Network throughput under dynamic user equilibrium: Queue spillback, paradox and traffic control☆

Kentaro Wada[a,b,*], Koki Satsukawa[c], Mike Smith[d], Takashi Akamatsu[e]

[a]*Institute of Industrial Science, The University of Tokyo, Tokyo, Japan*
[b]*Institute of Transportation Studies, University of California, Irvine, CA, United States.*
[c]*Department of Civil Engineering, The University of Tokyo, Tokyo, Japan*
[d]*Department of Mathematics, University of York, York, UK*
[e]*Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan*

Abstract

This study aims to analyze the relationship between a macroscopic fundamental diagram (MFD), which relates vehicle accumulation with throughput at the network level, and the spatial distribution of congestion (congestion pattern) in a general network with one-to-many origin-destination demands. In particular, we clarify the causes of a decreasing branch of MFDs and the influence of local signal controls on the (global) network throughput. For this aim, we present a new inverse problem of the dynamic user equilibrium assignment by using a periodic boundary condition, and an analytical formula of the network throughput for a fixed accumulation is derived by solving it. This enables us to incorporate the effects of network configurations and route choice behaviors into the analysis of the network throughput. By conducting a sensitivity analysis of this formula, we identify the types of congestion patterns that cause the decrease in the network throughput and examine a network signal control for improving network performance.

Keywords: network throughput, Macroscopic Fundamental Diagram, network exit function, dynamic user equilibrium, equilibrium paradox, network signal control

1. Introduction

Network-level relationships among traffic variables were introduced by Godfrey (1969) and their static properties (averages over long times) have been examined by several studies (e.g., Herman and Prigogine, 1979; Ardekani and Herman, 1987; Mahmassani et al., 1987). To predict the network performance dynamically, Daganzo (2007) later reintroduced a Macroscopic Fundamental Diagram (MFD) showing a steady-state functional relationship between vehicle accumulation (or average network density) and network throughput (or average network flow) within certain networks. Geroliminis and Daganzo (2008) then demonstrated that a well-defined

☆Acknowledgements. The authors express their gratitude to the anonymous referees for their careful reading of the manuscript and useful suggestions. The first author was supported by the Kajima Foundation's Assistance for Exchange of Researcher. The first and fourth authors were supported by JSPS KAKENHI Grants No.15H04053 and No.16H02368.

*Corresponding author. Tel.: +81-3-5452-6098 (ext. 58171); Fax: +81-3-5452-6420.

Email addresses: wadaken@iis.u-tokyo.ac.jp (Kentaro Wada), kouki@iis.u-tokyo.ac.jp (Koki Satsukawa), 152mikesmith@gmail.com (Mike Smith), akamatsu@plan.civil.tohoku.ac.jp (Takashi Akamatsu)

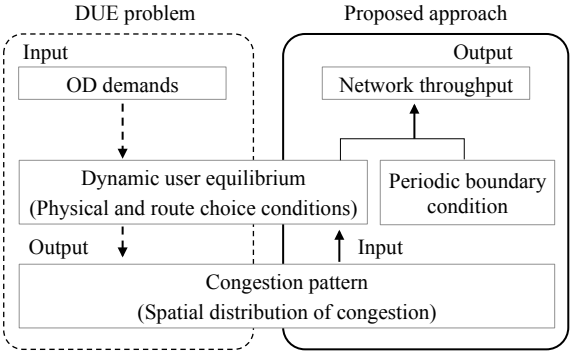


Figure 1: Framework of this study

MFD, which is reproducible and invariant when the demand changes, exists within a field experiment in downtown Yokohama, Japan. Although the MFD has been considered useful for establishing robust network-wide traffic control, the microscopic mechanisms behind the macroscopic behaviors of an MFD are not completely understood.

With regard to this issue, several studies have shown that the spatial distribution of congestion (hereon referred to as a "congestion pattern") in a network is one of the key factors in defining the shape of an MFD. Geroliminis and Sun (2011b) observed a well-defined MFD when the variance of link density was the same for different time periods with the same network density. For networks where users choose their routes randomly, Mazloumian et al. (2010) and Daganzo et al. (2011) showed that traffic congestion tends to distribute unevenly, which leads to a decrease in the network throughput. Because such an uneven congestion pattern may be induced by route choices, heterogeneous local capacities and controls, Leclercq and Geroliminis (2013) analyzed the effect of route choices on an MFD by combining an analytical approximation method of the MFD (Daganzo and Geroliminis, 2008) with equilibrium route choice models. However, their analysis was restrictive, i.e., they only considered a parallel route network. The relationship between a decrease in the network throughput and congestion patterns in general networks associated with route choice behaviors, local control and network topology is largely unknown.

The purpose of this study is to analyze the relationship between an MFD and the congestion patterns in a general network with one-to-many origin-destination demands. Of particular interests are the causes of a decreasing branches of MFDs and the influences of network signal control on MFDs. To achieve this purpose, we present a new inverse problem of the dynamic user equilibrium assignment using a periodic boundary condition (see Figure 1). The proposed problem is formulated as a system of linear equations; by solving this, we can derive an analytical formula for the network throughput that includes the effect of route choice. Thus, through congestion patterns, we can incorporate the effects of network configurations and route choice behaviors into the analysis of the network throughput. By conducting a sensitivity analysis of this formula, we identify the types of congestion patterns that cause a decreasing in network throughput, and further examine a signal control that could improve network performance.

The remainder of this paper is organized as follows. In section 2, we describe the dynamic user equilibrium assignment and its analytical solution for a given congestion pattern. Section 3 derives an analytical formula of the network throughput for a given congestion pattern and a fixed vehicle accumulation (i.e., steady-state condition). We also show that the network through-

put under a dynamic condition can be approximated by a similar formula under the steady-state. By sensitivity analyses of the proposed formula, Section 4 identifies the conditions for the decrease in the network throughput and for the occurrence of a capacity increasing paradox. Based on these conditions, we further explore a signal control policy for improving the network performance. Section 5 shows numerical examples for testing the present theory. Section 6 concludes the paper.

2. Preliminaries: Dynamic user equilibrium and its analytical solution

2.1. Network and notation

Consider a general network $\mathcal{G}[\mathcal{N}, \mathcal{L}, \mathcal{W}]$ with one-to-many origin-destination (OD) demands consisting of the set of nodes $\mathcal{N}$, the set of links $\mathcal{L}$ and the set of OD pairs $\mathcal{W}$. The elements of these sets are indicated by $i \in \mathcal{N}$, $(i, j) \in \mathcal{L}$ and $(o, d) \in \mathcal{W}$, respectively. A node-link incidence matrix $\mathbf{A}^*$ (N $\times$ L matrix) represents the structure of the network. We also define a reduced incidence matrix $\mathbf{A}$ by eliminating an arbitrary row of $\mathbf{A}^*$. We call the node corresponding to the elimination reference node. As in Akamatsu (2001), we select the unique origin node as the reference node. In addition, we define a matrix $\mathbf{A}_-$, which is obtained by replacing all the $+1$ elements of $\mathbf{A}$ with zero; similarly, $\mathbf{A}_+$ is a matrix obtained by letting all the -1 elements of $\mathbf{A}$ be zero.

We assume that the time-dependent demand for each OD pair is exogenous. Specifically, the cumulative flow departing from the origin o by time t and arriving at destination d is denoted by $Q_{od}(t)$. For a link model in our dynamic assignment, we employ the First-In-First-Out (FIFO) principle and the *point queue* concept. It is assumed that each link (i, j) comprises a free-flow section with a constant travel time m_{ij} and a bottleneck section with a constant capacity μ_{ij}. Then, the link travel time for a user entering the link at time t is given by

$$c_{ij}(t) = m_{ij} + \frac{x_{ij}(t + m_{ij})}{\mu_{ij}}, \qquad \text{where} \quad x_{ij}(t) = A_{ij}(t - m_{ij}) - D_{ij}(t), \tag{2.1}$$

$x_{ij}(t)$ is the amount of traffic in the queue at time t, $A_{ij}(t)$ and $D_{ij}(t)$ are the cumulative inflow and outflow for the link by time t, respectively. If $A_{ij}(t)$ and $D_{ij}(t)$ are differentiable the inflow and outflow rates are expressed as

$$\lambda_{ij}(t) \equiv \frac{\mathrm{d}A_{ij}(t)}{\mathrm{d}t}, \quad f_{ij}(t) \equiv \frac{\mathrm{d}D_{ij}(t)}{\mathrm{d}t}. \tag{2.2}$$

2.2. Decomposed formulation of dynamic user equilibrium assignment

The dynamic user equilibrium (DUE) is defined as the state, in which no user can reduce his/her travel time by changing his/her route (see Smith, 1993; Heydecker and Addison, 1996). The equilibrium concept along with the FIFO principle of the link model implies that users who depart from the origin at the same time have the same arrival time at any nodes, and that the order of departure from the origin must be kept at any nodes on the way to their destinations (Kuwahara and Akamatsu, 1993). Using this property, the DUE assignment with one-to-many pairs can be decomposed with respect to the departure time s from the origin.

For the decomposed DUE assignment, two variables are introduced: $\tau_i(s)$ is the earliest arrival time at node i for a user who departs from the origin at time s, and $y_{ij}(s)$ is the link flow rate with respect to the origin departure time s:

$$y_{ij}(s) \equiv \frac{\mathrm{d}A_{ij}(\tau_i(s))}{\mathrm{d}s} = \lambda_{ij}(\tau_i(s)) \cdot \dot{\tau}_i(s), \tag{2.3}$$

where $\dot{\tau}_i(s) \equiv \mathrm{d}\tau_i(s)/\mathrm{d}s$. Then, the equilibrium for every origin departure time s can be expressed with the following three conditions (for derivation, see Akamatsu, 2001). The first is the link travel time function:

$$c_{ij}(s) = \int_0^s \dot{c}_{ij}(s)\,\mathrm{d}s + c_{ij}(0) \qquad \forall (i,j) \in \mathcal{L},\ \forall s \tag{2.4}$$

$$\text{where}\ \ \dot{c}_{ij}(s) = \begin{cases} y_{ij}(s)/\mu_{ij} - \dot{\tau}_i(s) & \text{if}\ \ x_{ij}(\tau_i(s) + m_{ij}) > 0 \\ 0 & \text{if}\ \ x_{ij}(\tau_i(s) + m_{ij}) = 0 \end{cases} \tag{2.5}$$

where $\dot{c}_{ij}(s) \equiv \mathrm{d}c_{ij}(\tau_i(s))/\mathrm{d}s$. The second is the flow conservation condition at each node:

$$\textstyle\sum_{i \in I(k)} y_{ik}(s) - \sum_{j \in O(k)} y_{kj}(s) - \dot{Q}_{ok}(s) = 0 \qquad \forall k \in \mathcal{N} \setminus \{o\},\ \forall s, \tag{2.6}$$

where $I(k)$ [$O(k)$] is the set of upstream [downstream] nodes of links arriving at [leaving] the node k, and $\dot{Q}_{ok}(s) \equiv \mathrm{d}Q_{ok}(s)/\mathrm{d}s$. The final condition is the user's route choice condition:

$$\begin{cases} y_{ij}(s)\left\{c_{ij}(s) + \tau_i(s) - \tau_j(s)\right\} = 0 \\ c_{ij}(s) + \tau_i(s) - \tau_j(s) \geq 0,\ y_{ij}(s) \geq 0 \end{cases} \qquad \forall (i,j) \in \mathcal{L},\ \forall s. \tag{2.7}$$

For later convenience, we also show the vector-matrix form of the formulation. That is,

$$\mathbf{m} \leq \mathbf{c}(s) \perp (\dot{\mathbf{c}}(s) - \mathbf{M}^{-1}\mathbf{y}(s) - \mathbf{A}_+^T \dot{\boldsymbol{\tau}}(s)) \geq \mathbf{0} \qquad \forall s \tag{2.8}$$

$$\mathbf{A}\mathbf{y}(s) = -\dot{\mathbf{Q}}(s) \qquad \forall s \tag{2.9}$$

$$\mathbf{0} \leq \mathbf{y}(s) \perp (\mathbf{c}(s) + \mathbf{A}^T \boldsymbol{\tau}(s)) \geq \mathbf{0} \qquad \forall s \tag{2.10}$$

where $\mathbf{M}$ is the L × L diagonal matrix of the link capacities $\{\mu_{ij}\}$; $\mathbf{m}$, $\mathbf{c}(s)$, and $\mathbf{y}(s)$ are L-dimensional column vectors with elements m_{ij}, $c_{ij}(s)$, and $y_{ij}(s)$, respectively; $\boldsymbol{\tau}(s)$ and $\dot{\boldsymbol{\tau}}(s)$ are the N−1-dimensional column vector with elements[1] $\tau_i(s)$ and $\dot{\tau}_i(s)$, respectively; $\dot{\mathbf{Q}}(s)$ is the column vector with elements $\dot{Q}_{od}(s)$.

The important characteristic of this DUE formulation is that treating the complicated nested structure between link and route travel times in a network is not required, unlike the standard DUE formulation with respect to the absolute time t (for a detailed discussion on this issue, see Akamatsu et al., 2015). Consequently, the conditions (2.6) and (2.7) are almost the same as those of the static traffic assignment with an additive cost function. This property enables us to analyze the theoretical properties of the DUE assignment in a mathematically tractable way, as shown in this study (for other examples, see Akamatsu and Kuwahara, 1999; Akamatsu et al., 2015).

[1] Note that $\tau_o(s) = s$ and $\dot{\tau}_o(s) = 1$ by definition.

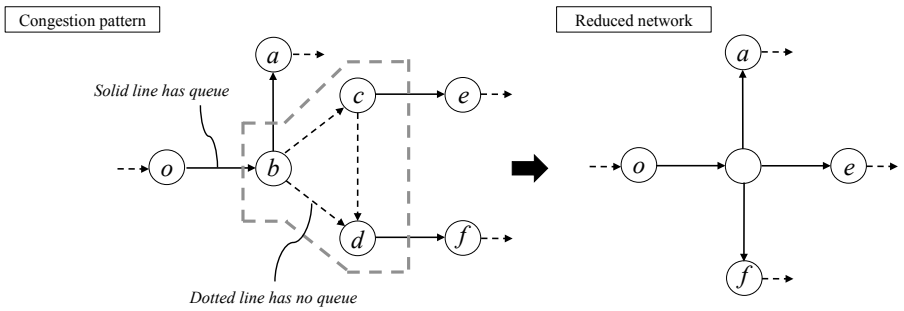


Figure 2: Example of constructing reduced network

2.3. Equilibrium solution for saturated networks

The DUE assignment is formulated as a variational inequality problem and cannot be analytically solved in general. However, there is a case that an analytical solution can be obtained. In particular, we consider *saturated networks* which satisfy the following two conditions: (i) all links in a network have positive inflows ($y_{ij}(s) > 0$), and (ii) all links have queues ($x_{ij}(\tau_i(s)) > 0$). In a saturated network, the decomposed DUE assignment is reduced to a system of linear equations, and thus an analytical solution can be obtained (Akamatsu and Heydecker, 2003b).

