Submitted to *Operations Research*
manuscript (Please, provide the manuscript number!)

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

**Sefishness need not be bad**

Zijun Wu  
Beijing Institute for Scientific and Engineering Computing (BISEC), College of Applied Sciences (CAS), Beijing University of Technology (BJUT), Pingleyuan 100, 100124, Beijing, P. R. China, and  
Institute for Applied Operations Research (IAOR), Department of Computer Science (DCS), Hefei University (HU), Jingxiu Road 99, 230091, Hefei, Anhui, P. R. China, zijunwu1984a@163.com  

Rolf H. Möhring  
Beijing Institute for Scientific and Engineering Computing (BISEC), College of Applied Sciences (CAS), Beijing University of Technology (BJUT), Pingleyuan 100, 100124, Beijing, P. R. China, rolf.moehring@me.com  

Yanyan Chen  
College of Metropolitan Transportation (CMT), Beijing University of Technology (BJUT), Pingleyuan 100, 100124, Beijing, P. R. China, cdyan@bjut.edu.cn
We devote this article to a study of the user behavior in non-atomic congestion games. Our study will focus on a particular class of non-atomic congestion games, called scalable games, which includes the well-known static model of road traffic from Wardrop \cite{Wardrop1952}.

We prove that the price of anarchy of scalable games tends to 1 as the number of users increases, see Theorem 3.2. This may imply that selfish routing need not be bad. Moreover, we prove that gaugeable routing games that were recently studied by Colini-Baldeschi et al. \cite{ColiniBaldeschi2017} are special cases of our scalable games, see Corollary 3.1, Theorem 3.3 and Theorem 3.4 for details. Thus, our results generalize those of Colini-Baldeschi et al. \cite{ColiniBaldeschi2017}. Although our results are much more general, the proofs appear simpler.

For routing games with BPR-travel time functions, we prove that social optimum strategy profiles are \(\epsilon\)-approximate Nash equilibria for a small \(\epsilon > 0\) tending to 0 as travel demands increase, see Theorem 3.5. Moreover, we prove for these particular games that the price of anarchy equals \(1 + O(T^{-\beta})\), where \(T\) is the total travel demand, and \(\beta\) is the degree of the BRP-functions that often equals 4 in practice, see Theorem 3.6. This confirms a conjecture proposed by O’Hare et al. \cite{OHare2016}. Thus, the price of anarchy for road traffic in practice converges very fast to 1 as the total travel demand increases. In addition, we show that, when the total travel demand \(T\) is large, then the distribution of users among OD pairs is a crucial factor impacting the costs of both, system optimum strategy profiles and Nash equilibria, see Theorem 3.7. This does not only supply an approximate method for computing these cost, but also give insights how to reduce the total travel time, when the total travel demand \(T\) is large.

To empirically verify our theoretical findings, we have taken real traffic data within the 2nd ring road of Beijing as an instance in an experimental study. Our empirical results definitely validate our findings. In addition, they show that the current traffic in Beijing within that area is already far beyond saturation, and no route guidance policy can significantly reduce the total travel time for the current huge total travel demand.

In summary, selfishness in a congestion game with a large number of users need not be bad. It may be the best choice in a bad environment.

\textit{Key words:} price of anarchy, user behavior, selfish behavior, non-atomic congestion game, road traffic

\textit{History:} This paper was first submitted on ...
1. **Introduction**

1.1. **Traffic in Beijing**

Since several years, traffic congestion has almost become a daily annoyance for every citizen in Beijing. To alleviate traffic congestion, the local government has taken a series of measures, including the even and odd license plate number rule, license plate lottery, encouraging the use of public transportation, and others. These measures definitely have prevented further deterioration of traffic, but have not completely cured the congestion.

Since 2015, voices calling for congestion pricing continued to come up from the local society. Congestion pricing is a measure of traffic demand management, which aims to guide the routing behavior of travelers by collecting tolls on some overloaded roads, see Cole *et al.* (2003) or Fleischer *et al.* (2004). It has been implemented in cities like Singapore, London, Stockholm, Milan and others. The implementation in Singapore showed that it can considerably reduce traffic congestion in rush hours in the city center, see Phang and Toh (2004).

To check whether congestion pricing will help in reducing congestion in the center of Beijing, we launched a study in May 2016. We first computed the price of anarchy (PoA, see Roughgarden and Tardos (2002)), so as to get an overall impression of the current traffic performance within that area. The concept of PoA stems from Papadimitriou (2001) and measures the inefficiency of user selfish behavior in a congestion game (Rosenthal 1973), i.e., the larger the PoA, the more inefficient the user behavior is. Specific to road traffic, the PoA measures the inefficiency of selfish routing (Roughgarden and Tardos 2004). Surprisingly, the computational outcome showed that the PoA within that area almost equals 1, see Section 4 for details. This means that selfish route choices within that area are almost the most efficient ones, which certainly shocked us! We were thus eager for a theoretical explanation of this peculiar result, which has motivated this article. This has led to a closer inspection of the PoA of congestion games presented in this paper.

1.2. **Congestion games**

*Congestion games*, see Rosenthal (1973) or Roughgarden and Tardos (2004), are non-cooperative games. In a congestion game, users (players) are usually classified into $K$ different groups for some
integer $K \in \mathbb{N}_+$. Associated with each group $k \in \{1, \ldots, K\}$, there is a user volume $d_k \geq 0$ and a finite set $S_k$ containing all strategies only available to the users belonging to that group. Each user needs to determine a strategy to follow, and each strategy consumes certain amounts of resources. Available resources are collected in a finite set $A$, and each resource $a \in A$ has a consumption price function $\tau_a(\cdot)$ depending only on the consumed volume of resource $a$. A feasible strategy profile is a vector $f = (f_s)_{s \in S}$ in which each component $f_s \geq 0$ represents the volume (or number) of users choosing strategy $s$. The social cost $C(f)$ of a feasible strategy profile $f$ is just the total expense of all users under profile $f$. Our study will only consider non-atomic congestion games, see Schmeidler (1973) or Nisan et al. (2007), in which users are assumed to be infinitesimal, i.e., they have negligible ability to affect the others, and the consumption price functions $\tau_a(\cdot)$ of resources $a$ are assumed to be non-negative, non-decreasing and continuous.

Popular examples of non-atomic congestion games are the static traffic routing games of Wardrop (1952), in which $A$ is the set of arcs (streets) of an underlying road network, the $K$ groups are the $K$ different origin-destination pairs, $S_k$ is the set of paths connecting the $k$-th origin-destination pair, $d_k$ is the travel demand from the $k$-th origin to the $k$-th destination, $\tau_a(\cdot)$ is the travel time function on arc $a$, a feasible strategy profile $f$ is a just feasible traffic flow assigning every user (traveler) to a feasible path, and the social cost of $f$ is the total travel time of all users.

### 1.3. The price of anarchy

Users are often completely selfish, i.e., they tend to use strategies minimizing their own cost. This selfish behavior will lead the underlying system into a so-called Wardrop equilibrium, see Wardrop (1952), at which every user adopts a strategy with minimum cost for him. In our study, Wardrop equilibria coincide with pure Nash equilibria, at which users are unlikely to change their strategies since an unilateral change in strategy will not introduce any extra profit. Throughout this paper, we shall only consider pure Nash equilibria and just use the term Nash equilibria or NE for simplicity. By Roughgarden and Tardos (2002) and Correa et al. (2005), all Nash (or equivalently, Wardrop) equilibria have equal cost under our setting of non-decreasing and continuous consumption price
functions. Besides Nash equilibria, feasible profiles having minimum total (social) cost are of great interest. Such profiles are called *system optimum strategy profiles* and any two of them also have the same cost. In our study, the PoA is just the ratio of the cost of a Nash equilibrium over the cost of a system optimum strategy profile. System optimum strategy profiles can be thought of as the best choice the users should take so as to achieve a social optimum, while Nash equilibria model possible choices that users will take in practice. Thus, the PoA indeed reflects the inefficiency of the practical (selfish) user behavior to a certain extent.

It is almost folklore that selfish user behavior is inefficient in congestion games, see Nisan et al. (2007). A prominent example demonstrating this inefficiency is Pigous’s example, see Nisan et al. (2007) or Figure 1 in which the PoA can be made as large as possible. To deepen the understanding of the inefficiency, worst-case upper bounds of the PoA of congestion games with particular types of consumption prices functions $\tau_a(\cdot)$ have been studied in recent years. Roughgarden and Tardos (2002) proved that the worst-case upper bound equals $4/3$, if all $\tau_a(\cdot)$ are affine linear. Roughgarden (2003) proved that the worst-case upper bound of the PoA of routing games actually does not depend on the underlying road network topology, but on the types of travel time functions $\tau_a(\cdot)$. Roughgarden and Tardos (2004) further obtained worst-case upper bounds for the PoA for some general classes of consumption price functions $\tau_a(\cdot)$. Particularly, they proved that the worst-case upper bound equals $\Theta(\beta/\ln \beta)$ if all $\tau_a(\cdot)$ are polynomials with the same maximum degree $\beta > 0$. For more results on worst-case upper bounds, readers may refer to Roughgarden and Tardos (2007).

### 1.4. The inadequacy of the PoA for large demand

Our study will not consider worst-case upper bounds. Actually, a worst-case upper bound is not a fair measurement of the inefficiency. Figure 1 below shows Pigous’s example from Nisan et al. (2007). It is a routing game with only $K = 1$ origin-destination pair $(o, t)$ and two arcs with travel time functions $x^\beta$ and 1, respectively, for some constant $\beta > 0$. The PoA of this example equals $T/(T - (\beta + 1)^{-1/\beta} + (\beta + 1)^{-1})$, where $T \geq 1$ is the total travel demand from origin $o$ to destination $t$. Obviously, considering all possible $\beta$, the worst-case upper bound of the PoA is infinity, since it
Figure 1  Pigou’s example

tends to ∞ as β → ∞ if T = 1. However, if we consider a large enough T, then the PoA will be very close to 1 for every β > 0. This means that selfish user behavior is efficient when the total user volume T is large. Thus, selfish user behavior actually need not be bad.

Inspired by Pigou’s example, we aim to understand the aforementioned PoA for Beijing by inspecting conditions which guarantee that the PoA approaches 1 as the total user volume T = \( \sum_{k=1}^{K} d_k \) increases. To obtain a general result, our analysis will not stick to the Beijing instance, but consider general non-atomic congestion games.

1.5. Related work

Parallel studies have been recently done by Colini-Baldeschi et al. (2016) and Colini-Baldeschi et al. (2017). Colini-Baldeschi et al. (2016) proved that for routing games with a network consisting of parallel arcs linking a single origin-destination pair, the PoA converges to 1 as the total travel demand (i.e. total user volume) T → ∞.

Colini-Baldeschi et al. (2017) continued the study of Colini-Baldeschi et al. (2016), and analyzed the asymptotic behavior of the PoA for gaugeable routing games. These are routing games in which all travel time functions \( \tau_a(\cdot) \) are gaugeable by a regularly varying function \( g(\cdot) \), which means that \( \lim_{t \to \infty} \frac{\tau_a(t)}{g(t)} = c_a \) for some constant \( c_a \in [0, +\infty] \). See Bingham et al. (1987) for a definition of regular variation. Colini-Baldeschi et al. (2017) proved that, if the underlying game is gaugeable, then the PoA approaches 1 with \( T = \sum_{k=1}^{T} d_k \to \infty \), see Theorem 4.1 of Colini-Baldeschi et al. (2017). The study of Colini-Baldeschi et al. (2017) assumes that there is a uniform gauge function \( g(\cdot) \), and that the user volume vector \( (d_k)_{k=1,\ldots,K} \) keeps a certain pattern when the total user volume
\[ T = \sum_{k=1}^{K} d_k \] increases, i.e., the total user volume of so-called tight groups always accounts for a non-negligible proportion of \[ T = \sum_{k=1}^{K} d_k, \] see Colini-Baldeschi et al. (2017) for details. As we will see, these two assumptions limit the generality of their results.

### 1.6. Our results

Our study does not assume these restrictions. We propose a new concept called \textit{scalability}, and study two kinds of games related to scalability, i.e., the class of \textit{scalable games} and its subclass of \textit{strongly scalable games}, see Definition 3.1 and Definition 3.2, respectively. The concept of scalability allows us to analyze the convergence of the PoA in a sequence-dependent style, i.e., for each sequence \( \{d^{(n)}\}_{n \in \mathbb{N}} \) of user volume vectors with total user volume \( T(d^{(n)}) := \sum_{k=1}^{K} d^{(n)}_k \rightarrow +\infty \), we independently discuss the convergence of the PoA with the help of a scaling sequence of positive numbers that is specific to \( \{d^{(n)}\}_{n \in \mathbb{N}} \). Thus, we do not need a uniform gauge function \( g(\cdot) \) for all sequences of user volume vectors. Moreover, our scalability focuses on the existence of a suitable limit game, which will avoid additional assumptions on the structure of the user volume vectors.

We show that the PoA of each scalable game converges to 1 as the total user volume \( T = \sum_{k=1}^{K} d_k \rightarrow +\infty \), see Theorem 3.2, and that every gaugeable game is scalable, see Corollary 3.1. Moreover, we show that gaugeable games and strongly scalable games coincide if the consumption price functions \( \tau_a(\cdot) \) are mutually comparable, see Theorem 3.3. However, when the consumption price functions \( \tau_a(\cdot) \) are not mutually comparable, our class of strongly scalable games is more extensive, see Theorem 3.4. As a side result, we show by examples that scalable games are indeed more general than strongly scalable games, see Example 3.7.

In the study of Colini-Baldeschi et al. (2017), the regular variation of the gauge function \( g(\cdot) \) plays a pivotal role, and their proofs heavily depend on the properties of regularly varying functions. In our study, we do not need the regular variation. The existence of limit games makes our proof of Theorem 3.2 simpler compared to that of Theorem 4.1 in Colini-Baldeschi et al. (2017), although the result is more general.
Our study of general non-atomic congestion games implies that the PoA of each routing game with U.S. Bureau of Public Road (BPR) travel time functions $\tau_a(\cdot)$, see Bureau of Public Roads (1964), converges to 1 as the total travel demand $T$ increases. This definitely explains our empirical finding about the PoA for the Beijing instance. To deepen the understanding of the convergence, we inspect routing games with BPR-travel time functions $\tau_a(\cdot)$ more closely. We show that system optimum strategy profiles of these games are $\epsilon$-approximate Nash equilibria for a small $\epsilon \in O\left(\min\{d_1^{-\beta}, \ldots, d_K^{-\beta}\}\right)$, see Theorem 3.5, where $\beta > 0$ is a parameter of the BPR-functions which often equals 4 in practice. The concept of $\epsilon$-approximate Nash equilibrium is due to Roughgarden and Tardos (2002) and models practical user behavior more closely by taking possible imprecise judgements of users on tiny differences between travel times of two paths into account. In practice, users may not follow the best Nash equilibrium, but an $\epsilon$-approximate Nash equilibrium. Theorem 3.5 indicates that users will unknowingly, by their selfish behavior, follow the paths of system optimum strategy profiles when user volumes are large.

