Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 10, No. 3, pp. 483–502, July 2014

A COMPARISON OF NETWORK INEFFICIENCY BETWEEN LOGIT-BASED STOCHASTIC USER EQUILIBRIUM AND DETERMINISTIC USER EQUILIBRIUM

Jing Jiao* and Jinxing Xie*†

Abstract: A traffic network without central authority often evolves efficiency loss issues, since its users try to minimize their own travel time. If users have perfect information about their accurate travel time, the traffic flow assignment becomes deterministic user equilibrium (UE), whereas the traffic flow assignment in a network whose users have incomplete information becomes logit-based stochastic user equilibrium (SUE). Results obtained from the literature provide the bounds of inefficiency of UE and SUE respectively. This study focuses on the comparison of inefficiency of UE and that of SUE. We define the relative performance ratio of a network as the ratio of its inefficiency of SUE to that of UE, and we present the lower bound and the upper bound of the relative performance ratio. Then we focus on a two-link network. We identify regions where the logit-based SUE leads to a system optimum. Furthermore, we determine a flow region in which the relative performance ratio is strictly less than one, which means SUE performs better than UE. The conditions under which the relative performance ratio reaches its lower bound are also achieved. We further extend our study of the relative performance ratio to networks with parallel routes.

Key words: *price of anarchy, stochastic user equilibrium, deterministic user equilibrium, nonatomic network game, relative performance ratio*

Mathematics Subject Classification: *90B15, 90B20*

---

1 Introduction

In a traffic network, due to network users' selfishness and the lack of central authority, network users will try to minimize their own individual perceived travel time. The resultant equilibrium flow pattern may behave inconsistently with that in a system optimum (SO) solution, which minimizes the total travel time of all users. Given a congested network, a well received assumption in the literature is that all users have perfect information about traffic conditions over the entire network and that they consistently make the correct decisions regarding route choice. This assumption will lead to deterministic user equilibrium (UE) assignments.

A large amount of literature is devoted to evaluate the inefficiency of UE. Koutsoupias and Papadimitriou[7] first analyzed the inefficiency of equilibria from a worst-case perspective, and introduced the concept "coordination ratio". The price of anarchy (PoA), which is defined as the worst-possible ratio of the total travel time of a UE to that of an SO solution,

*The authors wish to acknowledge the support from National Natural Science Foundation of China (Projects 71031005 and 71210002).
†Corresponding author

was first introduced by Papadimitriou[10] to depict network inefficiency. Roughgarden and Tardos[13] quantified PoA and Roughgarden[12] showed that PoA is independent of network topology, and a tight upper bound of PoA can be achieved by single-commodity instances in simple networks. This bound was later generalized by Correa et al.[1][2] by applying a geometric technique. A comprehensive survey of PoA can be found in[14].

In reality, however, the assumption that users have perfect information can hardly be achieved[6][15]. Horowitz[6] investigated the stability of the stochastic equilibrium in a two-link network and found that even when the equilibrium is unique, traffic flows may oscillate around the equilibrium perpetually, or converge to values that differ from the equilibrium. Selten et al.[15] conducted a route-choice experiment with a two-route choice scenario. They found that network users repeatedly choose between two congested routes, and the flow pattern fluctuations around UE always persist. Morgan et al.[9] also conducted experiments showing deviation of real world cases from the theoretical user equilibrium.

When network users have perception errors on their travel times, assuming that their perceived travel times are flow-dependent random variables, the resultant flow pattern is a stochastic user equilibrium (SUE) assignment[16]. By introducing the logit choice model[8] to describe users' choice behavior, one can get the logit-based SUE, which is generally used in literature to describe equilibrium flow patterns when network users can not perceive the accurate travel times.

The inefficiency of logit-based SUE has recently been studied by Guo et al.[5] for the first time. They derived an upper bound on the ratio of the total travel time of a logit-based SUE to that of an SO solution, which is formulated by a multiplication of the UE bound given by Roughgarden and Tardos[13] and a constant strictly larger than 1. An interesting question remains to be answered is, whether it is possible for an SUE assignment to behave more efficiently than a UE assignment, such that the network can benefit from users' inaccurate perception. Furthermore, if the conditions under which SUE behaves better than UE do exist, we want to identify these conditions.

This paper contributes to fill the above void. We define the relative performance ratio of a network as the inefficiency of SUE to that of UE. We find lower and upper bounds for the relative performance ratio. Then we study a two-link network, which is a generalization of Pigou's example[11], and we show that the logit-based SUE does benefit networks when the total flow amount in the network is in certain range. Regions where the logit-based SUE leads to SO, while UE cannot, are also found. We also derive the conditions under which the relative performance ratio reaches its lower bound. Finally we extend our study of the relative performance ratio to networks with parallel routes and we find that the relative performance ratio can still reach its lower bound in such networks.

The rest of this paper is organized as follows. Section 2 presents the network structure we consider and gives definitions of UE and the logit-based SUE. We define the relative performance ratio and present its bounds in Section 3. In Section 4, we characterize some properties of the logit-based SUE in the special network, and analyze how the relative performance ratio varies with network parameters. We extend our study in Section 5 and some concluding remarks are presented in Section 6.

2 Preliminaries

In this section, we describe the network topology and the fundamental results in the study of network inefficiency of UE and SUE. The network setting in this paper is the same as Correa et al.[2] and Guo et al.[5]. It is also similar to that of Roughgarden[12] except that the assumption of differentiable and standard travel time functions is extended to be continuous functions. The definitions of UE and SUE and their corresponding PoA are presented afterwards.

2.1 Network Setting

We consider a network given by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with vertex set $\mathcal{V}$, directed edge set $\mathcal{E}$, and k Origin-Destination (OD) pairs $(s_1, t_1), \ldots, (s_k, t_k)$. Let $\mathcal{P}_i \neq \emptyset$ denote the set of $s_i - t_i$ paths, and $\mathcal{P} = \cup_i \mathcal{P}_i$. A flow on path P is a function $f_P : \mathcal{P} \rightarrow \mathbb{R}_+$, where $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. The network flow $\mathbf{f}$ is a vector $\mathbf{f} = (\ldots, f_P, \ldots), P \in \mathcal{P}$. For a fixed network flow $\mathbf{f}$ we define the flow on edge $e \in \mathcal{E}$ is $f_e = \sum_{P:e \in P} f_P$. We denote the total flow amount between an OD pair (s_i, t_i) as D_i, and we say a flow $\mathbf{f}$ is *feasible* if $\sum_{P \in \mathcal{P}_i} f_P = D_i$ holds for each OD pair $i \in \{1, \ldots, k\}$. Each edge $e \in \mathcal{E}$ has a link travel time function $l_e(f_e) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ which is assumed to be nonnegative, nondecreasing and continuous. The travel time function of a path P with respect to a flow $\mathbf{f}$ is denoted by $l_P(\mathbf{f}) = \sum_{e \in P} l_e(f_e)$. Let $\mathbf{D} = (\ldots, D_i, \ldots), i \in \{1, \ldots, k\}$ and $\mathbf{l} = (\ldots, l_e, \ldots), e \in \mathcal{E}$,we call a triple $(\mathcal{G}, \mathbf{D}, \mathbf{l})$ as an *instance*.

Let $L(\mathbf{f}) = \sum_{P \in \mathcal{P}} l_P(\mathbf{f}) f_P$ denote the total travel time incurred by a given flow $\mathbf{f}$. By summing over the edges in a path P and reversing the order of summation, we can also write $L(\mathbf{f}) = \sum_{e \in \mathcal{E}} l_e(f_e) f_e$. Given an instance $(\mathcal{G}, \mathbf{D}, \mathbf{l})$, the system optimum (SO) solution $\mathbf{f}^{SO}$ is calculated from the minimization problem as below:

$$
\begin{aligned}
SOP: \quad \min_{\mathbf{f}} \quad & \sum_{P \in \mathcal{P}} l_P(\mathbf{f}) f_P \\
\text{s.t.} \quad & \sum_{P \in \mathcal{P}_i} f_P = D_i, \quad \forall i \in \{1, \ldots, k\}, \\
& f_P \geq 0, \quad \forall P \in \mathcal{P}.
\end{aligned}
$$

The corresponding system optimal total travel time is $L(\mathbf{f}^{SO})$, which is written as L^{SO} for short.

2.2 Deterministic user equilibrium

Consider a nonatomic selfish routing game, in which there are infinite users in the network, each of whom owns a negligible amount of flow. Each user will encounter a travel time $l_P(\mathbf{f})$ if he/she chooses route P. For a network without central authority, each user will choose routes that minimize his/her own travel time. Thus, a stable condition is reached only when no user can improve his/her travel time by unilaterally changing routes. This is the characterization of the deterministic user equilibrium (UE) condition[16]. Roughgarden and Tardos[13] give a precise definition of UE, which is referred to as *equilibrium* in their work.

Definition 2.1 (Roughgarden and Tardos[13])**.** Let $\mathbf{f}$ be a feasible flow for the nonatomic instance$(\mathcal{G}, \mathbf{D}, \mathbf{l})$. The flow $\mathbf{f}$ is an *equilibrium flow* if, for every $i \in \{1, \ldots, k\}$ and every pair $P, \tilde{P} \in \mathcal{P}_i$ with $f_P > 0$, $l_P(\mathbf{f}) \geq l_{\tilde{P}}(\mathbf{f})$.

