**Article:**
Iryo, T. orcid.org/0000-0003-0839-0835, Watling, D. orcid.org/0000-0002-6193-9121 and Hazelton, M. orcid.org/0000-0001-7831-725X (2024) Estimating Markov Chain Mixing Times: Convergence Rate Towards Equilibrium of a Stochastic Process Traffic Assignment Model. Transportation Science. ISSN 0041-1655

https://doi.org/10.1287/trsc.2024.0523

# Estimating Markov chain mixing times: convergence rate towards equilibrium of a stochastic process traffic assignment model

Takamasa Iryo[a,*], David Watling[b], and Martin Hazelton[c]

[a]*Tohoku University, Japan*
[b]*The University of Leeds, United Kingdom*
[c]*The University of Otago, New Zealand*

## Abstract

Network equilibrium models have been extensively used for decades. The rationale for using equilibrium as a predictor is essentially that (i) a unique and globally stable equilibrium point is guaranteed to exist, and (ii) the transient period over which a system adapts to a change is sufficiently short in time that it can be neglected. However, we find transport problems without a unique and stable equilibrium in the literature. Even if it exists, it is not certain how long it takes for the system to reach an equilibrium point after an external shock onto the transport system, such as infrastructure improvement and damage by a disaster. The day-to-day adjustment process must be analysed to answer these questions. Among several models, the Markov chain approach has been claimed to be the most general and flexible. It is also advantageous as a unique stationary distribution is guaranteed in mild conditions, even when a unique and stable equilibrium does not exist. In the present paper, we first aim to develop a methodology for estimating the Markov Chain Mixing Time (MCMT), a worst-case assessment of the convergence time of a Markov chain to its stationary distribution. The main tools are coupling and aggregation, which enable us to analyse MCMTs in large-scale transport systems. Our second aim is to conduct a preliminary examination of the relationships between MCMTs and critical properties of the system, such as travellers' sensitivity to differences in travel cost and the frequency of travellers' revisions of their choices. Through analytical and numerical analyses, we found key relationships in a few transport problems, including those without a unique and stable equilibrium. We also showed that the proposed method, combined with coupling and aggregation, can be applied to larger transport models.

*Keywords: Markov chain mixing time, day-to-day dynamics, stochastic traffic assignment process*

* Corresponding author. E-mail address: iryo@tohoku.ac.jp.

# 1 Introduction

Network equilibrium models have been extensively used for many decades, for example, to predict the impacts of anticipated exogenous changes in travel demand or to evaluate the consequences of hypothesised policy measures. Although it is known that real-life systems undergo many changes as travellers adapt their behaviour, the *rationale* for using a point equilibrium as a predictor is essentially that: (i) under some technical conditions, a unique and globally stable equilibrium point is guaranteed to exist, and (ii) the transient period over which a system adapts to a change is sufficiently short in time that it can be neglected. However, there are two limitations of this argument:

**Issue A**: Some transport problems do not have a unique and stable equilibrium, and for many others we are unsure. In the literature a range of examples have been reported where multiple equilibria exist or no stable equilibrium exists, such as with junction interactions, multiple vehicle types, responsive traffic signals, demand-responsive services and within-day dynamic interactions (e.g. Watling (1996); Iryo (2008); Iryo (2011); Iryo (2015); Cantarella et al. (2015); Guo et al. (2018); Iryo and Watling (2019)). In addition, Arieli and Young (2016) showed games with atomic users, in which the convergence time can grow exponentially with better response dynamics. Even without these examples, the typical convexity/monotonicity assumptions used to establish equilibrium uniqueness are known not to hold in many real-life cases.

**Issue B**: Even if a unique and stable equilibrium exists, it is not certain how long it takes for the system to reach an equilibrium point from some given initial condition. Evidence from real-life studies has demonstrated transient effects over varying time-scales (e.g. Zhu et al. (2010); Watling et al. (2012)). Evidence from modelling studies indicates that, depending on the behavioural, network and demand settings, transient effects may potentially persist for long periods (Watling (1996); Watling and Hazelton (2018)), bringing into question the premise of equilibrium as a reasonable near-term predictor.

The day-to-day adjustment process must be analysed to answer the above two questions. Different criteria for modelling such transport systems have been adopted in the literature. They may be classified into: continuous, deterministic models (flow-based dynamical systems; discrete, deterministic models (such as evolutionary games with atomic players), and discrete, stochastic models (such as a Markov chain)). The reader is referred to Cantarella and Watling (2016) for a review

and description of a wide class of such deterministic models, and Watling and Cantarella (2015) for an analogous review/overview of stochastic process models. Of these approaches, the Markov chain approach (e.g. Cascetta (1989), Cascetta and Cantarella (1991), Cantarella and Cascetta (1995), Hazelton (2002), Hazelton and Watling (2004), Watling and Cantarella (2013), Parry et al. (2016), and Watling and Hazelton (2018)) has claimed to be the most general and flexible, since it is able to capture a wide range of levels of aggregation and behavioural/traffic assumptions, down to the level of atomic decision-makers, while being able to represent both the uncertain and time-varying evolution of future system states. These features also imply that large-scale traffic simulations for the real world can also be constructed based on the Markov chain model.

Under mild conditions such Markov chain models are known to possess a unique stationary probability distribution, to which the system will converge regardless of the initial conditions, and which can be considered the counterpart of an equilibrium state in conventional steady-state models. This property resolves Issue A. In addition, the Markov approach has the additional property that we may also apply it to study the speed at which the system adjusts from some given initial conditions. This is of potentially great significance if the speed is so slow as to render stationary/equilibrium predictions practically irrelevant. It opens up two possible alternative approaches:

**a**. To apply the model in a 'dynamic planning' mode (as described in Watling and Cantarella (2015)), whereby the model-user needs to define both the initial point and number of days over which to run the model. In this case, both the initial point and the number of days since the initial condition are parameters that affect the prediction of the transport system, as are the assumed day-to-day dynamic adjustment process of travellers.

**b**. To aim to derive a result guaranteeing that the system will converge to the stationary distribution regardless of the initial point *and* within a practically reasonable period of time.

In the present paper, we will adopt the second of these approaches to resolve Issue B, i.e. to identify whether the stationary distribution is practically acceptable as the representative state of the transport system in which users travel recurrently, even after any kind of external shocks such as a network improvement or a major disruption changes conditions of the transport system. In other words, we wish to identify how long it takes for a transport system to 'settle down' after any arbitrary shock, or at least we wish to assess whether it can be bounded by a practically acceptable upper limit. We specifically will aim to estimate the *Markov Chain Mixing Time (MCMT)*, which is a worst-case assessment of the convergence time, and to study how it is affected by parameters

of the system. The MCMT '*measures the time required by a Markov chain for the distance to stationarity to be small*' (Levin and Peres, 2017).

We have two particular aims in the analysis of the MCMT. First, we wish to develop techniques to compute MCMTs, or at least bounds thereon, that are applicable for analysing large-scale transport systems. To that end, we observe that there is a well-developed theory on convergence times for Markov chains based on the second largest eigenvalue modulus of the transition matrix (Sinclair, 1992). However, the direct application of this mathematical machinery is infeasible for all but the smallest toy examples. As a consequence, we need to explore a combination of theoretical and simulation methods. The main tools we adopt are the *coupling* and *aggregation*. Coupling is known as a powerful tool to find an upper bound of MCMTs (Levin and Peres, 2017). Aggregation is a new concept (in the context of convergence times for day-to-day traffic models) that enables us to analyse MCMTs in large-scale transport systems. It reduces the number of states to consider when deriving the MCMT by bundling the states of the transport system, which are characterised by the choice behaviour of each individual traveller. The number of states of the transport system is equal to the number of choices that a user has multiplied for all users. This means that the number of states increases exponentially with respect to the number of users. The proposed aggregation approach avoids this problem and brings the possibility of applying the proposed method to large transport systems.

Our second, related aim, is to conduct a *preliminary* examination of the relationship between MCMTs – or their upper bounds – and critical properties of the system, such as travellers' sensitivity to differences in travel cost and frequency of travellers' revisions of their choices, through analytical and numerical assessments. Thus far, most existing studies dealing with the behaviour of day-to-day adjustment processes have focused on whether the dynamics of transport systems converge to a stable point or not, and very little attention has been made how long it takes. A study by Chien and Sinclair (2011) calculated the number of updates necessary to get from an arbitrary initial state to the $\varepsilon$-Nash equilibrium, and showed that in a simple congestion game, which is identical to a standard static traffic assignment problem with a single origin-destination pair, the number of updates to converge to an $\varepsilon$-Nash equilibrium is $O(n\log(n))$, where $n$ is the number of travellers. Regarding studies of MCMTs, Levin and Peres (2017) showed that the MCMT of the random walk on a hypercube $(0,1)^n$, which is regarded as a $n$ users' binomial choices with constant choice probabilities, is upper bounded by $O(n\log(n))$. While these outputs give us insights into the convergence time, they are based on simple models. More importantly, these studies are

based on the assumption that only a single user changes behaviour every day, which is not a realistic setup. The present study aims to add more practically useful insights to the MCMTs and their upper bounds. Owing to the analytical intractability, we have to resort to numerical simulations, and the generality of the findings will be limited. Nevertheless, these outputs should still give us useful information on the MCMTs.

The remainder of the paper is structured as follows. The backgrounds and motivations of this study were explained in this introduction section. Section 2 explains the model specification. Section 3 introduces the concept of aggregation. Section 4 introduces definitions of MCMTs. Section 5 explains the calculation methodologies of MCMTs. Section 6 investigates the relationships between MCMTs and critical properties of the system, as well as exhibits the performance of the proposed methodology, through an analytical assessment and numerical simulations of four example cases. Section 7 concludes the paper and suggests future directions.

# 2 Model Specification

Table 1 shows key notations used throughout the present paper, including Sections 2 and onwards.

## 2.1 Users' choices and travel costs in a transport system

Consider a transport system consisting of atomic users, in which the set of users is denoted by $I$ and the number of users is denoted by $n$. The choice of user $i$ is denoted by $c_i$, and the choice set of traveller $i \in I$ is denoted by $C_i$. The vector of user choices is $\mathbf{c} = (c_i)_{i \in I}$. The feasible set of $\mathbf{c}$ is denoted by $C$, i.e. $\mathbf{c} \in C$. $\mathbf{c}$ represents the state of the transport system and any properties of the system are uniquely determined once $\mathbf{c}$ is uniquely set. We call $\mathbf{c}$ the *disaggregated state*.

Users update their choices considering travel costs that were realised on the previous day. Let the disaggregated state of the transport system on the previous day be $\mathbf{c}^{\mathrm{prev}}$. Then, user $i$ considers that his/her travel cost of $c \in C_i$ will be $\pi_c(\mathbf{c}^{\mathrm{prev}})$ when updating his/her current choice.

Travel costs are the same for all users if $c$ is the same. When heterogeneity of users' preferences is considered, we use different $c$ for different users. For example, when users 1 and 2 can choose either car travel or a train ride, and they depart from different origins, we need to define e.g. $C_1 = \{\text{car1}, \text{train1}\}$ and $C_2 = \{\text{car2}, \text{train2}\}$ so that their choices are distinguished.

Table 1: Key notations

| Notation | Definition |
|---|---|
| $I$, $n$ | Set of users and number of users |
| $c_i$, $C_i$ | Choice and choice set of user $i \in I$ |
| $\mathbf{c}$ | Disaggregate state, which is a vector form of all users' choices, $\mathbf{c} = (c_i)_{i \in I}$ |
| $C$ | Set of all disaggregate states |
| $\pi_c(\mathbf{c})$ | Travel cost of choice $c \in C_i$ when the state is $\mathbf{c}$ |
| $t, r$ | Day and probability that a user revisits his/her current choice on a day |
| $A$ | Aggregation, a set consisting of aggregated states |
| $A^+ < A$ | Inequality $<$ indicates that $A^+$ is an aggregation of aggregation $A$. |
| $A(a^+)$ | Operator extracting all aggregated states in $A$ which are covered by $a^+ \in A^+$ |
| $a(\mathbf{c};A)$ | Operator finding $a \in A$ that contains $\mathbf{c}$ |
| $G$, $I^g$, $C^g$ | Set of groups, set of users in $g \in G$, and choice set of users in $I^g$ |
| $x_c$,$\mathbf{x}$ | Number of users choosing choice $c \in C^g$ and its vector form; $\mathbf{x} = (x_c)_{c \in C^g, g \in G}$ |
| $\mathbf{c}^0$, $a^0$ | Initial state of the Markov chain, in disaggregated and level 1 aggregated forms. |
| $\mathbf{c}(t;\mathbf{c}^0)$ | Disaggregated state on day $t$ |
| $p(\mathbf{c},t;\mathbf{c}^0)$ | Probability that state $\mathbf{c} \in C$ is realised on day $t$ |
| $p(a,t;\mathbf{c}^0)$ | Probability that aggregated state $a \in A$ is realised on day $t$ |
| $\mathbf{p}_C(t;\mathbf{c}^0)$ | Vector form of $p(\mathbf{c},t;\mathbf{c}^0)$; $\mathbf{p}_C(t;\mathbf{c}^0) = (p(\mathbf{c},t;\mathbf{c}^0))_{\mathbf{c} \in C}$ |
| $\mathbf{p}_A(t;\mathbf{c}^0)$ | Vector form of $p(a,t;\mathbf{c}^0)$; $\mathbf{p}_A(t;\mathbf{c}^0) = (p(a,t;\mathbf{c}^0))_{a \in A}$ |
| $\mathbf{p}^*_C$, $\mathbf{p}^*_A$ | Probability distribution of the stationary state on $C$ and $A$. |
| $\|\mathbf{p}^1_\square - \mathbf{p}^2_\square\|_{\mathrm{TV}}$ | TV distance between probability distributions $\mathbf{p}^1_\square$ and $\mathbf{p}^2_\square$ |
| $t^*_\square$ | o-MCMT (*o* stands for *original*) |
| $t^*_\square(\mathbf{c}^0)$ | si-MCMT (*si* stands for *specific initial point*) |
| $t^*_\square(\mathbf{c}^0_1, \mathbf{c}^0_2)$ | ti-MCMT (*ti* stands for *two specific initial points*) |

Note: $\square$ is replaced with $C$ or $A$. When $\square = A$, prefix 'ag-' is added to the name of the MCMT, e.g. ag-o-MCMT ('ag' stands for 'aggregated').

## 2.2 Markov-chain based day-to-day adjustment model

We need to describe the day-to-day behaviour adjustment process of users who travel in the transport system recurrently every day because, as stated in Section 1, we wish to evaluate how long it takes for the system to 'settle down' after any arbitrary shock. We describe this process as a Markov chain in which $\mathbf{c} \in C$ corresponds to each state. While the characteristics of the transport system in the real world may change over days, we simply assume that they are completely stationary over days *after the shock occurred* to focus on the effect of the users' behavioural adjustment process. Therefore, we first assume that the choice sets and travel cost functions do not change over days and all users in $I$ travel recurrently everyday.

We now model how users update their choices as follows, which is called the *update rule*:

**Definition.** On each day, each user revisits his/her current choice in the given and constant probability, which is called *update probability*, denoted by $r$.

While $r$ may depend on the situation of the system and users in the real world, e.g. it may be higher when people are keen to find their best choices after a major change in the transport system, we assume that it is constant and unique for all users for simplicity.

We first denote the states and their probabilities in the Markov chain. Let $t \geq 0$ be the day and $\mathbf{c}(t;\mathbf{c}^0) = (c_i(t;\mathbf{c}^0))_{i\in C_i}$ be the state on day $t$, where $c_i(t;\mathbf{c}^0)$ is user $i$'s choice on day $t$ and $\mathbf{c}^0 = (c_i^0)_{i\in C_i}$ is the initial state of the transport system on $t = 0$ (the epoch day). We consider that the shock was applied to the transport system only once, just before the epoch day, and the initial state reflects its effect on users' choices. Note that $\mathbf{c}^0$ does not necessarily have to be strictly defined in calculating the MCMT. Instead, we can set a number of possible states (or all states in $C$) as initial states to find the worst-case assessment of the convergence time, as explained later in Section 4. As the transitions are stochastic, $\mathbf{c}(t;\mathbf{c}^0)$ is a random variable that follows the probability distribution denoted by $p(\mathbf{c},t;\mathbf{c}^0) = \Pr(\mathbf{c}(t;\mathbf{c}^0) = \mathbf{c})$. We introduce the vector form that combines the probabilities of all states. Let $p(\mathbf{c})$ be a probability of state $\mathbf{c} \in C$. Its vector form is denoted by $\mathbf{p}_C = (p(\mathbf{c}))_{\mathbf{c}\in C}$. Note that the order of elements in $\mathbf{p}_C$ is based on the order of the states in $C$, which can be arbitrarily defined but must be consistent in all calculations. The vector form of $p(\mathbf{c},t;\mathbf{c}^0)$ is denoted by $\mathbf{p}_C(t;\mathbf{c}^0)$. The vector representation of probabilities in this manner is used throughout the present paper.