In addition, we prove that the PoA of these particular games equals $1 + O(T^{-\beta})$, see Theorem 3.6. This result proves a conjecture proposed by O’Hare et al. (2016) stating that the PoA of routing games with BPR travel time functions will with increasing total demand eventually enter a region of decay that can be characterized by a power law. So, the PoA will converge very quickly to 1, as the total volume $T$ increases. Our convergence rate $O(T^{-\beta})$ also improves the convergence rate of $O(T^{-1})$ shown in Colini-Baldeschi et al. (2017), since $\beta = 4$ in practice. Furthermore, we prove that the distribution of users among origin-destination pairs is a principal factor for the cost of both, Nash equilibria and system optimum strategy profiles when the total travel demand $T$ is large, see Theorem 3.7. This does not only supply an approximate method for computing these two cost values, but also brings some insight how to reduce the total travel time when the total travel demand $T$ is large, see the discussion following Theorem 3.7 in Subsection 3.2 for details.

Theorem 3.5, Theorem 3.6 and Theorem 3.7 actually indicate that congestion pricing has no effect when the total travel demand $T$ is large and not reduced. Generally, congestion pricing
concerns enforcing a certain traffic pattern (i.e., a traffic flow) through tolling streets (arcs), see Cole et al. (2003), Fleischer et al. (2004), and Harks et al. (2015). Preferred traffic patterns are often those at system optimum (Cole et al. 2003), or at constrained system optimum (Jahn et al. 2005). Our results show that when \( T \) is large, user behavior itself will naturally lead the system into a system optimum. Thus, we do not need to additionally employ congestion pricing to enforce the system optimum, when the travel demand \( T \) is already large. Constrained system optima (Jahn et al. 2005) aim at balancing total travel time and user fairness. Jahn et al. (2005) showed that the inherent unfairness of the system optimum can be reduced by restricting route choices while still improving the PoA. Unfairness means that the travel times of some users are significantly larger than their travel times at Nash equilibria. Nash equilibria can be thought of as the fairest traffic patterns, since all users follow the quickest path they could follow. According to our study, Nash equilibria themselves already perfectly balance the total travel time and user fairness when the total travel demand \( T \) is large enough. Therefore, we do not need to additionally employ congestion pricing to enforce other traffic patterns for the same demand, since they could only increase the total travel time.

To empirically verify our theoretical findings, we conducted a more detailed computational study with real traffic data within the second ring road of Beijing. The empirical results validate our theoretical findings. They show, in particular, that the current traffic in Beijing is already far beyond saturation, and any traffic guidance policy (particularly, congestion pricing) will fail in reducing the total travel time for the huge total travel demand of Beijing.

The remainder of this article is arranged as follows. Section 2 formalizes non-atomic congestion games. Section 3 reports our theoretical results, the proofs of which are collected in an Appendix so as to improve readability. Section 4 reports the empirical results from our experimental study of the Beijing instance. We conclude with a short summary in Section 5.
2. The model

2.1. Non-atomic congestion games

Our study considers non-atomic congestion games. Non-atomic congestion games are non-cooperative and complete information games, see Nisan et al. (2007), in which users need to choose strategies, and the utility of each user depends only on the volume (number) of users using the same or overlapping strategies. Definition 2.1 formally defines them.

Definition 2.1 A non-atomic congestion game (NCG) is a tuple

\[ \Gamma = \left( A, S, K, (r(a, s))_{a \in A, s \in S}, (\tau_a)_{a \in A}, (d_k^K)_{k=1} \right) \]

where

1. \( A \) is a finite set of available resources or public facilities,
2. \( K \in \mathbb{N}_+ \) represents the number of groups of users (players),
3. \( S = \bigcup_{k=1}^{K} S_k \) is a finite set of strategies, where each \( S_k \) contains strategies only available to users from group \( k \) for all \( k = 1, \ldots, K \),
4. each \( r(a, s) \geq 0 \) is a constant representing the volume of resource \( a \) demanded (consumed) by a user using strategy \( s \in S \), for all \( a \in A \) and \( s \in S \),
5. each \( \tau_a : [0, +\infty) \rightarrow [0, +\infty) \) is a continuous and nondecreasing consumption price function of resource \( a \), which depends only on the total demanded volume of resource \( a \), for all \( a \in A \),
6. \( d := (d_k)_{k=1, \ldots, K} \) is a user volume vector (also called user demand vector) such that each component \( d_k \geq 0 \) represents the volume (or demand) of users belonging to group \( k \), for all \( k = 1, \ldots, K \).

Our definition of NCGs is slightly different from the classic one in the literature, see e.g. Roughgarden and Tardos (2002), Roughgarden and Tardos (2004), and Bingham et al. (1987). Traditionally, every strategy \( s \in S \) is assumed to be a set of resources from \( A \). In our study, we do not assume this. The constants \( r(a, s) \) are employed to indicate the relation between resources \( a \in A \) and strategies \( s \in S \). Here, we assume that

\[ \sum_{a \in A} r(a, s) > 0 \quad \forall s \in S, \quad (2.1) \]
and
\[ \sum_{s \in S} r(a, s) > 0 \quad \forall a \in A, \tag{2.2} \]
so every strategy \( s \in S \) consumes a certain amount of resources and every resource \( a \in A \) is consumed by some strategies.

Given an NCG \( \Gamma \), we denote by \( T := \sum_{k=1}^{K} d_k \) the total user volume of \( \Gamma \). In our study, the user volume vector \( d = (d_k)_{k=1,\ldots,K} \) is not assumed to be fixed, and neither is the total volume \( T \).

A feasible strategy profile (simply called profile or strategy profile in the sequel) \( f \) of an NCG \( \Gamma \) is a vector \((f_s)_{s \in S}\), such that
\[
\sum_{s \in S_k} f_s = d_k \quad \forall k = 1, \ldots, K, \tag{2.3}
\]
\[
f_s \geq 0 \quad \forall s \in S, \tag{2.4}
\]
where \( f_s \) is the volume of users adopting strategy \( s \). \( \text{(2.3)} \) means that a feasible strategy profile fulfills all demands, i.e., it distributes all users to available strategies.

Given a feasible strategy profile \( f = (f_s)_{s \in S} \), the total demand of resource \( a \in A \) is defined by
\[
f_a := \sum_{s \in S} r(a, s) f_s = \sum_{k=1}^{K} \sum_{s \in S_k} r(a, s) f_s,
\]
where \( r(a, s)f_s \) is the demanded volume of resource \( a \) by users using strategy \( s \) w.r.t. profile \( f \). Then, the price of a strategy \( s \) (i.e., the cost of a user using strategy \( s \)) under profile \( f \) is
\[
\tau_s(f) := \sum_{a \in A} r(a, s) \cdot \tau_a(f_a).
\]
Finally, the corresponding social cost (or cost for simplicity) of the feasible strategy profile \( f \) is
\[
C(f) := \sum_{a \in A} \tau_a(f_a) f_a = \sum_{a \in A} \sum_{k=1}^{K} \sum_{s \in S_k} \tau_a(f_a) r(a, s) f_s = \sum_{k=1}^{K} \sum_{s \in S_k} f_s \cdot \tau_s(f),
\]
which is just the resulting total cost of the \( T \) many users of the NCG.

In NCGs, strategy profiles at system optimum and at Nash equilibrium are of great interest. Strategy profiles at system optimum are often thought of the most ideal profiles, which users are
hoped to adopt. Strategy profiles at Nash equilibrium reflect the user behavior in practice, and can be used to forecast the behavior of users. Formally, a feasible strategy profile is at social optimum (SO) if it minimizes the social cost. A feasible strategy profile is an Nash equilibrium (NE) if for each $k \in \{1, \ldots, K\}$, and any two strategies $s, s' \in S_k$, $\tau_s(f) \leq \tau_{s'}(f)$ if $f_s > 0$. So if $s$ is used, it should be one of the cheapest! In the sequel, to facilitate our discussion, we shall call a strategy profile at Nash equilibrium an NE-profile, and a strategy profile at social optimum an SO-profile.

2.2. The routing games for static road traffic

A typical NCG example is the routing game for static road traffic from [Wardrop 1952], where

(a) the resource set $A$ is the set of arcs (streets) of a directed graph (road network) $G = (V, A)$,

(b) the consumption price function $\tau_a : [0, +\infty) \mapsto [0, +\infty)$ is just the flow-dependent travel time function of arc $a \in A$, which is usually assumed to be continuous and non-decreasing,

(c) each user group $k$ corresponds to an origin-destination (OD) pair $(o_k, t_k)$, where $o_k, t_k \in V$, for $k = 1, \ldots, K$,

(d) $d_k$ corresponds to the travel demand from $o_k$ to $t_k$, for $k = 1, \ldots, K$,

(e) $S_k$ corresponds to the set of all paths from $o_k$ to $t_k$ of graph $G = (V, A)$, for $k = 1, \ldots, K$,

(f) the constant $r(a, s)$ is now the indicator function of the relation "$a \in s$", which is $\{0, 1\}$-valued, and equals 1 iff arc $a$ belongs to path $s$,

(g) a feasible strategy profile $f = (f_s)_{s \in S}$ is now a feasible traffic flow fulfilling all the travel demands, and the social cost of $f$ is just the total travel time of the users.

In road traffic, $\tau_a(\cdot)$ is often assumed to be a so-called BPR-function [Bureau of Public Roads 1964], i.e.,

$$\tau_a(x) = \tau_a(0) \left(1 + \alpha \left(\frac{x}{u_a}\right)^\beta\right) = \tau_a(0) + \frac{\alpha \tau_a(0)}{u_a^\beta} x^\beta := \gamma_a x^\beta + \eta_a, \quad (2.5)$$

where $\tau_a(0) := \eta_a$ is the free-flow travel time (the basic cost) of street $a \in A$, $u_a > 0$ represents the capacity (total volume) of street (resource) $a \in A$, and $\beta, \alpha > 0$ are constants independent of streets. Constants $u_a$ and $\tau_a(0)$ reflect road conditions such as street length, lane numbers, speed limits, etc. Constants $\beta, \alpha$ reflect the growth of the travel time with the increase of the traffic intensity. In practice, one often takes $\beta = 4$ and $\alpha = 0.15$. 
2.3. The price of anarchy

The natural (selfish) behavior of users in congestion games is usually considered to be inefficient, see, e.g., Ch. 18 of Nisan et al. (2007) for a survey. The inefficiency is usually measured by the PoA which, under our setting of non-decreasing and continuous consumption price functions, equals the ratio of the cost of an NE-profile over the cost of an SO-profile, i.e.,

$$\text{PoA} = \frac{C(\tilde{f})}{C(f^*)} \geq 1,$$

where $\tilde{f}$ is an NE-profile, and $f^*$ is an SO-profile.

Pigous’s example, see Nisan et al. (2007) or Figure 1, has already shown that selfish behavior need not be bad for large $T$. However, this example is too artificial to be a convincing evidence. To further understand the possible efficiency of user behavior, we still need a closer inspection of the PoA of NCGs. For this purpose, we introduce the following two definitions. The first one concerns the design of a game, in which the nature (selfishness) of users automatically minimizes social cost. Such games are “perfect” to a certain extent. However, they might be too restrictive for congestion games in practice. The second definition gives a more practical alternative. We will show later that many congestion games possess this property.

**Definition 2.2** We call an NCG $\Gamma$ a **well designed game** (WDG) if the cost of NE-profiles equals the cost of SO-profiles for all possible user volume vectors $d = (d_k)_{k=1,\ldots,K}$ with $T = \sum_{k=1}^{K} d_k > 0$. Throughout the paper, we shall denote by WDG also the class of all well designed NCGs.

Obviously, user behavior is completely efficient in a well designed game. Example 2.1 below gives examples of well designed games.

**Example 2.1** An NCG $\Gamma$, whose resource consumption price functions have the form $\tau_a(x) = \alpha_a x^\beta$ for some constants $\alpha_a \geq 0$ and a constant $\beta \geq 0$ independent of resources, is well designed. This can be easily seen by observing the necessary and sufficient conditions for NE-profiles and SO-profiles proposed in Roughgarden and Tardos (2002).
Definition 2.3 We call an NCG $\Gamma$ an asymptotically well designed game (AWDG) if the PoA of the game approaches 1 as $T = \sum_{k=1}^{K} d_k$ approaches infinity. In the sequel, we denote by AWDG also the class of all asymptotically well designed NCGs.

Obviously, WDG $\subseteq$ AWDG, but WDG $\neq$ AWDG. For instance, Pigous’s example belongs to AWDG, but not WDG. Note that Colini-Baldeschi et al. (2017) have already shown an example of NCGs which do not belong to AWDG. Thus, WDG $\subseteq$ AWDG $\subsetneq$ NCGs. Section 3 will explore more properties of AWDG.

3. User behavior need not be bad

User behavior in an AWDG need not be bad, but may eventually lead the underlying system into an equilibrium close to an social optimum. Thus, it is worth exploring which NCGs belong to AWDG. This will require us to analyze the convergence of the PoA of NCGs. We will devote this Section to such an analysis. We shall discuss conditions which can guarantee that the PoA converges to 1 as the total volume $T$ increases to infinity.

The first condition is the so-called gaugeability proposed by Colini-Baldeschi et al. (2017). They proved that every gaugeable routing game is an AWDG, see the Theorem 4.1 of Colini-Baldeschi et al. (2017). A gaugeable routing game is an NCG $\Gamma$ possessing the following four properties.

(G1) There exists a regularly varying function $g(\cdot)$ such that $\tau_a(x)/g(x) \to c_a$ with $x \to \infty$ for every resource $a \in A$, where $c_a \in [0, +\infty]$ is a constant depending only on resource $a$. A function $g(\cdot)$ is said to be regularly varying if $\lim_{t \to \infty} g(tx)/g(x)$ is finite and nonzero for all $x > 0$.

(G2) For each $a \in A$ and $s \in S$, $r(a, s) = 1_s(a)$ is the indicator function of the relation “$a \in s$”.