According to Sheffi[16], the UE flow pattern $\mathbf{f}^{UE}$ is equivalent to the optimal solution to the following minimization problem:

$$
\begin{aligned}
UEP: \quad \min_{\mathbf{f}} \quad & \sum_{e\in\mathcal{E}} \int_0^{f_e} l_e(\omega)d\omega \\
\text{s.t.} \quad & \sum_{P\in\mathcal{P}_i} f_P = D_i, \quad \forall i \in \{1,\ldots,k\}, \\
& f_e \geq 0, \quad \forall e \in \mathcal{E},
\end{aligned}
$$

where $f_e = \sum_{P:e\in P} f_P$ for all $e \in \mathcal{E}$.

The existence of a UE flow pattern in an instance $(\mathcal{G}, \mathbf{D}, \mathbf{l})$ is straight forward, since the objective function of UEP is continuous and the feasible region of $\mathbf{f}$ is compact. It is worth mentioning that, though it is possible to have more than one UE flow pattern, the minimization problem UEP has the property that all the UE flows commit the same value of L^{UE}, which equals to the minimization of UEP. Let $L^{UE} = L(\mathbf{f}^{UE})$ denote the total travel time corresponding to a UE flow pattern $\mathbf{f}^{UE}$, the definition of price of anarchy (PoA) given by Koutsoupias and Papadimitriou[7] is reduced to:

Definition 2.2. Given an instance $(\mathcal{G}, \mathbf{D}, \mathbf{l})$, PoA of UE is the ratio of the total travel time of UE to that of SO:

$$\rho^{UE} = L^{UE}/L^{SO}. \tag{2.1}$$

In the end of this subsection we present the tight upper bound of ρ^{UE} given by Correa et al.[1][2].

Theorem 2.3. *Let $\boldsymbol{f}^{UE}$ be the UE flow with separable link travel time functions drawn from a given class $\mathcal{L}$, and $\boldsymbol{f}^{SO}$ be an SO flow, then*

$$\rho^{UE} \leq \frac{1}{1-\gamma(\mathcal{L})}. \tag{2.2}$$

Here, the parameter $\gamma(\mathcal{L})$ satisfies

$$\gamma(\mathcal{L}) = \max_{l_e\in\mathcal{L}, z_e\geq 0} \gamma_e(l_e, z_e), \tag{2.3}$$

in which

$$\gamma_e(l_e, z_e) = \max_{f_e\geq 0} \frac{(l_e(z_e) - l_e(f_e))f_e}{l_e(z_e)z_e}, \quad e \in \mathcal{E}. \tag{2.4}$$

Remark 2.4. As stated by Correa et al. [1], if we add the assumptions that each travel time function $l_e(x)$ is differentiable and $xl_e(x)$ is convex, the network setting becomes the same as that given in Roughgarden [12]. In such a network setting, $\frac{1}{1-\gamma(\mathcal{L})}$ is equal to $\alpha(\mathcal{L})$, which is the upper bound of ρ^{UE} presented in Roughgarden [12].

2.3 Logit-based stochastic user equilibrium

For the sake of capturing behaviors of network users in the real world, we assume that users are not aware of their accurate travel time due to their perception errors. A logit-based choice model, which is generally used in literature to capture network users' behaviors[4][5][16], is used to characterize users' choices among different routes.

In a logit-based choice model, for an OD pair (s_i, t_i), $i \in \{1, \dots, k\}$, given a user's actual travel time l_P when he/she chooses route $P \in \mathcal{P}_i$, he/she will perceive it as $\hat{l}_P$, which is formulated as

$$\hat{l}_P = \theta l_P + \epsilon_P, \quad P \in \mathcal{P}_i. \tag{2.5}$$

The parameter $\theta \in \mathbb{R}_+$ is a constant that scales the perceived travel time. We assume that users are homogeneous, which means that they all share the same θ. It can be interpreted as a representation of all the users' systematical awareness of their accurate travel time. If θ is very large, i.e., $\theta \to +\infty$, the perception error is relatively small and users will tend to choose the minimum measured travel-time path. On the contrary, a small value of θ indicates a large perception variance, with travelers using many routes with different measured travel time. In the limit, when $\theta = 0$, the share of flow on all paths will be equal. The values $\epsilon_P, P \in \mathcal{P}_i$, which characterize the unobservable or immeasurable factors of users' perceived travel time, are assumed to be independently and identically distributed Gumbel variates[16].

Let q_P denote the probability of users choosing route P for $P \in \mathcal{P}_i$. The strategy for each user choosing his/her route is a mixed strategy:

$$\{(\dots, q_P, \dots) | \sum_{P \in \mathcal{P}_i} q_P = 1, q_P \geq 0, P \in \mathcal{P}\}. \tag{2.6}$$

By minimizing their perceived travel time, users calculate q_P as $q_P = Pr\{\hat{l}_P \leq \hat{l}_{\tilde{P}}, \forall \tilde{P} \in \mathcal{P}_i\}$. In other words, the probability that a given route is chosen is the probability that its travel time is perceived to be the lowest of all the alternative routes. q_P is further derived as

$$q_P = \frac{exp(-\theta l_P(\mathbf{f}^{SUE}))}{\sum_{\hat{P} \in \mathcal{P}_i} exp(-\theta l_{\hat{P}}(\mathbf{f}^{SUE}))}, \tag{2.7}$$

where the path flow assignment is given by $f_P^{SUE} = q_P D_i$, $P \in \mathcal{P}_i$, $i \in \{1, \dots, k\}$.

Definition 2.5. If for all OD pairs (s_i, t_i), $i \in \{1, \dots, k\}$, there exist $q_P \in [0, 1]$, $P \in \mathcal{P}_i$, $i \in \{1, \dots, k\}$ that satisfy Equation (2.7), then the strategy $(\dots, q_P, \dots), P \in \mathcal{P}$ is a logit-based SUE, or SUE for short.

Remark 2.6. As stated in Sheffi[16], SUE reduces to UE when $\theta \to +\infty$.

Remark 2.7. From Fisk[4], there exists a unique logit-based SUE for a congested network routing game.

Let $L^{SUE} = L(\mathbf{f}^{SUE})$ denote the total travel time corresponding to the SUE flow pattern $\mathbf{f}^{SUE}$. Similar to the definition of PoA of UE, the inefficiency of SUE defined by Guo et al.[5] is given as below:

Definition 2.8. (Guo et al.[5])The inefficiency of SUE compared to SO is the ratio of the total travel time of SUE to that of SO:

$$\rho^{SUE} = L^{SUE} / L^{SO}. \tag{2.8}$$

In the rest of this paper, we call ρ^{SUE} as PoA of SUE in coherence with PoA of UE. The main result of Guo et al.[5] is an upper bound of ρ^{SUE} as the theorem stated below.

Theorem 2.9. *Let L^{SUE} be the total system travel time under logit-based stochastic user equilibrium, and L^{SO} be the minimum total system travel time, then*

$$\rho^{SUE} \leq (\frac{1}{1 - \gamma(\mathcal{L})})(1 + \frac{\sqrt{6}}{\pi} \bar{\alpha} \varsigma). \tag{2.9}$$

Here, $\gamma(\mathcal{L})$ is the same as in Equation (2.3). The factor $\bar{\alpha}$ relates to the traffic demand D_i and the number of feasible paths $|P_i|$ between each OD pair $i \in \{1, \ldots, k\}$. Mathematically, $\bar{\alpha}$ satisfies $\bar{\alpha} = \sum_{i=1}^{k}(D_i/(\sum_{i=1}^{k} D_i))\alpha_i$, where α_i solves $\alpha_i e^{\alpha_i+1} = |P_i| - 1$. The factor $\varsigma = \pi/(\sqrt{6}\theta\bar{c_0})$ relates to the scaling parameter θ and the average free-flow travel time for all OD pairs $\bar{c_0}$. Apparently $\bar{\alpha}$ and $\bar{c_0}$ are nonnegative, thus ρ^{SUE} has a larger upper bound than ρ^{UE} does. Despite that the tightness of PoA of SUE still remains open[5], it is possible that ρ^{SUE} is larger than ρ^{UE}.

3 Comparison of the Inefficiency between UE and SUE

By comparing the upper bound of ρ^{UE} and ρ^{SUE} given in Sections 2.2 and 2.3, it is quite likely to have the intuition that the network behaves more efficiently in a UE flow assignment than that in an SUE flow assignment, which leads to the conjecture that the network performs better if users have more accurate information about the network. This attracts us to compare a network's PoA of SUE with its PoA of UE to investigate the relations between these two flow modes.

Given the definitions of PoA of UE and PoA of SUE, we define a network's relative performance ratio as the ratio of ρ^{SUE} to ρ^{UE}:

Definition 3.1. The relative performance ratio of a network is the ratio of PoA of SUE to PoA of UE:

$$\varphi = \rho^{SUE}/\rho^{UE}. \tag{3.1}$$

In our network setting, since ρ^{UE} is reduced to L^{UE}/L^{SO}, we can further rewrite φ as:

$$\varphi = L^{SUE}/L^{UE}. \tag{3.2}$$

The bounds of the value of φ can be deducted immediately by applying the results in Theorems 2.3 and 2.9.

