International Journal of Control
Vol. 00, No. 00, Month 200x, 1–19

Research Article

A big-data model for multi-modal public transportation with application to macroscopic control and optimisation

Mahsa Faizrahnemoon[a], Arieh Schlote[a], Lorenzo Maggi[b], Emanuele Crisostomi[c] *, and Robert Shorten[d]

[a] *Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare, Ireland*;
[b] *CREATE-NET, Trento, Italy*;
[c] *Department of Energy, Systems, Territory and Constructions Engineering, University of Pisa, Italy*;
[d] *IBM Research, Dublin, Ireland*

(*Received 00 Month 200x; final version received 00 Month 200x*)

This paper describes a Markov chain based approach to modelling multi-modal transportation networks. An advantage of the model is the ability to accommodate complex dynamics and handle huge amounts of data. The transition matrix of the Markov chain is built and the model is validated using data extracted from a traffic simulator. 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. Then, we use the Markov chain as a control tool to improve the overall efficiency of a transportation network, and some practical examples are described to illustrate the potentials of the approach.

Keywords: Markov chains; Transportation models; big-data models;

1 Introduction

1.1 *Motivation*

Recently, many papers have been published on modelling and optimizing public transportation networks. Motivated, in principle, by the availability of large amounts of data in real time, authors have focussed, for the most part, on problems such as: journey planning for individuals and the associated *optimization under uncertainty* problem [Hame and Hakula (2013)]; scheduling of public transport systems, and more recently, *on demand* public transport applications [Tsubochi et al. (2010)]; and niche topics such as bus-bunching [Bartholdi and Eisenstein (2012)]. Despite this intense interest many open questions remain to be resolved. These include: the development of simple tractable mathematical models that can accommodate the huge volume of real time data that is now available to public transport operators (GPS, mobile phone data, loop detectors, cameras, etc.); the development of mathematical models that allow to incorporate different modes of transport (walking, bikes, buses, taxis, trains, private cars, etc.); and the development of models that allow the extraction of key macroscopic design criteria.

A first important requirement is this latter objective of developing models that capture macroscopic properties of the network. Much of the current work in this area focusses on *quality of service metrics* as applied to a given individual. This latter aspect assumes that individuals operate essentially in a bath of noise, and ignores coupling between individuals.

*Corresponding author. Email: emanuele.crisostomi@unipi.it

ISSN: 0020-7179 print/ISSN 1366-5820 online

DOI: 10.1080/0020717YYxxxxxxxx
http://www.informaworld.com

For example, in multi-modal journey planning, every journey has the possibility of affecting every other journey. For example, if enough individuals are recommended a particular bus-leg, then the bus may be full, thereby invalidating other plans already calculated for other network users. Our objective in this paper is to develop models that do not focus on the individual, but rather focus on aggregated behavior. We are specifically interested in answering high-level questions that pertain to how accessible certain key spots in the city are (for example, hospitals), and how easy it is to travel, on average, from one part of a city to another. Such questions, we believe, are important characterizations of urban dynamics.

A second important requirement is that a useful model should have the ability to support transportation engineers in implementing control actions aiming at improving the efficiency of a transportation network. In this perspective, a model should be able to proactively predict the effect of some control actions (e.g., removing a bus stop, increasing the frequency of a train, adding an extra bus line) in an existing public transportation network. Given such a basic requirement, it is of paramount importance to adopt a model that can be constructed from (a big quantity of) real data obtained directly from the transportation network in near real time. For this purpose, our starting point is a recently proposed Markov-chain based framework for capturing macroscopic urban dynamics [Crisostomi et al. (2011a)].

1.2 *State of the art and related work*

In [Crisostomi et al. (2011a)], graph theory and Markov chain ideas were used to reveal non-trivial patterns of urban mobility and to support engineers with practical tools to solve a number of mobility applications; namely, routing, traffic light regulation, and road planning. The work of [Crisostomi et al. (2011a)] was well received by the transport community giving rise to further research along the same lines, see [Crisostomi et al. (2011b)], [Schlote et al. (2012)], [Jensen et al. (2014)], [Bekkerman et al. (2013)] and [Morimura et al. (2013)]. Also, the findings of [Crisostomi et al. (2011a)] were further validated using real data of Beijing [Moosavi and Hovestadt (2013)].

The objectives of this paper are twofold: (i) first, we want to illustrate how the approach of [Crisostomi et al. (2011a)] 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 paper. We later show how the same strategy can be applied to a general multi-modal transportation network, through a more realistic example pertaining 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. A preliminary 2-page draft paper outlining the proposed methodology had been first shown in [Faizrahnemoon et al. (2013)].

This paper is organised as follows: Section 2 reviews the basic notions of Markov chain theory that will be used in the remainder of the paper. Section 2.3 illustrates how to use collected data to build the Markov chain model. Section 4 validates the model using the mobility simulator SUMO (Simulator of Urban MObility) [Krajzewicz et al. (2012)], 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 5. Section 6 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. Finally, Section 7 concludes the paper and outlines current and future lines of research.

2 A primer on Markov chains

Markov chains are a standard tool for engineers and applied mathematicians and many of their properties can be found in classic references like [Kemeny and Snell (1960), Langville and Meyer (2006)]. We repeat here a discussion from [Crisostomi et al. (2011a)] that introduces some basic definitions and standard results that are needed for our discussion. Throughout this paper only discrete-time, finite-state, homogeneous Markov chains will be considered. In this situation, the Markov chain is a discrete time stochastic process $x_k, k \in \mathrm{N}$ and characterised by the equation

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

where $p(E|F)$ denotes the conditional probability that event E occurs given that event F occurs.