(G3) For each $k = 1, \ldots, K$, there exists a strategy $s \in S_k$ such that

$$c_s := \max\{c_a : a \in A, \text{ and } r(a, s) = 1\} \in [0, +\infty).$$

Such a strategy is called tight, and so (G3) says that every group has a tight strategy.

(G4) The total user volume of tight groups accounts for a non-negligible proportion of the total user volume $T$ as $T \to \infty$, i.e.,

$$\lim_{T \to \infty} \frac{T^{\text{tight}}}{T} > 0,$$
where a user group \( k \in \{1, \ldots, K\} \) is called tight if

\[
\min_{s \in S_k} c_s = \min_{s \in S_k} \max \{c_a : a \in A, \text{ and } r(a, s) = 1\} \in (0, +\infty),
\]

and \( T^{\text{tight}} := \sum_{k \text{ is tight}} d_k \).

Actually, one can easily provide examples of NCGs that are in AWDG, but not gaugeable, even in the context of road traffic. The proof of Theorem 3.1 presents such an example. Other examples are given in Theorem 3.4 and Example 3.1. Theorem 3.1 below indicates that there are many games in AWDG that are not gaugeable. Thus, we need a more general condition to analyze the convergence of the PoA.

**Theorem 3.1** There exist routing games in AWDG that are not gaugeable.

The results of Colini-Baldeschi et al. (2017) require the existence of a uniform gauge function \( g(\cdot) \) of the underlying game, i.e., a common regularly varying gauge function \( g(\cdot) \) for all \( a \in A \), see Colini-Baldeschi et al. (2017) or (G1). The regular variation of the gauge function \( g(\cdot) \) plays a crucial role in the study of Colini-Baldeschi et al. (2017). It indicates that the consumption price functions \( \tau_a(x) \) vary only moderately with the variation of \( x \). Moreover, the gaugeability in Colini-Baldeschi et al. (2017) heavily depends on the structure of the user volume vectors \( d = (d_k)_{k=1, \ldots, K} \).

It requires that the total user volume of tight groups accounts for a non-negligible proportion of the total volume \( T \), see (G4). These two features of gaugeability greatly restrict the generality of (G4). To derive a weaker condition, we must avoid these two restrictions. In Subsection 3.1, we will propose a new concept called scalability, and define two kinds of games related to scalability, namely, scalable games and strongly scalable games. Different from the gaugeability in Colini-Baldeschi et al. (2017), our notion of scalability does neither assume the existence of a uniform gauge function, nor a particular structure of user volume vectors \( d = (d_k)_{k=1, \ldots, K} \). This will make our results much more general.
3.1. Analyzing the convergence of the PoA in general non-atomic congestion games

Our approach will use a normalization of the model specified in Subsection 2.1 which stems from [Colini-Baldeschi et al. (2017)]. It concerns the distribution of users among groups, and also among strategies of an arbitrary strategy profile. Let us start with details of this normalization.

Let \( d = (d_k)_{k=1,\ldots,K} \) be a user volume vector with total volume \( T(d) = \sum_{k=1}^{K} d_k \). We employ the notation \( T(d) \) to explicitly note the dependence of \( T \) on \( d \). In the sequel, we may still directly use \( T \) instead of \( T(d) \) so as to simplify notation when the vector \( d \) is unambiguous.

We denote by \( d_k := d_k / T(d) \) the proportion of the \( k \)-th volume \( d_k \) in the total volume \( T(d) \) w.r.t. the given vector \( d \) for every \( k = 1,\ldots,K \). Then, the vector \( d := (d_k)_{k=1,\ldots,K} \) represents the distribution of the \( K \) user group volumes w.r.t. vector \( d \).

For a feasible strategy profile \( f = (f_s)_{s \in S} \) and a strategy \( s \in S \), we denote by \( f_s := f_s / T(d) \) the proportion of users choosing strategy \( s \) in profile \( f \) w.r.t. vector \( d \). Then, the vector \( f := (f_s)_{s \in S} \) represents the distribution of users among strategies in profile \( f \) w.r.t. vector \( d \). Obviously,

\[
\sum_{s \in S_k} f_s = d_k \quad \forall k = 1,\ldots,K, \tag{3.1}
\]

\[
f_s \geq 0 \quad \forall s \in S. \tag{3.2}
\]

With this normalization, we can further define the resource load rate vector, i.e., the normalized demand volume vector of resources. For every \( a \in A \), let \( f_a := \sum_{s \in S} r(a,s) \cdot f_s \) denote the load rate of resource \( a \) for profile \( f \) w.r.t. vector \( d \). Then, \( (f_a)_{a \in A} \) is the load rate vector corresponding to the demand volume vector \( (f_a)_{a \in A} \) of resources. Obviously,

\[
f_a = \sum_{a \in A} r(a,s) \cdot f_s = T(d) \cdot \sum_{a \in A} r(a,s) \cdot f_s = T(d) \cdot f_a \tag{3.3}
\]

for all resources \( a \in A \).

For a fixed distribution \( d = (d_k)_{k=1,\ldots,K} \) of users among groups, the load rate \( f_a \) of a resource \( a \in A \) is said to be admissible to the distribution \( d \) if there is a distribution \( f = (f_s)_{s \in S} \) of users among strategies such that \( f \) is feasible w.r.t. the distribution \( d \), i.e., \( f \) satisfies (3.1) and (3.2),
and \( f_a = \sum_{s \in S} r(a,s) \cdot f_s \). Note that values of \( f_a \) admissible to a distribution \( d \) form a closed interval \( I_a(d) \) on \( \mathbb{R}_+ \cup \{0\} \), i.e., \( I_a(d) = [v_a, w_a] \) for some constants \( w_a, v_a \) with \( 0 \leq w_a \leq v_a \). This can be easily proved by observing first that if \( f \) and \( f' \) are both feasible w.r.t. distribution \( d \), then \( (1 - \kappa) f + \kappa f' \) is also feasible w.r.t. \( d \) for any constant \( \kappa \in [0, 1] \), and second that if \( f_a(n) \rightarrow f_a \) as \( n \rightarrow \infty \) for a sequence \( \{f_a(n)\}_{n \in \mathbb{N}} \) of load rates of resource \( a \) that are admissible to \( d \), then \( f_a \) is again a load rate of \( a \) admissible to \( d \). For each \( a \in A \), let \( I_a : = \bigcup_d I_a(d) \) be the set of all admissible values of \( f_a \), where \( d \) ranges over all distributions. Similar arguments show that \( I_a \) is also a closed interval of \( \mathbb{R}_+ \cup \{0\} \).

By (3.3), the social cost of a feasible profile \( f \) can be rewritten as

\[
C(f) = \sum_{a \in A} f_a \cdot \tau_a(f_a) = \sum_{a \in A} T(d) \cdot f_a \cdot \tau_a(T(d) \cdot f_a)
\]

(3.4)

which will be frequently used in the sequel.

Given an NCG

\[
\Gamma_d = \left( A, K, S, (r(a,s))_{a \in A, s \in S}, (\tau_a)_{a \in A}, d \right)
\]

(where the subscript \( d \) is employed to show the dependence of \( \Gamma \) on the user volume vector \( d \) explicitly), and a constant \( c \in \mathbb{R}_+ \), we denote by

\[
\Gamma_d/c := \left( A, K, S, (r(a,s))_{a \in A, s \in S}, (\tau_a)_{a \in A}, d \right) \quad \text{with} \quad \tau_a(f_a) := \tau_a(T(d) \cdot f_a) / c
\]

the scaled game of \( \Gamma_d \) under the scaling factor \( c \). Here, \( \tau_a(f_a) \) is the scaled consumption price function for all \( a \in A \) and all \( f_a \in I_a(d) \). Obviously, the distribution \( f \) of a feasible profile \( f \) of \( \Gamma_d \) is a feasible profile of the scaled game \( \Gamma_d/c \). In the sequel, we will use bold-face symbols for scaled games. Table I list them and their counterparts in the original games.

By (3.4), the PoA of the scaled game \( \Gamma_d/c \) equals the PoA of the original game \( \Gamma_d \), since the distribution \( \tilde{f} \) of an NE-profile \( \tilde{f} \) of the original game \( \Gamma_d \) is an NE-profile of the scaled game \( \Gamma_d/c \),
Table 1  Symbols in original and scaled games

| Original                                      | Scaled with scaling factor \( c > 0 \)                  |
|-----------------------------------------------|--------------------------------------------------------|
| \( d = (d_k)_{k=1,\ldots,K} \)              | \( d = (d_k)_{k=1,\ldots,K} \)                        |
| \( d_k = \sum_{s \in S_k} f_s \)            | \( d_k = d_k/T(d) = \sum_{s \in S_k} f_s \)          |
| \( \sum_{k=1}^{K} d_k = T(d) \)            | \( \sum_{k=1}^{K} d_k = 1 \)                         |
| \( f = (f_s)_{s \in S} \)                   | \( f = (f_s)_{s \in S} \)                             |
| \( f_a = \sum_{a \in A} r(a,s) f_s = T(d) \cdot f_a \) | \( f_a = \sum_{a \in A} r(a,s) f_s \)                |
| \( \tau_a(f_a) \)                           | \( \tau_a(f_a) = \frac{\tau_a(T(d):f_a)}{c} \)     |
| \( C(f) = \sum_{a \in A} f_a \cdot \tau_a(f_a) \) | \( C(f) = \sum_{a \in A} f_a \cdot \tau_a(f_a) = \frac{C(f)}{c T(d)} \) |

and the distribution \( f^* \) of an SO-profile \( f^* \) of the original game \( \Gamma_d \) is an SO-profile of the scaled game \( \Gamma_{d/c} \).

Recall that every user volume vector \( d \) only has a unique social cost for NE-profiles, since we assume that all consumption price functions \( \tau_a(\cdot) \) are continuous and non-decreasing (see Roughgarden and Tardos (2002) or Smith (1979)). We can therefore consider the PoA of the original game \( \Gamma_d \) as a function of the user volume vector \( d \). To indicate this clearly, we will denote the PoA corresponding to a user volume vector \( d \) by PoA\((d)\). Note that PoA\((d)\) is also the PoA of the scaled game \( \Gamma_{d/c} \) for every scaling factor \( c > 0 \).

We will study the convergence of PoA\((d)\) by inspecting the convergence of scaled games under a suitable sequence of scaling factors. To this end, we need to formalize the scaling of NCGs. Definition 3.1 below defines the concept of scalable games. It roughly states that the limit of a sequence of scaled games exists for a suitable sequence of scaling factors.

**Definition 3.1** We say that an NCG

\[
\Gamma = \left( A, K, \mathcal{S}, (r(a,s))_{a \in A: s \in S}, (\tau_a)_{a \in A}, d \right)
\]
is scalable if for any sequence $\{d^{(n)} = (d^{(n)}_k)_{k=1,...,K}\}_{n \in \mathbb{N}}$ of user volume vectors with total volume $T(d^{(n)}) \to \infty$, there exist an infinite subsequence $\{n_i\}_{i \in \mathbb{N}}$, and a scaling sequence $\{g_i\}_{i \in \mathbb{N}}$ of positive numbers such that:

(S0) The distributions $d^{(n_i)} = (d^{(n_i)}_k)_{k=1,...,K}$ converge to $d = (d_k)_{k=1,...,K}$ for some distribution vector $d$ as $i \to \infty$.

(S1) There exists a limit price function vector $(l_a)_{a \in A}$ such that every $l_a(\cdot)$, defined on $I_a$, is either a non-decreasing function with range $[0, +\infty)$, or $l_a(f_a) \equiv +\infty$ for all $f_a \in I_a \cap (0, +\infty)$.

(S2) For all resources $a \in A$ and all $f_a \in I_a$, the scaled consumption price function $\tau_a(f_a)$ with scaling factor $g_i$ converges to the limit price function $l_a(f_a)$ as $i \to \infty$, i.e.,

$$\lim_{i \to \infty} \frac{\tau_a(T(d^{(n_i)}_i) \cdot f_a)}{g_i} = l_a(f_a).$$

Moreover, if $l_a(f_a) < +\infty$, then the limit price function $l_a(x)$ of resource $a$ is continuous at the point $x = f_a$.

(S3) Every group $k = 1, \ldots, K$ is either negligible, i.e., the total cost of the group $k$

$$\sum_{s \in S_k} f^{(n_i)}_s \sum_{a \in A} r(a, s) \cdot \tau_a(f^{(n_i)}_a) \cdot T(d^{(n_i)}_i) g_i \to 0 \quad \text{as } i \to +\infty,$$

where $f^{(n_i)}$ is an arbitrary feasible profile w.r.t. the user volume $d^{(n_i)}$ for each $i \in \mathbb{N}$, or has a tight strategy $s \in S_k$, where tight means that the limit price function $l_a(\cdot)$ of resource $a$ has range $[0, +\infty)$, i.e., $l_a(\cdot)$ is finite for every resource $a \in A$ with $r(a, s) > 0$.

(S4) NE-profiles and SO-profiles of the limit game

$$\Gamma_\infty := \left( A^{\text{tight}}, K, S^{\text{tight}}, (r(a, s))_{a \in A^{\text{tight}}, s \in S^{\text{tight}}}, (l_a)_{a \in A^{\text{tight}}}, d \right)$$

have equal cost under (the limit distribution) $d$ as user volume vector. Here, $S^{\text{tight}} := \bigcup_{k=1}^K S^{\text{tight}}_k$, $S^{\text{tight}}_k$ denotes the set of tight strategies belonging to $S_k$, and $A^{\text{tight}} \subseteq A$ is the set of all resources $a$ with $r(a, s) > 0$ for some $s \in S^{\text{tight}}$.

(S5) The cost of NE-profiles of the limit game is positive under (the limit distribution) $d$ as user volume vector.
In the sequel, we denote by SG the class of all scalable NCGs. We call the limit game \( \Gamma_\infty \) the limit of the scaled games \( \Gamma_{d(n_i)/g_i} \). Note that the existence of a convergent subsequence \( \{d^{(n_i)}\}_{i \in \mathbb{N}} \) of the distribution vector sequence \( \{d^{(n)}\}_{n \in \mathbb{N}} \) in (S0) of Definition 3.1 is not a real restriction, since the sequence \( \{d^{(n)}\}_{n \in \mathbb{N}} \) is bounded. Note also that every limit price function \( l_a(\cdot) \) is only required to exist on the interval \( I_a \), i.e., the set of all admissible load rate \( f_a \) of resource \( a \in \mathbb{A} \). Conditions (S1) and (S2) imply that every consumption price function \( l_a(\cdot) \) of \( \Gamma_\infty \) is non-decreasing and continuous on the interval \( I_a \). Conditions (S3) and (S5) require that the scaling sequence \( \{g_i\}_{i \in \mathbb{N}} \) should be moderately large compared to the consumption prices, which will guarantee that \( \Gamma_\infty \) is well defined. Condition (S4) means that Nash equilibria of \( \Gamma_\infty \) are social optima.

