

Markov modeling of traffic flow in Smart Cities*

Norbert Bátfai, Renátó Besenczi, Péter Jeszenszky,
Máté Szabó, Márton Ispány

Faculty of Informatics, University of Debrecen, Hungary

`besenczi.renato@inf.unideb.hu`

`jeszenszky.peter@inf.unideb.hu`

`szabo.mate@inf.unideb.hu`

`ispany.marton@inf.unideb.hu`

Submitted: January 15, 2021

Accepted: April 25, 2021

Published online: May 18, 2021

Dedicated to the kind memory of our colleague, Norbert Bátfai.

Abstract

Modeling and simulating the traffic flow in large urban road networks are important tasks. A mathematically rigorous stochastic model proposed in [8] is based on the synthesis of the graph and Markov chain theories. In this model, the transition probability matrix describes the traffic dynamics and its unique stationary distribution approximates the proportion of the vehicles at the segments of the road network. In this paper various Markov models are studied and a simulation method is presented for generating random traffic trajectories on a road network based on the two-dimensional stationary distribution of the models. In a case study we apply our method to the central region of the city of Debrecen by using the road network data from the OpenStreetMap project which is available publicly.

*Renátó Besenczi and Márton Ispány are supported by Project no. TKP2020-NKA-04 which has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2020-4.1.1-TKP2020 funding scheme. Péter Jeszenszky and Máté Szabó are supported by the EFOP-3.6.1-16-2016-00022 project. The project is co-financed by the European Union and the European Social Fund.

Keywords: Road network, traffic simulation, discrete time Markov chain, stationary distribution, OpenStreetMap

1. Introduction

Recently, the research and development of Smart City applications have become more important by providing services to inhabitants which can make everyday life easier [15]. These applications are based on emerging technologies such as big data analytics, cloud computing, and complex sensor systems (IoT) that can support their operation. By the year 2050, 70% of Earth's population is expected to live in cities [5] whose infrastructures will face new challenges, e.g., in the field of urban traffic. In the past few years, many developments have occurred in the automobile industry, e.g., autonomous (driverless) and pure electric cars are being introduced. Since more and more people live in urban areas, solutions for problems of dense traffic such as air pollution and congestion are highly demanding [20, 24, 30].

This research presented in this paper follows our development of a traffic simulation platform initiative called rObOCar World Championship (or OOCWC for short) [2, 3]. OOCWC is a multiagent-oriented environment for creating urban traffic simulations. The traffic simulations are performed by one of its components called Robocar City Emulator (RCE), which is an open source software released under the GNU GPL v3 and is available on GitHub.¹ RCE uses the OpenStreetMap (OSM) database and processes it with the Osmium Library. The traffic simulation model of RCE is based on the Nagel-Schreckenberg (NaSch) model [21]. The result of this processing is a routing map graph and a Boost Graph Library graph which can be visualized by various map viewers. For a detailed description of the operation of RCE, see [2]. There exist several traffic simulation platforms, e.g., Multi-Agent Transport Simulation [14], Simulation of Urban Mobility [18], Aimsun,² and PTV Vissim³. The main focus of their simulation algorithms is on microscopic traffic events, while our software system focuses only on the traffic flow on the road network of the whole city.

In [8] a mathematically rigorous stochastic model is proposed for investigating the traffic flow on a road network which is based on the synthesis of discrete time Markov chains and graph theory. In this model the transition probability matrix describes the dynamics of the traffic while its unique stationary distribution corresponds to the traffic equilibrium (or steady) state on the road network. In our previous paper [4], the concepts of Markov traffic and two-dimensional stationary distribution are introduced and a parameter estimation method is proposed by using the weighted least squares (WLS) approach. To investigate complex systems, the joint application of Markov chains and large graphs is well known, see [7, 10, 19].

Our contributions in this paper are as follows. Using the approach in [4], we

¹<https://github.com/nbatfai/robocar-emulator>

²<https://www.aimsun.com/>

³<http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/>

present various Markov models for modeling traffic flow on different road graph models based on, e.g., open or closed and digraph or line digraph views. We prove the existence and uniqueness of a stationary distribution as a solution of the global balance equation, see Theorem 3.1. We define the configuration space of Markov traffic, describe the transition mechanism and prove the ergodicity of Markov traffic, see Theorem 4.1. Finally, we propose a simulation method for generating random trajectories for a Markov traffic whose two-dimensional distribution is closest to a prescribed mask matrix in the least squares sense, see Theorem 5.1. The results of this paper together with those obtained in [4], which contains some additional proofs, show that the Markovian approach still works when the scale of the road graph is significantly enlarged compared to such small one as ‘De Uithof’, which is a district in the city of Utrecht in Netherlands, see [11].

Several approaches exist for traffic flow simulation and prediction, some recent surveys are [22, 27, 31], but a few of them are based on Markov models, see [8, 23].

This paper is structured as follows. In section 2 we present various graph models of road networks. Section 3 is devoted to the probability distributions and Markov kernels on road networks. Section 4 introduces the notion of Markov traffic, describes its stationary distribution and proves its ergodicity. A simulation method is presented in section 5. In section 6 we discuss our findings, and in section 7 we conclude the paper. The Appendix provides a toy example and a proof.

2. Graph modeling of road networks

Recall that the ordered pair $G = (V, E)$ is a directed graph (digraph), where V is a finite set of vertices and E is a set of ordered pairs, called directed edges, of vertices. In the sequel, vertices (or nodes) are denoted by u, v, w , edges (or arcs or arrows) are denoted by e, f, g . For a directed edge $e = (v, w) \in E$ we also use the notation $v \rightarrow w$. We suppose that G is a simple digraph, i.e., it does not contain multiple arrows. For details, see the textbook [1].

A road network G is defined as a simple directed graph, $G = (V, E)$, where V is a set of nodes representing the terminal points of road segments, and E is a set of directed edges denoting road segments, see [25]. A road segment $e = (v, w) \in E$ is a directed edge in a road network graph, with two terminal points v and w . The vehicles move on this edge from v to w . The road network G represents the road system of a city.

Let S denote the diagonal set of V , i.e., $S := \{(v, v) | v \in V\}$. From a practical point of view, we suppose that $E \cap S = \emptyset$, i.e., there is no loop $v \rightarrow v$ in the road network in order to avoid that a vehicle is able to move in an infinite cycle. For $v \in V$, define $v^- := \{e \in E | \exists u \in V : e = (u, v)\}$ and $v^+ := \{e \in E | \exists w \in V : e = (v, w)\}$, i.e., v^- and v^+ are the sets of edges in and out the node v , respectively. Then, $\deg^-(v) = |v^-|$ and $\deg^+(v) = |v^+|$ are the indegree and outdegree of node v , respectively.

Let $L(G) = (V', E')$ be the line digraph (line road network, network line graph, see [9]) associated to G , see Section 4.5 in [1]. Here, $V' = E$ and the set E' consists

of the ordered pairs (e, f) where $e, f \in E$ such that there exist $u, v, w \in V$ that $e = (u, v)$ and $f = (v, w)$, i.e., $u \rightarrow v \rightarrow w$ is a path of length 2 (dipath) in G . The elements of E' can be described by triplets (u, v, w) , where $u, v, w \in V$, $(u, v), (v, w) \in E$, and for a directed edge in $L(G)$ we use the notation $(u, v) \rightarrow (v, w)$ too.

The digraph model of a road network assigns the vehicles moving in a city to the vertices (first-order or primal network). Contrarily, the line digraph model assigns the vehicles to the edges (second-order or dual network), see [26, 29]. When we are studying issues that are associated with the crossings (vertices) we will be concerned with the adjacency relationships of crossings, and so with the road network. On the other hand, when we are studying issues that associated with road segments we will be concerned with the adjacency relationships of road segments, and so our analyses will involve the line road network.

The digraphs G and $L(G)$ can be characterized by their degree distributions. The pairs (i, n_i^+) form the frequency histogram for the outdegree distribution of G where $n_i^+ := |\{v \in V \mid \deg^+(v) = i\}|$. The indegree frequency histogram can be defined similarly as (i, n_i^-) , where $n_i^- := |\{v \in V \mid \deg^-(v) = i\}|$. The pairs (i, m_i^+) form the frequency histogram for the outdegree distribution of $L(G)$ where $m_i^+ := \sum_{v \in G_i^+} \deg^-(v)$ and $G_i^+ := \{v \in V \mid \deg^+(v) = i\}$. (Note that $n_i^+ = |G_i^+|$.) Similarly, the pairs (i, m_i^-) form the frequency histogram for the indegree distribution of $L(G)$ where $m_i^- := \sum_{v \in G_i^-} \deg^+(v)$ and $G_i^- := \{v \in V \mid \deg^-(v) = i\}$. For the city of Debrecen (described later in this paper), the above mentioned degree distributions can be seen in Fig. 6. These histograms corroborate the fact that Debrecen's road network is a sparse graph since there is no node with higher in- and outdegree than 4.