More specifically, the complementarity conditions (2.8) and (2.10) are reduced to the following equality conditions because all links in the network have positive flows and queues:

$$\dot{\mathbf{c}}(s) - \mathbf{M}^{-1}\mathbf{y}(s) - \mathbf{A}_{+}^{T}\dot{\boldsymbol{\tau}}(s) = \mathbf{0}, \tag{2.11}$$

$$\mathbf{c}(s) + \mathbf{A}^{T}\boldsymbol{\tau}(s) = \mathbf{0}. \tag{2.12}$$

By taking the derivative of the second equation with respect to s and substituting it into the first equation, we have

$$\mathbf{y}(s) = -(\mathbf{M}\mathbf{A}_{-}^{T})\,\dot{\boldsymbol{\tau}}(s). \tag{2.13}$$

Further, substituting this equation into the flow conservation (2.9), we obtain

$$(\mathbf{A}\mathbf{M}\mathbf{A}_{-}^{T})\,\dot{\boldsymbol{\tau}}(s) = \dot{\mathbf{Q}}(s). \tag{2.14}$$

From this equation, we can see that the DUE solution $\dot{\boldsymbol{\tau}}(s)$ is uniquely determined if the rank of the matrix $\mathbf{A}\mathbf{M}\mathbf{A}_{-}^{T}$ is N$-$1. As shown by Akamatsu (2000), for networks with one-to-many OD demands, the rank must be N$-$1 if we eliminate the origin as the reference node. Therefore, we can obtain the equilibrium solution analytically: $\dot{\boldsymbol{\tau}}(s) = (\mathbf{A}\mathbf{M}\mathbf{A}_{-}^{T})^{-1}\dot{\mathbf{Q}}(s)$. Note that the equilibrium link flows $\mathbf{y}(s)$ are determined by substituting this equation into Eq.(2.13).

The concept of saturated network can be applied to any non-saturated network by reducing the network to a saturated one, where links with no queue are appropriately removed: (a) unifying the initial and terminal nodes of each unsaturated link on the non-saturated network into a single node; (b) removing links with no flow (see Figure 2). This saturated network is called a *reduced network*. As proved in Akamatsu and Heydecker (2003a) (see also Appendix A), the solution $\dot{\boldsymbol{\tau}}(s)$ *defined on a reduced network* is also governed by Eq.(2.14) in which $(\mathbf{A}\mathbf{M}\mathbf{A}_{-}^{T})$ and

$\dot{\mathbf{Q}}(s)$ are also defined on the reduced network; the DUE solution is uniquely determined (but the equilibrium link flows for unsaturated links may not be unique).

Henceforth, we consider both non-saturated and saturated networks. When a network is non-saturated, we consider the corresponding reduced network to obtain the analytical solution of the DUE assignment. Meanwhile, when a network is saturated, the corresponding reduced network is exactly the same as the original saturated network. Thus, all the variables used in this manuscript must be considered as variables *defined on reduced networks*. Note that loops could arise in a reduced network, even if its original non-saturated network contains no loops[2]. If there is a link whose terminal node is the origin, the outflow information from its initial node disappears by eliminating the row of $\mathbf{A}^*$ corresponding to the origin node. This implies that the flow conservation at the initial node of this link does not hold. To resolve this problem, we need to regard this outflow as an OD demand-like flow. Let $\boldsymbol{\delta}$ be the column vector, whose i-th element is the link capacity μ_{io}, if the link (i, o) exists in the reduced network (zero otherwise). Then, Eq.(2.14) for the reduced network is modified as follows.

$$(\mathbf{AMA}_{-}^{T})\, \dot{\boldsymbol{\tau}}(s) = \dot{\mathbf{Q}}(s) + \boldsymbol{\delta}. \tag{2.15}$$

3. Analytical formulas of network throughput

In this section, we present the connection between a congestion pattern on a network (i.e., reduced network structure) and macroscopic traffic flow at a network level (i.e., network throughput). The basis of this theory is constructing an inverse problem of the DUE assignment by employing a periodic boundary condition. By solving this inverse problem, we can obtain an analytical formula of a steady-state network throughput (i.e., the sum of possible OD flows) that is consistent with a given congestion pattern and the route choice principle.

This analytical formula describes the network throughput for the situation where a fixed number of drivers circulate in the network indefinitely (or demands and traffic conditions change slowly with time), and can be interpreted as a steady-state relationship between vehicle accumulation and network throughput (or trip completion rate) within an underlying network (called a *Network Exit Function* (NEF) or outflow-MFD in the literature, e.g., Daganzo, 2007). Furthermore, the idea of giving the congestion pattern is motivated by the empirical finding indicating that a (reproducible) spatial distribution of congestion is a key component defining the shape of an MFD (or NEF) (see Geroliminis and Sun, 2011b). In brief, we here show a way to connect the conventional theory of traffic assignment to MFD theory via congestion patterns.

In Section 3.1, we formulate the inverse problem. We then present the analytical formula of the network throughput under the steady-state condition in Section 3.2. In Section 3.3, we derive the network throughput approximately by relaxing the assumption of a fixed vehicle accumulation and further investigate the effect of accumulation dynamics on the network throughput.

Note that we analyze the NEF throughout the paper instead of the relationship between accumulation and production (sometimes called a production MFD). This is because the former represents the network performance directly but the latter does not. Although the production (i.e., the total distance travelled per unit time in the network) can be a good proxy of the network throughput under a certain condition[3], this is not always the case for, especially inhomogeneous

[2]For saturated networks, no loops arise in the DUE flow patterns (see Akamatsu and Heydecker, 2003b).

[3]When the average trip length is the same for all ODs, the production-MFD is proportional to the NEF (Daganzo, 2007).

networks. An analysis of the relation between production and network throughput can be found, for example, in Yildirimoglu and Geroliminis (2014) and Leclercq et al. (2015).

3.1. Inverse problem

We first express macroscopic traffic variables, which are the cumulative network inflow $A_d(t)$ and outflow $D_d(t)$ for each destination d, with the variables used in the DUE formulation.

$$A_d(t) = Q_{od}(t), \quad D_d(t) = Q_{od}(t - C_d^*(t)) \qquad \forall d \in \mathcal{N}_d, \tag{3.1}$$

where $C_d^*(t)$ is the equilibrium OD travel time for users arriving at destination d at time t and $\mathcal{N}_d$ is the set of destinations. By using these variables, the vehicle accumulation for each destination $n_d(t)$ is expressed as $n_d(t) = A_d(t) - D_d(t)$. If these variables are differentiable, then the dynamics of vehicle accumulation are given by

$$\frac{\mathrm{d}n_d(\tau_d(s))}{\mathrm{d}t} = \lambda_d(\tau_d(s)) - f_d(\tau_d(s)) \qquad \forall d \in \mathcal{N}_d, \tag{3.2}$$

$$\text{where} \quad \lambda_d(\tau_d(s)) \equiv \frac{\mathrm{d}A_d(\tau_d(s))}{\mathrm{d}t} = \frac{\mathrm{d}Q_{od}(\tau_d(s))}{\mathrm{d}t} \tag{3.3}$$

$$f_d(\tau_d(s)) \equiv \frac{\mathrm{d}D_d(\tau_d(s))}{\mathrm{d}t} = \frac{\mathrm{d}Q_{od}(\tau_d(s) - (\tau_d(s) - s))/\mathrm{d}s}{\mathrm{d}\tau_d(s)/\mathrm{d}s} = \frac{\dot{Q}_{od}(s)}{\dot{\tau}_d(s)}. \tag{3.4}$$

Following, an inverse problem is constructed for a given congestion pattern. The congestion pattern is expressed by the reduced network; thus, the inputs in this problem are (i) the reduced node-link incidence matrix $\mathbf{A}$ (i.e., the topology of the reduced network) and (ii) the link capacity matrix $\mathbf{M}$ (i.e., the capacity pattern of the reduced network). For the topology of the reduced network, we assume that the origin and each destination are not unified into a single node in a reduced network. This assumption is introduced in order to exclude the situation in which a user can reach his/her destination with a free flow route travel time. This is because, in that case, the network throughput is mainly determined by the OD demands and not by the congestion patterns. In this sense, our DUE approach is suitable for analyzing the congested (or critical) regime of MFD but not the free flow regime.

As in the standard inverse problem of traffic assignments, the proposed inverse problem requires an additional condition to determine the network throughput uniquely[4]. We thus introduce a *periodic boundary condition*, which was employed in the previous studies (e.g., Daganzo and Geroliminis, 2008), for obtaining a steady-state network throughput for a fixed vehicle accumulation. This condition for each destination is expressed by

$$\frac{\mathrm{d}n_d(\tau_d(s))}{\mathrm{d}t} = 0 \quad \Leftrightarrow \quad \frac{\mathrm{d}Q_{od}(\tau_d(s))}{\mathrm{d}t} = \frac{\dot{Q}_{od}(s)}{\dot{\tau}_d(s)} \qquad \forall d \in \mathcal{N}_d. \tag{3.5}$$

We further assume that the input values change slowly compared with the system's relaxation time[5], which implies

$$\frac{\mathrm{d}Q_{od}(\tau_d(s))}{\mathrm{d}t} = \dot{Q}_{od}(s) \qquad \forall d \in \mathcal{N}_d. \tag{3.6}$$

[4]The standard inverse problem of traffic assignments requires a criterion such as error minimization and entropy maximization.

By combining Eq.(3.5) and Eq.(3.6), we obtain the additional condition for the problem.

$$\dot{\tau}_d(s) = 1 \qquad \forall d \in \mathcal{N}_d. \tag{3.7}$$

Let $\mathbf{f} \equiv \{f_i\}_{i\in\mathcal{N}}$ be the N−1-dimensional column vector whose i-th element represents the network throughput for OD pair (o, i), if node i is the destination in the reduced network (zero otherwise). Finally, the proposed inverse problem is formulated as (here we omit the time index s because the steady-state is considered)

$$\begin{cases} \mathbf{f} + \boldsymbol{\delta} = (\mathbf{A}\mathbf{M}\mathbf{A}_-^T)\,\dot{\boldsymbol{\tau}} \\ \dot{\boldsymbol{\tau}} = [\ \dot{\boldsymbol{\tau}}_i \mid \dot{\boldsymbol{\tau}}_d\ ]^T = [\ \dot{\boldsymbol{\tau}}_i \mid \mathbf{1}\ ]^T \end{cases}, \tag{3.8}$$

where $\dot{\boldsymbol{\tau}}_i$ and $\dot{\boldsymbol{\tau}}_d$ are sub-vectors of $\dot{\boldsymbol{\tau}}$ with respect to the transient node set $\mathcal{N}_i$ and the destination node set $\mathcal{N}_d$, respectively (i.e., $\mathcal{N} = \{o\} \cup \mathcal{N}_i \cup \mathcal{N}_d$). The first equation represents the DUE condition (2.15), wherein the OD demand vector $\dot{\mathbf{Q}}(s)$ is replaced by the vector $\mathbf{f}$. The second equation represents the condition (3.7). To see this problem more precisely, we also divide $\mathbf{f}$, $\boldsymbol{\delta}$, $\mathbf{A}$ and $\mathbf{A}_-$ into two blocks corresponding to $\mathcal{N}_i$ and $\mathcal{N}_d$, respectively:

$$\mathbf{f} = \left[\ \mathbf{f}_i \mid \mathbf{f}_d\ \right]^T,\quad \boldsymbol{\delta} = \left[\ \boldsymbol{\delta}_i \mid \boldsymbol{\delta}_d\ \right]^T,\quad \mathbf{A} = \left[\ \mathbf{A}_i \mid \mathbf{A}_d\ \right]^T,\quad \mathbf{A}_- = \left[\ \mathbf{A}_{i-} \mid \mathbf{A}_{d-}\ \right]^T,$$

where $\mathbf{f}_i$ is always $\mathbf{0}$ by definition. Rewriting (3.8) with these partitioned variables, we have

$$\left[\frac{\boldsymbol{\delta}_i}{\mathbf{f}_d + \boldsymbol{\delta}_d}\right] = \left[\frac{\mathbf{A}_i}{\mathbf{A}_d}\right]\mathbf{M}\left[\ \mathbf{A}_{i-}^T \ \middle|\ \mathbf{A}_{d-}^T\ \right]\left[\frac{\dot{\boldsymbol{\tau}}_i}{\mathbf{1}}\right] = \left[\frac{\mathbf{A}_i\mathbf{M}\mathbf{A}_{i-}^T \ \middle|\ \mathbf{A}_i\mathbf{M}\mathbf{A}_{d-}^T}{\mathbf{A}_d\mathbf{M}\mathbf{A}_{i-}^T \ \middle|\ \mathbf{A}_d\mathbf{M}\mathbf{A}_{d-}^T}\right]\left[\frac{\dot{\boldsymbol{\tau}}_i}{\mathbf{1}}\right]. \tag{3.9}$$

For notational brevity, we further define $\mathbf{V} \equiv \mathbf{A}\mathbf{M}\mathbf{A}_-^T$ and

$$\mathbf{V} = \left[\frac{\mathbf{V}_{ii} \ \middle|\ \mathbf{V}_{id}}{\mathbf{V}_{di} \ \middle|\ \mathbf{V}_{dd}}\right] \equiv \left[\frac{\mathbf{A}_i\mathbf{M}\mathbf{A}_{i-}^T \ \middle|\ \mathbf{A}_i\mathbf{M}\mathbf{A}_{d-}^T}{\mathbf{A}_d\mathbf{M}\mathbf{A}_{i-}^T \ \middle|\ \mathbf{A}_d\mathbf{M}\mathbf{A}_{d-}^T}\right].$$

3.2. Network throughput under steady-state

By solving the inverse problem (3.9), we have the following proposition for the network throughput, F.

Proposition 1. *For a given congestion pattern* $(\mathbf{A}, \mathbf{M})$ *and a fixed vehicle accumulation, the steady-state network throughput of a one-to-many network is given by the following formula:*

$$F \equiv \mathbf{1}^T\mathbf{f}_d = \mathbf{1}^T\mathbf{V}_{dd}\mathbf{1} - \mathbf{1}^T[\mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}(\mathbf{V}_{id}\mathbf{1} - \boldsymbol{\delta}_i) + \boldsymbol{\delta}_d]. \tag{3.10}$$

Proof. The equation (3.9) implies

$$\boldsymbol{\delta}_i = \mathbf{V}_{ii}\dot{\boldsymbol{\tau}}_i + \mathbf{V}_{id}\mathbf{1}, \tag{3.11}$$

$$\mathbf{f}_d + \boldsymbol{\delta}_d = \mathbf{V}_{di}\dot{\boldsymbol{\tau}}_i + \mathbf{V}_{dd}\mathbf{1}. \tag{3.12}$$

Because the matrix $\mathbf{V}_{ii}$ is invertible (Akamatsu and Heydecker, 2003a), Eq.(3.11) can be solved:

$$\dot{\boldsymbol{\tau}}_i = -(\mathbf{V}_{ii})^{-1}[\mathbf{V}_{id}\mathbf{1} - \boldsymbol{\delta}_i]. \tag{3.13}$$

By substituting (3.13) into (3.12) and summing the elements of vector $\mathbf{f}_d$, we obtain the Eq.(3.10). □

[5]The relaxation time is comparable with the travel time across the network (Daganzo, 2007).