The perturbed best response dynamics (Hofbauer and Sandholm, 2007) is employed to construct the Markov chain model describing the users' behaviour adjustment process. Let

$$K_c(\mathbf{c}^{\text{prev}}) = \frac{\exp\{-\theta\,\pi_c(\mathbf{c}^{\text{prev}})\}}{\sum_{d\in C_i}\exp\{-\theta\,\pi_d(\mathbf{c}^{\text{prev}})\}} \tag{1}$$

be the probability that user $i$ changes his/her choice to $c$ conditional on the fact that he/she reconsiders his/her route, when the previous day's state is $\mathbf{c}^{\text{prev}}$ and $\theta > 0$ is the logit parameter. Combined with the update probability $r$, we have the following transition rule:

$$c_i(t+1;\mathbf{c}^0) = \begin{cases} c_i^K(\mathbf{c}(t;\mathbf{c}^0)) & \text{if } r' \leq r \\ c_i(t;\mathbf{c}^0) & \text{otherwise,} \end{cases} \tag{2}$$

where $c_i^K(\mathbf{c}^{\text{prev}})$ is a choice that is randomly drawn from $C_i$ in the probability $K_c(\mathbf{c}^{\text{prev}})$ and $r'$ is

a uniform random number between 0 and 1. To explicitly denote the random draw from $C_i$, we consider other uniform random values $r_i^*$ $(0 \leq r_i^* \leq 1)$ for user $i \in I$, which is independent of $r'$ and $r_i^*$ of other users. In addition, we set a certain order of the choices in $C_i$ and let $z^i(\zeta)$ be the $\zeta$th choice in $C_i$, where $\zeta = 1, \ldots, |C_i|$ and $|C_i|$ is the number of choices. Using this, we define the cumulative distribution of $K_c^i(\mathbf{c}^{\text{prev}})$, denoted by $Z_\zeta^i(\mathbf{c}^{\text{prev}})$, as

$$Z_0^i(\mathbf{c}^{\text{prev}}) = 0 \ \text{ and } \ Z_\zeta^i(\mathbf{c}^{\text{prev}}) = \sum_{\zeta'=1}^{\zeta} K_{z^i(\zeta')}(\mathbf{c}^{\text{prev}}). \tag{3}$$

Using them, we define $c_i^K(\mathbf{c}^{\text{prev}})$ as

$$c_i^K(\mathbf{c}^{\text{prev}}) = z^i(\zeta) \quad \text{where } \ \zeta \ \text{ satisfies } \ Z_{\zeta-1}^i(\mathbf{c}^{\text{prev}}) \leq r_i^* < Z_\zeta^i(\mathbf{c}^{\text{prev}}). \tag{4}$$

The transition rule introduced above also invokes the formula calculating probabilities, which is

$$p(\mathbf{c}, t+1; \mathbf{c}^0) = \sum_{\mathbf{c}' \in C} q(\mathbf{c}', \mathbf{c}) p(\mathbf{c}', t; \mathbf{c}^0), \tag{5}$$

where $q(\mathbf{c}', \mathbf{c})$ are the entries of the transient matrix of the Markov chain, denoted by $Q_C$, and

$$q(\mathbf{c}', \mathbf{c}) = \prod_{i \in I} \left\{ r K_{c_i}(\mathbf{c}') + (1-r)\delta(c_i, c_i') \right\}, \tag{6}$$

where $\delta(c_i, c_i')$ is a Kronecker delta, which is 1 if $c_i = c_i'$ and 0 otherwise. Using $Q_C$, we have

$$\mathbf{p}_C(t; \mathbf{c}^0) = Q_C^t \mathbf{p}_C^0, \ \text{ where } \ \mathbf{p}_C^0 = (p^0(\mathbf{c}))_{\mathbf{c} \in C} \ \text{ and } \ p^0(\mathbf{c}) = \begin{cases} 1 & \text{if } \ \mathbf{c} = \mathbf{c}^0 \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

The non-zero choice probabilities of the perturbed best response dynamics (i.e. $K_c(\mathbf{c}) > 0$) implies the existence of unique stationary distribution of the Markov chain because both the irreducible and aperiodic properties hold. Let $\mathbf{p}_C^*$ be the stationary distribution. Then, we have

$$\mathbf{p}_C^* = Q_C \mathbf{p}_C^*. \tag{8}$$

The following relationship is known as a general property of the irreducible and aperiodic Markov chain:

$$\lim_{t \to \infty} \mathbf{p}_C(t; \mathbf{c}^0) = \mathbf{p}_C^* \ \ \forall \mathbf{c}^0 \in C. \tag{9}$$

# 3 Aggregation

## 3.1 Aim of aggregation and its definition

The states of the Markov chain describing the day-to-day dynamics of a transport system are characterised by the choices $c_i$ of all users. Therefore, the number of its states is the product of $|C_i|$ of all users, which grows exponentially with respect to the number of users. This property is harmful when calculating the MCMTs because their definitions include $\mathbf{p}_C(t;\mathbf{c}^0)$ as later explained in Section 4, and the size of this vector is the same as the number of states in the Markov chain.

To reduce the number of states and avoid the aforementioned issue, we now introduce the concept called *aggregation*. The definition of the aggregation is very simple, as stated below:

**Definition.** Consider a partition of set $C$, denoted by $A$. Then $A$ is called an *aggregation* of $C$, and elements in $A$ are called *aggregated states*.

We use the notation $a(\mathbf{c};A)$ to specify $a \in A$ containing $\mathbf{c}$.

As an example, let us consider $C = \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4\}$. Then, $A = \{a_1, a_2\} = \{\{\mathbf{c}_1, \mathbf{c}_2\}, \{\mathbf{c}_3, \mathbf{c}_4\}\}$ is an aggregation of $C$, in which $\mathbf{c}_1$ and $\mathbf{c}_2$ are aggregated into $a_1$ and $\mathbf{c}_3$ and $\mathbf{c}_4$ are aggregated into $a_2$. A different aggregation pattern for the same $C$ is possible, such as $A' = \{\{\mathbf{c}_1\}, \{\mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4\}\}$.

It should be noted that the aggregation introduced here is not intended to alter the dynamics of the transport system itself. Regardless of whether or how aggregation is conducted, the Markov chain for $\mathbf{c}(t;\mathbf{c}^0)$ defined in Section 2 represents the dynamics of the transport system. Aggregation is only intended to *describe* the probability distribution of the Markov chain with a smaller number of states. However, for a certain class of aggregation, which will be introduced in the next section as *level 1 aggregation*, it is possible to describe the dynamics for the aggregated states while keeping the above setting.

## 3.2 Level 1 / level 2 aggregations

We classify any possible aggregations of $C$ into two categories. i.e. *level 1 aggregation* and *level 2 aggregation*. To define them, we first consider the *groups* of users. The groups are defined so that

users in the same group have the same choice sets, i.e.

$$C_i = C_j = C^g \ \ \forall i, j \in I^g, \ \ \forall g \in G. \tag{10}$$

where $G$ is the set of groups, $I^g$ is the set of users in group $g$, and $C^g$ is the choice set for group $g$. Let $x_c(\mathbf{c})$ be the number of users choosing choice $c \in C^g$ when the state is $\mathbf{c}$. Its vector form is $\mathbf{x}(\mathbf{c}) = (x_c)_{c \in C^g, g \in G}$. Then, level 1 and level 2 aggregations are defined as follows:

**Definition.** $A$ is a level 1 aggregation if and only if the following two equations hold.

$$\mathbf{x}(\mathbf{c}) = \mathbf{x}(\mathbf{c}') \Leftrightarrow a(\mathbf{c}; A) = a(\mathbf{c}'; A) \ \ \forall \mathbf{c}, \mathbf{c}' \in C, \tag{11}$$

$$\pi_c(\mathbf{c}) = \pi_c(\mathbf{c}') \ \ \forall \mathbf{c}, \mathbf{c}' \in a, \forall a \in A, \forall c \in C^g, \forall g \in G. \tag{12}$$

Otherwise, $A$ is a level 2 aggregation.

The level 1 aggregation characterises the state of the transport system in terms of the number of users making each choice. This is a popular practice in many transport models, especially in traditional traffic assignment models, and the level 1 aggregation can be seen as a redefinition of this practice in the context of the present study. It should be noted, however, that this practice may not be universally applicable to all traffic models. For example, in traffic flow simulations, different departure times may be exogenously given to each user. In this case, even if the origin of every user is the same, the route travel time is different for different departure times, so it is not possible to combine several users in the same group owing to Equation (12). Such situations in which the level 1 aggregation cannot be applied are not uncommon. Of course, approximate approaches such as aggregating users with similar departure times into one group are possible, but in this case, it is no longer level 1 aggregation but level 2 aggregation.

The level 1 aggregation distinguishes which option in a group has been chosen by how many travellers, but not who selected it. It is possible to define level 1 aggregation on the basis of this property. Its formal definition and the equivalence between it and the aforementioned definition is given in Appendix 1.

We lastly define a special state called a *corner state* for the level 1 aggregation, in which all users in the same group choose the same choice:

**Definition.** The *corner state* is a disaggregated state $\mathbf{c}$ satisfying

$$c_i = c_j \ \ \forall i, j \in I^g \ \text{ and } \ \forall g \in G. \tag{13}$$

Note that any aggregated state including a corner state includes no other disaggregated state in it, namely $a(\mathbf{c};A) = \{\mathbf{c}\}$ if $\mathbf{c}$ is a corner state.

## 3.3 Aggregation of aggregated states

An aggregation can also be constructed by aggregating aggregated states. This operation is particularly useful when the level 1 aggregation is further aggregated to reduce the number of states by, e.g. truncating the lower digits of the numbers of a traffic volume. The following definition formalises this operation:

**Definition.** Consider aggregations denoted by $A$ and $A^+ \neq A$. If any element in $A^+$ can be obtained by combining element(s) in $A$, $A^+$ is called an aggregation of $A$.

This relationship is denoted by the inequality sign, i.e. $A^+ < A$ implies that $A^+$ is an aggregation of $A$. We also define an operator that extracts aggregated states covered by $a^+ \in A^+$ as follows:

$$A(a^+) = \{a | a^+ \cup a \neq \emptyset, a \in A\}. \tag{14}$$

## 3.4 An example of level 1 and level 2 aggregations

We now summarise the aggregation and related items with a simple example. Consider two choices, denoted by $\circ$ and $\bullet$, for 8 users, labelled as $1 \ldots 8$. All users are in the same group. This transport system has $2^8 = 256$ disaggregated states. Each state is denoted by specifying all users' choices, e.g. $(\circ\circ\circ\circ\circ\circ\circ\circ)$ and $(\bullet\bullet\circ\bullet\circ\circ\bullet\circ)$. The level 1 aggregation is performed using the number of users who choose choices $\circ$ and $\bullet$, denoted by $\mathbf{x} = (x_\circ, x_\bullet)$, where $x_\circ$ and $x_\bullet$ are the number of users choosing $\circ$ and $\bullet$, respectively, and $x_\circ + x_\bullet = 8$. We have 9 states in the level 1 aggregation $A$, namely $(x_\circ, x_\bullet)$ = (0,8),(1,7),...,(8,0). They are denoted by $a_0, a_1, \ldots, a_8$, respectively, and hence $A = \{a_0, \ldots, a_8\}$. We further consider a level 2 aggregation $A^+$ by aggregating $\{a_0, \ldots, a_2\}$, $\{a_3, \ldots, a_5\}$, $\{a_6, \ldots, a_8\}$ into three states denoted by $a^+_{0-2}$, $a^+_{3-5}$, $a^+_{6-8}$, respectively, which leads to $A^+ = \{a^+_{0-2}, a^+_{3-5}, a^+_{6-8}\}$.

Several examples of relationships are shown as follows:

1. Each element in $A$, namely $a_0,\ldots,a_8$, is a set including disaggregate states, for example, $a_0=\{(\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet)\}$ and $a_1=\{(\circ\bullet\bullet\bullet\bullet\bullet\bullet\bullet),(\bullet\circ\bullet\bullet\bullet\bullet\bullet\bullet),\ldots,(\bullet\bullet\bullet\bullet\bullet\bullet\bullet\circ)\}$.
2. Every element in $A$ is characterised by the number of users choosing $\circ$ or $\bullet$, e.g. $\mathbf{x}((\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet))=(0,8)$ and $\mathbf{x}((\bullet\circ\bullet\circ\bullet\circ\bullet\bullet))=(3,5)$.
3. Every element in $A^+$ is the union set of corresponding elements in $A$, e.g. $a^+_{0-2}=a_0\cup a_1\cup a_2$.
4. Function $a(\mathbf{c};A)$ returns an element in $A$ including $\mathbf{c}$, e.g. $a((\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet);A)=a_0$ and $a((\bullet\circ\bullet\circ\bullet\circ\bullet\bullet);A)=a_3$.
5. Function $A(a^+)$ returns a set of elements in $A$ corresponding to $a^+$, e.g. $A(a^+_{0-2})=\{a_0,a_1,a_2\}$.

We can also find two corner states in this case, namely $(\bullet\bullet\bullet\bullet\bullet\bullet\bullet\bullet)$ and $(\circ\circ\circ\circ\circ\circ\circ\circ)$.

## 3.5 Markov chain model for level 1 aggregation

The Markov-chain-based day-to-day dynamical model defined in Section 2.2 may be converted to a Markov chain model whose states are level 1 aggregated states. To facilitate it, we first define the sum of probabilities of the states in set $S\subseteq C$, called *aggregated probabilities*, as

$$p(S)=\sum_{\mathbf{c}\in S}p(\mathbf{c}), \tag{15}$$

where the probability distribution on $C$ is $p(\mathbf{c})$. It implies that, when the argument of the probability distribution function is a set, it returns the sum of probabilities of all states in the set.

To describe the level-1-aggregation-based dynamics, let $a(t;a^0)$ be an aggregated state on day $t$, where $a^0$ is the initial aggregated state and $\mathbf{p}_A(t;a^0)=(p(a,t;a^0))_{a\in A}$. Then, the following theorem holds, which guarantees the existence of the Markov chain describing the dynamics of the probability distribution of the level 1 aggregated states:

**Theorem 1.** Consider the set of disaggregated states $C$. Consider its level 1 aggregation denoted by $A$. Then, there exists the matrix, denoted by $Q_A$, that satisfies

$$\mathbf{p}_A(t;a^0)=Q_A^t\mathbf{p}_A^0,\ \text{where}\ \mathbf{p}_A^0=(p'(a))_{a\in A}\ \text{and}\ p'(a)=\begin{cases}1 & \text{if } a=a^0\\ 0 & \text{otherwise}\end{cases} \tag{16}$$

all all times.

*Proof.* See Appendix 2. □

The stationary distribution for this Markov chain is denoted by $\mathbf{p}_A^*$, which is defined by

$$\mathbf{p}_A^* = Q_A \mathbf{p}_A^*. \tag{17}$$

Finally, we define the *flat probability condition* and state a theorem and its corollary. This will be used to investigate the difference in the MCMTs in disaggregated Markov Chains and level 1 aggregated Markov Chains in Section 4. The flat probability condition is defined as follows:

**Definition.** Consider a level 1 aggregation denoted by $A$. Then, the following condition is defined as the *flat probability condition* on day $t$:

$$p(\mathbf{c}) = p(\mathbf{c}') \ \ \forall \mathbf{c}, \mathbf{c}' \in a, \ \ a \in \forall A, \tag{18}$$

which implies that the probabilities of all states in the same aggregated state are the same.

Then, the following theorem and its corollary hold:

**Theorem 2.** Consider a disaggregated Markov Chain. If the flat probability condition holds on day $t$, it also holds for all days after $t$.

**Corollary 2.1.** The flat probability condition holds in the stationary distribution.

*Proof.* See Appendix 3. □

# 4 Definitions of MCMT and their extensions

The MCMT and its extensions are defined in this section. We first define three types of MCMTs, namely *o-MCMT*, *si-MCMT*, and *ti-MCMT*, for disaggregated states, and then extend these MCMTs for aggregated states. The aggregated version of MCMTs are denoted by *ag-X-MCMT*, where X is replaced by either one of 'o', 'si', or 'ti'.

The difference of the o-MCMT and si-MCMT is made by the setting of initial states. The MCMT is defined as the time when the transient state is sufficiently close to the stationary state; therefore we need to set the initial state of the transient state. The original MCMT appearing in the literature such as Levin and Peres (2017) is defined so that it represents the 'worst' case, i.e. the longest time for all possible initial states. We call it o-MCMT. This setting is basically suitable for the aim of this study, but calculating MCMTs for all initial states can be impossible when the size of the problem is large. In such a case, we need to select a limited number of initial states. The si-MCMT, which is the MCMT with a specific initial point given, is used for this purpose. The si-MCMT is also useful for problems in which the initial point can be externally given.

The stationary distribution must be calculated when finding o-MCMT and si-MCMT, while it is sometimes impossible owing to the huge computational cost. The ti-MCMT is useful to address this issue because it does not require the stationary distribution. The ti-MCMT is defined as the time when two Markov chains starting from different initial states become close to each other.

Finding MCMTs for disaggregated states require the calculation of probabilities of all states. This is often impossible in the transport context, as the number of states is $2^n$ even for the binary choice case. The aggregated version of MCMTs uses $\mathbf{p}_A$ instead of $\mathbf{p}_C$ to avoid this issue. The aggregation reduced the number of states dramatically, even by level 1 aggregation. For instance, it decreases from $2^n$ to $n$ in the binary choice case mentioned above. We can also employ level 2 aggregation to gain more reduction in the number of states in a large-scale transport system.

One concern regarding ag-X-MCMTs would be how the aggregation affects them, i.e. how they are different from the disaggregated MCMT. We investigate this issue in this section and show that the impact is not necessarily high.

## 4.1 Total variation (TV) distance

Measuring the distance between two probability distributions is necessary to define the MCMTs. Among a number of definitions of such distances, the *total variation* distance (or TV distance in short) is convenient to state many theorems for calculating MCMTs and is mostly used for calculating the MCMTs in the literature. It is defined as follows:

**Definition.** Consider $C$ and its aggregation $A$. Define two probability distributions, denoted by $\mathbf{p}_C^1$ and $\mathbf{p}_C^2$ for $C$ and $\mathbf{p}_A^1$ and $\mathbf{p}_A^2$ for $A$. Then, the TV distances on $C$ and $A$ between $\mathbf{p}_C^1$ and $\mathbf{p}_C^2$ (for $C$)

and between $\mathbf{p}_A^1$ and $\mathbf{p}_A^2$ (for $A$) are defined as:

$$\|\mathbf{p}_C^1 - \mathbf{p}_C^2\|_{\text{TV}} = \max_{C' \subset C} \left| \sum_{\mathbf{c} \in C'} \{p_1(\mathbf{c}) - p_2(\mathbf{c})\} \right| \text{ and } \|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\text{TV}} = \max_{A' \subset A} \left| \sum_{a \in A'} \{p_1(a) - p_2(a)\} \right| \tag{19}$$

The TV distances can also be calculated by the following equations:

$$\|\mathbf{p}_C^1 - \mathbf{p}_C^2\|_{\text{TV}} = \frac{1}{2} \sum_{\mathbf{c} \in C} |p_1(\mathbf{c}) - p_2(\mathbf{c})| \text{ and } \|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\text{TV}} = \frac{1}{2} \sum_{a \in A} |p_1(a) - p_2(a)|. \tag{20}$$

See Levin and Peres (2017) for the proof of this equivalency.

