A Google-like Model of Road Network Dynamics and its Application to Regulation and Control

Emanuele Crisostomi, Steve Kirkland, Robert Shorten

August, 2010

Abstract

Inspired by the ability of Markov chains to model complex dynamics and handle large volumes of data in the successful experience of Google's PageRank algorithm, a similar approach is proposed here to model road network dynamics. The core of the Markov chain is the transition matrix which can be completely constructed by easily collecting traffic data. The proposed model is validated by checking the same results through the use of the popular mobility simulator SUMO. Markov chain theory and spectral analysis of the transition matrix are shown to reveal non-evident properties of the underlying road network and to correctly predict consequences of road network modifications. Preliminary results of possible applications of interest are shown and simple practical examples are provided throughout the paper to clarify and support the theoretical expectations.

1 Introduction

Intelligent traffic management is viewed as essential in reducing both congestion and harmful emissions within city limits [1]. This is recognised by both regulatory and federal authorities, and by industry; see the IBM smart city initiative [2] and Cisco's smart and connected communities initiative [3]. A key enabling technology in developing traffic management strategies are accurate traffic models that can be easily used for both prediction and control. A major objective in developing such models is to allow the development of smart traffic management systems that are proactive in predicting traffic flow and facilitate taking pre-emptive measures to avoid incidents (traffic build up, pollution peaks etc.) rather than reacting to traffic situations. Given this basic requirement, and given the trend in the automotive industry to instrument vehicles and infrastructure, a key feature of such models is that they should not only accurately model traffic flows and road dynamics, but also that they should be constructed from real data obtained directly from the road network that is obtained in near real-time.

Our contribution in this paper is to propose a new paradigm for modelling road network dynamics. The ability to use cars as sensors to harvest information in real time about the road network offers the possibility to deploy new tools to both model and engineer road networks. One such tool is the Markov chain, and here we propose to employ Markov chains to model congestion and emissions in a manner analogous to how Google employs these tools to model congestion in the Internet [4]. Markov chains offer considerable advantages over conventional road network simulators. They can be built from real data easily; they are fast and effective simulation tools. Also, Markov chains can be used to inform the design of control strategies that are suitable for regulating load in transportation networks; namely the design of load balancing strategies using infrastructure to shape the probabilities. Furthermore, Markov chain models allow users to glean structural information that is usually difficult to obtain using other modelling techniques. These include: identification of sensitive links in the network; identification 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 effects of failure of a link (i.e. due to road works or an unexpected event). Such information cannot be extracted easily from conventional simulators (most sensitive road junctions, speed of mixing, identification of subgraphs, and degree of graph connectivity). As such they are excellent traffic engineering tools and provide a mechanism to respond to congestion conditions in near real time in a preemptive manner.

Our paper is organised as follows. First we give a compact overview of road network models. Then, we review the main concepts of Markov chain theory and spectral analysis, with a special regard for the description of the key parameters that will be used to analyse the urban network. In section 0.4 we build a Markov chain model of a road network using measurements from the road network simulator. Finally, using this model, we give several applications in which we use our network model for prediction and control.

2 Stochastic models of road networks and related work

Our basic idea is to use tools from stochastic modelling to model road network dynamics. Specifically, our idea is very simple. By recording average vehicle speeds, and the average directions taken by cars when they reach a road junction, we shall use this information to build a Markov chain representing vehicular mobility patterns in an urban environment. While we believe our specific approach is novel for this application, stochastic mobility models have been employed for road network simulation in the past thanks to their inherent simplicity. A popular example is provided by the Constant Speed Motion model [5] where vehicles follow casual paths over a graph representing the road topology. The speed can be constant, or can be adjusted to take into account interactions with other vehicles. Stochastic models however usually fail to provide a sufficient level of realism for many applications of interest. For this reason *flow models* were introduced to provide a more realistic modelling of urban networks. Depending on the level of detail flow models are usually classified as *microscopic*, *mesoscopic* and *macroscopic* [6]. Mesoscopic models are at an intermediate detail level, as traffic flows are described at an aggregated level (e.g. through probability density functions), but interactions are at an individual level. Flow equations can be expressed as Partial Differential Equations (PDEs), discrete time equations or Cellular Automata (CA) models.

While vehicle motion patterns can be modeled as flows, it still remains to define the path followed by each single vehicle within the flow. In the literature, a distinction is made between a *trip* and a *route* (sometimes also called *path*). A trip is defined through a starting point (*origin*), a destination point and the departure time. A route is an expanded trip in the sense that also the sequence of roads to cover must be specified. Typically, trips are defined through Origin-Destination (OD) matrices, while 'optimal' routes can be computed according to Dijkstra-type graph algorithms, where optimality can be referred to shortest path or minimum time path. This approach clearly suffers from scalability issues when cities characterised by thousands of roads are investigated. An alternative approach consists in introducing turning probabilities at every junction to describe the probability of choosing one subsequent road segment over the others. This approach has advantageous scalability properties and can be made time variant to address different traffic behaviours (e.g. week days vs week ends or morning ingoing flows vs evening outgoing flows).

The most popular way of investigating the behaviour or the efficiency of a road network is to use mobility simulators that implement most (or all) of the previously described methods to create traffic. In this paper, the software SUMO (Simulation of Urban MObility) is used as a comparison tool [7]. SUMO is an open source, highly portable microscopic road traffic simulation package that developed at the Institute of Transportation Systems at the German Aerospace Center, and is licensed under the GPL.

We use SUMO here to illustrate the efficacy of our approach; SUMO is used both to generate data to build our Markov chain, to validate the outcomes of our modeling approach, and to illustrate other merits of the Markovian approach.

3 A Primer on Markov Chains and their Eigenspectra

The objective of this section is to present the mathematical tools that will be used to investigate the road network dynamics. The first definitions are basic and will not be described in detail, as they can be found on classic books such as [8] or [9]. More space will be dedicated to less conventional parameters related to Markov chains and spectral analysis that play an important role in the urban network counterpart. In particular, we use here the same notation of [4] as the Google's PageRank algorithm which has inspired this work on road networks.

The traffic flow will be described through a *Markov chain*, which is a stochastic process characterised by the important property

$$p(x_{k+1} = S_{k+1}|x_k = S_k, x_{k-1} = S_{k-1}, \ldots, x_0 = S_0) = p(x_{k+1} = S_{k+1}|x_k = S_k), \quad (1)$$

where accordingly to conventional notation $p(E|F)$ denotes the conditional probability that event E occurs given event F occurs. Equation (1) states that the probability that the random variable x is in state S_{k+1} at time step $k+1$ only depends on the state of x at time step k and not on preceding values. Throughout the paper only discrete-time, finite-state homogeneous Markov chains will be considered. We present no theoretical justification for this model, other than to state that Markov chains have a long history of providing compact representations of systems described by very complicated sets of dynamical equations [10], [11]; a good example is the recent work on meta-stability in molecular systems [12]. We also note that the Internet is, in spirit, similar to a road network, and Markov chains have been used with remarkable success in that application.

The Markov chain is completely described by the $n \times n$ *transition probability matrix* $\mathbb{P}$ whose entries $\mathbb{P}_{ij}$ denote the probability of passing from state S_i to state S_j, and n is the number of states. The matrix $\mathbb{P}$ is a row-stochastic nonnegative matrix, as the elements of each row are probabilities and they sum up to 1.

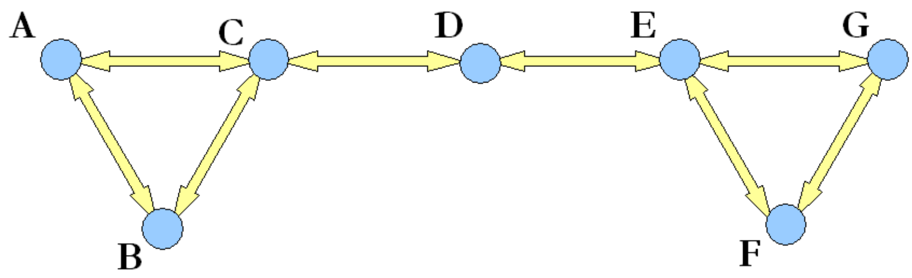

Figure 1: Example of the graph associated to a transition matrix.

Within Markov chain theory, there is a close relationship between the transition matrix $\mathbb{P}$ and a corresponding graph. A *graph* is represented by a set of *nodes* that are connected through *edges*. Therefore, the graph associated to the matrix $\mathbb{P}$ is a *directed* graph, whose nodes are represented by the states $S_i, i = 1, \ldots, n$ and there is a directed edge leading from S_i to S_j if and only if $\mathbb{P}_{ij} \neq 0$.

Example: Let us consider the transition matrix

$$\mathbb{P} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0.45 & 0.45 & 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 & 0.45 & 0.45 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}. \qquad (2)$$

The associated graph is shown in Figure 1, where nodes are enumerated from A to G and all edges are *bidirectional*. In this case all edges are bidirectional as in matrix (2) $\mathbb{P}_{ij} \neq 0 \Leftrightarrow \mathbb{P}_{ji} \neq 0$.

A graph is *strongly connected* if for each pair of nodes there is a sequence of directed edges leading from the first node to the second one. The matrix $\mathbb{P}$ is *irreducible* if and only if its directed graph is strongly connected. In the following some properties of irreducible transition matrices are shortly described, and most of them derive from the well-known Perron-Frobenius theorem:

- 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 algebraic multiplicity 1.

- The left-hand Perron eigenvector π is the unique vector defined by $\pi^T \mathbb{P} = \pi^T$, such that $\pi > 0, \|\pi\|_1 = 1$. Except for positive multiples of π there are no other non-negative left eigenvectors for $\mathbb{P}$.

In the last statement, by saying $\pi > 0$ it is meant that all entries of vector π are strictly positive.

One of the main properties of irreducible Markov chains is that the i^{th} component π_i of 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.

3.1 The Mean First Passage Time Matrix and the Kemeny constant

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 [13]. The group inverse of the singular matrix $(\boldsymbol{I} - \mathbb{P})^{\#}$ contains important information on the Markov chain and it will be often used; for this reason it is convenient to name it $Q^{\#}$ for short.