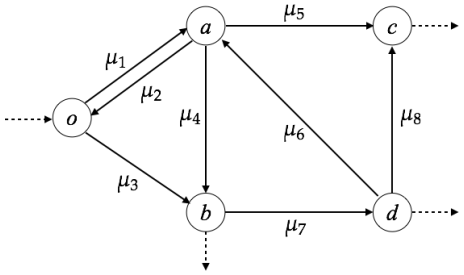


Figure 3: A example of network with single origin and three destinations

From the analytical formula (3.10), we see that the network throughput is determined by the topology (i.e., the relation of connection between saturated links) and capacity pattern of the reduced network and the locations of origin/destination nodes in the network[6]. To explore the formula (3.10) in more detail, let us look into its element form. If $[\mathbf{V}]_{kl}$ is the (k,l) element of $\mathbf{V}$, then

$$[\mathbf{V}]_{kl} = \begin{cases} \sum_{j\in I(k)} \mu_{jk} & \text{if } k = l \\ -\mu_{kl} & \text{if } k \neq l \text{ and } (k,l) \in \mathcal{L} \\ 0 & \text{otherwise} \end{cases}. \tag{3.14}$$

Thus Eq.(3.10) can be written as:

$$F = \sum_{d\in\mathcal{N}_d} \left[\sum_{k\in\{I(d)\cap\{\mathcal{N}_i\cup\{o\}\}\}} \mu_{kd} - \sum_{k\in\{O(d)\cap\{\mathcal{N}_i\cup\{o\}\}\}} \mu_{dk}\dot{\tau}_k \right] \tag{3.15}$$

The first term of this equation represents the sum of the capacities of links incoming from non-destination nodes to destination nodes. The second term represents the sum of outflows[7] from destination nodes to non-destination nodes (i.e., the sum of flows passing through the destinations). Thus, the formula (3.10) expresses the flow conservation at the destination nodes [8]. Note, however, that the formula describes not only the "local" conservation law but also the "global" phenomena such as route choice effects through the DUE solution $\dot{\tau}_k$ in the second term[9].

Example. Consider the reduced network shown in Figure 3. Node o is the origin and nodes $\{b,c,d\}$ are destinations. The capacity of link e ($e = 1,\ldots,8$) is given by μ_e. The matrices and

[6]The first two information can be obtained from the topology of the original network and the flow of each link (that is also required for drawing MFDs). Although the last is an OD related information, this is static and may be reasonably assumed, for example, based on the OD census in practice.

[7]Under the DUE state, the link flow $y_{ij}(s)$ should satisfy the following relationship: $y_{ij}(s) = \mu_{ij}\dot{\tau}_j(s)$ (i.e., the element form of Eq.(2.13)).

[8]We can obtain a "value" of the network throughput from flow conservation at the single origin under the steady-state. However, the destination-based formulation (3.10) is required for analyzing the decreasing mechanisms of the throughput (i.e., the interactions between users with *different destinations*), as we will discuss in Section 4.

[9]For single OD networks, where no loops occur in reduced networks such as parallel networks (e.g., Leclercq and Geroliminis, 2013), the network throughput is determined only by the first term (local condition), which implies that the network structure and/or route choice may not significantly affect the network throughput.

vectors in the formula (3.10) are given by

$$\mathbf{V}_{dd} = \begin{bmatrix} \mu_3 + \mu_4 & 0 & -\mu_7 \\ 0 & \mu_5 + \mu_8 & 0 \\ 0 & -\mu_8 & \mu_7 \end{bmatrix}, \ \mathbf{V}_{di} = \begin{bmatrix} 0 \\ 0 \\ -\mu_6 \end{bmatrix}, \ \mathbf{V}_{ii} = \begin{bmatrix} \mu_1 + \mu_6 \end{bmatrix},$$

$$\mathbf{V}_{id} = \begin{bmatrix} -\mu_4 & -\mu_5 & 0 \end{bmatrix}, \ \boldsymbol{\delta}_i = \begin{bmatrix} \mu_2 \\ 0 \end{bmatrix}, \boldsymbol{\delta}_d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

By substituting these into (3.10), we derive the network throughput:

$$F = \mu_3 + \mu_4 + \mu_5 - \mu_6 \dot{\tau}_d \qquad \text{where} \quad \dot{\tau}_d = \frac{\mu_2 + \mu_4 + \mu_5}{\mu_1 + \mu_6}.$$

From this equation, we see that the network throughput is affected not only by the links connected with the destinations but also by the links which are not connected with the destinations such as link 1 and link 2.

Note that, the proposed formula also holds *when there are queue spillbacks*. What we should do is only to replace the capacities of the links affected by the spillbacks by the reduced capacities (or actual link flows). This is because the traffic flow pattern of the point queue model with variable or time-dependent link capacities is same with that in a physical queue situation (or model) when the variable capacities coincide with the reduced ones. Furthermore, we can capture the dynamics of the congestion through the change of the congestion pattern $(\mathbf{A}, \mathbf{M})$, although each congestion pattern is assumed to be under the steady-state. Specifically, the dynamics is described as the change in the queuing state (queued or not) of the links and the resulting reduced network (see Figure 8, for example). The network throughput varies because of changes in the topology and the capacity pattern of the reduced network $(\mathbf{A}, \mathbf{M})$.

Before concluding this subsection, we mention the relation to the existing analytical method based on the variational theory (VT) (e.g., Daganzo and Geroliminis, 2008; Leclercq and Geroliminis, 2013; Laval and Castrillón, 2015). The proposed analytical method is consistent with it in the sense that both the methods analyze the network flow for a fixed number of vehicles in circulation. They, however, have different merits (and limitations). Thanks to a simple corridor network setting and homogeneous density assumption within it (i.e., the possible network states are limited), the VT-based method can evaluate the entire MFD (i.e., the flow for the entire range of the density) based on only a few parameters that represent the *static* average network characteristics (e.g., the link fundamental diagram and signal setting). This is useful for identifying an *ideal* MFD under the evenly distributed congestion and/or examining the effect of changes in average network characteristics on it. On the other hand, thanks to the explicit consideration of congestion patterns, our method can incorporate the spatially uneven congestion distribution, which is key to the characterization of a real world MFD (or NEF), into the determination of the network throughput. For example, we may analyze a hysteresis loop by applying the proposed method to different congestion patterns for the same vehicle accumulation (see Fig.3 in Geroliminis and Sun, 2011a, for example). However, since the possible network states are enormous, it is difficult to estimate the entire MFD only from the static network characteristics. Nevertheless, once the congestion pattern is assumed to be given, the proposed method can analyze the effect of local traffic changes on the global network throughput (see Sections 4.1–4.3) and this leads to developing a local control for improving the overall network performance (see Section 4.4).

3.3. Network throughput under dynamic condition

In the analysis above, we assumed that the congestion pattern is under the steady-state where a fixed number of vehicles circulate in the network. In this subsection, we relax this assumption, i.e., we consider the congestion pattern under a dynamic condition where vehicle accumulation and the OD demands could change over time. We then approximately derive an analytical formula of the network throughput under the dynamic condition. The comparison between the formulas under the steady-state and dynamic conditions demonstrates that the steady-state analysis would be valid for dynamic conditions even though some differences arise.

Suppose that the congestion pattern $(\mathbf{A}, \mathbf{M})$ is observed during a time interval $[t, t + \Delta t]$. Then, we define the time average of $\dot{Q}_{od}(s)$ and $\dot{\tau}_d(s)$ during the origin departure time interval $[s, s + \Delta s]$, which satisfies $\tau_d(s) = t$ and $\tau_d(s + \Delta s) = t + \Delta t$ for each destination d, as follows:

$$\overline{\dot{Q}}_{od} \equiv \frac{1}{\Delta s}\int_{s}^{s+\Delta s} \dot{Q}_{od}(s)\,\mathrm{d}s, \quad \overline{\dot{\tau}}_d \equiv \frac{1}{\Delta s}\int_{s}^{s+\Delta s} \dot{\tau}_d(s)\,\mathrm{d}s.$$

where the latter variable represents that the growth rate of the (equilibrium) OD travel time[10]:

$$\overline{\dot{\tau}}_d = 1 + \frac{C_d^*(\tau_d(s+\Delta s)) - C_d^*(\tau_d(s))}{(s+\Delta s) - s} = \frac{\Delta t}{\Delta s}. \tag{3.16}$$

Their column vectors are denoted by $\overline{\dot{\mathbf{Q}}}_d$ and $\overline{\dot{\boldsymbol{\tau}}}_d$, respectively. For these variables, we make the following assumption.

Assumption 1. *The DUE condition* (2.15) *holds approximately:*

$$\mathbf{V}\begin{bmatrix} \overline{\dot{\boldsymbol{\tau}}}_i \\ \hline \overline{\dot{\boldsymbol{\tau}}}_d \end{bmatrix} \approx \begin{bmatrix} \boldsymbol{\delta}_i \\ \hline \overline{\dot{\mathbf{Q}}}_d + \boldsymbol{\delta}_d \end{bmatrix}. \tag{3.17}$$

where $\overline{\dot{\boldsymbol{\tau}}}_i$ with element $\overline{\dot{\tau}}_i$ corresponding to the time average of $\dot{\tau}_i(s)$.

This assumption states that the DUE relationship between $\dot{\mathbf{Q}}_d(s)$ and $\dot{\boldsymbol{\tau}}_d(s)$ holds approximately in terms of their average values. The reason why Eq.(3.17) is an approximation, in general, is that the origin departure duration $[s, s + \Delta s]$ corresponding to the destination arrival duration $[t, t + \Delta t]$ is different for the different destinations (i.e., the OD travel times are different). However, the condition (3.17) becomes almost exact, if the changes in $\dot{\mathbf{Q}}_d(s)$ and $\dot{\boldsymbol{\tau}}_d(s)$ during $[t, t + \Delta t]$ are small because the DUE condition (2.15) is linear[11].

Let $\overline{f}_d \equiv \frac{1}{\Delta t}\int_t^{t+\Delta t} f_d(\tau_d(s))\mathrm{d}t$ be the average network throughput for destination d and $\overline{\mathbf{f}}_d$ be its column vector. Then, we have the following proposition:

Proposition 2. *Suppose that a congestion pattern $(\mathbf{A}, \mathbf{M})$ and $\overline{\dot{\boldsymbol{\tau}}}_d$ are given. Then, the average network throughput of a one-to-many network is approximately given by the following formula:*

$$\overline{F} \equiv \mathbf{1}^T\overline{\mathbf{f}}_d = \mathbf{1}^T\mathbf{T}^{-1}\mathbf{V}_{dd}\overline{\dot{\boldsymbol{\tau}}}_d - \mathbf{1}^T\mathbf{T}^{-1}[\mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}(\mathbf{V}_{id}\overline{\dot{\boldsymbol{\tau}}}_d - \boldsymbol{\delta}_i) + \boldsymbol{\delta}_d] \tag{3.18}$$

where $\mathbf{T}$ is the diagonal matrix, whose diagonal elements are the vector $\overline{\dot{\boldsymbol{\tau}}}_d$.

[10] Note that $C_d^*(\tau_d(s)) = \tau_d(s) - s$ and $C_d^*(\tau_d(s+\Delta s)) = \tau_d(s+\Delta s) - (s+\Delta s)$.

[11] The condition (3.17) exactly holds if the equilibrium travel times for different destinations are the same.

Proof. From Eq.(3.17), we have the following relation:

$$\overline{\dot{\mathbf{Q}}}_d \approx \mathbf{V}_{dd}\overline{\boldsymbol{\tau}}_d - [\mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}(\mathbf{V}_{id}\overline{\boldsymbol{\tau}}_d - \boldsymbol{\delta}_i) + \boldsymbol{\delta}_d]. \tag{3.19}$$

Meanwhile, the average network throughput (i.e., average outflow) for each destination can be expressed by

$$\overline{f}_d = \frac{D_d(t+\Delta t) - D_d(t)}{\Delta t} = \frac{Q_{od}(s+\Delta s) - Q_{od}(s)}{\Delta t} = \frac{\frac{1}{\Delta s}\int_s^{s+\Delta s} \dot{Q}_{od}(s)\,\mathrm{d}s}{\frac{1}{\Delta s}\int_s^{s+\Delta s} \dot{\tau}_d(s)\,\mathrm{d}s} = \frac{\overline{\dot{Q}}_{od}}{\overline{\dot{\tau}}_d}$$

$$\Rightarrow \quad \overline{\mathbf{f}}_d = \mathbf{T}^{-1}\overline{\dot{\mathbf{Q}}}_d. \tag{3.20}$$

By substituting Eq.(3.19) into Eq.(3.20) and summing the elements of vector $\overline{\mathbf{f}}_d$, we obtain Eq.(3.18). □

The element form of Eq.(3.18) is given by

$$\overline{F} = \textstyle\sum_{d\in\mathcal{N}_d}\left[\sum_{k\in I(d)}\mu_{kd} - \sum_{k\in O(d)}\mu_{dk}(\overline{\dot{\tau}}_k/\overline{\dot{\tau}}_d)\right] \quad \text{where} \quad \overline{\boldsymbol{\tau}}_i = -(\mathbf{V}_{ii})^{-1}[\mathbf{V}_{id}\overline{\boldsymbol{\tau}}_d - \boldsymbol{\delta}_i]. \tag{3.21}$$

From this equation, we see that the formula under the dynamic condition also expresses the flow conservation at the destinations, i.e., the essential structure that the congestion pattern characterizes the network throughput is the same for both formulas. Indeed, the formula (3.18) equals to the steady-state formula (3.10) if $\overline{\boldsymbol{\tau}}_d = \mathbf{1}$.

Note that the formula (3.18) requires $\overline{\boldsymbol{\tau}}_d$ as the input instead of the periodic boundary condition for deriving the steady-state formula (3.10). To examine their differences in detail, we subtract Eq.(3.21) from Eq.(3.15):

$$F - \overline{F} = \textstyle\sum_{d\in\mathcal{N}_d}\sum_{k\in O(d)}\mu_{dk}(\overline{\dot{\tau}}_k/\overline{\dot{\tau}}_d - \dot{\tau}_k/\dot{\tau}_d) \tag{3.22}$$

where $\dot{\tau}_d = 1$. From this equation, we see that the difference of the network throughputs under the steady-state and the dynamic conditions is caused by the difference of outflows from destination nodes. Furthermore, although we cannot identify whether Eq.(3.22) is positive or negative exactly, an estimate is identified as follows. From Eq.(2.3) and Eq.(2.13), we see that $\overline{\dot{\tau}}_k/\overline{\dot{\tau}}_d$ (or $\dot{\tau}_k/\dot{\tau}_d$) represents the proportion of the (average) inflow rate to the capacity of the link (d,k). Hence, the value of $\overline{\dot{\tau}}_k/\overline{\dot{\tau}}_d$ is greater than one, when the queue on the link is increasing; this may be the case if vehicle accumulation in the network (in particular, at downstream of the destinations) is increasing. Conversely, $\dot{\tau}_k/\dot{\tau}_d$ should be one, on an average, because the vehicle accumulation is fixed. Therefore, we can say that the network throughput under the dynamic condition tends to be lower [larger] compared with the steady-state network throughput when the vehicle accumulation is increasing [decreasing].

4. Sensitivity analysis and its applications

In this section, we investigate the mechanisms underlying the decrease in the network throughput through sensitivity analysis of the formula derived in the previous section. Specifically, we derive the sensitivity coefficient of the network throughput with respect to the change in a link capacity on a reduced network in Section 4.1. Then, in Section 4.2, we identify the necessary

and sufficient conditions for the decrease in network throughput by the decrease in a certain link capacity (due to queue spillbacks). Here, we also clarify the respective mechanisms. With the same method, in Section 4.3, we identify the conditions for the occurrence of a capacity increasing paradox that improving the capacity of a certain link on a network leads to a decrease in the network throughput. Finally, in Section 4.4, we explore a network signal control for improving the network performance based on the mechanisms of macroscopic flow behaviors clarified in Sections 4.2 and 4.3.

Note that we use the steady-state formula (3.10) throughout this section because it is simple and exact, and the essential structure is same with the dynamic one as we have discussed.

4.1. Sensitivity coefficient

In the derivation of the sensitivity coefficient of the formula (3.10) below, $\mathbf{M}(\boldsymbol{\mu})$ denotes the diagonal matrix, whose diagonal elements are link capacity vectors $\boldsymbol{\mu} \equiv [\dots, \mu_{kl}, \dots]$, and $\mathbf{V}(\boldsymbol{\mu})$ denotes the matrix $\mathbf{AM}(\boldsymbol{\mu})\mathbf{A}_-^T$; we also parameterized other variables by $\boldsymbol{\mu}$, i.e., $F(\boldsymbol{\mu})$, $\mathbf{f}_d(\boldsymbol{\mu})$, $\dot{\boldsymbol{\tau}}(\boldsymbol{\mu})$ and $\boldsymbol{\delta}(\boldsymbol{\mu})$.