## 4.2 o-MCMT: Original version of MCMT

To define the o-MCMT, we first define the *o-TV distance*, which is the maximum TV distance for all initial states $\mathbf{c}^0 \in C$ so that we can describe the 'worst' dissimilarity of the Markov chain from the stationary distribution without specifying $\mathbf{c}^0$. It is defined as

$$d_C^{\max}(t) = \max_{\mathbf{c}^0 \in C} \|\mathbf{p}_C(t; \mathbf{c}^0) - \mathbf{p}_C^*\|_{\text{TV}} \tag{21}$$

The smaller $d_C^{\max}(t)$ implies that the two probability distributions are similar. Therefore, we can set a threshold to check whether $\mathbf{p}_C(t; \mathbf{c}^0)$ is sufficiently close to the stationary distribution, i.e. $\mathbf{p}_C^*$, for all $\mathbf{c}^0 \in C$. Let the threshold be 0.25, as often employed in the literature (e.g. Levin and Peres (2017); see Appendix 4 for details). We then define the original version of the MCMT as follows:

**Definition.** The *o-MCMT*, denoted by $t_C^*$, is defined as the time at which the system is sufficiently close to the stationary distribution. This is formally defined by the following equation :

$$t_C^* = t_{\max}^{0.25}(d_C^{\max}(t)), \text{ where } t_{\max}^{\Theta}(f(t)) = \max\{t | f(t-1) \geq \Theta \text{ and } f(t) < \Theta\} \tag{22}$$

The max operator in Equation (22) may be replaced with min operator, and both give the same result when $d_C^{\max}(t)$ is monotonically decreasing with respect to $t$. However, this property does not necessarily hold for level 2 aggregations, which will be considered in Section 4.5. Considering that it is desirable in practice that TV distances are always below the threshold after the MCMT, it

is preferable to use max in the definition for aggregated states.

## 4.3 si-MCMT: MCMT with a specific initial point

The si-MCMT, denoted by $t_C^*(\mathbf{c}^0)$, is the MCMT with a specific initial state given. The definition is a simple variation of the o-MCMT, which is shown as follows:

**Definition.** The *si-MCMT* for the initial state $\mathbf{c}^0$, denoted by $t_C^*(\mathbf{c}^0)$, is defined as

$$t_C^*(\mathbf{c}^0) = t_{\max}^{0.25}(\|\mathbf{p}_C(t;\mathbf{c}^0) - \mathbf{p}_C^*\|_{\mathrm{TV}}) \tag{23}$$

We refer $\|\mathbf{p}_C(t;\mathbf{c}^0) - \mathbf{p}_C^*\|_{\mathrm{TV}}$ to as *si-TV distance*. The si-MCMT is a convenient replacement of the o-MCMT when the transient time from a known initial point to the stationary distribution is investigated or calculating the TV distances for all $\mathbf{c}^0$ in $C$ is not feasible. The second case is expected to happen when we calculate the MCMT by a numerical simulation, especially for a large-scale transport system, which consumes a large amount of calculation time. In such a case, choosing one or a few initial states that is expected to be far from the stationary distribution is practically acceptable to obtain a good approximation of the MCMT with a reasonable calculation time. For example, we may choose a state in which all users choose apparently costly choices (e.g. travelling at midnight in a departure time choice problem).

## 4.4 ti-MCMT: MCMT with two specific initial points

The ti-MCMT, denoted by $t^*(\mathbf{c}_1^0, \mathbf{c}_2^0)$, is the extension of the si-MCMT, which is defined as the TV distance between two Markov chains with different initial points. It is defined as follows:

**Definition.** The *ti-MCMT* for the initial states $\mathbf{c}_1^0$ and $\mathbf{c}_2^0$, denoted by $t_C^*(\mathbf{c}_1^0, \mathbf{c}_2^0)$, is defined as

$$t_C^*(\mathbf{c}_1^0, \mathbf{c}_2^0) = t_{\max}^{0.25}(\|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}_C(t;\mathbf{c}_2^0)\|_{\mathrm{TV}}) \tag{24}$$

We refer $\|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}_C(t;\mathbf{c}_2^0)\|_{\mathrm{TV}}$ to as *ti-TV distance*. The ti-MCMT indicates how many days it takes for the dependency of the initial state of two day-to-day dynamics starting from different initial values to decrease below a threshold specified by the TV distance. This can be interpreted as the number of days it takes for the initial state dependency to disappear to a certain level.

The ti-MCMT is mainly used to estimate the upper limit of the si-MCMT without knowing the stationary distribution. Calculating the stationary distribution is hard when there are huge numbers of states. The stationary distribution can be obtained by a method such as (a) finding the solution of Equation (17) or (b) performing the Monte Carlo simulation of the Markov chain. If the number of the states is huge, the size of the transient matrix is too big to solve Equation (17). In (b), the sample size necessary to obtain the stationary distribution with good accuracy can be huge.

Followings are the theorem/corollary facilitating the aforementioned application of ti-MCMT:

**Theorem 3.** The following inequality holds for all $\mathbf{c}_1^0 \in C$.

$$\|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}_C^*\|_{\mathrm{TV}} \leq \max_{\mathbf{c}_2^0 \in C} \|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}_C(t;\mathbf{c}_2^0)\|_{\mathrm{TV}} \leq 2\|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}_C^*\|_{\mathrm{TV}}. \tag{25}$$

**Corollary 3.1.**

$$t_C^*(\mathbf{c}_1^0) \leq \max_{\mathbf{c}_2^0 \in C} t_C^*(\mathbf{c}_1^0, \mathbf{c}_2^0) \quad \text{and} \quad t_C^* \leq \max_{\mathbf{c}_1^0 \in C, \mathbf{c}_2^0 \in C} t_C^*(\mathbf{c}_1^0, \mathbf{c}_2^0) \tag{26}$$

*Proof.* See Levin and Peres (2017) (p.53; Lemma 4.10). □

We used the first inequality of (25) to state the corollary. Its second inequality may be used to find a lower bound of the si-MCMT. Because our main interest is in the upper bound, we do not discuss its utilisation in the present paper.

We need to evaluate the TV distance for all $\mathbf{c}_2^0$ to apply Theorem 3 and Corollary 3.1. In practice, we may employ one or a few $\mathbf{c}_2^0$ which are considered to be sufficiently far from $\mathbf{c}_1^0$ to have an approximated value of the middle term of Inequality (25). When we estimate the upper bound of $t_C^*$, we may employ a pair of $\mathbf{c}_1^0, \mathbf{c}_2^0$ that are far from each other to have a good approximation.

## 4.5 ag-X-MCMT: Aggregated MCMTs

### 4.5.1 Definition

The *ag-X-MCMTs* are defined by replacing $C$ with $A$. They are called ag-X-MCMT, namely ag-o-MCMT, ag-si-MCMT, ag-ti-MCMT, denoted by $t_A^*$, $t_A^*(\mathbf{c}^0)$, and $t_A^*(\mathbf{c}_1^0, \mathbf{c}_2^0)$, respectively. The corresponding TV distances are called ag-o, ag-si, and ag-ti TV distances, respectively.

If there is a level 1 aggregation for $C$, denoted by $A'$, the initial states, i.e. $\mathbf{c}^0$, $\mathbf{c}_1^0$, and $\mathbf{c}_2^0$ can be replaced by $a^0$, $a_1^0$, and $a_2^0$, respectively, in which all of them are in $A'$. Note that $A'$ does not have to be equal to $A$. For example, we can formulate the Markov-chain dynamical model based on the level 1 aggregation and define an ag-X-MCMT using another aggregation $A$ such that $A < A'$.

There are a few benefits to using ag-X-MCMT. First, the similarity of two probability distributions can be measured on the basis of aggregate variables such as traffic volume or travel times, in which most analysts of transport systems are more interested than the detailed choices of individual users. In addition, numerical calculations become feasible for large-scale cases, in which the number of disaggregated states is huge. Both benefits support that the level 1 aggregation should be considered at least whenever it is applicable. Further aggregations by level 2 aggregation should also be considered for a large-scale transport system, which is of interest to practitioners.

### 4.5.2 Effect on TV distances and MCMTs by aggregation

The following theorem states a basic inequality of TV distances by different levels of aggregations:

**Theorem 4.** Consider aggregations $A$ and $A^+$ such that $A^+ < A$. Let $\mathbf{p}_A^1$ and $\mathbf{p}_A^2$ be the two probability distributions on $A$. Then, we have

$$\|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\mathrm{TV}} \geq \|\mathbf{p}_{A^+}^1 - \mathbf{p}_{A^+}^2\|_{\mathrm{TV}} \text{ and } \|\mathbf{p}_C^1 - \mathbf{p}_C^2\|_{\mathrm{TV}} \geq \|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\mathrm{TV}}, \tag{27}$$

i.e. the aggregation results in a reduced or unchanged TV distance.

*Proof.* See Appendix 5. □

Theorem 4 still leaves room for the possibility that the distance can be significantly reduced by the aggregation. Such impact on the TV distances and MCMTs by the aggregation should be a major interest of analysts who need to decide how to aggregate states when calculating MCMTs. Actually, this impact is not necessarily high, as suggested by the following two theorems.

**Theorem 5.** Consider a probability distribution on $C$, denoted by $\mathbf{p}_C$. Consider a level 1 aggregation of $C$, denoted by $A$. Then, if the flat probability condition holds in $\mathbf{p}_C$,

$$\|\mathbf{p}_C^1 - \mathbf{p}_C^2\|_{\mathrm{TV}} = \|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\mathrm{TV}} \tag{28}$$

holds, i.e. the aggregation has no impact on the TV distance.

*Proof.* See Appendix 6. □

**Corollary 5.1.** If $\mathbf{c}_1^0$ and $\mathbf{c}_2^0$ are corner states and $A$ is a level 1 aggregation,

$$\begin{aligned} \|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}^*\|_{\mathrm{TV}} &= \|\mathbf{p}_A(t;a_1^0) - \mathbf{p}_A^*\|_{\mathrm{TV}} \\ \|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}_C(t;\mathbf{c}_2^0)\|_{\mathrm{TV}} &= \|\mathbf{p}_A(t;a_1^0) - \mathbf{p}_A(t;a_2^0)\|_{\mathrm{TV}}, \end{aligned} \tag{29}$$

where $a_1^0 = \{\mathbf{c}_1^0\}$, $a_2^0 = \{\mathbf{c}_2^0\}$ are the initial aggregated states corresponding to disaggregated ones.

*Proof.* The flat probability condition holds for all $t$ by Theorem 2. Combined this fact with Theorem 5, the corollary is proven. □

**Theorem 6.** Consider aggregations $A$ and $A^+$ such that $A^+ < A$. Let $\mathbf{p}_A^1$ and $\mathbf{p}_A^2$ be the two probability distributions on $A$. Then, we define the difference of TV distances within $a^+ \in A^+$ as

$$\delta_{a^+}(\mathbf{p}_A^1, \mathbf{p}_A^2) = \frac{1}{2}\left\{ |p^1(a^+) - p^2(a^+)| - \sum_{a \in A(a^+)} |p^1(a) - p^2(a)| \right\}. \tag{30}$$

We also define the subset of $A^+$, denoted by $\tilde{A}^+$, as

$$\tilde{A}^+ = \{a^+ | p^1(a) \geq p^2(a) \quad \text{or} \quad p^1(a) \leq p^2(a) \ \ \forall a \in A(a^+)\}, \tag{31}$$

i.e. the signs of $p^1(a) - p^2(a)$ do not alternate within $\tilde{A}^+$. Then, $\delta_{a^+}(\mathbf{p}_A^1, \mathbf{p}_A^2) \geq 0$ for all $a^+ \in A^+$ and $= 0$ for all $a^+ \in \tilde{A}^+$, implying that only the states in $A^+ - \tilde{A}^+$ contribute the difference of TV distances between $A$ and $A^+$.

*Proof.* See Appendix 7. □

The implication of Theorem 6 is explained using a case in which there are 40 homogeneous users sharing the same choice set, denoted by $C^{\mathrm{all}} = \{1,2\}$. Let $x_1$ be the number of users choosing choice 1, which characterises the level 1 aggregation. Then, define a level 2 aggregation $A^+$ characterised by $y_1 = \lfloor x_1/5 \rfloor$ ($x_1 = 40$ is merged to $y_1 = 7$). Denote the level 2 aggregated states in $A^+$ by $a^+(y_1)$. If the cost does not rapidly change with respect to $x_1$, the probability distribution of the Markov chain dynamics, denoted by $p(a(x_1), t; \mathbf{c}^0)$, where $a(x_1)$ is the level 1 aggregated state

corresponding to $x_1$, does not rapidly change as well. This would cause a probability distribution like the one depicted in Figure 1, in which many $a^+(y_1)$ are included in $\tilde{A}^+$. This consideration implies that, unless $p(a(x_1),t;\mathbf{c}^0)$ and $p^*(a(x_1))$ are very close (in this case, TV distances will be small, and the difference between them is also small), the sign of $p(a(x_1),t;\mathbf{c}^0) - p^*(a(x_1))$ will not frequently change.

The aforementioned consideration is not a proposition with a concrete proof but an expectation. Therefore, we have to check how the TV distances on $A$ and $A^+$ are different when $A^+ < A$. Resolving this issue via an analytical approach is virtually impossible; we will numerically investigate this point in Section 6.4.3.

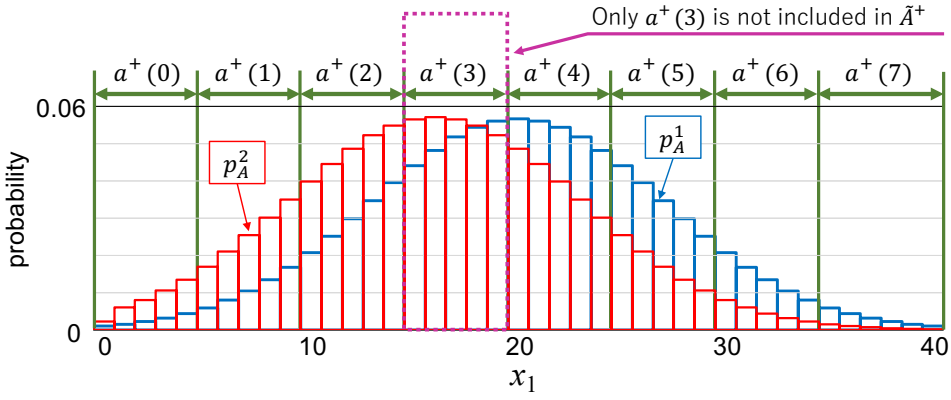


Figure 1: Comparison between level 1 aggregation and level 2 aggregation

# 5 Methodologies for calculating / estimating MCMT

This section introduces three methodologies for calculating MCMTs, namely the *direct matrix calculation*, *Monte Carlo method*, and *coupling technique*. Most of them are designed for numerical calculations. Calculation of the exact MCMT is mathematically intractable for all but tiny problems. We therefore resort to numerical methods.

## 5.1 Direct matrix calculation

Calculating the power of the transition matrix, as shown in Equation (7) and (16) is the most straightforward approach to finding the MCMT numerically. The standard diagonalisation tech-

nique to find the power of the transition matrix can be used. See Appendix 8 for details.

The main merit of the matrix calculation is that we can obtain the value of MCMTs precisely as accurately as the precision of the computer allows. Having said that, using the matrix calculation is not feasible for large-scale examples in which the number of states is large. First, all the entries of the transition matrix, of which $m^2$ exist ($m$ is the number of states), have to be determined. In addition, it is known that the calculation costs for the diagonalisation and matrix multiplication are $O(m^3)$. In real-world applications, $m$ can be huge even for a mid-size application with multiple user groups and multiple choices. In addition, because explicitly describing the transition matrix with the level 2 aggregation is not possible, we cannot resort to level 2 aggregation to reduce $m$. Therefore, the matrix calculation approach only applies to substantially smaller example cases. In Section 6, we use this approach to find exact MCMTs for small test cases.

## 5.2 Monte Carlo method for finding (ag-)si-MCMT

The Monte Carlo method estimates the probability distribution by simulating the dynamics of the transport system using equation (2) for a large number of times. Let $A$ be an aggregation, either level 1 or 2, and $f_A(a;t)$ be the frequency of occurrence of the aggregated state $a$ in the simulations and $N_{\mathrm{sim}}$ be the sample size of the Monte Carlo simulations. Then, we have $p_A(a;t) = f_A(a;t)/N_{\mathrm{sim}}$ as an estimated value of the probability of each state. The stationary distribution is estimated by letting $t$ be sufficiently large.

The Monte Carlo method is a powerful tool when the number of states is too large to handle by direct matrix calculation. In contrast, the calculation cost for obtaining a sufficiently accurate result may be large. This is particularly problematic when the number of states is large. Indeed, evaluating how accurately we can estimate the difference between two distributions $p_1(a)$ and $p_2(a)$ by the Monte Carlo method, we can conclude that we need to employ a number that is sufficiently greater than $p_{\mathrm{m}}^{-1}$ as $N_{\mathrm{sim}}$, where $p_{\mathrm{m}} = \max_{a\in A}\{p_1(a), p_2(a)\}$ (See Appendix 9 for details). $p_{\mathrm{m}}^{-1}$ is anticipated to be greater when the number of states is greater.

An effective approach to reducing the $N_{\mathrm{sim}}$ while maintaining accuracy is the level 2 aggregation. The high degree of aggregation is a straightforward answer if it is acceptable. Finding a strategy of an aggregation that can combine the states whose probabilities are not very different from each other is also a good approach, as suggested by Theorem 6.