Lemma 3.2. *Given the link travel time function class $\mathcal{L}$, the relative performance ratio of a network satisfies:*

$$1 - \gamma(\mathcal{L}) \leq \varphi \leq \Big(\frac{1}{1-\gamma(\mathcal{L})}\Big)\Big(1 + \frac{\sqrt{6}}{\pi}\bar{\alpha}\varsigma\Big). \tag{3.3}$$

Proof. By substituting $L^{UE} \geq L^{SO}$ and $L^{SUE} \geq L^{SO}$ into Equation (3.2), we have $\varphi \leq L^{SUE}/L^{SO} = \rho^{SUE}$ and $\varphi \geq L^{SO}/L^{UE} = 1/\rho^{UE}$. This completes the proof. □

Lemma 3.2 provides both the lower bound and the upper bound of φ. It is not surprising that $\varphi > 1$ because the perceived travel time of a user is the real travel time in the UE case, which enables users to make their right choice. What interests us is, since the lower bound $1 - \gamma(\mathcal{L}) < 1$, it is possible that $\varphi < 1$ holds. If this is true, the network would perform better given that users acquire less information about the network. In such cases, information providers such as navigation machines are actually playing a negative role. Such concern stimulates us to determine how much SUE can be better than UE, and to find the conditions under which SUE outperforms UE.

4 The Relative Performance Ratio in a Two-Link Network

In this section, we consider a special network topology which contains two parallel links (routes). We aim to show the tightness of the lower bound of φ, and to find out when SUE performs better than UE in such network topology. The reason why we pay attention to a two-link network is not only because of its simple structure, but also the fact that ρ^{UE} achieves its upper bound in a two-link network[13]. This result promotes us to start from the basis. Meanwhile, such network topology still reflects the real world. For example, for commuters who drive between their home and work place, it is very common that they make a decision on choosing between the highway and the urban road.

The network is designed to have a vertex set $\mathcal{V} = \{s, t\}$ and a directed edge set $\mathcal{E} = \{e_i | i = 1, 2\}$. Specifically, e_1 and e_2 are two parallel routes connecting the source vertex s and the sink vertex t. The network flow is $\mathbf{f} = (f_1, f_2)$, where f_i is the amount of flow using e_i for $i = 1, 2$. The total flow amount is $D = \sum_{i=1,2} f_i$. The travel time function on each edge e_i is denoted as $l_i(x)$. Let $l_1(x) = c$, which is a constant strictly larger than 0. Assume $l_2(x)$ is a nonnegative, strictly increasing, and strictly standard function, whose definition is given below, and it satisfies that $l_2(0) < c$ and $l_2(x) > c$ for some x sufficiently large. We denote the flow amount r satisfying $l_2(r) = c$ as the *cross point*. An instance of such network is denoted as $(\hat{\mathcal{G}}_2, D, \mathbf{l})$.

Definition 4.1. A travel time function $l_i(x)$ is *strictly standard* if it is differentiable and if the function $xl_i(x)$ is strictly convex on $\mathbb{R}_+$.

Lemma 4.2. *Given an instance* $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$, *there exists a unique* $\lambda \in (0, 1)$, *such that*

$$\frac{d}{dx}(xl_2(x)) \mid_{\lambda r} = c. \tag{4.1}$$

Proof. Since $l_2(x)$ is differentiable, it is continuous. Thus for any $\varepsilon > 0$, there exists a δ such that $l_2(0) < l_2(\delta) < l_2(0) + \varepsilon < c$. Noticing that $l_2(x)$ is strictly increasing, we have $l_2'(x) > 0$, which leads to $\frac{d}{dx}(xl_2(x)) \mid_r > c$. The existence of λ is immediately given by the intermediate value theorem. The uniqueness of λ is obtained since $\frac{d}{dx}(xl_2(x))$ is strictly increasing. □

Remark 4.3. Note that the definition of strictly standard is similar to the definition of standard given by Roughgarden and Tardos[13]. The only difference is that we change the assumption of $xl_i(x)$ from convex to strictly convex for simplicity. The strictness ensures the uniqueness of λ. However, if we remove the strictness and let it be just convex, the point λ changes to an interval $[\underline{\lambda}, \overline{\lambda}]$. In this case, all the analysis in this section won't change, and similar results can still be derived. We will leave it to the interested reader.

Comparing the assumptions of travel time function $l_2(x)$ here and in Section 2.1, we will find that the restriction to travel time functions is stronger here. However, according to Roughgarden[14], a large bunch of travel time functions still satisfy such assumptions. A description of the above network setting is given in Figure 1. When $c = 1$, $l_2(x) = x$ and $D = 1$, the instance becomes the Pigou's example[11].

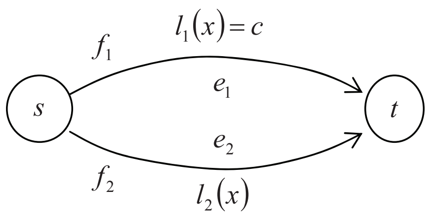

Figure 1: Description of two-link network setting.

For an instance $(\hat{\mathcal{G}}_2, D, \mathbf{l})$, SOP and UEP can be solved directly by using their first order conditions. Thus we present their solutions below without the proofs.

Lemma 4.4. *Given an instance $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$, the system optimum $\boldsymbol{f}^{SO}$ and its corresponding minimum total travel time L^{SO} are*

$$\boldsymbol{f}^{SO} = \begin{cases} (0, D), & if \quad 0 < D \leq \lambda r \\ (D - \lambda r, \lambda r), & if \quad D > \lambda r, \end{cases} \quad and \tag{4.2}$$

$$L^{SO} = \begin{cases} D \cdot l_2(D), & if \quad 0 < D \leq \lambda r \\ (D - \lambda r)c + \lambda r l_2(\lambda r), & if \quad D > \lambda r. \end{cases} \tag{4.3}$$

Lemma 4.5. *Given an instance $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$, the UE flow pattern $\boldsymbol{f}^{UE}$ and its corresponding total travel time L^{UE} are*

$$\boldsymbol{f}^{UE} = \begin{cases} (0, D), & if \quad 0 < D \leq r \\ (D - r, r), & if \quad D > r, \end{cases} \quad and \tag{4.4}$$

$$L^{UE} = \begin{cases} D \cdot l_2(D), & if \quad 0 < D \leq r \\ Dc, & if \quad D > r. \end{cases} \tag{4.5}$$

For an instance $(\hat{\mathcal{G}}_2, D, \mathbf{l})$, SUE is a strategy (q_1, q_2). For simplicity of exposition, we replace the strategy with $(1 - q^*, q^*)$. Let $M = c - l_2(q^* D)$. By Equation (2.7) we rewrite q^* as

$$q^* = \frac{1}{1 + exp(-\theta M)}, \tag{4.6}$$

which leads to

$$L^{SUE} = (1 - q^*)Dc + q^* D \cdot l_2(q^* D). \tag{4.7}$$

Next we start to verify how the relative performance ratio φ behaves in the network. Given the network topology, when total flow amount D varies, we aim to find out how φ changes with θ. Then we focus on the total flow amount region when $\varphi < 1$ holds to calculate the minimum value of φ as well as the conditions under which it reaches its minimum value. Since a trivial outcome of $L(\mathbf{f}) = 0$ will occur when $D = 0$, we assume $D > 0$ throughout this section.

4.1 Properties of SUE in the two-link network

Now we focus on some structural properties of SUE as building blocks for investigating the properties of the relative performance ratio φ.

Lemma 4.6. *For an instance $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$, $q^*|_{\theta \to +\infty} = 1$ for all $D \in (0, r]$, and $q^*|_{\theta \to +\infty} = r/D$ for all $D \in (r, +\infty)$.*

Proof. The value of $q^*|_{\theta \to +\infty}$ can be deduced from (4.4) since SUE reduces to UE when $\theta \to +\infty$[16]. □

Lemma 4.7. *For an instance $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$, $q^* = 1/2$ if and only if either $\theta = 0$ or $D = 2r$.*

Proof. Since q^* is the solution to Equation (4.6), it is obvious that $q^* = 1/2$ if and only if $\theta \cdot M = 0$, which is equivalent to $\theta = 0$, or $l_2(r) = l_2(q^*D)$. Next we prove that $l_2(r) = l_2(q^*D)$ is equivalent to $D = 2r$. If $l_2(r) = l_2(q^*D)$ holds, noticing that $l_2(x)$ is strictly increasing, we have $D = 2r$. On the other hand, when $D = 2r$, if $M > 0$, we have $q^* > 1/2$ according to Equation (4.6). However, since $M = c - l_2(q^*D) > 0$ and $l_2(x)$ is strictly increasing, we obtain that $q^* < 1/2$, which is equal to $M < 0$. The contradiction leads to $M = 0$ and ends our proof. □

Lemma 4.8. *For an instance $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$, q^* satisfies the following properties:*

(i) *for all total flow amount $D \in (0, +\infty)$, $\partial q^*/\partial D < 0$ holds;*

(ii) *for all constant travel time $c \in (0, +\infty)$, $\partial q^*/\partial c > 0$ holds;*

(iii) *for all scaling parameter $\theta \in (0, +\infty)$, $\partial q^*/\partial \theta \begin{cases} > 0, & if\ D < 2r \\ = 0, & if\ D = 2r\ holds; \\ < 0, & if\ D > 2r \end{cases}$ and*

(iv) *given total flow amount D and travel time c, q^* is continuous at $\theta = 0$.*

Proof. We rewrite Equation (4.6) as

$$F(q^*, c, \theta, D) = q^* - \frac{1}{1 + \exp(-\theta M)} = 0. \tag{4.8}$$

By using the implicit function theorem, we rewrite $\partial q^*/\partial D$,$\partial q^*/\partial c$ and $\partial q^*/\partial \theta$ as $\partial q^*/\partial \cdot = -(\partial F/\partial q^*)/(\partial F/\partial \cdot)$, where '$\cdot$' stands for D, c and θ. Taking partial derivative of F in (4.8) with respect to q^*, D and c respectively, we obtain that $\partial F/\partial q^* > 0$, $\partial F/\partial D > 0$ and $\partial F/\partial c < 0$. Thus (i) and (ii) are verified.