A Markov chain with n states is completely described by the $n \times n$ transition probability matrix P, whose entry P_{ij} denotes the probability of passing from state S_i to state S_j in exactly one step. P is a row-stochastic non-negative matrix, as the elements in each row are probabilities and they sum up to 1. Within Markov chain theory, there is a close relationship between the transition matrix P and a corresponding graph. The graph consists of a set of nodes that are connected through edges. The graph associated with the matrix P is a directed graph, whose nodes are given by the states $S_i, i = 1, ..., n$, and there is a directed edge leading from S_i to S_j if and only if $\mathrm{P}_{ij} \neq 0$. 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 P is irreducible if and only if its directed graph is strongly connected. Some important properties of irreducible transition matrices follow from the well-known Perron-Frobenius theorem [Langville and Meyer (2006)]:

- The spectral radius of P is 1; 1 also belongs to the spectrum of P, and has an algebraic multiplicity of 1;
- The left-hand Perron eigenvector π is the unique vector defined by $\pi^T \mathrm{P} = \pi^T$, 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 P.

One of the main properties of irreducible Markov chains 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.

We now discuss three properties of Markov chains that render them suitable for modelling large scale urban systems.

2.1 *Mean first passage times and the Kemeny constant*

A transition matrix P with 1 as a simple eigenvalue gives rise to a singular matrix $\boldsymbol{I} - \mathrm{P}$ (where the identity matrix $\boldsymbol{I}$ has appropriate dimensions) which is known to have a group inverse $(\boldsymbol{I} - \mathrm{P})^{\#}$. The group inverse is the unique matrix such that $(\boldsymbol{I} - \mathrm{P})(\boldsymbol{I} - \mathrm{P})^{\#} = (\boldsymbol{I} - \mathrm{P})^{\#}(\boldsymbol{I} - \mathrm{P})$, $(\boldsymbol{I} - \mathrm{P})(\boldsymbol{I} - \mathrm{P})^{\#}(\boldsymbol{I} - \mathrm{P}) = (\boldsymbol{I} - \mathrm{P})$, and $(\boldsymbol{I} - \mathrm{P})^{\#}(\boldsymbol{I} - \mathrm{P})(\boldsymbol{I} - \mathrm{P})^{\#} = (\boldsymbol{I} - \mathrm{P})^{\#}$. More properties of group inverses and their applications to Markov chains can be found in [Meyer (1975(@)]. The group inverse $(\boldsymbol{I} - \mathrm{P})^{\#}$ contains important information on the Markov chain and it will be often used in this paper. 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 times can be computed according to [Cho and Meyer

(2001)]

$$m_{ij} = \frac{q^{\#}_{jj} - q^{\#}_{ij}}{\pi_j} \qquad i, j = 1, ..., n, i \neq j. \tag{2}$$

We assume that $m_{ii} = 0$. The Kemeny constant is defined as

$$K = \sum_{j=1}^{n} m_{ij}\pi_j \tag{3}$$

where the right-hand side is independent of the choice of the origin state S_i [Kemeny and Snell (1960)]. 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 [Doyle (2009)]. Therefore, the Kemeny constant is an intrinsic measure of a Markov chain, and if the transition matrix P has eigenvalues $\lambda_1 = 1, \lambda_2, ..., \lambda_n$, then another way of computing K is [Levene and Loizou (2002)]

$$K = \sum_{j=2}^{n} \frac{1}{1 - \lambda_j}. \tag{4}$$

As can be seen from Equation (4), K is only related to the particular matrix P and it increases if one or more eigenvalues of P get close to 1.

2.2 *Clustering and the second eigenvector*

In this section we discuss another Markov chain characteristic, that we will use to investigate clusters in the bus network. It is well known that the eigenvectors of transition matrices for *undirected* graphs have good clustering properties, see for instance [Luxburg (2007)]. In [Crisostomi et al. (2011a)] it was further shown that the sign pattern of an eigenvector associated with an eigenvalue close to 1 can be also used to identify two different clusters. Since an irreducible transition matrix has only one eigenvalue equal to 1, such an eigenvector is called the second eigenvector (as it is associated with the eigenvalue of second largest modulus). However, the justification of the second eigenvector given in [Crisostomi et al. (2011a)] held under the assumption that it is real. In some cases, for instance when cyclic behavior occurs, (which could frequently occur if one thinks to typical bus routes), see Theorem 4.15 in [Huisinga (2003)], it is known that the eigenvalue of second largest modulus is actually a pair of complex conjugated eigenvalues. Accordingly, we extend below the justification initially given in [Crisostomi et al. (2011a)] to such a circumstance when we have two complex eigenvalues.