### 5.3 Coupling technique for finding (ag-)ti-MCMT

The coupling technique is often employed for assessing the theoretical upper bound of MCMTs in the literature. We utilise this technique not only for the theoretical assessment but also for reducing statistical errors in the Monte Carlo simulation. The coupling technique considers two correlated Markov chains starting at different initial points. Denote these Markov chains by $c_{\mathrm{corr}}(t;\mathbf{c}_1^0,\mathbf{c}_2^0)$, where $\mathbf{c}_1^0,\mathbf{c}_2^0$ are the initial points. Its probability distribution is given by $\Pr\{c_{\mathrm{corr}}(t;\mathbf{c}_1^0,\mathbf{c}_2^0)=(\mathbf{c}_1,\mathbf{c}_2)\}$, where $\mathbf{c}_1$ and $\mathbf{c}_2$ are the states of Markov chains starting from $\mathbf{c}_1^0$ and $\mathbf{c}_2^0$, respectively. In the coupling technique, the marginal distributions of $c_{\mathrm{corr}}(t;\mathbf{c}_1^0,\mathbf{c}_2^0)$ must be equal to the distributions of the single Markov chains of $\mathbf{c}_1$ and $\mathbf{c}_2$, i.e. $\mathbf{p}_C(t;\mathbf{c}_1^0)$ and $\mathbf{p}_C(t;\mathbf{c}_2^0)$.

The key point to estimating TV distances with good accuracy concerns the way in which we correlate the two Markov chains. In the present paper, we correlate the selection of users who revisit their choices on each day and the selection of choices of these users. We explicitly formulate it below. First, we let $I(t)$ be the set of users who revisit their choices on day $t$ in both chains. The probability that any user in $I$ is included in $I(t)$ is $r$ and they are randomly selected to be independent with respect to other users and other days. Then, we draw a sequence of independent random uniform numbers between 0 and 1, denoted by $r_i^*$ for all $i \in I(t)$. Using $I(t)$ and $r_i^*$, the transition rule defined in Equation (2) is replaced with

$$c_i(t+1;\mathbf{c}_k^0)=\begin{cases} c_i^K(\mathbf{c}(t;\mathbf{c}_k^0)) & \text{if } \ i\in I(t) \\ c_i(t;\mathbf{c}_k^0) & \text{otherwise,} \end{cases} \tag{32}$$

where $k=\{1,2\}$. We determine $c_i^K(\mathbf{c}(t;\mathbf{c}_k^0))$ by Equation (4) using the same random number $r_i^*$ for both $k$. Therefore, if $\pi_c(\mathbf{c}(t;\mathbf{c}_1^0))=\pi_c(\mathbf{c}(t;\mathbf{c}_2^0))\ \forall c\in C_i$, the same choices are drawn in both chains. This implies that, once $\mathbf{c}(t;\mathbf{c}_1^0)=\mathbf{c}(t;\mathbf{c}_2^0)$ holds at $t$ (i.e. if the states of two chains collide), it will also hold at all times after $t$. The collision of two states is an important concept for analytical assessments of MCMTs and will be utilised in Section 6.

## 6 Properties of MCMT

The properties of MCMTs are investigated in this section. We first show an analytical assessment wherever possible in Section 6.1 and then show the numerical simulations based on the four key

examples, namely 1. *constant travel costs*, 2. *homogeneous users and two choices*, 3. *two groups with two choices*, and 4. *multi-dimensional case*. All numerical simulations are conducted with several parameters in order to achieve one of the objectives of this study, i.e. 'to conduct a preliminary examination of the relationship between MCMTs and critical properties of the system', as mentioned in Section 1.

The four examples will have different roles. Example 1 is an analytically tractable case, and the numerical simulation is carried out to validate the analytical findings in Section 6.1. Examples 2 and 3 are designed as examples of cases with interactions among users. Example 4 employs the departure time choice problem as an example. It has multi-dimensional degrees of freedom in which a large number of departure times are the choices of users. The quantitative effects of interactions between users cannot be analytically assessed and are not explicitly considered in Section 6.1. Performing numerical analyses in Examples 2-4 is crucial to understanding this point.

Examples 3 and 4 also demonstrate the applicability of the proposed method to large-scale models, which is the other aim of the study. The transport system of Example 3 seems tiny, but there is a large number of states even when level 1 aggregation is performed. This example is just about the size of the problem that can be computed with level 1 aggregation. Therefore, it is possible to analyse in detail how level 2 aggregation reduces the computational complexity but also how it affects the estimation results. Example 4 can only be assessed with the level 2 aggregation. We demonstrate that the proposed method can be used in such cases using this example.

## 6.1 Analytical assessment

We start with the following theorem, whose proof is shown in Levin and Peres (2017) (Proposition 4.7).

**Theorem 7.** Consider two Markov chains. Then, regardless of how the two chains are coupled,

$$\|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}_C(t;\mathbf{c}_2^0)\|_{\mathrm{TV}} \leq 1 - \Pr\{\mathbf{c}(t;\mathbf{c}_1^0) = \mathbf{c}(t;\mathbf{c}_2^0)\}. \tag{33}$$

This theorem is a key to analytical assessments in the following two subsections.

### 6.1.1 Analytical estimation of the dependency of the MCMT on $r$ and $n$ in general cases

It is easy to understand that the MCMT is inversely proportional to $r$ when $r \ll 1$. In this case, the number of users who change their choices per day is substantially smaller than the total number of users, and the resulting changes in travel costs are also substantially smaller. Therefore, doubling $r$ simply implies that the speed of change of the system doubles, resulting in a halving of the MCMT.

Estimating how MCMTs change with respect to $n$ in general cases is hard. We only perform a rough estimation using the coupling technique under the following special settings:

**1.** We evaluate the change of the MCMT when the number of users in a transport system (called the *original transport system*) doubles by duplicating every user. We call it the *doubled transport system*. The duplicated users are called *shadows*. The probabilities of shadows' choices are calculated based on the travel cost calculated by the choices of the original users, while choices of shadows and original choices are completely independent.

**2.** We assume that $\mathbf{c}_1^0$ and $\mathbf{c}_2^0$ are far apart, and the probability that the two Markov chains do not collide for the first time in the original transport system is one for $t \leq t_0$. After that, it becomes a constant number, denoted by $0 < \alpha < 1$. Note that this assumption implies $\Pr\{\mathbf{c}(t;\mathbf{c}_1^0) = \mathbf{c}(t;\mathbf{c}_2^0)\} = 1 - \alpha^{t-t_0}\ \forall t > t_0$.

We first couple two Markov chains in the original transport system. It can be different from the way of coupling in Section 5.3, but the states in the two chains must be the same after they collide. Setting $\mathbf{c}_1^0$ and $\mathbf{c}_2^0$ so that $t_0$ becomes greatest and applying Corollary 3.1 and Theorem 7, we have

$$t_C^* \leq \frac{\ln 0.25}{\ln \alpha} + t_0. \tag{34}$$

Then, we couple two Markov chains in the doubled transport system. The probability that the two Markov chains collide at $t$ is $(1-\alpha^{t-t_0})^2$. Approximating this by $1-2\alpha^{t-t_0}$, we have

$$t_C^{*\prime} \leq \frac{\ln 0.25 + \ln 0.5}{\ln \alpha} + t_0, \tag{35}$$

where $t_C^{*\prime}$ is the o-MCMT of the doubled transport system. Equations (34) and (35) implies the log-linear relationship between the upper bound of the o-MCMT and $n$.

The doubled transport system is only a hypothetical one, and we have to assess if the transport system to be analysed is not substantially different from this. It may be a good approximation if

doubling the number of users halves the impact on each user's travel costs. This property holds in Examples 1-3; we will check this with numerical simulations.

### 6.1.2 Analytical assessment of o-MCMT in constant-travel-cost setting

When the travel costs of all choices are constant, the upper bound of the o-MCMT can be analytically calculated without any external assumptions. Let two Markov chains be coupled in the way explained in Section 5.3. Then, let $I_0$ be the set of users whose initial choices are different between $\mathbf{c}_1^0$ and $\mathbf{c}_2^0$ (i.e. $c_{1i}^0 \neq c_{2i}^0 \ \forall i \in I_0$ and $c_{1i}^0 = c_{2i}^0 \ \forall i \in I - I_0$). The constant-travel-cost setting implies $\pi_c(\mathbf{c}(t;\mathbf{c}_1^0)) = \pi_c(\mathbf{c}(t;\mathbf{c}_2^0)) \ \forall c \in C_i$. Thus, the choices of every user who has revisited his/her choice at least once are the same in two chains. This implies that $\mathbf{c}(t;\mathbf{c}_1^0) = \mathbf{c}(t;\mathbf{c}_2^0)$ when all users have updated their choices at least once. Therefore, we obtain

$$\Pr\{\mathbf{c}(t;\mathbf{c}_1^0) = \mathbf{c}(t;\mathbf{c}_2^0)\} = (1-(1-r)^t)^{n_0}, \tag{36}$$

where $n_0 = |I_0|$. Hence,

$$\|\mathbf{p}_C(t;\mathbf{c}_1^0) - \mathbf{p}_C(t;\mathbf{c}_2^0)\|_{\mathrm{TV}} \leq 1-(1-(1-r)^t)^{n_0}, \tag{37}$$

which yields

$$t_C^*(\mathbf{c}_1^0,\mathbf{c}_2^0) \leq \frac{\ln(1-0.75^{\frac{1}{n_0}})}{\ln(1-r)}. \tag{38}$$

Letting $n_0$ be as great as possible (i.e. letting $n_0 = n$) and applying Equation (3.1), we have

$$t_C^* \leq \frac{\ln(1-0.75^{\frac{1}{n}})}{\ln(1-r)}. \tag{39}$$

Performing linear approximations for both the numerator and denominator of Equation (39), we finally obtain

$$t_C^* \leq \frac{\ln n - \ln(-\ln 0.75)}{r}. \tag{40}$$

This result is consistent with the dependency of the o-MCMT on $n$ and $r$ estimated in Section 6.1.1.

## 6.2 Numerical simulation of Example 1

The constant-travel-cost setting is employed in Example 1 to verify the analytical result shown in Section 6.1.2. We employed the direct matrix calculation for this example. The number of groups is set to one. Users have two choices, i.e. choices 1 and 2, and the choice probability of choice 1, which is constant, is set to either 0.2 or 0.5, respectively, i.e. $(K_1, K_2) = (0.2, 0.8)$ or $(K_1, K_2) = (0.5, 0.5)$. Various $n$ and $r$ are employed, while only smaller $n$ are used for the disaggregated cases because the number of disaggregated states is $2^n$.

Figure 2 depicts the TV distances for the (ag-)o-MCMT, i.e. $d_C^{\max}(t)$ and $d_A^{\max}(t)$ for $n = 6$. The four parameter sets ($r = 0.1$ or 0.25 and choice probability of choice 1 = 0.2 or 0.5) are used. We can observe $d_C^{\max}(t) = d_A^{\max}(t)$ for all cases, implying that Corollary 5.1 holds in this case.

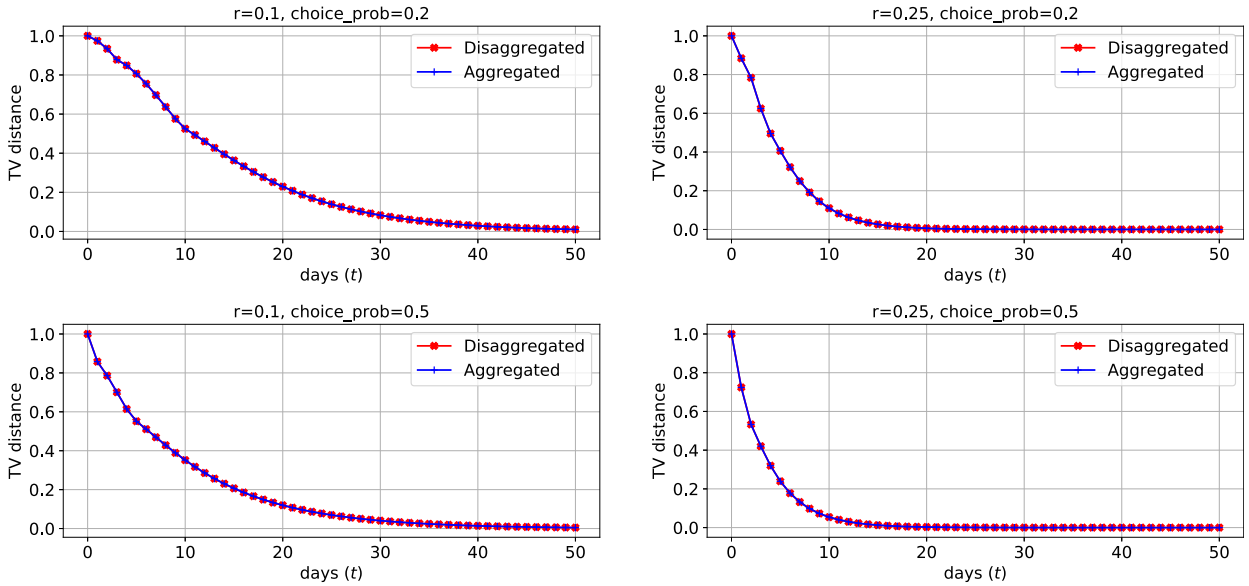


Figure 2: TV distances $d_C^{\max}(t)$ (disaggregated) and $d_A^{\max}(t)$ (aggregated)

Figure 3 depicts the ag-o-MCMTs with respect to $n$, where the choice probability of choice 1 = 0.2 or 0.5 and $r = 0.1$ or 0.25. The theoretical upper bound calculated by Equation (39) is also shown. The horizontal axis is shown in the log scale so that we can check the log-linear relationship to $n$. The results clearly show that this relationship holds in the numerical calculations.

Figure 4 depicts the ag-o-MCMTs with respect to $1/r$, where the choice probability of choice 1 = 0.2 or 0.5. The theoretical upper bound calculated by Equation (39) is also shown. $1/r$ is used instead of $r$ because Equation (40) implies the proportional relationship with respect to $1/r$. The results clearly show that this relationship holds in the numerical calculations.

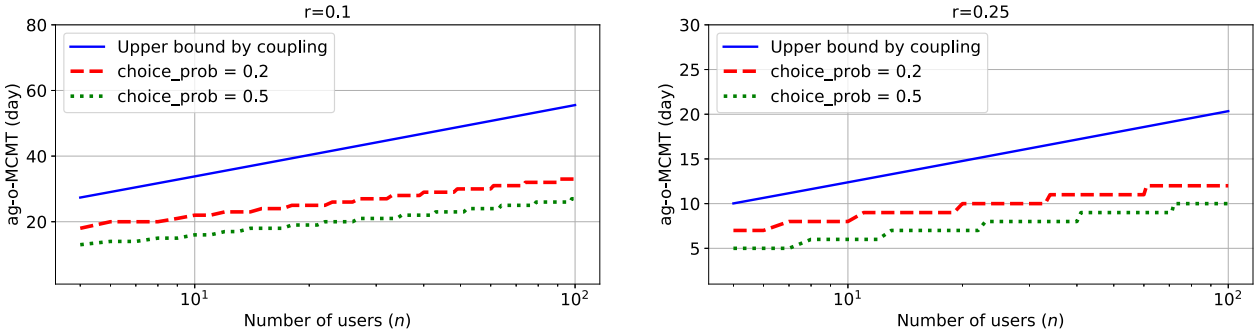


Figure 3: Numerically calculated ag-o-MCMT and its theoretical upper bound by coupling (w.r.t. $n$) Note: horizontal axis is in log scale.

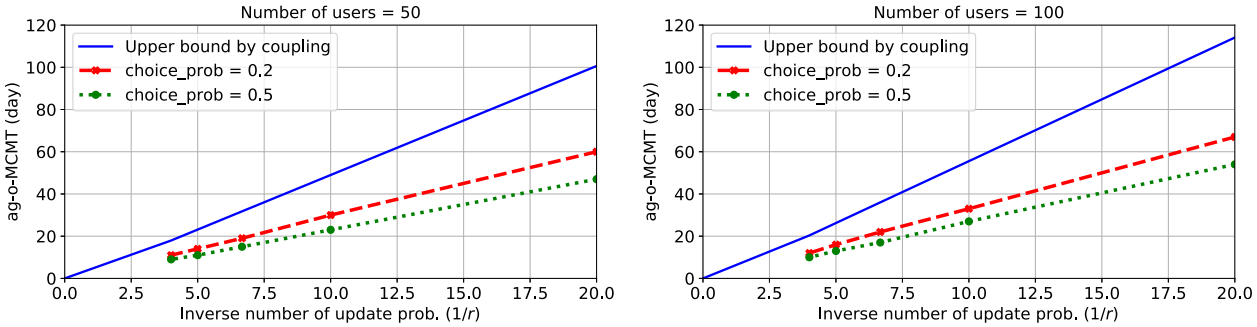


Figure 4: Numerically calculated ag-o-MCMT and its theoretical upper bound by coupling (w.r.t. $1/r$)

Another property that can be read from both figures is that the numerically calculated MCMTs are always approximately half of the theoretical upper bound. This indicates that the theoretical upper bound reflects the actual change in MCMTs with respect to $n$ and $1/r$. The MCMTs vary with the choice rate, but the effect seems not as large as the variation caused by other variables.

The following observation summarises the results of Example 1:

**Observation 1.** When the travel costs are constant, the (ag-)o-MCMTs are proportional to $1/r$ and have a linear relationship to $\log(n)$.

The properties stated in Observation 1 can be regarded as a 'reference point' for understanding the properties of MCMTs in other examples where interactions with the user are present. Comparing the results with this reference point, the properties of MCMTs will be better understood. Letting $\theta = 0$ will all lead back to Example 1 in subsequent calculation examples. We include the cases where $\theta = 0$ in most cases so that they serve as the reference points.

## 6.3 Numerical simulation of Example 2: One group with two choices

Example 2 is analysed mainly to understand the effect of users' interactions on MCMTs. It represents the simplest transport system with users' interactions: the single-OD-two-routes traffic assignment is the most typical example. All users have the same choice set $C^0 = \{1,2\}$. The cost of each choice is the same for all users.

The level 1 aggregation is employed. As there are one user group and two options, $\mathbf{x} = (x_1, n - x_1)$. This implies that there are $n+1$ states of level 1 aggregation with this $\mathbf{x}$. The cost of each option is defined as a function of $x_1$, denoted by $\pi_1(x_1), \pi_2(x_1)$.