Recall that a sequence $v_1, \dots, v_\ell \in V$, $\ell \in \mathbb{N}$, is called walk of length ℓ if $v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_\ell$. A walk is called path if its elements are different vertices. For a pair $u, v \in V$, $u \neq v$, it is said that v is reachable from u if there exists a walk v_1, v_2, \dots, v_ℓ such that $u = v_1$ and $v = v_\ell$. Clearly, if v is reachable from u , then there is a path from u to v . A digraph G is said to be strongly connected (diconnected) if every vertex is reachable from every other vertex. Clearly, the line digraph of a strongly connected digraph is also strongly connected. Namely, if $e = (u, v) \in V' (= E)$ and $f = (w, z) \in V'$ are arbitrary such that $e \neq f$, then, since G is strongly connected, there exists a walk (or a path) of length ℓ in G such that $v = v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_\ell = w$, where $v_1, \dots, v_\ell \in V$, and thus we have $e = (u, v) \rightarrow (v_1, v_2) \rightarrow \dots \rightarrow (v_{\ell-1}, v_\ell) \rightarrow (w, z) = f$, i.e., there exists a walk (or a path) of length ℓ in $L(G)$ between the vertices $e, f \in V'$. If $u \rightarrow v \rightarrow u$ for a pair $u, v \in V$ then we have $(u, v) \rightarrow (v, u) \rightarrow (u, v)$ in the line digraph, i.e., vehicles can turn back at vertex u into v . Sometimes the traffic regulations do not allow this kind of reversal, i.e., the edge set E' in $L(G)$ must not contain some triplet (u, v, u) , while some of these triplets are needed that $L(G)$ be strongly connected. By deleting all of the unnecessary triplets (u, v, u) , $u, v \in V$, such that the remaining line digraph be still strongly connected we get the minimal strongly connected line digraph of G . This line digraph is denoted by $ML(G)$. For example,

the vertices of $ML(G)$ for G in Fig. 1 are given in Table 1.

Recall that a cycle $C \subset V$ in digraph G is a path $v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_\ell \rightarrow v_1$. Here $\ell(C) = \ell$ is called the length of C . A digraph G is said to be aperiodic if the greatest common divisor of the lengths of its cycles is one. Formally, the period of G is defined as $per(G) := \gcd\{\ell > 0 : \exists C \subset V \text{ cycle such that } \ell(C) = \ell\}$. Then, G is called aperiodic if $per(G) = 1$. Clearly, if a digraph G is aperiodic then its line digraph $L(G)$ is also aperiodic. This statement follows from the following fact: if $v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_\ell \rightarrow v_1$ is a cycle then $(v_1, v_2) \rightarrow (v_2, v_3) \rightarrow \dots \rightarrow (v_\ell, v_1) \rightarrow (v_1, v_2)$ is a cycle in $L(G)$. Thus, if $\ell > 0$ and there exists a cycle $C \subset V$ such that $\ell(C) = \ell$ then there exists a cycle $C' \subset V'$ such that $\ell(C') = \ell$.

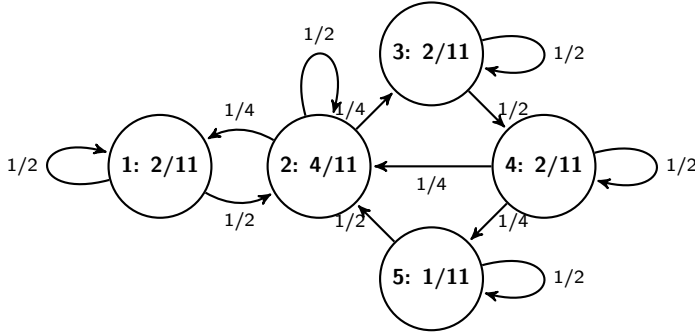


Figure 1. A Markov kernel (on edges) with its stationary distribution (on vertices with node's id) on a simple road network.

Table 1. An example for a Markov kernel on the minimal line digraph of the road network in Fig. 1.

	(1,2)	(2,3)	(3,4)	(4,2)	(2,1)	(4,5)	(5,2)
(1,2)	1/2	1/2	0	0	0	0	0
(2,3)	0	1/2	1/2	0	0	0	0
(3,4)	0	0	1/2	1/4	0	1/4	0
(4,2)	0	1/4	0	1/2	1/4	0	0
(2,1)	1/2	0	0	0	1/2	0	0
(4,5)	0	0	0	0	0	1/2	1/2
(5,2)	0	1/4	0	0	1/4	0	1/2

Let $A = (a_{uv})_{u,v \in V}$ denote the adjacency matrix of the digraph G , i.e., $a_{uv} = 1$ if and only if $(u, v) \in E$ and 0 otherwise. The number of directed walks from vertex u to vertex v of length k is the entry in the u -th row and the v -th column of the matrix A^k . For example, in Fig. 1, the number of directed walks of length 6 from vertex 2 to vertex 4 is 2, see Appendix 7. One can easily check that G is strongly connected if and only if there is a positive integer k such that the matrix $I + A + \dots + A^k$ is positive, i.e., all the entries of this matrix are positive. The

indegree and outdegree of a vertex v can be expressed by the adjacency matrix as $\deg^-(v) = \sum_{u \in V} a_{uv}$ and $\deg^+(v) = \sum_{u \in V} a_{vu}$. Let us introduce the vectors $\mathbf{d}^- := (\deg^-(v))_{v \in V}$ and $\mathbf{d}^+ := (\deg^+(v))_{v \in V}$. Then, we have $\mathbf{d}^- = A^T \mathbf{1}$ and $\mathbf{d}^+ = A \mathbf{1}$ where $\mathbf{1} := (1)_{v \in V}$ is the constant unit function. It is well known that the adjacency matrix A of an aperiodic, strongly connected graph G is primitive, i.e., irreducible and has only one eigenvalue of maximum modulus. Primitivity is equivalent to the following quasi-positivity: there exists $k \in \mathbb{N}$ such that the matrix $A^k > 0$, see Section 8.5 in [13].

In order to model the cases when vehicles leave or enter the city, we augment V by a new ideal vertex 0 and define $\bar{V} := V \cup \{0\}$, see [12]. Moreover, let \bar{E} denote the augmentation of E by directed edges $(0, v)$ and $(v, 0)$ for getting into and out of the city, respectively. Note that, for \bar{E} , it is not allowed to contain the loop $(0, 0)$. The augmentation $\bar{G} = (\bar{V}, \bar{E})$ of G is called the closure of the road network G . For $e = (v, w) \in \bar{E}$ we also use the notation $v \rightarrow w$. In what follows, we suppose that there exist $u, v \in V$ such that $u \rightarrow 0$ and $0 \rightarrow v$.

Each definition, including strong connectedness, periodicity, line digraph, given for G can be extended for \bar{G} in a natural way. Note that in the augmented line digraph $L(\bar{G}) = (\bar{V}', \bar{E}')$ the elements of the edge set \bar{E}' can be described by triplets (u, v, w) , where $u, v, w \in \bar{V}$ and if $v = 0$ then $u, w \neq 0$ and if u or w is 0 then $v \neq 0$ because triplets $(0, 0, v)$, $(v, 0, 0)$, and $(0, 0, 0)$ are excluded from \bar{E}' . One can easily see that if G is strongly connected then its closure \bar{G} is also strongly connected. Moreover, the strongly connected components of G , if there exist more than 1, can be connected through the ideal vertex 0 , resulting in a strongly connected \bar{G} . Thus, the augmented line digraph will also be strongly connected. Clearly, if G is aperiodic then \bar{G} is aperiodic too.

In the rest of this paper, it is assumed that the road network is closed by augmenting with the ideal vertex 0 .

3. Probability distributions and Markov kernels on road networks

On a road network, two kinds of probability distributions can be defined by considering the set V or E as the state space, respectively. However, the Markov kernels on the line road network must be defined with particular care.

A probability distribution (p.d.) on V is the vector $\boldsymbol{\pi} := (\pi_v)_{v \in V}$ where $\pi_v \geq 0$ for all $v \in V$ and $\sum_{v \in V} \pi_v = 1$. We may think of π_v as the proportion of all vehicles which drive through the crossing v with respect to all vehicles in the city. A Markov kernel or transition probability matrix on V is defined as a real kernel $P := (p_{uv})_{u, v \in V}$ such that $p_{uv} \geq 0$ for all $u, v \in V$ and $\sum_{v \in V} p_{uv} = 1$ for all $u \in V$. The quantity $p_{uv} \in [0, 1]$ is called the transition probability from vertex u to vertex v . In fact, P is a stochastic matrix on V and we assume that its support is the set

$E \cup S$. The sum condition for Markov kernel P can be rewritten as:

$$\sum_{w: v \rightarrow w} p_{vw} + p_{vv} = 1, \quad v \in V. \quad (3.1)$$

A p.d. π is a stationary distribution (s.d.) of the kernel P if $\sum_{u \in V} \pi_u p_{uv} = \pi_v$ for all $v \in V$. This so-called global balance equation can be expressed as:

$$\sum_{u: u \rightarrow v} \pi_u p_{uv} + \pi_v p_{vv} = \pi_v, \quad v \in V. \quad (3.2)$$

Fig. 1 presents a Markov kernel with its s.d. on a simple road network.