The rationale behind the clustering properties of the second eigenvector, which will be later used in Sections 4 and 6 is now anticipated through an illustrative example. Suppose that we have three irreducible stochastic matrices P_1, P_2 and P_3 of order k_1, k_2 and k_3 respectively. Assume that the last column of each of the three matrices is positive. Consider the transition matrix

$$A = \left[\begin{array}{c|c|c} P_1 & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & P_2 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & P_3 \end{array}\right];$$

note that A has 1 as an eigenvalue of multiplicity three (as we have diagonally combined three matrices that have 1 has eigenvalue, see Section 2). Suppose now that we perturb A slightly to obtain the matrix

$$\mathrm{B} = \mathrm{A} + \epsilon \left[\begin{array}{c|c|c} -\mathbf{1}_{k_1}\boldsymbol{e}_{k_1,k_1}^T & \mathbf{1}_{k_1}\boldsymbol{e}_{k_2,1}^T & \mathbf{0}_{k_1\times k_3} \\ \hline \mathbf{0}_{k_2\times k_1} & -\mathbf{1}_{k_2}\boldsymbol{e}_{k_2,k_2}^T & \mathbf{1}_{k_2}\boldsymbol{e}_{k_3,1}^T \\ \hline \mathbf{1}_{k_3}\boldsymbol{e}_{k_1,1}^T & \mathbf{0}_{k_3\times k_2} & -\mathbf{1}_{k_3}\boldsymbol{e}_{k_3,k_3}^T \end{array}\right],$$

where $\mathbf{1}_m$ represents an all ones vector of order m, $\mathbf{0}_{i\times l}$ represents an $i \times l$ matrix of all zeros, $\boldsymbol{e}_{i,l}$ is a vector of zeros of length i with a 1 in the l'th position, and ϵ is a small positive number. Positivity of the last column of $\mathrm{P}_i, i = 1, 2, 3$ guarantees that the perturbed matrix does not have negative entries for small ϵ. It is easily verified that B is a stochastic matrix which has complex eigenvalues $1 - 3/2\epsilon \pm j\sqrt{3}/2\epsilon$, with corresponding right eigenvector $\left[\begin{array}{c} \mathbf{1}_{k_1} \\ \hline \mathbf{1}_{k_2}\cdot(-0.5 \pm j\sqrt{3}/2) \\ \hline \mathbf{1}_{k_3}\cdot(-0.5 \mp j\sqrt{3}/2) \end{array}\right]$, where the eigenvectors are independent of the (small) value of ϵ.

In practice, the previous plausibility argument simply states that if we have three quasi-disconnected graphs, then we might have two complex eigenvalues of modulus very close to 1, whose corresponding eigenvectors have three main well-separated clusters of entries if plotted in the complex plane (note that either second eigenvector can be used as they are complex conjugates of each other). Thus, the second eigenvector can be used to identify clusters in the graph.

2.3 *Markov chains and big-data*

Markov chains are particularly suited to big-data applications for several reasons: *(i)* Microscopic behavior 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 MFPT matrices) of a Markov chain can be calculated in a recursive fashion using simple update formulae [Langville and Meyer (2006)]. The suitability of Markov chains for big-data application is discussed, for example, in the context of Google's PageRank algorithm [Langville and Meyer (2006)]. 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 [Schlote (2014)] and [Froyland (2001)].

To gain insight into fast recalculation of Markov chain quantities for changed data, regard the following theorem from [Langville and Meyer (2006)] addressing row updates to Markov chain transition matrices and their effect on the stationary distributions.

Theorem 2.1: *Let* P *and* $\tilde{\mathrm{P}}$ *be irreducible* $n \times n$ *Markov chain transition matrices that satisfy the relationship* $\tilde{\mathrm{P}} - \mathrm{P} = -e_i\delta^\top$, *where* e_i *is a vector of zeros of length* n *with a 1 in the* i*'th position and* $\delta \in \mathrm{R}^n$. *Let* π *and* $\tilde{\pi}$ *be the respective left Perron eigenvectors of* P *and* $\tilde{\mathrm{P}}$. *Let* $Q = (\boldsymbol{I} - \mathrm{P})$ *and let* $Q^\#$ *be its group inverse. Then*

$$\tilde{\pi}^\top = \pi^\top - \epsilon^\top, \tag{5}$$

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

$$(\boldsymbol{I} - \tilde{\mathrm{P}})^\# = Q^\# + \mathbf{1}\epsilon^\top (Q^\# - \frac{\epsilon^\top Q^\# e_i}{\pi_i}\boldsymbol{I}) - \frac{Q^\# e_i \epsilon^\top}{\pi_i}. \tag{6}$$

The above theorem allows 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 $(\boldsymbol{I} - \mathrm{P})$. It also allows 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 [Schlote et al. (2012)].

Theorem 2.2: *Let* P *and* $\tilde{\mathrm{P}}$ *be irreducible* $n \times n$ *Markov chain transition matrices such that* $\tilde{\mathrm{P}}$ *is obtained from* P *by multiplying the* i'*th diagonal entry with a factor* $w_i > 0$ *for each* $i = 1 \ldots, n$ *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* P *and* $\tilde{\mathrm{P}}$. *Then*

$$\tilde{\pi} = \kappa W \pi, \tag{7}$$

where $W = diag(w_1, \ldots, w_n)$ *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 that changing the diagonal entries corresponds to changing the public transport service frequency.

Figure 1. Area of Dublin city center used for analysis and simulations.

3 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 1. The blue icons in Figure 1 correspond to bus stops. We can depict them as in the graph in Figure 2 where consecutive bus stops are connected through an edge yielding 17 nodes connected by 21 edges.