Compared to the gaugeability in Colini-Baldeschi et al. (2017), our notion of scalability is more flexible. We do not require a uniform function \( g(\cdot) \) to gauge the underlying game. Instead, we allow different sequences \( \{d^{(n)}\}_{n \in \mathbb{N}} \) of user volume vectors to have different scaling sequences, and the existence of a scalable subsequence \( \{d^{(n_i)}\}_{i \in \mathbb{N}} \) of \( \{d^{(n)}\}_{n \in \mathbb{N}} \) is already sufficient for our results. In addition, our definition of scalability does not distinguish between user groups, and does not assume particular distributions of users among groups. Thus, our definition of scalability is indeed weaker than that of gaugeability.

Theorem 3.2 below shows that every scalable game is asymptotically well designed, i.e., belongs to AWDG. The proof of Theorem 3.2 is inspired by that of Theorem 4.1 in Colini-Baldeschi et al. (2017). However, we do not need any properties of regular varying functions. Because of conditions (S4) and (S5), we only need to show that the scaled costs of both, NE-profiles and SO-profiles converge, and that the limit distributions of NE-profiles under scaling sequences \( \{g_i\}_{i \in \mathbb{N}} \) are again NE-profiles of the limit games, see the Appendix for a detailed proof. Thus, our proof is much simpler compared to that of Theorem 4.1 in Colini-Baldeschi et al. (2017).

**Theorem 3.2** Every scalable game is asymptotically well designed, and so belongs to AWDG.

Although our condition is weaker, all results of Colini-Baldeschi et al. (2017) carry over to our study. Corollary 3.1 below states that every gaugeable game is scalable. Thus, Theorem 3.2
generalizes Theorem 4.1 of Colini-Baldeschi et al. (2017), although our proof of Theorem 3.2 appears to be much simpler.

**Corollary 3.1** Every gaugeable game is scalable.

Corollary 3.2 and Corollary 3.3 below give particular examples of scalable games. Moreover, Corollary 3.3 actually provides evidence that scalable games are indeed more general than gaugeable games. To see this, one can take a non-regularly varying travel time function for the dominating arc $b$ in Corollary 3.3 and see that the resulting game is not gaugeable, but still scalable.

**Corollary 3.2** Every NCG with polynomial resource consumption price functions $\tau_a(\cdot)$ of the same non-negative degree is scalable, and therefore belongs to AWDG.

We call an arc $b \in A$ of a routing game dominating if every feasible path $s \in S$ uses arc $b$, i.e., $b \in s$ for every path $s \in S$, and if arc $b$ is much slower than all other arcs, i.e.,

$$\lim_{t \to \infty} \frac{\tau_a(t \cdot f_a)}{\tau_b(t \cdot f_b)} = 0.$$  

for all feasible distributions $f = (f_s)_{s \in S}$ of users among groups with load rate vector $(f_a)_{a \in A}$, and for all arcs $a \in A$ with $a \neq b$. A routing game with a dominating arc models a traffic system such that all OD pairs share a common arc, which might be the case in large cities with an obvious separation in living and working areas, e.g., by a river with only one bridge. Corollary 3.3 states that such games belong to AWDG. Thus, the PoA of such games will tend to 1 as total user volume increases.

**Corollary 3.3** Every routing game with a dominating arc is scalable, and thus belongs to AWDG.

We will now inspect the relation between scalability and gaugeability more closely. Recall that our definition of scalability in Definition 3.1 requires that every sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ of user volume vectors has a subsequence $\{d^{(n_i)}\}_{i \in \mathbb{N}}$ such that the limit of the scaled games w.r.t. this subsequence exists. However, we actually proved a much stronger version of this scalability in the proofs of
Corollary 3.1 and Corollary 3.2. In particular, the proof of Corollary 3.1 shows that gaugeability is a special case of this stronger version. Thus, for a better understanding of the relation between gaugeability and scalability, we need to study the stronger version of scalability. Definition 3.2 formally defines this stronger version of scalability.

**Definition 3.2** We say that an NCG $\Gamma$ is **strongly scalable** if for every sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ of user volume vectors with total volume $T(d^{(n)}) \to +\infty$, there exists a scaling sequence $\{g_n\}_{n \in \mathbb{N}}$ of positive numbers, such that:

(SS1) The limit price

$$l_a(x) = \lim_{n \to \infty} \frac{\tau_a(T(d^{(n)}) \cdot x)}{g_n}$$

exists for every resource $a \in A$ and every $x \geq 0$, and $l_a(x)$ is either a non-negative, non-decreasing and continuous real function, or equals $+\infty$ for all $x > 0$.

(SS2) There exists an infinite subsequence $\{n_i\}_{i \in \mathbb{N}}$ such that:

(SS2.1) All groups without tight strategies are negligible for the subsequence $\{d^{(n_i)}\}_{i \in \mathbb{N}}$.

(SS2.2) The subsequence $\{d^{(n_i)}\}_{i \in \mathbb{N}}$ of the distribution vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ corresponding to the user volume vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ converges to a limit distribution $d$. Moreover, there exists a group $k_0$ with $d_{k_0} > 0$ such that for all strategies $s \in S_{k_0}$ there is a resource $a \in A$ with $r(a,s) > 0$ and $l_a(x) \in (0, +\infty)$ for all $x > 0$.

(SS2.3) NE-profiles and SO-profiles of the limit game $\Gamma_\infty$ have equal cost w.r.t. user volume vector $d$, where $\Gamma_\infty$ is defined as in Definition 3.1.

Obviously, a strongly scalable game is scalable, where we observe that (SS2.2) guarantees that the NE-profiles of $\Gamma_\infty$ w.r.t. $d$ are positive. There are two significant differences between scalability and strong scalability. Strong scalability requires that the limit game exists w.r.t. the whole sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ of user volume vectors and the limit price functions $l_a(\cdot)$ are defined on $[0, +\infty)$, while scalability only requires that the limit game exists w.r.t. to some subsequence $\{d^{(n_i)}\}_{i \in \mathbb{N}}$ of $\{d^{(n)}\}_{n \in \mathbb{N}}$ and that the limit price functions $l_a(\cdot)$ are defined on the closed intervals $I_a$. In the sequel, we shall
denote by SSG the class of all strongly scalable games. By Corollary 3.1 and Corollary 3.3, we obtain that

\[
gaugeable \text{ games} \subseteq \text{SSG} \subseteq \text{SG} \subseteq \text{AWDG} \subseteq \text{NCGs and gaugeable games} \subseteq \text{SG}.
\]

Now, we aim to further understand the relation between gaugeable and strongly scalable games.

We say that two non-negative real-valued functions \( h_1(\cdot), h_2(\cdot) \) are \( \text{asymptotically comparable} \) if there exists a constant \( c \in (0, +\infty) \) such that

\[
\lim_{x \to +\infty} \frac{h_1(x)}{h_2(x)} = c \quad \forall x > 0.
\]

An NCG is said to have \( \text{mutually comparable} \) consumption price functions if their consumption price functions \( \tau_a(\cdot) \) and \( \tau_b(\cdot) \) are asymptotically comparable for every two resources \( a, b \in A \). Theorem 3.3 below characterizes strongly scalable NCGs with mutually comparable consumption price functions \( \tau_a(\cdot) \). It states that an NCG with mutually comparable consumption price functions \( \tau_a(\cdot) \) is strongly scalable if and only if all the consumption price functions \( \tau_a(\cdot) \) are regularly varying with the same exponent \( \rho > 0 \), i.e., \( \lim_{x \to \infty} \frac{\tau_a(tx)}{\tau_a(x)} = x^\rho \) for all \( x > 0 \). Thus, when the consumption price functions are mutually comparable, strongly scalable games coincide with gaugeable games.

Moreover, Theorem 3.3 extends Corollary 4.8 of Colini-Baldeschi et al. (2017), in which they proved that routing games with mutually comparable and regularly varying travel time functions are gaugeable, if the proportion of the user volume of each group in the total volume is asymptotically bounded away from 0. We do not need this assumption.

**Theorem 3.3** Consider an NCG \( \Gamma \). If all consumption price functions \( \tau_a(\cdot) \) are mutually comparable, then \( \Gamma \) is strongly scalable if and only if all \( \tau_a(\cdot) \) are regularly varying and have the same exponent in the limit \( x^\rho \).

The condition that all consumption price functions are mutually comparable is crucial in the proof of Theorem 3.3. Otherwise, strongly scalable games need not be gaugeable. Theorem 3.4
below asserts this. It states that strongly scalable games are much more general than gaugeable games, when the consumption price functions are not mutually comparable. Hence,

\[
gaugeable\ games \subsetneq SSG \subsetneq SG \subseteq AWDG \subsetneq NCGs.
\]

**Theorem 3.4** Let \( h_1(\cdot), h_2(\cdot) \) be two polynomials with degrees \( \rho_1, \rho_2 > 0 \), respectively. If \( \rho_1 \neq \rho_2 \), then there is a strongly scalable game that is not gaugeable and has two resources \( a, b \in A \) with consumption price functions \( \tau_a(\cdot) = h_1(\cdot), \tau_b(\cdot) = h_2(\cdot) \), respectively.

By Theorem 3.4, strong scalability is indeed much more general than gaugeability. Example 3.1 below shows a scalable NCG with mutually comparable and non-regularly varying consumption price functions. By Theorem 3.3, this game is not strongly scalable, since the consumption price functions are not regularly varying. Thus, scalable games are more general than strongly scalable games. As a result, we now obtain that

\[
gaugeable\ games \subsetneq SSG \subsetneq SG \subseteq AWDG \subsetneq NCGs.
\]

**Example 3.1** Consider a routing game with only one OD pair and two parallel arcs with mutually comparable travel time functions as shown in Figure 2. Let \( \{b_i\}_{i \in \mathbb{N}} \) be a strictly increasing sequence, such that \( b_0 = 0 \) and \( b_{i+1}/b_i \to +\infty \) for \( i \to +\infty \). The function \( \tau(\cdot) \) shared by the travel time functions of the two arcs is continuous and non-decreasing. It is recursively defined as follows,

![Figure 2](image-url)  
An example of scalable games that are not strongly scalable
• \( \tau(x) \equiv 1 \) for all \( x \in [b_0, b_1) \),

• \( \tau(x) = [(x - b_{2i+1}) + 1] \tau(b_{2i}) \) for all \( x \in [b_{2i+1}, b_{2i+2}) \), for all \( i \in \mathbb{N} \),

• \( \tau(x) = \tau(b_{2i}) = \tau(b_{2i-2})[(b_{2i} - b_{2i-1}) + 1] \) for all \( x \in [b_{2i}, b_{2i+1}) \), for all \( i \geq 1 \),

Note that for every travel demand vector sequence \( \{d(n)\}_{n \in \mathbb{N}} \) with total travel demand \( T(d(n)) = b_n \), there is no appropriate scaling sequence \( \{g_n\}_{n \in \mathbb{N}} \) of positive numbers such that \((SS1)-(SS2)\) of Definition 3.2 hold. This follows since the sequence \( \{\tau(b_n x)/\tau(b_n)\}_{n \in \mathbb{N}} \) is divergent because the slope of the linear pieces on \([b_{2i+1}, b_{2i+2}]\) grows with \( b_n \) for \( x \neq 1 \), and so the function \( \tau(\cdot) \) is not regularly varying. Therefore, Theorem 3.3 implies that the game is not strongly scalable. However, the game is scalable and thus belongs to AWDG, see the Appendix for a proof.

The above discussion concerns general NCGs, and does not assume particular properties of the NCGs. In Subsection 3.2, we will focus the discussion on routing games with BPR travel time functions \([2.5]\). These games are often considered as static models of rush hour road traffic in practice. By Corollary 3.2 we know that such games are scalable, and thus belong to AWDG. This means that selfish routing need not be bad in a road traffic system with a huge total travel demand.

3.2. The convergence of the PoA in routing games with BPR travel time functions

In this Subsection, we consider routing games with BPR travel time functions. We will analyze the convergence rate of the PoA of these games.

We will first investigate the relation between SO-profiles and NE-profiles, and show that every SO-profile is an \( \epsilon \)-approximate NE-profile. The concept of \( \epsilon \)-approximate NE-profiles was proposed in Roughgarden and Tardos (2002). It models user behavior more precisely from a practical perspective.

**Definition 3.3** A feasible strategy profile \( f \) is an \( \epsilon \)-approximate NE-profile for a constant \( \epsilon > 0 \) if

\[
\tau_s(f) = \sum_{a \in A} r(a, s)\tau_a(f_a) \leq (1 + \epsilon) \sum_{a \in A} r(a, s')\tau_a(f_a) = (1 + \epsilon)\tau_{s'}(f)
\]

for any \( k = 1, \ldots, K \), and any \( s, s' \in S_k \) with \( f_s > 0 \).
In practice, it might be difficult for individual users to detect tiny differences between the travel times of paths (strategies). Two paths that have about the same travel times might be considered as equally good choices. Hence, in practice, users need not follow the best NE-profile, but an \( \epsilon \)-approximate NE-profile, due to their selfishness and imprecise judgement.

Theorem 3.5 below shows that every SO-profile is a \( O(d_{\text{min}}^{-\beta}) \)-approximate NE-profile, where \( d_{\text{min}} = \min \{d_k : k = 1, \ldots, K\} \) denotes the minimum travel demand in the travel demand vector and \( \beta \geq 0 \) is the degree of the BPR-functions (2.5) that is often 4 in practice. Although the relative ratio \( d_{\text{min}}/T \) may decrease as the total demand \( T = \sum_{k=1}^{K} d_k \) increases, \( d_{\text{min}} \) itself may still increase as the road network become more and more crowded. Thus, selfish behavior of a large number of non-cooperative individuals will automatically approximate social optimum cost.

**Theorem 3.5** Consider travel time functions \( \tau_a(t) = \gamma_a t^\beta + \eta_a \) for some constants \( \gamma_a > 0, \eta_a > 0, \beta \geq 0 \), and all \( a \in A \). Let \( d_{\text{min}} = \min \{d_k : k = 1, \ldots, K\} \). Then, every SO-profile is an \( O(d_{\text{min}}^{-\beta}) \)-approximate NE-profile.

