained the same results. This suggests that in this specific application clusters appear not to be time-varying.

The knowledge of the clusters can be used for a number of applications: since most trips occur within a cluster, and more rarely from one cluster to another cluster, the number of bikes within a cluster should be constant most of the time. Accordingly, one could plan two separate trucks performing relocation actions in the two clusters. Also, since there are two

(quite) separate flows of trips within the two clusters, two advertising campaigns could be conducted in parallel in the two areas.

2.7   Conclusion

In this chapter we introduced the Markov chain and we saw its properties in transportation networks. We also defined some of the quantities of Markov chain and interpreted them in transportation networks. After that, this chapter proposed a new Markov chain based model to describe the functioning of a bike-sharing network. The model required two states for each bike station (i.e., one to take into account the bikes parked at the bike station, and one to consider all trips starting from that given bike station). We used real data from the Boston bike-sharing network to build our Markov chain transition matrix, and to validate the model with respect to the data. In the next chapter we shall see how Markov chain can be used to model a multi-modal network and how its properties can control the transportation network properties.

Chapter 3

Macroscopic multi-modal transportation

Abstract

The contribution of this chapter is to develop a Markov chain based approach to model multi-modal transportation networks. A realistic test-case using multi-modal data from the city of London is given to further support the ability of the proposed methodology to handle big quantities of data. Apart from the validation of the model, some issues regarding the efficiency of the network and some other applications are described. This work is a joint work with Arieh Schlote, Lorenzo Maggi, Emanuele Crisostomi, and Robert Shorten and has been accepted to be published in International Journal of Control, see [32]. Some parts of it were presented in the International Conference on Connected Vehicles and Expo (ICCVE) 2014 [31].

3.1   Introduction

The objectives of this chapter are twofold: (i) first, we want to illustrate how the approach of [19] can be adapted to analyse a multi-modal public transport network. For the purpose of clarity, we illustrate our approach using a bus network in a small area of Dublin as an initial example throughout the chapter. We later show how the same strategy can be applied to a general multi-modal transportation network, through a more realistic example in the city of London; (ii) second, we want to show how the proposed model can be used to leverage appropriate control actions to improve the performance of a multi-modal transportation network.

Section 3.3 of this chapter validates the model using the mobility simulator SUMO (Simulator of Urban MObility) [69], which is well-known in the transportation community, and shows that the results from the Markov chain approach are consistent with those obtained via simulation. The same result is obtained when the model is validated over some data related from bus and Tube data from the transportation network of London in Section 3.4.

Section 3.5 describes how the developed big-data model can be used as a platform to deliver control actions to improve the quality of the public transportation service.

3.2   Models of transport networks

We now describe how to fill the entries of the Markov chain transition matrix to model a transportation network. For the sake of clarity, we shall use an example from the small area of Dublin shown in Figure 3.1. The blue icons in Figure 3.1 correspond to bus stops. We can depict them as in the graph in Figure 3.2 where consecutive bus stops are connected through an edge yielding 17 nodes connected by 21 edges.

Fig. 3.1 Area of Dublin city centre used for analysis and simulations.
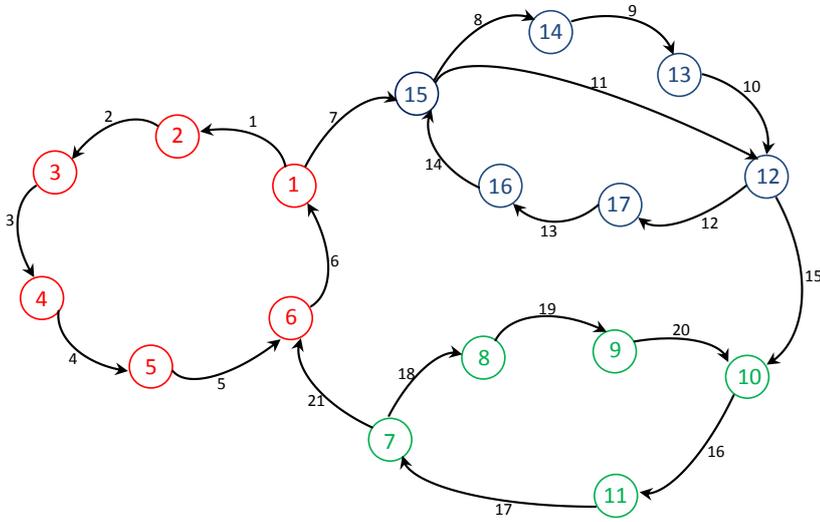
Fig. 3.2 Graph of the bus network.

3.2.1 Waiting graph

A Markov chain transition matrix corresponding to the graph shown in Figure 3.2 can easily be constructed from collected data according to the procedure outlined below:

(i) Each diagonal entry $\mathbb{P}_{ii}$ of the transition matrix is computed as $\mathbb{P}_{ii} = (t_i - 1)/t_i$, where $t_i > 1$ is the average time that people spend at the i'th bus stop waiting for the bus, so that the expected time before leaving the i'th state equals the average waiting time. The waiting time t_i can be expressed in any unit of measurement, for instance in seconds, in which case a step of the Markov chain corresponds to one second.

(ii) The value of each off-diagonal entry $\mathbb{P}_{ij}$ of the transition matrix is proportional to the proportion of passengers that travel from the bus stop i to bus stop j as the next bus stop. This implies that if bus stops i and j are not directly connected (e.g., bus stops 1 and 7 in Figure 3.2), then $\mathbb{P}_{ij} = 0$.

(iii) We add an extra state to denote the people that leave the bus network, and we call this the 'idle state' and denote it by S_{n+1}. Accordingly, entry $\mathbb{P}_{i,n+1}$ takes into account the proportion of people who leave the network after having reached bus stop i; similarly, entry $\mathbb{P}_{n+1,i}$ denotes the proportion of travellers who start their journey from the i'th bus stop; finally, we set the diagonal entry as $\mathbb{P}_{n+1,n+1} = (t_{n+1} - 1)/t_{n+1}$, where t_{n+1} corresponds to the inter-arrival time of passengers in the bus network.

(iv) We first set the diagonal entries of the transition matrix $\mathbb{P}$ as previously described. Then, we scale the off-diagonal entries in order to make matrix $\mathbb{P}$ row-stochastic.

The transition matrix constructed this way has the useful property that its Perron eigenvector corresponds exactly to the density of people at bus stops waiting for buses. This is analogous to what had been previously found in [19] in the case of vehicular density, and is validated through SUMO simulations in Section 3.3. The density of people at bus stops is computed by averaging the mean waiting times at bus stops weighted with the number of people waiting on average (i.e., if we have on average 3 people waiting for on average 10 minutes at bus stop *A*, and we have one person waiting for 30 minutes at bus stop *B*, then we have equal densities of people at the two bus stops).

The previous transition matrix can be built by collecting waiting times at bus stops (for the diagonal entries), and by checking how many passengers are on each bus (to build the off-diagonal entries and the entries of the idle state). However, we have not considered travel times so far. This information can be neglected if, for example, one is interested in making waiting times uniform all over the city. In other applications they have to be taken into account, as will be explained in Section 3.5. The next section illustrates how a transition matrix can be built to take travel times into account. To make a distinction, we will refer to the "waiting graph" (or "waiting transition matrix") when referring to the graph considered in this section, while the graph in the next section will be denoted as the "travel graph" (or "travel transition matrix").

3.2.2 Travel graph

Let us consider a new graph whose nodes are given by the existing direct connections between two consecutive bus stops, and the edges are given by the possibility to pass from one connection to a second connection. For instance, in the example of Figure 3.2, the new graph is shown in Figure 3.3. Accordingly, note that the new nodes in Figure 3.3 correspond to the edges in the previous Figure 3.2. This graph is sometimes denoted as the dual of the previous one [100].

A Markov chain transition matrix corresponding to the graph shown in Figure 3.3 can be easily constructed from collected data, according to the procedure outlined below. We shall denote such a second travel transition matrix as $\mathbb{P}^{(t)}$ for clarity.

(i) Each diagonal entry of the transition matrix $\mathbb{P}_{ii}^{(t)}$ is computed again as $\mathbb{P}_{ii}^{(t)} = (t_i - 1)/t_i$ where now t_i is the average time that people spend along the i'th bus connection, computed as the sum of the time spent waiting for the bus and the time to actually travel until the next
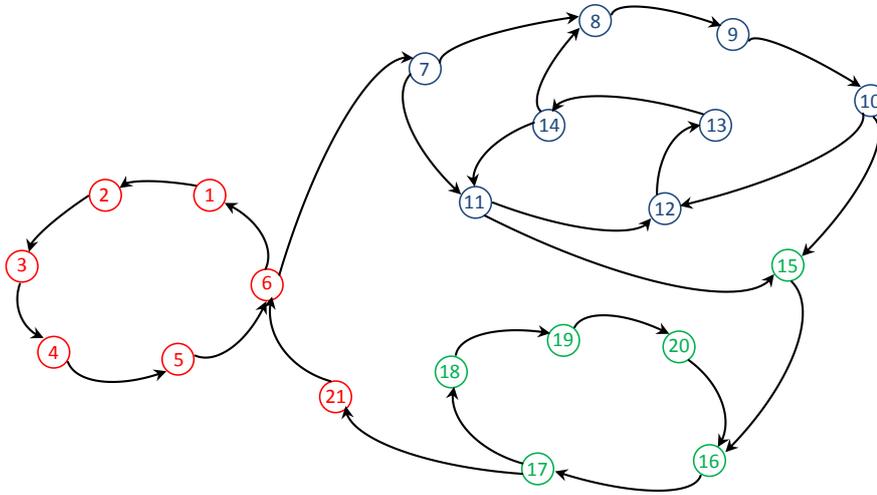
Fig. 3.3 Dual graph of the bus network shown in Figure 3.2.

bus stop.

(ii) The value of each off-diagonal entry $\mathbb{P}_{ij}^{(t)}$ of the transition matrix is proportional to the proportion of passengers that directly travel from the bus connection i to the bus connection j. This implies that if two bus connections i and j are not directly connected (e.g., connections 1 and 7, as can be seen from Figures 3.2 and 3.3), then $\mathbb{P}_{ij}^{(t)} = 0$.

(iii) We add an extra state to denote the people that leave the bus network. As before we call this the 'idle state' and denote it by S_{n+1}. Accordingly, entry $\mathbb{P}_{i,n+1}^{(t)}$ takes into account the proportion of people whose last travel in the bus network was connection i, and then they leave the network; similarly, entry $\mathbb{P}_{n+1,i}^{(t)}$ denotes the proportion of travellers who start their journey form the i'th bus connection; finally, we set $\mathbb{P}_{n+1,n+1}^{(t)} = (t_{n+1} - 1)/t_{n+1}$, where t_{n+1} corresponds to the inter-arrival time of passengers in the bus network.

(iv) We first set the diagonal entries of the transition matrix $\mathbb{P}^{(t)}$ as previously described. Then, we scale the off-diagonal entries in order to make matrix $\mathbb{P}$ row-stochastic.

The transition matrix constructed according to the previous procedure has the useful property that its Perron eigenvector corresponds exactly to the density of people along each bus connection. Such a density takes into account both people waiting for taking a given bus connection, and people currently travelling on that bus. Such a result is confirmed from

experimental results in Section 3.3. Note that this second transition matrix requires the same information of the waiting transition matrix, plus the information of the average travel times between (all pairs of) two consecutive bus stops.

Comment : We make the assumption that in a transportation network it is possible to get from every possible node to any other possible node. This implies for instance that from a particular bus stop, one can get to any other bus stop with an appropriate sequence of buses. Such an assumption is realistic and holds for most transportation networks, and allows us to obtain strongly connected graphs, and thus irreducible transition matrices [71].

3.2.3   Multimodality

One of the the main advantages of the Markov chain model is that it can accommodate different means of transport without introducing significant changes to the proposed theory. In particular, we do not have to know a priori if a given node in the graph is associated with a bus stop, rather than with a train station or a metro stop. Clearly, if one can take advantage of different transport modes, then the density of people at bus stops can be balanced by supporting the bus network with another means of transportation (e.g., taxis) instead of simply increasing the frequency of buses; analogously, accessibility to a critical destination (hospitals) can be realized by supporting the network with a dedicated service of shuttle buses. We shall have more to say on such issues in Section 3.5.

3.3   Validation

3.3.1   Simulation

We validate our approach simulating the bus network in the small area of Dublin city centre shown in Figure 3.1. For simplicity we focus on buses only; as explained above other modes of transport can be incorporated into the graph easily.

The bus network consists of 17 bus stops, 21 connections between the bus stops, and 4 bus lines, and the corresponding waiting and travel graphs were given in Figures 3.2 and 3.3. We use SUMO, a popular open source traffic simulation software [69], to simulate the bus network and extract the data required to build the corresponding Markov chain, and to compare the simulations results with those obtained through the Markov chain approach. In our simulation, we assume that a sensor is installed at each bus stop to collect information regarding the time of the day at which every single bus stops at that bus stop to collect

passengers. Such information can also be collected from GPS enabled mobile phones. We assume that people start their journey at a random bus stop, chosen with equal probability, according to a Poisson process with expected inter-arrival time of 2 minutes. We made this choice as Poisson processes are well established to model bursty traffic. We choose the final destination of the passenger in a uniform fashion in a first simulation (i.e., every node is equally likely to be the final destination), or according to a different probability distribution in a second simulation, as it will be explained later. If more than one sequence of buses can be used to get to the destination, we assume that the passenger minimizes the number of required buses, or, in case of a tie, minimizes the number of bus stops, and in case of further tie, the passenger would simply take the first bus.

3.3.2   Perron eigenvector

As previously explained, the Perron eigenvector of the waiting transition matrix corresponds to the long-run fraction of time that a person spends in a given state. Thus, we computed the Perron eigenvector, deleted its last entry (corresponding to the time spent in the idle state, i.e., not at a bus stop, which was not interesting in this case), and renormalised the remaining entries so that they would sum to 1. Note that this vector corresponds to the density of people waiting at each bus stop. We computed the same quantity from the SUMO simulation, and the two densities are shown in Figure 3.4.a. Similarly, we repeated the same procedure in the travel graph, in which case we obtain a density of people that is proportional to the time that people spend for a given connection (waiting for the bus to arrive, plus travelling on the bus) and results are given in Figure 3.4.b. As can be seen from the two figures, the two densities are clearly the same in both cases.

Comment : Note that the Perron eigenvector corresponds exactly to the density of people at bus stops as we have assumed that there is no noise in the sensors. In practice, if there is some noise in the measurements of the sensors (e.g., some people getting on the bus are not counted), then there will be a difference between the real density of people and the Perron eigenvector.

3.3.3   Clusters of the multimodal network

A useful information regarding flows of bus passengers involves the analysis of frequent patterns and the identification of clusters. Here we define a cluster as a set of bus stops from which people unlikely travel towards other sets of bus stops. To clarify our point of view, using the bus network example in Figure 3.2, we say that three clusters exist if, for instance,
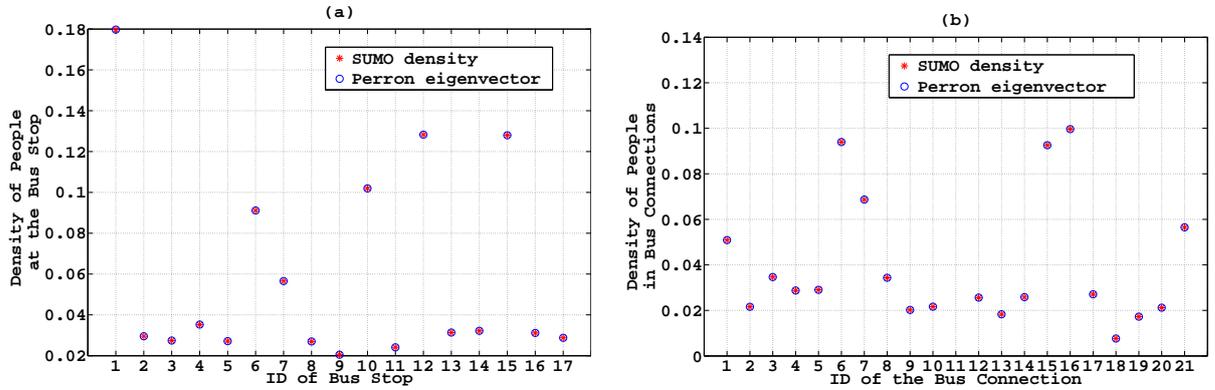
Fig. 3.4 The densities of people at bus stops (a), and at bus connections (b) computed from the Perron eigenvector of the Markov chain are the same of those computed in SUMO.

Entry	Value	Entry	Value	Entry	Value	Entry	Value	Entry	Value
(1,1)	0.9879	(1,2)	0.0041	(1,15)	0.0069	(1,18)	0.0011	(2,2)	0.9676
(2,3)	0.0262	(2,18)	0.0063	(3,3)	0.9646	(3,4)	0.0268	(3,18)	0.0085
(4,4)	0.9732	(4,5)	0.0218	(4,18)	0.005	(5,5)	0.9652	(5,6)	0.0262
(5,18)	0.0085	(6,1)	0.0218	(6,6)	0.9764	(6,18)	0.0018	(7,6)	0.0207
(7,7)	0.9686	(7,8)	0.0069	(7,18)	0.0039	(8,8)	0.9779	(8,9)	0.0160
(8,18)	0.0061	(9,9)	0.9697	(9,10)	0.0204	(9,18)	0.0098	(10,10)	0.9822
(10,11)	0.0163	(10,18)	0.0015	(11,7)	0.0657	(11,11)	0.9241	(11,18)	0.0101
(12,10)	0.0097	(12,12)	0.9856	(12,17)	0.0032	(12,18)	0.0015	(13,12)	0.0129
(13,13)	0.9802	(13,18)	0.0068	(14,13)	0.0128	(14,14)	0.9819	(14,18)	0.0053
(15,12)	0.0098	(15,14)	0.0028	(15,15)	0.9857	(15,18)	0.0017	(16,15)	0.0145
(16,16)	0.9801	(16,18)	0.0052	(17,16)	0.0142	(17,17)	0.9785	(17,18)	0.0072
(18,1)	0.0083	(18,2)	0.0090	(18,3)	0.0085	(18,4)	0.0089	(18,5)	0.0075
(18,6)	0.0117	(18,7)	0.0083	(18,8)	0.0089	(18,9)	0.0081	(18,10)	0.0068
(18,11)	0.0071	(18,12)	0.0078	(18,13)	0.0089	(18,14)	0.0099	(18,15)	0.0059
(18,16)	0.0088	(18,17)	0.0086	(18,18)	0.008570				

Table 3.1 This table represents the non-zero entries of the transition matrix of the graph in Figure 3.2. These values are rounded by Matlab. It should be noted that the matrix includes the idle state so it has 18 rows and 18 columns while the graph has 17 nodes. The perron eigenvector of this matrix is represented in Figure 3.4.a.

people mainly travel within the three sets of nodes having the same colour (i.e., within nodes 1 to 6, 7 to 11 and 12 to 17) and more rarely travel from one set of nodes to another set of nodes. This could happen if, for instance, we assume that each set of nodes belongs to a given neighbourhood that contains everything that the people need (e.g., shopping centres, cinemas, hospital, swimming pool), and more rarely people have the necessity to travel to another set of nodes.

If clusters exist, then the MFPTs should be very low among nodes belonging to the same clusters, and high if the origin and the destination belong to two different clusters. This simply follows from the fact that random walks are more likely to occur within the same set than from one set to another one. Also, if clusters exist, it should be possible to identify them using the second eigenvector. To simulate such a situation, we assumed that 90% (which is huge percentage) of the people would indeed choose their destination among one of the nodes belonging to the same cluster of origin.
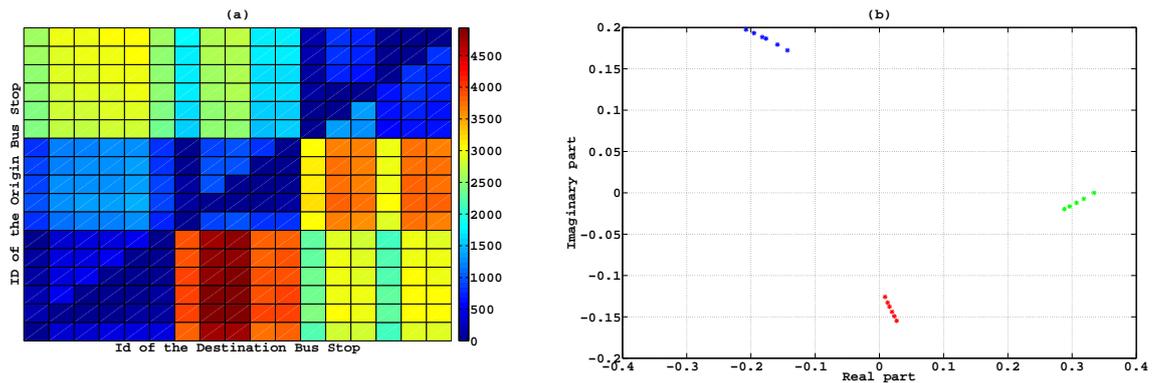
Fig. 3.5 The MFPT matrix and (especially) the second eigenvector are useful in identifying clusters. MFPTs are low among nodes belonging to the same clusters (a); similarly entries of the second eigenvector do identify what bus stops belong to what cluster (b).