The *mean first passage time* m_{ij} from the state S_i to state S_j is the expected number of steps to arrive at destination S_j when the origin is S_i. If we denote as $q_{ij}^{\#}$ the entries of the matrix $Q^{\#}$, then the mean first passage times can be computed easily according to (see [14] for example)

$$m_{ij} = \frac{q_{jj}^{\#} - q_{ij}^{\#}}{\pi_j}, \qquad i \neq j, \tag{3}$$

where it is intended that $m_{ii} = 0, \quad i = 1, \ldots, n$.
The *Kemeny constant* is defined as

$$K = \sum_{j=1}^{n} m_{ij} \pi_j, \tag{4}$$

where the right hand side is surprisingly independent of the choice of i [15]. An interpretation of this result is that the expected time to get from an origin state S_i to a destination state S_j selected randomly according to the equilibrium measure π does not depend on the starting point S_i [16]. Therefore the Kemeny constant is an intrinsic quantity of a Markov chain, and if the transition matrix

$\mathbb{P}$ has eigenvalues $\lambda_1 = 1, \lambda_2, \ldots, \lambda_n$ then another way of computing K is [17]

$$K = \sum_{j=2}^{n} \frac{1}{1-\lambda_j}. \tag{5}$$

Equation (5) emphasises the fact that K is only related to the particular matrix $\mathbb{P}$ and that it increases if eigenvalues of $\mathbb{P}$ (and in particular the second eigenvalue of largest modulus) are close to 1.

3.2 Spectral analysis of the transition matrix

Accordingly to the previous paragraph, the eigenvalues of the transition matrix determine the value of the Kemeny constant. However throughout this paper, we give to spectral analysis a broader meaning than just the spectrum of the matrix $\mathbb{P}$ and we generalise the term to include the associated eigenvectors. Eigenvectors of graph matrices are known to have good clustering properties, and a modern active area of research is called *spectral clustering*. An overview of spectral clustering algorithms is provided in [18], although most of the results presented therein hold for undirected graphs.

The rationale behind the clustering properties of the eigenvector associated to the second eigenvalue of largest modulus of $\mathbb{P}$ is now anticipated through an illustrative example. Suppose that we have two irreducible stochastic matrices P_1, P_2 of orders k and $n-k$, respectively. Assume that the last column of P_1 and the first column of P_2 are both positive. Consider the matrix $A = \left[\begin{array}{c|c} P_1 & \mathbf{0} \\ \hline \mathbf{0} & P_2 \end{array}\right]$; note that A has 1 as an eigenvalue of multiplicity two. Suppose now that we perturb A slightly to obtain the matrix $B = \left[\begin{array}{c|c} P_1 & \mathbf{0} \\ \hline \mathbf{0} & P_2 \end{array}\right] + \epsilon \left[\begin{array}{c|c} -\mathbf{1}_k \boldsymbol{e}_k^T & \mathbf{1}_k \boldsymbol{e}_1^T \\ \hline \mathbf{1}_{n-k} \boldsymbol{e}_k^T & -\mathbf{1}_{n-k} \boldsymbol{e}_1^T \end{array}\right]$, where $\mathbf{1}_m$ represents an all ones vector of order m, $\boldsymbol{e}_k$ is a vector of zeros with a 1 in k^{th} position, and ϵ is a small positive number. Positivity of the last column of P_1 and the first column of P_2 guarantees that the perturbed matrix does not have negative entries for small ϵ. It is straightforward to see then that B is a stochastic matrix which has $1-2\epsilon$ as an eigenvalue, with corresponding right eigenvector $\left[\begin{array}{c} \mathbf{1}_k \\ \hline -\mathbf{1}_{n-k} \end{array}\right]$. In particular, the sign pattern of this eigenvector for an eigenvalue close to 1 corresponds to the partition of A into 2 irreducible diagonal blocks. The idea previously outlined is formalised in the appendix, where the concept of graph cluster is stated in mathematical terms and formal arguments are provided as well.

4 Modelling a road network as a Markov chain

The connection between a road network and a Markov chain is straightforward if a city map is interpreted as a directed graph, where nodes correspond to junctions and edges to connecting roads. In the literature related to urban networks, this representation is sometimes called *primal* [19] in contrast to the dual representation where the role of streets and junctions is reversed (i.e. in the dual representation streets correspond to nodes and junctions to edges). The use of graph theory to analyse urban networks was proposed in the pioneering work of Hillier and Hanson [20] in the late eighties and further developed in the later works [21, 22]. An important achievement of the proposed theory was the establishment of a significant correlation between the topological accessibility of streets and urban properties like pedestrian and vehicular flows, human way-finding, safety against microcriminality, micro-economic vitality and social liveability [21]. The topology of the urban network is mathematically analysed by computing the *degree of the nodes*, 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 [23, 24, 25]. Although all approaches exploit well-established mathematical tools borrowed from graph theory, the starting point is a simple plain urban map (or its dual representation) which usually does not include quantitative data that are important to evaluate traffic. Therefore important variables like speed limits, street lengths, junction turning probabilities, numbers of lanes, presence of traffic lights and priority rules are neglected. The objective of this work is 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 clearly affect traffic flows.

4.1 From a road network to a Markov chain

This section shows how to construct the Markov chain transition matrix. Throughout the paper a simple road network will be used as a benchmark example to support and clarify the theoretical approach. The example is the road graph shown in Figure 1. The network was deliberately chosen to be simple, so that the behaviour of the proposed method can be evaluated easily. However, recall that the proposed procedure can be applied to more realistic maps with thousands of nodes, without scalability issues, as Google's PageRank algorithm has successfully proven. The network of Figure 1 was designed to represent (in a stylized way) a city with two main communities connected through junction D. This idea is consistent with cities like Dublin where the North and the South parts are separated by a river and are connected through bridges.

The first step to pass from a road network to a Markov chain is to transform the primal map into the dual one, where the nodes of the graph are represented by roads, as shown in Figure 2. The nodes of the dual network have been called XY intending that XY is the road that connects junction X to Y, where X and Y were nodes in the primal network. The dual network is more convenient than the primal because it includes more information:

- In the primal network some edges should be inhibited depending on the

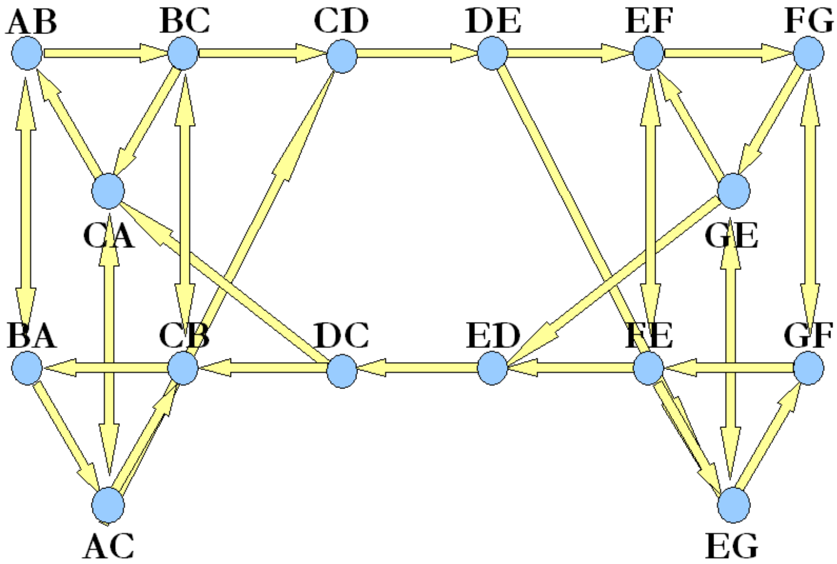

Figure 2: A possible example of dual network associated to the urban map shown in Figure 1.

edge of origin. For instance from Figure 1 it seems possible to go from node C to D and then to come back, while the more detailed dual network of Figure 2 shows that at the end of road CD turnaround is not permitted, and a longer route should be planned to enter road DC.

- A typical way of creating traffic flows is to exploit junction turning probabilities (see section 0.2). The probability of choosing an out-going road at a junction clearly depends on the road segment of origin. This information is lost in the primal network.

To represent traffic flow, we assume that turning probabilities at each junction are available. From a practical point of view, this implies that at each junction a webcam counts the number of vehicles that turn right, go straight or turn left. The collected data are then averaged to estimate mean probabilities. Alternatively, the same result can be obtained if each car stores its own route and then communicates its data to a central entity that collects and analyses data. At the end of the process, we assume that turning probabilities are available for each road. For instance, if we consider the road segment AC of Figure 2, we assume that probabilities of going from road AC to CA, CB and CD are available, as summarised in Figure 3.

A second step is used to take into account different travel times. In the

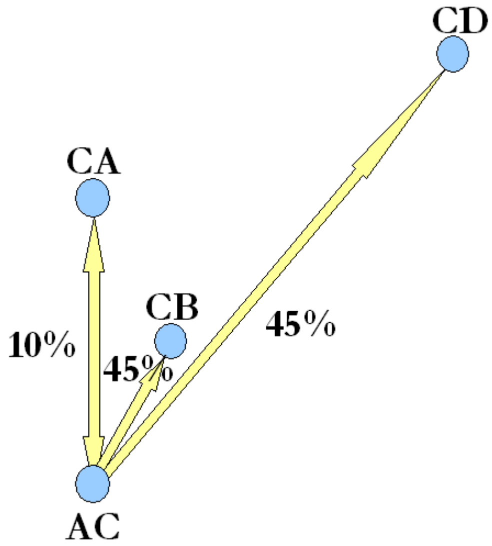

Figure 3: Turning probabilities from road segment AC. Cars coming from road AC will successively choose roads CB or CD with equal probability, while it is more unlikely that they will turnaround to take road CA.

Markov chain framework, transitions from one node (i.e. road) to the successive one take place in one time step. Clearly, independently of the time unit and even neglecting traffic, the time to cover single roads is not constant and it will generally depend on the length of the road, speed limits, conditions of the road surface, etc... For instance, let us consider an identical junction to that of Figure 3, with the only difference that the road AC is substituted by road $A^{'}C^{'}$ which has double length than AC and has half speed limit. Let us assume therefore that the time to cover it is four times the time required for road AC. A simple way to take into account the extra travel time is to include a self-loop in road $A^{'}C^{'}$ and adjust the other probabilities accordingly so that they still sum up to 1, as it is shown in Figure 4.