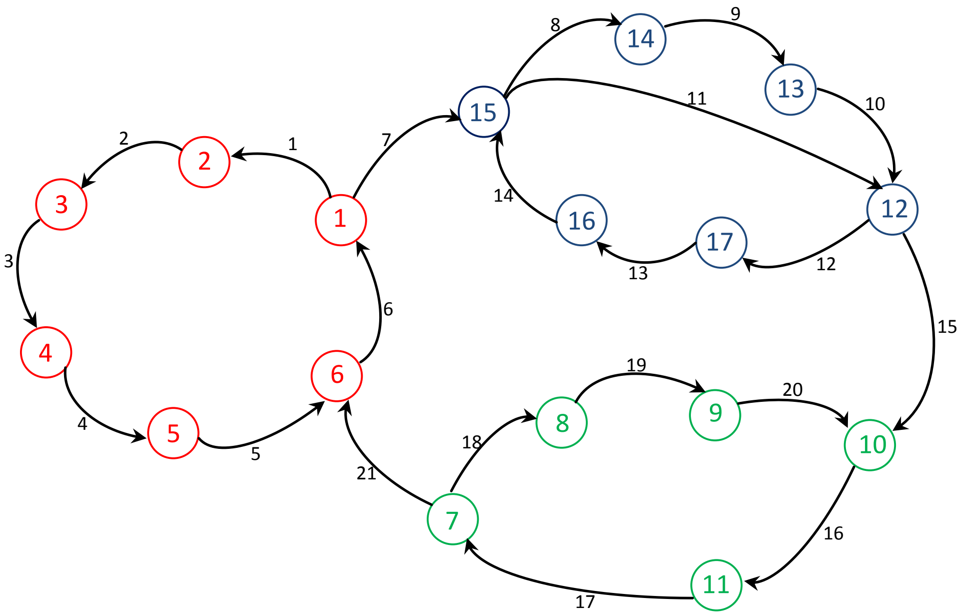


Figure 2. Graph of the bus network.

3.1 *Waiting graph*

A Markov chain transition matrix corresponding to the graph shown in Figure 2 can easily be constructed from collected data according to the procedure outlined below:
(i) Each diagonal entry P_{ii} of the transition matrix is computed as $\mathrm{P}_{ii} = (t_i - 1)/t_i$, 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 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 2), then $\mathrm{P}_{ij} = 0$.
(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 $\mathrm{P}_{i,n+1}$ takes into account the proportion of people who leave the network after having reached bus stop i; similarly, entry $\mathrm{P}_{n+1,i}$ denotes the proportion of travellers who start their journey from the i'th bus stop; finally, we set the diagonal entry as $\mathrm{P}_{n+1,n+1} = (t_{n+1} - 1)/t_{n+1}$, 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 P as previously described. Then,

we scale the off-diagonal entries in order to make matrix 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 [Crisostomi et al. (2011a)] in the case of vehicular density, and is validated through SUMO simulations in Section 4. 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 6. 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 *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 2, the new graph is shown in Figure 3. Accordingly, note that the new nodes in Figure 3 correspond to the edges in the previous Figure 2. This graph is sometimes denoted as the dual of the previous one [Porta et al. (2006)]. A Markov chain transition matrix corresponding to the graph shown in Figure 3 can

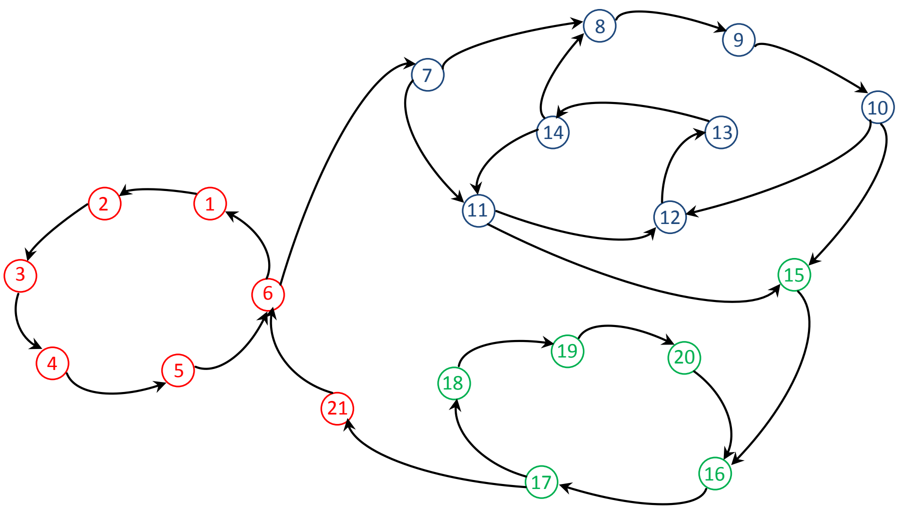


Figure 3. Dual graph of the bus network shown in Figure 2.

be easily constructed from collected data, according to the procedure outlined below. We shall denote such a second travel transition matrix as $\mathrm{P}^{(t)}$ for clarity.

(i) Each diagonal entry of the transition matrix $\mathrm{P}_{ii}^{(t)}$ is computed again as $\mathrm{P}_{ii}^{(t)} = (t_i - 1)/t_i$ 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 bus stop.
(ii) The value of each off-diagonal entry $\mathrm{P}_{ij}^{(t)}$ 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 2 and 3), then $\mathrm{P}_{ij}^{(t)} = 0$.
(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 $\mathrm{P}_{i,n+1}^{(t)}$ 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 $\mathrm{P}_{n+1,i}^{(t)}$ denotes the proportion of travellers who start their journey form the i'th bus connection; finally, we set $\mathrm{P}_{n+1,n+1}^{(t)} = (t_{n+1} - 1)/t_{n+1}$, 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 $\mathrm{P}^{(t)}$ as previously described. Then, we scale the off-diagonal entries in order to make matrix 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 4. 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. Note also that the transition matrices are primitive because they are irreducible and have at least one positive diagonal element by construction [Langville and Meyer (2006)].

3.3 *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 6.