As can be seen from Figure 3.5.a, mean first passage times are indeed low (blue colour) among nodes belonging to the same cluster, and are higher (brighter colours) among nodes belonging to different clusters. Analogously, Figure 3.5.b shows the entries of the second (complex) eigenvector in a complex plane, and nodes belonging to the same cluster (shown with the same colours of those used in Figure 3.2) are clearly separated in the complex plane.

Comment : While it is very simple to compute the second eigenvector and check the (possible) presence of clusters in passengers' flows, it is not straightforward to obtain the

Entry	Value	Entry	Value	Entry	Value	Entry	Value	Entry	Value
(1,1)	0.9774	(1,2)	0.0160	(1,15)	0.0018	(1,18)	0.0047	(2,2)	0.9702
(2,3)	0.0227	(2,18)	0.0072	(3,3)	0.9706	(3,4)	0.0232	(3,18)	0.0062
(4,4)	0.9754	(4,5)	0.0185	(4,18)	0.0062	(5,5)	0.9692	(5,6)	0.0216
(5,18)	0.0092	(6,1)	0.0159	(6,6)	0.9792	(6,18)	0.005	(7,6)	0.0038
(7,7)	0.9711	(7,8)	0.0174	(7,18)	0.0077	(8,8)	0.9763	(8,9)	0.0171
(8,18)	0.0066	(9,9)	0.9680	(9,10)	0.0212	(9,18)	0.0108	(10,10)	0.9664
(10,11)	0.0245	(10,18)	0.0091	(11,7)	0.0323	(11,11)	0.9581	(11,18)	0.0097
(12,10)	0.0022	(12,12)	0.9821	(12,17)	0.0120	(12,18)	0.0037	(13,12)	0.0127
(13,13)	0.9824	(13,18)	0.0048	(14,13)	0.0107	(14,14)	0.9833	(14,18)	0.0060
(15,12)	0.0050	(15,14)	0.0052	(15,15)	0.9862	(15,18)	0.0036	(16,15)	0.0187
(16,16)	0.9752	(16,18)	0.0061	(17,16)	0.0202	(17,17)	0.9726	(17,18)	0.0072
(18,1)	0.0084	(18,2)	0.0089	(18,3)	0.0092	(18,4)	0.0094	(18,5)	0.0077
(18,6)	0.0116	(18,7)	0.0083	(18,8)	0.0091	(18,9)	0.0080	(18,10)	0.0070
(18,11)	0.0073	(18,12)	0.0077	(18,13)	0.0088	(18,14)	0.0098	(18,15)	0.0059
(18,16)	0.0088	(18,17)	0.0087	(18,18)	0.8555						

Table 3.2 This table represents the non-zero entries of the transition matrix of the graph in Figure 3.2. These values are rounded by Matlab. It should be noted that the matrix includes the idle state so it has 18 rows and 18 columns while the graph has 17 nodes. The eigenvector corresponding the second largest modulus eigenvalue of this matrix is represented in Figure 3.5. The second largest eigenvalue of this matrix is complex and is equal to $0.9989 + 0.0005i$.

same information in another way, either from simulation results or even from collecting real data.

3.4   Big-data example

The objective of this section is to further validate the proposed methodology in a more realistic transportation network, consisting of 28 stops belonging to two tube lines and two bus lines of London, namely, Bakerloo Metro, Victoria Metro, Bus 13 and Bus 390. In particular, the chosen graph is shown in Figure 3.6.
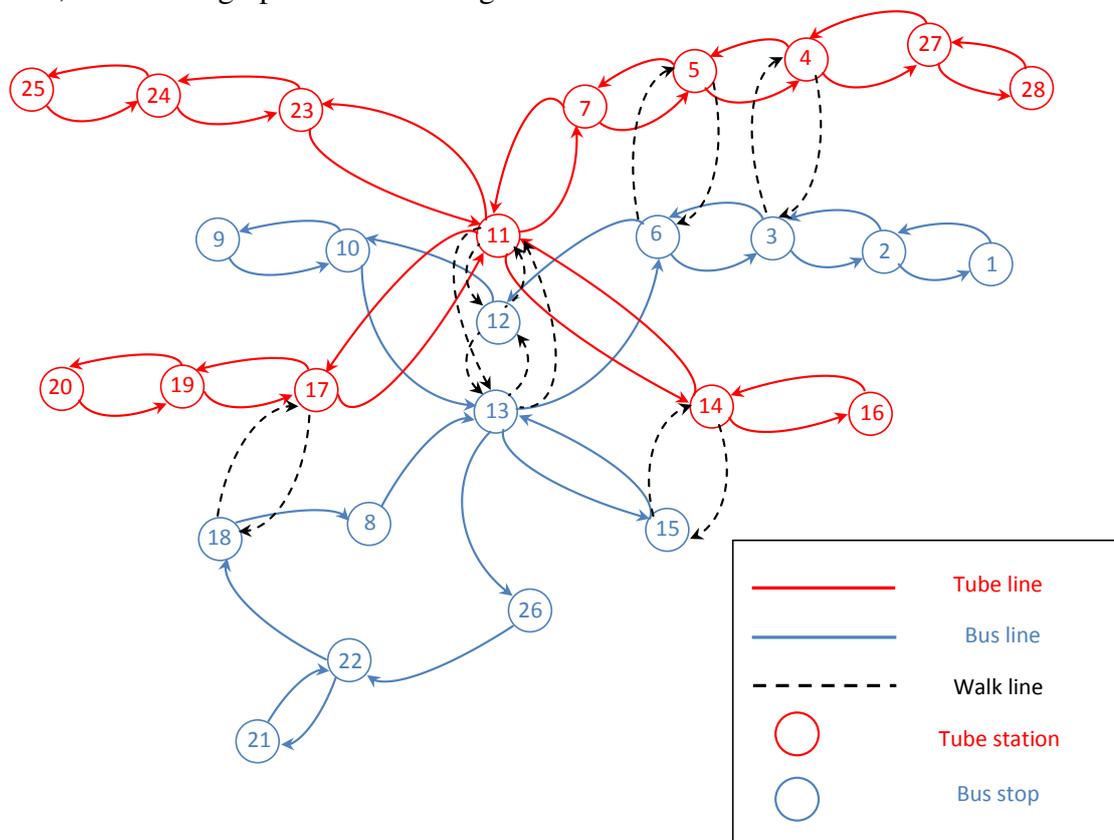
Fig. 3.6 Subgraph of London transportation network, consisting of two bus lines and two tube lines, used for a more realistic validation of the proposed methodology.

Note that we added some "walking edges" to connect bus stops with tube stops, and vice versa, to take into account the passengers that take a connection between the two different means of transportation, indicated with dashed lines in Figure 3.6. We then assumed that people would appear at bus/tube stations according to a Poisson process whose average

Table 3.3 Average number of people entering and exiting tube/bus stops in London in weekdays (data of year 2012). In Italic, the bus stops.

Node	Stop	Entries	Exits	Node	Stop	Entries	Exits
1	*Tufnell park station*	316	196	15	*Piccadilly circus/Regent street*	48	440
2	*Camden road/Breckrock road*	279	142	16	Waterloo/Waterloo East	139941	146314
3	*King's cross/st Pancras International station*	328	235	17	Baker Street	45141	43559
4	King's cross/st Pancras International Station	122769	120873	18	*Baker Street*	282	666
5	Eusten	56039	57197	19	Paddington	70505	72237
6	*Eusten*	131	341	20	Stone Bridge Park	4092	3929
7	Warren Street	28841	29195	21	*Firchly road Station*	132	736
8	*Orchard street/Selfridge*	722	625	22	*Swiss Cottage Cheese*	348	256
9	*Notting Hill gate/Palace Garden Terrace*	352	896	23	Green Park	51766	58893
10	*Marble arch Station*	359	372	24	Victoria	126275	130050
11	Oxford	119401	130658	25	Brixton	39375	36606
12	*Oxford*	359	517	26	*Dorset Square*	313	116
13	*Oxford*	215	514	27	Seven Sisters	21758	19300
14	Piccadilly circus/Regent street	59135	59279	28	Black Horse road	11956	10879

frequency of arrival was chosen consistently with the data provided by TFL[1], and summarised in Table 3.3. Similarly, the destination of the passengers was chosen proportionally to the exit data reported in Table 3.3. We then used a multi-modal journey planner to compute the shortest path (in terms of travel time) from origin to destination, with the only constraint that the path had to be fully contained in the map considered in Figure 3.6. We consider a period of two hours, namely, between 8 a.m. and 10 a.m., which according to data from TFL corresponds to a traffic of about 90000 people on average (i.e., 10 % of the total) among the considered bus/tube stops. As shown in Figure 3.7, we still have that the Markov chain well encapsulates the information related to the density of people at bus/tube stops, as, similarly to before, we are assuming that the sensors measuring people information are noiseless. Note that from Figure 3.7 we have a large number of people that spend a significant amount of time at stop 11. This is due to the fact that, for our choice of the subgraph of the London transportation network, stop 11 is the only stop that is common between the two chosen tube lines. Stop 11 is also connected with both the two bus lines (stops 12 and 13) via a short walk.

Finally, we show in Figure 3.8 the second eigenvector of the waiting graph, which makes a clear distinction between the stops of the tube network (red asterisks in figure), corresponding

---

[1]https://www.tfl.gov.uk/
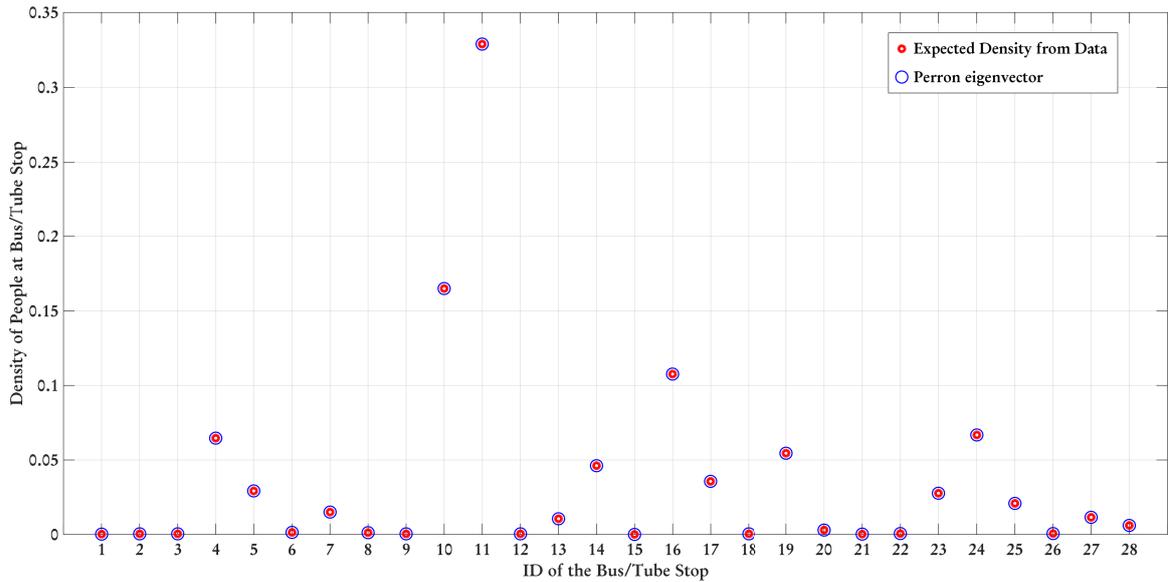
Fig. 3.7 Data from the simulation match data from the Markov chain model.

to the positive entries of the second (real) eigenvector and those of the bus network (blue dots in figure), corresponding to the negative entries of the eigenvector. This distinction is due to the fact that most of the trips consist of taking a single means of transportation, while more rarely a second different means of transportation is taken, thus it is possible to identify two main clusters of stops.
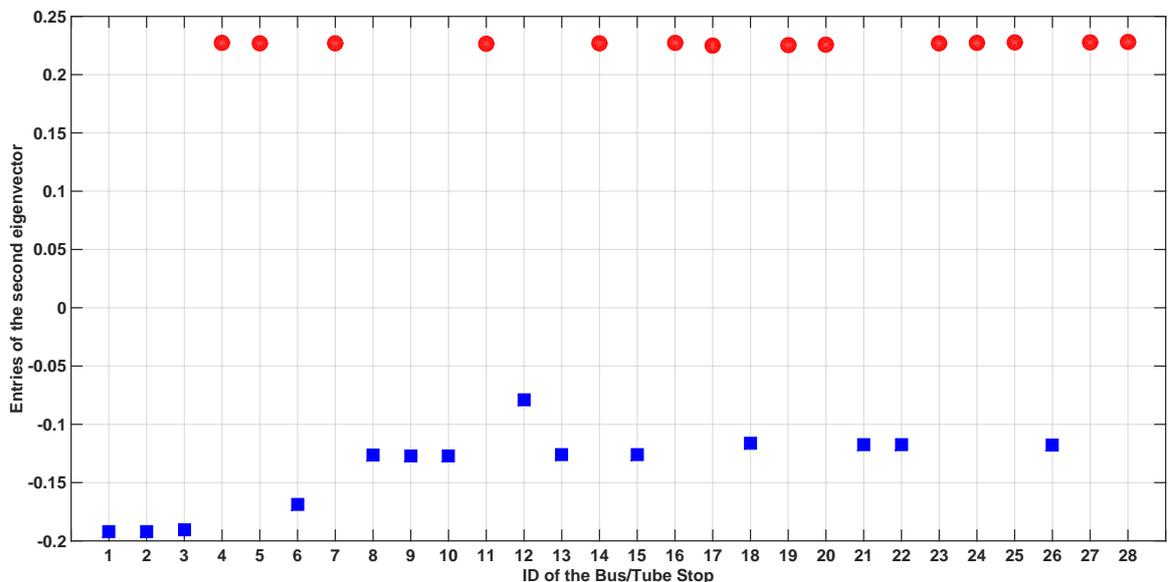
Fig. 3.8 The entries of the second (real) eigenvector distinguish tube stops (positive red circles) and bus stops (negative blue squares).

3.5 Markov chain based control applications to improve the public transportation network

The question as to how to measure a good network is a somewhat controversial topic. There are two basic stakeholders in the city. The first is the user of the transport network; he or she wants a good quality of service always (fast, clean, reliable service). On the other hand, the municipality is much more interested in aggregated average behaviour. Cities are concerned with issues such as:

(i) On average, how easy it is to travel from one part of the city to another?

(ii) On average, are certain spots accessible in an equitable manner from other parts of the city?

(iii) On average, are the travel times small between certain bus-stops?

(iv) Is it possible to identify emerging clusters in the bus network?

In the remainder of this section we focus on such issues and show how the proposed big-data model can be used as a platform to identify and implement practical control actions to improve the performance of the transportation network.

3.5.1 Node maintenance and control in the transportation network

The Kemeny constant is known to be a global indicator of the efficiency of a network [19, 88]. For instance, it can be used to evaluate the critical nodes of a transportation network. Typically, this can be done by picking out a node from the transportation network, and checking the efficiency of the residual network. This procedure can be carried out for all the nodes, and comparing the Kemeny constants obtained removing every single node. Those giving rise to highest Kemeny constants suggest that the removed nodes were indeed critical, as the residual network becomes less efficient.

The information on the most critical nodes of a mobility network is useful to implement a number of control actions:

- **Maintenance** - Special care should be devoted to maintain critical nodes always working properly, as a failure would greatly affect the efficiency of the remaining transportation network;

- **Road works** - When planning road works, one should be aware that temporarily disconnecting a critical node from the network would generally increase travelling times;

- **Strengthening critical nodes** - The public transportation network planner might want to strengthen critical nodes with redundant mobility services to improve their efficiency and their robustness.

To show the validity of the approach, we follow such a procedure for the usual bus network shown in Figure 3.2. We made the assumption that all the waiting times were the same at each bus stop, otherwise the Kemeny constant might give the obvious solution that the most efficient network is obtained by removing the least efficient node. In this way, the efficiency of the network is only given by its topology and by people's bus patterns. Accordingly, Figure 3.9 shows that the most critical bus stops of the bus networks are the bus stops from which people are allowed to change the bus. While such a solution could be easily expected from the simple bus network considered here by visual inspection, in realistic multi-modal transportation network the identification of critical nodes is not equally trivial.
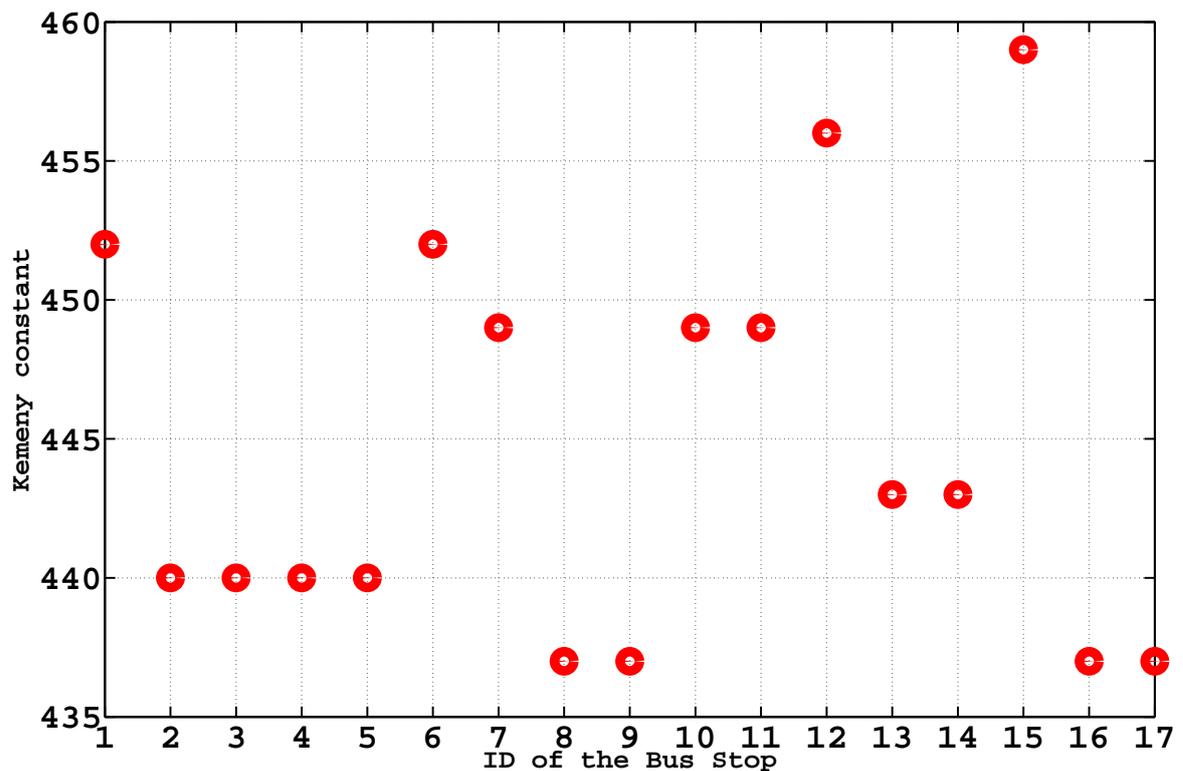
Fig. 3.9 The Kemeny constant shows that the most critical bus stops are 1, 6, 7, 10, 11, 12 and 15, i.e., those from which it is possible to change bus line.

3.5.2 Fair access control to critical areas

It is clear that in any functioning city some critical spots like hospitals should be easily accessible for all citizens. That is, they should be well-connected to all the neighbourhoods of the city, and accessible from any origin point. One way to ensure that hospitals are well-connected is to balance the average travel times from any point to the hospital, for instance making average travel times proportional to the distance (in meters) from the hospital; alternatively, one could use graph theory to increase the connectivity of the stops close to the hospital in a transportation network. As MFPTs are a way to take into account both average travel times (as they are an increasing function of average travel times) and also the topology of the network (a poorly connected network gives rise to high MFPTs and to a higher Kemeny constant as shown in the previous section), here we suggest that MFPTs can be used as an indicator of accessibility to some given areas.
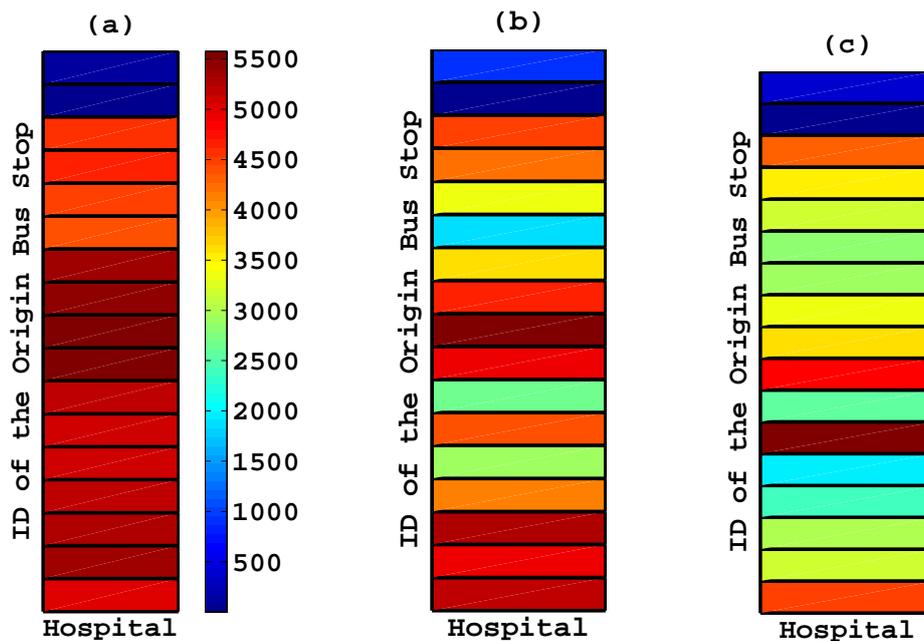
Fig. 3.10 In this example, the MFPTs to the hospital (a) are compared to the distances to the hospital (b). We say that the hospital is fairly connected if the two vectors have the same (normalised) values. We later try to improve the fairness of bus stops 5, 7 and 12 by supporting the bus network with two new lines of specific shuttle buses, obtaining a fairer result (c).