Since the vehicles are moving along the road segments of the road network G , it is natural to choose E to be the state space. In this case, to define probability distributions on the set of vertices again, we have to consider the line digraph $L(G)$ (or $ML(G)$). Formally, a probability distribution (p.d.) on $L(G)$ is the vector $\pi' := (\pi'_e)_{e \in E}$ where $\pi'_e \geq 0$ for all $e \in E$ and $\sum_{e \in E} \pi'_e = 1$. If we want to emphasize the vertices of the original road network G , instead of the edges, then the notation $\pi'_e = \pi'_{uv}$ is also used where $e = (u, v) \in E$. We may think of π'_e as the proportion of the vehicles at the road segment e with respect to all vehicles in the city. Note that G endowed with π' is a weighted digraph which is often called a network in itself as well.

A transition probability matrix (or Markov kernel) on E , i.e., on the line digraph $L(G)$, can be defined as a real kernel $P' := (p'_{ef})_{e, f \in E}$ such that $p'_{ef} \geq 0$ for all $e, f \in E$ and $\sum_{f \in E} p'_{ef} = 1$ for all $e \in E$. A p.d. π' on E is a s.d. of the kernel P' if $\sum_{e \in E} \pi'_e p'_{ef} = \pi'_f$ for all $f \in E$. Since G represents a road system we may suppose that if $e \neq f$ then $p'_{ef} > 0$ implies that $(e, f) \in E'$, i.e., there exist $u, v, w \in V$ such that $e = (u, v)$ and $f = (v, w)$, and hence, $u \rightarrow v \rightarrow w$ is a walk of length 2. In this case, we use the notation $p'_{ef} = p'_{uvw}$ as well. In fact, p'_{uvw} denotes the probability that a vehicle on the road segment (u, v) will go further to the road segment (v, w) in the next time point. Moreover, in the case of $e = f = (u, v)$, let $p'_{ee} = p'_{uv}$ be the probability that a vehicle remains on the same road segment in the next time point which can be non-zero as well. Thus, since P' is a Markov kernel, we have that, for all $u \rightarrow v$,

$$\sum_{w: v \rightarrow w} p'_{uvw} + p'_{uv} = 1 \quad (3.3)$$

and the global balance equation is given as:

$$\sum_{u: u \rightarrow v} \pi'_{uv} p'_{uvw} + \pi'_{vw} p'_{vw} = \pi'_{vw} \quad (3.4)$$

for all $v \rightarrow w$.

An example for the Markov kernel P' on the minimal line digraph $ML(G)$ of the road network G in Fig. 1 is shown in Table 1. Fig. 2 shows the unique s.d. π' of the Markov kernel P' .

Probability distributions and Markov kernels on the closure \overline{G} of an open road network G can be defined similarly by considering the set \overline{V} or \overline{E} as the state space,

respectively. Note that π_0 denotes the proportion of the number of vehicles which drive in or out of the city's roads at a time point. Moreover, for any Markov kernel P on \bar{V} it is supposed that $p_{00} = 0$, i.e., the vehicles cannot move from 0 to 0, thus they either enter to the road network or leave the road network. Equations (3.1) and (3.2) remain true, too. Equation (3.1) can be rewritten as

$$\sum_{w \in V: v \rightarrow w} p_{vw} + p_{v0} + p_{vv} = 1, \quad v \in V,$$

$$\sum_{w \in V: 0 \rightarrow w} p_{0w} = 1.$$

The global balance equation (3.2) for the s.d. can be rewritten as

$$\sum_{u \in V: u \rightarrow v} \pi_u p_{uv} + \pi_0 p_{0v} + \pi_v p_{vv} = \pi_v, \quad v \in V, \quad 0 \rightarrow v,$$

$$\sum_{u \in V: u \rightarrow v} \pi_u p_{uv} + \pi_v p_{vv} = \pi_v, \quad v \in V, \quad 0 \nrightarrow v,$$

$$\sum_{u \in V: u \rightarrow 0} \pi_u p_{u0} = \pi_0.$$

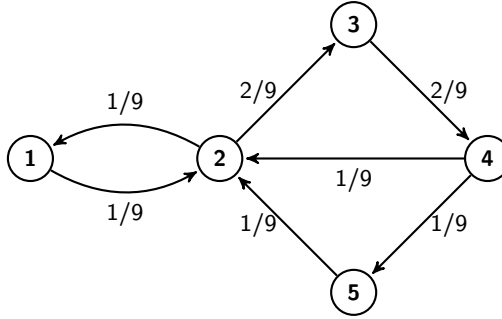


Figure 2. The stationary distribution of the Markov kernel in Table 1.

We can define Markov kernels on the line digraph $L(\bar{G})$ of the augmented road network \bar{G} , and thus on the augmented edge set \bar{E} similarly to the case of $L(G)$. Note that $(e, f) \in \bar{E}'$ implies that $e = (u, v)$ and $f = (v, w)$ where $u, v, w \in \bar{V}$ excluding the triplets $(0, 0, v)$, $(v, 0, 0)$, and $(0, 0, 0)$. We shall also use the notation $p'_{uvw} = p'_{ef}$ if $e = (u, v)$ and $f = (v, w)$ and $p'_{uv} = p'_{ee}$ if $e = (u, v)$. However, three additional conditions should be added. The first one is that $p'_{u0u} = 0$ for all $u \in V$ such that $u \rightarrow 0 \rightarrow u$. This means that if a vehicle is on the edge $(u, 0)$, i.e., it leaves the city at vertex u then it cannot be on the edge $(0, u)$ at the next time point, i.e., it cannot enter at vertex u in the road network again, immediately. The second one is that $p'_{0v0} = 0$ for all $v \in V$ such that $0 \rightarrow v \rightarrow 0$, i.e., vehicles can

enter and leave the city at node v . This means that if a vehicle enters the city then it cannot leave the city at the next time point. Finally, the third one is that $p'_{u0} = p'_{0v} = 0$ for all $u, v \in V$ such that $u \rightarrow 0$ and $0 \rightarrow v$. That is a vehicle cannot remain on the road network at the edge $(u, 0)$ after two consecutive time points and if a vehicle enters into the road network at the edge $(0, v)$ (or at the vertex v) the first time then it does not remain on this edge after the next time point and it goes further immediately in the road network. Under these conditions, equations (3.3) and (3.4) remain true. Equation (3.3) can be rewritten as:

$$\begin{aligned} \sum_{w \in V: v \rightarrow w} p'_{uvw} + p'_{uv0} + p'_{uv} &= 1, \quad u, v \in V, u \rightarrow v, \\ \sum_{w \in V: v \rightarrow w} p'_{0vw} &= 1, \quad v \in V, 0 \rightarrow v, \\ \sum_{v \in V \setminus \{u\}: 0 \rightarrow v} p'_{u0v} &= 1, \quad u \in V, u \rightarrow 0. \end{aligned}$$

Equation (3.4) can be rewritten as:

$$\begin{aligned} \sum_{u \in V: u \rightarrow v} \pi'_{uv} p'_{uvw} + \pi'_{0v} p'_{0vw} + \pi'_{vw} p'_{vw} &= \pi'_{vw}, \quad v, w \in V, v \rightarrow w, \\ \sum_{u \in V: u \rightarrow v} \pi'_{uv} p'_{uv0} + \pi'_{v0} p'_{v0} &= \pi'_{v0}, \quad v \in V, v \rightarrow 0, \\ \sum_{u \in V \setminus \{w\}: u \rightarrow 0} \pi'_{uw} p'_{u0w} + \pi'_{0w} p'_{0w} &= \pi'_{0w}, \quad w \in V, 0 \rightarrow w. \end{aligned}$$

The s.d. in all cases, i.e., for Markov kernels on road networks, line road networks and their closures, can be derived by solving the above appropriate linear equations numerically. It turns out that there is a direct connection between the existence and uniqueness of s.d. of the Markov kernels P and P' and the strongly connected property of the physical road network G if the Markov and graph structures are compatible with each other.

The Markov kernel P on V is called G -compatible if, for any $u, v \in V$ such that $u \neq v$, $p_{uv} > 0$ if and only if $(u, v) \in E$. Similarly, the Markov kernel P' on E is called G -compatible if it is $L(G)$ -compatible Markov kernel on $L(G)$, i.e., for any $e, f \in E$ such that $e \neq f$, $p'_{ef} > 0$ if and only if $(e, f) \in E'$. This is equivalent to the statement that $p'_{uvw} > 0$, $u, v, w \in V$, if and only if $(u, v), (v, w) \in E$. Since $(e, f) \in E'$ if and only if there exist $u, v, w \in V$ such that $e = (u, v)$ and $f = (v, w)$ we can define the G -compatibility of a Markov kernel P' as, for any $e, f \in E$ such that $e \neq f$, $p'_{ef} > 0$ if and only if there exist $u, v, w \in V$ such that $e = (u, v)$ and $f = (v, w)$.