When $\theta \in (0, +\infty)$, it follows from Lemma 4.7 and Lemma 4.8(i) immediately that

$$\begin{cases} q^* > 1/2 & \Leftrightarrow D < 2r \\ q^* = 1/2 & \Leftrightarrow D = 2r \\ q^* < 1/2 & \Leftrightarrow D > 2r. \end{cases} \tag{4.9}$$

Taking Equation (4.6) into consideration, we further obtain that

$$\begin{cases} M > 0 & \Leftrightarrow D < 2r \\ M = 0 & \Leftrightarrow D = 2r \\ M < 0 & \Leftrightarrow D > 2r. \end{cases} \tag{4.10}$$

Given the partial derivative of F with respect to θ as $\partial F/\partial \theta = \frac{-M \exp(-\theta M)}{[1+\exp(-\theta M)]^2}$ and $\partial F/\partial q^* > 0$, (iii) is proved.

Noticing that M is bounded when D and c are fixed, we have $\lim_{\theta \to 0^+} q^* = q^*|_{\theta=0} = 1/2$. This completes the proof of Lemma 4.8. □

Lemmas 4.6 and 4.7 show values of q^* at the two extremes of the domain of θ, whereas Lemma 4.8 shows how q^* varies with respect to θ as well as D and c. Lemma 4.8 also guarantees the continuity of q^* with respect to θ, D and c to their respective domains.

For any given network total flow amount D, we investigate the dependence of ρ^{SUE} on the scaling parameter θ.

Lemma 4.9. *Given an instance* $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$,

(i) *when* $D \in (0, \lambda r] \cup (2r, +\infty)$, *for all scaling parameter* $\theta \in (0, +\infty)$, $\partial\rho^{SUE}/\partial\theta < 0$ *holds;*

(ii) *when* $D \in (\lambda r, 2\lambda r]$, *there exists a unique* $\hat{\theta} = \hat{\theta}(D)$, *such that*

   i) $\partial\rho^{SUE}/\partial\theta|_{\hat{\theta}} = 0$ *holds;*

   ii) *for all* $\theta \in (0, \hat{\theta})$, $\partial\rho^{SUE}/\partial\theta < 0$ *holds; and*

   iii) *for all* $\theta \in (\hat{\theta}, +\infty)$, $\partial\rho^{SUE}/\partial\theta > 0$ *holds;*

(iii) *when* $D \in (2\lambda r, 2r)$, *for all* $\theta \in (0, +\infty)$, $\partial\rho^{SUE}/\partial\theta > 0$ *holds;*

(iv) *when* $D = 2r$, *for all* $\theta \in (0, +\infty)$, $\partial\rho^{SUE}/\partial\theta = 0$ *holds; and*

(v) *given fixed total flow amount* D *and travel time* c, ρ^{SUE} *is continuous at* $\theta = 0$.

Proof. Noticing that for any given total flow amount $D > 0$, L^{SO} is a constant, we take the partial derivative of ρ^{SUE} with respect to θ as

$$\begin{aligned}\partial\rho^{SUE}/\partial\theta &= (\partial L^{SUE}/\partial\theta)/L^{SO} \\ &= (D/L^{SO})(\partial q^*/\partial\theta)A(q^*D),\end{aligned} \tag{4.11}$$

where $A(q^*D) = l_2(q^*D) + (q^*D)l_2'(q^*D) - c$. Since $xl_2(x)$ is strictly convex, the function $l_2(x)+xl_2(x)$ is strictly increasing with x. Applying Equation (4.1) and taking the uniqueness of λ into consideration, we derive that

$$\begin{cases} A(q^*D) > 0 & \Leftrightarrow q^*D > \lambda r \\ A(q^*D) = 0 & \Leftrightarrow q^*D = \lambda r \\ A(q^*D) < 0 & \Leftrightarrow q^*D < \lambda r. \end{cases} \tag{4.12}$$

When $D \in (0, \lambda r]$, it is clear that $q^*D < D \leq \lambda r$. By applying (4.12) and Lemma 4.8(iii), it follows immediately from Equation (4.11) that $\partial\rho^{SUE}/\partial\theta < 0$ for all $\theta \in (0, +\infty)$. When $D \in (2r, +\infty)$, we have $M < 0$ from (4.10). Thus we have $q^*D > r > \lambda r$. By following the same method as for proving the case of $D \in (0, \lambda r]$, we obtain that $\partial\rho^{SUE}/\partial\theta < 0$ holds.

From Lemma 4.6 we obtain that, for any $D \in (\lambda r, min(r, 2\lambda r))$, $\lim_{\theta\to+\infty} q^* = 1 > \lambda r/D$ holds; and for any $D \in [min(r, 2\lambda r), 2\lambda r]$, $\lim_{\theta\to+\infty} q^* = r/D > \lambda r/D$ holds. Together with Lemma 4.7, we conclude that for any $D \in (\lambda r, 2\lambda r]$, there exists a unique $\hat{\theta} = \hat{\theta}(D)$, such that $q^*|_{\hat{\theta}} = \lambda r/D$ satisfying $A(q^*D) = 0$. Thus we have $(\partial\rho^{SUE}/\partial\theta)|_{\hat{\theta}} = 0$. Furthermore, the inequations $\partial\rho^{SUE}/\partial\theta > 0$ for $\theta > \hat{\theta}$ and $\partial\rho^{SUE}/\partial\theta < 0$ for $\theta < \hat{\theta}$ holds.

When $D \in (2\lambda r, 2r)$, it follows from (4.9) that $q* > 1/2$. Thus we have $q^*D > \lambda r$. By using the same method as for proving the case of $D \in (0, \lambda r]$, we find $\partial\rho^{SUE}/\partial\theta > 0$ holds for all $\theta \in (0, +\infty)$.

The proofs of (iv) and (v) follow directly from Lemma 4.8(iii) and (iv). This completes our proof. □

In the rest of this paper, we denote $\hat{\theta}(D)$ as the best response scaling parameter. It is indeed a response function of total flow amount D.

The next theorem identifies the situations where the PoA of SUE is 1, which is its best possible value.

Theorem 4.10. *Given an instance* $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$, ρ^{SUE} *is strictly larger than 1 except that*

(i) *when* $D \in (0, \lambda r]$, *we have* $\lim_{\theta\to+\infty} \rho^{SUE} = 1$ *holds; and*

(ii) *when* $D \in (\lambda r, 2\lambda r]$, *we have* $\rho^{SUE}|_{\hat{\theta}} = 1$ *holds.*

Proof. According to Lemma 4.9, it suffices to check the minimum value of ρ^{SUE} when D belongs to different regions. When $D \in (0, \lambda r] \cup (2r, +\infty)$, ρ^{SUE} reaches its minimum value when $\theta \to +\infty$. By following Lemma 4.6, we obtain that $\lim_{\theta\to+\infty} \rho^{SUE} = 1$ for any $D \in (0, \lambda r]$ and $\lim_{\theta\to+\infty} \rho^{SUE} = (cD)/(c(D - \lambda r) + \lambda r l_2(\lambda r)) > 1$ for any $D \in (2r, +\infty)$. Similarly, we can derive that when $D \in (2\lambda r, 2r]$, ρ^{SUE} reaches to its minimum value when $\theta = 0$, and $\rho^{SUE}|_{\theta=0} > 1$. Finally, when $D \in (\lambda r, 2\lambda r]$, it follows directly from the proof of Lemma 4.9 that ρ^{SUE} reaches its minimum value when $\theta = \hat{\theta}$ with $q^*|_{\hat{\theta}} = \lambda r/D$. Thus $\rho^{SUE}|_{\hat{\theta}} = 1$. □

Theorem 4.10 reveals that, despite the trivial case that $\rho^{SUE} = 1$ when SUE goes to UE, SUE is possible to reach SO when UE fails. Especially, such cases happen when total flow amount D is near the cross point r.

4.2 The relative performance ratio in a two-link network

In this section we aim at identifying when and by how much the SUE assignment can outperform that of UE. Following from the definition of φ, we know that the smaller the minimum of φ is, the better SUE behaves.

Theorem 4.11. *Given an instance* $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$, *the relative performance ratio* φ *satisfies*

(i) $\varphi > 1$ *holds for all scaling parameter* $\theta \in \mathbb{R}_+$, *when* $D \in (0, \lambda r] \cup (2r, +\infty)$;

(ii) $\varphi = 1$ *holds for all* $\theta \in \mathbb{R}_+$, *when* $D = 2r$;

(iii) $\varphi < 1$ *holds for all* $\theta \in \mathbb{R}_+$, *when* $D \in (\min(r, 2\lambda r), 2r)$; *and*

(iv) *when* $D \in (\lambda r, \min(r, 2\lambda r)]$, *there exist a* $\bar{\theta} = \bar{\theta}(D) \in [0, \hat{\theta})$, *such that* $\varphi < 1$ *for all* $\theta \in (\bar{\theta}, +\infty)$, *and* $\varphi \geq 1$ *for all* $\theta \in [0, \bar{\theta}]$.