To give a practical example of how to control and ensure a fair accessibility to a critical key spot in the city, let us assume that node 16 in Figure 3.2 corresponds to the bus stop close

to the only hospital in the area. Figure 3.10.(a) then compares the $16'$th column of the MFPT matrix (MFPTs from any other bus stop to bus stop 16) with (b) the distance (in meters) from any bus stop to the $16'$th bus stop (the same scale was used for comparison purposes). Comparing the two figures, one can easily note that bus stops 5, 7 and 12 are those that are not well-connected to the hospital (i.e., MFPTs are high despite the path being relatively short). We now assume that the public transportation planner wishes to implement a control action to increase the fairness of connectivity to the hospital. The previously mentioned MFPT data can be used to predict that a fair access to the hospital can be achieved if, for instance, two fast shuttle-buses are added, one along the loop $2 - 5 - 16 - 2$ and another one along the loop $7 - 16 - 12 - 7$. The new primal graph is now shown in Figure 3.11.
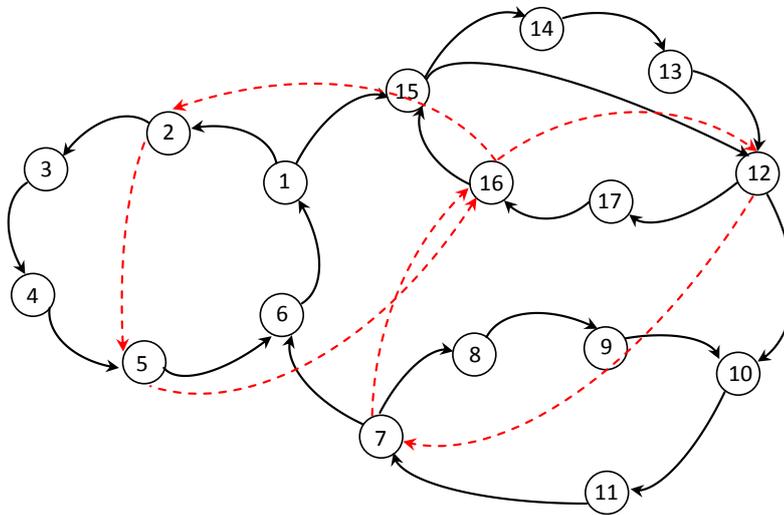
Fig. 3.11 The new transportation network where the bus network is further supported by two new shuttle-buses lines (shown with red dashed lines) to improve the connectivity of the hospital.

Correspondingly, we have that the accessibility of bus stops 5, 7 and 12 has increased, as predicted by the MFPT analysis, as shown in Figure 3.10.(c).

3.5.3 Balanced control of waiting and travel times

Another concern of network transport planners is to ensure that the travel time between stops is fairly distributed between destinations. This ensures, on average, a fair QoS delivered to network participants. Clearly, achieving fairness requires to control that the average waiting

times are the same at each bus stop. More realistically, one could balance *aggregated* waiting times (i.e., one takes into account how many people take the bus, and accordingly waiting times should be smaller where more people take the bus, and larger where fewer people take the bus). This last control objective exactly corresponds to balancing the entries of the Perron eigenvector of the waiting transition matrix. Also, one could expect that doubling the frequency of a bus at the i'th bus stop should imply that average waiting times at the i'th bus stop should be halved as well, and this is exactly what happens to the i'th entry of the Perron eigenvector. Accordingly, the entries of the Perron eigenvector automatically give to the mobility planner the expected optimal relative frequency of buses to achieve perfectly balanced aggregated waiting times.

We now give an example of this by trying to improve the balance of the bus network shown in Figure 3.4.a. As can be noticed from the figure, entries 1, 6, 10, 12 and 15 were the largest entries of the Perron eigenvector. Accordingly, we now double the frequency of the bus line serving stations (1-15-12-10-11-7-6), and the new, more balanced, Perron eigenvector is shown in Figure 3.12. Note that since a single bus line serves more stations, it is obviously impossible to arbitrarily control the frequency with which single bus stops are served (unless we assume that single buses can be used for point-to-point connections).
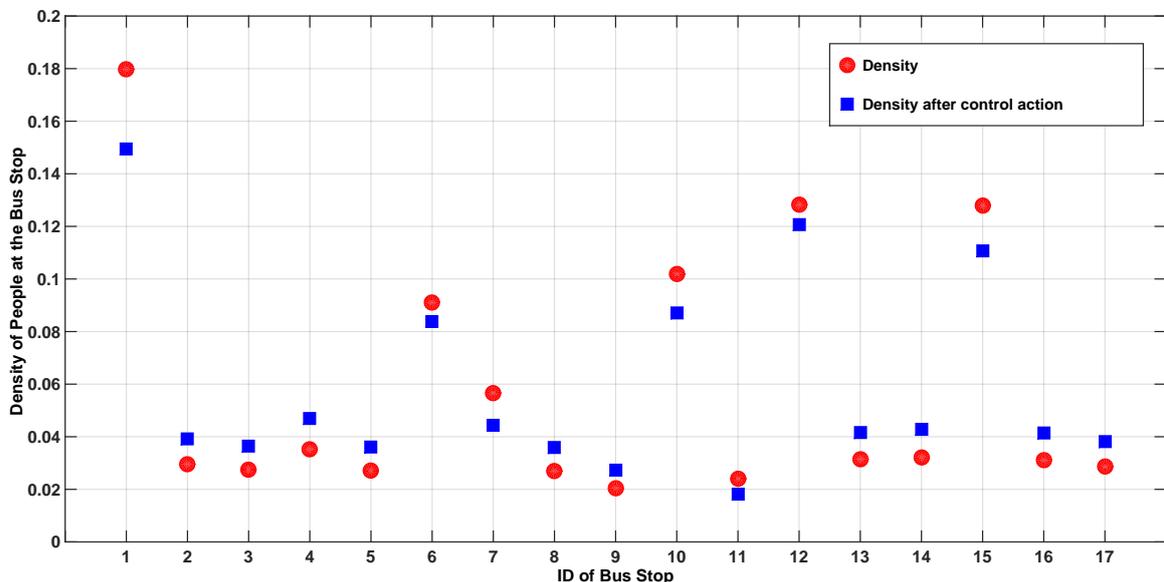
Fig. 3.12 After the control action (i.e., doubling the frequency of the bus line 1-15-12-10-11-7-6), the densities of people at bus stops are better balanced.

Also notice that more bus lines might serve the same bus stops (e.g., bus stop 1). Accordingly, changing the frequency of a single bus line only affects a subset of the people waiting

for the bus at bus stop 1.

In some cases, the network planner might be interested in obtaining another specific distribution of waiting times, for instance to take into account queues of people at bus stops, but in such a way that a given threshold of waiting time is never exceeded. In fact, the bus network would not be efficient if in some circumstances a small number of people would have to wait an unacceptablly long time for the bus. In such cases, it is not obvious what the optimal frequency of buses and the optimal network topology are to achieve a target density of people at bus stops. However, some Markov chain tools are available for finding such results. For example one of these tools is load balancing for Markov chains which is discussed in [64].

3.5.4 Clustering, services and advertising control

As a final control application of our model we now consider the identification of clusters in the network. Recall, clusters are a function of networks, bus routes, population movement, and demographic information. By filtering the population appropriately, information can be extracted from the population about the behaviour of demographic groups. This information can be used to provision bus services, or as part of targeted advertising campaigns. Here the basic idea is to control the adequate spread of information in the network among a particular group. Clearly, clusters are important in this context and should be targeted by advertisers to ensure rapid information dissemination. Similarly, critical nodes can be use as part of targeted health campaigns (flu vaccination).

Section 2.6 provided a justification for the use of the second eigenvector to identify clusters in a transport network, and we recall here that clusters do not only depend on the topology of the network, but also on how people do use the transport network (e.g., how often they travel from one area to other areas). The information of clusters can be conveniently used for a number of control applications, and referring to network planning and city management, they can be also used

- to design transport routes within clusters, and to minimize the use of transport resources to connect the clusters. Clusters would correspond to sub-cities within the whole city (e.g., neighbourhoods);

- when planning the construction of new facilities, one could focus on what facilities are missing in what neighbourhoods (e.g., if one plans to open a new pharmacy, it could be convenient to check if one cluster is missing a pharmacy);

- finally, the same information regarding clusters could be given to some interested service providers as a means to link clusters (e.g., taxi companies, car rental companies, adverstising companies, etc.).

3.6   Conclusion

In this chapter a Markov chain approach was developed to model multi-modal transport networks. According to the network two graphs were designed, so Markov chain transition matrices were constructed corresponding these graphs. Some information collected from the transport network like waiting times of people at the stations, travel times between the stations, the number of people entering and leaving stations etc) was used to build the transition matrix of our model. The model was then validated using the mobility simulator SUMO, and some data available from the multi-modal transportation network in London. Then, some applications of effcient network control like identifying the critical stations, fair access control to critical area and balanced control of waiting and travel times were outlined to demonstrate the potentials of the proposed model. Future work will further investigate the described applications, and extend the model to incorporate data over multiple time-scales. In the next chapter we investigate the quality of the data to be able to know how trustworthy the data is.

Chapter 4

An algorithm for extracting quality of traffic data

Abstract

In this chapter, a set of indicators are defined to extract the quality of traffic data to support the provision of a high level transport service in Intelligent Transportation Systems. Bootstrap, fuzzy techniques, and the fundamental diagram are used to define the indicators to be able to evaluate different aspects of traffic data. This is a joint work with Francesco Alesiani and the patent is filed in NEC Laboratories Europe in Heidelberg. This work described here has been presented at the 18th IEEE Conference on Intelligent Transportation Systems (ITSC) 2015.

4.1   Introduction

In Chapters 2 and 3 of this thesis, we have been using real data from Boston and London to validate our models and propose some applications. The importance of the quality of the data should not be ignored in the transportation research. Intelligent Transportation Systems (ITS) are daily used to optimize transportation in the cities and highways with respect to various criteria based on data measured by sensors. Managing and controlling the transportation systems is only possible with reliable data. Nowadays technologies allow for very accurate measurements, but errors are unavoidable since there are multiple sources, e.g. human configuration errors, communication outages, electromechanical failures, external factors, local environmental conditions to cite a few. The quality of data should be known by the users because they need to know how reliable their data and consequently their interpretations are. The purpose of this research is to extract the quality of the traffic data. We present here a set of techniques that are integrated to build up an overall traffic quality indicator. Figure 4.1 presents the overall overview of the proposed approach.

In order to distinguish any mistake of a loop system which is a system that saves traffic data, to check the detection accuracy, and to deal with sensitivity of detector card, an independent source is required as a baseline data to be able to compare with the loop
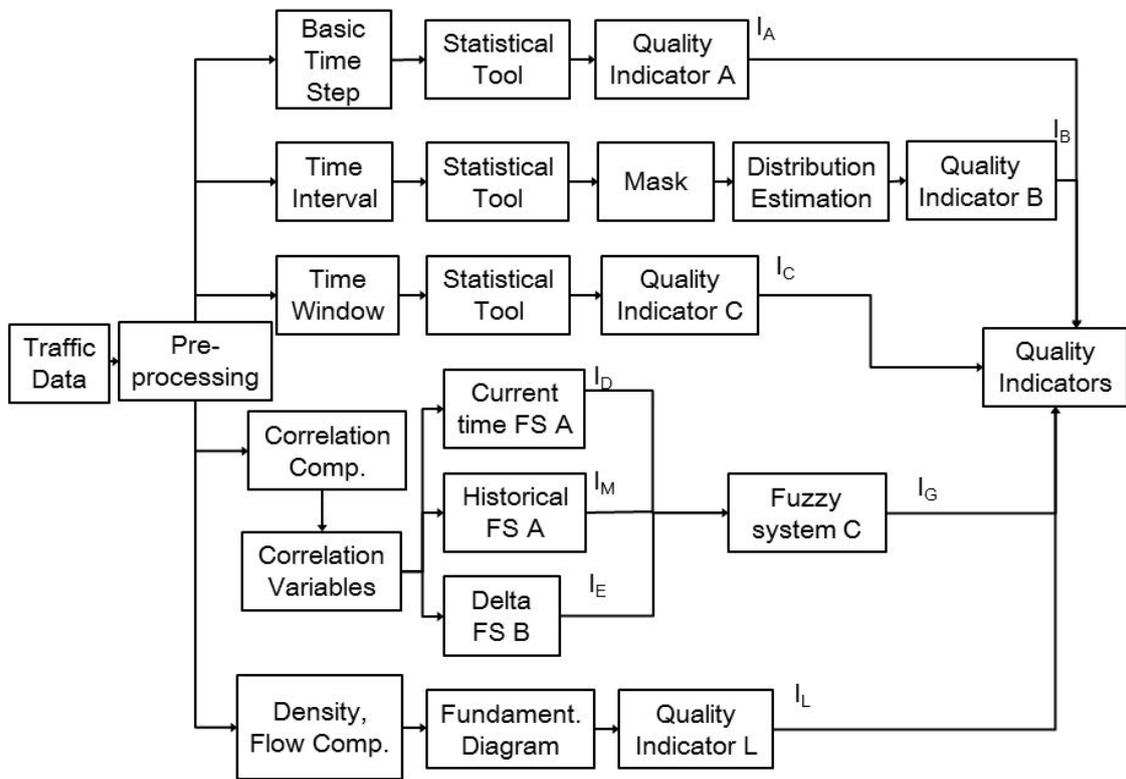
Fig. 4.1 The Indicators structure which is able to verify the traffic data against different hypotheses

detection system output. This comparison is an evaluation of the loop system. In [76], a development of a prototype system as the independent source is presented to achieve these goals.

A constrained nonlinear optimization approach is presented in [124] to identify the errors and modify loop detector data. The objective function and the constraints are defined such that the result follows the vehicle conservation principle.

A method that detects erroneous data samples of single-loop detector data is proposed in [15]. The algorithm used in this approach detects erroneous single-loop detectors from their volume and occupancy measurement. Decisions are made based on time series of many samples, not on a single sample, and an overall statistics of the measured day is reported based on four type of errors, namely zero-occupancy, zero-flow, steady-occupancy and randomness. The work also proposes a method for data imputation based on neighbor loop measurements. The model which is defining the relationship between neighboring loops is linear, and linear regression is used to estimate the value of missing or erroneous samples.

Many techniques have been developed that assign a single value for a missing data. These methods provide simple and fast estimates but the results produced might be biased. A multiple imputation method is designed in [93] which provides multiples estimates for a missing value by simulating multiple draws from a population.

Two spatial approaches are presented in [131] to detect the errors of loop detector data. The relationship between individual detector flows and station flows is built based on linear regression in the first method. The second method uses Kernel regression, see [91], to incorporate lane use percentages.

Practices for addressing traffic data quality are reviewed in [123]. Three types of user communities, real time traffic data collection, historical traffic data collection and other industries such as geospatial data sharing are considered.

In [80], the practice in the quality of traffic data extracted by Intelligent Transportation System is viewed from the perspective of both operations and planning personel which uses data for real time and not time intensive applications respectively. Quality issues, causes of the poor quality and possible solutions for low quality of traffic data are also discussed.

The white paper, [85], describes approaches that improve data quality through quality control. Contracting methods, standards and training for data collection are included in [85]. Data sharing between agencies and states, and traffic data detection methods are also considered.

The project described in [127] considers how such data can be best sorted and managed to accommodate multiple users. A method for quantifying data reliability is described and some techniques are proposed for imputing missing data. A general reliability index is defined which is indicating the consistency of each data measurement with fundamental traffic relations, historical data, and upstream/downstream measurements. The reliability index is an integer between zero and ten and it is obtained based on Continuous set theory which is developed by Lotfi Zadeh; see [134].

In [3] the authors proposed different schemes to verify the reliability of the messages sent over the Traffic Message Channel (TMC) by comparing it with infrastructure based data or with floating car data (FCD). The methods proposed are not aimed at assessing the quality of the single system alone, but by comparison with some reference system, either the traffic estimation derived by fix measures or by probe vehicles.

4.2   Problem definition

Nowadays Intelligent Transportation Systems are widely used to improve the performance of the transport system and increase the safety. The accuracy and reliability of the information affects the transport systems. Intelligent Transportation systems have evolved from single manned systems in which the system is closely managed; see Figure 4.2, to open market systems in which the information is exchanged by different users; see Figure 4.3. In both cases the data should be evaluated to be able to provide a reliable system. Generally the

quality aspect of the data may change depending on the user requirements. Data quality can be evaluated based on different criteria (such as accuracy or completeness) because users may have multiple expectations.
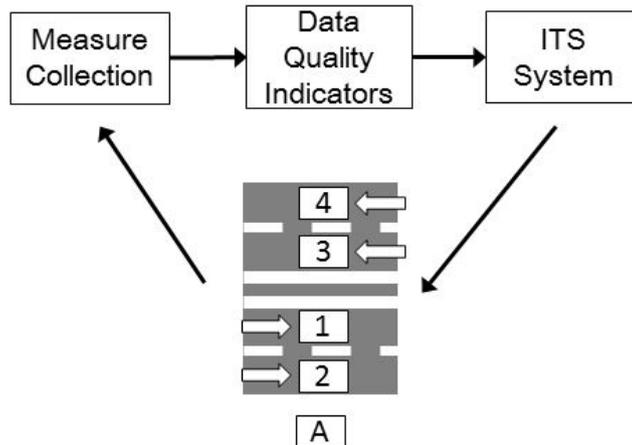
Fig. 4.2 Single operated system
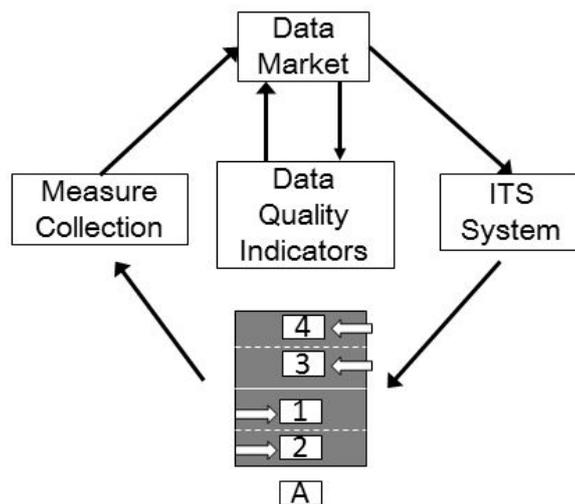
Fig. 4.3 Open System, Multiple actors acting in a Data Market System

Typical criteria are considered at the design of a road traffic measuring system and for its nominal behavior. Some quality criteria are evaluated only using some reference system

either in a closed testing facility or performing dedicated onsite campaigns. During operation the sensor or some part of the whole system may not function at the nominal levels. For this reason other online technology should be developed. Manual checks of the status of the sensor is a typical approach but is lengthy and costly. Further it does not scale with the size of the system and further prevents the use of more advanced ITS systems. In modern ITS systems, accuracy is increasingly important and performance of the system is directly connected to the provided quality. An indication of the reliability of the provided data is crucial in this case. Having recognized this requirement, this work addresses the online computation of the Confidence or Reliability Quality indicators of the provided data based on some consistency or rate of variance tests.

Data quality for ITS systems can be measured with respect to different aspects. These inclued the following:

- Accuracy: This concerns how close the current measure is to the real underlaying physical phenomenon;

- Completeness: This describes the presence of missing data;

- Validity: Validity is the period of measurement or validity of the provided data;

- Timeliness: It is the delay introduced between the physical phenomenon and the availability of the data;

- Coverage: This is the spatial availability of the data or the portion of the road that is monitored;

- Accessibility: It is the level at which the data is made available;

Existing online indicators include the check of some unrealistic conditions. For example, at a certain time step, the flow and the occupancy can have huge values and one of them cannot rule most of them out. During a day most detectors represent more or less the same pattern that shows high flow and occupancy in rush hours and low flow and occupancy at night. Most loops have this pattern as output but some loops behave very differently for example they show zero flow but high value of occupancy for several hours. It is obvious that these values are incorrect. Four types of these abnormal behaviour that are mentioned in [15] are:

- Non zero measure with zero occupancy;

- Non zero flow with zero density;

- Samples with high occupancy, especially with non zero flow;

- Entropy of the density;

More basic indicators are defined in [124]. These indicators measure:

- The delay in the provision of the data (online latency variation);

- The actual available sample over the expected (online availability).

4.3 Proposed solution

In this section different indicators with different purposes are defined to determine the quality of the data. The range of the quality indicators is $[0, 1]$. High and low value of the indicators represent high and low quality of the data respectively. In the first part, a single section (500 meter of road) is under consideration and the second part is about evaluating the data of a corridor (two or more sections).