Clearly, if P is G -compatible then the strong connectivity of G implies that the Markov kernel (the transition matrix) P is irreducible. Thus, by Theorem 1 in [16], see also Theorem 3.1 and 3.3 in Chapter 3 of [6] the following theorem holds.

Theorem 3.1. *If a road network G is strongly connected then there is a unique stationary distribution π (π') to any G -compatible Markov kernel P (P'). Moreover, this distribution satisfies $\pi_v > 0$ for all $v \in V$ ($\pi'_{uv} > 0$ for all $(u, v) \in E$).*

The main consequence of this theorem is that, in case of any physical road network augmented by the ideal vertex 0, all of the Markov kernels defined on the road network that has positive transition probability on all roads have unique s.d.

4. Markov traffic on road networks

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A sequence $\{X_t\}_{t \in \mathbb{Z}_+}$ of V -valued r.v.'s is a Markov chain on the state space V if the Markov property holds:

$$\mathbb{P}(X_t = v_t | X_{t-1} = v_{t-1}, \dots, X_0 = v_0) = \mathbb{P}(X_t = v_t | X_{t-1} = v_{t-1})$$

for all $t \in \mathbb{N}$, $v_0, \dots, v_t \in V$. If X, X' are V -valued r.v.'s then for the conditional distribution $P = (p_{vv'})_{v, v' \in V}$, $p_{vv'} := \mathbb{P}(X = v | X' = v')$, $v, v' \in V$, we shall also use the notation $X|X'$. Clearly, $X|X'$ is a Markov kernel on V . Similarly, a Markov chain $\{Y_t\}_{t \in \mathbb{Z}_+}$ of E -valued r.v.'s can also be defined through the Markov kernel $Y|Y'$ on the state space E .

In what follows, we suppose that the road network G is strongly connected and the Markov kernel P is G -compatible on V with unique s.d. π . The Markov chain $\{X_t\}_{t \in \mathbb{Z}_+}$ on V is called Markov random walk on the road network G with Markov kernel P if for its initial distribution $\pi_{X_0} = \pi$ and transition probabilities $X_t | X_{t-1} \sim P$ for all $t \in \mathbb{N}$. The set of k ($k \in \mathbb{N}$) mutually independent Markov random walks on G with Markov kernel P is called Markov traffic of size k and it is denoted by the quadruple (G, P, π, k) . Similarly, $\{Y_t\}_{t \in \mathbb{Z}_+}$ is a Markov random walk on the line road network if it is a Markov chain on the state space E such that $\pi'_{Y_0} = \pi'$ and $Y_t | Y_{t-1} \sim P'$ for all $t \in \mathbb{N}$.

A Markov random walk is the movement of a random vehicle which follows the stochastic rules defined by the Markov kernel. For a pair $u, v \in V$, the notation $u \Rightarrow v$ means that $(u, v) \in E \cup S$, i.e., either $u \rightarrow v$ or $u = v$. One can see that $X_t \Rightarrow X_{t+1} \Rightarrow \dots \Rightarrow X_{t+n}$ for all t and $n \in \mathbb{N}$. $\{X_t\}_{t \in \mathbb{Z}_+}$ is also called a first-order random walk on the road network where a vehicle moves from vertex u to vertex v with probability p_{uv} . On the other hand, $\{Y_t\}_{t \in \mathbb{Z}_+}$ may be referred as a second-order random walk where the vehicles move from edge to edge, i.e., we have to consider where the vehicle came from, the vertex visited before the current vertex. The second-order random walk has also been considered in graph analysis, see [29].

The state space of a first-order Markov traffic can be modeled by the function space \mathcal{F} where $\mathbf{f} \in \mathcal{F}$ is a non-negative integer valued function on V , i.e., $\mathbf{f} = (f_v)_{v \in V}$ such that $f_v \in \{0, 1, 2, \dots\}$ for all $v \in V$. The function \mathbf{f} is called a traffic configuration or a counting function and f_v measures the number of vehicles at vertex $v \in V$. Let $|\mathbf{f}|$ denote the size of the traffic configuration \mathbf{f} defined by $|\mathbf{f}| := \sum_{v \in V} f_v$. The size of a traffic configuration counts the number of vehicles on the road network at a time. Let \mathcal{F}_k ($k \in \mathbb{N}$) denote the subset of traffic

configurations of size k . A p.d. ϱ on \mathcal{F} is a function $\varrho : \mathcal{F} \rightarrow [0, 1]$ such that $\sum_{\mathbf{f}} \varrho(\mathbf{f}) = 1$. For a p.d. π on the road network G , let ϱ denote a multinomial distribution on \mathcal{F}_k with parameters k and π , see Chapter 35 in [17]. Thus, we have

$$\varrho(\mathbf{f}) := k! \prod_{v \in V} \frac{\pi_v^{f_v}}{f_v!} \quad (4.1)$$

for all $\mathbf{f} \in \mathcal{F}_k$. In fact, ϱ is the k -fold convolution of π . By formula (4.1) the probability of any complex event of the traffic can be computed.

A Markov kernel R on \mathcal{F}_k is a function $\mathcal{F}_k \times \mathcal{F}_k \rightarrow [0, 1]$ such that, for all $\mathbf{f} \in \mathcal{F}_k$, $\sum_{\mathbf{g} \in \mathcal{F}_k} R(\mathbf{f}, \mathbf{g}) = 1$. We demonstrate that every Markov kernel P induces a natural Markov kernel on \mathcal{F}_k . The matrix $K = (k_{uv})_{u, v \in V}$ is called transport matrix from traffic configuration \mathbf{f} to \mathbf{g} on the road network G if $K : V \times V \rightarrow \mathbb{N}_0$ such that $k_{uv} > 0$ implies $u \Rightarrow v$, $\sum_{v \in V} k_{uv} = f_u$ for all $u \in V$, and $\sum_{u \in V} k_{uv} = g_v$ for all $v \in V$. In fact, K has row and column marginals \mathbf{f} and \mathbf{g} , respectively, and, heuristically, K defines a way for transporting the vehicles from configuration \mathbf{f} into \mathbf{g} on the road network. An example for a transport matrix can be seen in Fig 3. For a pair $\mathbf{f}, \mathbf{g} \in \mathcal{F}_k$ let $\mathcal{M}(\mathbf{f}, \mathbf{g})$ denote the set of all transport matrices from \mathbf{f} to \mathbf{g} . Define the Markov kernel R on \mathcal{F}_k in the following way:

$$R(\mathbf{f}, \mathbf{g}) := \prod_{u \in V} f_u! \sum_{K \in \mathcal{M}(\mathbf{f}, \mathbf{g})} \prod_{u, v: u \Rightarrow v} \frac{p_{uv}^{k_{uv}}}{k_{uv}!} \quad (4.2)$$

where $\mathbf{f}, \mathbf{g} \in \mathcal{F}_k$. Then, R maps a p.d. ϱ into the p.d. $R\varrho$ on the state space \mathcal{F}_k in the following way:

$$(R\varrho)(\mathbf{g}) := \sum_{\mathbf{f} \in \mathcal{F}_k} \varrho(\mathbf{f}) R(\mathbf{f}, \mathbf{g}) \quad (4.3)$$

for all $\mathbf{g} \in \mathcal{F}_k$. To check that R is a Markov kernel indeed we note that, by the multinomial theorem,

$$\sum_{\mathbf{g} \in \mathcal{F}_k} R(\mathbf{f}, \mathbf{g}) = \prod_{u \in V} f_u! \sum_{\substack{\mathbf{g} \in \mathcal{F}_k \\ \sum_{v \in V} g_v = f_u}} \prod_{u, v: u \Rightarrow v} \frac{p_{uv}^{k_{uv}}}{k_{uv}!} = \prod_{u \in V} \left(\sum_{v \in V} p_{uv} \right)^{f_u} = 1. \quad (4.4)$$

Moreover, one can easily see similarly to (4.4), by the multinomial theorem, that if π is a s.d. of the Markov kernel P , then the p.d. ϱ defined by (4.1) is the s.d. of the induced Markov kernel R defined by (4.2). Namely, we have the global balance equation

$$\sum_{\mathbf{f} \in \mathcal{F}_k} \varrho(\mathbf{f}) R(\mathbf{f}, \mathbf{g}) = \varrho(\mathbf{g}) \quad (4.5)$$

for all $\mathbf{g} \in \mathcal{F}_k$. (For the proof see Appendix.)

Note that the concepts of traffic configuration and induced Markov kernel on them can be extended to the case of second-order Markov traffic by using the function space of non-negative integer valued functions on E as state space.