4 Validation

4.1 *Simulation*

We validate our approach simulating the bus network in the small area of Dublin city center shown in Figure 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 2 and 3. We use SUMO, a popular open source traffic simulation software [Krajzewicz et al. (2012)], 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.

4.2 *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 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 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.

4.3 *Clusters*

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 2, we say that three clusters exist if, for instance, 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.

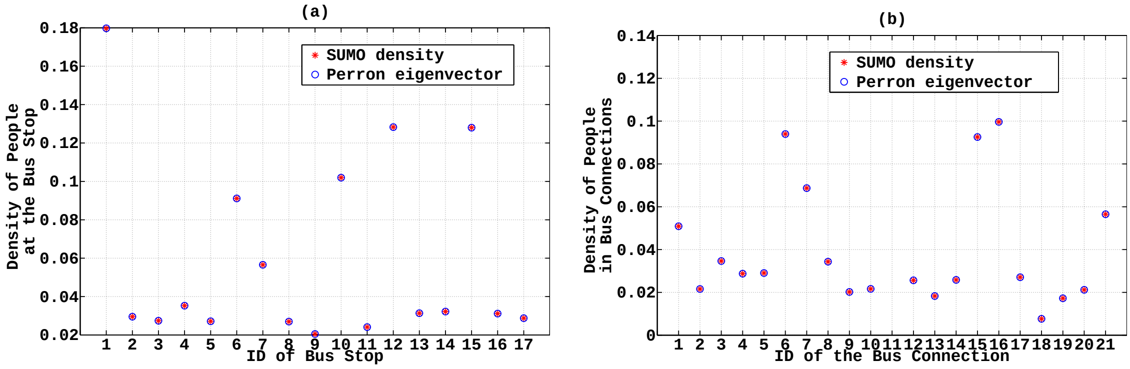


Figure 4. The densities of people at bus stops (a), and at bus connections (b) computed from the Perron eigenvector of the Markov chain are the same of those computed in SUMO.

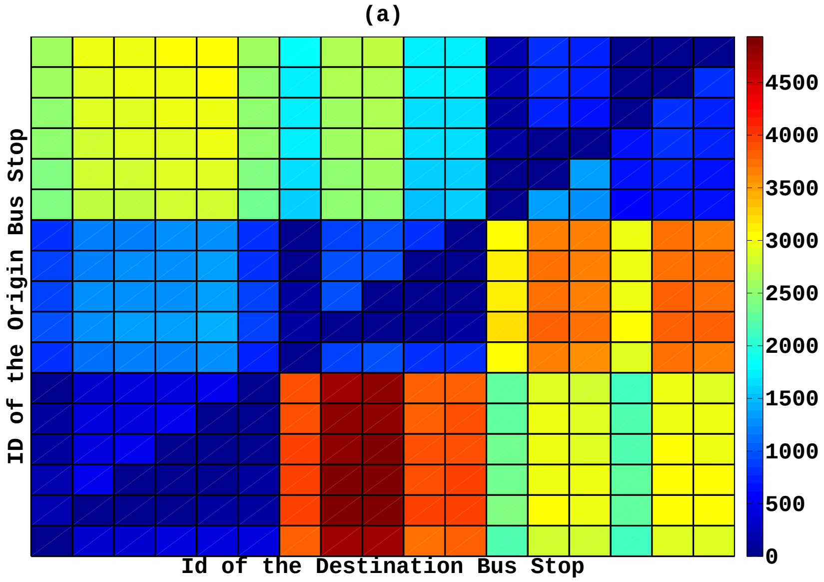
If clusters exist, then the MFPTs should be very low among nodes belonging to the same clusters, and high if the origin and the destination belong to two different clusters. This simply follows from the fact that random walks are more likely to occur within the same set than from one set to another one. Also, if clusters exist, it should be possible to identify them using the second eigenvector as claimed in Section 2.2. To simulate such a situation, we assumed that 90% of the people would indeed choose their destination among one of the nodes belonging to the same cluster of origin. As can be seen from Figure 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 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 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 same information in another way, either from simulation results or even from collecting real data.

5 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 6. 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 6. We then assumed that people would appear at bus/tube stations according to a Poisson process whose average frequency of arrival was chosen consistently with the data provided by TFL[1], and summarised in Table 1. Similarly, the destination of the passengers was chosen proportionally to the exit data reported in Table 1. 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 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 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

[1] `https://www.tfl.gov.uk/`



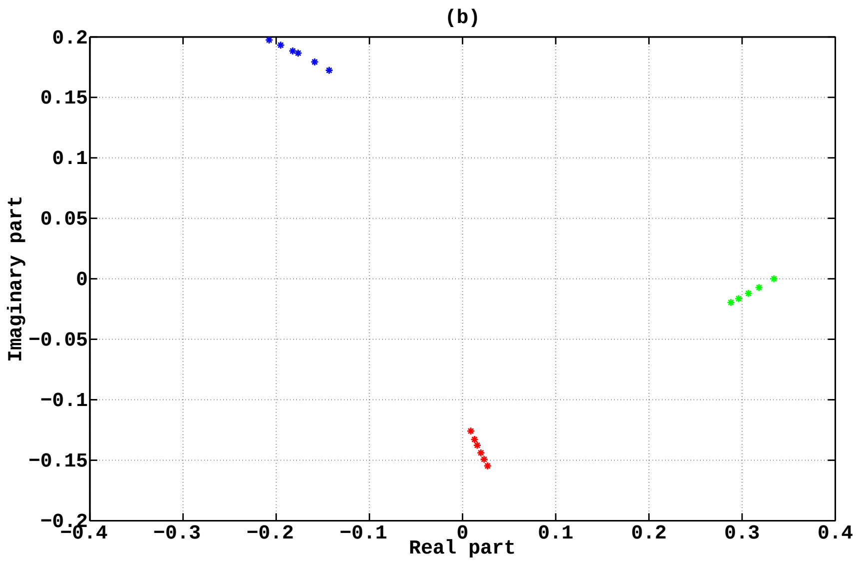


Figure 5. The MFPT matrix and (especially) the second eigenvector are useful in identifying clusters. MFPTs are low among nodes belonging to the same clusters (a); similarly entries of the second eigenvector do identify what bus stops belong to what cluster (b).

the sensors measuring people information are noiseless. Note that from Figure 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 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 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.