Traffic measures are collected per road section. $x_s(i, d)$ represents the traffic measure on section s on time instant i on day d. The measured traffic quantity can be: q: traffic flow, ρ: vehicle density, v: speed or tt: travel time. Some of the proposed methods can be applied to some of these measures, while others use relationships among traffic quantities. $\mu_s(i) = \frac{1}{N_d} \sum_{d=1}^{N_d} x_s(i, d)$ is the mean with respect to the days of the same type. The single sample is also indicated as $X_i = x_s(i, d)$.

The bootstrap is a powerful method for quantifying the quality of estimators, (see [28] for a description of the technique), and is applied as the statistical tool to derive quality indicators. The basic idea of this method is to infer statistical properties about a population by re-sampling the sample data. If a statistical model of the data is not available and normal theory is not applicable, it is not possible to evaluate the statistical error of the sample against the underlying population. Bootstrap then uses sub-sampling of the data to generate the reference distribution of the underlying population. For this reason bootstrap has been applied in social related studies, since a detailed model of human behavior is not available.

4.3.1 Single section

Three indicators are developed in this section to quantify the quality of the data of a single section. The first and the second indicators extract the quality of the data at a single time step and during a time interval along the days respectively. The third indicator measures the consistency inside a time window around the original data.

Indicator of single time step

The indicator of the data of time step i is defined based on the mean and standard deviation, μ_{ib} and σ_{ib}, which are extracted by implementing the bootstrap on the data of the time step i along the days; see Figure 4.4. The indicator is then built as

$$I_A = \frac{\mu_{bi}}{\mu_{bi} + \sigma_{bi}} \tag{4.1}$$
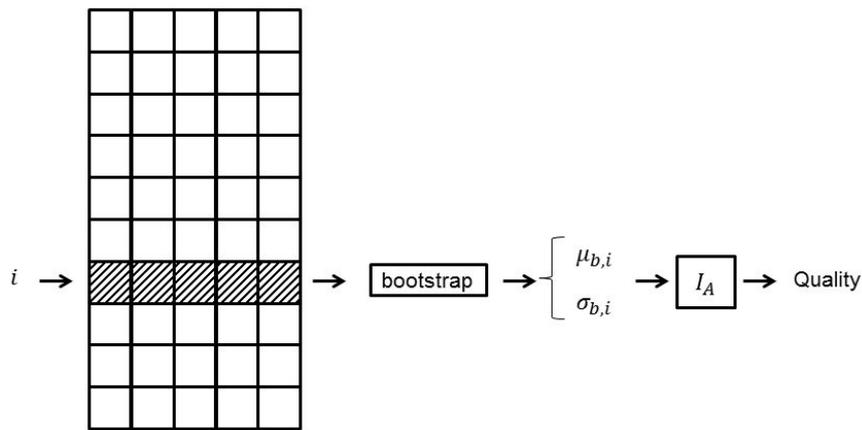
Fig. 4.4 The quality indicator of data of time step i along the days is defined based on the output of bootstrap.

It is clear that when the mean of the data is more accurate, then σ_{ib} has a lower value. Therefore, I_A is closer to 1 and it is representing a higher quality of the data.

This indicator cannot determine the low quality of data when there is a constant error in everyday data of the time step i, so the indicator of a time interval data, that shows low quality when there is such an error, is defined in the next section.

Indicator of time interval

Before defining the quality indicator of a time interval data, it is required to know how fast the data is changing. Formula (4.2) is defined to indicate the rate of fluctuation of the means of the data in a time interval, indicated by μ_i, μ_{i+1}, ..., μ_{i+k} in Figure 4.5.

$$F = \frac{\mu_{b\mu}}{\mu_{b\mu} + \sigma_{b\mu}} \tag{4.2}$$

$\mu_{b\mu}$ and $\sigma_{b\mu}$ are the mean and the standard deviation extracted by bootstrap when the inputs are the means of the data of basic time steps in the interval(μ_i, μ_{i+1}, ..., μ_{i+k}); see Figure 4.5.

When the data changes more rapidly, $\sigma_{b\mu}$ has a higher value. Therefore, the fluctuation value decreases. Fast change in the data can be related to a measurement error or to a physical phenomenon. These two cases must be distinguished to be able to quantify the quality of the data. We propose to build the mask either based on the history of the section or even from some reference section that sees similar traffic. In general the distribution of the mask is not symmetric.

In order to find the estimated not symmetric distribution, see equation (4.3), the standard deviations of the left and the right side, σ_L and σ_R, of the mean of the means of the fluctuation values extracted by bootstrap, μ_M, are estimated.
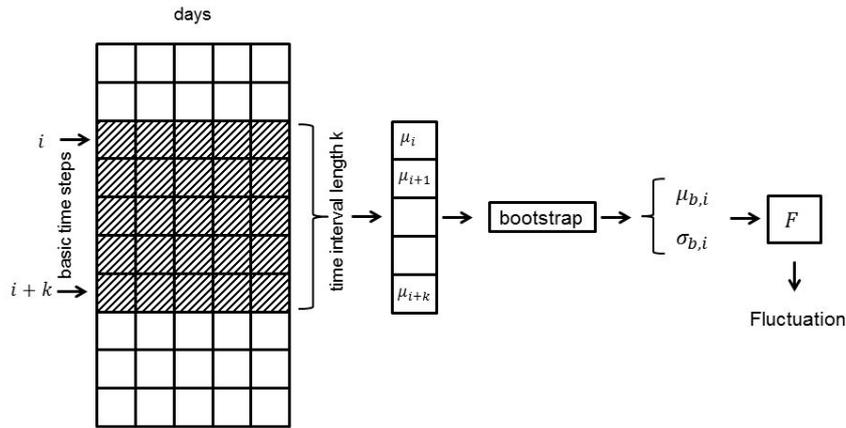
Fig. 4.5 The rate of fluctuation of the means of the data in a time interval. Implementing bootstrap is arbitrary.

$$
f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma_L} e^{\frac{-(x-\mu_M)^2}{2\sigma_L^2}}, & x < \mu_M \\[2ex] \dfrac{1}{\sqrt{2\pi}\sigma_R} e^{\frac{-(x-\mu_M)^2}{2\sigma_R^2}}, & x \geq \mu_M \end{cases} \tag{4.3}
$$

The asymmetric distribution is estimated using a function that computes density estimate by implementing a Kernel smoothing method [7] on the histogram of the mask. σ_L and σ_R are the values of the variable that equal $e^{-\frac{1}{2}}$ the maximum value of the estimated distribution.

Finally the quality indicator is a probability function and it is defined as

$$
I_B(F) = \begin{cases} 2\int_{-\infty}^{F} \frac{1}{\sqrt{2\pi}\sigma_L} e^{\frac{-(x-\mu_M)^2}{2\sigma_L^2}} \, dx, & F < \mu_M \\[2ex] 2\int_{F}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_R} e^{\frac{-(x-\mu_M)^2}{2\sigma_R^2}} \, dx, & F \geq \mu_M \end{cases}
\tag{4.4}
$$

which is the probability of the sample F to be compatible with the distribution of the mask.

Indicator of time window consistency

The indicator $I_C(x_i)$ evaluates the consistency of the data x_i within a time window and is defined as

$$
I_C(x) = 2Pr(X \leq x) = 2\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma_W} e^{\frac{-(X-\mu_W)^2}{2\sigma_W^2}} \, dX
\tag{4.5}
$$

in which μ_W and σ_W are the outputs of bootstrap when the input is the data of the window excluding the data under consideration; see Figure 4.6. Since the distribution of data is symmetric, the estimated normal distribution is used to define the indicator in (4.5).
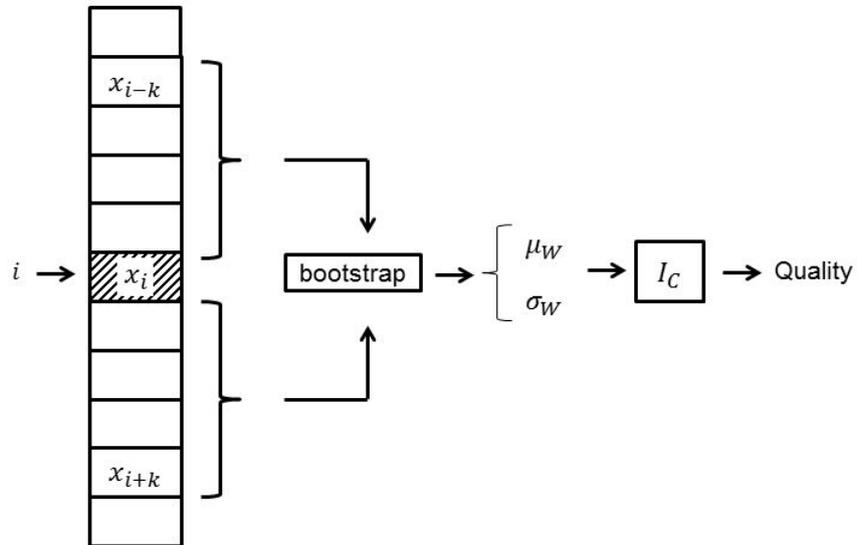
Fig. 4.6 Quality indicator of time window consistency

4.3.2 Corridor

In this section, analysis of data quality is based on the observation of two or more different sections. The indicator exploits the similarity of the measure between sections and it monitors abnormal variations of statistically relevant quantities. The indicator is defined using fuzzy sets based on the information derived from the correlation of two sections.

The correlation between sections S_1 and S_2 is calculated based on the following formula.

$$C_{S_1, S_2} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.6}$$

in which $z_i = x_{S_1}(i, d)$ and $y_i = x_{S_2}(i - \tau, d)$ are the vectors of data, for sections S_1 and S_2 respectively, of day d at time step i, while $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. The time delay between the data of the two sections is τ and n is the length of the vectors.
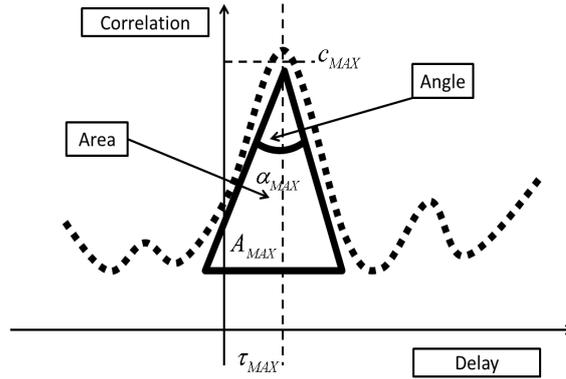
Fig. 4.7 C_{Max}, α_{Max}, A_{Max} and τ_{Max} are the information, obtained from the correlation between two sections, useful to find the quality of the data.

From the correlation function, see Figure 4.7, it is possible to identify some parameters. These include:

1. the maximum correlation value C_{Max};

2. the angle of the maximum correlation α_{Max};

3. the area of the maximum correlation A_{Max};

4. the delay, τ_{Max}, at which the maximum correlation happens.

Based on these parameters, some fuzzy sets and a fuzzy logic system are defined and the indicator is generated. τ_{Max} can be used to infer the delay when the data of the two sections is reliable.

Fuzzy system indicator

The fuzzy logic systems are defined by a set of fuzzy membership functions for the inputs and outputs and a set of rules to extract outputs from the inputs. The rules can be changed depending on the requirements and the criteria of the final indicator. In other words, the user may change the membership functions or the rules of the Fuzzy systems based on the aspect of the data that he is considering (for example here we are simply using triangular membership functions, but the user may want to have another type of function based on the application that he has). The quality indicator, which has a value in [0 1], is extracted from the intermediate outputs of three fuzzy systems. The structure of the fuzzy systems is depicted in Figure 4.8.

Fuzzy system A has output indicators of a single day, I_D, and the average of some days, e.g. a week, I_M. As input, C_{Max}, $cos(\alpha_{\text{Max}})$, and A_{Max} are used. Each of them has three membership function of type A, as depicted in Figure 4.9.

The output of the fuzzy system B, I_E, determines if the quality of a specific day is higher or lower than the quality of the average of the data of a week. The inputs of this fuzzy system are the differences of the correlation information between the considered day and the average values: ΔC_{Max}, $\Delta cos(\alpha_{\text{Max}})$, and ΔA_{Max}. The input ΔC_{Max} and $\Delta \alpha_{\text{Max}}$ have membership functions of type B; see Figure 4.10, and ΔA_{Max} has a membership function of type C; see Figure 4.11. The membership functions of the outputs of both fuzzy systems A and B are of type A. The range of the outputs is [0 1].

The fuzzy system C is proposed to integrate the output of the other fuzzy systems: 1) the indicator associated with the historical data; 2) the current data or 3) the difference between the current and the previous ones. The output of the last fuzzy system is the final indicator of the quality of the two sections, I_G. All the inputs and the output of this fuzzy system have the membership function of type A.
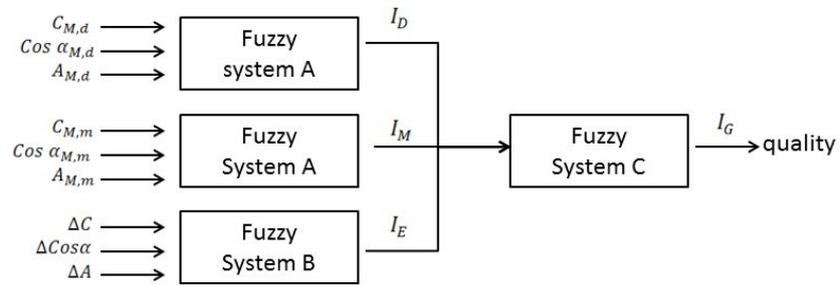
Fig. 4.8 The outputs of the fuzzy systems A and B are the inputs of the fuzzy system C which extracts the quality.
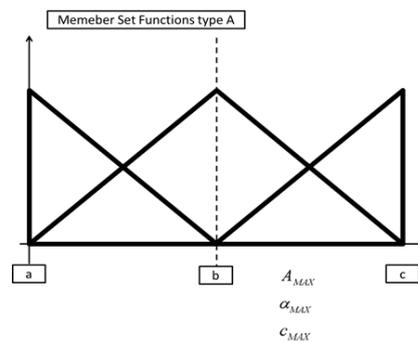
Fig. 4.9 Membership function type A.
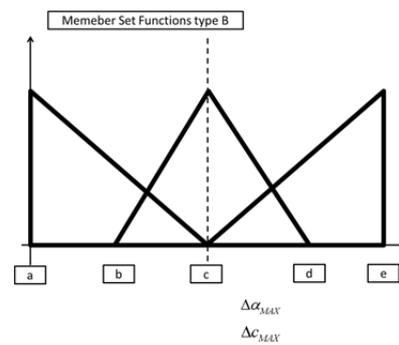
Fig. 4.10 Membership function type B,
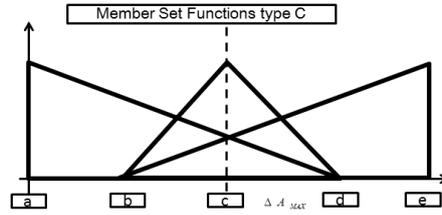
Fig. 4.11 Membership function type C.

Input 1	Input 2	Input 3	Output
$\|C_{M,d}\|$	$A_{M,d}$	$cos(\alpha_{M,d})$	I_D
$\|C_{M,m}\|$	$A_{M,m}$	$cos(\alpha_{M,m})$	I_M
$\Delta\|C\|$	ΔA	$\Delta cos(\alpha)$	I_E
1	1	1	1
2	2	2	2
3	3	3	3
1	3	3	3
2	3	3	3
3	1	1	1
2	1	1	1
3	2	1	3
3	1	2	3
3	3	1	3

Table 4.1 Rules of the fuzzy systems A and B

Input 1	Input 2	Input3	Output
I_E	I_D	I_M	I_G
2	1	1	1
2	2	2	2
2	3	3	3
3	2	1	3
3	1	1	3
3	2	2	3
1	3	3	2
1	1	3	1
1	1	2	1
3	3	1	3

Table 4.2 Rules of the fuzzy systems C

4.4   Simulation and Analysis

The proposed indicators are studied using highway traffic data. The data was confidential, so here we can only mention that the area of study is a road stretch including 4 lanes. The length of each section is 500 meters. Data is available for each minute of each lane for each day of the week excluding the weekend. Before extracting the quality of the data using the quality indicators defined, the data is aggregated at road level for two directions. The data has also been preprocessed by a finite impulse response (FIR) filter of $l = 10$. The formula used to smooth the data is

$$\frac{1}{(\sum_{j=i}^{i+l} a_j) + 1} \sum_{j=i}^{i+l} a_j x_j \tag{4.7}$$

where $a_j = 1$ if x_j is valid data and 0 otherwise. In the following sections the single indicators are shown graphically, in order to comparing the quality indicators in different traffic measurement conditions.

It should be mentioned that in this section we are not using the data of the same section. The quality of the data of different sections are represented here because as we said we define a set of indicators and each indicator has its own capabilities to show deficiencies of different aspects of data. In order to show the power of indictors for different aspects we picked appropriate section for the indictors to show the quality of data.

4.4.1   Single section quality indicator

Let us first focus on the single section quality indicators. Two sections of different quality level are used to show the capability of the indicators in describing the quality of the data.

One time step

Figure 4.12 shows the quality of the data of a time step along the days of one week according to the indicator (4.4) for the flow data of two different sections. One section has more invalid values. The indicator shows that the data of the section with more invalid data has lower quality compared to the other section.

Adding constant artificial error to the data of a few minutes but not for everyday of the week shows that the quality of the data decreases; see Figure 4.13. The problem is that adding the constant artificial error to the data of the whole week does not affect the quality extracted by the indicator (4.4) because it is extracting the quality of one minute independent from the other minutes. Therefore, the constant noise added to the data of a specific minute for the whole week is not recognizable by indicator (4.4); see Figure 4.13.
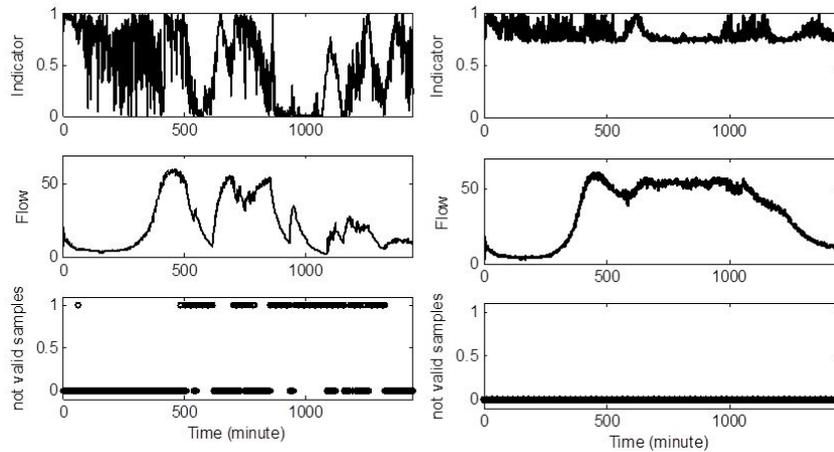
Fig. 4.12 Quality of each minute data along the days for two different sections. The number of negative values is the number of not valid data. Therefore, the negative values are not represented and not ised to plot the flow here.
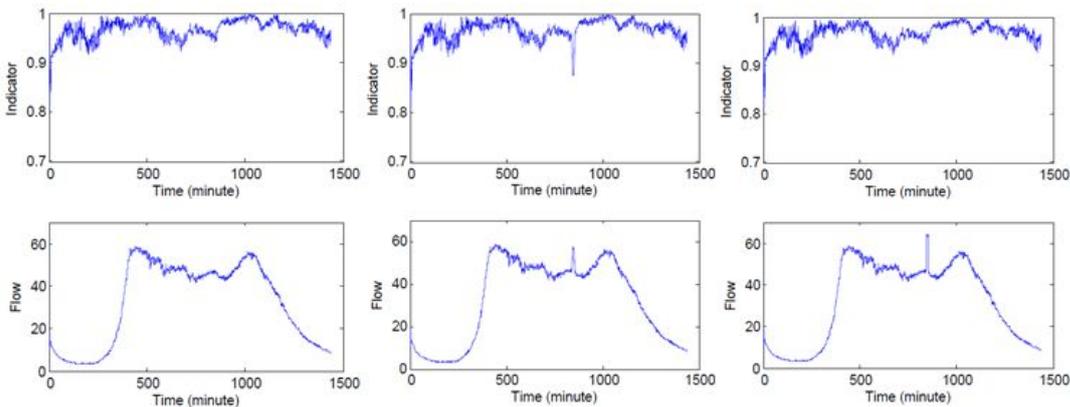
Fig. 4.13 The first column shows the flow with no artificial error and its quality based on the indicator I_A, the second column shows the flow with artificial constant error added not to the all days of the week. The indicator plot shows that the error is recognized by the indicator of a single minute data, I_A. In the third column, a constant artificial error is added to everyday data of the week. It can be seen that the indicator of a single minute is not capable of recognizing this error because the indicator has the same plot as in the first column.