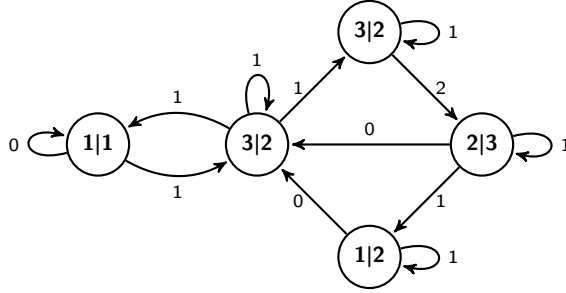


Figure 3. A transport matrix (on edges) on the road network in Fig. 1 from configuration $\mathbf{f} = (1, 3, 3, 2, 1)$ (left in vertices) to configuration $\mathbf{g} = (1, 2, 2, 3, 2)$ (right in vertices) with $k = 10$.

The applicability of the Markov traffic model is based on its ergodicity. Let \mathbf{q}_0 be an initial p.d. on \mathcal{F}_k and let us define the n th absolute p.d. \mathbf{q}_n on \mathcal{F}_k by the recursion $\mathbf{q}_n := R\mathbf{q}_{n-1}$, $n \in \mathbb{N}$, where R is a Markov kernel on \mathcal{F}_k induced by a G -compatible Markov kernel P on G , see formula (4.2). One can prove that the irreducibility and aperiodicity of P imply the same properties for R , respectively.

Our main result on ergodicity of Markov traffic, which follows from the ergodicity of irreducible aperiodic Markov chains, is the following theorem. Note that the n th power of R is defined recursively as $R^n \mathbf{q} := R(R^{n-1} \mathbf{q})$, $n = 2, 3, \dots$, by formula (4.3).

Theorem 4.1. *Let G be a strongly connected and aperiodic road network and P be a G -compatible Markov kernel. Then, there is a unique stationary distribution \mathbf{q} to the Markov traffic described by the Markov kernel R on \mathcal{F}_k induced by P which has the form (4.1).*

Moreover, the Markov traffic is ergodic in the sense that we have

$$R^n(\mathbf{f}, \mathbf{g}) \rightarrow \mathbf{q}(\mathbf{g})$$

as $n \rightarrow \infty$ for all $\mathbf{f}, \mathbf{g} \in \mathcal{F}_k$ and, for all initial p.d. \mathbf{q}_0 on \mathcal{F}_k ,

$$\mathbf{q}_n(\mathbf{f}) \rightarrow \mathbf{q}(\mathbf{f})$$

as $n \rightarrow \infty$ for all $\mathbf{f} \in \mathcal{F}_k$.

By the ergodic theorem, Theorem 4.1 implies that the p.d. π on G can be unfolded by the limit of state space averages in time as

$$\frac{1}{k} \sum_{\mathbf{f} \in \mathcal{F}_k} f_v \mathbf{q}_n(\mathbf{f}) \rightarrow \pi_v$$

as $n \rightarrow \infty$ for all $v \in V$. This formula follows from the well-known fact that the expectation vector of a multivariate distribution with parameters k and π is equal

to $k\pi$, see formula (35.6) in [17]. Similar results hold for any G -compatible Markov kernel P' on $V' = E$.

These results guarantee that the unique s.d. of a G -compatible Markov kernel can be approximated and thus explored by long run behavior of absolute p.d.'s on the traffic configurations of the road network.

5. Simulation by two-dimensional stationary distribution

A Markov traffic can be reparametrized by using its two-dimensional stationary distribution. Let us define the two-dimensional distribution $Q = (q_{uv})$ on $V \times V$ as $q_{uv} := \pi_u p_{uv}$, $u, v \in V$. One can see that Q satisfies the following properties: (i) $q_{uv} \geq 0$ for all $u, v \in V$ and $q_{uv} = 0$ for all $u, v \in V$ such that $(u, v) \notin E \cup S$; (ii) $\sum_{u,v \in V} q_{uv} = 1$ (i.e., Q is a normalized matrix on V); and (iii) $\sum_{v \in V} q_{uv} = \sum_{v \in V} q_{vu}$ for all $u \in V$ (i.e., Q has equidistributed marginals). Q is called the two-dimensional stationary distribution (2D s.d.) of the Markov traffic. Clearly, if P is G -compatible, then Q is positive on E , i.e., $q_{uv} > 0$ for all $(u, v) \in E$.

Q can also be considered as a p.d. on the state space $E \cup S$, i.e., if we extend the set V' of vertices of $L(G)$ as $V' = E \cup S$, on the line road network. Thus, we can think of Q as the distribution of the vehicles on the edges of the road network, see formula (11) in [8]. The distribution Q , similarly to traffic trajectories, can also be visualized on the edges, see Fig. 8.

For a positive Q on E , let us define

$$\begin{aligned} \pi_u &:= \sum_{v \in V} q_{uv} = \sum_{v \in V} q_{vu}, \quad u \in V, \\ p_{uv} &:= \frac{q_{uv}}{\pi_u}, \quad u, v \in V. \end{aligned} \tag{5.1}$$

Note that $\pi_v > 0$ for all $v \in V$ by Theorem 3.1. Then, $P = (p_{uv})$ defines a G -compatible Markov kernel with s.d. π on G . Thus, a Markov traffic defined by the quadruple (G, P, π, k) can be introduced by an equivalent way through the triplet (G, Q, k) .

With the help of 2D s.d., we can assign a p.d. to any Markov traffic on the space of traffic configurations which are defined on the edges of the road network. Namely, let the traffic configuration $\mathbf{h} = (h_{uv})_{u \Rightarrow v}$ be a non-negative integer valued function on $E \cup S$. Here, h_{uv} denotes the number of vehicles on the edge (u, v) where $u, v \in V$ such that $u \Rightarrow v$. We define the two-dimensional distribution σ on the set of traffic configurations \mathbf{h} with size k ($k \in \mathbb{N}$), i.e., where $\sum_{u \Rightarrow v} h_{uv} = k$. Similarly to (4.1), the two-dimensional distribution σ induced by a p.d. π on G as its k -fold convolution has a multinomial distribution with parameter k and Q , i.e., for all \mathbf{h} , we have

$$\sigma(\mathbf{h}) := k! \prod_{u \Rightarrow v} \frac{q_{uv}^{h_{uv}}}{h_{uv}!}.$$

In fact, σ describes the 2D s.d. of a Markov traffic with size k . One can easily see that the concept of 2D s.d. can also be extended for the second-order Markov traffic.

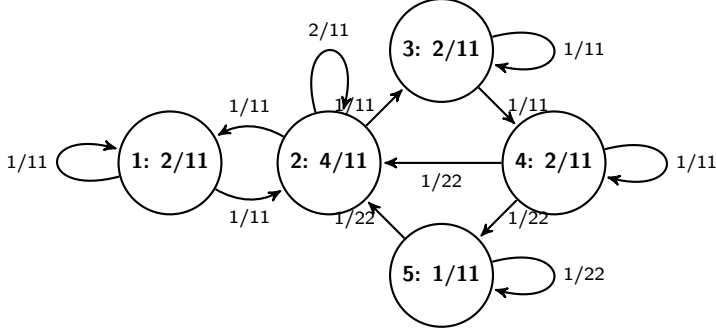


Figure 4. The two-dimensional stationary distribution (on edges) with its equidistributed marginals (on vertices) for the Markov kernel in Fig. 1. One can easily check that the sums of probabilities written on the edges in and out each vertex are equal, respectively.

The simulation algorithm presented in this paper is based on the 2D s.d. defined on the road graph. However, it is not an easy task to find a matrix Q which satisfies properties (i)-(iii) on a sparse graph. Hence, at first, we propose a method for finding such Q which is closest to a given mask matrix M on G in the least square sense. The role of the mask matrix is to specify the weight of edges by modeling the odds of consecutive occurrences of cars on the terminal points of edges in the road network. For example, these weights may stem from observed trajectories for the traffic in a time period.

Let us observe a random sample of trajectories $\{X^i\}$, $i = 1, \dots, k$, of size k defined by $X_1^i \Rightarrow X_2^i \Rightarrow \dots \Rightarrow X_{n_i}^i$, $i = 1, \dots, k$, where n_i denotes the length of the i th trajectory. The total sample size is given by $n := n_1 + \dots + n_k$. Define the total two-dimensional consecutive empirical frequencies as:

$$n_{uv} := \sum_{i=1}^k \sum_{j=1}^{n_i-1} I(X_j^i = u, X_{j+1}^i = v), \quad (5.2)$$

$u, v \in V$, where I denotes the indicator function. Plainly, n_{uv} is the number of consecutive pairs (u, v) ($u, v \in V$) in the trajectories. One can see that the support of the two-dimensional frequency matrix $N := (n_{uv})_{u,v \in V}$ is a subset of $E \cup S$. Clearly, $\mathbf{1}^\top N \mathbf{1} = n - k$, where $n - k$ is the corrected sample size. One can also see that the vectors $N^\top \mathbf{1} - N \mathbf{1}$ and $\mathbf{1}$ are orthogonal. In this case, the matrix N is a good candidate for the role of the mask matrix M .