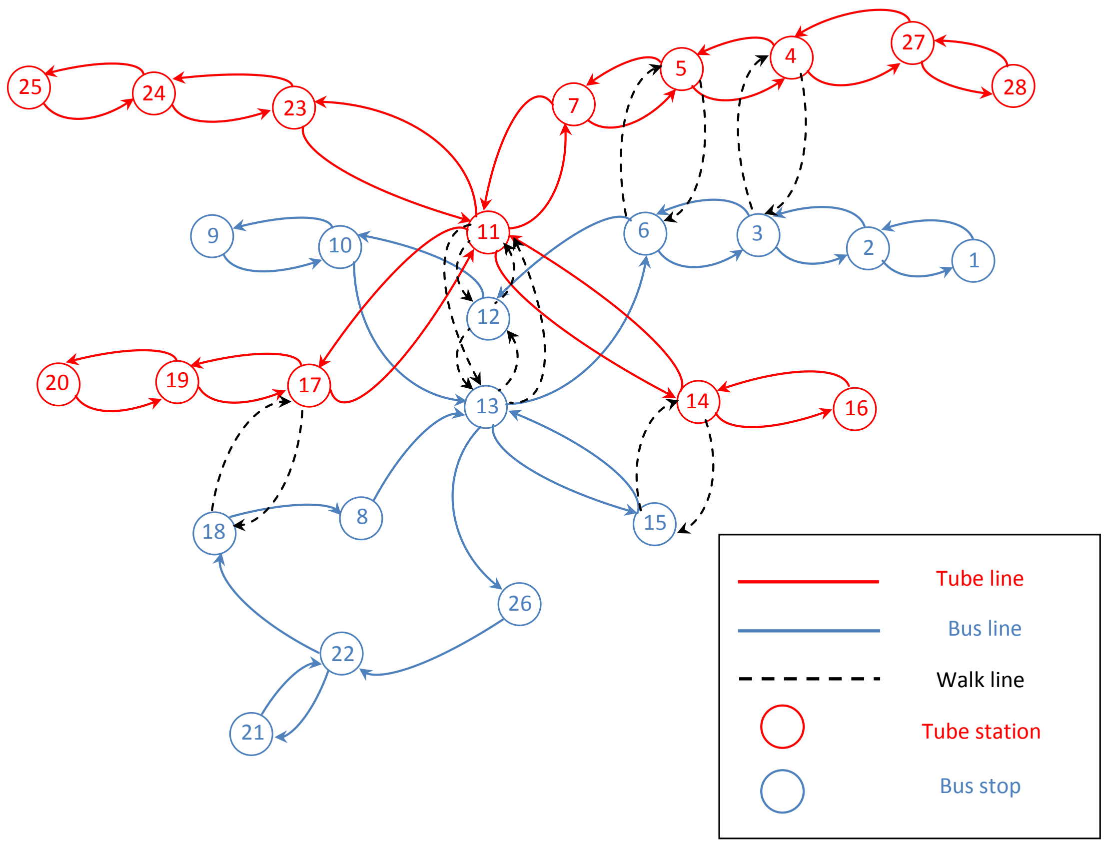


Figure 6. Subgraph of London transportation network, consisting of two bus lines and two tube lines, used for a more realistic validation of the proposed methodology.

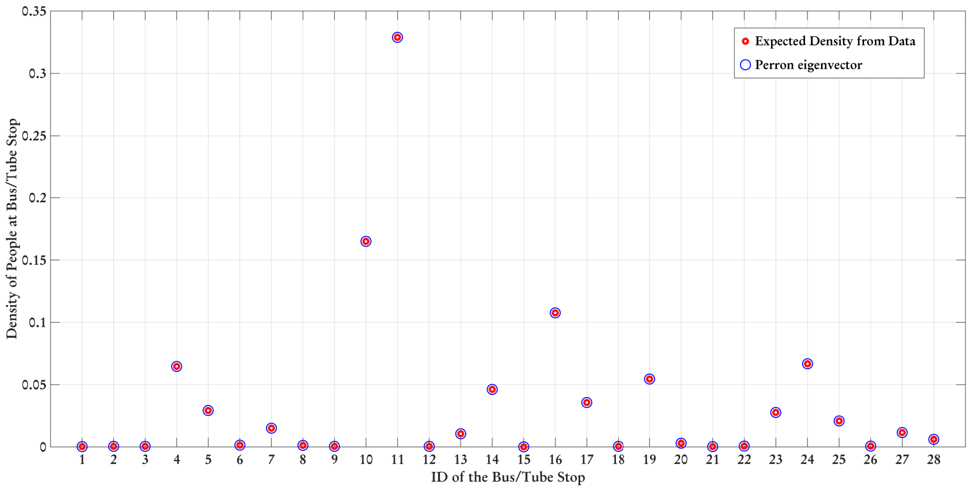


Figure 7. Data from the simulation match data from the Markov chain model.