Let us compare the two network throughputs $F(\boldsymbol{\mu})$ and $F(\boldsymbol{\mu}')$ for two capacity patterns $\boldsymbol{\mu}$ and $\boldsymbol{\mu}' \equiv \boldsymbol{\mu} + \Delta\boldsymbol{\mu}$. By considering Eq.(3.11) for two capacity patterns and taking the difference between them, we have

$$\dot{\boldsymbol{\tau}}_i(\boldsymbol{\mu}') - \dot{\boldsymbol{\tau}}_i(\boldsymbol{\mu}) = -(\mathbf{V}_{ii}(\boldsymbol{\mu}))^{-1}[\mathbf{V}_{ii}(\Delta\boldsymbol{\mu})\dot{\boldsymbol{\tau}}_i(\boldsymbol{\mu}') + \mathbf{V}_{id}(\Delta\boldsymbol{\mu})\mathbf{1} - \boldsymbol{\delta}_i(\Delta\boldsymbol{\mu})]. \tag{4.1}$$

where we used the fact that $\mathbf{V}(\boldsymbol{\mu})$ and $\boldsymbol{\delta}(\boldsymbol{\mu})$ are linear in $\boldsymbol{\mu}$: $\mathbf{V}(\boldsymbol{\mu}') = \mathbf{V}(\boldsymbol{\mu}) + \mathbf{V}(\Delta\boldsymbol{\mu})$, $\boldsymbol{\delta}(\boldsymbol{\mu}') = \boldsymbol{\delta}(\boldsymbol{\mu}) + \boldsymbol{\delta}(\Delta\boldsymbol{\mu})$. In the same way, from Eq.(3.12), we have

$$\mathbf{f}_d(\boldsymbol{\mu}') - \mathbf{f}_d(\boldsymbol{\mu}) = \mathbf{V}_{dd}(\Delta\boldsymbol{\mu})\mathbf{1} - \boldsymbol{\delta}_d(\Delta\boldsymbol{\mu}) + \mathbf{V}_{di}(\Delta\boldsymbol{\mu})\dot{\boldsymbol{\tau}}_i(\boldsymbol{\mu}') + \mathbf{V}_{di}(\boldsymbol{\mu})[\dot{\boldsymbol{\tau}}_i(\boldsymbol{\mu}') - \dot{\boldsymbol{\tau}}_i(\boldsymbol{\mu})] \tag{4.2}$$

By substituting Eq.(4.1) into Eq.(4.2), we obtain (we omit $\boldsymbol{\mu}$ here)

$$\begin{aligned}\mathbf{f}_d(\boldsymbol{\mu}') - \mathbf{f}_d =& \mathbf{V}_{dd}(\Delta\boldsymbol{\mu})\mathbf{1} - \boldsymbol{\delta}_d(\Delta\boldsymbol{\mu}) + \mathbf{V}_{di}(\Delta\boldsymbol{\mu})\dot{\boldsymbol{\tau}}_i(\boldsymbol{\mu}') \\ &- \mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}[\mathbf{V}_{ii}(\Delta\boldsymbol{\mu})\dot{\boldsymbol{\tau}}_i(\boldsymbol{\mu}') + \mathbf{V}_{id}(\Delta\boldsymbol{\mu})\mathbf{1} - \boldsymbol{\delta}_i(\Delta\boldsymbol{\mu})].\end{aligned} \tag{4.3}$$

Finally, consider the case where $\Delta\boldsymbol{\mu} = [0, \dots, 0, \Delta\mu_{kl}, 0, \dots, 0]$. Multiplying Eq.(4.3) with $\mathbf{1}^T$ from both sides, dividing both sides of Eq.(4.3) by $\Delta\mu_{kl}$, and taking the limit as $\Delta\mu_{kl} \to 0$, we obtain the sensitivity coefficient:

$$\begin{aligned}\frac{\partial F}{\partial \mu_{kl}} =& \mathbf{1}^T\left\{\frac{\partial \mathbf{V}_{dd}(\Delta\boldsymbol{\mu})}{\partial \mu_{kl}}\mathbf{1} - \frac{\partial \boldsymbol{\delta}_d(\Delta\boldsymbol{\mu})}{\partial \mu_{kl}} + \frac{\partial \mathbf{V}_{di}(\Delta\boldsymbol{\mu})}{\partial \mu_{kl}}\dot{\boldsymbol{\tau}}_i\right. \\ &\left. - \mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}\left[\frac{\partial \mathbf{V}_{ii}(\Delta\boldsymbol{\mu})}{\partial \mu_{kl}}\dot{\boldsymbol{\tau}}_i + \frac{\partial \mathbf{V}_{id}(\Delta\boldsymbol{\mu})}{\partial \mu_{kl}}\mathbf{1} - \frac{\partial \boldsymbol{\delta}_i(\Delta\boldsymbol{\mu})}{\partial \mu_{kl}}\right]\right\}\end{aligned} \tag{4.4}$$

$$\text{where} \quad \frac{\partial \mathbf{V}_{ab}(\Delta\boldsymbol{\mu})}{\partial \mu_{kl}} \equiv \lim_{\Delta\mu_{kl}\to 0}\frac{\mathbf{A}_a\mathbf{M}(\Delta\boldsymbol{\mu})\mathbf{A}_{b-}^T}{\Delta\mu_{kl}} = \mathbf{A}_a\mathbf{I}_{kl}\mathbf{A}_{b-}^T, \quad a, b \in \{i, d\}$$

$$\frac{\partial \boldsymbol{\delta}_a(\Delta\boldsymbol{\mu})}{\partial \mu_{kl}} \equiv \lim_{\Delta\mu_{kl}\to 0}\frac{\boldsymbol{\delta}_a(\Delta\boldsymbol{\mu})}{\Delta\mu_{kl}} = \begin{cases} \mathbf{e}_k & \text{if} \quad l = o \\ \mathbf{0} & \text{otherwise} \end{cases}, \quad a \in \{i, d\},$$

$\mathbf{I}_{kl}$ is the L × L matrix, whose diagonal element corresponding to link (k, l) is one and all other elements are zero; $\mathbf{e}_k$ is the ($|\mathcal{N}_i|$ or $|\mathcal{N}_d|$ dimensional) unit vector, whose the element corresponding to node k is one.

The first term of Eq.(4.4) is the sensitivity coefficient of the first term of the formula (3.10), and the other terms are the sensitivity coefficients of the second term of the formula. Note that we can obtain a simple expression of the sensitivity coefficient (4.4) by recalling the structure of the matrix $\mathbf{V} = \mathbf{AMA}_{-}^{T}$, i.e., each link capacity μ_{kl} is included in at most two elements of the matrix $\mathbf{V}$ as an incoming and/or outgoing link (see also Eq.(3.14)). Thus, Eq.(4.4) has at most two nonzero terms depending on the types of initial and terminal nodes of link (k, l):

$$\frac{\partial F}{\partial \mu_{kl}} = \begin{cases} \mathbf{1}^T(\mathbf{A}_d\mathbf{I}_{kl}\mathbf{A}_{d-}^T)\mathbf{1} & \text{if } k,l \in \mathcal{N}_d \text{ or } k = o \wedge l \in \mathcal{N}_d \\ -\mathbf{1}^T\mathbf{e}_k & \text{if } k \in \mathcal{N}_d \wedge l = o \\ \mathbf{1}^T\mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}\mathbf{e}_k & \text{if } k \in \mathcal{N}_i \wedge l = o \\ -\mathbf{1}^T\mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}(\mathbf{A}_i\mathbf{I}_{kl}\mathbf{A}_{i-}^T)\dot{\boldsymbol{\tau}}_i & \text{if } k = o \wedge l \in \mathcal{N}_i \text{ or } k,l \in \mathcal{N}_i \\ \mathbf{1}^T(\mathbf{A}_d\mathbf{I}_{kl}\mathbf{A}_{d-}^T)\mathbf{1} - \mathbf{1}^T\mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}(\mathbf{A}_i\mathbf{I}_{kl}\mathbf{A}_{d-}^T)\mathbf{1} & \text{if } k \in \mathcal{N}_i \wedge l \in \mathcal{N}_d \\ \mathbf{1}^T(\mathbf{A}_d\mathbf{I}_{kl}\mathbf{A}_{i-}^T)\dot{\boldsymbol{\tau}}_i - \mathbf{1}^T\mathbf{V}_{di}(\mathbf{V}_{ii})^{-1}(\mathbf{A}_i\mathbf{I}_{kl}\mathbf{A}_{i-}^T)\dot{\boldsymbol{\tau}}_i & \text{if } k \in \mathcal{N}_d \wedge l \in \mathcal{N}_i \end{cases} \tag{4.5}$$

where the logical operation "$\wedge$" means "and." Note that the sensitivity coefficient (4.5) (and (4.4)) can be calculated easily by matrix operations, although it still looks complicated. The inputs required for calculating the sensitivity coefficient are the same with those for the formula (3.10), i.e., the topology and capacity pattern of the reduced network and the locations of origin/destination nodes in the network.

4.2. Network throughput decreasing conditions and their mechanisms

We can now identify the conditions for the decrease in the "global" network throughput due to the decrease in a "local" link capacity (i.e., $\partial F/\partial \mu_{kl} > 0$). Using the sensitivity coefficient (4.5), we make the following proposition:

Proposition 3. *Suppose the capacity of link* (k, l) *in a one-to-many reduced network decreases. Then, the steady-state network throughput decreases, if and only if one of the following conditions is satisfied.*

(i) node k *is the origin and node* l *is a destination.*

(ii) node k *is the origin and node* l *is a transient node; there exists a route on the reduced network from the origin to node* l *so that the route passes through at least one destination.*

(iii) both nodes k *and* l *are transient nodes; the following condition is satisfied.*

$$\sum_{i\in\mathcal{N}_i}\sum_{d\in\{I(i)\cap\mathcal{N}_d\}}\mu_{di}[v_{il}^{-1} - v_{ik}^{-1}]\dot{\tau}_l > 0. \tag{4.6}$$

(iv) node k *is transient node and node* l *is a destination; the following condition is satisfied.*

$$1 - \sum_{i\in\mathcal{N}_i}\sum_{d\in\{I(i)\cap\mathcal{N}_d\}}\mu_{di}v_{ik}^{-1} > 0. \tag{4.7}$$

where v_{kl}^{-1} *represents the* (k, l) *element of the matrix* $(\mathbf{V}_{ii})^{-1}$.

Proof. See Appendix B. □

The four conditions in this proposition are different with respect to which terms of the formula (3.10) change when the underlying link capacity changes. Following, we show the relationship between the types of the underlying link and the changes in the formula through illustrated

examples shown in the left column of Figure 4. Each example shows one of the minimal reduced networks that satisfy Proposition 3, when the link capacity (red arrows) decreases due to a *queue spillback*[12]. The corresponding (original) non-saturated networks are also shown in the right column of Figure 4, illustrating the physical mechanism under each condition.

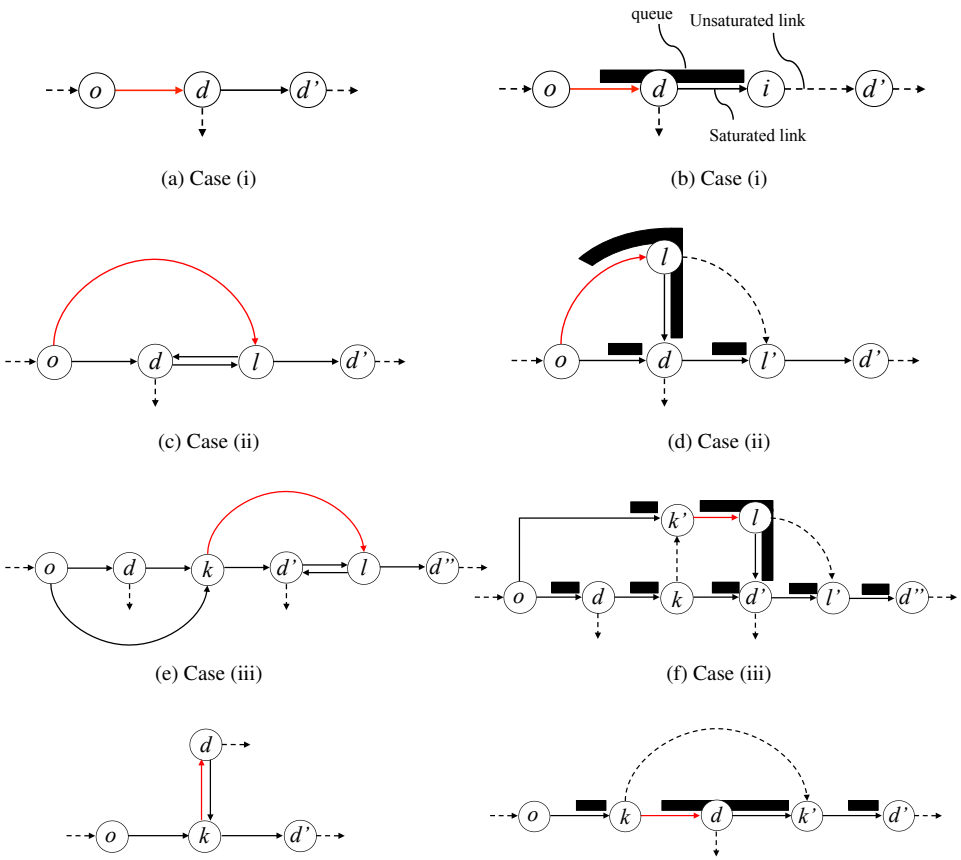
A capacity decrease in the link that satisfies Proposition 3-(i) leads to a decrease in the first term of the formula (3.10), and thus a decrease in inflow to a destination. In Figures 4(a) and (b), link (o, d) is the corresponding link. Suppose the queue on link (d, i) spillovers to link (o, d) and its capacity decreases. Then, the queue, which is caused by users with more downstream destination d', is *blocking* the flow exiting to destination d; then, the network throughput decreases. This phenomenon is almost obvious and may likely occur on congested urban networks, where destinations are close to each other.

Proposition 3-(ii) describes a more complex phenomenon. The capacity decrease in the underlying link leads to the increase in the second term of the formula (3.10), which means the increase in flows passing through a destination. In this case, two types of routes exist on the reduced network (see Figure 4(c)) from the origin to the transient node l: one is a route that passes through at least one destination, and the other is a route that does not pass through any destination; the underlying link (o, l) is on the latter route. The reduced network may look strange, but the corresponding non-saturated network (see Figure 4(d)) describes an intuitive and common situation. Suppose the queue on link (l, d) spillovers to link (o, l) and its capacity decreases. Then, the queue spillback alters the route choice pattern, increasing the sum of flows on the routes that pass through destination d, because the flow on the route that does not pass through d decreases. However, the capacity of link (l', d') does not change, which means the throughput from destination d' does not change. In contrast, users with different destinations d and d' compete with each other for the constant capacities of link (o, d) and (l, d), respectively, which implies that the change in route choice pattern mentioned above decreases the capacity share for users with destination d. Thus, the decrease in the flow exiting to destination d leads to the decrease in the network throughput. This can be interpreted as the phenomenon, where a road that bypasses the congested area does not work well due to congestion.