In principle, it is not easy to compute the average time required to cover a road, as it is related through unknown functions of physical quantities like the conditions of the road surfaces and other relevant variables as for instance the presence of pedestrian crosses, bus stops and so on. More generally, traffic conditions have a very strong impact on average travel times. In order to include all these variables, we assume that average travel times for each road are also available. Again, they can be computed on-board and communicated to some central collection point, or it could be required that two sensors at the beginning and at the end of a road record passing times. If travel times are computed for

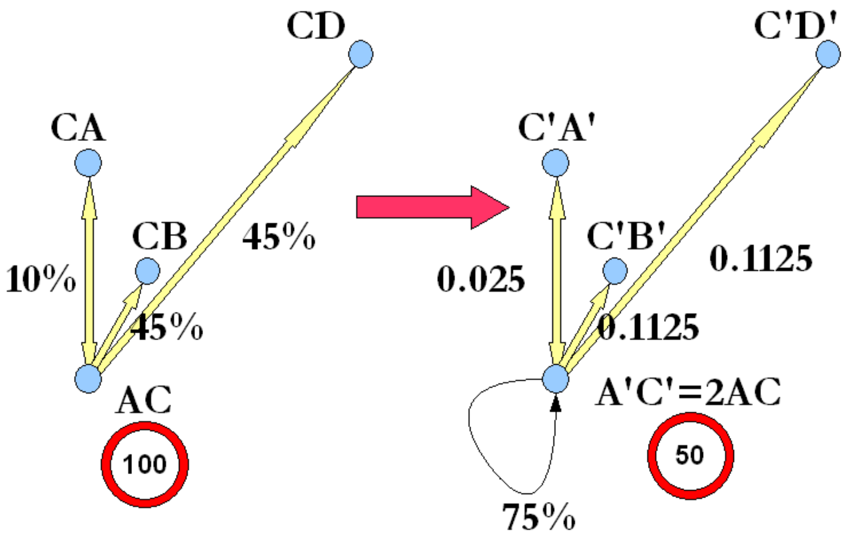

Figure 4: Let us consider an identical junction to that of Figure 3, with the difference that road $A^{'}C^{'}$ requires four times the time required to cover AC (because the length doubles and the speed limit is reduced). The problem is solved by introducing a self-loop and adjusting the remaining probabilities accordingly.

all segment roads, and they are normalised so that the smallest travel time is 1, then the probability value associated to each self-loop is

$$\mathbb{P}_{ii} = \frac{tt_i - 1}{tt_i}, \qquad i = 1, \dots, n \tag{6}$$

where tt_i is the average travel time (estimated from collected data) for the i^{th} road. The proof of equation (6) and its rationale are provided in the Appendix. At this point, the off-diagonal elements of the transition matrix $\mathbb{P}$ can be obtained as

$$\mathbb{P}_{ij} = (1 - \mathbb{P}_{ii}) \cdot tp_{ij}, \qquad i \neq j, \tag{7}$$

where tp_{ij} is the turning probability (estimated from collected data) of going from road i to road j.

In conclusion of this section, we emphasise the fact that the transition matrix $\mathbb{P}$ can be obtained after gathering the **average travel times** and **junction turning probabilities**. The nodes of the graph associated to the matrix $\mathbb{P}$ (and therefore the size of the matrix) are obtained from the dual representation of the road network. The diagonal and off-diagonal entries of the transition matrix are then found according to equations (6) and (7). We also note that the transition matrix $\mathbb{P}$ is always irreducible, otherwise the corresponding digraph would not be strongly connected, which implies that it would be impossible to go from one particular road to another road. Of course, this does not occur in road networks. Finally, we also add the following remark regarding the proposed Markov chain model.

Remark:
In practice, according to the previous model, the vehicle is represented as a particle following a random walk (although probabilities are given from real data) through a directed graph. We are aware of course that vehicles do not actually follow a random walk; however, by analogy with similar approaches like the PageRank model already mentioned (which posits a computer user performing a random walk on the world wide web), the underlying stochastic process is viewed as a mechanism for obtaining useful information, rather than a literal representation of the behaviour of a single vehicle.

4.2 Validation

The procedure previously outlined is a data driven approach and can be applied provided that junction turning probabilities and road travel times are available. Therefore, the road network of Figure 1 is now simulated with the software SUMO and the required data is collected from the simulation.

Remark:
In this work, the simulated traffic scenario plays the role of the real environment, so from one point of view this corresponds to collecting data from the

simulation rather than from the real world. From another point of view however, a realistic simulation requires in practice the same data as the Markov chain approach: one must decide a priori the traffic flows over the road network, and then provide the resulting data to both the simulator and to the Markov chain transition matrix.

In order to perform the desired simulation, four steps must be performed in SUMO:

1. The desired network shown in Figure 1 is created in SUMO by fixing the position of nodes and edges.

2. Several flows of cars are created inside the road network. This choice affects the traffic load.

3. Junction turning probabilities are fixed, so that car routes around the network can be planned.

4. SUMO provides several interesting statistics regarding average and overall network properties. The desired output files with the associated statistics must be chosen.

One of the main ideas of the proposed work is to substitute simulations with the Markov chain transition matrix, therefore a comparison of the outcomes of both approaches is required.

4.2.1 Stationary distribution of cars

The stationary distribution corresponds to the long-run fraction of time that cars will be along a particular road. Although it does not carry information about the traffic load (i.e. the number of cars), it is still valuable as it is possible to evaluate whether traffic is balanced and what roads are particularly busy, and eventually crucial. One output provided by SUMO is the occupancy of each road (measured as vehicles/Kilometers), from which it is possible to compute the relative density. In the Markov chain approach, this information is the left Perron eigenvector.

We assume that the following transition matrix $\mathbb{P}$ is extracted from the collected data

$$
\mathbb{P} = \begin{bmatrix}
0 & 0 & 0.1 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.1 & 0.8 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.1 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.8 & 0.1 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.9 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.1 & 0.8 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.8 & 0.1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 & 0 & 0
\end{bmatrix}. \tag{8}
$$

The nodes are the 16 roads represented in Figure 2 and are taken in alphabetical order. The entries of the transition matrix reveal that turnarounds are unlikely and that most cars tend to travel within the two sets of junctions A-B-C and E-F-G and rarely pass from one set to the other; in this first example we also notice that travel times are considered constant as the diagonal elements are all zero. Figure 5 compares the stationary distribution obtained from the simulator with the left Perron eigenvector of matrix (8). We also remark that the simulated distribution is of course prone to statistical variations caused by the considered set of cars.

4.2.2 Road clusters

In the proposed example cars tend to travel within sets A-B-C and E-F-G of Figure 1, therefore it is interesting to evaluate the signs of the entries in the eigenvector associated to the second largest eigenvalue, provided that it is real. As it could be expected, the second eigenvalue is close to 1 (0.9216), and the entries of the associated eigenvector are shown in Figure 6. It is not straightforward to extract the same information from the simulator. This information is however very valuable because it can be used to find hidden communities within an urban network more complicated than the one of Figure 1. A formal definition of community from a mathematical point of view is given in the Appendix.

4.2.3 Mean first passage times

It is very simple to compute the mean first passage times from the transition matrix $\mathbb{P}$, accordingly to equation (3). It is very complicated to compute the

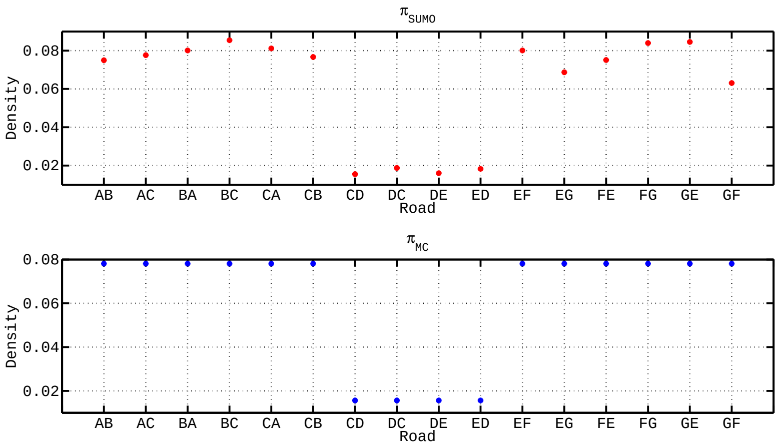

Figure 5: Comparison between the stationary distribution of cars estimated from the Markov chain approach (below) and computed from the simulation (above).

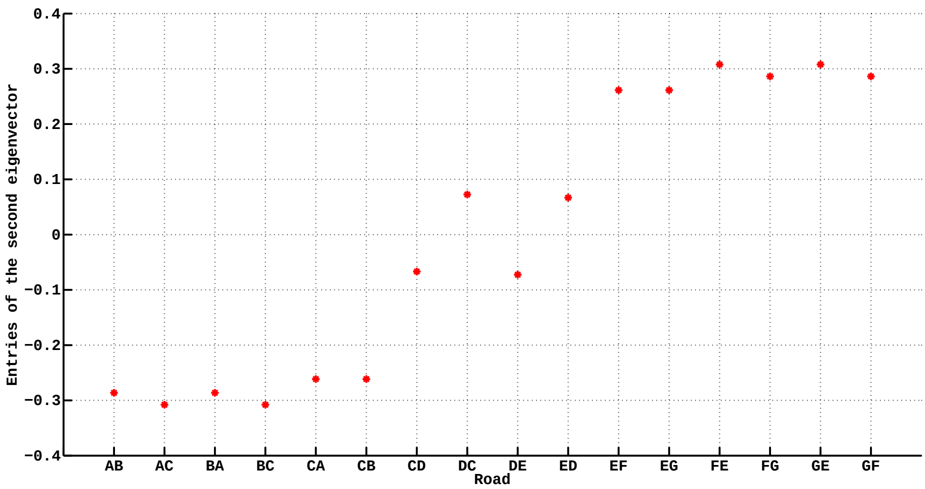

Figure 6: The second eigenvector clearly separates the first 6 roads belonging to the first cluster A-B-C, the last 6 roads of cluster E-F-G and the four roads that connect the two clusters (CD, DC, DE, ED).

same quantity from the simulator SUMO. For this purpose, a different simulation was performed for each possible entry of the 16×16 mean first passage time matrix M. A flow of cars was simulated to start from the origin road until the destination road, and the average required time with respect to all the chosen routes was computed. The two matrices are compared in Figure 7, where the origin and destination roads are represented on the x-y axes and times are on the z-axis. The resemblance between the two mean first passage time matrices is

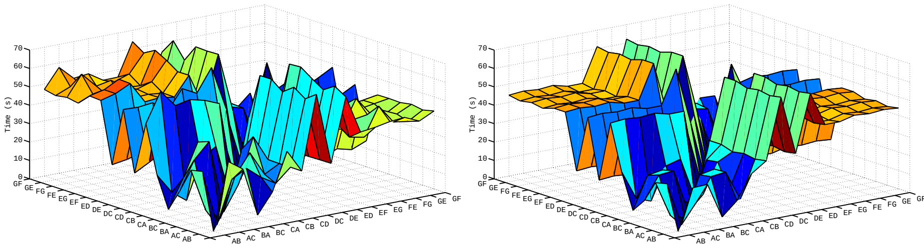

Figure 7: Mean first passage times extracted through an ensemble of simulations (left) and computed from the Markov chain matrix (right). The x-y axes contain the origin and destination roads, while the z-axis contains the average first passage travel time.

further shown in Figure 8 which compares the contour lines of the two matrices (entries of the matrices that have similar values).