The proof of Theorem 3.5 actually indicates that

\[
\tau_s(f^*) \leq (1 + O(d_k^{-\beta})) \cdot \tau_{s'}(f^*)
\]

for every SO-profile \( f^* = (f_s^*)_s \in S \), in which \( s \in S_k \) is a strategy with positive flow under \( f^* \), and \( s' \in S_k \) is a strategy other than \( s \), for each \( k \in \{1, \ldots, K\} \). Thus, in practice, the users of OD pairs with large demands \( d_k \) will approximately follow paths of an SO-profile, and their choices will be independent of the choices of other users. In particular, when all OD pairs have large travel demands, an SO-profile is an \( O(T^{-\beta}) \)-approximate NE-profile.

We will show second in Theorem 3.6 that the PoA of routing games with BPR travel time functions is of order \( 1 + O(T^{-\beta}) \). Thus, the PoA of routing games converges very fast to one, since \( \beta \) usually equals 4 in practice. This greatly improves the convergence rate \( O(T^{-1}) \) shown in Theorem 5.1 of [Colini-Baldeschi et al. (2017)](https://example.com), and also proves a conjecture of [O’Hare et al. (2016)](https://example.com) that the PoA follows a power law for large demands.
Theorem 3.6 Assume that $\tau_a(t) = \gamma_a t^\beta + \eta_a$ for any $a \in A$. Then, the PoA is $1 + O(T^{-\beta})$, where $T = \sum_{k=1}^{K} d_k$.

Theorem 3.7 below deepens the understanding of the convergence of the PoA in road traffic, and provides a method for estimating the cost of both, NE-profiles and SO-profiles, when the total travel demand $T$ is large. In that case, the distribution of users among OD pairs will play a pivotal role. It turns out to be the “unique” principal factor for the cost of NE-profiles and SO-profiles. Moreover, Theorem 3.7 b) states that the distribution $\tilde{f}$ of an NE-profile $\tilde{f}$ and the distribution $f^*$ of an SO-profile $f^*$ will be almost identical when the total travel demand $T$ is large enough.

Theorem 3.7 Assume that $\tau_a(x) = \gamma_a \cdot x^\beta + \eta_a$ for all $a \in A$ with some constants $\gamma_a, \eta_a > 0$ and $\beta \geq 0$ independent of $a$, and consider a sequence $\{d^{(n)} = (d^{(n)}_k)_{k=1, \ldots, K}\}_{n \in \mathbb{N}}$ of travel demand vectors such that the total demand $T(d^{(n)}) = \sum_{k=1}^{K} d^{(n)}_k \to +\infty$ as $n \to +\infty$. If the distribution vector $d^{(n)} = (d^{(n)}_k)_{k=1, \ldots, K}$ corresponding to the given demand vector $d^{(n)}$ converges to a limit distribution $d = (d_k)_{k=1, \ldots, K}$, i.e., $d^{(n)}_k \to d_k$ for all OD pairs $k = 1, \ldots, K$, then:

a) \[
\lim_{n \to +\infty} \frac{C(f^*(n))}{(T(d^{(n)}))^{\beta+1}} = L(d) = \lim_{n \to +\infty} \frac{C(\tilde{f}^{(n)})}{(T(d^{(n)}))^{\beta+1}} > 0,
\]
where $L(d)$ is the cost of the unique NE-profile of the limit game $\Gamma_\infty$ under scaling sequence $\{T(d^{(n)})^\beta\}_{n \in \mathbb{N}}$ and user volume vector $d$, and where $\tilde{f}^{(n)}, f^*(n)$ are the NE-profile and the SO-profile corresponding to the demand vector $d^{(n)}$ for all $n \in \mathbb{N}$, respectively.

b) For $n$ large enough, the distribution $f^*(n)$ of an SO-profile $f^*(n)$ and the distribution $\tilde{f}^{(n)}$ of an NE-profile $\tilde{f}^{(n)}$ are almost identical, i.e., for every $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that

\[
\max_{s \in S} \left| \tilde{f}^{(n)}_s - f^*_s \right| < \epsilon \quad \text{for all } n \geq N.
\]

Theorem 3.7 a) states that the social cost $L(d)$ of the unique NE-profile of the limit game

$\Gamma_\infty = (A, K, S, (r(a,s))_{a \in A, s \in S}, (l_a(x) = \gamma_a \cdot x^\beta)_{a \in A}, d)$
depends only on the limit distribution \( d \) of users among OD pairs, and that
\( L(d) \cdot T^{\beta+1} \) approximates the cost of both, NE-profiles and SO-profiles, for an arbitrary travel demand vector
\( d = (d_k)_{k=1,\ldots,K} \) with a large enough total demand \( T \), where \( d = (d_k)_{k=1,\ldots,K} \) is the distribution of users among OD pairs corresponding to the vector \( d \). Note that NE-profiles model user choices in practice. Thus, the travel demand distribution \( d \) among OD pairs is pivotal in a road traffic system with a large total travel demand, as it approximately determines the total travel time (social cost) of users.

The coefficients \( \alpha_a \) of the BPR-functions equal \( \frac{\alpha \tau_a(0)}{u_a} \), see (2.5), where \( \tau_a(0) \) is the free-flow travel time and \( u_a \) is the capacity of arc \( a \) for all \( a \in A \). Thus,

\[
L(d) \leq \sum_{s \in S} \sum_{a \in A} r(a, s) \frac{\alpha_a \cdot \tau_a(0)}{u_a^\beta} =: L,
\]

since \( f_s \in [0,1] \) and \( f_a \in [0,1] \) for each feasible distribution \( f = (f_s)_{s \in S} \) of users among OD pairs. Note that the constant \( L \) depends only on road conditions, i.e., on free-flow travel times \( \tau_a(0) \) and capacities \( u_a \), and it can thus be reduced by improving road conditions. It also provides the upper bound \( L \cdot T^{\beta+1} \) to the total travel time when the total travel demand \( T \) is large.

In practice, the travel demand distribution \( d \) of users among OD pairs is actually determined by the location of facilities in the underlying city, where facilities refer to working and living places, hospitals, shopping malls, schools, government offices, etc. Thus, Theorem 3.7 implies that facility locations are crucial factors for the total travel time in networks with large total travel demand \( T \). Suitably re-locating facilities might be an effective solution to reduce the total travel time when the city gets crowded. So city planning should take the impact of facility locations on the travel demand distribution into account.

4. Experimental study

Our experimental study was done with real traffic data during rush hour (7:00 a.m.–9:00 a.m.) within the second ring road of Beijing. The demands and OD pairs were gathered from GPS data of mobile phones. After a suitable calibration of the demands, we obtained \( K = 33,426 \) OD pairs with total travel demand \( T = \sum_{k=1}^{K} d_k = 101,074 \). Figure 3 shows the road network within that
area of Beijing, which is taken from OpenStreetMap. The network contains $|V| = 4,716$ nodes, and $|A| = 10,267$ arcs.

To determine the PoA, we computed an SO-profile and an NE-profile for this instance, i.e., we solved two convex programs (4.1) (for SO) and (4.2) (for NE), respectively.

$$\begin{align*}
\min & \quad C(f) = \sum_{a \in A} f_a \cdot \tau_a(f_a) \\
\text{s.t.} & \quad \sum_{s \in S_k} f_s = d_k, \quad \forall k = 1, \ldots, K, \\
& \quad f_a - \sum_{s \in S} r(a, s) f_s = 0, \quad \forall a \in A, \\
& \quad f_s \geq 0, \quad \forall s \in S.
\end{align*} \tag{4.1}$$

$$\begin{align*}
\min & \quad \sum_{a \in A} \int_0^{f_a} \tau_a(x) \, dx \\
\text{s.t.} & \quad \sum_{s \in S_k} f_s = d_k, \quad \forall k = 1, \ldots, K, \\
& \quad f_a - \sum_{s \in S} r(a, s) f_s = 0, \quad \forall a \in A, \\
& \quad f_s \geq 0, \quad \forall s \in S.
\end{align*} \tag{4.2}$$

In our experiment, every $\tau_a(\cdot)$ is a BPR function (2.5), where we put $\beta = 4$ and $\alpha = 0.15$. The capacity $u_a$ of an arc (street) $a$ is computed as

$$\frac{\text{street length} \times \text{lane numbers}}{7.5},$$
where 7.5 is our estimation of the space occupied by a vehicle. The free flow travel time $\tau_a(0)$ on a street $a$ is computed as

$$\frac{\text{street length}}{\text{allowed maximum driving speed on the street}}$$

We used the software “cmcf” developed by the COGA group at Technical University of Berlin to solve (4.1) and (4.2). The software has been successfully applied in Jahn et al. (2005) and Harks et al. (2015) to compute SO-profiles, NE profiles, tolls for congestion prices and different traffic patterns. For solving the convex programs, the software uses a variant of the Frank-Wolfe algorithm (Fukushima 1984) together with Dijkstra’s algorithm (Dijkstra 1959) for shortest paths in each iteration.

Our implementation was done under Mac OS Sierra on a Laptop with a 2.7 GHz Intel Core i7 CPU. In the implementation, we stopped the software once the current solution has an objective value within 1% of the optimal value.

The experiment was actually carried out in two separate phases. The first phase had already been done before conceiving this paper, and only computed the empirical PoA within the second ring road of Beijing. Table 2 below reports the result from the first phase, which shows that the PoA within that area of Beijing almost equals 1.0. This certainly shocked us at that time, and motivated the study done in this paper.

| PoA  | SO cost        | NE cost        | CPU_SO (s) | CPU_NE (s) |
|------|----------------|----------------|------------|------------|
| 1.0  | 1.23093000E+15 | 1.23083000E+15 | 29287.245  | 29307.265  |

Table 2 The PoA within the 2nd ring road of Beijing. Column “PoA” reports the price of anarchy, column “So cost” reports the cost of SO-profiles, column “NE cost” reports the cost of NE-profiles, and the last two columns report the CPU time for computing SO and NE, respectively.

The second phase was done after obtaining the theoretical results of this article, and aims to empirically verify our theoretical findings. To this end, we took 65 different fractions of the total
33,426 OD pairs, and ran the algorithm for every of them. To save space, we only report the implementation results for some of the 65 fractions. Table 3 shows the results, where column “Perc.” lists the percentage of the 33,426 OD pairs contained in a fraction, column “$K$” lists the corresponding number of OD pairs of that fraction, and column “$T$” lists the corresponding total travel demands of these OD pairs. For instance, for the first row in Table 3 we took 0.01% of the 33,426 OD pairs, which results in $K = \lceil 33,426 \times 0.01\% \rceil = 4$ OD pairs with total travel demand $T = 15$. Table 3 shows that the PoA has already converged to 1 when $K \geq 1,003$ (which accounts for only 3% of the 33,426 OD pairs). This empirically verifies Theorem 3.2.

| Perc.  | SO cost  | NE cost  | PoA | $K$ | $T$ |
|--------|----------|----------|-----|-----|-----|
| 0.01%  | 5.92E+03 | 5.92E+03 | 1.00| 4   | 15  |
| 0.05%  | 1.45E+04 | 1.61E+04 | 1.11| 17  | 51  |
| 0.08%  | 1.90E+04 | 2.11E+04 | 1.11| 27  | 77  |
| 0.10%  | 2.91E+04 | 3.30E+04 | 1.13| 34  | 90  |
| 0.13%  | 3.45E+04 | 3.84E+04 | 1.11| 44  | 108 |
| 0.15%  | 3.76E+04 | 4.16E+04 | 1.11| 51  | 116 |
| 0.20%  | 4.65E+04 | 5.14E+04 | 1.10| 67  | 146 |
| 0.25%  | 6.16E+04 | 6.75E+04 | 1.10| 84  | 193 |
| 0.30%  | 7.56E+04 | 8.32E+04 | 1.10| 101 | 216 |
| 0.35%  | 1.39E+05 | 1.51E+05 | 1.08| 117 | 264 |
| 0.40%  | 1.54E+05 | 1.68E+05 | 1.09| 134 | 343 |
| 0.45%  | 1.73E+05 | 1.89E+05 | 1.09| 151 | 392 |
| 0.50%  | 2.62E+05 | 2.90E+05 | 1.11| 168 | 483 |
| 0.55%  | 2.74E+05 | 3.05E+05 | 1.11| 184 | 506 |
| 0.60%  | 3.12E+05 | 3.48E+05 | 1.12| 201 | 550 |
| 0.65%  | 3.37E+05 | 3.75E+05 | 1.11| 218 | 626 |

Table 3: Convergence of the PoA (To be continued on the next page)
| Perc. | SO cost | NE cost | PoA | K  | T   |
|-------|---------|---------|-----|----|-----|
| 0.70% | 3.51E+05| 3.90E+05| 1.11| 234| 647 |
| 0.75% | 6.15E+05| 6.99E+05| 1.14| 251| 766 |
| 0.80% | 6.63E+05| 7.48E+05| 1.13| 268| 792 |
| 0.85% | 6.96E+05| 7.86E+05| 1.13| 285| 824 |
| 0.90% | 3.65E+06| 3.74E+06| 1.02| 301| 1030|
| 0.95% | 3.75E+06| 3.85E+06| 1.03| 318| 1111|
| 1.00% | 3.84E+06| 3.94E+06| 1.03| 335| 1149|
| 1.50% | 5.12E+06| 5.22E+06| 1.02| 502| 1531|
| 2.00% | 7.73E+06| 7.82E+06| 1.01| 669| 1938|
| 2.50% | 1.43E+07| 1.44E+07| 1.01| 836| 2276|
| 3.00% | 3.81E+07| 3.81E+07| 1.00| 1003| 2726|
| 3.50% | 6.65E+07| 6.65E+07| 1.00| 1170| 3280|
| 15.00% | 1.06E+11| 1.06E+11| 1.00| 5014| 14944|
| 50.00% | 3.71E+13| 3.71E+13| 1.00| 16714| 50038|
| 90.00% | 7.18E+14| 7.18E+14| 1.00| 30084| 90302|
| 100.00% | 1.23E+15| 1.23E+15| 1.00| 33426| 101074|

Table 3: Convergence of the PoA

Figure 4 shows the plot of the PoA w.r.t. the total volume $T$ with the data from our implementation results of the 65 fractions. Figure 4 (a) depicts the PoA with $T$ up to 101,074, which shows that it quickly converges to 1 as $T$ increases. In particular, when $T$ becomes large, the PoA suddenly takes a steep drop and then never rebounds. This empirically verifies Theorem 3.6.
Figure 4 (b) depicts the PoA with $T$ below 3,000, and provides a closer look at the peak part of Figure 4 (a). Table 3 Figure 4 (a) and Figure 4 (b) show that the PoA increases quickly with the growth of $T$ when $T$ is small, i.e., $T \leq 100$. However, when $T$ gets moderately large, i.e., $100 \leq T \leq 1,200$, the PoA gets choppy. After these oscillations, i.e., $T \geq 1,200$, the PoA decreases very fast to 1.0.