Proof. Because SUE reduces to UE when $\theta \to +\infty$[16], the first two items can be deduced directly from items (i), (iv), and (v) in Lemma 4.9. Similarly, we acquire that when $D \in (2\lambda r, 2r)$, $\varphi < 1$ holds for any $\theta \in \mathbb{R}_+$. Consider the case where $D \in (\min(r, 2\lambda r), 2\lambda r]$. The interval is nonempty only when $r < 2\lambda r$. Thus by applying Equation (4.10), we have $M > 0$. By comparing L^{UE} with L^{SUE} in Equations (4.5) and (4.7), we have $\varphi < 1$. To prove the last case, since $\min(r, 2\lambda r) \leq 2\lambda r$, by Theorem 4.10 and Lemma 4.9 (ii), accompanying with that SUE reduces to UE when $\theta \to +\infty$, we obtain that $\varphi < 1$ when $\theta \geq \hat{\theta}$. On the other hand, when $0 \leq \theta < \hat{\theta}$, because $\rho^{UE} > 1$, $\rho^{SUE}|_{\hat{\theta}} = 1$, and $\partial\rho^{SUE}/\partial\theta < 0$ for $\theta < \hat{\theta}$, we conclude that there exists $\bar{\theta} \in [0, \hat{\theta})$, such that for any $\theta > \bar{\theta}$, $\varphi < 1$ holds; and for any $\theta \leq \bar{\theta}$, $\varphi \geq 1$ holds. This completes the proof. □

Considering the proof in Theorem 4.11, if $\rho^{SUE}|_{\theta=0} \geq \rho^{UE}$, we can further obtain that $\bar{\theta} > 0$. Otherwise $\bar{\theta} = 0$.

Theorem 4.11 classifies the domain of total flow amount into several subsets based on the relation of φ and 1. When $\varphi > 1$, UE achieves a better outcome of total travel time and vise versa. Next we only consider the case $\varphi < 1$ to study when and by how much SUE outperforms UE, which reduces the domain of D to $(\lambda r, 2r)$. It is worth mentioning that, the domain of D satisfying $\varphi < 1$ is the one that covers the cross point r. When the total flow amount is in the proximity of the cross point, the difference of travel time between e_1 and e_2 is small. In this circumstance, the network benefits from the split of all users choosing e_1 and e_2 other than all the users taking the same route.

Theorem 4.12. *Given an instance $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$ with $\lambda \in (0, 0.5)$.*

(i) *If $l_2'(D/2) \leq 2l_2'(D)$ holds for all $D \in [2\lambda r, r]$, then $\varphi \geq 1/2 + l_2(r/2)/(2c)$, and "=" is achieved when $D = r$ at $\theta = 0$.*

(ii) *If $l_2'(D/2) \geq 4l_2'(D)$ holds for all $D \in [2\lambda r, r]$, then $\varphi \geq (c + l_2(\lambda r))/(2l_2(2\lambda r))$, and "=" is achieved when $D = 2\lambda r$ at $\theta = 0$.*

(iii) *Otherwise there exist an $D_0 \in [2\lambda r, r]$, such that φ achieves its minimum when $D = D_0$ at $\theta = 0$.*

Proof. We divide the region $(\lambda r, 2r)$ into three intervals with overlapping boundaries: $(\lambda r, 2\lambda r]$, $[2\lambda r, r]$ and $[r, 2r)$.

When $D \in (\lambda r, 2\lambda r]$, it follows from Theorem 4.10 that $\min_\theta \rho^{SUE} = 1$. Thus $\varphi \geq 1/\rho^{UE} = L^{SO}/L^{UE}$. Since L^{SO}/L^{UE} is decreasing in D when $D > \lambda r$ holds, we obtain that L^{SO}/L^{UE} reaches its minimum when $D = 2\lambda r$. Taking $\rho^{SUE}|_{\theta=0,D=2\lambda r} = 1$ into consideration, we obtain that φ achieves its minimum when $D = 2\lambda r$ at $\theta = 0$.

When $D \in [2\lambda r, r]$, by Lemma 4.9 (iii), the minimum of φ must be achieved on the boundary of $\theta = 0$. Thus we consider $\varphi|_{\theta=0} = L^{SUE}_{\theta=0}/L^{UE}$. If $l_2'(D/2) \leq 2l_2'(D)$ holds for all $D \in [2\lambda r, r]$, we have that $L^{SUE}_{\theta=0}/L^{UE}$ is decreasing in D and reaches its minimum when $D = r$. Thus φ reaches its minimum when $D = r$ at $\theta = 0$. If $l_2'(D/2) \geq 4l_2'(D)$ holds for all $D \in [2\lambda r, r]$, then $L^{SUE}_{\theta=0}/L^{UE}$ is increasing in D and reaches its minimum when $D = 2\lambda r$, and φ reaches its minimum when $D = 2\lambda r$ at $\theta = 0$. Otherwise, since $[2\lambda r, r]$ is a compact set, there exists an $D_0 \in [2\lambda r, r]$ such that $L^{SUE}_{\theta=0}/L^{UE}$ reaches its minimum when $D = D_0$, and φ reaches its minimum when $D = D_0$ at $\theta = 0$.

When $D \in [r, 2r)$, we obtain that φ reaches its minimum when $D = r$ at $\theta = 0$ similarly with that in the case $D \in [2\lambda r, r]$.

By comparing the three minimum values in their corresponding intervals, we find the minimum value of φ for $D \in (\lambda r, 2r)$ and complete the proof of Theorem 4.12. ☐

Theorem 4.13. *Given an instance $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$ with $\lambda \in [0.5, 1)$, the relative performance ratio $\varphi \geq 1 - \lambda(1 - l_2(\lambda r)/c)$, and "=" is achieved when $D = r$ at $\theta = \hat{\theta}$.*

Proof. When $\lambda \in [0.5, 1)$, we divide $(\lambda r, 2r)$ into two intervals $(\lambda r, r]$ and $[r, 2r)$. Theorem 4.13 can be proved by using similar approach as in the proof of Theorem 4.12. ☐

Theorems 4.12 and 4.13 point out the minimum value of φ in the two-link network as well as the conditions for φ to reach its lower bound. In Theorem 4.12, it is interesting to find that φ reaches its minimum when $\theta = 0$, which means that the network performs best even when all the users randomly choose any of their routes with probability 0.5.

When the travel time function $l_2(x)$ is assumed to be a convex function, we can further get the corollary as below.

Corollary 4.14. *Given an instance $(\hat{\mathcal{G}}_2, D, \boldsymbol{l})$ with $l_2(x)$ being a convex function, the relative performance ratio $\varphi \geq 1/2 + l_2(r/2)/(2c)$ if $\lambda \in (0, 0.5)$ and $\varphi \geq 1 - \lambda(1 - l_2(\lambda r)/c)$ if $\lambda \in [0.5, 1)$, and "=" is achieved when $D = r$.*

Proof. When $l_2(x)$ is a convex function, the assumption of Theorem 4.12 (i) is satisfied. Thus Corollary 4.14 is achieved by combining Theorems 4.12 and 4.13. □

Furthermore, if we compare the upper bound of UE given by Roughgarden[12] with the minimum value of φ in Corollary4.14, we find that they are exactly the same. That is to say, the lower bound of φ as in Lemma 3.2 is tight. Thus we have Corollary 4.15.

Corollary 4.15. *Let $\mathcal{L}$ be a strictly standard class of travel time functions containing the constant functions. If $\mathcal{J}_2$ denotes the set of all single-commodity instances with underlying network $\hat{\mathcal{G}}_2$ and travel time functions in $\mathcal{L}$, then*

$$\inf_{(\hat{\mathcal{G}}_2, D, \mathbf{l}) \in \mathcal{J}_2} \varphi = 1 - \gamma(\mathcal{L}). \tag{4.13}$$

At the end of this section, we restrict $l_2(x)$ to some widely used travel time functions and calculate their corresponding minimum relative performance ratio φ, denoted as φ_0. Suppose $l_2(x)$ is a polynomial function. Let $l_2(x) = ax^m + b$, with $m \in \mathcal{R}^+$, $a > 0$ and $c > b \geq 0$, where c is the travel time function of link e_1. We find that $\lambda = (1+m)^{-1/m}$. Thus $\lambda \in [0.5, 1)$. According to Theorem 4.13, we have $\varphi_0 = 1 - m(1+m)^{-(1+1/m)}(1-b/c)$. When $b = 0$, φ_0 is deduced to $1 - m(1+m)^{-(1+1/m)}$. Furthermore, we obtain that φ_0 decreases when m increases, and $\lim_{m \to +\infty} \varphi_0 = 0$. In particular, when $m = 1$, $l_2(x)$ is a linear function. We have $\varphi_0 = 0.75 + 0.25b/c$. When $m = 4$, $l_2(x)$ is a BPR (Bureau of Public Roads) type function, where b is the free flow travel time on link e_2. We have $\varphi_0 = 0.465 + 0.535b/c$.