Time interval

Figure 4.14 shows that the fluctuation rate, computed every 20 minutes, decreases when a constant error is present for all days or is present in only one day. The typical flow variations due to the night low traffic condition has also an effect on the fluctuation, but this is not related to low quality in the data.
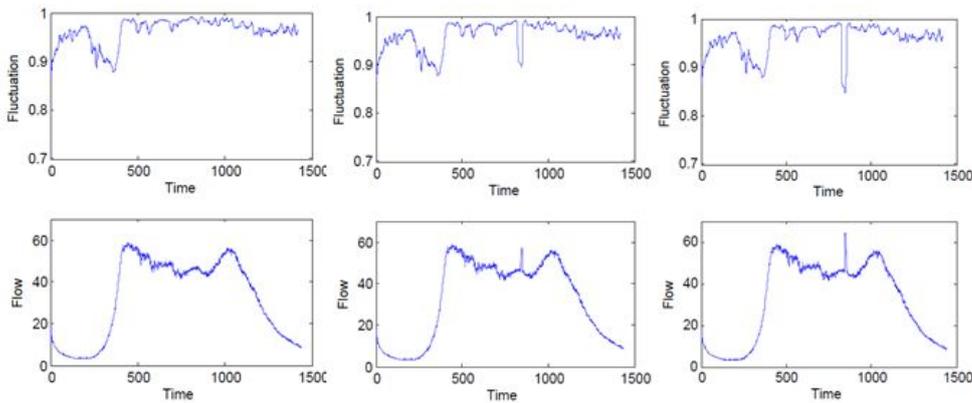
Fig. 4.14 The fluctuation value decreases when there is artificial error either in everyday data or not everyday data. Moreover, sharp change of flow which happens because of normal phenomenon cause decrease in the fluctuation rate.

In order to distinguish between normal traffic conditions and the effects due to external factors, a mask is constructed according to the history of other sections. In our test we used 5 other sections. The mask, generated by the average of the fluctuation values of some reference sections, and the quality indicator I_B of two different sections is represented in Figure 4.17.

The histogram of the mask for the 30th time interval and its density function, whose range is in [0 1] and is useful to estimate the σ_L and σ_R, are shown in Figures 4.15 and 4.16.

Window

In Figure 4.19 the quality indicator (4.5) of the data is plotted. The indicator is computed on the third day of the week, while the other days of the week are used to create the reference. On the analysis only week days are considered. From Figure 4.19 it is possible to notice that when the flow of the considered day deviates from the week average, the consistency indicator decreases.
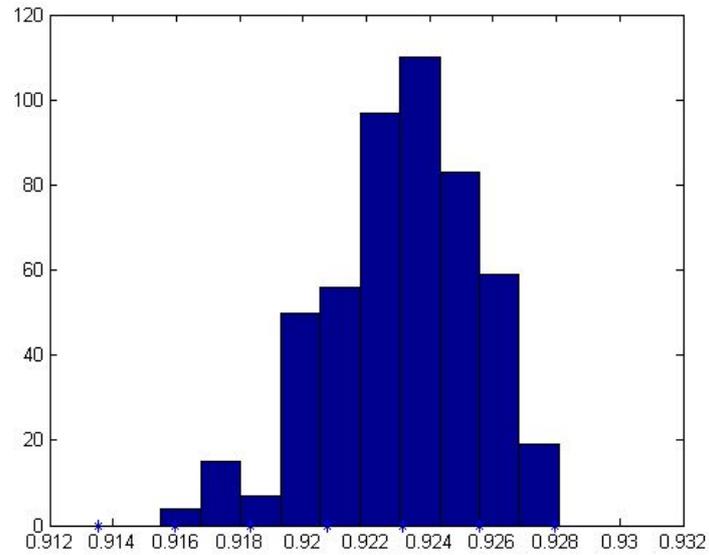
Fig. 4.15 Histogram of the means extracted by bootstrap implemented on the fluctuation values of the five sections for the 30th time interval. It is obvious that the distribution is not symmetric.
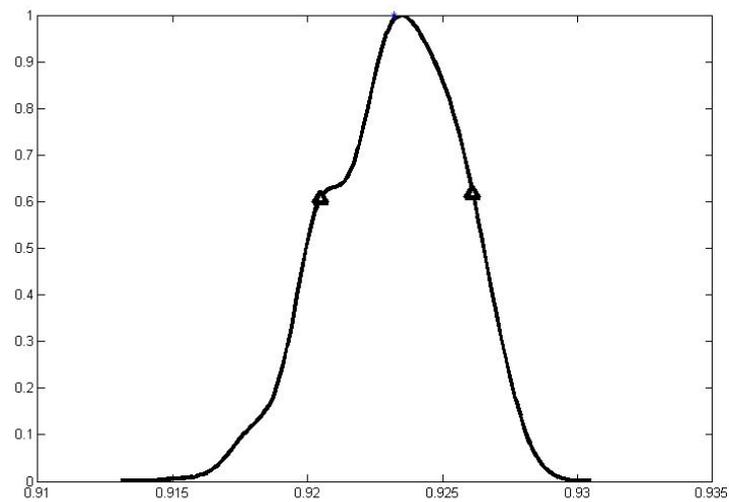
Fig. 4.16 The density function of the histogram represented in Figure 4.15 with range in [0 1]. The stars are the points with second coordinates equal to $e^{\frac{-1}{2}}$.
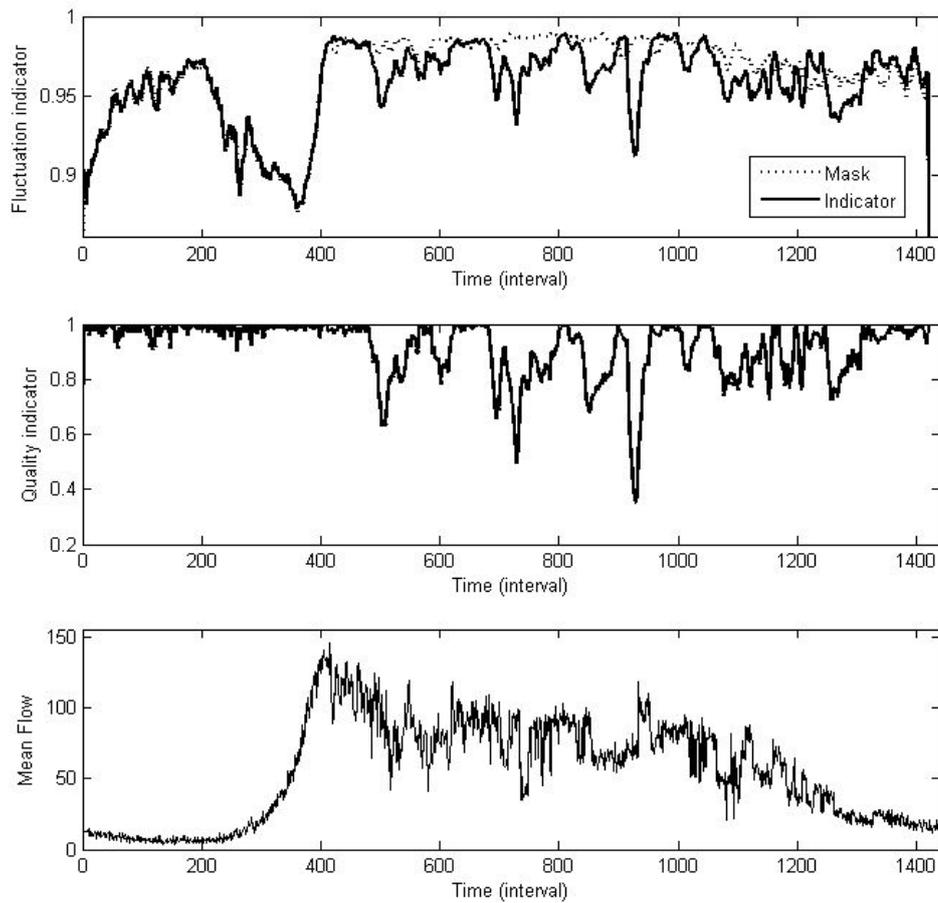
Fig. 4.17 The probability indicator, I_B, of a section with erroneous measurements. In some time interval of this section the fluctuation indicator is under the mask, due to a lower quality. The quality indicator is also representing low quality in those time intervals.
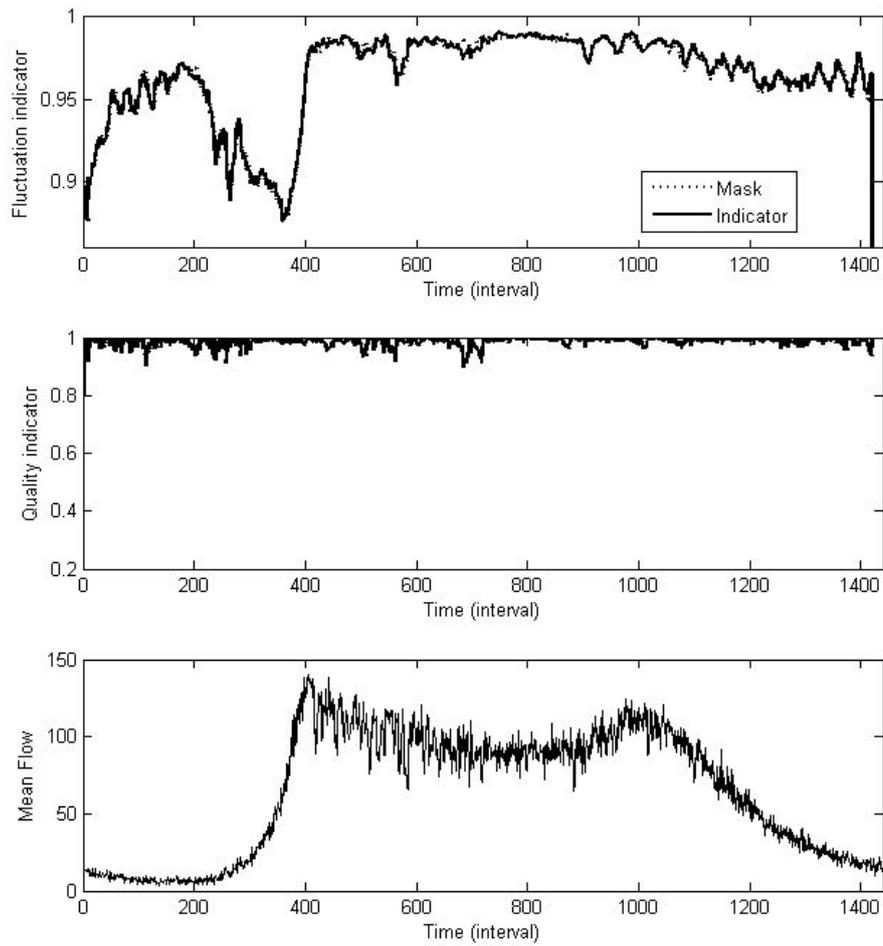
Fig. 4.18 The probability indicator, I_B, for a normal section. The fluctuation value of the flow for normal section is closer to the mask, since the quality is high.
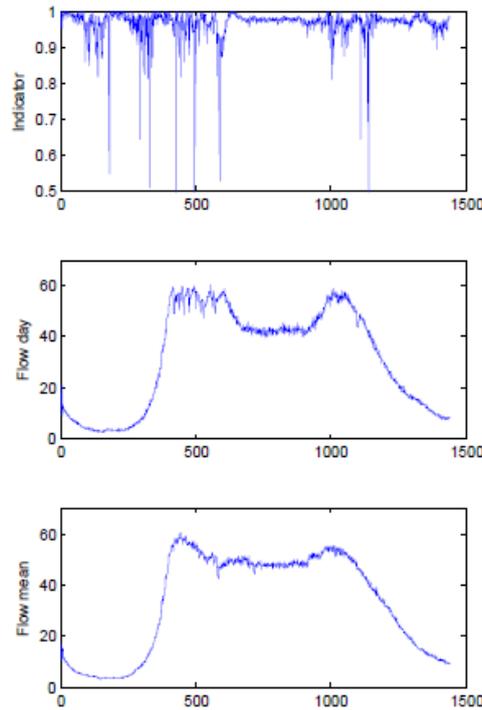
Fig. 4.19 Window consistency indicator for a critical day. The indicator decreases when flow of the day is substantially different from the average flow of the window.

4.4.2   Corridor

In Figure 4.20, an example of correlation function between two consecutive sections is presented. The vector of the data of the first section consists of one hour data, 10:00 am-11:00 am, and the time delay between two sections is assumed to be between $[-10, 10]$ minutes. The correlation is computed either within the same day or between the week averages.

As described in the previous sections, the input to the Fuzzy Systems are:$|C_{\mathrm{Max}}|$, A_{Max} and $cos(\alpha_{\mathrm{Max}})$ The values of the inputs can be between $[0, 1]$, $[0, 20]$, and $[-1, 0]$ respectively, since the angle of the peak of the correlation cannot be less than 90 degrees (and of course not more than 180 degrees). The inputs of the fuzzy system B are $|C_{\mathrm{Max},d}| - |C_{\mathrm{Max},m}|$, $A_{\mathrm{Max},d} - A_{\mathrm{Max},m}$ and $cos(\alpha_{\mathrm{Max},d}) - cos(\alpha_{\mathrm{Max},m})$. The ranges in this case are $[-1, 1]$, $[-20, 20]$, and $[-2, 2]$. The fuzzy system C uses the outputs of the other two systems as described in section 4.3.2.

Figure 4.21 shows the three indicators extracted from the three fuzzy systems. The figure on the top represents the indicator of the day (I_D) and the indicator of the mean of one week data (I_M), they are outputs of fuzzy system A, see Figures 4.8, 4.9 and Table 4.1. It can be seen that in some parts the quality of the data of the day in greater than the quality of the
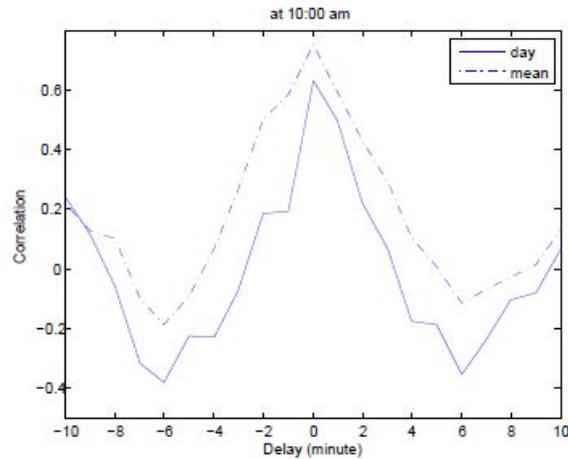
Fig. 4.20 The correlation of two sections in a row for one day data and the average data of a week.

average data of the week, and in some parts it is less. It can also be seen that in some parts they are closer to each other compare to the other parts. Based on these observations fuzzy system B is defined. The second figure in the middle is the indicator of comparison (I_E) which determines if the quality of the day is greater or less than the quality of the average data of the week. It is based on Fuzzy system B, see Figures 4.8, 4.10 and Table 4.1. According to this Figure we can see that when the value of I_D is greater (or less) than the value of I_M, then the value of I_E is greater (or lower) than 0.5 which means that when the quality of the data of the day under consideration is greater (or less) than the quality of the average data of the week, then the quality indicator I_E has greater (or less) value than 0.5. The final indicator (I_G) is defined based on the Fuzzy system C, see Figures 4.8, 4.11 and Table 4.2. This Fuzzy system is defined to make sure that all the indicators are used. In other words if I_E is showing high quality which means that teh quality of the data of teh day is higher than the average data of the week, this is not enought to say that the quality of the data of this day is good, because may be both I_M and I_D are representing bad quality. Therefore, if I_M, I_D and I_E are representing high quality then I_G also represenst high quality (for I_G value greater than 0.5 mean high quality).
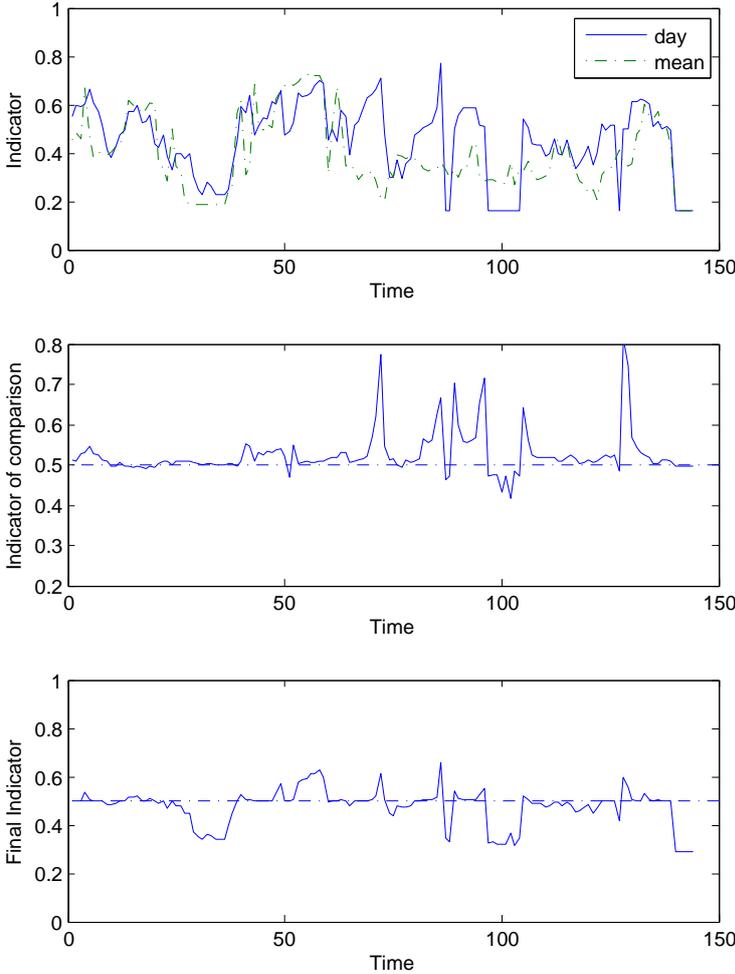
Fig. 4.21 The outputs of the fuzzy systems A, B and C are plotted.

Chapter 5

A framework for real-time emissions trading in large scale vehicle fleets

Abstract

In this chapter a framework for the real-time trading of budgeted emission rights between a fleet of participating vehicles is presented. The trading problem is formulated as a utility maximisation and as a utility fairness problem. The problem is solved in real time in centralised, decentralized, distributed and hybrid manners. It is a joint work with Florian Häusler, Emanuele Crisostomi, Arieh Schlote, Ilja Radusch, and Robert Shorten and was published in [47]. This work won the best scientific paper award of the ITS conference 2013 in Dublin [48]. The content of this chapter follows closely [67] but goes beyond this work by developing algorithms that have faster convergence rates. Our (Faizrahnemoon and Shorten) main contributions in this chapter have been the design of algorithms with higher speed of convergence in section 5.4, and running their corresponding simulations (Figures 5.2 and 5.3). Florian Häusler, (with Schlote and Shorten) formulated the constraints given in Section 5.2, applied ideas from Stanojevic to the emissions trading problem, and completed the simulations from Berlin.

5.1 Introduction

In the research that has been done on smart-cities [52, 55], collaborative mobility is becoming one of the main topics in relation to transportation networks. Since the road networks in cities are often inefficient and new technology continues to emerge, it is expected that the provision of products related to smart mobility will increase over the next decade [30, 128]

Collaborative mobility is not a new topic in transportation research. For example, the notion of V2X is related to the transport telematic [13]. Researchers have been working on transport telematics but the impact of the research has been limited because the technology is not sufficiently advanced. Apart from technological limitations, the strict guidelines from regulatory bodies can also prevent collaborative mobility growing as an area of research. In general there are four main factors in transport related regulations: vehicular safety;

greenhouse gas emissions; and the quality of air in our cities. They are all important considerations. Congestion is very unpleasant and it is a major problem related to economic growth [26]. Car-related safety is an issue for people in cars as well as other road users [97].Transportation accounts for 20% of green house gas emissions, whose control is currently an important issue for the European Union. However, perhaps the most important reason to consider this issue is that the internal combustion engine (ICE) is harmful for human health [42]. By-products of the ICE include: CO; NOx; SO; ozone; benzene; PM10; and PM25

Some governments have been trying to address the problem of air-quality control. In some cities certain vehicles are banned from densely populated areas. Another method which is used to control the air-quality to define speed limits according to the pollution peaks [2].Shifting mobility away from dirty modes of transportation and encouraging the adaptation of bike sharing systems are two strategies suggested to tackle the traffic-related air pollution problem [14, 23]. Another related method is via multimodality to simultaneously increase the efficiencies of different transport modes [5].We already discussed the bike sharing systems and multimodal transportation networks in Chapters 2 and 3. An alternative focus is on developing automated driving systems that can take the sometimes erratic and wasteful human behaviour out of the picture [40].

In [113] the aggregate effect of vehicles in a geographic region has been considered. Here we extend their approach and introduce a method to incorporate fairness into the problem. In our approach, individual utility functions are assigned to the vehicles and all the vehicles experience the same utility from the system. Moreover, we optimize a vehicle fleet's emission without considering their geographical positions. It is assumed that a fleet operator wants to control the aggregate effect of a group of vehicles (e.g. pollution) such that the total sum of the utility functions is maximized. Many problems in intelligent transoprtation systems (ITS) can be formulated in such a fashion. Companies like DHL, Schenker or UPS are good examples because they are usually asked to monitor and control their fleets in regard to their ecological impact. Such companies can provide information about the emissions of a single trip [25]. The problem is constrained in a number of ways. Customers may want the shipping to be done within a limited time. Sometimes there is speed limit along the route meaning that the total pollution must be less than a certain amount. Advances in V2X technologies, communications and in distributed optimisation, allow us to implement such a system in real time without requiring a significant amount of centralised processing, and also to respond to changing levels of vehicle density. The demand for such algorithms is likely to increase as equipment manufacturers and cities are asked to provide information about the environmental impact of products during their lifecycle [41] (and for municipalities, information on air quality). Furthemore, the customers of such companies may pay more for a higher quality

of service. This will allow vehicles to borrow emissions from each other to maintain an allowable overall level of emissions [43].In this chapter, we describe a number of algorithms that can be utilised for this purpose.