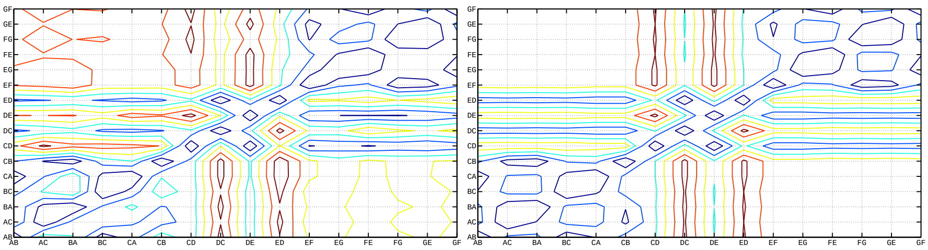

Figure 8: Contour lines of the mean first passage time matrix extracted through an ensemble of simulations (left) and computed from the Markov chain (right). The x-y axes contain the origin and destination roads.

4.2.4 Kemeny constant

We remind the reader for convenience that the Kemeny constant can be computed as in (4) in section 0.3.1

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

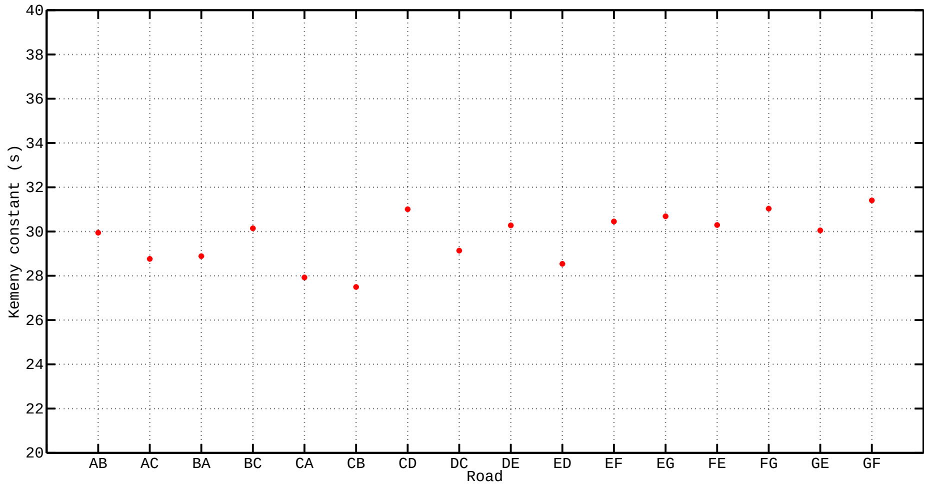

Figure 9: The Kemeny constant computes the average time required to go from one source road to a random destination chosen according to the stationary distribution. Surprisingly, this quantity does not depend on the starting road, and is a global parameter of the road network. The set of simulations show that also in the provided example, the "Kemeny constants" are rather close to being independent of the starting road, represented in the horizontal axis.

and the important result is that it is independent from the choice of the road of origin i. Therefore we used the mean first passage time matrix previously extracted from the simulations to check if the Kemeny constant is indeed constant also in practice (or, at least, in simulations). The result is shown in Figure 9.

Remark:
The mean first passage time matrix computed from SUMO simulations contains times expressed in seconds, while, at least in principle, the entries of the same matrix from the Markov chain are expressed in number of steps. Therefore, also the Kemeny constants are given in seconds and steps respectively. The multiplicative factor to convert steps to seconds can be chosen in such a way the Markov chain Kemeny constant corresponds to the average of those computed via simulations (shown for instance in Figure 9). Alternatively, should the Kemeny constant from simulation (or from real data) not be available, travel times can be used. In the proposed example, for instance, all cars coming from road CD must continue to DE (see for instance Figure 2). Therefore the travel time of CD corresponds directly to one step in the Markov chain framework.

4.2.5 Comparison in presence of traffic

In this paragraph we just intend to show that the approach works also in the presence of traffic. The main difference with the previous example is that travel times are larger because of traffic (traffic was simply simulated in SUMO by increasing the number of cars). The diagonal elements of the new transition matrix are computed according to equation (6) while the non-diagonal elements maintain the previous relative ratios. (For simplicity we assumed that traffic only slowed cars, but did not affect junction turning probabilities. As junction turning probabilities are collected from data, this assumption is not necessary). According to simulation data, the new matrix becomes

$$\mathbb{P}=\begin{bmatrix}
0.168 & 0 & 0.083 & 0.749 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.621 & 0 & 0 & 0.038 & 0.304 & 0.038 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.040 & 0.361 & 0.598 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.338 & 0.529 & 0.066 & 0.066 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.345 & 0.038 & 0 & 0 & 0.616 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.717 & 0.080 & 0 & 0.203 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.049 & 0 & 0.951 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.469 & 0.469 & 0 & 0.062 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.935 & 0 & 0.065 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.099 & 0 & 0.090 & 0.811 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.236 & 0 & 0 & 0.076 & 0.688 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.016 & 0.016 & 0.126 & 0.843 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.144 & 0.771 & 0.086 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.073 & 0.586 & 0.073 & 0 & 0 & 0.268 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.838 & 0.093 & 0 & 0.069
\end{bmatrix},$$

where entries are rounded to three decimal places. In Figure 10 the stationary distributions predicted by the Markov chain (above) and computed through the simulation (below) are compared. It is interesting to note that they are completely different from the distribution shown in Figure 5, but bear a strong resemblance to each other.

5 Traffic Management & Control

In this section we claim that the Markov chain traffic model is particularly suitable to control and regulate traffic, and to both engineer and investigate road networks. In the following we give an overview of several novel control theoretic applications that are easily realised using the Markov chain approach. These include applications to control and regulate traffic such as the following.

(i) Novel approaches to routing based on: (a) First Mean Passage Times; (b) Emissions based Markovian models;

(ii) Load balancing in traffic networks to improve traffic flows (e.g. timing of traffic lights);

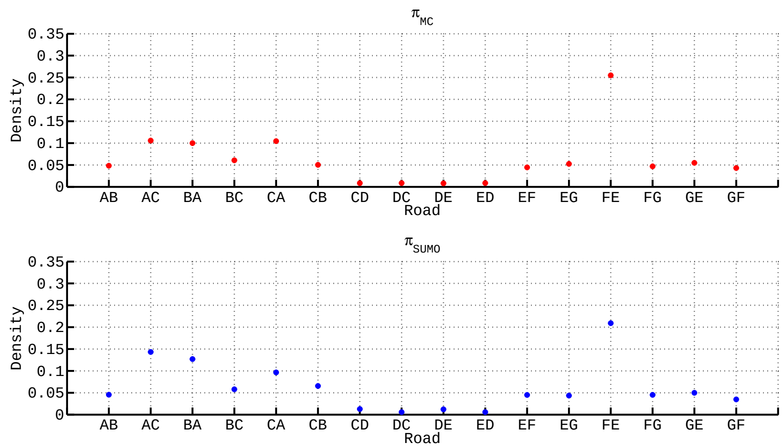

Figure 10: Comparison between the stationary distribution of cars estimated from the Markov chain approach (above) and computed from the simulation (below).

(iii) Identification of critical links in road networks.

Further applications can be outlined to support road network designers include designing road networks to:

(iv) Minimizing or decreasing the Kemeny constant;

(v) Control of the stationary distribution to balance traffic load;

(vi) Conditioning of the second eigenvector to improve robustness of network.

In the remainder of this section we give a flavour of how the Markov modelling paradigm can be used to achieve these objectives.

5.1 Control and regulation of traffic

5.1.1 Routing

Smart routing of traffic is seen as a major enabler of reduced carbon transport []. It enables more efficient use of the road network, and it can be used proactively to avoid pollution peaks in certain urban areas. Currently, in practice, given a pair of origin/destination roads, the optimal route is computed, usually in terms of minimum time or minimum distance based on map information. On the other hand, the Markov chain transition matrix is constructed from real traffic data, including average travel times of each single road, and expected congestion spots in the form of the Perron vector. This information can be easily exploited to plan shortest time routes accordingly to popular algorithms like Dijkstra [26] or other dynamic programming algorithms [27]. Dijkstra's algorithm is widely used for routing problems and solves the shortest path problem for a graph with nonegative edge path costs. Also notice that as travel times are taken as edge costs, then the solution takes traffic conditions into account as well.

An alternative solution is obtained if Dijkstra's algorithm is used to minimise a different cost function. Here we describe the solution obtained if the edge costs correspond to mean first passage times and we compare it with the minimum time solution. Although the mean first passage time matrix takes traffic delays into account as well, the main difference is that it also includes information related to junction turning probabilities. Therefore it also accounts for the possibility that the driver takes a different path from the scheduled one and for how much time is wasted due to the wrong (or alternative) choice. Indeed, the route suggested by the mean first passage time approach and the minimum time path coincide in the case of $(0, 1)$ transition matrices.