Proposition 3-(iii) is similar to Proposition 3-(ii) in the sense that the capacity decrease in the underlying link leads to the increase in the second term of the formula (3.10), and it describes almost the same phenomenon. Suppose the queue on (l, d') in Figure 4(f) spillovers to link (k', l), and its capacity decreases. Then, as in the second case, the flow exiting to destination d' decreases. In contrast, the flow exiting to destination d increase because the flows on the routes that pass through destination d decrease. The RHS of Eq.(4.6) represents the total change in the throughputs for destinations d and d'; it is negative if Eq.(4.6) is satisfied, i.e., the network throughput decreases.

Proposition 3-(iv) describes the combination of the above two phenomena. Specifically, a capacity decrease in the underlying link leads to a decreases in both the first and second terms of the formula (3.10). Suppose the queue on (d, k') in Figure 4(h) spillovers to link (k, d) and its capacity decreases. The queue spillback alters the route choice pattern and *indirectly* decreases the flow on the route that passes through destination d as well as is *directly* blocking the flow that exits to destination d. The condition (4.7) states that the latter effect is dominant, i.e., the throughput from destination d and the network throughput decrease. This reflects a situa-

[12]Note that the proposition states a general condition for the decrease in the network throughput due to the decrease in a link capacity; thus various factors may cause the decrease in a link capacity (e.g., capacity drops, traffic accidents).



(a) Case (i)

(b) Case (i)

(c) Case (ii)

(d) Case (ii)

(e) Case (iii)

(f) Case (iii)

(g) Case (iv)

(h) Case (iv)

Figure 4: Minimal reduced networks (left column) and their corresponding non-saturated networks (right column) for Proposition 3

tion where a bypass road works relatively well but is not fully utilized because it may not be a convenient route.

In summary, Proposition 3 shows that the *interactions between users with different destinations* are the main causes for a decrease in the network throughput due to *queue spillbacks*. This insight on the mechanisms of the macroscopic traffic flow behaviors is not completely new. For example, our result is consistent with the discussion presented by Daganzo (2007), who suggested that the *blocking* and *increasing of through traffic* of congested destinations should be avoided using a certain strategy (i.e., *AB strategy for cities*) to maximize the network throughput. However, the present analysis describes the "local-global relationships" in a more concrete form (for general networks with one-to-many demands). This could be useful for developing a local control policy based on (stable) congestion patterns to avoid the decrease in the network throughput or improve it. The subsequent two subsections explore this possibility.

4.3. Capacity increasing paradox

This subsection identifies a *capacity increasing paradox* such as Braess (1968) and Smith (1978), which is the basis of a traffic control shown in the next subsection. We here refer to the situation as a "paradox" if an increase [decrease] in the capacity of a certain link leads to a decrease [increase] in the network throughput (i.e., $\partial F/\partial \mu_{kl} < 0$).

As in the previous subsection, we have the following proposition:

Proposition 4. *Suppose the capacity of link* (k, l) *on a one-to-many reduced network increases. Then, the capacity increasing paradox occurs if and only if one of the following conditions is satisfied.*

(i) node k *is a destination and node* l *is the origin.*

(ii) node k *is a transient node and node* l *is the origin; there exists a route on the reduced network from the origin to node* k *so that the route passes through at least one destination.*

(iii) both nodes k *and* l *are transient nodes; the following condition is satisfied.*

$$\sum_{i\in\mathcal{N}_i}\sum_{d\in\{I(i)\cap\mathcal{N}_d\}}\mu_{di}[v_{il}^{-1}-v_{ik}^{-1}]\dot{\tau}_l<0. \tag{4.8}$$

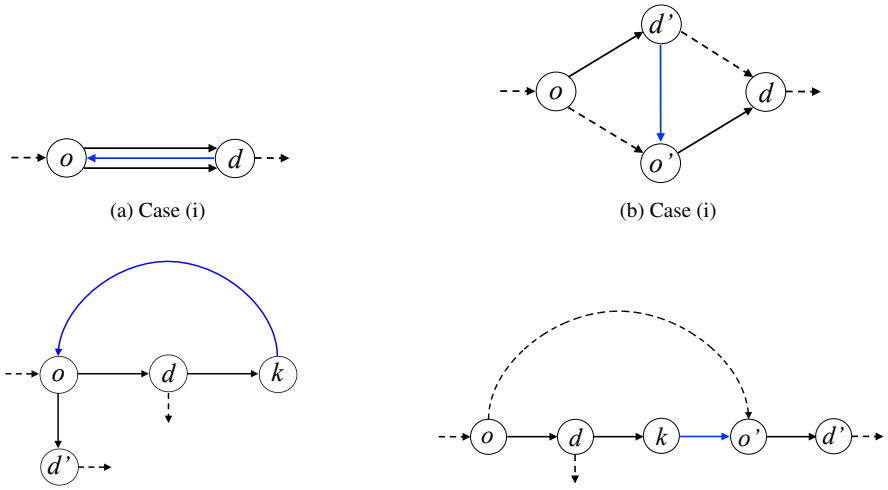
(iv) node k *is a destination and node* l *is a transient node; the following condition is satisfied.*

$$\sum_{i\in\mathcal{N}_i}\sum_{d\in\{I(i)\cap\mathcal{N}_d\}}\mu_{di}v_{il}^{-1}-1<0 \tag{4.9}$$

Proof. See Appendix B. □

Minimal reduced networks and their corresponding non-saturated networks for some cases are shown in Figure 5. Although four different conditions are identified in Proposition 4, their meanings are almost the same in the sense that only the second term of the formula (3.10) changes. The increase in the capacity of the underlying link (blue arrow in each figure) increases the through traffic of a congested destination on the reduced network; then, the network throughput decreases (the paradox occurs). However, two types of interactions can be found: the interaction between users with different routes for the same destination and the interaction between users with different destinations as we mentioned in the previous subsection.

Proposition 4-(i) describes the former interaction (Figure 5(a)). In this network, the capacity increase in the underlying link increases the "through traffic" of the unique destination (a self-destruction occurs), which interestingly shows the situation known as *Braess' paradox*: the



(a) Case (i) (b) Case (i)

(c) Case (ii) (d) Case (ii)

Figure 5: Minimal reduced networks (left column) and their corresponding non-saturated networks (right column) for Proposition 4

improvement of link (d', o') in Figure 5(b) cannot increase the throughput from link (o', d) but reduces that from link (d', d). Note that this type of interaction may not be sufficiently important to mitigate the decrease in the network throughput due to queue spillbacks because the latter interaction is essential.

The other three cases describe the interaction between users with different destinations. For Proposition 4-(ii), the improvement of link (k, o) in Figure 5(c) (or link (k, o') in Figure 5(d)) decreases the flow exiting to destination d. Proposition 4-(iii) is the opposite situation of Proposition 3-(iii), which was examined by Akamatsu (2000); Akamatsu and Heydecker (2003a,b) extensively[13]. Proposition 4-(iv) can be illustrated by Figure 4(f): the improvement of link (d', l') decreases the flows exiting to destination d' due to the increase in the passing flow through destination d' as in Proposition 3-(iii); simultaneously, the flow exiting to destination d increases because the flows on route $\{(o, d), (d, k), (k, k'), (k', l), (l, l')\}$ decreases with the decrease in the flow on link (l, l'); the network throughput decreases because the former effect is dominant (i.e., Eq.(4.9) is satisfied).

From the paradox, we can easily see that degrading (or restricting the outflow of) a certain link can improve the network performance. Note however that the links that should be controlled changes dynamically depending on the evolution of congestion pattern. In this sense, the conditions in Proposition 4 suggest dynamic traffic control logics. However, it may be usually difficult to control a single link independently, particularly in urban networks. For example, the decrease in the capacity of a link increases the capacity of other links for a merge situation. Thus, as a more realistic control policy, the next subsection develops a signal control policy based on the mechanisms of macroscopic flow behaviors clarified so far.

[13]They obtain the similar condition, although they do not consider the steady-state and the throughput (to measure the efficiency of networks).

Table 1: Sensitivities when the green split of link (i,k) increases and the merge node is a transient node ($k \in \mathcal{N}_i$)

No.	upstream nodes	dF	causes of sensitivity	additional condition
1	$i = o \ \wedge j \in \mathcal{N}_i$	≥ 0	second term	–
2	$i = o \ \wedge j \in \mathcal{N}_d$	> 0	second term	–
3	$i, j \in \mathcal{N}_i$	≥ 0	second term	There is no route from the origin to node i so that it passes through a destination.
4	$i \in \mathcal{N}_i \ \wedge \ j \in \mathcal{N}_d$	≥ 0	second term	–
5	$i, j \in \mathcal{N}_d$	0	second term	–

4.4. Network signal control

To simplify the exposition, here, we consider a simpler situation that each node in the original network has at most two incoming links and no loop exists in reduced networks. For a signal control policy, we also assume that the green splits for incoming links at a merge node (in the original network) are controlled by a signal control policy and the policy is activated when both incoming links are saturated. We then examine the following sensitivity:

$$\mathrm{d}F = \frac{\partial F}{\partial \mu_{ik}}\mathrm{d}\mu_{ik} + \frac{\partial F}{\partial \mu_{jk}}\mathrm{d}\mu_{jk} = \left(s_{ik}\frac{\partial F}{\partial \mu_{ik}} - s_{jk}\frac{\partial F}{\partial \mu_{jk}}\right)\mathrm{d}g_{ik}, \tag{4.10}$$

where s_{ik} and g_{ik} are the saturation flow rate and the green split for link (i,k), respectively, and we use the facts: $\mu_{ik} = s_{ik}g_{ik}$ and $\mathrm{d}g_{ik} = -\mathrm{d}g_{jk}$. The sensitivities for all combinations of link types are summarized in Table 1 ($k \in \mathcal{N}_i$) and Table 2 ($k \in \mathcal{N}_d$). We omit the proof of these tables because it is a straightforward calculation by the same way as in Appendix B. In Tables 1 and 2, "$\mathrm{d}F$" shows the sensitivity when the green split of link (i,k) increases. The indicated "causes of sensitivity" refer to changing terms of the formula (3.10). The "additional condition" describes a sufficient condition to obtain the sign of $\mathrm{d}F$.

The result can be understood intuitively by combining Proposition 3 and Proposition 4. Because the only second term of the formula (3.10) changes when the merge node is a transient node (Table 1), the basic strategy is to decrease the through traffic of congested destinations. That is,

> *Strategy for transient nodes*: increase the green split of the link whose initial node on the *reduced network* is the origin node (i.e., Proposition 3-(ii)) or decrease the green split of the link whose initial node on the reduced network is a destination node (i.e., Proposition 4-(iv)).

Note that, sensitivity becomes zero when there are no routes from the origin to nodes i, j passing through destinations (i.e., the second term of the formula (3.10) is zero) or there are no routes from the origin to both nodes i, j that do not pass through any destination (i.e., the changes in through traffic of congested destinations are cancelled out each other).

For the case that the merging node is a destination (Table 2), the strategy is simplified because there is no link that leads to the paradox (only the cases for Proposition 3-(i) and (iv) occur). That is,

> *Strategy for destination nodes*: increase the green split of the link whose initial node on the reduced network is not a destination node[14].

[14] If both the initial and terminal nodes of a link are destinations, the sensitivity coefficient is zero (see Appendix B).

Table 2: Sensitivities when the green split of link (i,k) increases and the merge node is a destination node ($k \in \mathcal{N}_d$)

No.	upstream nodes	dF	causes of sensitivity	additional condition
1	$i = o \ \wedge j \in \mathcal{N}_i$	> 0	first & second terms	$s_{ik} - s_{jk} > 0.$
2	$i = o \ \wedge j \in \mathcal{N}_d$	> 0	first term	–
3	$i, j \in \mathcal{N}_i$	> 0	first & second terms	$s_{ik} - s_{jk} > 0$
4	$i \in \mathcal{N}_i \ \wedge \ j \in \mathcal{N}_d$	≥ 0	first & second terms	–
5	$i, j \in \mathcal{N}_d$	0	first term	–

However, in the case that the initial nodes of both incoming links are not destination nodes, it is sufficient to increase the green split of the link that has the larger saturation flow rate.

The interesting feature of the proposed signal control policy is that it determines the adjustment directions of green splits from only information on the types of upstream nodes of the merge node on the reduced network (and saturation flow rates for some cases). Thus, it can be regarded as a *local* and *distributed* signal control policy. Meanwhile, we have to determine a *step size* (how much green split we should change for the direction of the proposed policy). To do this without global information and also to smooth the evolution of the congestion pattern, we here set a sufficiently small step size (i.e., change the green split gradually) as in Smith et al. (2015). This is consistent with the sensitivity analysis, where the congestion pattern is assumed to be fixed, i.e., the control direction is valid for the congestion pattern in the near future.

A few remarks are in order. First, the strategy for the same merge node on the original network is not static (or pre-determined) because its node type and/or its upstream node types on reduced networks change depending on the congestion patterns. Second, in principle, the proposed policy cannot eliminate the decrease in the network throughput caused by *blocking* exiting flows to destinations due to queue spillbacks. Furthermore, there is a possibility that the queue spillbacks become worse with the proposed policy as discussed by Daganzo (1998). Therefore, to avoid such problems due to queue spillbacks, it is better to turn off the strategy when the queue length of each link exceeds a pre-determined threshold. Alternatively, it would be also useful to combine it with an MFD-based perimeter control (e.g., Daganzo, 2007; Yoshii et al., 2010; Keyvan-Ekbatani et al., 2012; Haddad and Shraiber, 2014) that maintains the vehicle accumulation at an optimal level and may prevent the occurrence of queue spillbacks.

Conversely, from the viewpoint of MFD-based perimeter control, such a hybrid control can be viewed as a hierarchical control that enhances a pure macroscopic (or upper-level) control by incorporating the local (or lower-level) control to improve the (global) network throughput. More specifically, under spatially uneven congestion patterns, (single region) perimeter controls can be inefficient[15]; however, we may expect that the hybrid control works better, because the proposed control intends to increase the throughputs of congested destinations and the resulting congestion distribution may become more even[16].

[15]For this issue, the perimeter controls for heterogeneous multi-region networks have been explored (see Haddad and Geroliminis, 2012; Aboudolas and Geroliminis, 2013; Keyvan-Ekbatani et al., 2015; Kouvelas et al., 2017, for examples and references therein).

[16]Appropriate route guidance strategies might lead to the similar result (e.g., Knoop et al., 2012; Mahmassani et al., 2013; Yildirimoglu et al., 2015).

5. Numerical examples

In this section, we test the proposed method for a more general condition in which the congestion pattern vary over time through numerical experiments. In section 5.1, we present the comparison of the simulated and analytical network throughputs and the validation of the decreasing mechanisms of the network throughput clarified in Section 4.2. In sections 5.2 and 5.3, we investigate the effectiveness of the signal control policy developed in Section 4.4.

5.1. Example 1: Behaviors of network throughput

We considered a one-to-many network comprising of an arterial road and two bypasses, as shown in Figure 6, to demonstrate clearly the correspondence between the theory and numerical examples. Node o is the origin, and nodes $\{d_1, d_2, d_3, d_4\}$ are destinations. In addition, nodes $\{k, l, n, p, d_1, d_3\}$ are signal-controlled junctions: the green splits of links incoming to these nodes are controlled (the sum of green splits assumed to be 1 at each junction). Each link has a bottleneck section with a bottleneck capacity. The saturation flows of links on bypasses are greater than on arterial roads. Note that bottleneck capacity of links incoming to the junction is the product of a green split and saturation flow rate. We assume that all the green splits are 0.5 for the case with no signal control. The link settings are summarized in Table 3. The total OD flow rate (the departure flow rate at origin) is given by

$$\sum_{d\in\mathcal{N}_d} \mathrm{d}Q_{od}(t)/\mathrm{d}t = \begin{cases} 1 & t < 180, \\ 1 + \frac{1}{720}(t - 180) & 180 \le t < 1980, \\ 3.5 & 1980 \le t < 3780, \\ 3.5 - \frac{1}{720}(t - 3780) & 3780 \le t < 5580, \\ 1 & 5580 \le t < 5760. \end{cases}$$

We set the ratio of users for destinations as: $d_1 : d_2 : d_3 : d_4 = 1 : 1 : 2 : 3$. The average network throughputs and vehicle accumulations under the DUE state are calculated for every time slot (we assume 3 min here) until the vehicle accumulation decreases; the congestion patterns are also identified from the averaged link flows.