We consider two types of travel cost functions in this section, i.e. the congested transport system and the transport system with positive interactions. The main point of the analysis is how MCMTs change by parameters $n$, $r$, and $\theta$. All analysis in this section is based on aggregated states; hence, all MCMTs are aggregated versions. We employed the direct matrix calculation.

### 6.3.1 Case A: Congested transport system

We consider a traffic assignment problem with a single origin-destination pair and two routes, denoted by Routes 1 and 2. The travel cost of Route 1 is assumed to be constant, while that of Route 2 depends on the number of users choosing this route. They are denoted by

$$\pi_1(x_1) = 2 \quad \text{and} \quad \pi_2(x_2) = \frac{2x_2}{n} + 1. \tag{41}$$

Therefore, the choice probability of route 2, denoted by $K_2(x_2)$, is

$$K_2(x_2) = \frac{\exp(-\theta\,(2x_2/n - 1))}{1 + \exp(-\theta\,(2x_2/n - 1))}. \tag{42}$$

Numerical simulations calculating ag-o-MCMTs were carried out for the parameters sets of $n = 50, 100$, $\theta = 0, 1, \ldots, 30$, $1/r = 3, \ldots, 100$ (i.e. $r = 0.01, \ldots, 0.333$), and $n = 10, 20, \cdots, 100$, $\theta = 0, 10, 30$, $1/r = 3, 10$. The results are depicted in Figures 5-7. They indicate that the greater $\theta$ causes a smaller MCMT, except for greater $r$ ($r = 0.25$ and $0.333$), in which greater $\theta$ results in greater ag-o-MCMTs. The linear relationship between $1/r$ and the ag-o-MCMT is found unless $1/r$ is very low. In contrast, the log-linear relationship between $n$ and ag-o-MCMT is not always found, especially when $r$ and $\theta$ are large.

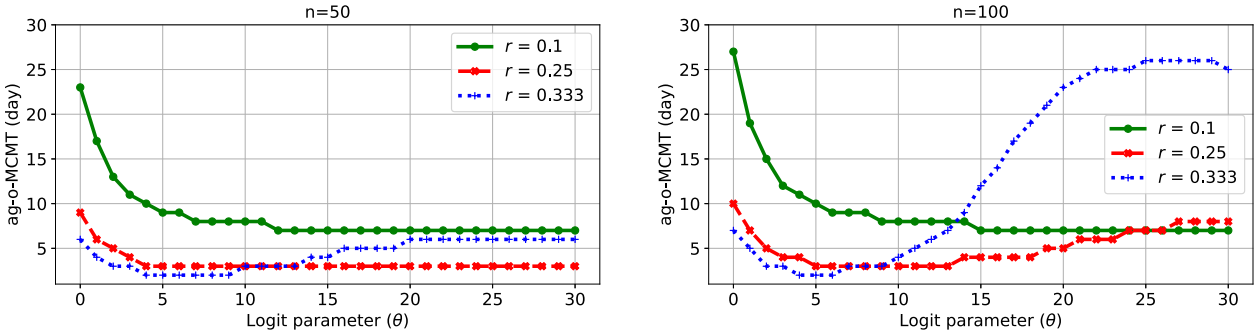


Figure 5: ag-o-MCMT w.r.t. θ in the congested case

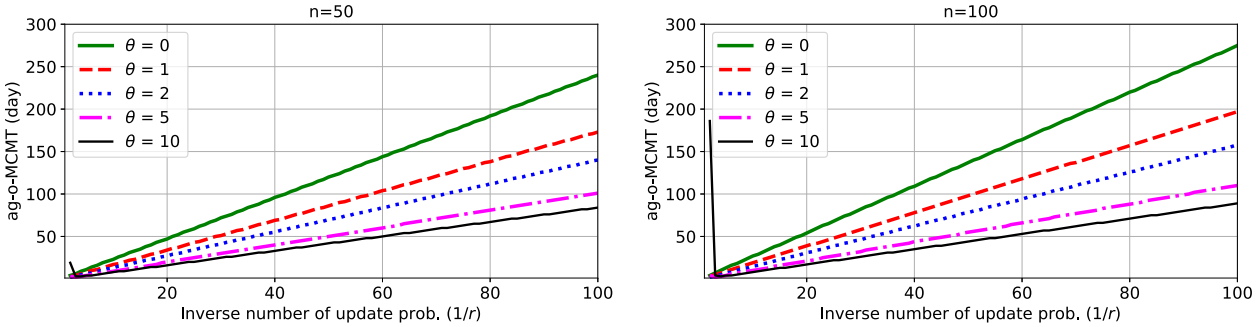


Figure 6: ag-o-MCMT w.r.t. $1/r$ in the congested case

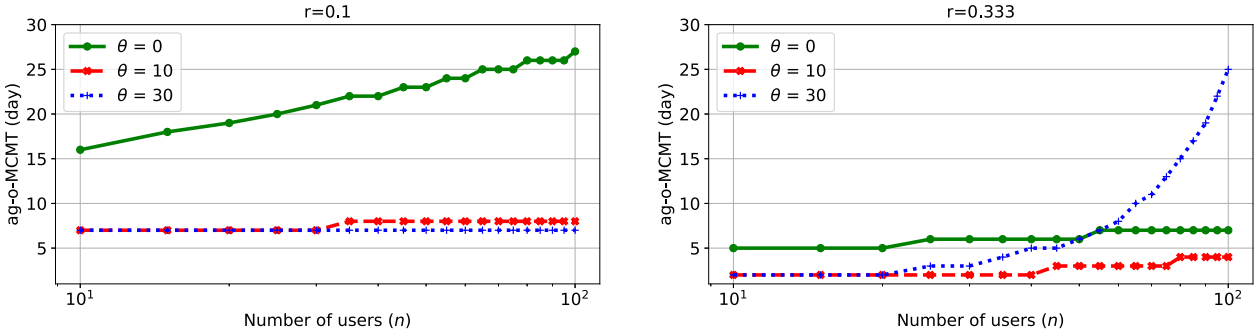


Figure 7: ag-o-MCMT w.r.t. $n$ in the congested case

To investigate the situations where $r$ is greater, we calculated the stationary distributions for θ = 5,10 and several greater values of $r$, as depicted in Figure 8. Transient states ($t$ = 1 - 6) were also calculated for θ =5,10 and $r = 0.333$ as depicted in Figure 9. We can observe that the stationary distributions for $(r, \theta)$ = (0.25,20) and (0.333,20) have two peaks, while those in other cases have a single peak. In addition, Figure 9 indicates that the distribution oscillates when θ = 20, while it quickly converges to the stationary distribution when θ = 5. Combined with the result of MCMTs shown in Figure 5, the MCMTs attain greater values than other cases when the oscillation occurs.

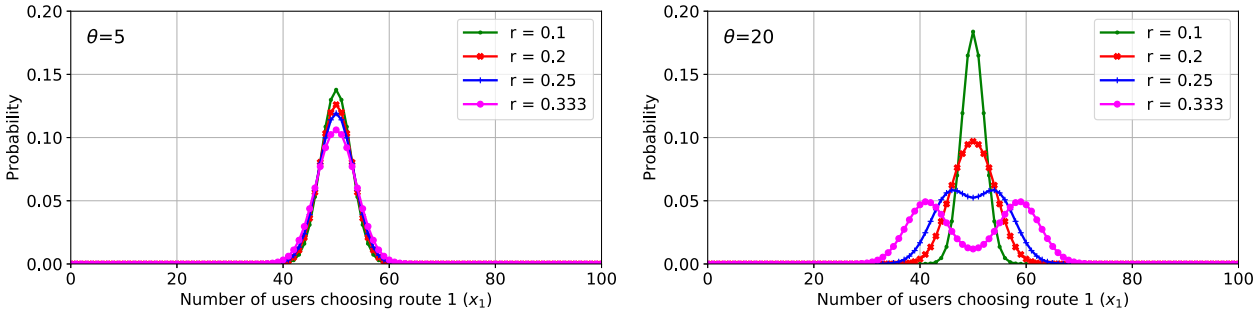


Figure 8: Stationary distributions of $x_1$ in the congested case

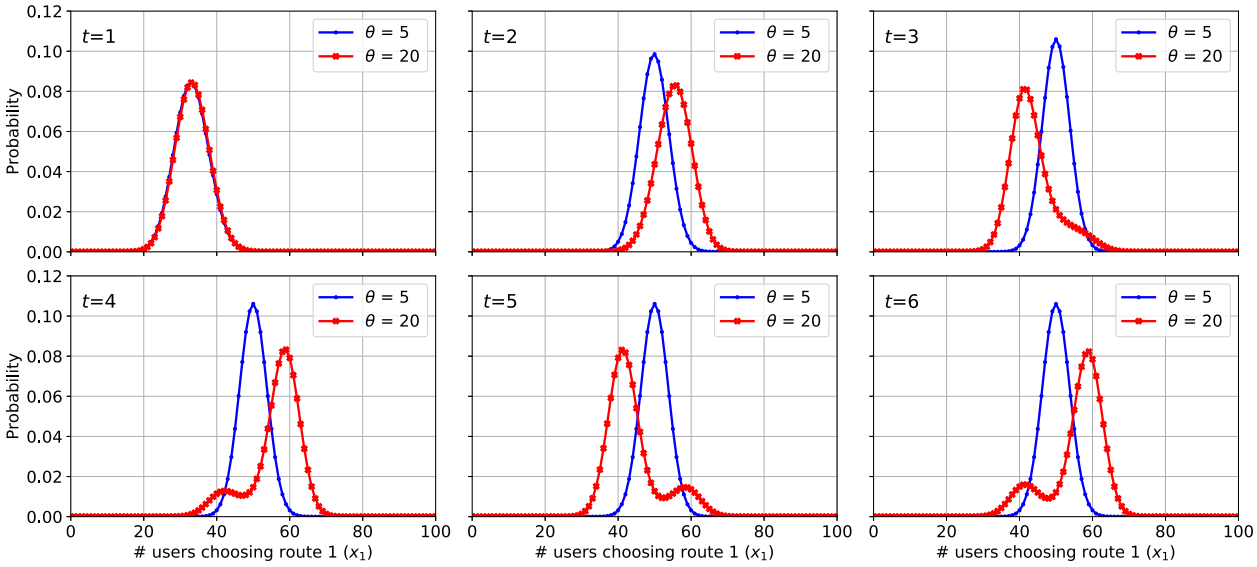


Figure 9: Probability distributions of $x_1$ in transient states in the congested case; $r = 0.333$

We summarise the results of Example 2 by the following observation:

**Observation 2.** The following properties are found in the congested case of Example 2:

**1.** The ag-o-MCMT is smaller than in the constant-travel-cost setting, especially for greater θ.

**2.** The ag-o-MCMT is proportional to $1/r$. The log-linear relationship to $n$ is not common.

**3.** Properties 1 and 2 do not hold when both $r$ and θ are so large that the probability distribution oscillates. The ag-o-MCMT in such cases can be substantially larger than in other cases.

In the above-reported experiments two parameters were altered which affect the 'reactivity' of the system to past events: increasing θ means that drivers behave more similarly to each other, conditional on the past, and increasing $r$ means that more drivers will adapt their choice on any

one day. The observed impacts of such parameters on the shape of the stationary distribution are consistent with those noted in Hazelton and Watling (2004). The impacts on the dynamics of the process are consistent with those noted in the auto-correlation analysis of Balijepalli et al. (2007), and in the simulation experiments reported in Watling and Hazelton (2018)

### 6.3.2 Case B: Transport system with positive interactions

We consider a transport system with positive interactions, in which users prefer using a more popular choice. The positive interactions may occur by users influencing each other, e.g. Fukuda and Morichi (2007). They may exist in a transport system with the scale of economics (e.g. the Mohring effect (Mohring, 1972)). See Iryo and Watling (2019) for a detailed discussion on them.

We consider a transport system in which users choose either one of two transport modes, denoted by Modes 1 and 2. Mode 1 is a traditional transport system with no user interactions, while the travel cost of Mode 2 decreases with respect to the number of users using it due to the users' positive interactions. Their travel costs are defined as

$$\pi_1(x_1) = 0 \quad \text{and} \quad \pi_2(x_2) = -\frac{2x_2}{n} + 1. \tag{43}$$

Therefore, the choice probability of route 2, denoted by $K_2(x_2)$, is

$$K_2(x_2) = \frac{\exp(\theta\,(2x_2/n - 1))}{1 + \exp(\theta\,(2x_2/n - 1))}. \tag{44}$$

A numerical simulation was carried out for parameters $n = 50, 100$, $\theta = 0, 0.5, \ldots, 2.5$ and $1/r = 3, ..., 100$. The results are shown in Figures 10 – 12. A greater $\theta$ causes a greater ag-o-MCMT, which is different from the result of the congested-transport-system case explained in Section 6.3.1. The linear relationship between $1/r$ and the ag-o-MCMT is established in all cases in the numerical test. The relationship between $n$ and the ag-o-MCMT depends on $\theta$. When, $\theta = 1$, we observe log-linear relationships. When $\theta > 1$, the ag-o-MCMT increases faster than $\log(n)$.

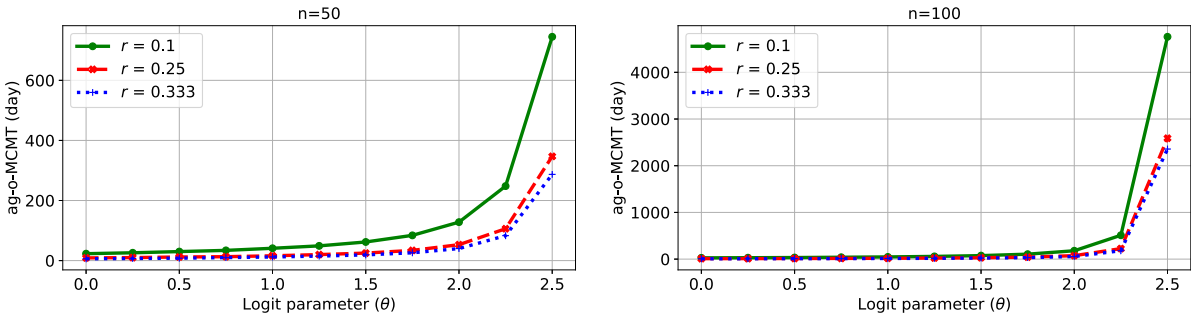


Figure 10: ag-o-MCMT w.r.t. θ in the positive interaction case

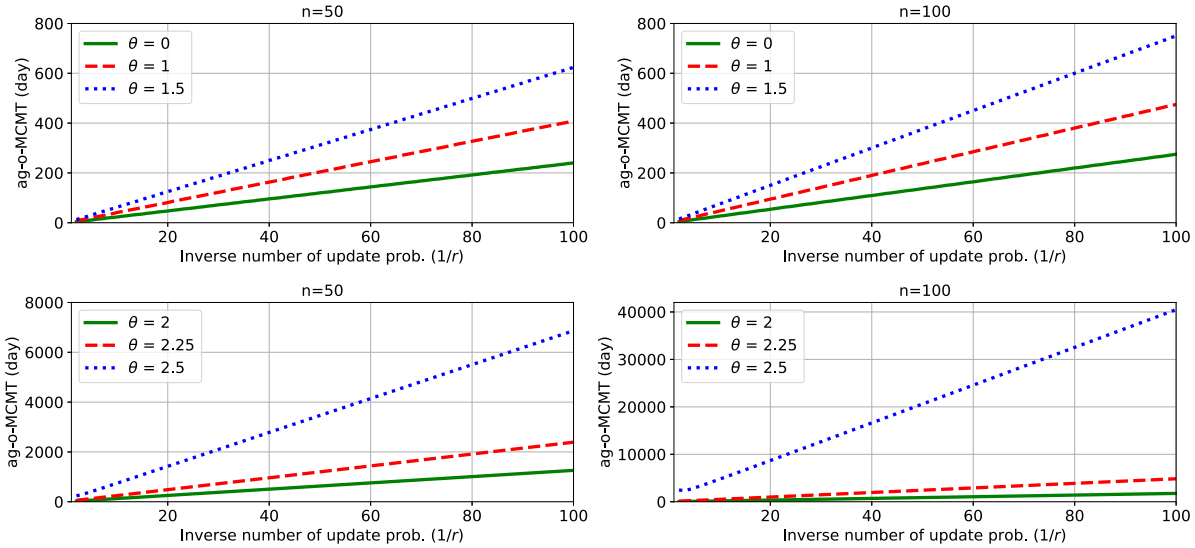


Figure 11: ag-o-MCMT w.r.t. $1/r$ in the positive interaction case

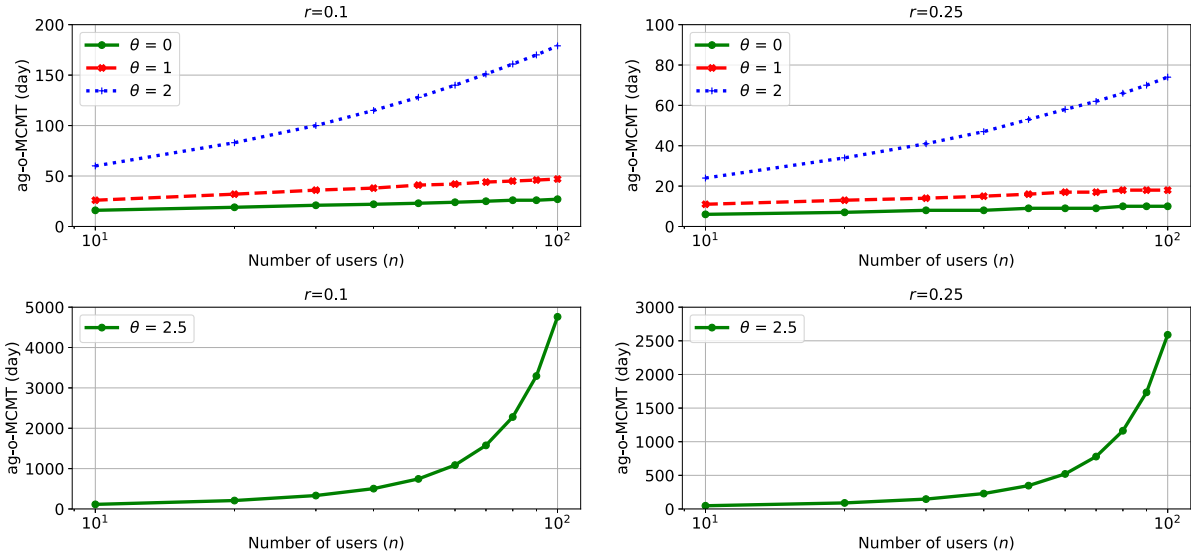


Figure 12: ag-o-MCMT w.r.t. $n$ in the positive interaction case

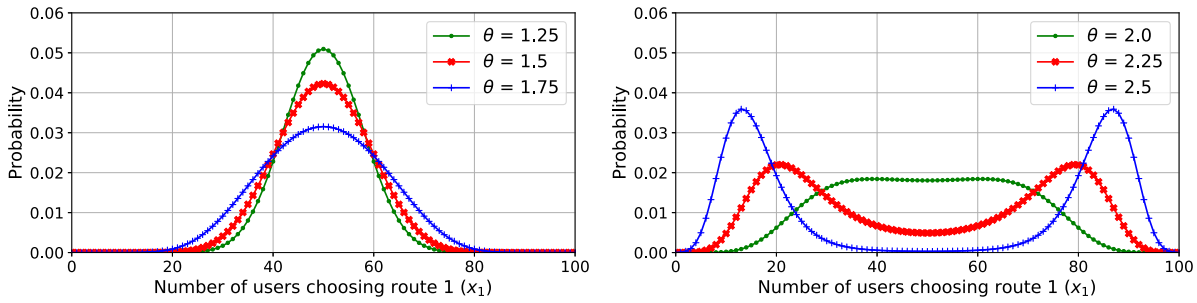


Figure 13: Stationary distributions of $x_1$ in states in the positive interaction case

To investigate why greater θ causes greater ag-o-MCMTs, we plotted the stationary distribution of several parameter settings in Figure 13. The result indicates that the two peaks are present when $\theta \geq 2.0$. It imply that the states in which users either choose Mode 1 more often or Mode 2 more often are realised because of the influence of positive interactions between users. In such cases, transitions between states occur infrequently because the more 'popular' choice is stuck in either of them. This is considered as a reason for the large increase in MCMTs in the present case.