In summary, our main objective in this chapter is to describe a method that can regulate the aggregate effect (e.g. pollution) of a fleet of vehicles, so that the utility (benefit) to the vehicle owners is at all times maximised. In the particular application examples considered, we study a fleet of hybrid vehicles that has a total budget of CO_2 that they can emit. Thus, we are aiming to control the aggregate emission of CO_2 . Clearly, there are many possible ways to achieve this, and this non-uniqueness creates an opportunity to regulate and optimise simultaneously [67, 112, 120, 121]. The present setting offers the chance of not only regulating, say, aggregate CO_2 , but also of doing this in a way that allows another objective (e.g. maximisation of QoS, minimisation of CO or PM production) to be achieved at the same time.

We specifically: (i) formulate the emissions problem mathematically; (ii) discuss the communication con- straints of the problem; (iii) present algorithms to solve the utility maximization and utility fairness problems; and (iv) present some simulations. Our work follows closely [67] but goes beyond this work by developing algorithms that have much faster convergence rates. The rest of the chapter is structured as follows. In Section 5.2, we shall discuss the typical constraints that arise in cooperative mobility settings. These constraints guide the type of algorithms that can be deployed in such environments. In Section 5.3, we formally state our problem and in Section 5.4, we will discuss algorithms for solving utility maximisation problems in vehicular environments. Simulation results are presented in Section 5.5, where we also develop the case study given in [113] to include utility fairness constraints.

5.2   Background

We consider fleets of plug-in hybrid vehicles that can use two engines to achieve the goal of controlling the pollution, D_i. In this context, we study the following optimisation problem

$$max \quad \sum_{i=1}^{N} f_i(D_i) \qquad (5.1)$$

subject to the constraint that

$$\sum_{i=1}^{N} D_i = C \qquad (5.2)$$

Where the scalars D_i are mapped to some utility scalars by the functions $f_i(D_i)$. As we already said, many problems in ITS map to this scenario. In our context, C is a pollution budget assigned to a group of vehicles, D_i is the pollution budget assigned to vehicle i and $f_i(D_i)$ is the benefit of the ith vehicle using the allocation D_i. Different techniques are available to deal with such problems and many of them have already been applied in other application domains [119]. Under reasonable assumptions on the f_i, it is possible to deploy centralised optimization methods; however, problems that arise in the context of ITS present several specific challenges. We now summarise some of the basic issues considering the specific ITS context.

1. *Non-homogeneous levels of vehicle participation:* ITS problems typically involve varying levels of vehicle participation as vehicles leave and join the system defining the optimisation problem.

   Algorithms deployed to solve the maximisation problem must also respect constraints arising from the communication limitations of the ITS system.

2. *Scale:* ITS problems often involve a large-scale level of participation. Optimisation problems involving very-large graphs are considered among the most challenging in the field of operations research [44].

3. *V2I unicast:* In contrast to unicast I2 V, communicating in the other direction is technically straightforward to implement. A standard cellular communication network is required and the communication effort increases linearly with the number of vehicles.

4. *I2V unicast:* Direct (IP) communication from the centre (e.g. a fleet management back-office) to individual users over cellular networks cannot be implemented easily today. The main reason for this is network address translation (NAT) traversal problems. Current technologies for NAT traversal need extra infrastructure and communication protocols or control over network infrastructure.

5. *Infrastructure-to-vehicle (I2V) broadcast:* This scenario is probably the most common and well-established communication link in the present context. Vehicles listen to a broadcast from a central node. The technologies needed to implement I2V broadcasts are currently available in most vehicles and a feature of broadcasting is that the total communication effort scales well and independently of the number of vehicles in the fleet.

6. *Computation:* This is a fundamental issue. Is it possible for the central infrastructure to solve large-scale optimisation problems and communicate bidirectionally with vehicles at relatively high sampling rates.

7. *Asymmetric and symmetric Vehicle-to-vehicle (V2V):* V2V communication using IEEE 802.11p technology is currently in the process of standardisation but has not yet been deployed. V2V communication can be symmetric and asymmetric. In the symmetric case, vehicle i communicates with vehicle j if and only if j can communicate with i.

   Another issue that arises in ITS applications is the need to solve large-scale optimisation algorithms. For example, it may not always be possible or desirable to solve optimisation problems in a centralised manner.

   Finally, ITS applications give rise to several important performance constraints, including the following:

8. *Speed of convergence:* Algorithms should converge to the optimal solution in a reasonably short time.

9. *Robustness to failure and uncertainty:* Optimization algorithms should be robust with respect to failures in the system. Also, it should be noted that in many situations, utility functions are not known exactly.

10. *Scalability:* Deployed algorithms need to function satisfactorily irrespective of the level of vehicle participation.

    Given the above considerations, the question now arises as to the type of solution which is suitable for ITS applications. Roughly speaking, three choices are available to us:

1. *Fully centralised solutions:* involve solving the optimisation problem defined in (5.1) and (5.2) centrally. In order to solve the optimization problem by using such solution we need perfect knowledge of the $f_i(D_i)$, bi-directional communication between the infrastructure and all vehicles, and computational ability of the centralized node. Advantages of the centralised approach include speed of convergence (instantaneous) and that there is no V2V communication requirement. Disadvantages of this approach include sensitivity to node failure, sensitivity of solutions to noise and uncertainty, and the feasibility of centralised approaches in large-scale environments

2. *Fully decentralised solutions:* involve solving the optimisation problem defined in (5.1) and (5.2) using local computation and communication only.In this approach it is

required to have V2V communication. The advantages of this method are that there is no need for centralized optimization step, the ability to deal with uncertainty in the $f_i(D_i)$, and the robustness with respect to node failures. Disadvantages include the very-slow convergence rates – especially in large-scale applications

3. *Hybrid solutions:* involve using a small amount of information from the center in order to increase the rate of convergence of the decentralised algorithm and to mitigate some of the constraints of the ITS environment. In the next section, we will present two hybrid algorithms.

5.3   Problem statement and assumptions

We have a network of N nodes (vehicles). Individual utility functions $f_i(D_i)$ are assigned to the vehicles. There is a constraint $\sum_{i=1}^{N} = C$ and the goal is to find an allocation D_i that solves $max \sum_{i=1}^{N} f_i(D_i)$ subject to $\sum_{i=1}^{N} D_i = C$.

Comment: In some applications, we may be interested in solving a minimisation problem.

Assumptions

The following assumptions appear to be a reasonable reflection of current and near future state of the art in ITS systems, and suggest a hybrid solution whereby the optimisation is solved using decentralised computation with aid of limited signaling from the centre.

In order to solve the problem we assume that

- The utility functions $f_i(D_i)$ are concave because we are maximizing the objective function. If the optimization problem is minimizing the sum of the utuility functions (e.g. minimizing the pollution), then the assumption is that the utility functions are convex. The exact nature of the utility functions may not be known but the vehicles are aware of the value of $f_i(D_i)$ at each time step. In some cases, for example, the utility fairness case, the utility functions are increasing (decreasing) when the problem is maximizing (minimizing) the sum of the utility functions.

- There exists at least one feasible solution which is also optimal.

- All vehicles can communicate directly (or indirectly through a measurement) with a centralised infrastructure.

- The centralised infrastructure broadcast information to the vehicles but does not communicate directly with individual vehicles.

5.3.1 Solution

We can solve the optimization problem given by equations (5.1) and (5.2) using the classical technique of Lagrange multipliers. Let $H(D_1, \ldots, D_n, l)$ be the Lagrangian associated with the problem, that is

$$H(D_1, ..., D_n, \lambda) = \sum_{i=1}^{N} f_i(D_i) - \lambda \left(\sum_{i=1}^{N} D_i - C \right). \tag{5.3}$$

The optimal solution of the problem is the allocation D_i^* such that $\lambda^* = \frac{\partial f_i(D_i^*)}{\partial D_i}, \forall i = 1, ..., N$, subject to the linear constraint being satisfied. Two hybrid approaches can be used to solve the maximization problem. First, the maximisation problem can be solved using a combination of feedback and local on-vehicle computational power. Namely, by regulating the outputs of the vehicle functions in such a way that they follow the λ, which is broadcast from a centralised infrastructure. Therefore, the task of the central infrastructure is to find the value of multiplier λ so that the utility maximization problem can be solved. We will see that this can be solved by embedding the multiplier as part of a feedback loop as is done in subgradient methods. This approach does not need any V2V or I2V unicast capabilities or any centralized computation. It is even able to deal with uncertainty in the utility functions. A disadvantage of the approach is that it is not robust to failure of the centralised node and selecting the control gains can be an issue. The second approach is based on the fact that optimality is also achieved when each of the vehicles achieves consensus of the values of the functions $g_i(D_i) = \frac{\partial f_i(D_i)}{\partial D_i}$. This observation was first exploited in [119] in the context of symmetric communication graphs, and later extended to the asymmetric case in [65], and to the case where the outputs of the vehicle are adjusted as part of a feedback loop. Advantages of these approaches are that no I2V unicast capabilities are required, and the techniques can deal with uncertainty in the utility functions. In addition no centralised computation is required, and the technique is robust with respect to node failure.

5.4 Algorithms

5.4.1 Algorithm 1. Feedback control and local computation

Algorithm 1 uses an integral control to asymptotically find the Lagrange multiplier λ^*. The centralised infrastructure then updates iteratively according to

$$\bar{\lambda}(k+1) = \bar{\lambda}(k) - \mu \left(C - \sum_{i=1}^{N} D_i(k) \right) \tag{5.4}$$

where we have assumed that the centralised infrastructure either receives communication of the $D_i(k)$ from each vehicle, or measures it. Each node (vehicle) receives a broadcast of $\bar{\lambda}(k)$ and updates its utility (implicitly) according to

$$D_i(k+1) = D_i(k) - \beta_i \left(\bar{\lambda}(k) - \frac{\partial f_i(D_i(k))}{\partial D_i(k)} \right). \tag{5.5}$$

Roughly speaking, this corresponds to a classic subgradient method from optimisation theory, see [8, 92] and the gains can be selected using methods from nonlinear control theory. The convergence of the algorithm, and the choice of the model parameters is beyond the scope of this thesis. The convergence argument and choice of gains follows from standard systems theory ideas [63, 118, 119]. Note that Algorithm 1 is a variant on a standard optimisation algorithm, the dual subgradient descent algorithm, where we replaced the primal optimisation step with a second ascend type update.

5.4.2   Algorithm 2. Implicit consensus with input

We have previously noted that it was shown in [119] that utility maximization problems may be formulated as consensus problems where all agents achieve the same value of the derivative $g_i(D_i) = \frac{\partial f_i(D_i)}{\partial D_i}$ of their utility functions. This solution is interesting because it can be realised in a fully decentralised manner, but consensus based solutions usually require symmetric communication graphs, to guarantee that $\sum_{i=1}^{N} D_i(k) = C$ for all k, and might suffer from other convergence related issues. Some of these issues are addressed in [65], in which a basic feedback error signal, $C - \sum_{i=1}^{N} D_i(k)$, from the infrastructure is used to yield a modified integral control of the following form

$$D_i(k+1) = D_i(k) + \varepsilon \sum_{j \in \mathcal{N}_i} \left(g_i(D_i(k)) - g_j(D_j(k)) \right) + \gamma \left(C - \sum_{q=1}^{N} D_q(k) \right) \tag{5.6}$$

in which $\mathcal{N}_i$ is the set of nodes that send information to node i. Note that if node i has $\mathcal{N}_i$ neighbors that can send information to node i, then (5.6) can be rewritten as

$$D_i(k+1) = D_i(k) + \varepsilon \left(N_i \cdot g_i(D_i(k)) - \sum_{j \in \mathcal{N}_i} g_j(D_j(k)) \right) + \gamma \left(C - \sum_{q=1}^{N} D_q(k) \right) \tag{5.7}$$

Thus, by denoting $\varepsilon_i = \varepsilon \cdot N_i$, the problem can be stated in a consensus-like form, as

$$D_i(k+1) = D_i(k) + \varepsilon_i \left(g_i(D_i(k)) - \frac{1}{N_i} \sum_{j \in \mathcal{N}_i} g_j(D_j(k)) \right) + \gamma \left(C - \sum_{q=1}^{N} D_q(k) \right) \quad (5.8)$$

where j ranges over all neighbours of i. Roughly speaking, the first term of the update (5.8) is used to achieve a consensus around the optimal value, while the second term ensures that the global equality constraint is satisfied.

The stability and convergence properties of Algorithm 2, and the choice of γ and ε can again be determined using ideas from systems theory (see [66]). However, a proper discussion of this issue is again beyond the scope of this thesis. It is important to note that Algorithm 2 requires V2V communication, and has slow rate of convergence [67].

Comment: It should be noticed that the sign of the parameter ε in (5.6) - (5.8) depends on the choice of the functions g. The sign will be positive if functions g are decreasing, and negative if functions g are increasing. Note that different policies can be implemented by replacing functions $g_i(D_i) = \frac{\partial f_i(D_i)}{\partial D_i}$ in (5.8) with other appropriate functions. For example, if we use $g_i(D_i) = f_i(D_i)$ instead of the partial derivatives, a utility fairness policy will be implemented. A discussion on the properties of the candidate functions g_i is beyond the purposes of this thesis, but conditions on the functions g_i and a related discussion can be found in [65, 119].

5.4.3   Some details of Algorithm 3

Modified algorithm: My main contribution in this chapter is to modify Algorithm 2 to speed up convergence. In the modified algorithm, which is denoted as Algorithm 3, the vehicles communicate the maximum and minimum values of $\frac{\partial f_i(D_i(k))}{\partial D_i(k)}$ to the central infrastructure at each time step. These values are then broadcast to all nodes who then augment their neighborhood information to encorporate the new values. In the simulations it will be shown that the rate of convergence of the modified algorithm increases compared to Algorithm 2. In the next paragraph we give a plausibility argument of why this occurs.

We will see in the simulations that when the center broadcasts the maximum and the minimum values of the second term in (5.8) to the nodes, convergence to implicit consensus occurs at a much faster rate.. It is well known that in undirected graphs, the second smallest eigenvalue of the graph Laplacian, also called *algebraic connectivity*, quantifies the speed of convergence of consensus algorithms, see for instance [94]. Also, it is well known that in

this case the algebraic connectivity of a new graph G_2, $a(G_2)$, obtained from adding a new edge to graph G_1 is characterised by

$$a(G_1) \leq a(G_2) \leq a(G_1) + 2 \tag{5.9}$$

where v and w are two nodes of the graph G, can be used to state that increasing the out-degree d_0 of some nodes of the graph (as in practice occurs when the algorithm is modified to circulate the minimum and the maximum values) does also increase (or at least, does not decrease) an upper bound of the algebraic connectivity of the graph.

Finally, we also mention that similar strategies to improve the convergence of network consensus algorithms in directed graphs were also devised in [11], where the authors claim that having as many as possible vertices with the maximum out-degree of n_1, and having the in-degree of each vertex around m/n, where m is the number of edges and n that of nodes, can improve the convergence speed.

5.5   Simulations

5.5.1   Algorithm 1: Utility minimisation

In our first simulation we assume that there is a total budget of CO_2 and we have a fleet of 20 hybrid vehicles that are not allowed to emit more than the budget. The objective here is to minimize the total quantity of emitted *PM*. Furthermore, as in [65], we assume that vehicles can regulate the quantity of emissions by deciding which fraction of the desired speed is supplied by the *ICE* or by the electric motor. This is consistent with a power split hybrid vehicle. In our simulations, we compute emissions using a simple 'average speed model', and data taken from [6] set in an urban setting where vehicle speeds vary between 10 and 50 Km/h. Then, according to the data from [6] (vehicle code $R019$, corresponding to petrol *LDV*s with an *ICE* equivalent of class *EURO*4, data can be found on page 164 and 106 for CO_2 and *PM*, respectively), one can approximate the relationship between *PM* and CO_2 as

$$PM(CO_2) = 50CO_2^2 - 19CO_2 + 6.4 \tag{5.10}$$

where *PM* is expressed in mg/km and CO_2 in kg/km, which is a good approximation for speeds included between 10 and 50 km/h. This leads to the optimisation problem *min* $\sum_{i=1}^{N} f_i(CO_2^{(i)})$ subject to $\sum_{i=1}^{N} CO_2^{(i)} = C$, where C is the overall CO_2 budget allocated to the fleet of vehicles, and $CO_2^{(i)}$ is the amount of CO_2 allocated to the ith vehicle. Note that this problem is equivalent to the one described by equations (5.1) and (5.2) which has already

been discussed in the previous sections. We assume that the vehicles have slightly different parameters, due to characteristics specific of the particular vehicle (e.g. brand, age), and we use the parameter values $\gamma = 0.1$ and $\varepsilon = 0.01$. Figure 5.1 shows how the CO_2 budget is shared among the fleet of vehicles in order to minimise PM emissions, using Algorithm 1.
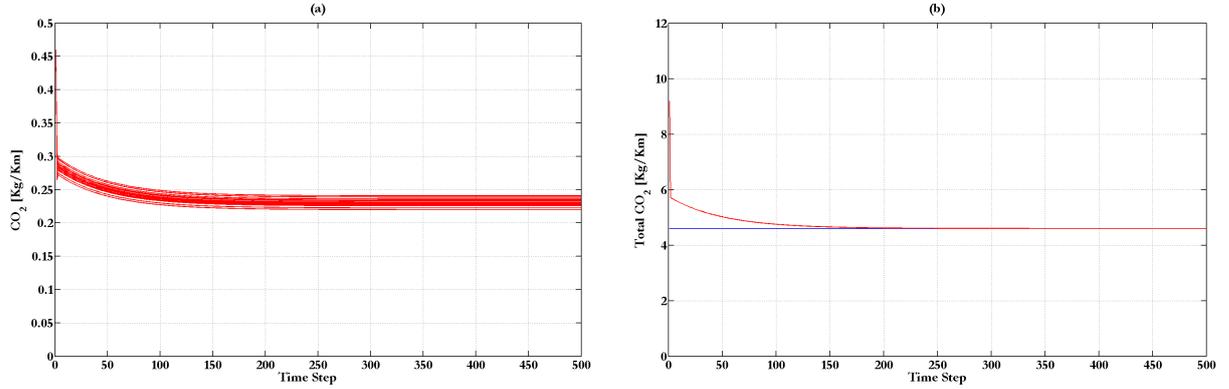
Fig. 5.1 Here, Figure 5.1.a. shows how the CO_2 budget is shared among the fleet of vehicles in order to minimise PM emissions. At the same time, the CO_2 budget is asymptotically respected as shown in Figure 5.1.b.

5.5.2   Utility fairness

In the first simulation the goal was to minimize the production of PM while the total amount of CO_2 was the same as the budget. However, in some circumstances such a solution is unfair and it is desired to have equal amount of pollution for all the vehicles. In the next simulation the idea is to make the vehicles produce the same amount of PM while the total CO_2 that they emit should be equal to the budget. Such a solution is obviously not an optimal solution for the environment because more PM can be produced compare to the previous solution but it is a fair solution for the vehicles.

We now compare the three algorithms discussed in Section 5.4. We assume that the utility function of node i is given by a function of the form $f_i = \alpha_i log D_i$, which is increasing and concave. The idea of choosing the logarithm as the utility function is motivated by dispatch type problems where economic utility is an increasing function of allocated emissions. In the first iteration it is assumed that $g_i(D_i(k)) = f_i(D_i(k))$.