5 Extension

In this section, we generalize the special two-link network analysed in Section 4 to networks that consist of $m+n$ parallel links. The network contains a vertex set $\mathcal{V} = \{s, t\}$ and a directed edge set $\mathcal{E} = \{e_i | i = 1, \ldots m+n\}$, in which all the edges are parallel routes connecting s and t. The network flow is $\mathbf{f} = (\ldots, f_i, \ldots)$, and the total flow amount is $D = \sum_{i=1}^{m+n} f_i$. Each travel time function $l_i(x)$ associated with edge e_i is still assumed to be nonnegative, strictly increasing and strictly standard as we did in Section 4. Furthermore, we assume that there are only two different travel time functions $l_1(x)$ and $l_2(x)$ in the network. Without loss of generality, we denote the travel time functions on edges e_1 to e_m as $l_1^i(x), i \in \{1, \ldots, m\}$ with $l_1^i(x) = l_1(x)$, and the travel time functions on edges e_{m+1} to e_{m+n} as $l_2^j(x), j \in \{m+1, \ldots, m+n\}$ with $l_2^j(x) = l_2(x)$. A cross point of $l_1^i(x)$ and $l_2^j(x)$ is redefined as:

Definition 5.1. For any $c \in \mathbb{R}_+$, if there exists $r \in (0, +\infty)$ such that $l_1^i(r) = l_2^j(r) = c$, then r is a *cross point* of $l_1^i(x)$ and $l_2^j(x)$.

Note that $l_1^i(x)$ and $l_2^j(x)$ may have more than one cross point. We further denote the set of all cross points of $l_1^i(x)$ and $l_2^j(x)$ as $\mathcal{R}$. Clearly $\mathcal{R}$ is a closed set. It may contain $+\infty$, or consist of several subsets. When there is no cross point for $l_1^i(x)$ and $l_2^j(x)$, $\mathcal{R}$ is an empty set. For convenience, the origin 0 is denoted as r_0. Finally we denote an instance of the above network as $(\hat{\mathcal{G}}_{m+n}, D, \mathbf{l})$. A description of such networks is given in Figure 2.

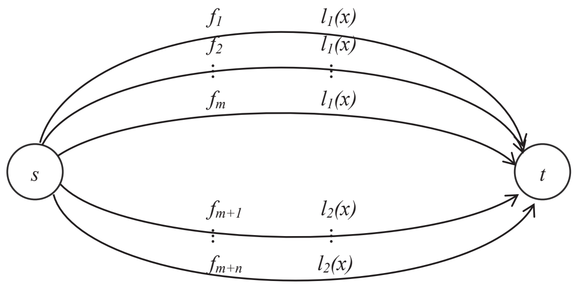

Figure 2: Description of networks with parallel links.

For an instance $(\hat{\mathcal{G}}_{m+n}, D, \mathbf{l})$, according to Definition 2.1, the UE flow pattern $\mathbf{f}^{UE}$ has the form

$$\mathbf{f}^{UE} = (\underbrace{f_1^{UE}, \ldots, f_1^{UE}}_{m}, \underbrace{f_2^{UE}, \ldots, f_2^{UE}}_{n}), \tag{5.1}$$

with $D = mf_1^{UE} + nf_2^{UE}$ and $l_1^i(f_1^{UE}) = l_2^j(f_2^{UE})$. Next we claim that the SUE of the instance $(\hat{\mathcal{G}}_{m+n}, D, \mathbf{l})$ satisfies the following lemma.

Lemma 5.2. *For an instance $(\hat{\mathcal{G}}_{m+n}, D, \boldsymbol{l})$, its SUE has the following form:*

$$\begin{aligned} q_1, \ldots, q_m &= \frac{1}{m + n\exp(\theta M)}, \quad \textit{and} \\ q_{m+1}, \ldots, q_{m+n} &= \frac{1}{n + m\exp(-\theta M)}, \end{aligned} \tag{5.2}$$

where $M = l_1(q_1 D) - l_2(q_{m+1} D)$.

Proof. Since a flow pattern $\mathbf{f} = (q_1 D, \ldots, q_m D, q_{m+1} D, \ldots, q_{m+n} D)$ satisfies Equation (2.7) in Definition 2.5, it is an SUE. On the other hand, by Fisk [4], we know the SUE for a congested network routing game is unique. This ends our proof. □

In the rest of this section, let f_1^{UE}, f_{m+1}^{UE}, f_1^{SUE} and f_{m+1}^{SUE} represent the UE flows and SUE flows on edges e_1 to e_m and e_{m+1} to e_{m+n} respectively. And we use $l_1(x)$ and $l_2(x)$ to denote $l_1^i(x)$ and $l_2^j(x)$ for simplicity. Now we present some properties of SUE and analyse the relative performance ratio φ in such networks.

Lemma 5.3. *For an instance $(\hat{\mathcal{G}}_{m+n}, D, \boldsymbol{l})$, $q_{m+1} = 1/(m+n)$ if and only if either $\theta = 0$ or $D = (m+n)r$ for some $r \in \mathcal{R}$.*

Proof. Since q_{m+1} is the solution to Equation (5.2), it is obvious that $q_{m+1} = 1/(m+n)$ if and only if $\theta \cdot M = 0$, which is equivalent to $\theta = 0$, or $M = 0$. Note that if $\mathcal{R}$ is empty, $M = 0$ has no solution. Thus we focus on the cases where $\mathcal{R}$ is not empty.

We show the equivalence of $M = 0$ and $D = (m+n)r$ for some $r \in \mathcal{R}$. If $M = 0$ holds, following by $q_{m+1} = 1/(m+n)$, we obtain that there exist an $r \in \mathcal{R}$, such that $D = (m+n)r$ holds. In the contrary, for a given r, if $D = (m+n)r$ holds, we show that $M \neq 0$ is impossible. Suppose $M > 0$, From Equation (5.2), we have $q_1 < 1/(m+n)$ and $q_{m+1} > 1/(m+n)$. Given that $l_1(x)$ and $l_2(x)$ are strictly increasing, we have $l_1(q_1 D) < l_1(r) = l_2(r) < l_2(q_{m+1} D)$. This is a contrary. The proof of the case $M < 0$ is exactly the same. This completes our proof. □

Lemma 5.4. *For an instance* $(\hat{\mathcal{G}}_{m+n}, D, \boldsymbol{l})$ *where* $D/(m+n) \notin \mathcal{R}$*, we can find two adjacent cross points* $r_k < r_{k+1}, r_k, r_{k+1} \in \mathcal{R}$*, such that for all* $D \in ((m+n)r_k, (m+n)r_{k+1})$*,*

(i) *if* $f_1^{UE} < f_{m+1}^{UE}$*, then for all* $\theta > 0$*, we have* $l_1(f_1^{SUE}) > l_2(f_{m+1}^{SUE})$*; and*

(ii) *if* $f_1^{UE} > f_{m+1}^{UE}$*, then for all* $\theta > 0$*, we have* $l_1(f_1^{SUE}) < l_2(f_{m+1}^{SUE})$*.*

Proof. Assume that $l_1(0) \geq l_2(0)$. We only prove the first item. The idea for proving the second item is the same, so we omit it here. When $f_1^{UE} = 0$ and $f_{m+1}^{UE} > 0$, since $f_1^{SUE} > 0$, we obtain that $l_1(f_1^{SUE}) > l_1(0) \geq l_2(f_{m+1}^{UE}) > l_2(f_{m+1}^{SUE})$ immediately. Otherwise, we have $f_1^{UE} > 0$ and $f_{m+1}^{UE} > 0$. Let $c = l_1(f_1^{UE}) = l_2(f_{m+1}^{UE})$. Since $l_1(D/(m+n)) \neq l_2(D/(m+n))$, obviously $l_1(f_1^{SUE}) = l_2(f_{m+1}^{SUE})$ happens only if $\theta = 0$. Suppose $l_1(f_1^{SUE}) < l_2(f_{m+1}^{SUE})$, from Equation (5.2) we have $q_1 > 1/(m+n)$, thus $l_1(f_1^{SUE}) \geq l_1(D/(m+n)) > c > l_2(D/(m+n)) \geq l_2(f_{m+1}^{SUE})$. This is a contradiction, which completes our proof. □

Lemma 5.5. *For an instance* $(\hat{\mathcal{G}}_{m+n}, D, \boldsymbol{l})$ *where* $D/(m+n) \notin \mathcal{R}$*, we can find two adjacent cross points* $r_k < r_{k+1}, r_k, r_{k+1} \in \mathcal{R}$*, such that for all* $D \in ((m+n)r_k, (m+n)r_{k+1})$*, SUE satisfies the following properties:*

(i) *if* $f_1^{UE} < f_{m+1}^{UE}$*, then for all* $\theta > 0$*, we have* $\partial q_1/\partial\theta < 0$ *with* $f_1^{UE} < f_1^{SUE} \leq D/(m+n)$*, and* $\partial q_{m+1}/\partial\theta > 0$ *with* $f_{m+1}^{UE} > f_{m+1}^{SUE} \geq D/(m+n)$*;*

(ii) *if* $f_1^{UE} > f_{m+1}^{UE}$*, then for all* $\theta > 0$*, we have* $\partial q_1/\partial\theta > 0$ *with* $f_1^{UE} > f_1^{SUE} \geq D/(m+n)$*, and* $\partial q_{m+1}/\partial\theta < 0$ *with* $f_{m+1}^{UE} < f_{m+1}^{SUE} \geq D/(m+n)$*; and*

(iii) q_i*,* $i \in \{1, \ldots, m+n\}$ *is continuous at* $\theta = 0$*.*

Proof. We rewrite the first equation of Equation (5.2) as

$$F(q_1, \theta, D) = q_1 - \frac{1}{m + n\exp(\theta M)} = 0. \tag{5.3}$$

When $\theta \in (0, +\infty)$, by using the implicit function theorem, we rewrite $\partial q_1/\partial\theta$ as $\partial q_1 = -(\partial F/\partial q_1)/(\partial F/\partial\theta)$. Taking partial derivative of F in (5.3) with respect to q_1, we obtain that $\partial F/\partial q_1 > 0$. From Lemma 5.4, we obtain that, when $f_1^{UE} < f_{m+1}^{UE}$, we have $M > 0$ holds. Given the partial derivative of F with respect to θ as $\partial F/\partial\theta = \frac{nM\exp(\theta M)}{[m+n\exp(\theta M)]^2}$, we have $\partial F/\partial\theta > 0$. This leads to $\partial q_1/\partial\theta < 0$. Since $f_1^{SUE}|_{\theta=0} = D/(m+n)$ and $f_1^{SUE}|_{\theta\to+\infty} \to f_1^{UE}$, the front half part of the first item is proved. By using the same technique, we can prove the second half part of the first item as well as the second item.