We summarise these findings as the following observation:

**Observation 3.** The following properties are found in the positive-interaction case of Example 2:

**1.** The ag-o-MCMT is greater than in the constant-travel-cost setting, especially for greater θ. Moreover, it grows rapidly after θ exceeds a threshold (around 2.0 in the present example), at which the stationary distribution has two peaks.
**2.** The ag-o-MCMT is proportional to $1/r$. The log-linear relationships to $n$ are observed when $\theta = 1$, while the ag-o-MCMT increases faster than $\log(n)$ when $\theta > 1$.

## 6.4 Numerical simulation of Example 3: Two groups with two choices

### 6.4.1 Problem setting and calculation methods

This is a more complex example compared to example 2, in which users are categorised into two groups. The degree of freedom in this transport system is two, while Examples 1 and 2 have only one degree of freedom. Iryo and Watling (2019) analysed the behaviour of such a system by a deterministic and continuous model and found that the two-dimensional setting generates different types of solutions. They can be classified into nine types of behaviour, depending on the cost

function. These can have, for example, a single equilibrium point, multiple equilibrium points or no stable equilibrium point but only periodic trajectories.

The cost functions used in the numerical example in Iryo and Watling (2019) were used as an example of a case without a stable equilibrium solution. Note that while this example generates a periodic trajectory in the deterministic and continuous setting, the trajectory of the Markov chain is always aperiodic, as mentioned in Section 2.

The purpose of analysing Example 3 is twofold. First, same as with Examples 1 and 2, we investigate the relationships between MCMTs and parameters such as $n$,$r$, and $\theta$. We also investigate how the coupling technique and level 2 aggregations work for larger cases. The number of the states of Example 3 can be huge; if there are 100 users in group $a$ and other 100 users in group $b$, the number of states will be $(100+1)^2 = 10,201$ even after the level 1 aggregation is applied. We will show that the coupling technique will provide an upper bound of MCMTs with a substantially small number of samples.

The detailed setting of the cost functions is as follows: two groups are denoted by $a,b$, and two choices for them are denoted by $C^a = \{a_1, a_2\}$ and $C^b = \{b_1, b_2\}$. The number of users in each group is $n_a = n_b = n$ (i.e. $2n$ in total). Their travel costs are

$$\begin{aligned} &\pi_{a_1}(\mathbf{x}) = -0.5(x_a/n_a) - (x_a/n_a) + 0.75, \;\; \pi_{a_2}(\mathbf{x}) = 0 \\ &\pi_{b_1}(\mathbf{x}) = 1.5(x_a/n_a) - 0.5(x_a/n_a) - 0.5, \;\; \pi_{b_2}(\mathbf{x}) = 0 \end{aligned} \tag{45}$$

where $x_a$ and $x_b$ are number of users in groups $a$ and $b$ and choose choice $a_1$ and $b_1$, respectively. The level 1 aggregation is applied to calculate the TV distances and MCMTs.

We used the Monte Carlo method to calculate the TV distances and MCMTs of Example 3. We first calculated the MCMTs using the Monte Carlo method without coupling, which was explained in Section 5.2, as the size of the problem prohibits us from performing the direct matrix calculation. The Monte Carlo simulation was carried out for parameters $\theta = 0, 1, \ldots, 10$ and $1/r = 10, \ldots, 100$ (i.e. $r = 0.01 \ldots 0.1$). To obtain accurate results with fewer stochastic errors, we let the number of samples be very big, i.e. 1 million. Then, we calculated the MCMTs using the coupling technique explained in Section 5.3. Because we intend to demonstrate how this technique reduced the number of samples in the Monte Carlo simulation, we employed 10, 100, 1000 and 10000 as the number of samples for the calculations using the coupling technique.

Because the number of level 1 aggregated states is substantially greater than in Examples 1 and 2, we did not calculate the ag-o-MCMT as it demands calculations from all states. Instead, we calculated the ag-si-MCMT and ag-ti-MCMT as surrogates of the ag-o-MCMT. For this purpose, choosing states that are far from a state with a high probability of the stationary distribution is preferable. We can confirm from the stationary distribution that the probabilities of the corner states are low. Therefore, we set $x_a = x_b = 0$ as the initial state of the ag-si- and -ti-MCMT and $x_a = x_b = n$ as the second initial state of the ag-ti-MCMT.

We first show the stationary states in a few settings in Figure 14. Parameters are $n_a = n_b = 100$, $r = 0.1$, and $\theta = 5, 8, 10$. We can observe a doughnut-shaped distribution for $\theta = 8, 10$.

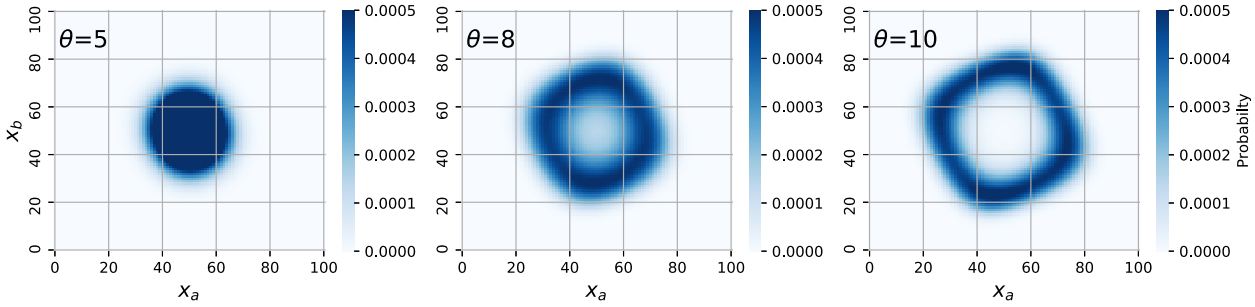


Figure 14: Stationary distributions of Example 3

### 6.4.2 Relationship between MCMTs and parameters

We examined how the ag-si-MCMT changes with respect to $n$, $r$, and $\theta$. All results in this section were obtained by the Monte Carlo method without coupling. The level 1 aggregation was employed, in which each aggregated state is characterised by $x_a$ and $x_b$.

Figrues 15 – 17 depict the ag-si-MCMT with respect to $1/r$, $\theta$, and $n$, respectively. The following observation summarises the results:

**Observation 4.** The following properties are found in Example 3:

**1.** The ag-si-MCMT is proportional to $1/r$. The relationships with $n$ and the ag-si-MCMT are approximately log-linear when $\theta = 5$, while it increases faster than $\log(n)$ when $\theta = 10$.
**2.** The ag-si-MCMT is monotonically increasing with respect to $\theta$.

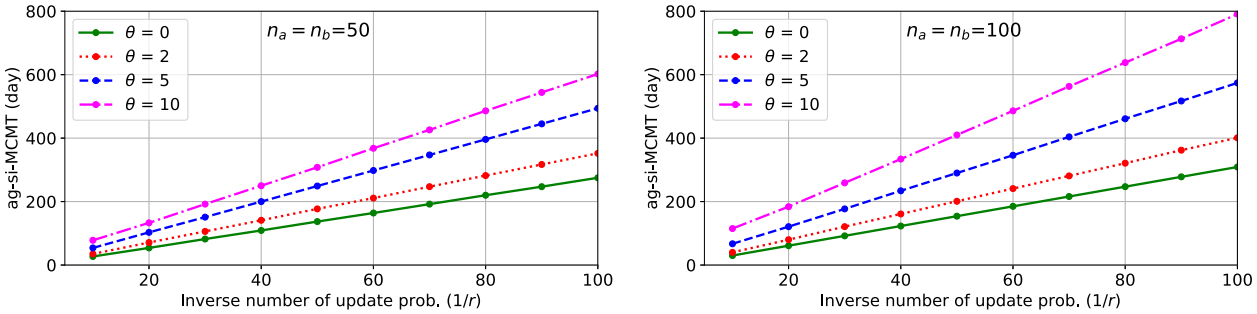


Figure 15: Ag-si-MCMTs w.r.t. to $1/r$ in Example 3

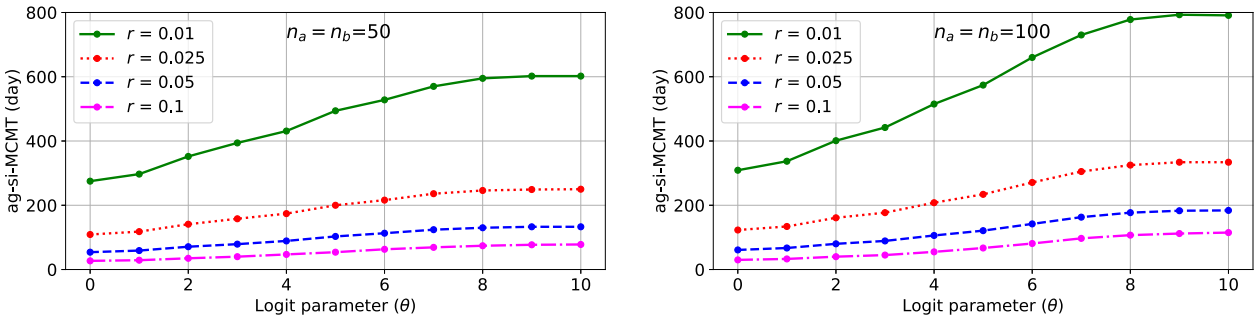


Figure 16: Ag-si-MCMTs w.r.t. $\theta$ in Example 3

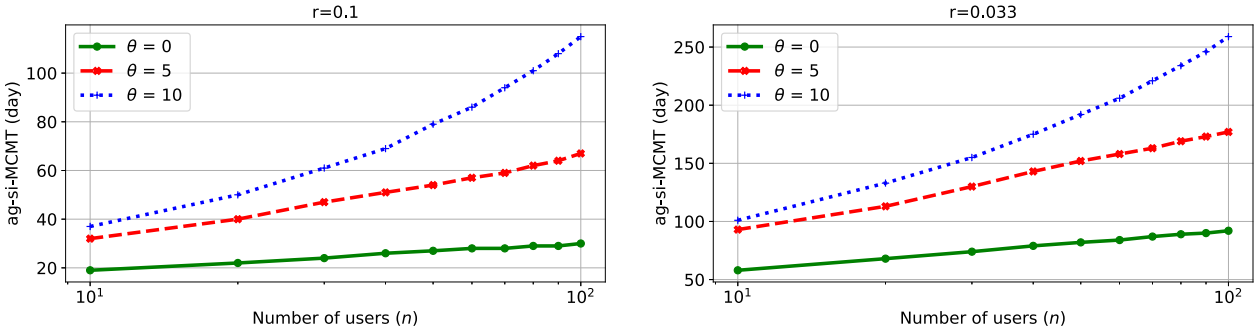


Figure 17: Ag-si-MCMTs w.r.t. $n$ in Example 3

### 6.4.3 Performance comparison of MCMT calculations with small sample size

We examined how the level 2 aggregation and the coupling are effective when the sample size in the Monte Carlo simulation is small. Throughout this section, we used $\theta = 10$ and $r = 0.1$.

The comparison was made for direct calculations of the ag-si-MCMTs and calculations of ag-ti-MCMT with the coupling technique introduced in Section 5.3. The level 2 aggregations were

performed by combining the level 1 aggregated states with respect to $(x_a^{\mathrm{Agg}}, x_b^{\mathrm{Agg}})$, defined as

$$(x_a^{\mathrm{Agg}}, x_b^{\mathrm{Agg}}) = (\lfloor x_a / n^{\mathrm{Agg}} \rfloor, \lfloor x_b / n^{\mathrm{Agg}} \rfloor). \tag{46}$$

For example, when $n^{\mathrm{Agg}} = 10$, all states are grouped together by multiples of 10 of their number. We employed $n^{\mathrm{Agg}} = 1$ (i.e. identical to the level 1 aggregation), 3, and 10 in the numerical test. The numbers of states are 10,201, 1,156, and 121, respectively.

We first examine how the level 2 aggregations affect the TV distances and the ag-si-MCMTs. For this purpose, the TV distances for different aggregation levels are depicted in Figure 18, which were calculated by the Monte Carlo simulation with the same setting in Section 6.4.2. We can observe that the changes in TV distances by different level 2 aggregations are very small, as anticipated by Theorem 6. On the other hand, the ag-ti-TV distance is greater than the ag-si-TV distance, implying that the ag-ti-MCMTs are greater than the ag-si-MCMTs, as stated by Corollary 3.1.

Figure 19 shows TV distances when the numbers of samples are small, which are 10, 100, and 1000. We can observe that the ag-si-TV distances do not drop smaller than 0.25 when $n^{\mathrm{agg}} = 1$, implying that estimating MCMTs with level 1 aggregation is impossible with these sample sizes. In contrast, ag-ti-TV distances drop to zero, which can be used to estimate MCMTs.

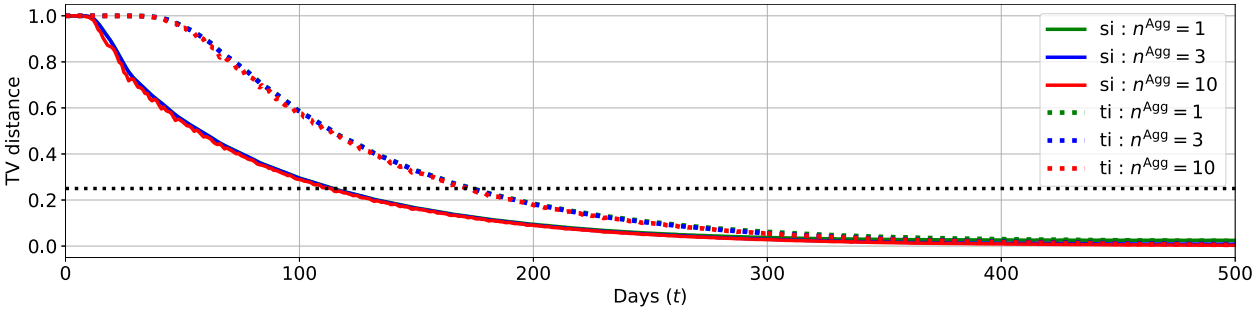


Figure 18: TV distances for different aggregation levels. si: ag-si-TV dist., ti: ag-ti-TV dist.

To observe the estimations of ag-ti-MCMTs in different numbers of samples, we performed 100 trials with four sample sizes, namely 10, 100, 1000, 10000, for $n^{\mathrm{Agg}} = 1, 10$. Figure 20 depicts the distributions of estimated MCMTs in 100 trials. We can observe that, while the level 1 aggregation (i.e. $n^{\mathrm{Agg}} = 1$) cases derive overestimated results, the level 2 aggregation cases provide good estimations, especially when the sample size is greater than 100. This result implies that appropriate level 2 aggregations can provide accurate estimations of MCMTs.