In figure 5.2 the performances of Algorithm 1 (Figure 5.2.b.), Algorithm 2 (Figure 5.2.c.) and Algorithm 3 (Figure 5.2.d.) are compared. Our simulation example has 41 vehicles, and the communication graph is shown in Figure 5.2.a. The parameters ε and γ of (5.6) in the distributed algorithms are chosen as 0.002 and 0.0024 respectively. In the case of the centralised algorithm, μ in (5.4) is chosen as 0.01, while β_i of (5.5) is chosen
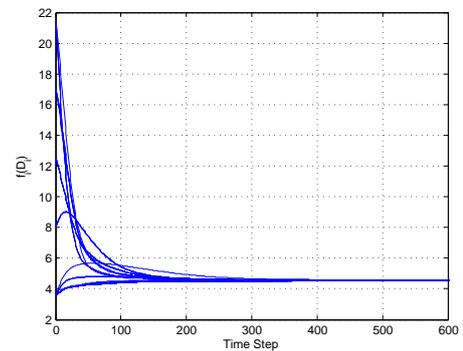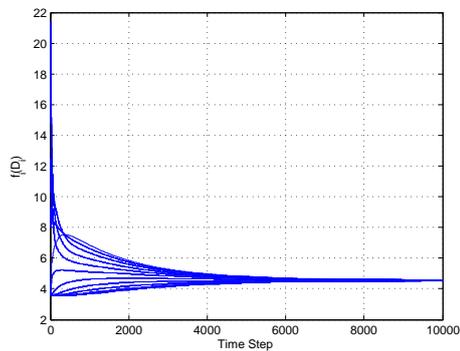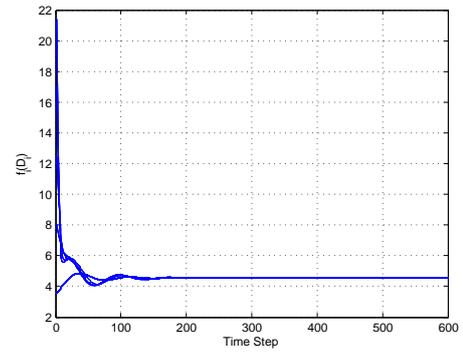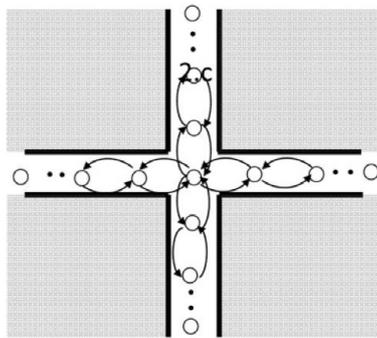
Fig. 5.2 CO_2 budget is shared in such a way to equalise the production of *PM*. In *b*, the infrastructure communicates the value of $\lambda(k)$ as explained in Algorithm 1, and the vehicles do not communicate among themselves; in *c* the infrastructure only communicates the produced CO_2, and vehicles can communicate only with the neighbouring vehicles, in *d* vehicles further receive some more information from the infrastructure as explained in Algorithm 3. Note that this last opportunity algorithm improves the time of convergence of the hybrid algorithm with respect to *c*.
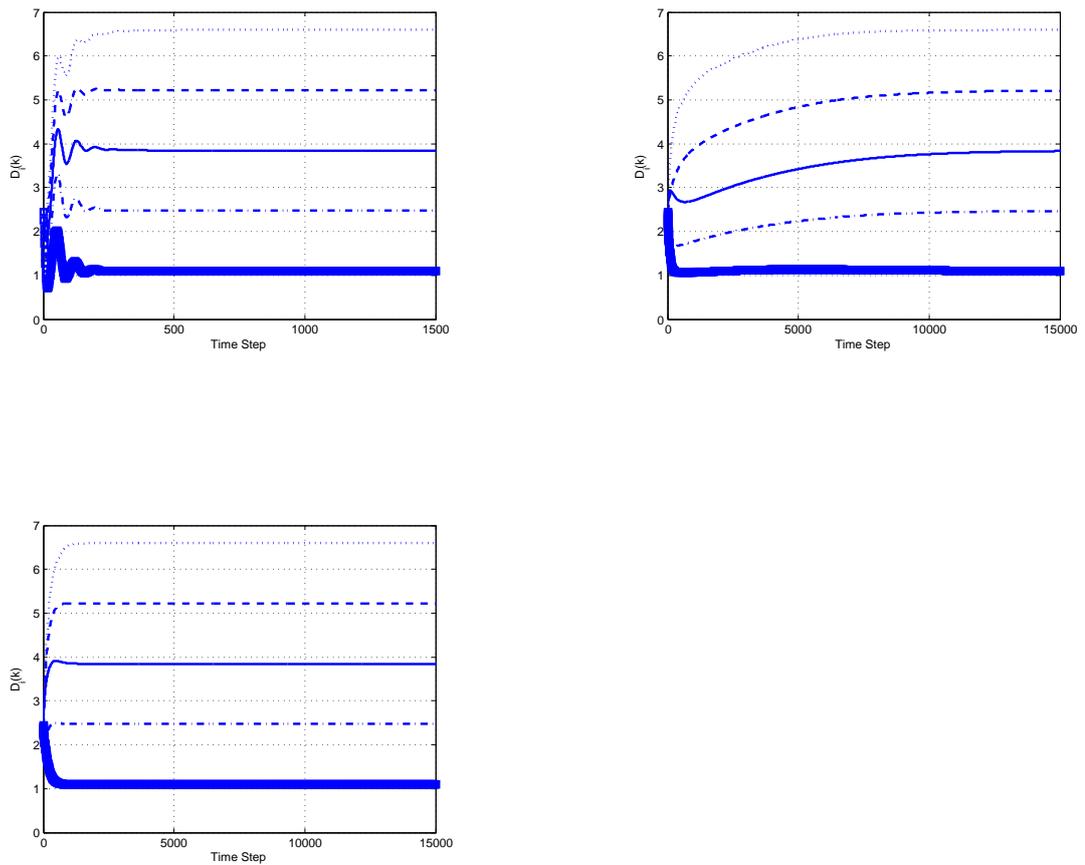
Fig. 5.3 Time evolution of the pollution of some of the vehicles is depicted in the case we are interested in equalising the derivatives of the utility functions

equal to 0.02. We further assume that the budget C is initially equally divided between the vehicles. In the shown example, we assume that vehicles are able to communicate along the roads, while buildings do not allow vehicles travelling along different roads to communicate with each other. Note the increased convergence speed in Figure 5.2.d over that in Figure 5.2.c. However, making all vehicles produce the same amount of pollution might not be itself a fair choice in some circumstances. For instance, it is reasonable to allocate a larger pollution budget to vehicles that have to travel longer distances. In this case, we take a different fairness approach, and choose $g_i(D_i) = \frac{\partial f_i(D_i)}{\partial D_i}$, as discussed in Section 3.2 and 4.2. Accordingly, Figure 5.3 compares again Algorithms 1, 2 and 3 in Figures 5.3.a., 5.3.b. and 5.3.c., respectively. The multipliers ε and γ of (5.4), (5.5) and (5.6) are the same of Figure 5.2.

5.5.3   Utility fairness: a case study

We now present a more realistic application of the utility fairness paradigm. The case study here is an extension of a previous work within the so-called TwinLIN initiative [113]. In order to control the CO pollution in urban areas, the TwinLIN work implemented a central control approach. The idea was to maintain the control variable below, or close to, a desired level and the approach was forcing hybrid vehicles to travel either in electric or ICE mode. The objective was achieved by assuming that the infrastructure could broadcast a probability p_{ev} to all vehicles. This probability was then used locally by the vehicles, and possibly modified to achieve other objectives at the same time, to decide whether they should travel in electric or ICE mode.

By extending the Twinlin framework we now implement the utiltiy fairness strategy and we use SUMO with its HBEFA-emission model and TRACI interface [68], to test our idea. Figure 5.4 shows an arbitrary area of Berlin that we simulate by SUMO. We divided this area into nine equivalent rectangular zones of the same size (in a 3×3 grid).

Then we apply a simplified version of Algorithm 2 to all these sub-areas. We assume that in each sub-area there is a central infrastructure that broadcasts the feedback error signal $C - \sum_{i=1}^{N} D_i(k)$ in order to keep the level of CO_2 equal to the budget. At the same time, each vehicle in the sub-area uses the error signal to further adjust the proportion of speed generated from the ICE in order to minimise the overall quantity of emitted CO, according to the algorithm the (5.8). SUMO simulation results are given in Figure 5.5, which shows the level of CO_2 in each sub-zone of the Berlin area of interest, and in Figure 5.6 which shows the normalised CO produced by vehicles belonging to 5 different classes of vehicles (with ICEs equivalent to EURO 1 to 5 vehicles). We approximately have the same level of CO_2 in each sub-zone of Berlin and each vehicle produces the same amount of CO.

Fig. 5.4 Road network imported from OpenStreetMap [79], area around Rosenthaler Platz in Berlin, Germany
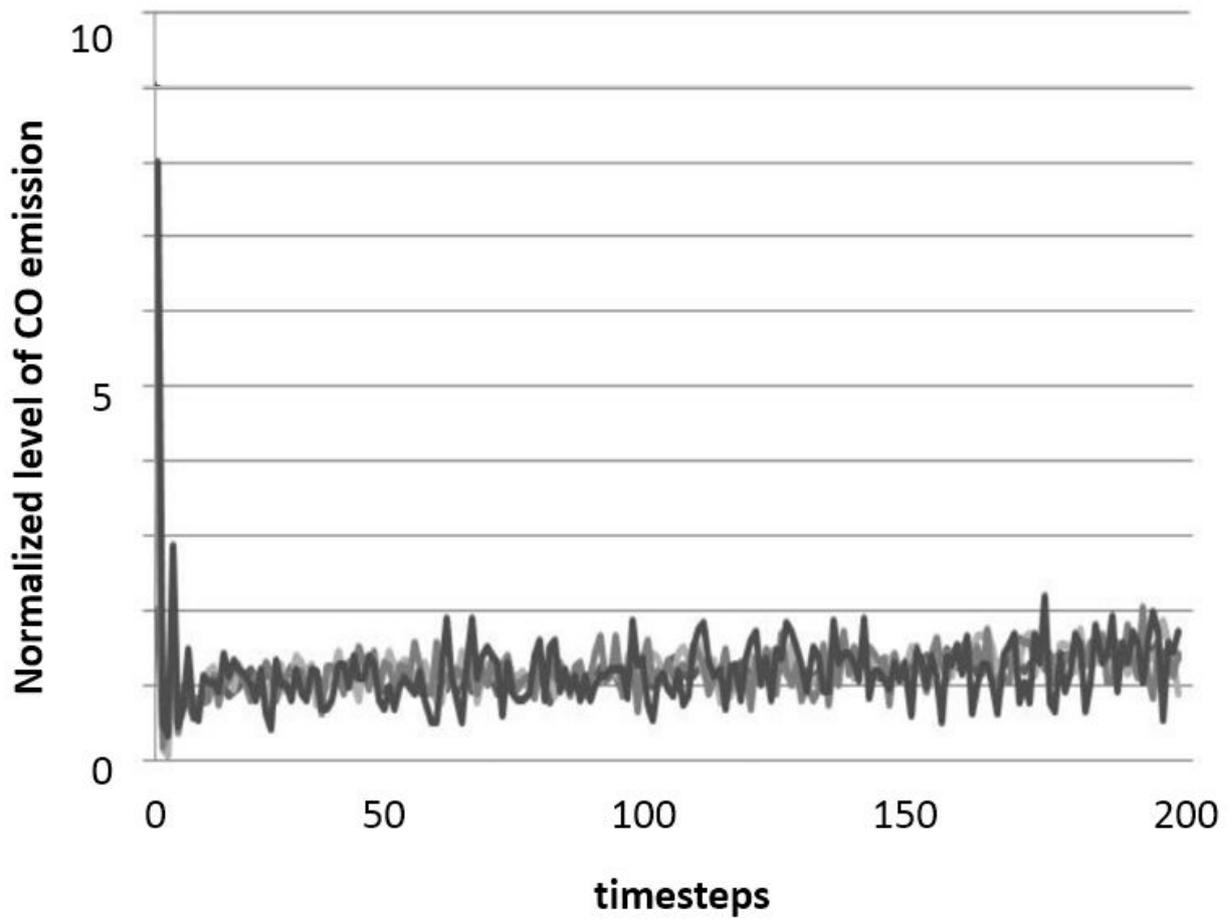
Fig. 5.5 Vehicles belonging to different classes produce the same quantity of *CO*.
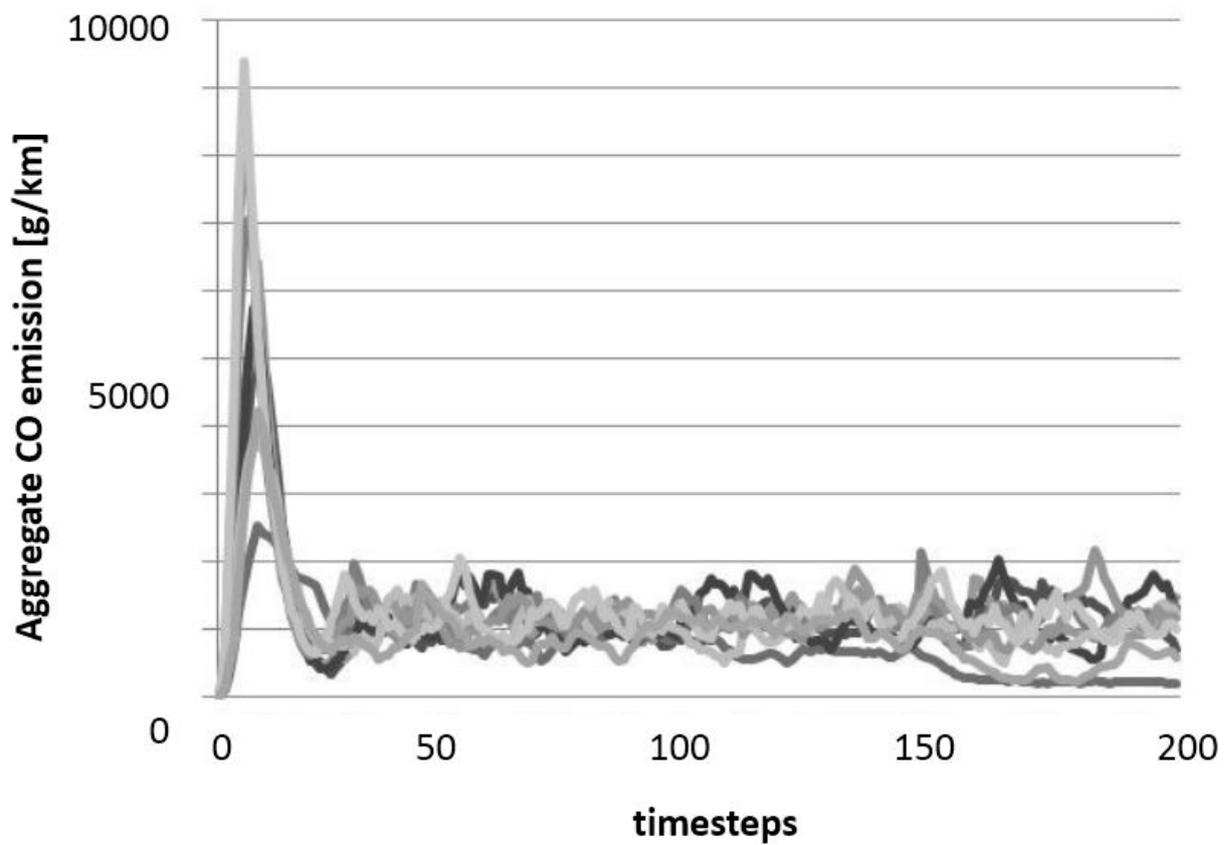
Fig. 5.6 Level of CO_2 is maintained equal to the CO_2 budget in each sub-zone of the Berlin area of interest.

Chapter 6

Conclusion and future work

6.1 Conclusion

We began such thesis by describing the basic idea of Markov chains and defining some quantities as the mean first passage time and the Kemeny constant. These quantities are helpful to extract useful information about the network.

Then a theorem from [19] that reveals the clustering properties of the eigenvector that is associated with the eigenvalue of second largest modulus of the transition matrix of a Markov chain is provided. In the theorem in [19]it is necssary to assume that the eigenvalue is real. Therefore, we extended the theorem to the case that the eigenvalue is complex, and proved that the eigenvector corresponding to the complex eigenvalue can reveal clusters of the network. To be more specific, we provided a theorem which says that when the transition matrix has an eigenvalue whose real part and imaginary part are close to one and zero respectively, the eigenvector corresponding to that eigenvalue reveals clusters of the network.

Some applications of Markov chains in transportation networks were described. For example we said how Markov chains have been used to model road networks, how they have been used in emission applications, and how suitable and useful they have been in big- data applications. In this thesis we used Markov chains to model different transportation networks (bike sharing systems, bus networks, and multi-modal transportation systems). We started by modeling the bike sharing systems and explained how to build the transition matrix of the Markov chain. Then we validated our model by using some real data of the bike network in Boston. We represented the clusters of the bike network identified by the eigenvector corresponding the eigenvalue of the second largest modulus.

We also used Markov chains to model a bus network. We designed two graphs (waiting graph and travel graph) for the bus network and we modelled the bus network in two different

ways based on these graphs. In order to validate the Markov chain models, we simulated a small part of Dublin city centre to extract data (we used SUMO as the tool). We validated the Markov chain models by the data extracted by SUMO and extracted some quantities of the Markov chains. These quantities provide useful information. For example, bus stops with high density of people waiting for a bus, and the clusters of the bus network were identified. In the Markov chain, we faced the situation in which the eigenvalue of the second largest modulus was complex and we showed that the corresponding eigenvector was able to identify the clusters of the network (the theory was mathematically proved before).

In order to extend our approach, we used the idea of Markov chains to model a multi-modal transportation network, not just bike or just bus network, We used the same techniques (as we used to model the bus network) to model the multi-modal network. We successfully validated the models by using some real data from London. We introduced some Markov chain based control applications to improve the network. For example, having access to a critical area (like a hospital) should be reasonable such that the travel time should be proportional to the distance. Balanced control of the waiting time at the stations and travel time, and some clustering applications are other applications that we considered.

Since we used real data from Boston and London to validate our bike network and our multi-modal network models, we need to know how trustworthy the data is. For this purpose, we defined several indicators (for a single time step, a time step interval, single section of the road, a corridor of the road etc) to determine the quality of data. The indicators are defined such that they give us a value between zero and one. If the value is close to one, it means that the quality of the data is high and if the value is close to zero, it means that the quality is low. We tested these indicators on some data from London highways and we represented the quality of the flow.

At the end we presented a framework for real time trading of emission rights between a fleet of vehicles. Two solutions were discussed. The first one assumes, that a central instance (e.g. the back-office of the fleet operator) broadcasts a simple information to listening vehicles. The vehicles use this value as an input target value, which guides their behavior. The second solution assumes additional V2V communication. It exploits the idea of (fully decentralised) consensus of utility derivatives - thus implementing an optimal utility for the fleet - with additional information given from the infrastructure to meet the global constraint. We extended this idea with one more message exchanged between vehicles and infrastructure, A framework for optimal real-time emissions trading in large scale vehicle fleets which incorporates the maximum and minimum values of the utility derivatives to speed up the convergence.

6.2   Future work

Future work will involve extending the applications of Markov chains, mathematical proofs, and the quality indicators.

We believe our result about clustering properties of the eigenvector corresponding to a complex eigenvalue can be further investigated and developed in the future. For example we can identify the nodes in different clusters, but identifying all the nodes of each cluster and also identifying all the clusters in the network need more investigation. Moreover, our complex eigenvalue is assumed to be a specific one which is very close to the $1 + 0i$. We need to know how close the eigenvalue should be to $1 + 0i$. In other words if the imaginary part of the eigenvalue becomes bigger and bigger, does the corresponding eigenvector still reveal clustering properties?

The bike sharing systems application can be extended by relocating the bikes during the day (and night) to avoid unpleasant situations. For example if a user goes to a station to get a bike, the station should not be empty or if a user goes to a station to park a bike the station should not be full. The solution can be an optimized route for the truck that relocates the bikes.

In our applications we identified the nodes with high density of people waiting for a bus or tube. One way to improve the network in the future can be increasing the frequency of the buses or the tubes for those stops. Optimal frequency and optimal route for the tubes and buses must be found such that the density of people decreases in those stops. Another way to improve the network can be integrating taxis in the bus network or in the multi-modal public transport network. Taxis can help the public transportation network by being available in the areas that people must wait long at the stations. One of the applications that can be enhanced is the accessibility of a critical area. In this thesis we added a few connections to the network to have reasonable accessibility to the critical area. This solution can be further investigated to find a new network which is optimized in the sense of accessibility to a specific area.

Clustering properties of the eigenvector corresponding the eigenvalue of the second largest modulus can create applications for the future work. When a graph has clusters, it is nearly decoupled. One can consider a converse problem. Sometimes a graph may be too connected and problems in one part may spread too quickly to the other parts of the network. Therefore, the challenge is designing networks where the spread of information is controlled.

The quality indicators that we defined in Chapter 4 are able to identify that the quality of data is low which means that there is an error or a problem in the data. However, they are not able to identify the reason (which can be a problem with the sensor , or an accident in the road, or a real phenomena like storm or rain). The suggestion for future work is to define indicators that are able to identify the reason of low quality data.

Future work will also involve quantifying the rate of convergence of both methods in Chapter 5 and mathematically proving that the method that we proposed in which we have extra information from the centre in addition to the V2V communication is faster convergent.

References

[1] Bekkerman, E., Kataev, S., and Kataeva, S. (2013). Heuristic approximation method for a random flow of events by an mc-flow with arbitrary number of states. *Automation and Remote Control*, 74(9):1449–1459.

[2] Bel, G. and Rosell, J. (2013). Effects of the 80 km/h and variable speed limits on air pollution in the metropolitan area of barcelona. *Transp. Res. D, Transp. Environ.*, 23:90–97.

[3] Bogenberger, K. and Weikl, S. (2012). Quality management methods for real-time traffic information. *Procedia-Social and Behavioral Sciences*, 54:936–945.

[4] Bos, B. K. (2014). Improving traffc mathematically: Expanding on the work of crisostomi et al. (2011). Master's thesis, Department of Mathematics, Utrecht University.

[5] Botea, A., Berlingerio, M., and Nikolova, E. (2013). Multi-modal journey planning in the presence of uncertainty. *Proc. of the Int. Conf. on Automated Planning and Scheduling*, pages 20–28.

[6] Boulter, P. G., Barlow, T. J., and McCrae, I. S. (2009). Emission factors 2009: Report 3 – exhaust emission factors for road vehicles in the united kingdom. *Project Report PPR356, TRL Limited*.

[7] Bowman, A. W. and Azzalini, A. (1997). Applied smoothing techniques for data analysis. *Oxford University Press*.

[8] Boyd, S. and Vandenberghe, L. (2004). Convex optimization. *Cambridge University Press*.

[9] Buehler, R. and Pucher, J. (2010). Cycling to sustainability in Amsterdam. *Sustain: A Journal of Environmental and Sustainability issues*, 21:36–40.