Noticing that M is bounded when D is fixed, we have $\lim_{\theta\to 0^+} q_i = q_i|_{\theta=0} = 1/(m+n)$. This completes the proof of Lemma 5.5. □

Note that the value r_{k+1} may not exist, i.e., $r_{k+1} = +\infty$. however, it does not change the conclusion in Lemmas 5.4 and 5.5. We illustrate the ideas of Lemmas 5.4 and 5.5 by Figure 3. Given a fixed demand D, the corresponding UE flow pattern can be calculated by UEP, and each route encounters a travel time c. Lemma 5.4 reveals that, The SUE flow pattern has an effect that it pulls all the UE flows toward their middle point (the average flow amount $D/(m+n)$). Lemma 5.5 further reveals that the larger the users' perception error is, the closer the SUE flows on different routes are. However, the SUE flow on a route corresponding to a small UE flow is always no more than the average, while those corresponding to large UE flows are always no less than the average.

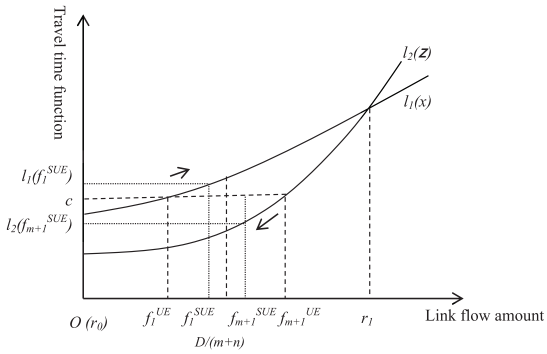

Figure 3: Description of Lemmas 5.4 and 5.5.

Suppose the UE travel time cost of a network $(\hat{\mathcal{G}}_{m+n}, D, \mathbf{l})$ is c, to study the properties of φ in this network, we modify the presentation of φ to:

$$\varphi = L^{SUE}/L^{UE} = 1 + (L^{SUE} - L^{UE})/L^{UE}. \tag{5.4}$$

In this case, the conditions for $\varphi < 1$ is equivalent to that for $L^{SUE} - L^{UE} < 0$. We further derive $L^{SUE} - L^{UE}$ as

$$\begin{aligned} L^{SUE} - L^{UE} &= mf_1^{SUE} l_1(f_1^{SUE}) + nf_{m+1}^{SUE} l_2(f_{m+1}^{SUE}) - c(\sum_{i=1}^{m+n} f_i^{UE}) \\ &= mf_1^{SUE} l_1(f_1^{SUE}) + nf_{m+1}^{SUE} l_2(f_{m+1}^{SUE}) - c(\sum_{i=1}^{m+n} f_i^{SUE}) \\ &= mf_1^{SUE}(l_1(f_1^{SUE}) - c) + nf_{m+1}^{SUE}(l_2(f_{m+1}^{SUE}) - c). \end{aligned} \tag{5.5}$$

According to Lemma 5.5, it is clear that $(l_1(f_1^{SUE}) - c) \cdot (l_2(f_{m+1}^{SUE}) - c) \leq 0$, and " = " is achieved only when $\theta \to +\infty$. Thus the relationship of φ and 1 is decided by the trade off of the two groups of area. Those above the c line have a positive effect, as well as those under the c line have a negative effect. $\varphi < 1$ holds only when the summation of the negative area is larger. From Lemma 5.3 we immediately obtain a sufficient condition under which $\varphi = 1$.

Lemma 5.6. *For an instance $(\hat{\mathcal{G}}_{m+n}, D, \boldsymbol{l})$, if there exist an $r \in \mathcal{R}$, such that the total flow amount $D = (m+n)r$ holds, then we have $\varphi = 1$.*

Proof. The lemma holds since all the link flows of UE and SUE are equal to $D/(m+n)$ in such circumstances. □

In most of the cases, it is difficult to find the conditions under which φ is less than 1. In the following lemma along with its corollary, we present two sufficient conditions with more specific assumptions on the network structure and the travel time functions.

Lemma 5.7. *For an instance $(\hat{\mathcal{G}}_{m+n}, D, \boldsymbol{l})$ with concave travel time functions, when $m \leq n$, $0 < f_1^{UE} < f_{m+1}^{UE}$ and $l_1'(f_1^{UE}) \leq l_2'(f_{m+1}^{UE})$ holds, then for all θ, the relative performance ratio φ is less than 1.*

Proof. Since the travel time functions are concave, we have $l_1(f_1^{SUE}) \leq l_1(f_1^{UE}) + l_1'(f_1^{UE})(f_1^{SUE} - f_1^{UE})$ and $l_2(f_{m+1}^{SUE}) \leq l_2(f_{m+1}^{UE}) + l_2'(f_{m+1}^{UE})(f_{m+1}^{SUE} - f_{m+1}^{UE})$. By applying them to Equation (5.5), we obtain that $L^{SUE} - L^{UE} < 0$. The assumptions $m \leq n$ and $0 < f_1^{UE} < f_{m+1}^{UE}$ completes our proof of the lemma. □

Corollary 5.8. *For an instance $(\hat{\mathcal{G}}_{m+n}, D, \boldsymbol{l})$ with linear travel time functions, when $m \leq n$, $0 < f_1^{UE} < f_{m+1}^{UE}$ and $l_1^{'}(f_1^{UE}) \leq l_2^{'}(f_{m+1}^{UE})$ holds, then for all θ, the relative performance ratio φ is less than 1.*

Remark 5.9. Note that in both Lemma 5.7 and Corollary 5.8 we assume $f_1^{UE} > 0$. This assumption is very critical and can not be removed. If $f_1^{UE} = 0$, c is possible to be less than $l_1(0)$. In this case, $l_1(f_1^{SUE}) - c$ could be larger than $l_2(f_{m+1}^{SUE}) - c$. Even if $m \leq n$ holds, we can not guaranty that $L^{SUE} - L^{UE} < 0$ holds.

At the end of this section, we use the same method as in Roughgarden [12], and verify the tightness of the lower bound of φ in networks with parallel routes. First we consider networks with single OD pair and parallel links. Set $n = 1$, we build an instance $(\hat{\mathcal{G}}_{m+1}, D, \mathbf{l})$ and claim that the lower bound of φ is achieved in such networks. Then we generalize the case to allow the networks contain disjoint paths. Our results indicates that, for those network topologies which admit the worst case of ρ^{UE} in Roughgarden [12], they also admit the lower bound of φ.

Definition 5.10. (Roughgarden [12]) A class $\mathcal{L}$ of travel time functions is diverse if for each $c > 0$, there is a travel time function $l \in \mathcal{L}$ satisfying $l(0) = c$.

Theorem 5.11. *Let $\mathcal{L}$ be a strictly standard and diverse class of travel time functions. If $\mathcal{J}_{m+1}$ denotes the set of all single-commodity instances with underlying network $\hat{\mathcal{G}}_{m+1}$ and travel time functions in $\mathcal{L}$, then*

$$\inf_{(\hat{\mathcal{G}}_{m+1}, D, \boldsymbol{l}) \in \cup_m \mathcal{J}_{m+1}} \varphi = 1 - \gamma(\mathcal{L}). \tag{5.6}$$

Proof. For any $\varepsilon > 0$, choose a nonzero travel time function $l_2(x) \in \mathcal{L}$, a positive number $r > 0$ with $l_2(r) > 0$ and its associated $\lambda \in [0, 1]$ satisfying Equation (4.1) such that $\lambda \frac{l_2(\lambda r)}{l_2(r)} + (1 - \lambda) \leq 1 - \gamma(\mathcal{L}) + \varepsilon/2$. It follows from Lemma 4.2 that given $l_2(x)$ and r there always exists a unique λ.

From Section 4 we know that these parameter choices correspond to an instance $(\hat{\mathcal{G}}_2, D, \mathbf{l})$ with $D = r$, $l_1(x) = c$ and $l_2(x)$, and φ is at most $1 - \gamma(\mathcal{L}) + \varepsilon/2$. Since an instance $(\hat{\mathcal{G}}_{m+1}, D, \mathbf{l})$ does not contain constant latency functions, we will transform this instance into $(\hat{\mathcal{G}}_{m+1}, D, \mathbf{l})$ by simulating $l_1(x) = c$ with parallel links.