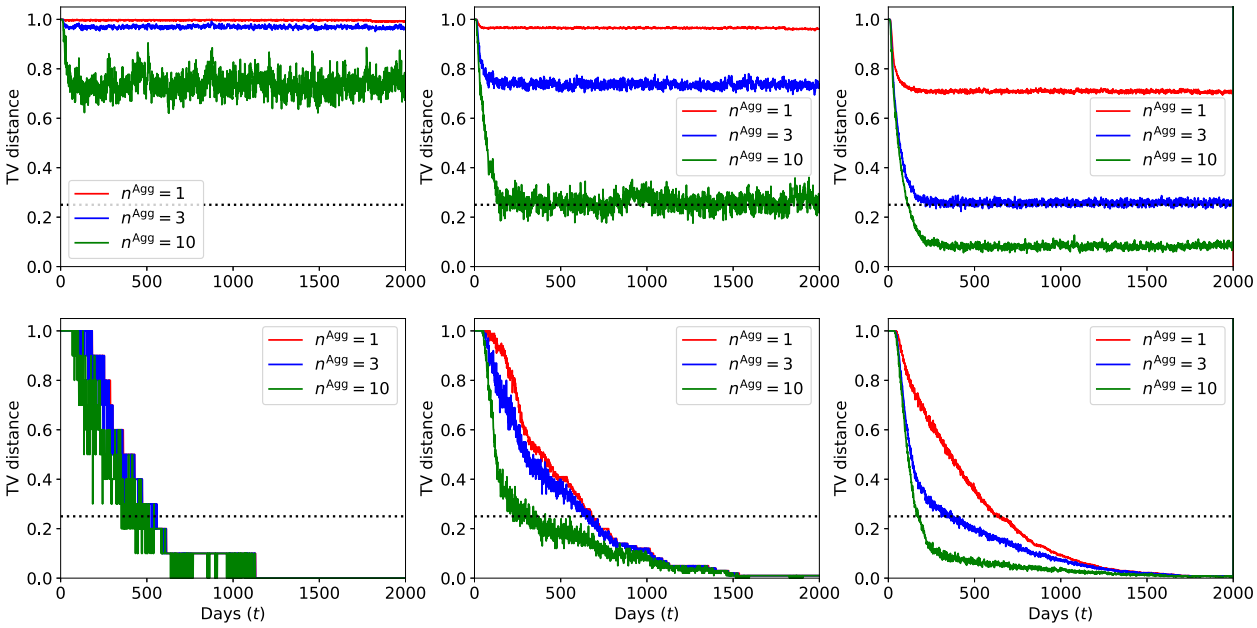


Figure 19: ag-si-TV (up) and ag-ti-TV (down) distances: Sample sizes = 10 (left), 100 (middle), 1000(right)

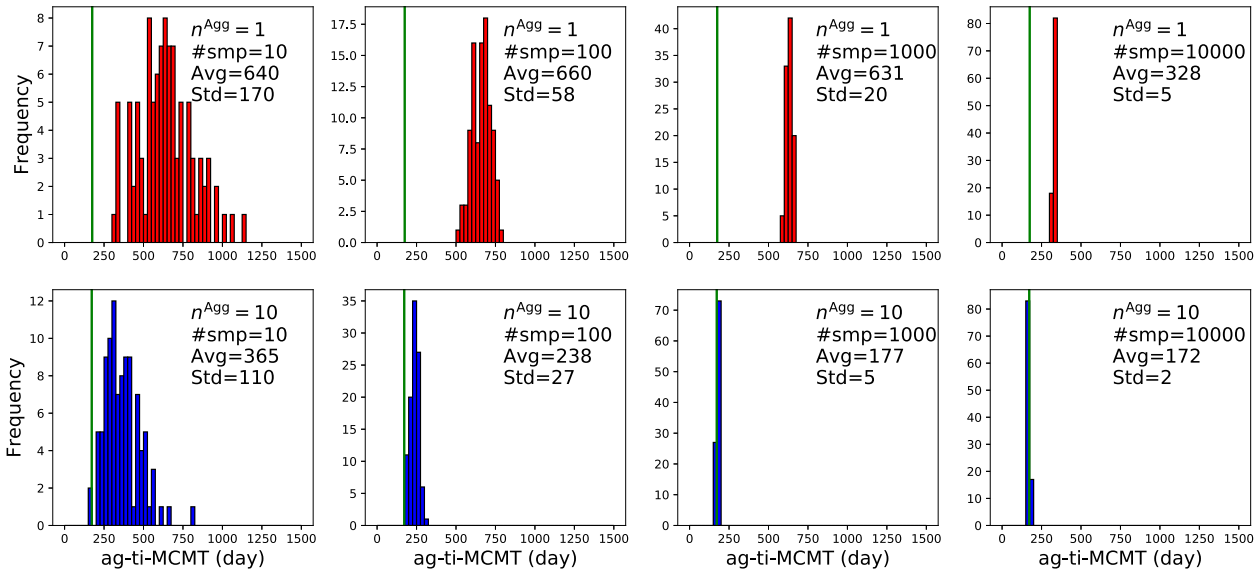


Figure 20: Histogram of estimated ag-ti-MCMTs for 100 trials: #smp = sample size; green horizontal lines indicate the ag-ti-MCMT of $n^{\text{Agg}} = 1$

## 6.5 Numerical simulation of Example 4: Multi-dimensional case: the departure time choice problem

This example is employed to observe whether the proposed method is useful for the analysis of transport systems with a high degree of freedom. The basic setup of the problem is the same as in Vickrey (1969), in which several studies (e.g. Iryo (2008), Guo et al. (2018), Iryo (2019), Jin (2018, 2021)) have analysed the stability of the equilibrium solution. In this study, for simplicity, all users are assumed to have the same schedule cost function. In addition, all users travel between the same origin and destination. The only choices made by the users are the departure times. By discretising the departure time in the model, the users' choices are also discretised. There is only one bottleneck with constant capacity between the origin and destination points, at which the delay occurs. Travel cost is defined as the sum of the delay and the schedule cost at the destination.

All users are assumed to have the same schedule cost function for simplicity. In addition, all users travel between the same origin and destination. There is only one bottleneck with constant capacity between the origin and destination points, at which the delay occurs. The free flow travel time is set to zero without loss of generality. Its capacity is 500 vehicles per hour. The number of users is set to 1000. Therefore it takes at least two hours for all users to pass through the bottleneck. The study duration is 10 hours, divided into 100 time slots (i.e. one time slot = 6 mins). Users' desired arrival time is set as the middle point of the study duration. Users choose their departure times so as to minimise the generalised travel cost consisting of the bottleneck delay and the schedule cost. The schedule cost is $0.3\Delta T_S$ for early arrivals and $5\Delta T_S$ for late arrivals, where $\Delta T_S$ is the difference between the desired arrival time and the actual arrival time. The units of both costs are in hours. We set $1/r = 10, 20, \ldots, 100$ and $\theta$ = 12 for the numerical test.

The ag-ti-MCMT was calculated in the present example. One of two initial states is fixed to the first time slot, i.e. everyone chooses the first slot in this state. The departure times of other initial states are set to $3.2, 3.6, \ldots$, or 4.8 (hour), where 0.0 is the beginning of the study duration and 5.0 is the desired arrival time. In the earlier time cases, no congestion occurs near the desired arrival time, and consequently, the user first selects a time near the desired arrival time. As the congestion grows, they select other time slots to avoid congestion. In the later time cases, congestion already exists around the desired arrival time in the initial state, and consequently, the user chooses time slots considering congestion from the epoch day. As these user behaviours are very different, the ag-ti-MCMTs can be used to find an approximated result of the ag-o-MCMT.

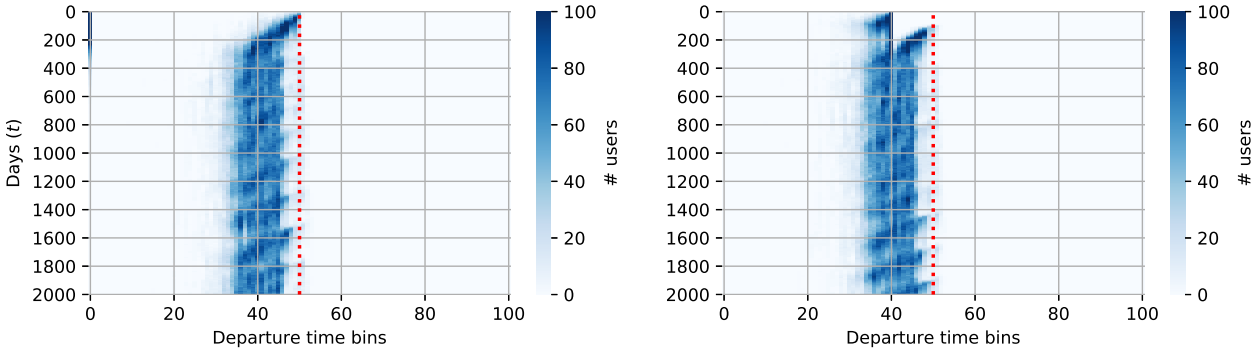


Figure 21: A sample of two coupled Markov chains in Example 4; initial states are 0.0 (left) and 4.0 (right). Red dotted line = desired arrival time

An example of two coupled Markov chains with $1/r = 10$ (i.e. $r = 0.1$) is shown in Figure 21 by visualising the number of people selecting each time slot time for each of the initial states. An oscillation can be seen in both cases.

The level 2 aggregation is essential to calculate the MCMT in this example. Observing the oscillation in Figure 21, we would be interested in how long it takes for the departure times to converge to the stationary distribution. Therefore, we chose the average departure time as the index for the level 2 aggregation. This index is calculated by the following formula:

$$t_{\mathrm{Avg}}(t) = \left\lfloor \frac{\sum_i^{1000} \tau_i(t)}{1000} \times 60 \right\rfloor \quad \text{where } \tau_i(t) \text{ is the departure time of user } i \text{ in hour,} \tag{47}$$

i.e. $t_{\mathrm{Avg}}(t)$ is the average departure time shown by an integer in minutes. Any states having the same $t_{\mathrm{Avg}}(t)$ are aggregated into the same level 2 aggregated state.

The calculation results of ag-ti-TV distances from different initial states are shown in Figures 22 and 23. As the level 2 aggregation is used, the TV distances do not always decrease monotonically. We especially observe the large oscillations in the latter case. Nevertheless, as the magnitude of oscillations rapidly damp to zero, and we can find the MCMT as the latest day at which the TV distance crosses the threshold line at 0.25.

The relationship between $1/r$ and ag-ti-MCMT is shown in Fig. 24. We can observe the proportional relationship between them, which is formally stated as the following observation:

**Observation 5.** The ag-ti-MCMT in Example 4 is proportional to $1/r$ when it is greater (i.e. $r$ is smaller). For smaller $1/r$, the MCMTs are slightly away from a linear relationship, although to a lesser extent.

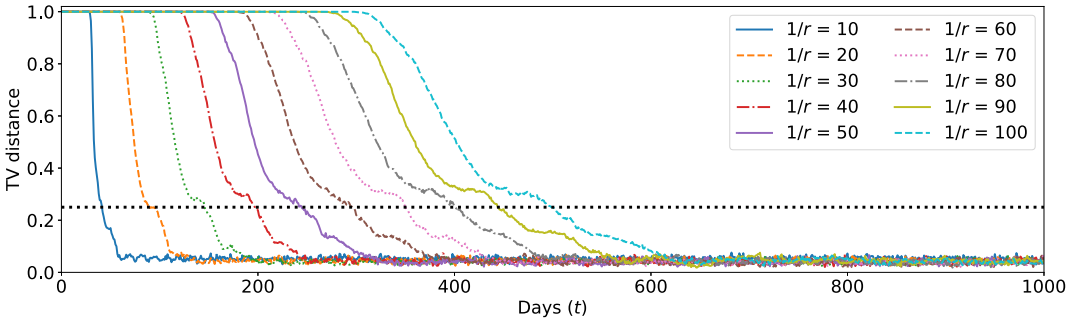


Figure 22: Ag-ti-TV distances (initial dep. time = 3.2) in Example 4

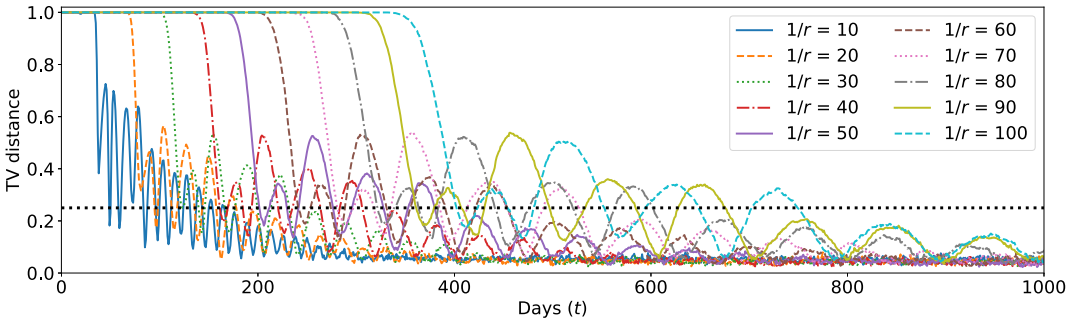


Figure 23: Ag-ti-TV distances (initial dep. time = 4.4) in Example 4

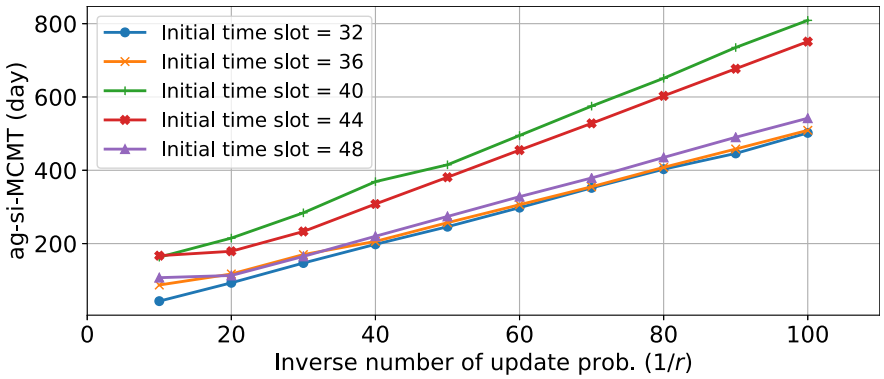


Figure 24: Ag-ti-MCMTs in Example 4

## 6.6 Summary of findings

As a main output from all examples, we first state the following conjecture:

**Conjecture 1.** The ag-o-MCMT (or the ag-si/ti-MCMT as a surrogate of the ag-o-MCMT) is proportional to $1/r$ with coefficient $c_{\text{MCMT}}$ when $r < r_{\max}$, where $c_{\text{MCMT}}$ and $r_{\max}$ is a positive number that may depend on other system parameters such as $\theta$ and $n$.

It is consistent with the theoretical assessment shown in Section 6.1.1. As far as observed in the examples, $r_{\max}$ is not less than 0.1 (except for Example 4, while an approximated proportional relationship is observed for $r \leq 0.1$). Therefore, the conjecture should hold in the wide range of $r$.

The other analytical assessment, i.e. the relationship between the ag-o-MCMT and $n$, does not always fit the numerical results well compared to $r$. This discrepancy would occur owing to the special settings in the analytical assessment. On the other hand, the existence of several cases having (near-)log-linear relationships implies that improving these settings would lead to more accurate analytical outputs.

Approximated values of $c_{\mathrm{MCMT}}$ and its dependency on $\theta$ and $n$ vary in different situation, which is summarised in Table 2. It suggests that, except for the positive interaction case (i.e. Example 2B), $c_{\mathrm{MCMT}}$ is around 1-10, implying that the transport system is close to the stationary distribution after up to 10 times revisions of users' choices. It is surprising that such results can be attained for unstable transport systems like Examples 3 and 4. This implies that the stationary distribution of the Markov Chain can be achieved within a reasonable period of time, even for transport systems with complex interactions between users. On the other hand, comparing the results of Example 2A with those of Examples 2B, 3 and 4 suggests that complex interactions tend to delay convergence. In these examples, there is also a tendency for convergence to get slower as the number of users increases. These facts must be taken into account when MCMTs of transport systems with complex interactions are to be assessed.

The large MCMTs were observed in Example 2B owing to the positive interactions between users. While we could qualitatively forecast such a situation (sometimes called 'lock-in') by a deterministic model, we now have more quantitative results of this situation through MCMTs, i.e. their dependencies on system parameters. Such information may be helpful for evaluating how strongly the transport system falls into a locked state and how we can let it move towards another state, which will attain a better situation, e.g. a public-transport-oriented transport system instead of a car-oriented one.

Table 2: Summary of results of Examples 1 – 4

| Example | Approx. range of $c_{\mathrm{MCMT}}$ [*a] | Dependency on $\theta$ | Dependency on $n$ [*b] | Source figures |
|---|---|---|---|---|
| 1 | 2 – 4 | NA | logarithm | Figs. 3, 4 |
| 2A | $<$ 1 – 3 | decreasing [*c] | almost constant [*c] | Figs. 5 – 7 |
| 2B | 2 – 400 | increasing | faster than log. [*d] | Figs. 10 – 12 |
| 3 | 2 – 8 | increasing | faster than log. [*d] | Figs. 15 – 17 |
| 4 | 4 – 8 [*e] | (not observed) | (not observed) | Fig. 24 |

[*a] subject to parameter settings (especially it should be noted that $n \leq 100$ in Examples 1-3)

[*b] increasing w.r.t. $n$ in all cases

[*c] unless $\theta$ and $r$ are so large that two peaks exist in the stationary distribution (see Figure 8)

[*d] when $\theta$ is greater. Note that linear relationships approximately hold when $\theta$ is smaller.

[*e] upper-bound estimation by level 2 aggregation and coupling

# 7 Conclusion and Future Directions

This study achieved the following two points for MCMT, the time taken for a transport system to converge to a steady state, based on the Markov-chains-based day-to-day dynamics model:

**1**. The method for estimating the upper bounds of the MCMTs was developed by combining the coupling technique and the aggregation approach. It is applicable to large transport systems with the level 2 aggregation. We exhibited this by the departure-time-choice example with a huge amount of level 1 states.

**2**. Numerical and analytical methods were used to observe how the MCMTs depend on several parameters of the day-to-day dynamics. It was generally observed that the MCMTs were inversely proportional to the update probability when it was small, which is consistent with the analytical assessment. The relationship with $n$ varies among problems. The log-linear relationship was found in many cases, while the MCMTs grew faster than that in other cases when the double-peak or doughnuts-style stationary distributions are found in Examples 2 and 3. In addition, MCMTs were not very large in the examples except for Example 2B, implying that it is not unusual for stationary distributions to be achieved within a reasonable period of time.

The achievements of the present study are useful for predicting the convergence time – or its upper bound – to reach a steady state after an arbitrary shock. Such situations can occur in many conditions, such as developing transport infrastructure, introducing a new transport mode, or after a disaster. The ability to predict the time to reach the steady state and evaluate the impact of various parameters on that time is important for planning and operating transport measures in these situations. The proposed method has the potential to be applied to such practical issues in large-scale transport systems.