We define the optimality criteria for determining Q by means of the least squares distance between matrices over G . Let $A = (a_{uv})_{u,v \in V}$ and $B = (b_{uv})_{u,v \in V}$ such that $a_{uv} = b_{uv} = 0$ for all $u, v \in V$ where $u \nrightarrow v$. The least square distance between

A and B is defined as

$$\|A - B\|_G^2 := \sum_{u,v: u \Rightarrow v} |a_{uv} - b_{uv}|^2.$$

In fact, $\|\cdot\|_G$ is the Frobenius norm of the matrices of dimension $|V| \times |V|$ which vanish on the entries outside of $E \cup S$.

To formulate our main result, we need some basic facts on the spectral theory of directed graphs, see [28] for details. The symmetric unnormalized graph Laplacian matrix L of a digraph G is defined as $L := D - A - A^\top$, where A denotes the adjacency matrix of G and $D := \text{diag}\{\mathbf{d}^+ + \mathbf{d}^-\}$.

Theorem 5.1. *Let M be a non-negative matrix on G . Then, there is a unique pair (Q, \varkappa) , where the matrix Q on G satisfies properties (i)-(iii) and $\varkappa \geq 0$, which minimizes the error function $\|\varkappa Q - M\|_G^2$. Moreover, the unique solution to this optimization problem is derived as*

$$\begin{aligned} \varkappa &:= \mathbf{1}^\top M \mathbf{1} + (\mathbf{d}^- - \mathbf{d}^+)^\top \boldsymbol{\lambda}, \\ Q &:= \varkappa^{-1} (M + (\mathbf{1} \boldsymbol{\lambda}^\top - \boldsymbol{\lambda} \mathbf{1}^\top) \circ A), \end{aligned}$$

where $\boldsymbol{\lambda} = (\lambda_v)_{v \in V}$ is called Lagrange vector and defined as a unique solution to the vector linear equation $L\boldsymbol{\lambda} = (M - M^\top)\mathbf{1}$ which satisfies the constraint $\mathbf{1}^\top \boldsymbol{\lambda} = 0$ (i.e., $\sum_{v \in V} \lambda_v = 0$), and \circ denotes the entrywise (Hadamard) product of matrices.

The proof of Theorem 5.1 is based on the Lagrange method, see Appendix in [4]. One can easily see that the error function at the optimum equals to the sum of squared differences (SSD) of the Lagrange vector defined by

$$\text{SSD} := \sum_{u \rightarrow v} (\lambda_u - \lambda_v)^2.$$

The fundamental statement of Theorem 5.1, as one of the main results of this paper, is that the optimal 2D s.d. Q is a low-dimensional perturbation of the mask matrix M . This perturbation term and the normalizing constant \varkappa depend on two components through a unique solution to a vector linear equation. The coefficient matrix of the linear equation is the Laplacian matrix L of the road graph which depends only on the graph structure of the road network and independent from the mask matrix. Thus, L can be computed and stored in advance for a given road network. Contrarily, the constant vector of the linear equation depends only on the marginals of the mask matrix, however, it does not depend on its entries and mainly on the road network itself.

After having defined or determined a 2D s.d. Q on a road network G , a simple simulation algorithm for generating random trajectories on G is the following. A trajectory t of length ℓ is a generalized path $v_0 \Rightarrow v_1 \Rightarrow \dots \Rightarrow v_{\ell-1}$, $v_i \in V$, $i = 0, 1, \dots, \ell-1$, which is stored in an ordered list as $t = [v_0, v_1, \dots, v_{\ell-1}]$. Note that $v_i = v_{i+1}$ is also allowed for any index i , i.e., a vehicle may stay in place after a timestep. The temporary set of generated trajectories is stored in a dictionary D

which consists of key-value pairs (v, T_v) . Here, the key $v \in V$ identifies a node in the road graph, and the value $T_v = [t_0, t_1, \dots, t_{n_v-1}]$ is an ordered list of trajectories $t_j, j = 0, 1, \dots, n_v - 1$, of length n_v such that the last element of all trajectories in T_v is v , i.e., the trajectories end at the node v . In fact, T_v is a list of lists for each $v \in V$. Let T denote the final set of trajectories as the output of the algorithm. By generating random pairs (u, v) from Q successively, D is updated, and then T is derived in the following way. If T_u is not empty, then let the trajectory t be given by appending v to the first trajectory in T_u . Moreover, let us delete this trajectory from the list T_u . If the length of t is large enough, then let us add it to T , otherwise add it to the list T_v . If T_u was empty then append the list $[u, v]$ to T_v .

Algorithm 1: Trajectory simulation.

Input: Q : two-dimensional stationary distribution
 m : maximum trajectory length
 n : number of simulated consecutive pairs

Output: T : list of trajectories

```

/* initialization */
D = {}; /* temporary dictionary */
T = [];
/* iterating over simulated pairs */
for i = 1 to n do
    pick a random pair  $(u, v) \sim Q$ ;
    if  $D[u]$  is not empty then
         $t = D[u][0]$ ; /* temporary trajectory */
        append node  $v$  to  $t$ ;
        delete the first element of  $D[u]$ ;
    else
         $t = [u, v]$ ;
    end
    if  $\text{length}(t) = m$  then
        append  $t$  to  $T$ ;
    else
        if  $v \in D$  then
            append  $t$  to  $D[v]$ ;
        else
            append  $(v, t)$  to  $D$ ;
        end
    end
end
end
/* appending the trajectories in temporary dictionary to the output */
for  $v$  in  $D$  do
    append  $D[v]$  to  $T$ ;
end

```

Having finished the random generation of pairs, let us append the trajectories of whole D to the final set T . One can easily see that the longer trajectories are at the head of T . A pythonic pseudo-code of the above procedure is in Algorithm 1. After the simulation, the generated trajectories can be visualized by using a digital map system, e.g., Google Maps or OpenStreetMap. Finally, we note that, in a typical step of the algorithm, a trajectory moves from the first position of a trajectory list to the last position of an other one. This is a kind of mixing which helps to avoid the formation of very unbalanced trajectories.

6. Results

In our work, OpenStreetMap (OSM) was used which is a community project to build a free map of the world. OSM data is available under the Open Data Commons Open Database License (ODbL). The representation and storing of map data is based on only three modeling primitives: nodes, ways, and relations.⁴ A node represents a geographical entity with GPS coordinates. A way is an ordered list of at least two nodes. A relation is an ordered list of nodes, ways, and/or relations. Users can export map data at the OSM web site manually, selecting a rectangular region of the map. OSM uses OSM XML and PBF formats for exporting map data. Software libraries for parsing and working with OSM data are available for several programming languages.⁵

We started our processing by building a graph from the OSM map of Debrecen in the bounding box defined by the coordinates N47.4771, W21.5565, N47.571, W21.6918, see Fig. 5. Because we only need those nodes that can be reached by vehicles, we had to filter the OSM file and collect only specific types of way nodes. In the OSM file, a way is a sequence of OSM nodes, so naturally, the nodes of ways become nodes in the graph. For every node we store the node's OSM ID and its coordinates. We also insert an edge into the graph to connect each pair of nodes that follow each other in a way. We used the PyOsmium library for processing the OSM files and the NetworkX Python library for building the graph. The result of this processing is an aperiodic strongly connected road network of Debrecen augmented by the ideal vertex 0. The descriptive statistics of edges of the road graph are: Min=0.3395, Q1=10.7906, Med=24.7830, Mean=49.9052, Q3=67.6021, Max=1167.4902 (in meters). The degree distributions of this road network are visualized in Fig. 6.

To evaluate the performance of the proposed algorithm a simple simulation study was conducted at different sample sizes for the road network of Debrecen. In the simulations, we kept the length of trajectories low and the number of trajectories high compared to the size of the road network. By our experience, the real traffic trajectories possess these properties. All simulations were carried out in Python. The codes and datasets of our simulation are available upon request.

We have also implemented the model in the OOCWC system. Regarding RCE,

⁴<http://wiki.openstreetmap.org/wiki/Elements>

⁵<https://wiki.openstreetmap.org/wiki/Frameworks>

this validation technique is one of our future work.

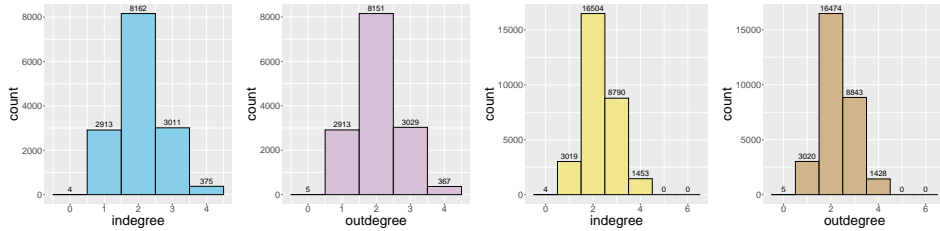


Figure 6. The degree distribution (first: in-vertices, second: out-vertices, third: in-edges, fourth: out-edges) histograms of the Debrecen map road graph.