Figure 5 shows the empirical plots of the SO cost and the NE cost w.r.t. $T$, which further verifies the convergence of the PoA. Figure 5 (a) depicts the cost curves of SO-profiles and NE-profiles, for $T \leq 101,074$. Since the cost differences are too small compared to the corresponding demand values, the two curves almost coincide. Figure 5 (b) depicts the two curves for $T \in (200, 2,000)$, and shows that they gradually become identical as $T$ increases. Thus, Figure 5 empirically verifies Theorem 3.7 (b).

Figure 6 shows the empirical plot of the ratio of SO cost over $T^\beta+1$ w.r.t. $T$ with data from the implementation results for the 65 fractions, for $\beta = 4$. It demonstrates that the ratio converges quickly to a constant, as the total travel demand $T$ increases. This empirically verifies Theorem 3.7 (a). Moreover, when $T$ reaches about $2 \cdot 10^4$, the ratio has already converged to the constant $1.18 \cdot 10^{-10}$, which is an estimator of $L(d)$ in Theorem 3.7 (a). The CPU time for computing the SO cost for $T = 101,074$ is about 29,307 seconds, and the CPU time for computing the SO cost
for $T = 20,098$ is about 5,967 seconds. Thus, when we use the approximation method indicated in Theorem 3.7 (a), we can save about $\frac{29,307 - 5,967}{29,307} \approx 79.6\%$ of time to compute the SO cost within the second ring road of Beijing.

Besides a convincing verification of our theoretical findings, our empirical study also shows that there is a threshold value (a saturation point), beyond which the PoA has already decreased to 1. This saturation point seems to be at about $T = 1,200$, which is far below the current total travel demand of 101,074. This means that the current traffic in Beijing is far beyond saturation, and
user guidance policies (e.g., congestion pricing) that do not considerably change the total demand will fail in reducing the total travel time for the huge total demand of Beijing.

5. Conclusion

We proved that every scalable game is asymptotically well designed, see Theorem 3.2. This result generalizes the main result of Colini-Baldeschi et al. (2017) for gaugeable routing games. Moreover, we proved that for some particular cases, gaugeable games coincide with strongly scalable games, see Theorem 3.3 and that, in general, strongly scalable games are much more general than gaugeable games, see Theorem 3.4. In addition, we showed by examples that scalable games are more general than strongly scalable games, see Example 3.1. Altogether, our results extend and enrich the study of Colini-Baldeschi et al. (2017).

For routing games with BPR travel time functions, we proved that an SO-profile is an $\epsilon$-approximate NE-profile, see Theorem 3.5. Moreover, we showed for these particular games that $\text{PoA} = 1 + O(d^{-\beta})$, see Theorem 3.6. This improves the convergence rate $O(T^{-1})$ of Colini-Baldeschi et al. (2017), and also proves a conjecture of O’Hare et al. (2016). In addition, we proved that the distribution of users among groups is a crucial factor for the cost of both, SO-profiles and NE-profiles, when $T$ is large, see Theorem 3.7.

Finally, we have empirically verified our theoretical findings with real traffic data from Beijing. Our empirical results show that the current traffic volume in Beijing is far beyond its saturation point, which indicates that a reduction of the total travel time without reducing the travel demand is not feasible.

Appendix. Proofs of the Theorems and Corollaries

A. Proof of Theorem 3.1

Proof of Theorem 3.1 We prove this theorem by providing a class of routing games that are well designed (thus in AWDG), but not gaugeable. Figure 7 shows such an example. The game consists of two groups of users, the users traveling from $D$ to $E$, and the users traveling from $F$ to $H$. Except for the arc $B1B2$, all other edges in the graph have constant travel times that are
shown in Figure 7. Obviously, the PoA in such a game is always 1, independent of the travel time (or consumption price) function of arc $B1B2$, and independent of the user volumes of the two groups. Therefore, such routing games are well designed. However, once the travel time function of arc $B1B2$ is not regularly varying like $e^x$, then the routing game is no longer gaugeable. For $e^x$, there is no regularly varying function $g$ such that assumption (G3) holds, since no strategy (path) of the graph will be tight in this case.

\[\text{Figure 7}\quad \text{An example of routing games that are not gaugeable}\]

B. Proof of Theorem 3.2

**Proof of Theorem 3.2** Let $\Gamma$ be a scalable NCG, and let $\{d^{(n)}\}_{n \in \mathbb{N}}$ be a sequence of user volume vectors such that the total volume $T(d^{(n)})$ increases to infinity, and

\[
\lim_{n \to \infty} \text{PoA}(d^{(n)}) = \lim_{t \to \infty} \sup \{ \text{PoA}(d) : T(d) = t \} \geq \lim_{t \to \infty} \inf \{ \text{PoA}(d) : T(d) = t \} \geq 1.
\]

To show that $\Gamma \in \text{AWDG}$, we only need to prove that

\[
\lim_{n \to \infty} \text{PoA}(d^{(n)}) \leq 1. \tag{B.1}
\]

Since the limit of PoA($d^{(n)}$) has been assumed to exist, (B.1) can be proved by simply showing that (B.1) holds for an infinite subsequence of $\mathbb{N}$.

For each $n \in \mathbb{N}$, let $\tilde{f}^{(n)}$, $f^*(n)$ be, resp., an NE-profile and an SO-profile w.r.t. $d^{(n)}$. Recall that $d^{(n)} = (d_k^{(n)})_{k=1,...,K}$ is the distribution of users among groups w.r.t. $d^{(n)}$. Then, every $d_k^{(n)} = d_k^{(n)}/T(d^{(n)}) \in [0,1]$, $f^{(n)} = (f_s^{(n)})_{s \in S}$ is the distribution of users among strategies w.r.t. $\tilde{f}^{(n)}$ (i.e., every $\tilde{f}_s^{(n)} = \tilde{f}_s^{(n)}/T(d^{(n)}) \in [0,1]$), and $(\tilde{f}_a^{(n)})_{a \in A}$ is the resulting load rate vector corresponding to $(\tilde{f}_a^{(n)})_{a \in A}$, i.e.,

\[
\tilde{f}_a^{(n)} = \sum_{s \in S} r(a, s) \cdot \tilde{f}_s^{(n)}, \quad \forall a \in A.
\]

This holds similarly for the distribution vector $f^*(n) = (f_s^*(n))_{s \in S}$ and its load rate vector $(f_a^*(n))_{a \in A}$.

Since all the above normalized sequences are bounded, we can assume that:
a1) \( d^{(n)} \to d = (d_k)_{k=1,...,K} \) with \( n \to \infty \), for some limit distribution \( d \) of users among groups.

a2) \( \tilde{f}^{(n)} \to \tilde{f} = (\tilde{f}_s)_{s \in S} \), with \( n \to \infty \), for some feasible distribution \( \tilde{f} \) of users among strategies.

a3) \( f^{*(n)} \to f^* = (f^*_s)_{s \in S} \), with \( n \to \infty \), for some feasible distribution \( f^* \) of users among strategies.

Otherwise, we can take an infinite subsequence fulfilling \( a1\)-\( a3 \).

By \( a2 \) and \( a3 \), we obtain

\[
\begin{align*}
\tilde{f}_a^{(n)} & \to \tilde{f}_a := \sum_{s \in S} r(a,s) \cdot f_s \quad \text{for all } a \in A, \quad \text{and} \\
f_a^{*(n)} & \to f^*_a := \sum_{s \in S} r(a,s) \cdot f^*_s \quad \text{for all } a \in A.
\end{align*}
\]

Then, \((\tilde{f}_a)_{a \in A}\) and \((f^*_a)_{a \in A}\) are the load rate vectors corresponding to \( \tilde{f} \) and \( f^* \), respectively. Since \( \Gamma \) is scalable, we can assume that there is a scaling sequence \( \{g_n\}_{n \in \mathbb{N}} \) such that \( (S0)-(S5) \) hold, and \( \Gamma_\infty \) is the corresponding limit game. Otherwise, we can again take an infinite subsequence fulfilling \( (S0)-(S5) \).

To prove \((B.1)\), we only need to show because of \((S3)\) and \((S5)\) that the scaled cost \( C(\tilde{f}^{(n)})/T(d^{(n)})g_n \) and the scaled cost \( C(f^{*(n)})/T(d^{(n)})g_n \) converge to the cost of \( \tilde{f} \) and the cost of \( f^* \), respectively, and that \( \tilde{f} \) is an NE-profile of the limit game \( \Gamma_\infty \).

To this end, we claim first that

\[
\frac{\tau_a(\tilde{f}_a^{(n)})}{g_n} = \frac{\tau_a(T(d^{(n)})) \cdot \tilde{f}_a^{(n)}}{g_n} \to l_a(\tilde{f}_a) \quad \text{as } n \to \infty, \quad \text{and}
\]

\[
\frac{\tau_a(f^*_a^{(n)})}{g_n} = \frac{\tau_a(T(d^{(n)})) \cdot f^*_a^{(n)}}{g_n} \to l_a(f^*_a) \quad \text{as } n \to \infty,
\]

for all \( a \in A \) with limit price function \( l_a(\cdot) \neq +\infty \). Here, we only prove \((B.2)\). \((B.3)\) follows with an almost identical argument.

Note that \( \tilde{f}_a, f_a^{(n)} \in I_a \) for all \( n \in \mathbb{N} \) and all \( a \in A \), where we recall that \( I_a \) is the closed interval including all admissible load rates of resource \( a \). By \( a4 \), for any \( \epsilon > 0 \), there exist \( f_a^-, f_a^+ \in I_a \) such that \( f_a^+ - f_a^- < \epsilon, f_a^- \leq \tilde{f}_a \leq f_a^+ \), and \( f_a^- \leq f_a^{(n)} \leq f_a^+ \) for large enough \( n \in \mathbb{N} \). Therefore, we obtain for large enough \( n \in \mathbb{N} \)

\[
\frac{\tau_a(T(d^{(n)})) \cdot f_a^-}{g_n} \leq \frac{\tau_a(\tilde{f}_a^{(n)})}{g_n} \leq \frac{\tau_a(T(d^{(n)})) \cdot f_a^+}{g_n},
\]

since \( \tau_a(\cdot) \) is nondecreasing. Letting \( n \to \infty \), we get that

\[
l_a(f_a^-) \leq \lim_{n \to \infty} \frac{\tau_a(\tilde{f}_a^{(n)})}{g_n} \leq \lim_{n \to \infty} \frac{\tau_a(f_a^{(n)})}{g_n} \leq l_a(f_a^+).
\]

Then \((B.2)\) is proved by letting \( \epsilon \to 0 \), and observing that \( f_a^+ - f_a^- < \epsilon, f_a^- \leq \tilde{f}_a \leq f_a^+ \), and that \( l_a(x) \) is continuous at \( x = \tilde{f}_a \), see assumption \((S2)\).
We will now prove first that the scaled cost
\[
\frac{C(\tilde{f}^{(n)})}{T(d^{(n)})g_n} = \sum_{k=1}^{K} \sum_{s \in \mathcal{S}_k} T(d^{(n)}_s) \cdot \tilde{f}_s^{(n)} \cdot \sum_{a \in A} r(a, s) \cdot \tau_a(T(d^{(n)}) \cdot \tilde{f}_a^{(n)})
\]
converges to the limit cost \(L(\tilde{f})\), where
\[
L(\tilde{f}) := \sum_{k=1}^{K} \sum_{s \in \mathcal{S}_k^{tight}} \tilde{f}_s \cdot \sum_{a \in A^{tight}} r(a, s) \cdot l_a(\tilde{f}_a) := \sum_{k=1}^{K} \sum_{s \in \mathcal{S}_k^{tight}} \tilde{f}_s \cdot L_s(\tilde{f}) < +\infty, \quad (B.4)
\]
and second that \(\tilde{f} = (\tilde{f}_s)_{s \in \mathcal{S}}\) is actually an NE-profile of the limit game \(\Gamma_\infty\).

To this end, consider a non-negligible group \(k \in \{1, \ldots, K\}\) and a non-tight \(s \in \mathcal{S}_k \setminus \mathcal{S}^{tight}\) with \(\tilde{f}_s > 0\). Then, (2.1) and the definition of a tight strategy, see (S3), yield that the limit price of \(s\) w.r.t. \(\tilde{f}\) is \(+\infty\), i.e.,
\[
L_s(\tilde{f}) = \sum_{a \in A} r(a, s) \cdot l_a(\tilde{f}_a) = +\infty.
\]
By (B.2) and assumption (S3), the limit price \(L'_s(\tilde{f}) = \lim_{n \to \infty} \frac{\tau_s(\tilde{f}^{(n)})}{g_n} < +\infty\) for every tight strategy \(s' \in \mathcal{S}^{tight}\). Therefore, \(\tilde{f}_s = \lim_{n \to \infty} \tilde{f}_s^{(n)} = 0\) for each \(k = 1, \ldots, K\) and each non-tight strategy \(s \in \mathcal{S}_k \setminus \mathcal{S}^{tight}\), since all \(\tilde{f}^{(n)}\) are NE-profiles and we assumed that every non-negligible group \(k\) has at least one tight strategy.

Moreover, for all \(k = 1, \ldots, K\), the total limit cost of a non-tight strategy \(s \in \mathcal{S}_k \setminus \mathcal{S}^{tight}\) must be 0, i.e.,
\[
\lim_{n \to \infty} \frac{\tilde{f}_s^{(n)} \cdot \sum_{a \in A} r(a, s) \cdot \tau_a(T(d^{(n)}) \cdot \tilde{f}_a^{(n)})}{g_n} = 0. \quad (B.5)
\]
Otherwise, there must exist a subsequence \(\{n_i\}_{i \in \mathbb{N}}\) such that for all tight strategies \(s' \in \mathcal{S}_k^{tight}\),
\[
L_{s'}(\tilde{f}) < L_s(\tilde{f}) = \lim_{i \to \infty} \frac{\sum_{a \in A} r(a, s) \cdot \tau_a(T(d^{(n_i)}) \cdot \tilde{f}_a^{(n_i)})}{g_{n_i}} = +\infty, \quad (B.6)
\]
\[
\lim_{i \to \infty} \frac{\tilde{f}_s^{(n_i)} \cdot \sum_{a \in A} r(a, s) \cdot \tau_a(T(d^{(n_i)}) \cdot \tilde{f}_a^{(n_i)})}{g_{n_i}} > 0. \quad (B.7)
\]
Then, (B.6) implies that, for large enough \(i\),
\[
\tilde{f}_s^{(n_i)} = \tilde{f}_s^{(n_i)} / T(d^{(n_i)}) \equiv 0,
\]
again since all \(\tilde{f}^{(n_i)}\) are NE-profiles and every group has at least one tight strategy (see (S3)). This contradicts (B.7). Therefore, (B.5) holds, which means that we can ignore non-tight strategies when we consider NE-profiles of the limit game \(\Gamma_\infty\). Note also that we can ignore these negligible groups. Altogether, we obtain that (B.4) holds, and that \(\tilde{f}\) is a feasible profile of the limit game \(\Gamma_\infty\).