This study employs a Markov chain model to represent day-to-day dynamics. The Markov chain model is guaranteed to converge to a steady state under mild conditions. This allows the proposed method to be applied to various problems in which convergence to a unique equilibrium solution is not guaranteed. In particular, applications to within-day dynamic traffic assignments are promising. In addition to Vickrey's departure time choice problem, stochastic modelling in DTA (e.g. Satsukawa et al. (2019)) would also be an application of the proposed methodology.

A few future tasks are listed below. The theoretical analysis of MCMTs is limited to simple models in existing studies, and numerical calculations were used in many situations in the present paper. Future theoretical analyses with less dependency on numerical calculations will be demanded. It is also important to implement practical applications to large-scale transport systems. The proposed method can be combined with a traffic simulator with a day-to-day adjustment process, such as MATSim, and would enable us to assess the convergence property of large-scale real-world problems. Accumulating knowledge of the convergence speed in these cases is also demanded to understand its general properties in real-world systems. An important issue for this is to establish a systematic policy for the level 2 aggregation. This point was not discussed much in this paper. It will be necessary in the future to establish a policy that is both practically useful and has good theoretical properties.

It is not easy to investigate how well the proposed MCMTs fit with real data. Hence, we did not address this issue in this study. There are existing studies that have attempted to apply Markov-chain-based day-to-day dynamical models to real data (e.g. Parry et al. (2016)) and have developed positive outputs. This gives a certain justification for the evaluation of convergence time based on the approach taken in the present study. However, this does not directly indicate that the MCMTs are consistent with the real data. It is an important issue that should be addressed in future studies to increase the practical usefulness of the proposed method.

# Acknowledgement

This study was financially supported by JSPS Grant-in-Aid #20H00265.


# References


Arieli, I. and Young, H. P. (2016). Stochastic learning dynamics and speed of convergence in population games. *Econometrica*, 84(2):627–676.

Balijepalli, N., Watling, D., and Liu, R. (2007). Doubly dynamic traffic assignment: Simulation modeling framework and experimental results. *Transportation Research Record*, 2029:39–48.

Cantarella, G., Velon, P., and Watling, D. (2015). Day-to-day dynamics and equilibrium stability in a two-mode transport system with responsive bus operator strategies. *Networks and Spatial Economics*, 15:485–506.

Cantarella, G. and Watling, D. (2016). Modelling road traffic assignment as a day-to-day dynamic, deterministic process: a unified approach to discrete- and continuous-time models. *EURO Journal on Transportation and Logistics*, 5(1):69–98.

Cantarella, G. E. and Cascetta, E. (1995). Dynamic processes and equilibrium in transportation networks: Towards a unifying theory. *Transportation Science*, 29(4):305–329.

Cascetta, E. (1989). A stochastic process approach to the analysis of temporal dynamics in transportation networks. *Transportation Research Part B*, 23(1):1–17.

Cascetta, E. and Cantarella, G. E. (1991). A day-to-day and within-day dynamic stochastic assignment model. *Transportation Research Part A*, 25(5):277–291.

Chien, S. and Sinclair, A. (2011). Convergence to approximate Nash equilibria in congestion games. *Games and Economic Behavior*, 71(2):315–327.

Fukuda, D. and Morichi, S. (2007). Incorporating aggregate behavior in an individual's discrete choice: An application to analyzing illegal bicycle parking behavior. *Transportation Research Part A*, 41(4):313–325.

Guo, R.-Y., Yang, H., and Huang, H.-J. (2018). Are we really solving the dynamic traffic equilibrium problem with a departure time choice? *Transportation Science*, 52(3):603–620.

Hazelton, M. L. (2002). Day-to-day variation in Markovian traffic assignment models. *Transportation Research Part B*, 36(7):637–648.

Hazelton, M. L. and Watling, D. P. (2004). Computation of equilibrium distributions of markov traffic-assignment models. *Transportation Science*, 38(3):331–342.

Hofbauer, J. and Sandholm, W. H. (2007). Evolution in games with randomly disturbed payoffs. *Journal of Economic Theory*, 132(1):47–69.

Iryo, T. (2008). An analysis of instability in a departure time choice problem. *Journal of Advanced Transportation*, 42(3):333–356.

Iryo, T. (2011). Multiple equilibria in a dynamic traffic network. *Transportation Research Part B*, 45(6):867–879.

Iryo, T. (2015). Investigating factors for existence of multiple equilibria in dynamic traffic network. *Networks and Spatial Economics*, 15(3):599–616.

Iryo, T. (2019). Instability of departure time choice problem: A case with replicator dynamics. *Transportation Research Part B*, 126:353–364.

Iryo, T. and Watling, D. (2019). Properties of equilibria in transport problems with complex interactions between users. *Transportation Research Part B*, 126:87 – 114.

Jin, W.-L. (2018). Stable day-to-day dynamics for departure time choice. *https://arxiv.org/abs/1801.09653*.

Jin, W.-L. (2021). Stable local dynamics for day-to-day departure time choice. *Transportation Research Part B*, 149:463–479.

Levin, D. A. and Peres, Y. (2017). *Markov chains and mixing times*, volume 107. American Mathematical Society.

Mohring, H. (1972). Optimization and scale economies in urban bus transportation. *American Economic Review*, 62(4):591–604.

Parry, K., Watling, D. P., and Hazelton, M. L. (2016). A new class of doubly stochastic day-to-day dynamic traffic assignment models. *EURO Journal on Transportation and Logistics*, 5(1):5–23.

Satsukawa, K., Wada, K., and Iryo, T. (2019). Stochastic stability of dynamic user equilibrium in unidirectional networks: Weakly acyclic game approach. *Transportation Research Part B*, 125:229–247.

Sinclair, A. (1992). Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1(4):351–370.

Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, 59(2):251–260.

Watling, D., Milne, D., and Clark, S. (2012). Network impacts of a road capacity reduction: Empirical analysis and model predictions. *Transportation Research Part A: Policy and Practice*, 46(1):167–189.

Watling, D. P. (1996). Asymmetric problems and stochastic process models of traffic assignment. *Transportation Research Part B*, 30(5):339–357.

Watling, D. P. and Cantarella, G. E. (2013). Modelling sources of variation in transportation systems: theoretical foundations of day-to-day dynamic models. *Transportmetrica B: Transport Dynamics*, 1(1):3–32.

Watling, D. P. and Cantarella, G. E. (2015). Model representation and decision-making in an ever-changing world: The role of stochastic process models of transportation systems. *Networks and Spatial Economics*, 15:843–882.

Watling, D. P. and Hazelton, M. L. (2018). Asymptotic approximations of transient behaviour for day-to-day traffic models. *Transportation Research Part B*, 118:90–105.

Zhu, S., Levinson, D., Liu, H. X., and Harder, K. (2010). The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. *Transportation Research Part A*, 44(10):771–784.

# Appendix

## Appendix 1: Identical definition for level 1 aggregation

As noted in the main text, the level 1 aggregation distinguishes disaggregated states by how many users select a choice but not who select it. To characterise this property, we first define an operator called *swapper* that swaps choices of users in the same group. When choices of two users $i, j$ in the same group are swapped, this operation is denoted by $\mathbf{c}^{\text{swap}}(\mathbf{c}, [(i, j)])$. The iterative operation of an arbitrary number of swaps is denoted by $\mathbf{c}^{\text{swap}}(\mathbf{c}, \sigma)$, where $\sigma$ denotes how the choices of users are swapped (e.g. if $\sigma = [(i, j), (k, l)]$, choices of users $i, j$ are swapped first, and then choices of users $k, l$ are swapped). Note that users in the same parenthesis must belong to the same group, while those not in the same parenthesis need not belong to the same group (e.g. users $k, l$ need not belong to the same group as users $i, j$ in the example). We call $\sigma$ the *sequence of swaps* and let $\Sigma_{\text{Swp}}$ be the set of all available sequences of swaps. Note that $\Sigma_{\text{Swp}}$ constitutes a symmetric group, whose elements are permutations that swaps choices of users belonging to the same group.

The following proposition can be stated using the sequence of swaps:

**Proposition 1.** *$A$ is the level 1 aggregation if and only if*

$$\mathbf{c}' = \mathbf{c}^{\text{swap}}(\mathbf{c}, \sigma) \;\; \Leftrightarrow \;\; a(\mathbf{c}; A) = a(\mathbf{c}'; A) \;\; \forall \mathbf{c}, \mathbf{c}' \in C, \;\; \exists \sigma \in \Sigma_{\text{Swp}}. \tag{48}$$

and Equation (12) holds.

*Proof.* Because any swapper does not change $x(\mathbf{c})$, we have

$$\mathbf{c}' = \mathbf{c}^{\text{swap}}(\mathbf{c}, \sigma) \Leftrightarrow \mathbf{x}(\mathbf{c}) = \mathbf{x}(\mathbf{c}') \;\; \exists \sigma \in \Sigma_{\text{Swp}}. \tag{49}$$

This implies that Equations (11) and (48) are identical. □

## Appendix 2: Proof of Theorem 1

Combining Equations (5) and (6), we have

$$p(a, t+1; \mathbf{c}^0) = \sum_{\mathbf{c} \in a} p(\mathbf{c}, t+1; \mathbf{c}^0)$$

$$= \sum_{a' \in A} \sum_{\mathbf{c}' \in a'} \sum_{\mathbf{c} \in a} \prod_{i \in I} \left\{ rK_{c_i}(a') + (1-r)\delta(c_i, c_i') \right\} p(\mathbf{c}', t; \mathbf{c}^0). \tag{50}$$

$K_{c_i}(\mathbf{c}')$ is now replaced with $K_{c_i}(a')$ owing to Equation (12). In addition, Equation (48) implies that the innermost summation of the right-hand side of Equation (50) takes the sum for all combinations of swaps. Therefore, it does not depend on $\mathbf{c}'$ but $a'$ and can be denoted by

$$q_A(a', a) = \sum_{\mathbf{c} \in a} \prod_{i \in I} \left\{ rK_{c_i}(a') + (1-r)\delta(c_i, c_i') \right\}. \tag{51}$$

Substituting $q_A(a', a)$ into Equation (50), we have

$$p(a, t+1; \mathbf{c}^0) = \sum_{a' \in A} q_A(a', a) \sum_{\mathbf{c}' \in a'} p(\mathbf{c}', t; \mathbf{c}^0) = \sum_{a' \in A} q_A(a', a) p(a', t; \mathbf{c}^0). \tag{52}$$

Letting $p(a, t; a^0) = \sum_{\mathbf{c}^0 \in a} p(a, t; \mathbf{c}^0)$, we finally obtain

$$p(a, t+1; a^0) = \sum_{a' \in A} q_A(a', a) p(a', t; a^0), \tag{53}$$

which derives Equation (16). □

## Appendix 3: Proof of Theorem 2 and Corollary 2.1

Consider any pair of $\mathbf{c}^1, \mathbf{c}^2 \in a \in A$, where $A$ is a level 1 aggregation. Let $p(\mathbf{c}^1, t; \mathbf{c}^0) = p(\mathbf{c}^2, t; \mathbf{c}^0)$ and denote them by $p(a, t; \mathbf{c}^0)$. Then,

$$\begin{aligned} &p(\mathbf{c}^1, t+1; \mathbf{c}^0) - p(\mathbf{c}^2, t+1; \mathbf{c}^0) = \\ &\sum_{a' \in A} p(a, t; \mathbf{c}^0) \sum_{\mathbf{c}' \in a'} \left[ \prod_{i \in I} \left\{ rK_{c_i^1}(a') + (1-r)\delta(c_i^1, c_i') \right\} - \prod_{i \in I} \left\{ rK_{c_i^2}(a') + (1-r)\delta(c_i^2, c_i') \right\} \right]. \end{aligned} \tag{54}$$

Because the right-hand side of Equation (54) takes the sum for all combinations of swaps in $a'$, the term in the square brackets is zero.

To prove the corollary, let the initial state of a disaggregated Markov chain be a corner state. It lets the flat probability condition hold when $t = 0$, which leads to the corollary by letting $t \to \infty$ in Theorem 2. □

## Appendix 4: Mixing time with different thresholds of the TV distance

A smaller threshold than 0.25 yields greater $t_C^*$. How $t_C^*$ increases by decreasing the threshold depends on the property of the TV distance, but according to Levin and Peres (2017) (Section 4), it is upper bounded by a value that decays exponentially with respect to $t$. Thus, if this upper bound is a good approximation of the TV distance, using $0.25^k$ as a threshold (where $k \geq 1$) will increase the MCMT by $\alpha(k-1)$, where $\alpha$ is a positive constant.

## Appendix 5: Proof of Theorem 4

The following equations/inequalities, which is derived by Equation (20), prove the theorem:

$$\begin{aligned}
&\|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\mathrm{TV}} = \frac{1}{2}\sum_{a\in A}|p_1(a) - p_2(a)| = \frac{1}{2}\sum_{a^+\in A^+}\sum_{a\in A(a^+)}|p_1(a) - p_2(a)| \\
&\geq \frac{1}{2}\sum_{a^+\in A^+}\left|p_1(a^+) - p_2(a^+)\right| \|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\mathrm{TV}} = \|\mathbf{p}_{A^+}^1 - \mathbf{p}_{A^+}^2\|_{\mathrm{TV}} \qquad (55)
\end{aligned}$$

This equation also holds when $A^+$ and $A$ are replaced by $A$ and $C$, respectively. □

## Appendix 6: Proof of Theorem 5

The flat probability condition implies

$$p(a) = |C(a)|p(\mathbf{c}) \ \ \forall \mathbf{c} \in C(a), \qquad (56)$$

where $C(a)$ is a set of disaggregated states combined to aggregate set $a$ and $|C(a)|$ the number of disaggregated sets in $a$. Therefore, we have

$$\sum_{\mathbf{c}\in C(a)}|p^1(\mathbf{c}) - p^2(\mathbf{c})| = |p^1(a) - p^2(a)|, \qquad (57)$$

which constitutes the theorem. □

## Appendix 7: Proof of Theorem 6

$\delta_{a^+}(\mathbf{p}_A^1, \mathbf{p}_A^2) \geq 0$ owing to the triangle inequality. In addition,

$$\sum_{a \in A(a^+)} |p^1(a) - p^2(a)| = \left| \sum_{a \in A(a^+)} p^1(a) - \sum_{a \in A(a^+)} p^2(a) \right| = |p^1(a^+) - p^2(a^+)| \tag{58}$$

holds for all $a^+ \in \tilde{A}^+$ as either $p^1(a) \geq p^2(a)$ or $p^1(a) \leq p^2(a)$ holds for all $a \in A(a^+)$ by the definition of $\tilde{A}^+$, which leads to $\delta_{a^+}(\mathbf{p}_A^1, \mathbf{p}_A^2) = 0$. □

## Appendix 8: Matrix calculation

The equations applicable for the level 1 aggregation denoted by $A$ is shown below. All the equations can also be applied to disaggregated states $C$ by replacing $A$ with $C$.

While we can use the standard diagonalisation technique to find the power of the transition matrix, the following algorithm is easier to implement and faster when the number of multiplication to find the stationary distribution is not many.

1. Let $k = 1$ and $Q_{A1} = Q_A$.
2. Let $Q_{Ak+1} \leftarrow (Q_{Ak})^2$
3. Check if every pair of columns in $Q_{Ak+1}$ are sufficiently close each other. If so, finish the algorithm. Any column of $Q_{Ak+1}$ includes the stationary distribution.
4. Let $k \leftarrow k+1$ and go back to Step 2.

We can use the bisection method to find an MCMT because the TV distances of disaggregated and level 1 aggregated states are monotonically decreasing with respect to $t$ (see Exercise 4.2 of Levin and Peres (2017)). We can reduce the calculation cost of the bisection search by keeping $Q_{Ak}$ in the memory when finding the stationary distribution. For example, when the TV distance becomes less than 0.25 between $k = 3$ (i.e. $Q_{A3} = Q_A^4$) and $k = 4$ (i.e. $Q_{A4} = Q_A^8$), we can find the TV distance at $t = 6$ by $Q_A^6 = Q_A^2 Q_A^4 = Q_{A2} Q_{A3}$.

## Appendix 9: Accuracy estimation of the Monte Carlo method

To roughly estimate how the error is correlated with the number of states, consider two probability distributions $p_1(a)$ and $p_2(a)$ and the frequencies of occurrence by these probabilities, denoted by $f_1(a)$ and $f_2(a)$, respectively. Approximating the distributions of $f_1(a)$ and $f_2(a)$ by the binomial distribution, i.e. $f_1(a) \sim \mathrm{Bi}(p_1(a), N_{\mathrm{sim}})$ and $f_2(a) \sim \mathrm{Bi}(p_2(a), N_{\mathrm{sim}})$, we have

$$\frac{f_1(a) - f_2(a)}{N_{\mathrm{sim}}} \sim \mathrm{N}\left( p_1(a) - p_2(a), \frac{p_1(a) + p_2(a)}{N_{\mathrm{sim}}} \right) \tag{59}$$

as an estimated difference between $p_1(a)$ and $p_2(a)$, when $N_{\mathrm{sim}}$ is large and $p_1(a)$ and $p_2(a)$ are small so that the binomial distributions can be approximated by normal distributions. The relative standard deviation of this value, denoted by

$$\sigma_{\mathrm{rd}}(a) = \frac{1}{|p_1(a) - p_2(a)|} \sqrt{\frac{p_1(a) + p_2(a)}{N_{\mathrm{sim}}}} \tag{60}$$

should be sufficiently less than one to maintain accuracy in estimating TV distances and MCMTs. This $\sigma_{\mathrm{rd}}$ can be lower bounded as follows.

$$\sigma_{\mathrm{rd}} \geq \frac{1}{\sqrt{0.5 N_{\mathrm{sim}} p_{\mathrm{m}}(a)}}. \tag{61}$$

This result implies that we need to employ a number that is sufficiently greater than $p_{\mathrm{m}}^{-1}$ as $N_{\mathrm{sim}}$.