Table 2. Transitions of intersection 26755459.

Neighbor node	Transition Probability
1402222987	0.24
1402222861	0.32
1534652124	0.26
7834632455	0.18
Sum	1

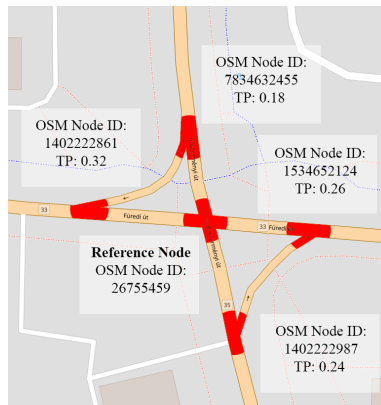


Figure 7. A visual explanation of transitions of intersection 26755459. TP means transition probability, nodes are highlighted with red. Base map and data from OpenStreetMap and OpenStreetMap Foundation. © OpenStreetMap contributors. Annotated by the authors.

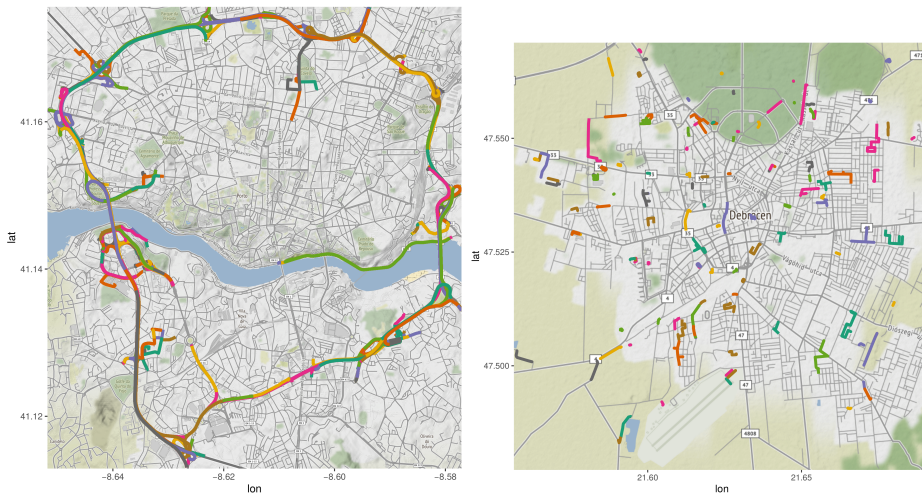


Figure 8. Generated trajectories in Porto (left: $n = 200,000,000$; $m = 75$) and Debrecen (right: $n = 50,000,000$; $m = 35$) simulated with Algorithm 1.

7. Conclusions

In this paper we have described various graph models for proper road networks and introduced the concept of Markov traffic. By tools of Markov chain theory, we have proven the existence and uniqueness of a stationary distribution for any Markov traffic on strongly connected and aperiodic road networks. We have also derived an explicit formula for the stationary distribution and the two-dimensional stationary distribution. Finally, we have proposed a simulation algorithm for generating random trajectories which follows the two-dimensional stationary distribution which being closest to a given mask matrix on the road network.

To test our theories, we have implemented the proposed model in our simulation program (RCE) using OpenStreetMap. The whole project (including RCE) is available for download.⁶

Future work will focus on the further improvements and the possible applications of our simulation algorithms, e.g., modelling the pollution or energy consumption in Smart Cities.

Acknowledgements. The authors would like to thank all actual and former members of the Smart City group of the University of Debrecen. We are especially grateful to all of the participants of the OOCWC competitions and the students of the BSc courses of “High Level Programming Languages” at the University of Debrecen.

⁶<https://github.com/rbesenczi/Crowd-sourced-Traffic-Simulator/blob/master/justine/install.txt>

References

- [1] J. BANG-JENSEN, G. Z. GUTIN: *Digraphs: Theory, Algorithms and Applications*, Berlin Heidelberg New York: Springer Science & Business Media, 2008, DOI: <https://doi.org/10.1007/978-1-84800-998-1>.
- [2] N. BÁTFAI, R. BESENCZI, A. MAMENYÁK, M. ISPÁNY: *OOVCW: The Robocar World Championship initiative*, in: Telecommunications (ConTEL), IEEE 13th International Conference on, ed. by T. PLANCK, 2015, pp. 1–6, DOI: <https://doi.org/10.1109/ConTEL.2015.7231223>.
- [3] N. BÁTFAI, R. BESENCZI, A. MAMENYÁK, M. ISPÁNY: *Traffic simulation based on the Robocar World Championship initiative*, Infocommunications Journal 7.3 (2015), pp. 50–58.
- [4] R. BESENCZI, N. BÁTFAI, P. JESZENSZKY, R. MAJOR, F. MONORI, M. ISPÁNY: *Large-scale Simulation of Traffic Flow using Markov Model*, PLoS ONE 16.2 (2021), DOI: <https://doi.org/10.1371/journal.pone.0246062>.
- [5] P. BOCQUIER: *World Urbanization Prospects: An alternative to the UN model of projection compatible with the mobility transition theory*, Demographic Research 12 (2005), pp. 197–236, DOI: <https://dx.doi.org/10.4054/DemRes.2005.12.9>.
- [6] P. BRÉMAUD: *Markov chains. Gibbs fields, Monte Carlo simulation, and queues*, vol. 31, Texts in Applied Mathematics, New York: Springer-Verlag, 1999, DOI: <https://doi.org/10.1007/978-1-4757-3124-8>.
- [7] M. CAVERS, K. VASUDEVAN: *Spatio-temporal complex Markov Chain (SCMC) model using directed graphs: Earthquake sequencing*, Pure and Applied Geophysics 172.2 (2015), pp. 225–241, DOI: <https://doi.org/10.1007/s00024-014-0850-7>.
- [8] E. CRISOSTOMI, S. KIRKLAND, R. SHORTEN: *A Google-like model of road network dynamics and its application to regulation and control*, International Journal of Control 84.3 (2011), pp. 633–651, DOI: <https://doi.org/10.1080/00207179.2011.568005>.
- [9] M. CROVELLA, E. KOLACZYK: *Graph wavelets for spatial traffic analysis*, in: IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428), vol. 3, 2003, pp. 1848–1857, DOI: <https://doi.org/10.1109/INFCOM.2003.1209207>.
- [10] C. DABROWSKI, F. HUNT: *Using Markov chain and graph theory concepts to analyze behavior in complex distributed systems*, tech. rep., U.S. National Institute of Standards and Technology, 2011.
- [11] M. FAIZRAHNEEMOON: *Real-data modelling of transportation networks*, PhD thesis, Hamilton Institute, National University of Ireland Maynooth, 2016.
- [12] M. FAIZRAHNEEMOON, A. SCHLOTE, E. CRISOSTOMI, R. SHORTEN: *A Google-like model for public transport*, in: International Conference on Connected Vehicles and Expo (ICCVE), 2013, pp. 612–613, DOI: <https://doi.org/10.1109/ICCVE.2013.6799864>.
- [13] R. A. HORN, C. R. JOHNSON: *Matrix Analysis*, Cambridge: Cambridge University Press, 2012.
- [14] A. HORN, K. NAGEL, K. W. AXHAUSEN: *The Multi-Agent Transport Simulation MATSim*, Ubiquity Press London, 2016, DOI: <https://doi.org/10.5334/baw>.
- [15] E. ISMAGILOVA, L. HUGHES, Y. K. DWIVEDI, K. R. RAMAN: *Smart cities: Advances in research—An information systems perspective*, International Journal of Information Management 47 (2019), pp. 88–100, DOI: <https://doi.org/10.1016/j.ijinfomgt.2019.01.004>.