We now prove that \(\tilde{f}\) is an NE profile of \(\Gamma_\infty\). For any \(k = 1, \ldots, K\), and any two strategies \(s, s' \in \mathcal{S}_k\) with \(\tilde{f}_s > 0\), we need to show that \(L_s(\tilde{f}) \leq L_{s'}(\tilde{f})\) for the limit price, i.e., \(s\) is cheaper
than $s'$ w.r.t. the feasible profile $\tilde{f}$ of the limit game. Since $\tilde{f}_s > 0$, we obtain that $\tilde{f}_s^{(n)} > 0$ for large enough $n$. Therefore, $\tilde{f}_s^{(n)} = T(d^{(n)}) \cdot \tilde{f}_s^{(n)} > 0$ for large enough $n$. Since all $f^{(n)}$ are NE-profiles, $\tau_a(\tilde{f}^{(n)}) \leq \tau_a(\tilde{f}^{(n)})$ for large enough $n$. Then $\tau_a(\tilde{f}^{(n)})/g_n \leq \tau_{s'}(\tilde{f}^{(n)})/g_n$ for large enough $n$. Letting $n \to \infty$, we obtain that $L_s(\tilde{f}) \leq L_{s'}(\tilde{f})$. Therefore, $\tilde{f}$ is an NE-profile of the limit game $\Gamma_{\infty}$. By (S4), $\tilde{f}$ is also an SO-profile of the limit game. So, $L(\tilde{f}) \leq L(\mu)$ for every feasible profile $\mu$ of the limit game. By (S5), the limit cost $L(\tilde{f}) > 0$.

By the facts that $C(\tilde{f}^{(n)}) \geq C(f^{(n)})$ for all $n \in \mathbb{N}$, that $0 < L(\tilde{f}) < +\infty$, that $l_a(\cdot)$ is continuous, and by (B.3), we obtain through an argument similar to the above for non-tight strategies that the scaled cost

$$
\frac{C(f^{(n)})}{T(d^{(n)})g_n} = \sum_{k=1}^{K} \sum_{s \in S_k} T(d^{(n)}) \cdot f^{(n)} \cdot \sum_{a \in A} r(a, s) \cdot \tau_a(T(d^{(n)}) \cdot f^{(n)})
$$

converges to the limit cost $L(f^*)$, where

$$
L(f^*) = \sum_{k=1}^{K} \sum_{s \in S_k^{tight}} f^*_s \sum_{a \in A^{tight}} r(a, s) l_a(f^*_s) = \sum_{k=1}^{K} \sum_{s \in S_k^{tight}} f^*_s \cdot L_s(f^*) < +\infty,
$$

and that $f^*$ is a feasible profile (actually an SO-profile) of the limit game $\Gamma_{\infty}$. Hence, $0 < L(\tilde{f}) \leq L(f^*) \leq L(\tilde{f}) < +\infty$. Therefore,

$$
\lim_{n \to \infty} \text{PoA}(d^{(n)}) = \lim_{n \to \infty} \frac{C(\tilde{f}^{(n)})}{C(f^{(n)})} = \lim_{n \to \infty} \frac{C(f^{(n)})}{T(d^{(n)})g_n} = \frac{L(\tilde{f})}{L(f^*)} \leq 1,
$$

which proves (B.1). \hfill \Box

C. Proof of Corollary 3.1

Proof of Corollary 3.1 We now consider a gaugeable game $\Gamma$ with a regularly varying gauge function $g$. Let $\{d(n)\}_{n \in \mathbb{N}}$ be a sequence of user volume vectors with $\lim_{n \to \infty} T(d^{(n)}) = +\infty$. We define a scaling sequence $\{g_n\}_{n \in \mathbb{N}}$ by putting $g_n := g(T(d^{(n)}))$. Using Colini-Baldeschi et al. (2017) and the regular variability of the function $g(\cdot)$, we obtain that

$$
l_a(x) := \lim_{n \to \infty} \frac{\tau_a(T(d^{(n)}) \cdot x)}{g_n} = c_a \cdot x^{\rho}
$$

for all $x \geq 0$ and all $a \in A$, where

$$
c_a = \lim_{x \to \infty} \frac{\tau_a(x)}{g(x)}
$$

is a non-negative constant or $+\infty$, and $\rho \geq 0$ is a constant such that for all $x > 0$

$$
\lim_{t \to \infty} \frac{g(tx)}{g(x)} = x^\rho.
$$

We need to show (S3)-(S5).
By (G3) and the gaugeability of $\Gamma$, we obtain that every group has a tight strategy $s \in S_k$ such that, for all resources $a \in A$ with $r(a, s) > 0$, $l_a(x) = c_a \cdot x^\rho \in [0, +\infty)$ for all $x \geq 0$. Thus, (S3) holds. Let $\Gamma_\infty$ be the limit game as defined in (S4). Obviously, $\Gamma_\infty$ is in WDG, since all consumption prices functions of $\Gamma_\infty$ are of the form $c_a \cdot x^\rho$ for all $a \in A^{\text{tight}}$ and some constant $\rho$ independent of resources $a$. Thus, (S4) holds for every user volume vector that is a limit of the distribution vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ corresponding to the above fixed user volume vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$.

Since the proportion of tight groups in the total travel volume is non-negligible, see (G4), $\Gamma$ has at least one tight group $k$ such that $\lim_{n \to +\infty} d^{(n)}_k > 0$. Recall that a group $k \in \{1, \ldots, K\}$ is called tight if

$$\min \{c_s : a \in A, \text{ and } r(a, s) > 0\} \in (0, +\infty),$$

see (G4) for more details. Let $k_0$ be a tight group with $\lim_{n \to +\infty} d^{(n)}_{k_0} > 0$. Then, $S_{k_0} \cap S^{\text{tight}} \neq \emptyset$, and there must exist a resource $a \in S_{k_0}$ for each $s \in S_{k_0}$ such that $r(a, s) > 0$ and $c_a \in (0, +\infty]$. Therefore, the total limit cost of group $k_0$ is positive w.r.t. any profile $f = (f_s)_{s \in S^{\text{tight}}}$ that is feasible under some limit $d$ of the distribution vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$, i.e.,

$$\sum_{s \in S_{k_0}} f_s \cdot \sum_{a \in A^{\text{tight}}} r(a, s) \cdot l_a(f_s) = \sum_{s \in S_{k_0}} f_s \cdot \sum_{a \in A^{\text{tight}}} r(a, s) \cdot c_a \cdot f_s^\rho > 0.$$

Recall that $f_s$ is the load rate of $a$ w.r.t. $f_s$. Thus, for every user volume vector $d$ that is a limit of the distribution vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$, the NE-profiles of $\Gamma_\infty$ w.r.t. the user volume vector $d$ have positive cost. So, (S5) holds for every user volume vector $d$ that is a limit of the distribution vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$.

Since (S1)-(S3) hold for the whole sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$, and (S4)-(S5) hold for every subsequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ for which the corresponding distribution vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ converges to a limit distribution $d$, we obtain that $\Gamma_d$ is scalable. \hfill \Box

D. Proof of Corollary 3.2

Proof of Corollary 3.2 We assume, w.l.o.g., that $\tau_a(x) = \alpha_a x^\beta + \sum_{i=1}^{m} \alpha_{a,i} x^{\beta_i} + c_0$ for all $a \in A$, and some constants $\alpha_a > 0, \beta \geq 0, 0 \leq \beta_i < \beta$. For any user volume vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ with total volume $T(d^{(n)}) \to +\infty$ as $n \to \infty$, we can define a scaling sequence $\{g_n\}_{n \in \mathbb{N}}$ by putting $g_n := (T(d^{(n)}))^\beta$ for all $n \in \mathbb{N}$. Then, we obtain that

$$l_a(x) = \lim_{n \to \infty} \frac{\tau_a(T(d^{(n)}))}{g_n} = \alpha_a \cdot x^\beta \in (0, +\infty)$$

for all $a \in A$ and all $x \geq 0$. By a similar argument to the proof of Corollary 3.1 we can then show that the given game is scalable. \hfill \Box
E. Proof of Corollary 3.3

Proof of Corollary 3.3 Let \( \{d^{(n)}\}_{n \in \mathbb{N}} \) be a sequence of travel demand vectors such that

\[
\lim_{n \to \infty} T(d^{(n)}) = +\infty.
\]

Let \( b \) be a dominating arc of the given game, and set \( g_n := \tau_b(T(d^{(n)})) \) for each \( n \in \mathbb{N} \). Since \( b \) is a dominating arc, it follows that the load rate \( f_b = 1 \), that

\[
l_a(f_a) = \lim_{n \to \infty} \frac{\tau_a(T(d^{(n)})) f_a}{g_n} = \lim_{n \to \infty} \frac{\tau_a(T(d^{(n)}))}{\tau_b(T(d^{(n)})) f_b} = 0 \quad \forall a \in A \setminus \{b\} \forall f_a \in I_a, \quad \text{and that}
\]

\[
l_b(f_b) = \lim_{n \to \infty} \frac{\tau_b(T(d^{(n)}))}{g_n} = 1.
\]

Since \( b \in s \) for all paths \( s \in S \), all paths and all OD pairs are tight. The limit game has constant travel times on all arcs. Therefore, it belongs to WDG, and Theorem 3.2 applies.

F. Proof of Theorem 3.3

Proof of Theorem 3.3 (\( \Leftarrow \)) This direction is similar to the proof of Corollary 3.1 when we observe that all \( \tau_a(\cdot) \) are mutually comparable and have the same exponent \( \rho > 0 \), and are thus gaugeable.

(\( \Rightarrow \)) To show that all \( \tau_a(\cdot) \) are regularly varying and have the same exponent \( \rho > 0 \), we only need to show by Karamata’s Characterization Theorem (see [Bingham et al. 1987]) that

\[
\lim_{t \to \infty} \frac{\tau_a(tx)}{\tau_a(t)} = h(x) \in (0, +\infty)
\]

(F.1) for all \( x \in (0, +\infty) \), all \( a \in A \), and some function \( h(\cdot) \) independent of resources \( a \).

Let \( a^* \in A \) be an arbitrarily fixed resource. Since the consumption price functions are mutually comparable, (F.1) follows already if we show that there exists a function \( h_{a^*}(x) \) for \( a^* \in A \) such that

\[
\lim_{t \to +\infty} \frac{\tau_{a^*}(tx)}{\tau_{a^*}(t)} = h_{a^*}(x) \in (0, +\infty) \quad \forall x > 0.
\]

(F.2)

Then, we obtain for all \( x > 0 \) and any \( a \in A \) that

\[
\lim_{t \to +\infty} \frac{\tau_a(tx)}{\tau_a(t)} = \lim_{t \to +\infty} \frac{\tau_a(tx)}{\tau_{a^*}(tx)} \cdot \frac{\tau_{a^*}(tx)}{\tau_{a^*}(t)} \cdot \frac{\tau_{a^*}(t)}{\tau_a(t)} = \lim_{t \to +\infty} \frac{\tau_{a^*}(tx)}{\tau_{a^*}(t)} = h_{a^*}(x) \in (0, +\infty),
\]

where mutual comparability gives

\[
\lim_{x \to +\infty} \frac{\tau_a(x)}{\tau_{a^*}(x)} = c_{a,a^*} \quad \text{and} \quad \lim_{x \to +\infty} \frac{\tau_{a^*}(x)}{\tau_a(x)} = \frac{1}{c_{a,a^*}}
\]

for some constant \( c_{a,a^*} \in (0, +\infty) \). Thus, we only need to find a suitable \( a^* \in A \) such that the function \( h_{a^*}(\cdot) \) required by (F.2) exists.
Since the game $\Gamma$ is strongly scalable, there exist for every user volume vector sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} T(d^{(n)}) = +\infty$ a scaling sequence $\{g_n\}_{n \in \mathbb{N}}$ of positive numbers and a limit price function vector $\{(l_a)_{a \in A}\}$ such that

$$\lim_{n \to \infty} \frac{\tau_a(T(d^{(n)}) \cdot x)}{g_n} = l_a(x), \quad \forall a \in A \quad \forall x > 0,$$

where $l_a(\cdot)$ is either a non-negative, non-decreasing and continuous real function, or $l_a(x) \equiv +\infty$ for all $x > 0$, see (SS1). Now, let $\{d^{(n)}\}_{n \in \mathbb{N}}$ be an arbitrarily fixed sequence of user volume vectors of the game $\Gamma$, $\{g_n\}_{n \in \mathbb{N}}$ be a scaling sequence, and $\{(l_a)_{a \in A}\}$ be the corresponding vector of limit price functions.

By (SS3.1), there exists at least one resource $a \in A$ such that $l_a(x) \in (0, +\infty)$ for all $x > 0$. Let $a^*$ be such a resource. We will now construct the limit $h_{a^*}(\cdot)$ satisfying (F.2). For the fixed sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$, we obtain that

$$\lim_{n \to \infty} \frac{\tau_{a^*}(T(d^{(n)}) x)}{\tau_{a^*}(T(d^{(n)}))} = \lim_{n \to \infty} \frac{\tau_{a^*}(T(d^{(n)}) x)}{g_n} = \frac{l_{a^*}(x)}{l_{a^*}(1)} = h_{a^*}(x) \in (0, +\infty) \quad \forall x > 0.$$

To complete the proof, we still need to show that $h_{a^*}$ is independent of the choice of $\{d^{(n)}\}_{n \in \mathbb{N}}$. The following argument asserts this.