6 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:

Table 1. 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*	107	625	22	*Swiss Cottage Cheese*	348	256
9	*Notting Hill gate/Palace Garden Terrace*	200	200	23	Green Park	51766	58893
10	*Marble arch Station*	352	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

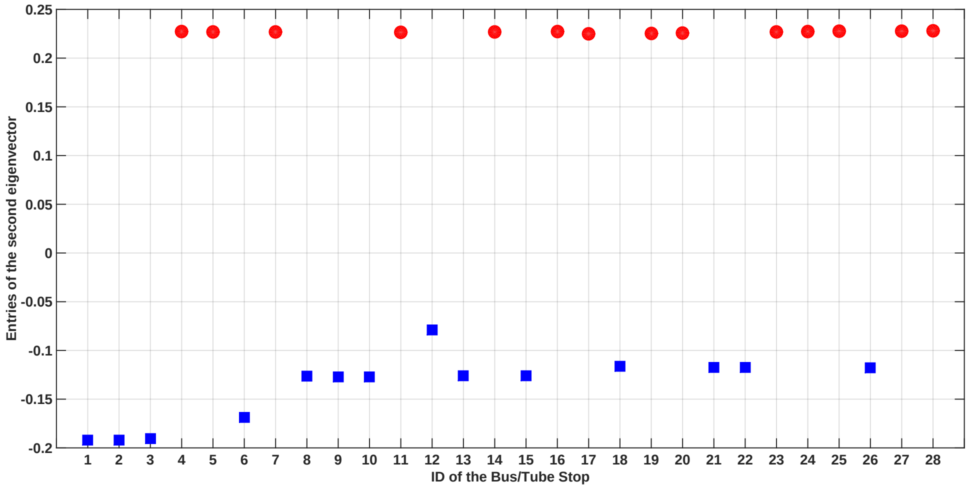


Figure 8. The entries of the second (real) eigenvector distinguish tube stops (positive red circles) and bus stops (negative blue squares).

(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.

6.1 *Node maintenance and control in the transportation network*

The Kemeny constant illustrated in Section 2.1 is known to be a global indicator of the efficiency of a network [Crisostomi et al. (2011a), Moosavi and Hovestadt (2013)]. 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 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 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.

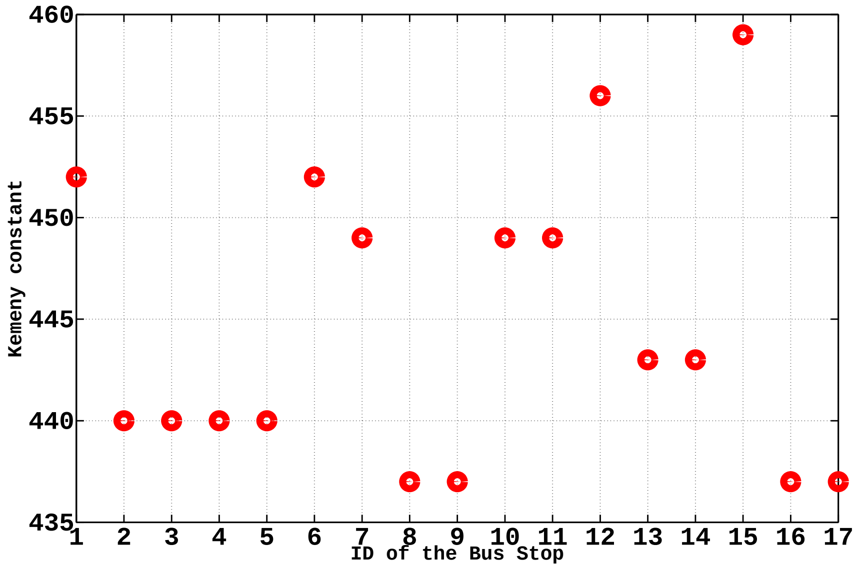


Figure 9. The Kemeny constant shows that the most critical bus stops are 1, 7, 12 and 15, i.e., those from which it is possible to change bus line.

6.2 *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.

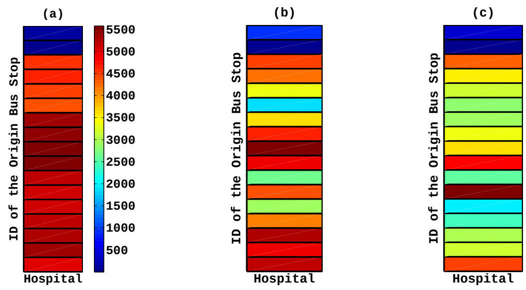


Figure 10. In this example, the MFPTs to the hospital (a) are compared to the distances to the hospital (b). We say that the hospital is fairly connected if the two vectors have the same (normalised) values. We later try to improve the fairness of bus stops 5, 7 and 12 by supporting the bus network with 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 2 corresponds to the bus stop close to the only hospital in the area. Figure 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 11. Correspondingly, we have that the accessibility of bus stops 5,

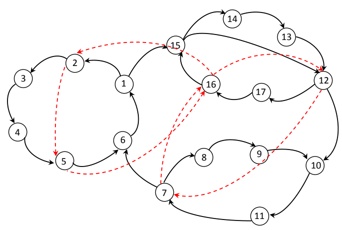


Figure 11. The new transportation network where the bus network is further supported by two new shuttle-buses lines (shown with red dashed lines) to improve the connectivity of the hospital.