The road network example shown in Figure 11 clarifies the differences between the two methods. For simplicity it is assumed that all turning probabilities are the same, and that the same time is required to travel along each road of the network. It is therefore obvious that the minimum-time paths from AB to DA are $AB-BC-CD-DA$ and $AB-BI-ID-DA$. On the other hand, the optimal path accordingly to mean first passage times is $AB-BG-GF-FD-DA$. The reason is that the time-optimal routes are more prone to mistakes that increase the overall travel time.

Minimum pollution path The objective of this paragraph is to show how the proposed Markov chain model can be easily modified to take environmental issues into account. Let us assume that the average emissions per unit time per road are available. Accordingly to our usual approach, this implies that average CO_2 emissions (for instance) are measured along each road. Let us denote by e_j the emission per unit time along road j, therefore the contribution of pollution of a single car that takes tt_j seconds to travel along road j is $e_j \cdot tt_j$. We assume that the emission coefficient e_j depends on the road j because roads characterised by frequent changes of speed limits or by average heavy traffic have a stronger impact on pollution. We also consider normalised values obtained after dividing by a scale factor, so that $\min_j \{e_j \cdot tt_j\} = 1$.

Let us now change the nominal transition matrix $\mathbb{P}$ where travel times have not been taken into account yet (i.e. with zeros in the diagonal) into a new transition matrix $\hat{\mathbb{P}}$ that further includes emissions. With an analogous argument to that of equation (6), further detailed in the Appendix, it is sufficient to include non-zero diagonal elements, for instance for road i we have

$$\hat{\mathbb{P}}_{ii} = \frac{tt_i \cdot e_i - 1}{tt_i \cdot e_i}, \tag{9}$$

so that the expected number of emissions before leaving the road is $e_i \cdot tt_i$. The

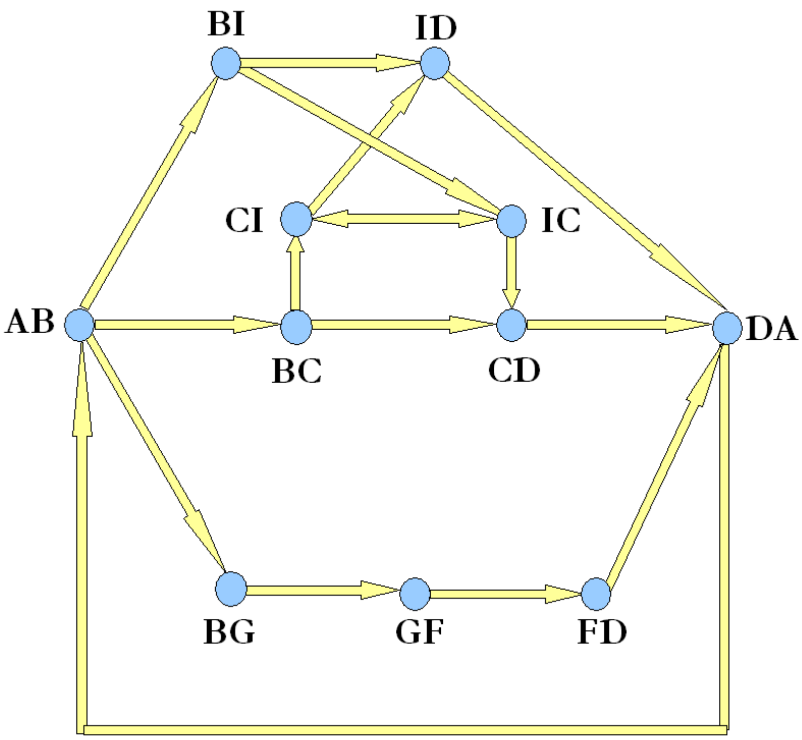

Figure 11: The time-optimal routes from AB to DA are $AB-BC-CD-DA$ and $AB-BI-ID-DA$. If the sum of first mean passage times is considered as a cost function for optimality, the best solution is the path $AB-BG-GF-FD-DA$, as there are fewer chances to take wrong turns and to increase the travel time.

off-diagonal terms of $\hat{\mathbb{P}}$ must change accordingly so that the matrix remains row-stochastic:

$$\hat{\mathbb{P}}_{ij} = \left(1 - \hat{\mathbb{P}}_{ii}\right) \mathbb{P}_{ij}, \quad \forall i, j, i \neq j \tag{10}$$

Then the matrix $\hat{\mathbb{P}}$ describes a new Markov chain where the transition step is not a unit time but a unit of pollution emissions. Accordingly, the stationary distribution represents the long-run fraction of emissions along each single road, while mean first passage times correspond to the mean first passage emissions.

In a similar way to section 0.5.1, Dijkstra's algorithm can be used to compute the minimum pollution path or the minimum mean first passage pollution path. With an analogous argument to that of 0.5.1, the second optimal path takes into account the possibility that pollution might increase due to deviations from the nominal path.

5.1.2 Timing traffic lights

In this section it is shown that the Markov chain traffic model can provides simple tools to tune traffic lights (in particular the ratio of green with respect to red times) with a view of improving traffic flow. Optimal timings are computed accordingly to theoretical expectations and SUMO is used to confirm the traffic improvements. As timing traffic lights directly affects all the streets involved in the junction, in this application it is convenient to use also the primal network representation where each node corresponds to a junction. The beginning of this section is therefore dedicated to illustrate how it is possible to pass from a dual representation to a primal one, while in the second part the traffic application is illustrated.

In the primal representation nodes correspond to road intersections, thus the stationary distribution refers to the long-run fraction of time that cars will be at a particular junction. In the following, we will make a distinction between $\pi(X)$, which represents the stationary distribution at junction X, and therefore in the primal network, and $\pi(XY)$, which represents the stationary distribution along road XY from intersection X to intersection Y, and therefore in the dual network. It is possible to derive the stationary distribution of the primal network easily from the stationary distribution of the dual network as

$$\pi(X) = \sum_{*} \pi(*X) \tag{11}$$

where the symbol "$*$" denotes all the roads that end at junction X. For instance, for junction C of Figure 1 one obtains $\pi(C) = \pi(AC) + \pi(BC) + \pi(DC)$. Similarly, the entries of the primal transition matrix $\mathbb{P}_p$ can be computed from dual data as

$$\mathbb{P}_p(XY) = \frac{\sum_{*} \pi(*X)\, \mathbb{P}(*X, XY)}{\sum_{*} \pi(*X)}. \tag{12}$$

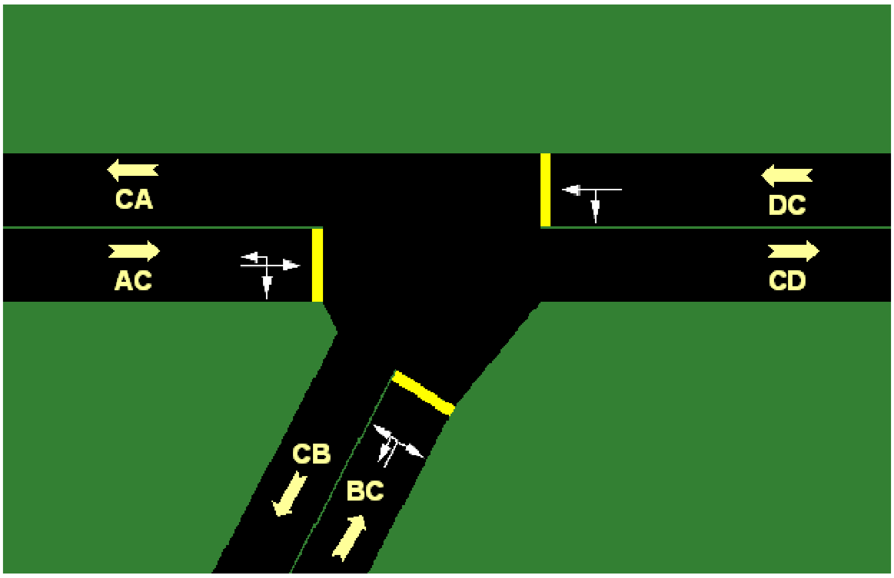

Figure 12: Junction C from Figure 1. All connection between roads are permitted except from DC to CD. The junction can be easily realised with a roundabout where flows of cars from AC, BC and DC in turn have a green light.

For completeness, we also remark that as the dual network is richer in information than the primal network, it is not generally possible to compute the transition matrix $\mathbb{P}$ of the dual using only information of the primal, but it is possible to infer the stationary distribution of the dual as

$$\pi(XY) = \pi(X)\,\mathbb{P}_p(XY) \tag{13}$$

"Optimal" green times can be obtained from the knowledge of the number of cars queuing at each traffic light and their next destination. This information can be easily recovered from the dual stationary distribution (i.e. density of cars along each road) and average junction turning probabilities. A simple method to time traffic lights is now given through the example of junction C taken from Figure 1. The junction is shown in Figure 12, which is a snapshot from SUMO, and it is formed by the three in-going roads AC, BC and DC and the three out-going roads CA, CB and CD. Note that the only impossible connection is from DC to CD. In practice, a sensible way to implement the junction is to build a roundabout, and to give green light in turn to flows of cars coming from AC, BC and DC (this is also motivated by the fact that all roads have one only lane). If no further information is considered, a first solution is to have equal periods of green light for all the three flows. Let us denote by T this green

period and by π_{AC}, π_{BC} and π_{DC} the components of the Perron eigenvector corresponding to the roads of interest. Then, we set the green period of each road accordingly to this theoretical distribution, for instance the green time of road AC becomes $\frac{\pi_{AC}}{\pi_C}T$ where π_C is computed according to (11).
The same network is then simulated again in SUMO with the new traffic light timings and improvements are measured in terms of the different stationary density at junction C. As all the cars take exactly the same route in both simulations, the difference in the long-run fraction of time spent at junction C *only* depends on the different traffic light timings.
With the same green periods the primal stationary distribution computed from the simulation was

$$\pi_{SUMO}^T = [0.0696 \quad 0.0661 \quad 0.4601 \quad 0.0766 \quad 0.2525 \quad 0.0323 \quad 0.0428] \quad (14)$$

while after changing the green periods, it becomes

$$\pi_{SUMO}^T = [0.0599 \quad 0.0706 \quad 0.3278 \quad 0.0979 \quad 0.3652 \quad 0.0336 \quad 0.0450] \quad (15)$$

with an evident decrease in the third entry which corresponds to junction C, illustrating the fact that there are less cars queuing at the traffic light. We remark that improvements are planned on the basis of theoretical data (the Perron eigenvector) and are confirmed by SUMO simulations.
By comparing (14) with (15) it is also possible to notice that the density of cars vanished from junction C spreads to other junctions of the network, and most of it at the following junction E of Figure 1 (as could be easily expected by visual inspection of the road network), which corresponds to the the fifth entry of the stationary distribution vector. In the later section 0.5.2 it is shown that this result could be predicted as well without performing simulations.