Given $\mathcal{L}$ is diverse, there exists a function $l(x) \in \mathcal{L}$ such that $l(0) = l_2(r)$. We set the first m links with travel time function $l(x)$, and the last link as $l_2(x)$. Let m be sufficiently large such that $1/(m + 1) < \lambda$ and $l((1 - \lambda)r/m) \leq l_2(r) + \delta$ holds, where δ is a sufficiently small positive number (depending on ε) to be chosen later. Clearly, the total travel time incurred by the UE flow is $rl_2(r)$ with all flow using $l_2(x)$. Thus we also have $f_1^{UE} < f_{m+1}^{UE}$. By applying Lemmas 5.4 and 5.5, there exist a $\theta < +\infty$ (depending on m) such that $f_1^{SUE} = (1 - \lambda)r/m$ and $f_{m+1}^{SUE} = \lambda r$. The corresponding total travel time $L^{SUE} \leq mf_1^{SUE}(l_2(r)+\delta)+\lambda rl_2(\lambda r) = rl_2(r)[\lambda \frac{l_2(\lambda r)}{l_2(r)}+(1-\lambda)+(1-\lambda)\delta/l_2(r)]$. Choosing δ sufficiently small, we obtain an instance with φ at most $1 - \gamma(\mathcal{L}) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we have $\inf_{(\hat{\mathcal{G}}_{m+1}, D, \mathbf{l}) \in \cup_m \mathcal{J}_{m+1}} \varphi \leq 1 - \gamma(\mathcal{L})$. Together with Lemma 3.2, our proof completes. □

Next we prove that in a network with parallel routes, the lower bound of φ can also be achieved. A description of the network is given in Figure 4. The same as in Roughgarden [12], we replace the assumption of diversity with weaker condition that some available latency function is positive when its flow amount is 0. From the definition given below, it is clear that a diverse class of travel time functions is inhomogeneous. Thus the assumption is a relaxation.

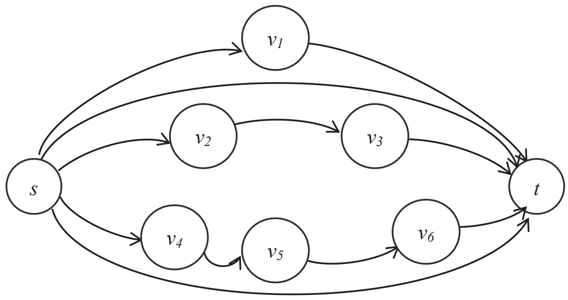

Figure 4: Description of networks with parallel routes.

Definition 5.12. (Roughgarden [12]) A class $\mathcal{L}$ of travel time functions is homogeneous if $l(0) = 0$ for all $l(x) \in \mathcal{L}$ and inhomogeneous otherwise.

Theorem 5.13. *Let $\mathcal{L}$ be a strictly standard and inhomogeneous class of travel time functions. If $\mathcal{J}_u$ denotes the set of all single-commodity instances with underlying network a union of paths and travel time functions in $\mathcal{L}$, then*

$$\inf_{(\mathcal{G},D,\mathbf{l})\in\mathcal{J}_u} \varphi = 1 - \gamma(\mathcal{L}). \tag{5.7}$$

Proof. The idea is to work with a larger class of travel time functions and then argue that any travel time function in the class can be simulated with a collection of edges all possessing travel time functions in $\mathcal{L}$.

Let $\bar{\mathcal{L}} = \{\beta l(x) : l(x) \in \mathcal{L}, \beta > 0\}$ denote the closure of $\mathcal{L}$ under multiplication by positive scalars. Clearly $\bar{\mathcal{L}}$ is strictly standard and diverse. Since for each $l(x) \in \mathcal{L}$ and $\beta > 0$, $\gamma(l) = \gamma(\beta l)$, thus we have $\gamma(\bar{\mathcal{L}}) = \gamma(\mathcal{L})$. For any $\varepsilon > 0$, from Theorem 5.11 we acquire an instance $(\bar{\mathcal{G}}_{m+1}, \bar{D}, \bar{\mathbf{l}})$ on a network $\bar{\mathcal{G}}_{m+1}$ of parallel link functions in $\bar{\mathcal{L}}$ and φ value at most $\gamma(\mathcal{L}) + \varepsilon/2$. For each link e of $\bar{\mathcal{G}}_{m+1}$, the ratio φ is continuous in β_e. So we may replace each β_e by a sufficiently close positive rational number β'_e to obtain a new instance with φ value at most $\gamma(\mathcal{L}) + \varepsilon$. For convenience, we assume that β'_e are integers.

Define $\mathcal{G}$ by replacing each link e of $\bar{\mathcal{G}}_{m+1}$ by a directed path of β'_e new edges, each endowed with travel time function $l_e(x)$. Clearly, the SUE flow pattern and UE flow pattern as well as the travel time on each route remains the same. In this way, we construct a new instance in $(\mathcal{G}, D, \mathbf{l})$ with φ value at most $\gamma(\mathcal{L}) + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we have $\inf_{(\mathcal{G},D,\mathbf{l})\in\mathcal{J}_u} \varphi \le 1 - \gamma(\mathcal{L})$. Together with Lemma 3.2, our proof completes. □

6 Conclusions

This paper defines the relative performance ratio of a network as the ratio of PoA of SUE to PoA of UE. We identify that the relative performance ratio has an upper bound which equals to PoA of SUE and a lower bound which equals to the inverse of PoA of UE. The lower bound is further proved to be tight in some special networks. In a two-link network which is a generalization of Pigou's example, we study the properties of SUE and point out circumstances where the outcome of SUE equals to that of SO. Based on these properties, we further find the total flow regions in which SUE performs better than UE. We also point out conditions under which the relative performance ratio reaches its lower bound. It is interesting to discover that, the relative performance ratio is less than 1 when the total flow amount is near the cross point r. Under such circumstances, the network can be better off even if users choose their routes with equal probability. We further extend our study on relative performance ratio to networks with parallel routes, and we discover that the lower bound of relative performance ratio is still tight in these cases.

These exciting results motivate us to investigate the relative performance ratio in more complicated networks. It would be interesting to check if SUE can still outperform UE. Under such circumstances, we have concern that if they are still near some cross points. It is also attractive to study the upper bound of the relative performance ratio and pay more attention to the circumstances where UE outperforms SUE. Other behaviors of network users are also worth adding to the network analysis to depict the real world situations.

Acknowledgements

The authors thank the guest editors and three anonymous referees for their constructive feedback that helped to significantly improve this paper.

References

[1] J.R. Correa, A.S. Schulz and N.E. Stier-Moses, Selfish routing in capacitated networks, *Math. Oper. Res.* 29 (2004) 961–976.

[2] J. R. Correa, A.S. Schulz and N.E. Stier-Moses, On the inefficiency of equilibria in congestion games. Extended abstract. In: *Proceedings of the 11th Conference on Integer Programming and Combinatorial Opitimization, Lecture Notes in Computer Science*, Springer-Verlag, Berlin Heidelberg, 2005, pp. 167–181.

[3] S. Dafermos, Traffic equilibrium and variational inequalities, *Transp. Sci.* 14 (1980) 42–54.

[4] C. Fisk, Some developments in equilibrium traffic assignment, *Transp. Res. Pt. B-Methodol.* 14 (1980) 243–255.

[5] X. Guo, H. Yang and T.L. Liu, Bounding the inefficiency of logit-based stochastic user equilibrium, *European J. Oper. Res.* 201 (2010) 463–469.

[6] J. L. Horowitz, The stability of stochastic equilibrium in a two-link transportation network, *Transp. Res. Pt. B-Methodol.* 18 (1984) 13–28.

[7] E. Koutsoupias and C. H. Papadimitriou, Worst-case equilibria, in: *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, C. Meinel, S. Tison (Eds.), Springer-Verlag, Berlin Heidelberg, 1999, pp. 404–413.

[8] R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis*, John Wiley and Sons, Inc., New York, 1959.

[9] J. Morgan, H. Orzen and M. Sefton, Network architecture and traffic flows: Experiments on the Pigou-Knight-Downs and Braess Paradoxes, *Games Econom. Behav.* 66 (2009) 348–372.

[10] C.H. Papadimitriou, Algorithms, games, and the Internet, in: *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, ACM Press, New York, 2001, pp. 749–753.

[11] A. C. Pigou, *The Economics of Welfare*, Macmillan and Co., London, 1920.

[12] T. Roughgarden, The price of anarchy is independent of the network topology, *J. Comput. Syst. Sci.* 67 (2003) 341–364.

[13] T. Roughgarden and É. Tardos, Bounding the inefficiency of equilibria in nonatomic congestion games, *Games Econom. Behav.* 47 (2004) 389–403.

[14] T. Roughgarden, *Selfish Routing and the Price of Anarchy*, MIT Press, Cambridge, 2005.

[15] R. Selten, T. Chmura, T. Pitz, S. Kube and M. Schreckenberg, Commuters route choice behavior, *Games Econom. Behav.* 58 (2007) 394–406.

[16] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1985.

Manuscript received 30 September 2013
revised 25 March 2014
accepted for publication 26 April 2014

Jing Jiao
Department of Mathematical Sciences, Tsinghua University
Beijing 100084, P.R. China
E-mail address: jiaoj05@mails.tsinghua.edu.cn

Jinxing Xie
Department of Mathematical Sciences, Tsinghua University
Beijing 100084, P.R. China
E-mail address: jxie@math.tsinghua.edu.cn