[10] C2C-CC (2007). Car to car communication consortium manifesto: Overview overview of the c2c-cc system. *Car to Car Communication Consortium, Tech. Rep.*

[11] Cao, M. and Wu, C. W. (2007). Topology design for fast convergence of network consensus algorithms. *IEEE Int. Symp. on Circuits and Systems*.

[12] Chandler, R., Herman, R., and Montroll, E. (1958). Traffic dynamics: Studies in car following. *Operations Research*, 6:165–184.

[13] Chatterjee, K., Mcdonald, M., Paulley, N., and Taylor, N. B. (1999). Modelling the impacts of transport telematics: current limitations and future developments. *Transp. Rev.*, 19(1):57–80.

[14] Chen, B., Pinelli, F., Sinn, M., Botea, A., and Calabrese, F. (2013). Uncertainty in urban mobility: Predicting waiting times for shared bicycles and parking lots. *Proc. of the 16th Int. IEEE Annual Conf. on Intelligent Transportation Systems, The Hague, Netherlands.*

[15] Chen, C., Kwon, J., Rice, J., Skabardonis, A., and Varaiya, P. (2003). Detecting errors and imputing missing data for single-loop surveillance systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1855(1):160–167.

[16] Chen, M., Liu, J., and Tang, X. (2008). Clustering via random walk hitting time on directed graphs. *Proceedings of the twenty third AAAI Conference on Artificial Intelligence*, pages 616–621.

[17] Cho, G. E. and Meyer, C. D. (2001). Comparison of perturbation bounds for the stationary distribution of a markov chain. *Linear Algebra and its Applications*, 335:137–150.

[18] Crisostomi, E., Kirkland, S., Schlote, A., and Shorten, R. (2011a). Markov chain based emissions models: a precursor for green control. *Green IT: Technologies and Applications*, pages 381–400.

[19] Crisostomi, E., Kirkland, S., and Shorten, R. (2011b). A google-like model of road network dynamics and its application to regulation and control. *International Journal of Control*, 84(3):633–651.

[20] Crisostomi, E., Kirkland, S., and Shorten, R. (2011c). Robust and risk-averse routing algorithms in road networks. *In proceedings of the 18th IFAC world congress, Milano, Italy*, pages 9512–9517.

[21] D2.2, C. (2006). Use cases and system requirements. *IST CVIS Project, CVIS IST-4-027293-IP deliverable D2.2 version 1.0,*.

[22] D4.1, P. (2008). Detailed description of selected use cases and corresponding technical requirements. *IST PreDrive C2X project, Tech. Rep. PreDrive C2X deliverable D4.1.*

[23] de Weerdt, M., Gerding, E., Stein, S., Robu, V., and Jennings, N. (2013). Intention-aware routing to minimise delays at electric vehicle charging stations. *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence*, pages 83–89.

[24] DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Transport Reviews: A Transnational Transdisciplinary Journal*, 12(4):41–56.

[25] DHL, D. (2013). Go green. *http://www.dhl.com/en/logistics/freight transportation/go green.html*

[26] Dirks, S., Gurdgiev, C. C., and Keeling, M. (2010). Smarter cities for smarter growth. *IBM Institute for business value, White Paper*.

[27] Doyle, P. (2009). The kemeny constant of a markov chain. *http://www.math.dartmouth.edu/ doyle/docs/kc/kc.pdf*

[28] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, pages 1–26.

[29] ETSITR102638 (2009). Etsitr102638, intelligent transport system (its); vehicular communications; basic set of applications; definition. *ETSI Std. ETSI ITS Specification TR 102 638 version 1.1.1, June 2009.*

[30] Ezell, S. (2010). Intelligent transportation systems. *http://www.itif.org/files/2010 1 27 ITS Leadership.pdf*

[31] Faizrahnemoon, M., Schlote, A., Crisostomi, E., and Shorten, R. (2013). A google-like model for public transport. *International Conference on Connected Vehicles and Expo (ICCVE).*

[32] Faizrahnemoon, M., Schlote, A., Maggi, L., Crisostomi, E., and Shorten, R. (2015). A big-data model for multi-modal public transportation with application to macroscopic control and optimisation. *International Journal of Control*, 88(11):2354–2368.

[33] Fishman, E. (2011). The impact of public bicycle share schemes on transport choice. In *Asia-Pacific Cycle Congress*, pages 18–21, Berlin, Germany.

[34] Fishman, E., Washington, S., and Haworth, N. (2013). Bike share: A synthesis of the literature. *Transport Reviews: A Transnational Transdisciplinary Journal*, 33(2):148–165.

[35] Franz, W., Hartenstein, H., and Mauve, M. (2005). Inter vehicle communications based on ad hoc networking principles the fleetnet project. *University of Karlsruhe.*

[36] Froyland, G. (2001). Extracting dynamical behavior via markov models. *Nonlinear dynamics and statistics*, pages 281–321.

[37] Gao, J. (2006). Gps/ins/ g sensors/ yaw rate sensor/ wheel speed sensors integrated vehicular positioning system. *Proceedings of ION GNSS*, pages 1427–1439.

[38] Gazis, D., Herman, R., and Rothery, R. (1961). Nonlinear follow-the-leader models of traffic flow. *Operations Research*, 9:545–567.

[39] Gipps, P. (1981). A behaviorl car following model for computer simulation. *Transportation esearch board*, 15:105–111.

[40] Göhring, D., Latotzky, D., Wang, M., and Rojas, R. (2013). Semi-autonomous car control using brain computer interfaces. *Intelligent Autonomous Systems 12, Advances in Intelligent Systems and Computing, (Springer Berlin Heidelberg).*

[41] Golden, J. S., Dooley, K. J., Anderies, J. M., Thompson, B. H., Gereffi, G., and Pratson, L. (2011). Sustainable product indexing: navigating the challenge of ecolabeling. *Ecol. Soc.*, 15(3).

[42] Guerreiro, C., de Leeuw, F., and Foltescu, V. (2013). Air quality in europe – 2013 report. *European Environment Agency, Copenhagen, Denmark.*

[43] Haase, H. (2011). Logistics emission trading system for green optimization. *http://www.emissionshandel letsgo.ovgu.de/letsgo media/Downloads/Schlussbericht LETSGO ILM2012.pdf*

[44] Hager, W., Hearn, D., and Pardolas, P. (1994). Large scale optimization: state of the art (kluwer academic).

[45] Hartenstein, H. and Laberteaux, K. P. (2010). *VANET vehicular applications and inter networking technologies*. WILEY.

[46] Hartman, K. and Strasser, J. (2005). Saving lives through advanced vehicle safety technology intelligent vehicle initiative final report.

[47] Häusler, F., Faizrahnemoon, M., Crisostomi, E., Schlote, A., Radusch, I., and Shorten, R. (2013a). A framework for real time emissions trading in large scale vehicle fleets. *IET Intelligent Transportation Systems*, 9(3):275–284.

[48] Häusler, F., Faizrahnemoon, M., Crisostomi, E., Schlote, A., Radusch, I., and Shorten, R. (2013b). A framework for real time emissions trading in large scale vehicle fleets. *9th ITS European congress*.

[49] Hillier, B. (1996). Space is the machine: a configurational theory of architecture. *Cambridge University Press*.

[50] Hillier, B. (1999). The hidden geometry of deformed grids: or, why space syntax works, when it looks as though it shouldn't. *Environment and Planning B: Planning and Design*, 26:169–191.

[51] Hillier, B. and Hanson, J. (1984). The social logic of space. *Cambridge University Press, Cambridge*.

[52] Hodgkinson, S. (2011). Is your city smart enough. *Is your city smart enough Ovum Analyst Insights.pdf*.

[53] Horn, R. A. and Johnson, C. R. (1985). *Matrix analysis*. Cambridge University Press.

[54] Huang, H., Zhang, D., Zhu, Y., Li, M., and Wu, M. (2012). A metropolitan taxi mobility model from real gps traces, journal of universal computer science 01/2012; 18:1072- 1092. pages 1072–1092.

[55] IBM (2007). Delivering smarter transportation systems. transport-systems-white-paper.pdf. *http://www-935.ibm.com/services/us/igs/pdf/*.

[56] Ji, S., Cherry, C., Han, L., and Jordan, D. (2013). Electric bike sharing: simulation of user demand and system availability. *Journal of Cleaner Production*, 85:250–257.

[57] Jiang, B. (2009). Ranking spaces for predicting human movement in an urban environment. *International Journal of Geographical Information Science*, 23(823–837).

[58] Jiang, B. and Liu, C. (2009). Street-based topological representations and analyses for predicting traffic flows in gis. *International Journal of Geographical Information Science*, 23(9):1119–1137.

[59] Jiang, B., Zhao, S., and Yin, J. (2008). Self-organised natural roads for predicting traffic flow: a sensitivity study. *Journal of Statistical Mechanics: Theory and Experiment*, 7(P07008).

[60] JPO, I. (2008). Vehicle safety applications. *US DOT IntelliDrive(sm) Project - ITS Joint Program Office, Tech. Rep.*

[61] Karagiannis, G., Altintas, O., Ekici, E., Heijenk, G., Jarupan, B., Lin, K., and Weil, T. (2011). Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions. *IEEE Communications Surveys and Tutorials*, 13(4):584–616.

[62] Kemeny, J. G. and Snell, J. L. (1960). *Finite Markov Chains*. Van Nostrand, Princeton.

[63] Khalil, H. (2001). Nonlinear systems. *Prentice Hall*.

[64] Kirkland, S. (2014). Load balancing for markov chains with a specified directed graph. *Linear and Multilinear Algebra*.

[65] Knorn, F. (2011). *Topics in Cooperative Control*. PhD thesis, Hamilton Institute, NUI Maynooth.

[66] Knorn, F., Corless, M., and R., S. (2011a). Results in cooperative control and implicit consensus. *Int. J. Control*, 84.

[67] Knorn, F., Corless, M. J., and Shorten, R. (2011b). A result on implicit consensus with application to emissions control. *Proc. of 50th IEEE Conf. on Decision and Control and European Control Conf. (CDC-ECC)*.

[68] Krajzewicz, D., Bonert, M., and Wagner, P. (2006). The open source traffic simulation package sumo. *RoboCup 2006 Infrastructure Simulation Competition, Germany*.

[69] Krajzewicz, D., Erdmann, J., Behrisch, M., and Bieker, L. (2012). Recent development and applications of sumo - simulation of urban mobility. *International Journal On Advances in Systems and Measurements*, 5(3):128–138.

[70] Langville, A. and Meyer, C. D. (2006a). Updating Markov chains with an eye on Google's PageRank. *SIAM Journal of Matrix Analysis and Applications*, 27(4):968–987.

[71] Langville, A. N. and Meyer, C. D. (2006b). *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, Princeton.

[72] Lasky, T. and Ravani, B. (1993). A review of research related to automated highway systems (ahs). advanced highway maintenance and construction technology research center. *UCD APR 93 10 25 01, Dept of Mechanical and Aeronautical engineering, University of California, Davis, October 1993*.

[73] Lathia, N., Ahmed, S., and Capra, L. (2012). Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C*, 22:88–102.

[74] Levene, M. and Loizou, G. (2002). Kemeny's constant and the random surfer. *American Mathematical Monthly*, 109:741–745.

[75] Liu, N. and Stewart, W. J. (2011). Markov chains and spectral clustering. *Performance Evaluation of Computer and Communication Systems*, pages 87–98.

[76] Lu, X. Y., Kim, Z., Cao, M., Varaiya, P. P., and R.Horowitz (2010). *Deliver a Set of Tools for Resolving Bad Inductive Loops and Correcting Bad Data*. California PATH Program, Institute of Transportation Studies, University of California at Berkeley.

[77] Ma, S., Zheng, Y., and Wolfson, O. (2013). T-share: A large- scale dynamic taxi ridesharing service. *In proceedings of IEEE International Conference on Data Engineering (ICDE 2013)*.

[78] Maile, M. (2009). Vehicle safety communications applications vsc a project crash scenarios and safety applications. In *SAE government industry meeting*, Washington D. C.

[79] Map, O. S. (2013).

[80] Margiotta, R. (2002). State of the practice for traffic data quality. *Traffic Data Quality Workshop, Work Order Number BAT-02-006. Washington, DC: Batelle for Federal Highway Administration*.

[81] McHale, G. (2007). Cicas program update. In *ITS world congress Beijing China*.

[82] Meyer, C. D. (1975a). The role of group generalized inverse in the theory of finite markov chains. *Siam review*, 17:443–464.

[83] Meyer, C. D. (1975b). The role of group generalized inverse in the theory of finite markov chains. *SIAM Review*, 17(3):443–464.

[84] Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge Mathematical Library.

[85] Middleton, D., Gopalakrishna, D., and Raman, M. (2003). Advances in traffic data collection and management. *One of three white papers that support a series of workshops on Data Quality, which were held in March*.

[86] Midgley, P. (2009). The role of smart bike-sharing systems in urban mobility. *Journeys*, pages 23–31.

[87] Midgley, P. (2011). Bicycle-sharing schemes: enhancing sustainable mobility in urban areas. *United Nations, Department of Economic and Social Affairs*.

[88] Moosavi, V. and Hovestadt, L. (2013). Modeling urban traffic dynamics in coexistence with urban data streams. *Proc. of the 2nd ACM SIGKDD International Workshop on Urban Computing*.

[89] Morimura, T., Osogami, T., and Idé, T. (2013). Solving inverse problem of markov chain with partial observations. *IBM Research Report RT0952*.

[90] Murphy, E. and Usher, J. (2015). The role of bicycle-sharing in the city: Analysis of the Irish experience. *International Journal of Sustainable Transportation*, 9(2):116–125.

[91] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

[92] Nedic, A. and A., O. (2009). Subgradient methods for saddle-point problems. *Optim. Theory Appl*, 142(1):205–228.

[93] Ni, D., Leonard, J. D., Guin, A., and Feng, C. (2005). Multiple imputation scheme for overcoming the missing values and variability issues in its data. *Journal of transportation engineering*, 131(12):931–938.

[94] Olfati-Saber, R., Fax, J. A., and Murray, R. M. (2007). Consensus and cooperation in networked multi-agent systems. *Proc. IEEE*, 95(1).

[95] Ordonez-Hurtado, R. H. and Shorten, R. N. (2014). Using stationary vehicles to enhance cooperative positioning in vehicular ad-hoc networks. In *International Conference on Connected vehicles and Expo (ICCVE)*, pages 863–864, Vienna, Austria.

[96] Papoulis, A. and Pillai, S. U. (2002). *Probability, Random Variables and Stochastic Processes*. International Edition.

[97] Peden, M., McGee, K., and Sharma, G. (2002). The injury chart book: a graphical overview of the global burden of injuries. *World Health Organization, Geneva*.

[98] Petovello, M. (2033). *Real-time Integration of a Tactical Grade IMU and GPS for High-Accuracy Positioning and Navigation*. PhD thesis, University of Calgary.

[99] Pfrommer, J., Warrington, J., Schildbach, G., and Morari, M. (2014). Dynamic vehicle redistribution and online price incentives in shared mobility systems. *IEEE Transactions on Intelligent Transportation Systems*, 15(4):1567–1578.

[100] Porta, S., Crucitti, P., and Latora, V. (2006). The network analysis of urban streets: A dual approach. *Physica A*, 369:853–866.

[101] Pucher, J. and Buehler, R. (2008). Making cycling irresistible: Lessons from the Netherlands, Denmark and Germany. *Transport Reviews: A Transnational Transdisciplinary Journal*, 28(4):495–528.

[102] Pucher, J., Buehler, R., and Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A*, 45(6):451–475.

[103] Pucher, J., Dill, J., and Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine*, 50(Supplement):S106—-S125.

[104] Pucher, J. E. and Buehler, R. E. (2012). *City cycling*. MIT Press, Cambridge, MA.

[105] Raney, B., Çetin, N., Vollmy, A., Vrtic, M., Axhausen, K., and Nagel, K. (2003). An agent-based microsimulation model of swiss travel: First results networks and spatial economics. 3(1):23–42.

[106] Resch, B., Calabrese, F., Biderman, A., and Ratti, C. (2008). An approach towards real-time data exchange platform system architecture (concise contribution). *In Proceeding of the 2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications*, pages 153–159.

[107] Rosen, D., Mammano, F., and Favout, R. (1970). An electronic route guidance system for highway vehicles. *IEEE Transactions on vehicular technology*, 19(1):143–152.

[108] SAFESPOTD8.4.4 (2008). Use cases, functional specifications and safety margin applications for the safespot project. *IST Safespot Project, Tech. Rep. Safespot IST-4-026963-IP deliverable D8.4.4*, pages 1–54.

[109] Schlote, A. (2014). *New Perspectives on Modelling and Control for Next Generation Intelligent Transport Systems*. PhD Thesis, Hamilton Institute, National University of Ireland Maynooth.

[110] Schlote, A., Crisostomi, E., Kirkland, S., and Shorten, R. (2011). A google-like model of road network dynamics. *In The 15th Yale workshop on adaptive and learning systems, Yale University, New Haven*.

[111] Schlote, A., Crisostomi, E., Kirkland, S., and Shorten, R. (2012a). Traffic modelling framework for electric vehicles. *International Journal of Control*, 85(7):880–897.

[112] Schlote, A., Häusler, F., Hecker, T., Bergmann, A., Crisostomi, E., Radusch, I., and Shorten, R. (2012b). Cooperative regulation of emissions using plug-in hybrid vehicles. *ACM/IEEE/IFAC/TRB Int. Conf. on Connected Vehicles and Expo, Beijing, China*.

[113] Schlote, A., Häusler, F., Hecker, T., Bergmann, A., Crisostomi, E., Radusch, I., and Shorten, R. (2013). Cooperative regulation and trading of emissions using plug-in hybrid vehicles. *IEEE Trans. Intell. Transp. Syst*, 14(4):1572–1585.

[114] Shaheen, S., Guzman, S., and Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: Past, present and future. *Transportation Research Record*.

[115] Shaheen, S., Martin, E., Cohen, A., and Finson, R. (2012). Public bikesharing in North America: Early operator and user understanding. *Research Report 11-26, Mineta Transportation Institute*.

[116] Shladover, S. (1989). Research needs in roadway automation. *Vehicle highway automation technology and plicy issues*, pages 89–104.

[117] Sichitiu, M. L. and Kihl, M. (2008). Inter-vehicle communication systems: a survey. *IEEE Communication Surveys and Tutorials,*, 10(2):88–105.

[118] Siljak, D. (2007). Large-scale dynamic systems: stability and structure (north-holland).

[119] Stanojevic, R. and R., S. (2009). Load balancing vs. distributed rate limiting:an unifying framework for cloud control. *Proc. of ICC*.

[120] Stüdli, S., Crisostomi, E., Middleton, R., and Shorten, R. (2012a). Aimd-like algorithms for charging electric and plug-in hybrid vehicles. *IEEE Int. Conf. on Electric Vehicles, Greenville*.

[121] Stüdli, S., Crisostomi, E., Middleton, R., and Shorten, R. (2012b). A flexible distributed framework for realising electric and plug-in hybrid vehicle charging policies. *Int. J. Control*, 85(8):1130–1145.

[122] Sun, S., Yu, G., and Zhang, C. (2004). Short-term traffic flow forecasting using sampling markov chain method with incomplete data. *In proceeding of Intelligent Vehicles Symposium IEEE.*

[123] Turner, S. (2002). Defining and measuring traffic data quality. *Proceedings of the Traffic Data Quality Workshop, Washington, DC.*

[124] Vanajakshi, L. and Rilett, L. (2004). Loop detector data diagnostics based on conservation-of-vehicles principle. *Transportation Research Record: Journal of the Transportation Research Board*, 1870(1):162–169.

[125] VSC (2006). Final report. *US DOT, Vehicle Safety Communications Project DOT HS 810 591, April 2006.*

[126] VSC-A (2009). Final report. *US DOT, Vehicle Safety Communications Applications (VSC-A) Project DOT HS 810 073, 2009 January 2009.*

[127] Waller, S. T., Kockelman, K. M., Sun, D., Boyles, S., Lin, D., Ng, M., Seraj, S., Tassabehji, M., Valsaraj, V., and Wang, X. (2008). Archiving, sharing, and quantifying reliability of traffic data. *Tech. Rep.*

[128] Woodburn, A., Allen, J., and Leonardi, J. (2010). International road and rail freight transport: the impact of globalisation on activity levels. *OECD, Globalisation, Transport and the Environment, OECD Publishing, Paris.*

[129] Wu, Y., Zhu, Y., and Li, B. (2011). Trajectory improves data delivery in vehicular networks. *NFOCOM 2011:*, pages 2183–2191.

[130] Yang, B., Kaul, M., and Jensen, C. S. (2014). Using incomplete information for completeweight annotation of road networks. *IEEE Transactions on Knowledge and Data Engineering*, 26(5):1267–1279.

[131] Yin, W., Murray-Tuite, P., and Rakha, H. (2012). Imputing erroneous data of single-station loop detectors for nonincident conditions: Comparison between temporal and spatial methods. *Journal of Intelligent Transportation Systems*, 16(3):159–176.

[132] Yuan, J. and Zheng, Y. (2010). T-drive: Driving directions based on taxi trajectories. *In proceedings of ACM SIGSPATIAL GIS 2010.*

[133] Yuan, J., Zheng, Y., and Xie, X. (2012). Discovering regions of different functions in a city using human mobility and pois. *In proceedings of 18th SIGKDD conference on Knowledge Discovery and Data Mining (KDD 2012).*

[134] Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3):338–353.