7 and 12 has increased, as predicted by the MFPT analysis, as shown in Figure 10.(c).

6.3 *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 (apart from normalisation constraints, see Theorem 2.2 in Section 2.3). 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 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 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). Also notice that

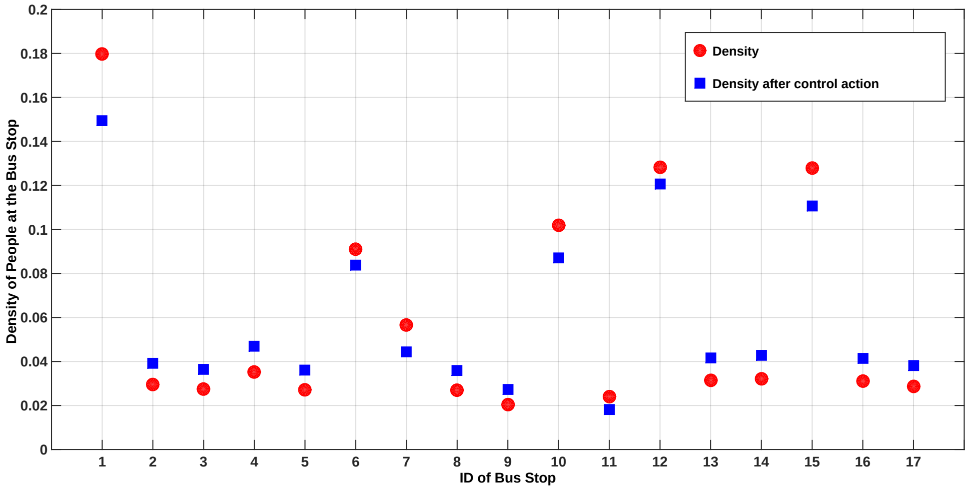


Figure 12. After the control action (i.e., doubling the frequency of the bus line 1-15-12-10-11-7-6), the densities of people at bus stops are better balanced.

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 unacceptable 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 [Kirkland (2014)].

6.4 *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 specific 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 2.2 provided a justification for the use of the second eigenvector to identify clusters in a transport network, and we remind 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.).

7 Conclusions

In this paper a Markov chain approach was developed to model multi-modal transport networks. Some information collected from the transport network (e.g., waiting and travel times) 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 efficient network control 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.

References

Häme, L., and Hakula, H. (2013), "Dynamic Journeying in Scheduled Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 360–369.

Tsubouchi, K., Yamato, H., and Hiekata, K. (2010), "Innovative on-demand bus system in Japan," *IET Intelligent Transport Systems*, vol. 4, no. 4, pp. 270–279.

Bartholdi III, J.J., and Eisenstein, D.D. (2012), "A self-coordinating bus route to resist bus bunching," *Transportation Research Part B*, vol. 46, no. 4, pp. 481–491.

Crisostomi, E., Kirkland, S., and Shorten, R. (2011), "A Google-like model of road network dynamics and its application to regulation and control," *International Journal of Control*, vol. 84, no. 3, pp. 633–651.

Crisostomi, E., Kirkland, S., Schlote, A., and Shorten, R. (2011), "Markov Chain based Emissions Models: a Precursor for Green Control," *Green IT: Technologies and Applications*, Springer Verlag, pp. 381–400.

Schlote, A., Crisostomi, E., Kirkland, S., and Shorten, R. (2012), "Traffic Modelling Framework for Electric Vehicles," *International Journal of Control*, vol. 85, no. 7, pp. 880–897.

Yang, B., Kaul, M., and Jensen, C.S. (2014), "Using Incomplete Information for Complete Weight Annotation of Road Networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 5, pp. 1267–1279.

Bekkerman, E.N., Kataev, S.G., and Kataeva, S.S. (2013), "Heuristic approximation method for a random flow of events by an MC-flow with arbitrary number of states," *Automation and Remote Control*, vol. 74, no. 9, pp. 1449–1459.

Morimura, T., Osogami, T., and Idé, T. (2013), "Solving inverse problem of Markov chain with partial observations," *IBM Research Report RT0952.*

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.*

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).*

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

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

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.

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*, vol. 335, pp. 137–150.

Doyle, P. (2009), "The Kemeny constant of a Markov chain," *preprint*, available online at url `http://www.math.dartmouth.edu/~doyle/docs/kc/kc.pdf` (last checked in October 2014).

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

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

Huisinga, W. (2003), "Metastability of Markovian systems: A transfer operator based approach in application to molecular dynamics," PhD Thesis, available online at url: `http://compphysiol.math.uni-potsdam.de/pub/HuisingaDiss.pdf` (last checked in October 2014).

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*, vol. 5, no. 3 & 4, pp. 128–138.

von Luxburg, U. (2007), "A tutorial on spectral clustering," *Statistic and Computing*, vol.17, no. 4, pp. 395–416.

Schlote, A. (2014), "PhD Thesis," Hamilton Institute, National University of Ireland Maynooth, 2014.

Langville, A., and Meyer, C. (2006), "Updating Markov Chains With an Eye on Google's PageRank," *SIAM Journal of Matrix Analysis and Applications*, vol. 27, no. 4, pp. 968–987.

Froyland, G. (2001), "Extracting dynamical behavior via Markov models," *Nonlinear dynamics and statistics*, Birkhäuser Boston, pp. 281–321.

Kirkland, S. (2014), "Load balancing for Markov chains with a specified directed graph," *Linear and Multilinear Algebra*, accepted for publication.