- [16] J. P. JARVIS, D. R. SHIER: *Graph-theoretic analysis of finite Markov chains*, in: Applied mathematical modeling: A multidisciplinary approach, ed. by D. R. SHIER, K. T. WALLENIUS, CRC Press, 1996, pp. 85–102.
- [17] N. L. JOHNSON, S. KOTZ, N. BALAKRISHNAN: *Discrete Multivariate Distributions*, New York: Wiley, 1997.
- [18] D. KRAJZEWICZ, J. ERDMANN, M. BEHRISCH, L. BIEKER: *Recent Development and Applications of SUMO - Simulation of Urban MObility*, International Journal On Advances in Systems and Measurements, International Journal On Advances in Systems and Measurements 5.3&4 (Dec. 2012), pp. 128–138, URL: <http://elib.dlr.de/80483/>.
- [19] A. LESNE: *Complex Networks: from Graph Theory to Biology*, Letters in Mathematical Physics 78 (2006), pp. 235–262, DOI: <https://doi.org/10.1007/s11005-006-0123-1>.
- [20] H. MENOVAR, I. GUVENC, K. AKKAYA, A. S. ULUAGAC, A. KADRI, A. TUNCER: *UAV-Enabled Intelligent Transportation Systems for the Smart City: Applications and Challenges*, IEEE Communications Magazine 55.3 (2017), pp. 22–28, DOI: <https://doi.org/10.1109/MCOM.2017.1600238CM>.
- [21] K. NAGEL, M. SCHRECKENBERG: *A cellular automaton model for freeway traffic*, Journal de Physique I France 2.12 (1992), pp. 2221–2229.
- [22] A. M. NAGY, V. SIMON: *Survey on traffic prediction in smart cities*, Pervasive and Mobile Computing 50 (2018), pp. 148–163, DOI: <https://doi.org/10.1016/j.pmcj.2018.07.004>.
- [23] E. NECULA: *Dynamic traffic flow prediction based on GPS data*, in: IEEE 26th International Conference on Tools with Artificial Intelligence, 2014, pp. 922–929, DOI: <https://doi.org/10.1109/ICTAI.2014.140>.
- [24] L. E. OLMOS, S. ÇOLAK, S. SHAFIEL, M. SABERI, M. C. GONZÁLEZ: *Macroscopic dynamics and the collapse of urban traffic*, Proceedings of the National Academy of Sciences 115.50 (2018), pp. 12654–12661, DOI: <https://doi.org/10.1073/pnas.1800474115>.
- [25] B. PAN, Y. ZHENG, D. WILKIE, C. SHAHABI: *Crowd sensing of traffic anomalies based on human mobility and social media*, in: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2013, pp. 344–353, DOI: <https://doi.org/10.1145/2525314.2525343>.
- [26] S. PORTA, P. CRUCITTI, V. LATORA: *The Network Analysis of Urban Streets: A Dual Approach*, Physica A 369 (2006), pp. 853–866, DOI: <https://doi.org/10.1016/j.physa.2005.12.063>.
- [27] E. VLAHOGIANNI, M. KARLAFTIS, J. GOLIAS: *Short-term traffic forecasting: Where we are and where we were going*, Transportation Research Part C Emerging Technologies 43.1 (2014), DOI: <https://doi.org/10.1016/j.trc.2014.01.005>.
- [28] U. VON LUXBURG: *A tutorial on spectral clustering*, Statistics and Computing 17.4 (2007), pp. 395–416, DOI: <https://doi.org/10.1007/s11222-007-9033-z>.
- [29] Y. WU, X. ZHANG, Y. BIAN, Z. CAI, X. LIAN, X. LIAO, F. ZHAO: *Second-order random walk-based proximity measures in graph analysis: Formulations and algorithms*, The VLDB Journal 27.1 (2018), pp. 127–152, DOI: <https://doi.org/10.1007/s00778-017-0490-5>.
- [30] Y. XU, S. ÇOLAK, E. C. KARA, S. J. MOURA, M. C. GONZÁLEZ: *Planning for electric vehicle needs by coupling charging profiles with urban mobility*, Nature Energy 3.6 (2018), pp. 484–493, DOI: <https://doi.org/10.1038/s41560-018-0136-x>.

- [31] X. YIN, G. WU, J. WEI, Y. SHEN, H. QI, B. YIN: *A Comprehensive Survey on Traffic Prediction*, tech. rep., <https://arxiv.org/abs/2004.08555>, Arxiv, 2020.

Appendix

In order to demonstrate the results of this paper we present a simple toy example implemented in Python. Consider the road network $G = (V, E)$ on Fig. 1, where $V := \{1, 2, 3, 4, 5\}$ and $E := \{(1, 2), (2, 1), (2, 3), (3, 4), (4, 2), (4, 5), (5, 2)\}$. Then $|V| = 5$ and $|E| = 7$. The adjacency matrix A of G , where we denote the vertices as well, can be derived as:

$$A := \begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 0 & 1 & 0 \\ 4 & 0 & 1 & 0 & 0 & 1 \\ 5 & 0 & 1 & 0 & 0 & 0 \end{array}.$$

Clearly, G is a strongly connected digraph. Since $1 \rightarrow 2 \rightarrow 1$ and $2 \rightarrow 3 \rightarrow 4 \rightarrow 2$ are cycles of length 2 and 3, respectively, we have $\text{per}(G) = 1$ and thus G is aperiodic. The first power k that $A^k > 0$ is $k = 6$ and

$$A^6 := \begin{bmatrix} 2 & 2 & 2 & 1 & 1 \\ 2 & 4 & 2 & 2 & 1 \\ 2 & 3 & 2 & 1 & 1 \\ 3 & 4 & 3 & 2 & 1 \\ 2 & 2 & 2 & 1 & 1 \end{bmatrix}.$$

The entries of this matrix are the number of directed walks of length 6 between the pairs of vertices. One can see that the in- and outdegree of vertices are given as $\mathbf{d}^- = (1, 3, 1, 1, 1)^\top$ and $\mathbf{d}^+ = (1, 2, 1, 2, 1)^\top$, respectively.

Define the Markov kernel P on the road network G as:

$$P := \begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & 1/2 & 1/2 & 0 & 0 & 0 \\ 2 & 1/4 & 1/2 & 1/4 & 0 & 0 \\ 3 & 0 & 0 & 1/2 & 1/2 & 0 \\ 4 & 0 & 1/4 & 0 & 1/2 & 1/4 \\ 5 & 0 & 1/2 & 0 & 0 & 1/2 \end{array}.$$

Fig. 1 displays the Markov kernel P denoting the transition probabilities on the edges and its s.d. π denoting on the vertices. Note that $\pi = 1/11(2, 4, 2, 2, 1)^\top$ and the 2D s.d. is given by:

$$Q = \frac{1}{22} \begin{bmatrix} 2 & 2 & 0 & 0 & 0 \\ 2 & 4 & 2 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 \\ 0 & 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

One can easily check that the marginals of Q coincide to π .

We compute the 2D s.d. least square approximation of the adjacency matrix A by Theorem 5.1. The symmetric unnormalized graph Laplacian matrix L of the road network G is given as:

$$L = \begin{bmatrix} 2 & -2 & 0 & 0 & 0 \\ -2 & 5 & -1 & -1 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & -1 & -1 & 3 & -1 \\ 0 & -1 & 0 & -1 & 2 \end{bmatrix}.$$

The eigenvalues are $(0, 1.55, 2, 4, 6.45)$. The multiplicity of the smallest eigenvalue 0 is 1 which shows that the road network is strongly connected. The generalized (Moore-Penrose) inverse of L can be derived as

$$L^{-1} = \begin{bmatrix} 0.44 & 0.04 & -0.16 & -0.16 & -0.16 \\ 0.04 & 0.14 & -0.06 & -0.06 & -0.06 \\ -0.16 & -0.06 & 0.365 & -0.01 & -0.135 \\ -0.16 & -0.06 & -0.01 & 0.24 & -0.01 \\ -0.16 & -0.06 & -0.135 & -0.01 & 0.365 \end{bmatrix}.$$

Then, by solving the vector linear equation $L\lambda = \mathbf{d}^+ - \mathbf{d}^-$, we have Lagrange multipliers $\lambda = (-0.2, -0.2, 0.05, 0.3, 0.05)$. One can see that the sum of multipliers is 0. Thus, the 2D s.d. Q_A to the adjacency matrix A is

$$Q_A = \frac{1}{26} \begin{bmatrix} 0 & 4 & 0 & 0 & 0 \\ 4 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 2 & 0 & 0 & 3 \\ 0 & 3 & 0 & 0 & 0 \end{bmatrix}$$

with stationary marginals $\pi_A = 1/26(4, 9, 5, 5, 3)^\top$. The error square of the approximation is $\text{SSD} = 0.5$.

Proof of formula (4.5). For all $\mathbf{g} \in \mathcal{F}_k$ we have by formulas (4.1) and (4.2) and the multinomial theorem that

$$\begin{aligned} \sum_{\mathbf{f} \in \mathcal{F}_k} \varrho(\mathbf{f}) R(\mathbf{f}, \mathbf{g}) &= k! \sum_{\mathbf{f} \in \mathcal{F}_k} \prod_{u \in V} \pi_u^{f_u} \sum_{K \in \mathcal{M}(\mathbf{f}, \mathbf{g})} \prod_{u, v: u \Rightarrow v} \frac{p_{uv}^{k_{uv}}}{k_{uv}!} \\ &= k! \sum_{\mathbf{f} \in \mathcal{F}_N} \sum_{K \in \mathcal{M}(\mathbf{f}, \mathbf{g})} \prod_{u, v: u \Rightarrow v} \frac{(\pi_u p_{uv})^{k_{uv}}}{k_{uv}!} = k! \sum_{\substack{\sum_{u \in V} k_{uv} = g_v}} \prod_{u, v: u \Rightarrow v} \frac{(\pi_u p_{uv})^{k_{uv}}}{k_{uv}!} \\ &= k! \prod_{v \in V} (g_v!)^{-1} \left(\sum_{u \in V} \pi_u p_{uv} \right)^{g_v} = k! \prod_{v \in V} \frac{\pi_v^{g_v}}{g_v!} = \varrho(\mathbf{g}). \quad \square \end{aligned}$$