We compute the DUE state by the algorithm proposed by Iryo (2011) (slightly revised by Satsukawa and Wada, 2017). This algorithm assigns the users to their shortest routes, in the order of departure from the single origin. With the FIFO property of the link model, this implies that users departing later (following users) cannot overtake those departing earlier (leading users). By combining this relationship with the causality (Carey et al., 2003) of the dynamic traffic flow, it is guaranteed that the route choices of the following users do not affect the travel times of the shortest routes of the leading users; all users choose the ex-post shortest routes. Consequently, an exact equilibrium solution is obtained without heuristics. Within the algorithm, we can employ any appropriate traffic flow model that satisfies FIFO and causality, including physical queue models. Among these, we employ the simplified car following model of Newell (2002).

The results for the no-signal control case are shown in Figures 7 and 8. Figure 7 shows the simulated and analytical NEFs[17]. We obtained analytical NEFs for both the dynamic and steady conditions by applying the analytical formulas (3.10) and (3.18) to the congestion patterns

[17]Simulated NEF shows zero network throughput around 100 [veh] because the state occurs immediately after the caluclation starts, and the vehicles do not reach their destinations by this time.

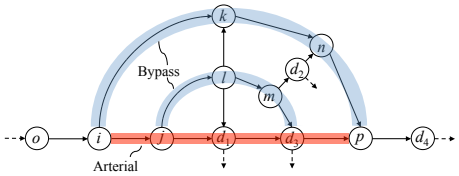


Figure 6: Network comprising of an arterial road and bypasses

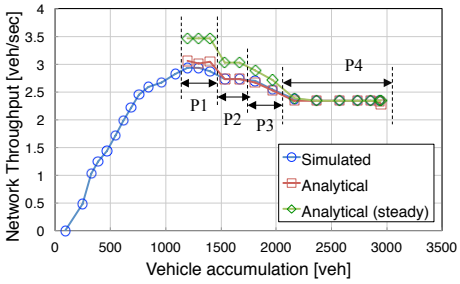


Figure 7: Simulated and analytical NEFs

Table 3: Physical condition of links in the one-to-many network (FFTT: Free Flow Travel Time, BC: Bottleneck Capacity, SF: Saturation Flow)

Link	FFTT [sec]	BN [veh/sec]	SF [veh/sec]	Link	FFTT [sec]	BN [veh/sec]	SF [veh/sec]
(o,i)	143.88	6	6	(i,k)	71.94	3	6
(i,j)	28.78	2.53	4	(k,n)	21.58	1.5	3
(j,d_1)	28.78	0.83	1.67	(n,p)	50.36	0.93	1.87
(d_1,d_3)	28.78	0.5	1	(l,k)	28.78	0.5	1
(d_3,p)	28.78	0.55	1.1	(l,d_1)	21.58	0.5	1
(j,l)	28.78	4	4	(m,d_2)	14.39	0.67	1.0
(l,m)	14.39	4	4	(d_2,n)	14.39	0.33	0.67
(m,d_3)	28.78	0.92	1.83	(p,d_4)	57.56	1	2.5

obtained in the experiment. Note that we did not calculate the analytical NEFs when the origin and destination nodes are unified into a single node in a reduced network. Figure 8 shows the congestion patterns that correspond to the range of P1–P4 in Figure 7. The red arrow represents the link whose capacity decreases because of queue spillback.

From Figure 7, we see that the analytical NEF for the dynamic condition agrees well with the simulated NEF. In contrast, the analytical network throughputs for the steady-state tend to overestimate the simulated ones during P1–P3, although the behaviors of both network throughputs are consistent. This is because the queues on the links between destinations increase with vehicle accumulation. However, during P4, in which the queue spillback reached link (o,i), the queues between destinations do not change significantly; thus, the analytical network throughputs for the steady condition also agree well with the simulated ones. This is consistent with the discussion in Section 3.3.

Next, we validate the decreasing mechanisms identified in Section 4.2. From Figure 8, we can see that the network throughput first decreases due to the change of users route choice pattern during P2 (i.e., increase in the flows on the route, represented by the green block arrow, that passes through destinations). It further decreases by blocking flows exiting to destinations during P3. To investigate these phenomena in detail, let us calculate the sensitivities. During P2, the sensitivity of the network throughput due to the decrease in the capacity of link (o',p) is as

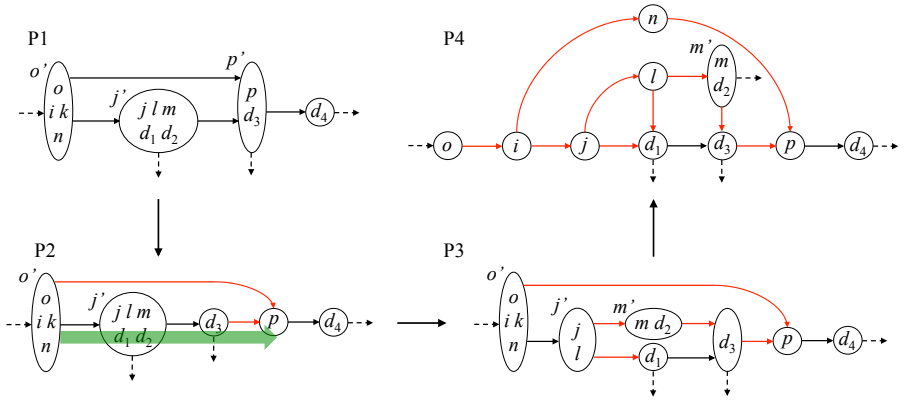


Figure 8: Congestion patterns. Black arrow: saturated link; red arrow: link whose capacity decreases due to queue spillback. Each reduced node is denoted by a subscript on the left upper side of the node when two or more nodes are unified (e.g., node $\{o, i, k, n\}$ during P1 is denoted by o').

follows (Proposition 3-(ii)):

$$(\partial F/\partial \mu_{o'p}) \cdot \mathrm{d}\mu_{o'p} = ((\mu_{d_3p}\mu_{pd_4})/(\mu_{d_3p} + \mu_{o'p})^2) \cdot \mathrm{d}\mu_{o'p} = 0.25 \cdot -0.43 \simeq -0.1.$$

Similarly, the sensitivity respect to the decrease in the capacity of link (d_3, p) is as follows (Proposition 4-(iv)):

$$(\partial F/\partial \mu_{d_3p}) \cdot \mathrm{d}\mu_{d_3p} = (-(\mu_{o'p}\mu_{pd_4})/(\mu_{d_3p} + \mu_{o'p})^2) \cdot \mathrm{d}\mu_{d_3p} = -0.42 \cdot -0.05 \simeq 0.02.$$

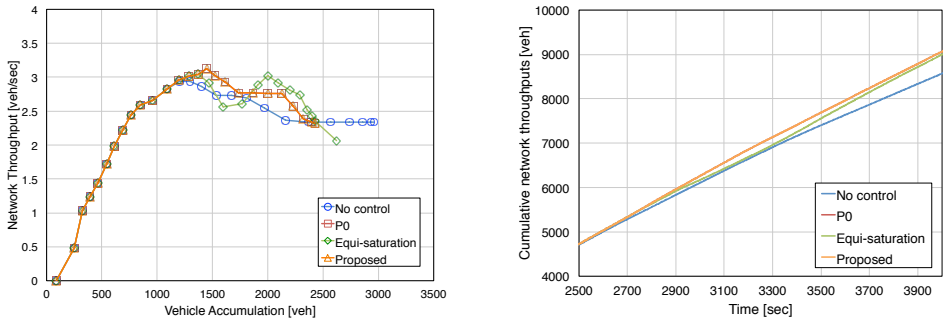
As a result, the network throughput decreases because the sensitivity is negative as a total. During P3, the decrease in the flows to destinations is caused by the decrease in the capacities of two links due to queue spillbacks: links (j', m') and (j', d_1). In this case, the sensitivity coefficients of these links are 1 (i.e., only the direct effect of Proposition 3-(iv) works); thus, the sensitivity respect to the decrease in the capacity of these links is negative. During P4 (i.e., more than 2,000[veh]), the NEF exhibits flat parts because the destinations $\{d_1, d_2, d_3\}$ have already been influenced by the queue spillback. In this situation, the network throughput cannot decrease further while the vehicle accumulation can increase on the link (o, i).

5.2. *Example 2: Performance of network signal controls*

We next investigate the effectiveness of the proposed network signal control. Specifically, we compare the proposed policy with the existing network signal control policies (see Smith, 2015, for a review). In particular, we consider the *policy* P_0 (Smith, 1979a,b, 1980) and a "generalization" of the *equisaturation policy* (Webster, 1958). The policy P_0 determines the green splits to equalize pressures for all the approaches, where a *pressure* is defined as (saturation flow rate) $\times$ (queuing delay) for a point queuing quasi-dynamic traffic assignment model (Smith, 1987, 2015). On the other hand, the natural generalization of the equisaturation policy under oversaturated conditions determines the green splits to equalize the queuing delays for all approaches (Smith, 1987; Mounce, 2009). More specifically, suppose the congestion and the queuing delay patterns are given, the adjustment directions of the green splits for the policies are expressed as

[Policy P_0] *more pressured link has more green split,* (5.1)

[Equisaturation] *more delayed link has more green split.* (5.2)



(a) Network exit functions

(b) Cumulative network throughputs

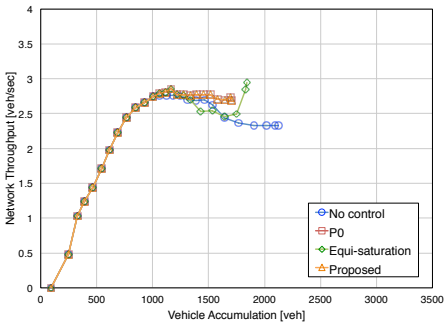
Figure 9: Network exit functions and cumulative network throughputs for high demand case

In the experiment, the policy P_0 [equisaturation policy] gradually adjusts (step size is 0.05) current green splits according to (5.1) [(5.2)] every 3 min when both incoming links are congested. Similarly, the proposed policy adjusts green splits according to the *strategies for transient and destination nodes*. Note that the green split of each link g is constrained by $0.3 \leq g \leq 0.7$. To confirm the robustness of the policies, we consider two demand levels: a demand level, with the same level as that Section 5.1 (hereafter referred to as high demand level) and a medium level with peak of 3.1[veh/sec].

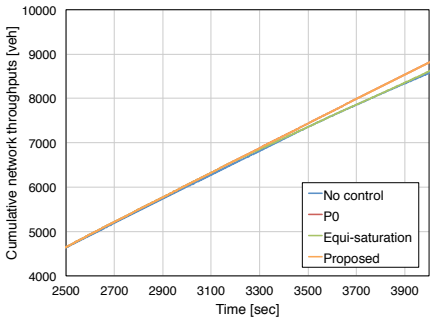
Figure 9(a) shows the NEF for high demand level. The policy P_0 and the proposed policy are observed to achieve higher network throughputs than those under no signal control for most of the accumulation levels. Besides, there are different characteristics among policies in the congested regime. Specifically, the network throughput under policy P_0 and the proposed policy decreases almost monotonously with increase in vehicle accumulation. In contrast, such a monotonicity is not observed under the equisaturation policy. This is because the allocation of green splits and users' route choice patterns under the equisaturation policy oscillate with the increase in vehicle accumulation, even though we gradually adjusted the green splits. This result may imply that the policy P_0 and the proposed policy are consistent with user's route choice; however the equisaturation policy is not.

Figure 9(b) shows cumulative network throughputs under different signal policies. In this figure, the cumulative curves of policy P_0 and the proposed policy are overlapped. It is clear that all the signal control policies, especially the policy P_0 and the proposed policy, improve the network performance. The cumulative network throughput until $4,000$ [sec] (the time when vehicle accumulation decreases under no signal control) under policy P_0 and the proposed policy is about 5.8% higher than that under no signal control. The cumulative network throughput under the equisaturation policy is about 5.0% higher. Moreover, the total time spent by users until 4000 [sec] is improved by about 7.4% under policy P_0 and the proposed policy and by about 4.6% under equisaturation policy.

Figure 10 shows the NEFs and the cumulative network throughputs for the medium demand level. The figure shows similar results for the high demand level. However, the differences in the NEFs among the signal control policies and the degree of improvements by the policies become small: the improvements of the cumulative network throughputs and total time spent are about 2.8% and 3.4% under policy P_0 and the proposed policy, and those under the equisaturation policy are about 0.5% and 1.2%. This is because the network becomes less congested.



(a) Network exit functions



(b) Cumulative network throughputs

Figure 10: Network exit functions and cumulative network throughputs for medium demand case

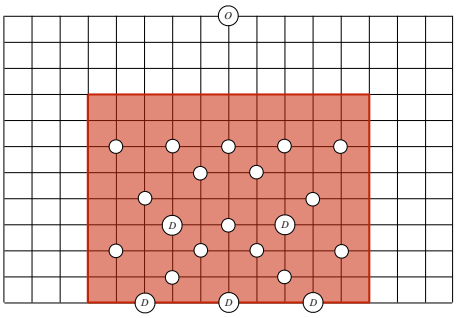


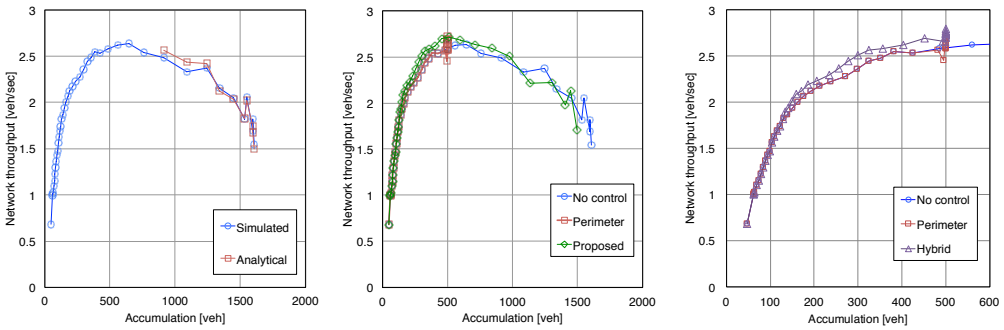
Figure 11: Medium-size network

5.3. Example 3: Combining with a perimeter control

We finally investigate a hybrid control policy of combining the proposed network control with an MFD-based perimeter control. As mentioned in Section 4.4, this is considered as a policy that compliments the proposed control policy so as to prevent the queue spillbacks and from blocking exiting flows to destinations. In the experiment, we considered a medium-size (two-way) network (204 noes, 758 links, 21 destinations) shown in Figure 11 and implemented the simplest "bang-bang" perimeter control by Daganzo (2007) that restricts inflows to the target area (specified by a red-shaded rectangle) so that the vehicle accumulation does not exceed a predetermined threshold.