5.1.3 Identification of critical links

The objective of this section is to illustrate the use the Markov chain transition matrix to find critical links within the road network. Criticality of a road can be defined accordingly to several points of view: in this section we consider a road critical if travel times increase when the road is missing (i.e. due to road works or an unexpected event). This problem can be even more serious if criticality is not expected at all, for instance because in normal traffic conditions the road is not particularly busy.

In the usual example of Figure 1, if some of the roads are closed, connectivity of the road network is lost. This is by far the worst scenario, as depending on the starting point, some destinations would not be reachable any more. In realistic and complicated road networks it is possible to predict the loss of connectivity as a consequence of removing a road by computing the rank of the incidence matrix C. In the directed dual network, the incidence matrix C is the $(0, -1, 1)$-matrix

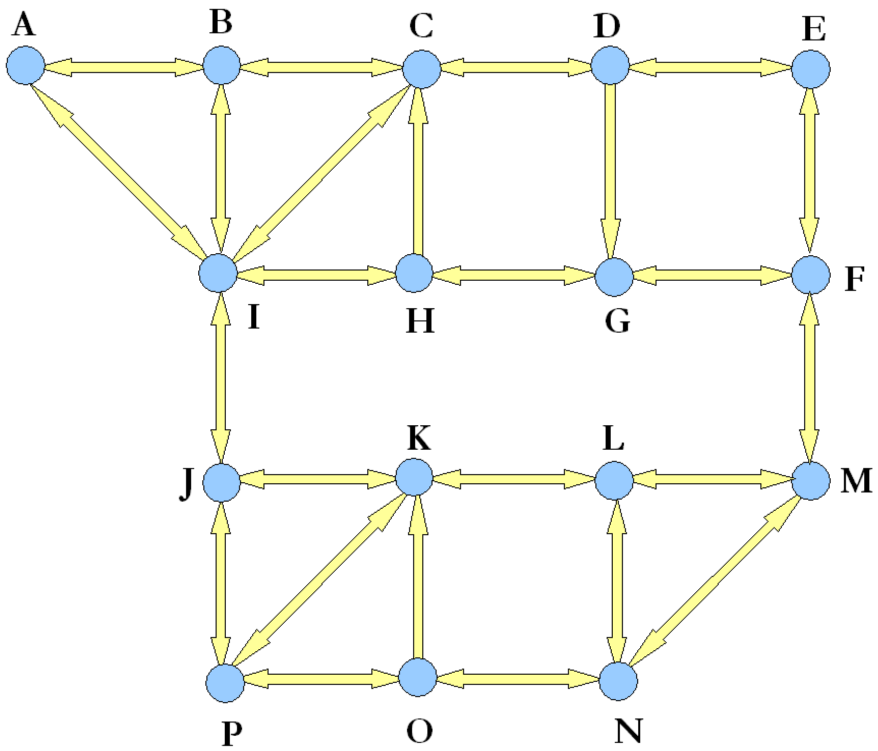

Figure 13: A new example of road network where connectivity is preserved if any one of the roads is closed to traffic. The primal network is represented only for the sake of simplicity, since the dual network consists of 49 nodes.

having [4]

$$c_{ij} = \begin{cases} 1 & \text{if edge } E_j \text{ is directed toward node } N_i \\ -1 & \text{if edge } E_j \text{ is directed away from node } N_i \\ 0 & \text{if edge } E_j \text{ neither begins nor ends at node } N_i \end{cases}$$

A directed graph with n nodes and m edges and $n \times m$ incidence matrix C is connected if and only if $rank(C) = n - 1$.

A new example shown in Figure 13 is now provided where the graph remains strongly connected no matter which link is deleted, so that another parameter than the loss of connectivity is required to identify critical roads. Figure 13 represents the primal network, while the dual is not shown for the sake of clarity, as it consists of 49 nodes. It is assumed that all turnarounds are prohibited, except those required to preserve connectivity in case of closed roads. For

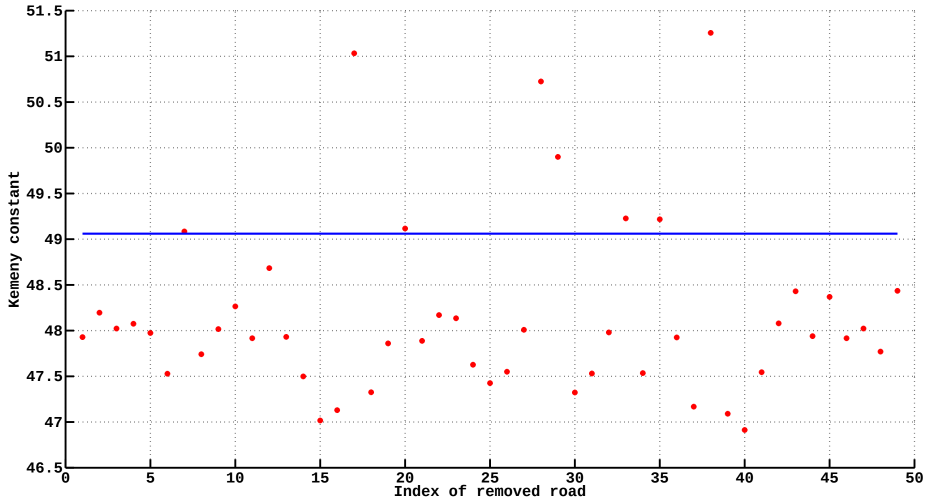

Figure 14: Values of the Kemeny constant as each road of the original road network is closed. Accordingly to visual inspection of Figure 13 the most critical roads are IJ, JI, FM and MF. Roads are numbered according to alphabetical order.

simplicity, equal turning probabilities and constant travel times are considered for all the roads (therefore there is no traffic, or at least it is uniformly spread). The Kemeny constant is computed for the whole network in the case that each road is individually closed (so that 49 Kemeny constants are obtained). Kemeny constants represent a global performance cost of the network and obviously it is desired that they are as small as possible as they measure the average time required to reach a destination chosen randomly accordingly to the stationary distribution. In the example of Figure 13 the largest values of the Kemeny constants are obtained if roads IJ, JI, FM and MF are deleted, as shown in Figure 14, which is the expected outcome by visual inspection of the network. It is interesting to remark that the result was found without taking traffic into account, and to emphasise that critical links occupy respectively positions 32, 39, 23 and 13 in the rank of most trafficked roads in the stationary distribution, and therefore it is not trivial to predict their criticality.

5.2 Road network engineering

In this section we use properties of the Markov transition matrix $\mathbb{P}$ as an indicator of the quality of the road network. For the sake of clarity the properties of $\mathbb{P}$ and their interpretation in the road network context are briefly summarised in Table 1. In this section we show how road network designers can use these properties as control variables to design networks.

Table 1: Direct comparison of average $|u|_{max}$, IAU, IADU, ISE performance indices and convergence time T 100 random initial conditions x_0.

Property	**Meaning**
Perron Eigenvector	Congested roads in the network
Second Eigenvalue	(a) Rate of convergence to stationary distribution (b) If it is real, it identifies the presence of weakly-connected sub-communities
Second Eigenvector	In case (b) it associates nodes to sub-communities
First Mean Passage Times	Average travel time from origin to destination
Kemeny Constant	Average travel time in the network
Perron Eigenvector (Primal)	Congested junctions in the network

5.2.1 What type of road network has the smallest Kemeny constant?

In section 0.3.1 we defined the Kemeny constant as a global indicator of the network, because it depends solely on the eigenvalues of the Markov chain transition matrix. As the Kemeny constant also corresponds to the average time to travel from an arbitrary road to a destination chosen randomly accordingly to the stationary distribution, it suggests that networks characterised by small values of the Kemeny constant should be more efficient in terms of traffic flow. This section provides the intuitive interpretation of which type of road network is characterised by the smallest Kemeny constant.

In paper [28] it is shown that given an irreducible matrix T, there exists a matrix of minimum Kemeny constant T^* such that its non-zero entries are a subset of the non-zero entries of T (and therefore it can be obtained without adding extra links to the original graph). In particular it is also shown that an optimal $(0, 1)$ matrix T^* can be found. Following the procedure outlined in the same paper [28], it can be found that the minimum Kemeny constant of the road network of Figure 2 can be obtained for instance from the graph shown in Figure 15, which corresponds to the well-known concept of *ring road*. In practice, this traffic solution can be obtained from the original network by closing all roads that do not appear explicitly in Figure 15, and by forcing all cars to follow the only available ring path. Although this "optimal" solution circumvents the need of junctions, it clearly has the main drawback that it

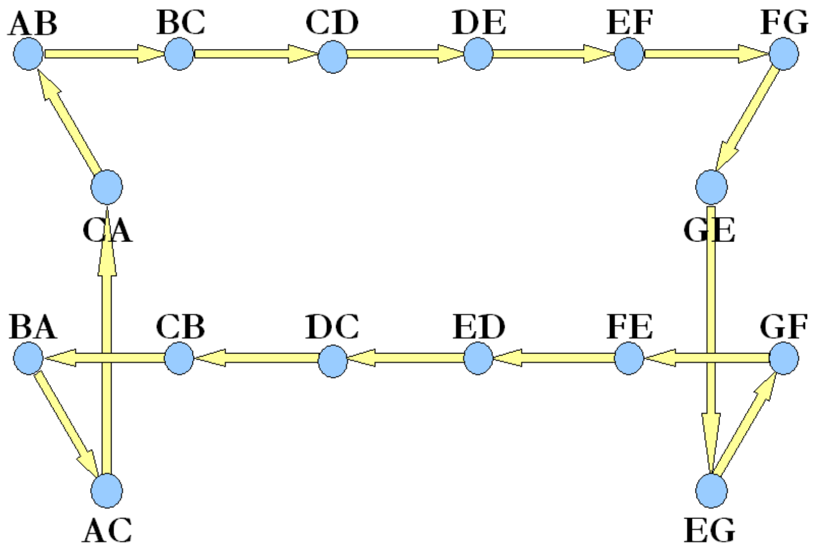

Figure 15: The smallest Kemeny constant is achieved when roads are connected as in a ring road. Although this solution avoids junctions, it might still be required to drive along the whole ring road before reaching the final destination (for instance to go from road BC to AB).

might be required to drive along the whole ring road before reaching the final destination (for instance to go from road BC to AB). We do not discuss further the optimality of this solution, which is only the interpretation of the minimum Kemeny constant road network. We also note that the abstract mathematical solution of the minimum Kemeny constant may not always be applied in practice, as depending on the original graph it might give rise to a non-connected network. Thus, it may be desirable to decrease the Kemeny constant subject to maintaining certain connectivity properties.

Comment: As a conclusion of this section we note that the Kemeny constant can reduce if appropriate roads are closed. Indeed the optimal solution shown in Figure 15 is a subgraph of the original network of Figure 2. It is interesting to remark that this result is in agreement with a well known paradox in road networks, also known as Braess's paradox, which states that adding extra capacity to a network can in some cases reduce overall performance [29, 30].

5.2.2 Decreasing the Kemeny constant

In practice, road network designers are usually interested in finding the simple modifications of a pre-existing road network that have positive effects in mitigating traffic and reducing average travel times. The first problem is that it might not be clear which point of the network would provide the most benefits to the overall traffic flow. We find here the best solution in terms of the Kemeny constant again. The idea is that of computing which small modification in the Markov chain transition matrix $\mathbb{P}$ provides the highest local reduction of the overall Kemeny constant.

Mathematically, this corresponds to evaluating the derivative of the Kemeny constant with respect to small perturbations of entries of the matrix $\mathbb{P}$. Following the same approach of [28] let us check what happens if the p^{th} element of the i^{th} row of matrix $\mathbb{P}$ is decreased of a quantity equal to $t \cdot \mathbb{P}_{ip}$, where $t \in [0,\ 1]$. The original transition matrix $\mathbb{P}$ is perturbed into $\tilde{\mathbb{P}} = \mathbb{P} + E_t$, where $E_t = t\frac{\mathbb{P}_{ip}}{1-\mathbb{P}_{ip}}\boldsymbol{e}_i\left[\boldsymbol{e}_i^T\mathbb{P} - \boldsymbol{e}_p^T\right]$, where $\boldsymbol{e}_k$ is a vector of zeros with a 1 in k^{th} position. As we have that $K(\mathbb{P}+E_t) = trace\left((Q-E_t)^{\#}\right)$ (where we recall that the definition of Q was given in 0.3.1) then the derivative corresponds to

$$\begin{aligned}\frac{\partial K}{\partial t} &= \lim_{t\to 0} \frac{t\frac{\mathbb{P}_{ip}}{1-\mathbb{P}_{ip}}}{1-\left(\boldsymbol{e}_i^T\mathbb{P}-\boldsymbol{e}_p^T\right)Q^{\#}t\frac{\mathbb{P}_{ip}}{1-\mathbb{P}_{ip}}\boldsymbol{e}_i} \cdot \frac{\left(\boldsymbol{e}_i^T\mathbb{P}-\boldsymbol{e}_p^T\right)Q^{\#}Q^{\#}\boldsymbol{e}_i}{t} = \\ &= \frac{\mathbb{P}_{ip}}{1-\mathbb{P}_{ip}}\left(\boldsymbol{e}_i^T\mathbb{P}-\boldsymbol{e}_p^T\right)Q^{\#}Q^{\#}\boldsymbol{e}_i \end{aligned} \tag{16}$$

Such a procedure refers to a particular entry of the matrix $\mathbb{P}$, and it can be computed for all (non-zero) elements of $\mathbb{P}$, thus obtaining a matrix of derivatives with the same size as $\mathbb{P}$. The derivative can be assumed to be zero in positions corresponding to zero elements of $\mathbb{P}$.

The highest negative entries of the derivative matrix indicate that immediate benefits in terms of reductions of average travel times can be obtained if the corresponding entries of $\mathbb{P}$ can be decreased. A non-diagonal entry $\mathbb{P}_{ij}$ can be set to zero easily by closing the connection from road i to road j. In contrast, the diagonal terms also include information regarding the road viability (e.g. the length of the road) and therefore they can not be set to zero. However, they can be decreased by raising speed limits, or by timing traffic lights or changing priority rules appropriately.

5.2.3 Conditioning of the stationary distribution under perturbations of the transition matrix

A valuable property of the Markov chain road network model is that it is possible to predict the effects of modifications of the original road network in terms of road dynamics. This feature is useful for instance to plan road works or to predict the propagation of the traffic density as a consequence of different traffic light timings.

Mathematically this corresponds to predicting changes in the stationary distribution as some entries of the transition matrix are slightly perturbed. This topic has been investigated for instance in references [31] and [32]. Let us denote with $\tilde{\mathbb{P}} = \mathbb{P} + E$ the transition matrix after the perturbation E. For instance, in the example of timing a traffic light, E has all zero rows except for the rows corresponding to the roads involved in the junction of interest. In these roads, the diagonal elements are positive or negative depending on whether their ratio of green period has been decreased or increased. As the matrix $\mathbb{P} + E$ has to remain row-stochastic, elements of E corresponding to off-diagonal non-zero elements of $\mathbb{P}$ have non-zero value as well, and opposite sign with respect to the diagonal entry of E. Then

$$\begin{array}{ll} & \tilde{\pi}^T \mathbb{P} + \tilde{\pi}^T E = \tilde{\pi}^T \Rightarrow \\ \Rightarrow & \tilde{\pi}^T E \left(\boldsymbol{I} - \mathbb{P}\right)^{\#} = \tilde{\pi}^T \left(\boldsymbol{I} - \mathbf{1}\pi^T\right) = \tilde{\pi}^T - \pi^T \Rightarrow \\ \Rightarrow & \pi^T = \tilde{\pi}^T \left[\boldsymbol{I} - E\left(\boldsymbol{I} - \mathbb{P}\right)^{\#}\right] \Rightarrow \\ \Rightarrow & \tilde{\pi}^T = \pi^T \left[\boldsymbol{I} - E\left(\boldsymbol{I} - \mathbb{P}\right)^{\#}\right]^{-1} \end{array}, \qquad (17)$$

where $\mathbf{1}$ is the column vector of ones of appropriate dimensions. In the last passage the inverse exists provided that both $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are irreducible [33].

The ability of predicting the new stationary distribution was tested for the example shown in section 0.5.1 where green times were modified to improve traffic flow. The theoretical results were in accordance with the new stationary distribution (15) extracted from the SUMO simulation.

6 Conclusions

Inspired by the success of Google's PageRank algorithm, a Markov chain approach was proposed to model road network dynamics. The major difference with conventional road traffic models is that all the network information is concentrated inside the transition matrix, which is constructed from collected data, specifically road travel times and junction turning probabilities. The proposed approach circumvents the requirement of modelling complex road dynamics and car interactions, or the necessity of extensive Monte Carlo simulations, but at the same time it still provides an accurate and realistic representation of the road network dynamics.

One of the main objectives of this paper is to validate the proposed model. Extensive simulations over several road networks of different shapes have confirmed the theoretical expectations, and results concerning one particular network were shown in detail throughout the paper as a benchmark example. Although only initial results are reported, they are very promising for two main reasons. The first one is that cars can be easily equipped to start collecting real data to build the Markov transition matrix. From this point of view, validation of the proposed model with real data is expected to be one of the very next steps. The second advantage is that from the mathematical analysis of the Markov chain it is possible to infer hidden properties of the underlying road network which can be hardly revealed even by tailored ad-hoc simulations. In addition it is possible to predict road dynamics, for instance the propagation of the traffic density in consequence of different traffic light timings, or to the closure of a road for road works.

The ability of predicting road dynamics paves the way to a wide variety of applications, some of which have been briefly discussed. The Markov chain model apparently correctly identifies critical links and offers a very simple and intuitive way to optimise traffic light timings. The analysis of mean first passage times and the use of the Kemeny constant serve to quantify the efficiency of the road network. Although it might be argued that the "ring road" can not obviously be always the best solution in practice, still the idea that simple and schematic urban networks should be preferred against tangled and complicated networks is rather intuitive and interesting. Obviously each of the outlined applications should be further investigated, while several other applications can be easily stated within the same framework. This is subject of current study and will be taken into explicit account in future work.

Appendix

6.1 Clustering properties of the second eigenvector

Theorem:
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 $\left[\begin{array}{c|c|c} \mathbb{P}_{11} & \mathbb{P}_{12} & \mathbb{P}_{13} \\ \hline \mathbb{P}_{21} & \mathbb{P}_{22} & \mathbb{P}_{23} \\ \hline \mathbb{P}_{31} & \mathbb{P}_{32} & \mathbb{P}_{33} \end{array}\right]$ and label the subsets of the partition as S_1, S_2 and S_0 respectively. Then:

1. $\rho(\mathbb{P}_{11}), \rho(\mathbb{P}_{22}) \geq \lambda$.

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 $\sum_{i \in \tilde{S}_1} \tilde{w}_1(i) \sum_{j \notin \tilde{S}_1} \mathbb{P}_{ij} = 1 - \rho(\mathbb{P}_{11}) \leq 1 - \lambda$ and $\sum_{i \in \tilde{S}_2} \tilde{w}_2(i) \sum_{j \notin \tilde{S}_2} \mathbb{P}_{ij} = 1 - \rho(\mathbb{P}_{22}) \leq 1 - \lambda$.

3. For any $j \in \tilde{S}_2$, $\sum_{i \in \tilde{S}_1} \tilde{w}_1(i)\, m_{ij} \geq \dfrac{1}{1-\rho(\mathbb{P}_{11})} \geq \dfrac{1}{1-\lambda}$ and for any $j \in \tilde{S}_1$, $\sum_{i \in \tilde{S}_2} \tilde{w}_2(i)\, m_{ij} \geq \dfrac{1}{1-\rho(\mathbb{P}_{22})} \geq \dfrac{1}{1-\lambda}$, 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 $\mathbb{P}_{11}v_1 = \lambda v_1 + \mathbb{P}_{12}v_2$. Let z^T be a Perron vector for $\mathbb{P}_{11}$. Then $\rho(\mathbb{P}_{11})\, z^T v_1 = z^T \mathbb{P}_{11} v_1 = \lambda z^T v_1 + z^T \mathbb{P}_{12} v_2 \geq \lambda z^T v_1$. The inequality follows because all terms are positive (either because Perron vectors or because parts of the nonnegative stochastic matrix $\mathbb{P}$). Therefore, comparing the first and last term of the chain of inequalities, we obtain $\rho(\mathbb{P}_{11}) \geq \lambda$. An analogous argument applies for $\rho(\mathbb{P}_{22})$.