The detailed experimental setup is as follows. Node o is the origin, and other circles in the figure are destinations. All links have the same length of 100 [m] and each link has a section with a bottleneck as in Section 5.1; the saturation flow rates of links where are far from the origin are set smaller than those of links near the origin so that the interactions among destinations due to queue spillbacks occur. A single peaked total OD flow rate is given as in Section 5.1; the ratio of users for destinations, which indicated by "D" in the network, is set double to users for other destinations. Note that we calculate the average network throughputs and vehicle accumulations for every 6 min here until the vehicle accumulation decreases.

Figure 12(a) shows the simulated and analytical NEFs for the no-control case. It demonstrates that the (dynamic) analytical NEF agrees well with the simulated NEF also for the medium-size network. In this scenario, the network throughput decreases more significantly than that of



(a) Simulated and analytical NEFs (b) Individual controls (c) Hybrid control

Figure 12: Network exit functions for the medium size network

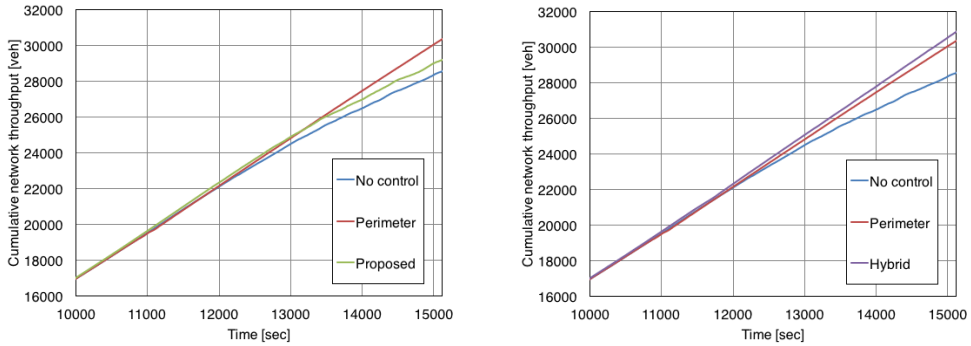
Section 5.1 because more destinations affected by queue spillbacks.

Figure 12(b) shows the NEFs under the perimeter control and proposed network control (no control case is also presented for comparison). The threshold value of the accumulation for the perimeter control was set 500 [veh] to maintain the free-flow (near capacity) state; the proposed control was implemented only inside the target area of the perimeter control. From this figure, we see that both the perimeter and proposed controls work effectively but in a different way. Specifically, the perimeter control *prevents the network performance from degrading* by avoiding the blocking of destinations; the proposed control *improve the network performance* by reducing the through traffic of congested destinations, but the effect disappears due to queue spillbacks for high accumulation (i.e., more than 1000 [veh]). Note that the improvement of the network performance is obtained even for low vehicle accumulation (i.e., 200 – 500 [veh]), which means that the proposed control is activated (i.e., both incoming links for some destinations are congested) before the occurrence of queue spillbacks.

Figure 12(c) shows the NEF under the hybrid control (and the pure perimeter control). As expected, its shape reflects the above-mentioned complementary features of the perimeter and proposed controls. Figures 13(a) and 13(b) show cumulative network throughputs until 15,120 [sec] (the time when vehicle accumulation decreases under no control case). The network throughputs under the perimeter control, proposed control and hybrid control are about 6.3%, 2.2% and 8.1% higher than no control case, respectively. The total time spent by users is significantly improved by about 20.5% (perimeter), 15.8% (proposed) and 32.3% (hybrid). Hence, in this particular scenario, we see that the effect of the hybrid control policy is almost the sum of individual controls.

6. Concluding remarks

In this study, we considered the relationship between an MFD and the congestion patterns in a general network with one-to-many OD demands. We first derived a network throughput analytically by solving an inverse problem of DUE assignment for a fixed vehicle accumulation. This enabled us to incorporate the effects of network configuration and route choice behaviors into the analysis of the network throughput. We then conducted sensitivity analysis of the formula and clarified the following: (1) the mechanisms underlying the decrease in network throughput; (2) the condition for occurrence of the capacity increasing paradox; (3) local and distributed signal control strategy for improving the network throughput. Finally, we validated the proposed



(a) Individual controls

(b) Hybrid control

Figure 13: Cumulative network throughputs for the medium size network

method and examined the effectiveness of signal control policies (including a hybrid one combining with a perimeter control) through numerical examples.

This study shows that the interactions between users with different destinations are the main causes for a decrease in network throughput. This implies that the same decreasing mechanisms may not work in a network with many-to-one OD demands. In fact, Satsukawa and Wada (2015) showed the simple examples that the network throughputs for many-to-one networks do not decrease because of queue spillbacks. Also, we can analyze such a property and the difference between the network performances with one-to-many/many-to-one OD demands through the proposed sensitivity analysis. For many-to-many OD demands, the present decreasing mechanisms should work, but the investigation of analytical properties of the network throughput is not trivial because the analysis of DUE itself is difficult. However, we may expect to extend our analysis to many-to-many networks because the present method does not need to treat a complex variational inequality problem for the DUE explicitly (i.e., congestion pattern is given).

We investigated signal control policies for a certain type of network, but their properties could depend on the network structure. Thus, a systematic numerical experiment should be conducted for different types of networks. The stability analysis of congestion patterns is also important to be implemented for controls. Moreover, the present theory implicitly assumes a stable relationship between an MFD and a congestion pattern. However, the empirical studies on this issue are limited because most focused only on the aggregated index (e.g., the variance of link density). Thus, exploring such a relationship is an important topic for future studies (see Wang et al., 2016, for a recent attempt).

Appendix A. The DUE solution on a reduced network

By following the proof in Akamatsu and Heydecker (2003a), we will show that Eq.(2.14) holds for the variables and matrices defined on any reduced network (here we use subscript R to indicate these explicitly). That is,

$$(\mathbf{A}_R \mathbf{M}_R \mathbf{A}_{R-}^T)\, \dot{\boldsymbol{\tau}}_R(s) = \dot{\mathbf{Q}}_R(s). \tag{A.1}$$

Let $\mathcal{L}_Q$ and $\mathcal{L}_F$ be the set of saturated links and set of non-saturated links with positive flows of the original networks, respectively. We also divide $\mathbf{y}$, $\mathbf{c}$, $\mathbf{M}$ and $\mathbf{A}$ into two blocks

corresponding to $\mathcal{L}_Q$ and $\mathcal{L}_F$:

$$\mathbf{y} = \left[\begin{array}{c|c} \mathbf{y}_Q & \mathbf{y}_F \end{array} \right]^T, \quad \mathbf{c} = \left[\begin{array}{c|c} \mathbf{c}_Q & \mathbf{c}_F \end{array} \right]^T, \quad \mathbf{M} = \left[\begin{array}{c|c} \mathbf{M}_R & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{M}_F \end{array} \right], \quad \mathbf{A} = \left[\begin{array}{c|c} \mathbf{A}_Q & \mathbf{A}_F \end{array} \right].$$

where $\mathbf{M}_R = \mathbf{M}_Q$. Then Eqs.(2.8)–(2.10) reduce to

$$\begin{cases} \dot{\mathbf{c}}_Q(s) - \mathbf{M}_R^{-1}\mathbf{y}_Q(s) - \mathbf{A}_{Q+}^T\dot{\boldsymbol{\tau}}(s) = \mathbf{0} \\ \dot{\mathbf{c}}_F(s) = \mathbf{0} \end{cases} \quad \forall s \tag{A.2}$$

$$\mathbf{A}_Q\mathbf{y}_Q(s) + \mathbf{A}_F\mathbf{y}_F(s) = -\dot{\mathbf{Q}}(s) \quad \forall s \tag{A.3}$$

$$\begin{cases} \mathbf{c}_Q(s) + \mathbf{A}_Q^T\boldsymbol{\tau}(s) = \mathbf{0} \\ \mathbf{c}_F(s) + \mathbf{A}_F^T\boldsymbol{\tau}(s) = \mathbf{0} \end{cases} \quad \forall s. \tag{A.4}$$

By taking the derivative of the third equation with respect to s and substituting it into the first equation, we have

$$\begin{cases} \mathbf{y}_Q(s) = -(\mathbf{M}_R\mathbf{A}_{Q-}^T)\dot{\boldsymbol{\tau}}(s) \\ \mathbf{A}_F^T\dot{\boldsymbol{\tau}}(s) = \mathbf{0} \end{cases} \tag{A.5}$$

The second equation indicates that $\dot{\tau}_i(s)$ and $\dot{\tau}_j(s)$ are identical if link (i, j) is non-saturated, which is the evidence that the procedure (a) of constructing a reduced network is valid. This also implies that $\dot{\boldsymbol{\tau}}(s)$ can be recovered from that of the corresponding reduced network, $\dot{\boldsymbol{\tau}}_R(s)$, as follows.

$$\dot{\boldsymbol{\tau}}(s) = \mathbf{R}^T\dot{\boldsymbol{\tau}}_R(s) \tag{A.6}$$

where an ($|\mathcal{N}|$ - $|\mathcal{L}_F|$)× $|\mathcal{N}|$ matrix $\mathbf{R}$ whose the element in the i-th row and j-th column is 1 if node i on the reduced network corresponds to node j on the original network, 0 otherwise. The similar identities hold for other matrices: (i) $\mathbf{R}\dot{\mathbf{Q}}(s) = \dot{\mathbf{Q}}_R(s)$; (ii) $\mathbf{R}\mathbf{A}_Q = \mathbf{A}_R$; (iii) $\mathbf{R}\mathbf{A}_F = \mathbf{0}$. Further note that $[\mathbf{R}^T\mathbf{R}]$ and $[\mathbf{R}\mathbf{R}^T]$ are invertible, and thus we have (i') $\dot{\mathbf{Q}}(s) = [\mathbf{R}^T\mathbf{R}]^{-1}\mathbf{R}^T\dot{\mathbf{Q}}_R(s)$ and (ii') $\mathbf{A}_Q = [\mathbf{R}^T\mathbf{R}]^{-1}\mathbf{R}^T\mathbf{A}_R$.

By substituting Eq.(A.5) into Eq.(A.3) and multiplying it with $[\mathbf{R}^T\mathbf{R}]$ from both sides,

$$[\mathbf{R}^T\mathbf{R}]\left\{(\mathbf{A}_Q\mathbf{M}_R\mathbf{A}_{Q-}^T)\dot{\boldsymbol{\tau}}(s) - \mathbf{A}_F\mathbf{y}_F(s)\right\} = [\mathbf{R}^T\mathbf{R}]\dot{\mathbf{Q}}(s)$$

is obtained. This is true because $[\mathbf{R}^T\mathbf{R}]$ is a one-to-one mapping from $|\mathcal{N}|$ dimensional Euclid space to itself. By using the above identity (iii), this reduces to

$$[\mathbf{R}^T\mathbf{R}]\left\{(\mathbf{A}_Q\mathbf{M}_R\mathbf{A}_{Q-}^T)\dot{\boldsymbol{\tau}}(s)\right\} = [\mathbf{R}^T\mathbf{R}]\dot{\mathbf{Q}}(s) \tag{A.7}$$

The LHS of this equation can be rewritten as

$$\begin{aligned} [\mathbf{R}^T\mathbf{R}]\left\{(\mathbf{A}_Q\mathbf{M}_R\mathbf{A}_{Q-}^T)\dot{\boldsymbol{\tau}}(s)\right\} &= [\mathbf{R}^T\mathbf{R}]\left\{[\mathbf{R}^T\mathbf{R}]^{-1}\mathbf{R}^T\mathbf{A}_R\right\}\mathbf{M}_R\mathbf{A}_{Q-}^T\left\{\mathbf{R}^T\dot{\boldsymbol{\tau}}_R(s)\right\} \\ &= \mathbf{R}^T\mathbf{A}_R\mathbf{M}_R\mathbf{A}_{R-}^T\dot{\boldsymbol{\tau}}_R(s) \end{aligned} \tag{A.8}$$

where we use the relationships (ii'), (ii) and Eq.(A.6). We thus finally have

$$\mathbf{R}^T\mathbf{A}_R\mathbf{M}_R\mathbf{A}_{R-}^T\dot{\boldsymbol{\tau}}_R(s) = \mathbf{R}^T\dot{\mathbf{Q}}_R(s), \tag{A.9}$$

which implies that Eq.(A.1) holds true. Note that if there is a link whose terminal node is the origin, Eq.(A.1) needs to be modified as we discussed in Subsection 2.3.

Appendix B. Proofs of Proposition 3 and Proposition 4

By the calculation of the matrix elements, the sensitivity coefficient (4.5) is reduced to the following equation:

$$\frac{\partial F}{\partial \mu_{kl}} = \begin{cases} 0 & \text{if} \quad k,l \in \mathcal{N}_d & \text{[Case 1]} \\ 1 & \text{if} \quad k = o \ \wedge \ l \in \mathcal{N}_d & \text{[Case 2]} \\ -1 & \text{if} \quad k \in \mathcal{N}_d \ \wedge \ l = o & \text{[Case 3]} \\ \sum_{i\in\mathcal{N}_i} \sum_{d\in\{I(i)\cap\mathcal{N}_d\}} \mu_{di} v_{il}^{-1} \dot{\tau}_l & \text{if} \quad k = o \ \wedge \ l \in \mathcal{N}_i & \text{[Case 4]} \\ -\sum_{i\in\mathcal{N}_i} \sum_{d\in\{I(i)\cap\mathcal{N}_d\}} \mu_{di} v_{ik}^{-1} & \text{if} \quad k \in \mathcal{N}_i \ \wedge \ l = o & \text{[Case 5]} \\ \sum_{i\in\mathcal{N}_i} \sum_{d\in\{I(i)\cap\mathcal{N}_d\}} \mu_{di} [v_{il}^{-1} - v_{ik}^{-1}] \dot{\tau}_l & \text{if} \quad k,l \in \mathcal{N}_i & \text{[Case 6]} \\ 1 - \sum_{i\in\mathcal{N}_i} \sum_{d\in\{I(i)\cap\mathcal{N}_d\}} \mu_{di} v_{ik}^{-1} & \text{if} \quad k \in \mathcal{N}_i \ \wedge \ l \in \mathcal{N}_d & \text{[Case 7]} \\ [\sum_{i\in\mathcal{N}_i} \sum_{d\in\{I(i)\cap\mathcal{N}_d\}} \mu_{di} v_{il}^{-1} - 1] \dot{\tau}_l & \text{if} \quad k \in \mathcal{N}_d \ \wedge \ l \in \mathcal{N}_i & \text{[Case 8]} \end{cases}. \tag{B.1}$$

where v_{kl}^{-1} represents (k,l) element of the matrix $(\mathbf{V}_{ii})^{-1}$, and we use the fact that the (m,n) element of the matrix $\mathbf{A}_a \mathbf{I}_{kl} \mathbf{A}_{b-}^T$ $(a,b \in \{i,d\})$ is given by

$$[\mathbf{A}_a \mathbf{I}_{kl} \mathbf{A}_{b-}^T]_{mn} = \begin{cases} 1 & \text{if} \quad \{a = b \ \wedge \ l \in \mathcal{N}_a\} \ \wedge \ \{m = n = l\} \\ -1 & \text{if} \quad \{k \in \mathcal{N}_a \ \wedge \ l \in \mathcal{N}_b\} \ \wedge \ \{m = k \ \wedge \ n = l\} \quad a,b \in \{i,d\}. \\ 0 & \text{otherwise} \end{cases} \tag{B.2}$$

Let us examine whether the sensitivity coefficient (B.1) for each link type is positive (i.e., Proposition 3) or negative (i.e., Proposition 4). The first three cases are obvious: nothing occurs in Case 1 because the changes in the throughputs of different destinations by changes in the link capacity are cancelled out each other; Case 2 corresponds to Proposition 3-(i); and Case 3 corresponds to Proposition 4-(i).