2. Let w_1^T be a left Perron vector for $\mathbb{P}_{11}$, 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 $\mathbb{P}_{22}$. Let us write $\mathbb{P}$ in partitioned form as

$$\left[\begin{array}{c|c|c|c|c}
\mathbb{P}_{\tilde{S}_1\tilde{S}_1} & \mathbb{P}_{\tilde{S}_1\bar{S}_1} & \mathbb{P}_{\tilde{S}_1\tilde{S}_2} & \mathbb{P}_{\tilde{S}_1\bar{S}_2} & \mathbb{P}_{\tilde{S}_1 S_0} \\ \hline
\mathbb{P}_{\bar{S}_1\tilde{S}_1} & \mathbb{P}_{\bar{S}_1\bar{S}_1} & \mathbb{P}_{\bar{S}_1\tilde{S}_2} & \mathbb{P}_{\bar{S}_1\bar{S}_2} & \mathbb{P}_{\bar{S}_1 S_0} \\ \hline
\mathbb{P}_{\tilde{S}_2\tilde{S}_1} & \mathbb{P}_{\tilde{S}_2\bar{S}_1} & \mathbb{P}_{\tilde{S}_2\tilde{S}_2} & \mathbb{P}_{\tilde{S}_2\bar{S}_2} & \mathbb{P}_{\tilde{S}_2 S_0} \\ \hline
\mathbb{P}_{\bar{S}_2\tilde{S}_1} & \mathbb{P}_{\bar{S}_2\bar{S}_1} & \mathbb{P}_{\bar{S}_2\tilde{S}_2} & \mathbb{P}_{\bar{S}_2\bar{S}_2} & \mathbb{P}_{\bar{S}_2 S_0} \\ \hline
\mathbb{P}_{S_0\tilde{S}_1} & \mathbb{P}_{S_0\bar{S}_1} & \mathbb{P}_{S_0\tilde{S}_2} & \mathbb{P}_{S_0\bar{S}_2} & \mathbb{P}_{S_0 S_0}
\end{array}\right]. \tag{18}$$

We have

$$\begin{aligned} 1 &= \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\tilde{S}_1}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\bar{S}_1}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\tilde{S}_2}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\bar{S}_2}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1 S_0}\mathbf{1} = \\ &= \rho(\mathbb{P}_{11}) + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\bar{S}_1}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\tilde{S}_2}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\bar{S}_2}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1 S_0}\mathbf{1} \end{aligned}$$

so that $1-\lambda \geq 1-\rho(\mathbb{P}_{11}) = \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\bar{S}_1}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\tilde{S}_2}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1\bar{S}_2}\mathbf{1} + \tilde{w}_1^T \mathbb{P}_{\tilde{S}_1 S_0}\mathbf{1}$ and the desired inequality follows. An analogous argument applies to $(\mathbb{P}_{22})$.

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

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

The last inequality follows as $\mathbb{P}_{\tilde{S}_1 \tilde{S}_1}$ is a submatrix of $\mathbb{P}_{(j)}$. Hence, $\sum_{i \in \tilde{S}_1} \tilde{w}_1(i)\, m_{ij} \geq \tilde{w}_1^T \left(\boldsymbol{I} - \mathbb{P}_{\tilde{S}_1 \tilde{S}_1}\right)^{-1} \mathbf{1} = \frac{1}{1 - \rho(\mathbb{P}_{11})} \geq \frac{1}{1-\lambda}$. A similar argument establishes the desired inequality for $j \in \tilde{S}_1$.

6.2 Motivation of equation (6)

Let us assume that all roads can be covered in the same time, therefore all diagonal elements of the matrix $\mathbb{P}$ are zero and we have $\pi^T \mathbb{P} = \pi^T$, where π^T is the left Perron eigenvector. As in reality all roads have different travel times, self loops will now be added accordingly.
Let us assume that the diagonal entry in the i^{th} position is $\mathbb{P}_{ii}$, then the probability of leaving the road in exactly j steps is $\mathbb{P}_{ii}^{j-1}(1 - \mathbb{P}_{ii})$, thus the expected number of steps before leaving the road is

$$\sum_{i}^{\infty} j \mathbb{P}_{ii}^{j-1}(1 - \mathbb{P}_{ii}) = \frac{1}{1 - \mathbb{P}_{ii}}. \tag{19}$$

As described in section 0.4.1, we assume that travel times for each single road are available; if tt_i is the observed travel time for road i, and travel times are normalised so that $\min\{tt_i\} = 1, i = 1, \ldots, n$, then we have $\frac{1}{1 - \hat{\mathbb{P}}_{ii}} = tt_i$ and equation (6) follows. $\hat{\mathbb{P}}_{ii}$ represents the updated entry of the transition matrix required to take different travel times into account.

Once diagonal terms are changed, off-diagonal terms have to be updated to keep the transition matrix row-stochastic, without affecting turning probabilities ratios. That is

$$\hat{\mathbb{P}}_{ij} = \left(1 - \hat{\mathbb{P}}_{ii}\right) \mathbb{P}_{ij}, \tag{20}$$

where $\hat{\mathbb{P}}_{ij}$ are the updated off-diagonal terms of the transition matrix. It is interesting to notice that the (non-normalised) Perron eigenvector $\hat{\pi}^T$ associated to matrix $\hat{\mathbb{P}}$ is related to π^T of matrix $\mathbb{P}$, where travel times had not been taken into account, through $\hat{\pi}^T = [tt_1\pi_1 \quad tt_2\pi_2 \quad \ldots \quad tt_n\pi_n]$.

References

[1] Barth, M., and Boriboonsomsin, K. (2009), "Traffic congestion and greenhouse gases," *Access: the magazine of the UC transportation center*, 35, 2–9.

[2] Dirks, S., Gurdgiev, C., and Keeling, M., (2010), "Smarter cities for smarter growth, How cities can optimize their systems for the talent-based economy," *IBM Global Business Services Executive Report*, available on ftp://public.dhe.ibm.com/common/ssi/pm/xb/n/gbe03348usen/GBE03348USEN.PDF.

[3] Villa, N., and Mitchell, S., (2009), "Connecting cities: achieving sustainability through innovation," *Fifth Urban Research Symposium 2009: Cities and Climate Change: Responding to an Urgent Agenda*, Marseille, France.

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

[5] Fiore, M., and Härri, J. (2008), "The networking shape of vehicular mobility," *ACM MobiHoc*, Hong Kong, China, May 2008.

[6] Hartenstein H., and Laberteaux, K. Eds., (2010), "VANET - Vehicular applications and inter-networking technologies," Wiley.

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

[8] Meyn, S.P., and Tweedie, R.L., (2009), "Markov Chains and stochastic stability," *Cambridge University Press.*

[9] Papoulis, A., and Pillai, S.U., (2002), "Probability, random variables and stochastic processes," *McGraw-Hill Higher Education.*

[10] Dellnitz, M., and Junge, O., (1999), "On the approximation of complicated dynamical behaviour," *SIAM Journal on Numerical Analysis*, vol. 36, no. 2, pp. 491–515.

[11] Froyland, G., (2001), "Extracting dynamical behaviour via Markov models," *In Alistar Mees, editor, Nonlinear Dynamics and Statistics: Proceedings, Newton Institute, Cambridge*, Birkhauser, pp. 283–324.

[12] Huisinga, W., Meyn, S.P., and Schütte,C., (2004), "Phase transition and metastability in Markovian and molecular systems," *Ann. Appl. Probab.*, vol. 14, no. 1, pp. 419–458.

[13] Meyer, C.D., (1975), "The role of group generalized inverse in the theory of finite Markov Chains," *SIAM Review*, vol. 17, no. 3, pp. 443–464.

[14] 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, pp. 137–150.

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

[16] Doyle, P., (2009), "The Kemeny constant of a Markov chain," preprint, available on http://math.dartmouth.edu/∼doyle/docs/kc/kc.pdf.

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

[18] von Luxburg, U., (2007), "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4.

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

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

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

[22] 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*, vol. 26, pp. 169–191.

[23] Jiang, B., (2008), "Ranking spaces for predicting human movement in an urban environment," *International Journal of Geographical Information Science*.

[24] 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*.

[25] Jiang, B., and Liu, C., (2009), "Street-based topological representations and analyses for predicting traffic flows in GIS," *International Journal of Geographical Information Science*.

[26] Dijkstra, E.W., (1959), "A note of two problems in connexion with graphs,", *Numer. Math. 1*, pp. 269–271.

[27] Bertsekas, D.P.(2000), "Dynamic programming and optimal control,", *Athena Scientific*, 2nd edition.

[28] Kirkland, S., (2010), "Fastest expected time to mixing for a Markov chain on a directed graph,", *Linear Algebra and its Applications*, to appear.

[29] Braess, D., (1968), "Über ein Paradoxon aus der Verkehrsplanung," *Unternehmensforschung*, vol. 12, pp. 258–268.

[30] Braess, D., Nagurney, A., and Wakolbinger, T., (2005), "On a paradox of traffic planning," *Transportation Science*, vol. 39, pp. 446–450 (translation from the original German).

[31] Kirkland, S., (2003), "Conditioning properties of the stationary distribution for a Markov chain," *Electronic Journal of Linear Algebra*, vol. 10, pp. 1–15.

[32] Kirkland, S., (2004), "A combinatorial approach to the conditioning of a single entry in the stationary distribution for a Markov chain," *Electronic Journal of Linear Algebra*, vol. 11, pp. 168–179.

[33] Meyer, C.D., (1980), "The condition of a finite Markov chain and perturbation bounds for the limiting probabilities," *SIAM Journ. Alg. Disc. Meth.*, vol. 1, no. 3, pp. 273–283.