The sensitivity coefficients in Cases 4 and 5 are nonzero if there exists a route on the reduced network from the origin to node l or k that passes through at least one destination (i.e., the second term of the formula (3.10) is nonzero). Furthermore, because the matrix $(\mathbf{V}_{ii})^{-1}$ is a non-negative matrix (see Appendix C) and $\dot{\tau}_l$ is positive, the sensitivity coefficient is positive in Case 4, which corresponds to Proposition 3-(ii). Similarly, in Case 5, the sensitivity coefficient is negative, which is corresponding to Proposition 4-(ii).

In Case 6, the sensitivity coefficient could be either positive or negative, which is dependent of the congestion pattern (i.e., the topology and capacity pattern of the reduced network), corresponding to Proposition 3-(iii) or Proposition 4-(iii), respectively.

In Case 7, the sensitivity coefficient could become either positive or negative in general. However, if we consider the situation in which the reduced network has no loops, we can prove that the sensitivity coefficient must be non-negative. This means that Case 7 cannot be a necessary condition for the occurrence of the paradox. Thus, we can conclude that this case corresponds to Proposition 3-(iv).

For the proof, we use the following formula of the element of the matrix $\mathbf{V}^{-1}$, which is applicable to saturated networks (or reduced networks with no loops).

Lemma 5. (Eq.(25) in Akamatsu and Heydecker, 2003b) *The* (i,k) *element of the matrix* $\mathbf{V}^{-1}$

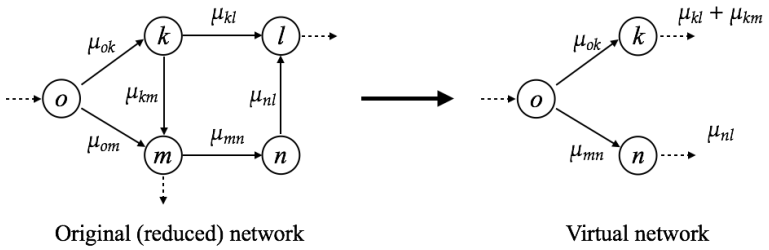


Figure B.14: An example of virtual network

for a one-to-many saturated network, which is denoted by $[\mathbf{V}^{-1}]_{ik}$*, is given by*

$$[\mathbf{V}^{-1}]_{ik} = \frac{1}{\sum_{j\in I(i)} \mu_{ji}} \sum_{r\in\hat{\mathcal{R}}(i,k)} \prod_{(a,b)\in r} \frac{\mu_{ab}}{\sum_{a'\in I(b)} \mu_{a'b}} \tag{B.3}$$

where $\hat{\mathcal{R}}(i,k)$ *is the set of directed routes from node* k *to node* i*.*

This formula cannot be applied directly to the element of $(\mathbf{V}_{ii})^{-1}$ because the matrices $\mathbf{V}$ and $\mathbf{V}_{ii}$ are different. However, let us consider a virtual one-to-many reduced network comprising only transient nodes (see Figure B.14): (a) the initial nodes of incoming links to transient nodes from non-transient nodes (i.e., the origin or destination nodes) are unified into one dummy origin; (b) the capacities of outgoing links to destination nodes from transient nodes are regarded as OD demands. Then, the matrices $\mathbf{V}_{ii}$ play the same role as $\mathbf{V}$ for the virtual reduced network. Thus, we can use formula (B.3) in terms of directed routes that pass through only transient nodes (we denote the set of these routes by $\mathcal{R}(i,k)$).

By substituting the formula into the sensitivity coefficient in Case 7, we have

$$\frac{\partial F}{\partial \mu_{kl}} = 1 - \sum_{i\in\mathcal{N}_i} \sum_{d\in\{I(i)\cap\mathcal{N}_d\}} \left\{ p_{id} \sum_{r\in\mathcal{R}(i,k)} \prod_{(a,b)\in r} p_{ba} \right\} \quad \text{where} \quad p_{ba} \equiv \frac{\mu_{ab}}{\sum_{a'\in I(l)} \mu_{a'b}}. \tag{B.4}$$

Now let us prove that Eq. (B.4) is non-negative. As shown by Akamatsu and Heydecker (2003b), $\sum_{r\in\mathcal{R}(i,k)} \prod_{(a,b)\in r} p_{ab}$ in Eq. (B.4) can be interpreted as the probability P_{ki} from node k to node i when vehicles move subject to the Markov chain rule on a *reverse network*, which is obtained by reversing the direction of all links in the reduced network, with the transition probability from node b to node a defined as p_{ba}. Using this concept, Eq. (B.4) can be written as

$$\frac{\partial F}{\partial \mu_{kl}} = 1 - \sum_{i\in\mathcal{N}_i} \sum_{d\in\{I(i)\cap\mathcal{N}_d\}} \{p_{id} \cdot P_{ki}\} = 1 - \sum_{d\in\mathcal{N}_d} P_{kd}. \tag{B.5}$$

Because there is no loop in the reduced network, the probability from node k to the dummy origin node on the virtual network must be one: $P_{ko} + \sum_{d\in\mathcal{N}_d} P_{kd} = 1$, which implies that $\sum_{d\in\mathcal{N}_d} P_{kd} \leq 1$. Thus, Eq. (B.5) is always non-negative if the reduced network has no loops.

Similarly, the sensitivity coefficient for Case 8 becomes non-positive when the reduced network has no loops, which implies that Case 8 cannot be a necessary condition for decreasing in the network throughput. Thus, we can conclude that this case corresponds to Proposition 4-(iv). □

Appendix C. Properties of the matrix $\mathbf{V}_{ii}$

We prove that matrix $(\mathbf{V}_{ii})^{-1}$ is a non-negative matrix. For evidence, we use the following property of a non-negative matrix: *the inverse of any non-singular M-matrix is a non-negative matrix*. An M-matrix is a Z-matrix with eigenvalues whose real parts are positive, and a Z-matrix is a matrix whose non-diagonal elements are less than or equal to zero. Here, matrix $\mathbf{V}_{ii}$ is always a non-singular matrix because it plays the same role as matrix $\mathbf{V}$ for a virtual one-to-many network comprising only transient nodes, and matrix $\mathbf{V}$ for the reduced network is invertible (see Akamatsu and Heydecker, 2003a). In addition, it is obvious that matrix $\mathbf{V}_{ii}$ is a Z-matrix from Eq. (3.14).

Thus, it is sufficient to prove that the eigenvalues of matrix $\mathbf{V}_{ii}$ are always positive. In order to do this, we use the *Gershgorin circle theorem* (e.g., Horn and Johnson, 1990). From this theorem, we see that all the eigenvalues of matrix $\mathbf{V}_{ii}$ are positive or zero, if the following inequality is satisfied for each column i: $v_{ii} - \sum_{j \neq i} |v_{ji}| \geq 0$, where v_{ji} is element (j, i) of matrix $\mathbf{V}_{ii}$. The non-diagonal elements in the i-th column of matrix $\mathbf{V}_{ii}$ represent the capacities of links incoming to transient node i from other "transient" nodes, and the diagonal element represents the sum of the capacities of links incoming to node i from other nodes. Because the sum of the capacities of links incoming to transient node i should be greater than or equal to that of links incoming from only transient nodes, the diagonal element is always greater than or equal to the sum of the non-diagonal elements for each column. Moreover, matrix $\mathbf{V}_{ii}$ does not have zero eigenvalue because it is a non-singular matrix. Thus, we can conclude that all eigenvalues of matrix $\mathbf{V}_{ii}$ are always positive and matrix $(\mathbf{V}_{ii})^{-1}$ is a non-negative matrix. □

References

Aboudolas, K., Geroliminis, N., 2013. Perimeter and boundary flow control in multi-reservoir heterogeneous networks. Transportation Research Part B 55, 265–281.

Akamatsu, T., 2000. A dynamic traffic equilibrium assignment paradox. Transportation Research Part B 34 (6), 515–531.

Akamatsu, T., 2001. An efficient algorithm for dynamic traffic equilibrium assignment with queues. Transportation Science 35 (4), 389–404.

Akamatsu, T., Heydecker, B., 2003a. Detecting dynamic traffic assignment capacity paradoxes: Analysis of non-saturated networks. Working Paper, Tohoku University, <http://www.plan.civil.tohoku.ac.jp/~akamatsu/Publications/PDF/TS-Dynmc_Prdx-Prt2(030422).pdf>.

Akamatsu, T., Heydecker, B., 2003b. Detecting dynamic traffic assignment capacity paradoxes in saturated networks. Transportation Science 37 (2), 123–138.

Akamatsu, T., Kuwahara, M., 1999. A capacity increasing paradox for a dynamic traffic assignment with departure time choice. In: Ceder, A. (Ed.), Proceedings of the 14th International Symposium on Transportation and Traffic Theory. Pergamon, Jerusalem, pp. 301–324.

Akamatsu, T., Wada, K., Hayashi, S., 2015. The corridor problem with discrete multiple bottlenecks. Transportation Research Part B 81, 808–829.

Ardekani, S., Herman, R., 1987. Urban network-wide traffic variables and their relations. Transportation Science 21 (1), 1–16.

Braess, D., 1968. Über ein Paradoxon aus der Verkehrsplanung. Mathematical Methods of Operations Research 12 (1), 258–268.

Carey, M., Ge, Y. E., McCartney, M., 2003. A whole-link travel-time model with desirable properties. Transportation Science 37 (1), 83–96.

Daganzo, C. F., 1998. Queue spillovers in transportation networks with a route choice. Transportation Science 32 (1), 3–11.

Daganzo, C. F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B 41 (1), 49–62.

Daganzo, C. F., Gayah, V. V., Gonzales, E. J., 2011. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. Transportation Research Part B 45 (1), 278–288.

Daganzo, C. F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B 42 (9), 771–781.
Geroliminis, N., Daganzo, C. F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B 42 (9), 759–770.
Geroliminis, N., Sun, J., 2011a. Hysteresis phenomena of a Macroscopic Fundamental Diagram in freeway networks. Transportation Research Part A 45 (9), 966–979.
Geroliminis, N., Sun, J., 2011b. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transportation Research Part B 45 (3), 605–617.
Godfrey, J. W., 1969. The mechanism of a road network. Traffic Engineering and Control 11, 323–327.
Haddad, J., Geroliminis, N., 2012. On the stability of traffic perimeter control in two-region urban cities. Transportation Research Part B 46 (9), 1159–1176.
Haddad, J., Shraiber, A., 2014. Robust perimeter control design for an urban region. Transportation Research Part B 68, 315–332.
Herman, R., Prigogine, I., 1979. A two-fluid approach to town traffic. Science 204 (4389), 148–151.
Heydecker, B. G., Addison, J. D., 1996. An exact expression of dynamic equilibrium. In: Lesort, J. B. (Ed.), Proceedings of the 13th International Symposium on Transportation and Traffic Theory. Pergamon-Elsevier, Lyon, pp. 359–384.
Horn, R. A., Johnson, C. R., 1990. Matrix Analysis. Cambridge University Press.
Iryo, T., 2011. Solution algorithm of Nash equilibrium in dynamic traffic assignment with discretised vehicles. Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management) 67 (1), 70–83.
Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., Papageorgiou, M., 2012. Exploiting the fundamental diagram of urban networks for feedback-based gating. Transportation Research Part B 46 (10), 1393–1403.
Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., Papageorgiou, M., 2015. Multiple concentric gating traffic control in large-scale urban networks. IEEE Transactions on Intelligent Transportation Systems 16 (4), 2141–2154.
Knoop, V., Hoogendoorn, S., Van Lint, J., 2012. Routing strategies based on macroscopic fundamental diagram. Transportation Research Record 2315 (2315), 1–10.
Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. Transportation Research Part B 96, 26–45.
Kuwahara, M., Akamatsu, T., 1993. Dynamic equilibrium assignment with queues for one-to-many OD pattern. In: Daganzo, C. F. (Ed.), Proceedings of the 12th International Symposium on Transportation and Traffic Theory. Elsevior, Berkeley, pp. 185–204.
Laval, J. A., Castrillón, F., 2015. Stochastic approximations for the macroscopic fundamental diagram of urban networks. Transportation Research Part B 1, 1–13.
Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in simple networks with route choice. Transportation Research Part B 57, 468–484.
Leclercq, L., Parzani, C., Knoop, V. L., Amourette, J., Hoogendoorn, S. P., 2015. Macroscopic traffic dynamics with heterogeneous route patterns. Transportation Research Part C 59, 292–307.
Mahmassani, H., Williams, J., Herman, R., 1987. Performance of urban traffic networks. In: Gartner, N., Wilson, N. (Eds.), Proceedings of the 10th International Symposium on Transportation and Traffic Theory. Elsevior, Amsterdam, The Netherlands, pp. 1–20.
Mahmassani, H. S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. Transportation Research Part C 36, 480–497.
Mazloumian, A., Geroliminis, N., Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. Philosophical Transactions of Royal Society A 368 (1928), 4627–47.
Mounce, R., 2009. Existence of equilibrium in a continuous dynamic queueing model for traffic networks with responsive signal control. In: Lam, W. H. K., Wong, S. C., Lo, H. K. (Eds.), Proceedings of the 18th International Symposium on Transportation and Traffic Theory. Springer, Hong Kong, pp. 327–344.
Newell, G., 2002. A simplified car-following theory: a lower order model. Transportation Research Part B 36 (3), 195–205.
Satsukawa, K., Wada, K., 2015. Effect of origin-destination structures on network performance: Some simple examples. In: Proceedings of the 20th International Conference of Hong Kong Society for Transportation Studies. pp. 526–534.
Satsukawa, K., Wada, K., 2017. A note on the solution algorithm of Nash equilibrium in dynamic traffic assignment for single destination networks. Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management) 73 (1), 103–108.
Smith, M., 2015. Traffic signal control and route choice: A new assignment and control model which designs signal timings. Transportation Research Part C 58, 451–473.
Smith, M. J., 1978. In a road network, increasing delay locally can reduce delay globally. Transportation Research 12 (6), 419–422.
Smith, M. J., 1979a. The existence, uniqueness and stability of traffic equilibria. Transportation Research Part B 13 (4), 295–304.

Smith, M. J., 1979b. Traffic control and route-choice; a simple example. Transportation Research Part B 13 (4), 289–294.
Smith, M. J., 1980. A local traffic control policy which automatically maximises the overall travel capacity of an urban road network. Traffic Engineering and Control 21 (6), 298–302.
Smith, M. J., 1987. Traffic control and traffic assignment in a signal-controlled network with queuing. In: Gartner, N. H., Wilson, N. H. M. (Eds.), Proceedings of the 10th International Symposium on Transportation and Traffic Theory. Elsevior, Cambridge, Massachusetts, pp. 61–77.
Smith, M. J., 1993. A new dynamic traffic model and the existence and calculation of dynamic user equilibria on congested capacity-constrained road networks. Transportation Research Part B 27 (1), 49–63.
Smith, M. J., Liu, R., Mounce, R., 2015. Traffic control and route choice: Capacity maximisation and stability. Transportation Research Part B 81, 863–885.
Wang, P., Wada, K., Akamatsu, T., Sugita, M., Nagoya, T., Sumi, H., 2016. Characterization of Macroscopic Fundamental Diagrams based on long-term detectors data: Case studies of Sendai and Kyoto cities. JSTE Journal of Traffic Engineering 2 (7), 11–20.
Webster, F. V., 1958. Traffic signal settings. Road Research Technical Paper No. 39.
Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. Transportation Research Part B 70, 186–200.
Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. Transportation Research Part C 59, 404–420.
Yoshii, T., Yonezawa, Y., Kitamura, R., 2010. Evaluation of an area metering control method using the macroscopic fundamental diagram. In: Proceedings of the 12th World Conference of Transport Research.