Arbitrarily fix two user volume vector sequences $\{d_1^{(n)}\}_{n \in \mathbb{N}}$ and $\{d_2^{(n)}\}_{n \in \mathbb{N}}$ such that $T(d_1^{(n)}) \to +\infty$ and $T(d_2^{(n)}) \to +\infty$ as $n \to +\infty$. We consider two limit functions, $h_{a^*}^{(i)}(\cdot)$, $i = 1, 2$, such that for $i = 1, 2$,

$$\lim_{n \to \infty} \frac{\tau_{a^*}(T(d_1^{(n)}) x)}{\tau_{a^*}(T(d_1^{(n)}))} = h_{a^*}^{(i)}(x) \in (0, +\infty) \quad \forall x > 0.$$

We now define a new user volume sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ by merging the two user volume vector sequences, i.e., $d^{(2m)} = d_1^{(m)}$ and $d^{(2m+1)} = d_2^{(m)}$ for all $m \in \mathbb{N}$. Then, also $T(d^{(n)}) \to +\infty$ as $n \to +\infty$. By the above discussion, there also exists a limit function $h_{a^*}(\cdot)$ such that

$$\lim_{n \to \infty} \frac{\tau_{a^*}(T(d^{(n)}) x)}{\tau_{a^*}(T(d^{(n)}))} = h_{a^*}(x) \in (0, +\infty) \quad \forall x > 0.$$

This implies that $h_{a^*}^{(1)}(x) = h_{a^*}(x) = h_{a^*}^{(2)}(x)$ for all $x > 0$, since $\{d_i^{(n)}\}_{n \in \mathbb{N}}, i = 1, 2$, are subsequences of $\{d^{(n)}\}_{n \in \mathbb{N}}$. Therefore, the limit $h_{a^*}(\cdot)$ is independent of the choice of $\{d^{(n)}\}_{n \in \mathbb{N}}$. This completes the proof.

G. Proof of Theorem 3.4

Proof of Theorem 3.4 We give a proof for $h_1(x) = 2x + 1$ and $h_2(x) = 4x^2 + 1$ that are generic for general polynomials with different degrees. Figure 8 shows a routing game $\Gamma$ with two OD pairs and four parallel paths (arcs). The four polynomials above the four arcs are their travel time
functions, respectively. The user volumes of the game Γ have a particular property that \( d_1 \cdot d_2 = 0 \) for every user volume vector \( d = (d_1, d_2) \), where \( d_1 \) denotes the user volume of the top OD pair and \( d_2 \) denotes the user volume of the bottom OD pair. The following argument shows that Γ is strongly scalable, but not gaugeable.

Consider now a sequence \( \{d^{(n)} = (d_1^{(n)}, d_2^{(n)})\} \) of user volume vectors with total volume

\[
T(d^{(n)}) = d_1^{(n)} + d_2^{(n)} \to \infty \quad \text{as} \quad n \to \infty,
\]

where \( d_1^{(n)} \) denotes the user volume of the top OD pair and \( d_2^{(n)} \) denotes the user volume of the bottom OD pair for each \( n \in \mathbb{N} \). Then, \( d_1^{(n)} \cdot d_2^{(n)} = 0 \) holds for every \( n \in \mathbb{N} \).

Gaugeability requires a uniform gauge function \( g \) for all possible user volume vector sequences. (G3) and (G4) then imply that there is at least one tight group accounting for a negligible proportion of the total user volume as the total user volume approaches infinity, and that every OD pair has at least one tight strategy. A legal gaugeable function \( g \) of Γ thus must satisfy the condition that the limit

\[
\lim_{x \to +\infty} g(x)/x^2 \exists \quad \text{and} \quad (0, +\infty).
\]

In particular, this requires that the second (bottom) OD pair must be tight and

\[
\lim_{n \to \infty} d_2^{(n)} = \lim_{n \to \infty} d_2^{(n)}/T(d^{(n)}) > 0.
\]

However, our assumptions permit that

\[
\lim_{n \to \infty} d_2^{(n)} = 0.
\]

Thus, Γ is not gaugeable.

We now show that Γ is strongly scalable. Since \( d_1^{(n)} \cdot d_2^{(n)} = 0 \) and \( T(d^{(n)}) = d_1^{(n)} + d_2^{(n)} \to \infty \) as \( n \to \infty \), we can an infinite subsequence \( \{n_i\}_{i \in \mathbb{N}} \) such that either

- \( d_1^{(n_i)} \equiv 0 \) \( \forall i \in \mathbb{N} \), \( T(d^{(n_i)}) = d_2^{(n_i)} \to \infty \) as \( i \to \infty \) \quad \text{or} \quad (G.1)

- \( d_2^{(n_i)} \equiv 0 \) \( \forall i \in \mathbb{N} \), \( T(d^{(n_i)}) = d_1^{(n_i)} \to \infty \) as \( i \to \infty \) \quad \text{or} \quad (G.2)

The following will prove the strong scalability of Γ according to the two cases (G.1) and (G.2).

**Case I:** \( d_1^{(n_i)} \equiv 0 \) We define scaling factors \( g_n = T(d^{(n)})^2 \) for each \( n \in \mathbb{N} \). Then, the limit price functions of the first (top) OD pair equal the constant function \( 0 \), and the limit price functions of the second (bottom) OD pair equals \( 4x^2 \) and \( 5x^2 \) respectively. Thus, \( (SS1) \) holds. It is easy to check that \( (SS2) \) also holds w.r.t. the subsequence \( \{d^{(n_i)}\}_{i \in \mathbb{N}} \).
(Case II: $d_2^{(n_i)} \equiv 0$) We set the scaling factor $g_n = T(d^{(n)})$ for every $n \in \mathbb{N}$. Then, the limit price functions of the first (top) OD pairs are

$$\lim_{n \to +\infty} \frac{2(d_1^{(n)} + d_2^{(n)})x}{d_1^{(n)}} = 2x, \quad \text{and} \quad \lim_{n \to +\infty} \frac{3(d_1^{(n)} + d_2^{(n)})x}{d_1^{(n)}} = 3x.$$ 

The limit price functions of the second (bottom) OD pair equal $+\infty$. Thus, the second OD pair does not have tight strategies under this scaling sequence $\{g_n\}_{n \in \mathbb{N}}$. However, the second OD pair is negligible for this subsequence $\{d^{(n_i)}\}_{i \in \mathbb{N}}$ since $d_2^{(n_i)} \equiv 0$ for all $i \in \mathbb{N}$. Thus, (SS1)-(SS2) hold.

In summary, we can find a scaling sequence $\{g_n\}_{n \in \mathbb{N}}$ for any sequence $\{d^{(n)}\}_{n \in \mathbb{N}}$ of user volume vectors of $\Gamma$ such that (SS1)-(SS2) hold. Thus $\Gamma$ is strongly scalable. 

\[ \square \]

H. Proof of Example 3.1

Proof of Example 3.1 Let $\{d^{(n)}\}_{n \in \mathbb{N}}$ be a sequence of travel demand vectors with total demand $T(d^{(n)}) \to +\infty$ as $n \to +\infty$. Then, there is an infinite subsequence $\{n_q\}_{q \in \mathbb{N}}$ such that either

$$\forall q \in \mathbb{N} \exists i \in \mathbb{N} \quad T(d^{(n_q)}) \in [b_{2i}, b_{2i+1})$$

or

$$\forall q \in \mathbb{N} \exists i \in \mathbb{N} \quad T(d^{(n_q)}) \in [b_{2i+1}, b_{2i+2}).$$

To facilitate the discussion, we assume w.o.l.g. that $T(d^{(2i)}) \in [b_{2i}, b_{2i+1})$ for all $i \in \mathbb{N}$.

We now consider the subsequence $\{d^{(2i)}\}_{i \in \mathbb{N}}$. Note that

$$\overline{\lim}_{i \to +\infty} \frac{T(d^{(2i)})}{b_{2i}} =: \kappa,$$

for some constant $\kappa \geq 1$ or $+\infty$. We assume that

$$\lim_{i \to +\infty} \frac{T(d^{(2i)})}{b_{2i}} = \overline{\lim}_{i \to +\infty} \frac{T(d^{(2i)})}{b_{2i}} = \kappa.$$

Otherwise, we can take a suitable subsequence and this does not affect our further discussion. We will now construct a scaling sequence $\{g_i\}_{i \in \mathbb{N}}$ and the resulting two limit price functions for the subsequence $\{d^{(2i)}\}_{i \in \mathbb{N}}$.

There are three cases: $\kappa = 1$, $\kappa = +\infty$, and $\kappa \in (1, +\infty)$.

(Case 1: $\kappa = 1$) If $\kappa = 1$, then we obtain for any $x \in (0, 1)$ that $b_{2i-1} < xT(d^{(2i)}) < b_{2i}$ for $i$ large enough, since $b_{2i}/b_{2i-1} \to +\infty$ and $T(d^{(2i)})/b_{2i} \to 1$ as $i \to +\infty$. Therefore, for all $x \in (0, 1)$,

$$\lim_{i \to +\infty} \frac{\tau(T(d^{(2i)})x)}{\tau(b_{2i})} = \lim_{i \to +\infty} \frac{\tau(b_{2i-1})[(T(d^{(2i)})x - b_{2i-1}) + 1]}{\tau(b_{2i})}$$

$$= \lim_{i \to +\infty} \frac{\tau(b_{2i-1})[(T(d^{(2i)})x - b_{2i-1}) + 1]}{\tau(b_{2i-1})[(b_{2i} - b_{2i-1}) + 1]}$$

$$= \lim_{i \to +\infty} \frac{[(T(d^{(2i)})x - b_{2i-1}) + 1]}{(b_{2i} - b_{2i-1}) + 1} = x,$$
where we use again that \( \lim_{i \to \infty} b_{2i}/b_{2i-1} = +\infty \) and \( \kappa = \lim_{i \to \infty} T(d^{(2i)}/b_{2i} = 1 \). Therefore, we obtain with scaling factor \( g_i = \tau(b_{2i}) \) that \( \tau(T(d^{(2i)})x)/g_i \to x \) with \( i \to \infty \) for all \( x \in [0, 1] \), which, in turn, implies that the limit game is in WDG.

(Case 2: \( \kappa = +\infty \)) We can show similarly that \( \lim_{i \to +\infty} \tau(T(d^{(2i)})x)/\tau(T(d^{(2i)})) = 1 \) for all \( x \in [0, 1] \), since \( b_{2i} \leq T(d^{(2i)})x \leq T(d^{(2i)}) < b_{2i+1} \) for large enough \( i \) for all \( x \in (0, 1] \) in this case. Therefore, we put \( g_i = \tau(T(d^{(2i)})) \) in this case. Then \( l(x) \equiv 1 \) for all \( x \in [0, 1] \) and the limit game is in WDG.

(Case 3: \( \kappa \in (1, +\infty) \)) In this case, \( b_{2i} < T(d^{(2i)})x < T(d^{(2i)}) < b_{2i+1} \) for all \( i \) large enough for \( x \geq 1/\kappa \), and \( b_{2i-1} < T(d^{(2i)})x < b_{2i} < T(d^{(2i)}) < b_{2i+1} \) for all \( i \) large enough for \( 0 < x < \kappa \). For all \( x > 1/\kappa \), we obtain that

\[
\lim_{i \to +\infty} \frac{\tau(T(d^{(2i)})x)}{\tau(T(d^{(2i)}))} = \lim_{i \to +\infty} \frac{\tau(b_{2i-2})[(T(d^{(2i)})x - b_{2i-1}) + 1]}{\tau(b_{2i-2})[(b_{2i} - b_{2i-1}) + 1]} = \kappa \cdot x.
\]

For \( 0 < x < 1/\kappa \), we obtain that

\[
\lim_{i \to +\infty} \frac{\tau(T(d^{(2i)})x)}{\tau(T(d^{(2i)}))} = \frac{\tau(T(d^{(2i)})x)}{\tau(T(d^{(2i)}))}
\]

For \( x = 1/\kappa \), we use the fact that

\[
\frac{\tau(T(d^{(2i)})x_1)}{\tau(T(d^{(2i)}))} \leq \frac{\tau(T(d^{(2i)}))}{\tau(T(d^{(2i)}))} \leq \frac{\tau(T(d^{(2i)})x_2)}{\tau(T(d^{(2i)}))}
\]

for all \( x_1 \in [0, 1/\kappa) \) and \( x_2 \in (1/\kappa, 1] \), and obtain that

\[
\lim_{i \to +\infty} \frac{\tau(T(d^{(2i)})x)}{\tau(T(d^{(2i)}))} = 1.
\]

So, if we take \( g_i = \tau(T(d^{(2i)})) \), then

\[
l(x) = \begin{cases} 
\kappa \cdot x & \text{if } x \in [0, 1/\kappa), \\
1 & \text{if } x \in [1/\kappa, 1],
\end{cases}
\]

which is again continuous and nondecreasing. The limit game then has travel time functions \( 4l(x) \) and \( 3l(x) \), and PoA = 1. \( \square \)

1. **Proof of Theorem 3.5**

Proof of Theorem 3.5 Suppose that \( f^* \) is an SO-profile. Then, we obtain for any \( k = 1, \ldots, K \) and any \( s, s' \in S_k \) with \( f^*_s > 0 \) that

\[
\sum_{a \in A} r(a, s) \cdot (\tau_a(f^*_a) + f^*_a \cdot \tau'_a(f^*_a)) \leq \sum_{a \in A} r(a, s') \cdot (\tau_a(f^*_a) + f^*_a \cdot \tau'_a(f^*_a)),
\]

where \( \tau'_a(\cdot) \) is the first-order derivative of \( \tau_a(\cdot) \), see Roughgarden and Tardos (2002) for a proof of the above condition. Note that

\[
\tau_a(f^*_a) + f^*_a \cdot \tau'_a(f^*_a) = \gamma_a \cdot (f^*_a)\beta + \eta_a + \beta \cdot \gamma_a \cdot (f^*_a)\beta = (1 + \beta) \cdot \tau_a(f^*_a) - \beta \cdot \eta_a
\]

for any \( a \in A \). Therefore, we obtain further that

\[
(1 + \beta) \cdot \tau_s(f^*) - \beta \cdot \tau_s(0) \leq (1 + \beta) \cdot \tau_{s'}(f^*) - \beta \cdot \tau_{s'}(0),
\]
for any \( k = 1, \ldots, K \) and any \( s, s' \in \mathcal{S}_k \) with \( f^*_s > 0 \), where \( 0 = (0)_{p \in \mathcal{S}} \) represents the zero flow.

Regrouping terms, we obtain that

\[
\tau_s(f^*) \leq \tau_{s'}(f^*) \cdot \left( 1 + \frac{\beta \cdot (\tau_s(0) - \tau_{s'}(0))}{(1 + \beta) \cdot \tau_s(f^*)} \right) \leq \tau_{s'}(f^*) \cdot \left( 1 + \frac{\beta \cdot |\tau_s(0) - \tau_{s'}(0)|}{(1 + \beta) \cdot \tau_s(f^*)} \right),
\]

for any \( k = 1, \ldots, K \), any \( s, s' \in \mathcal{S}_k \) with \( f^*_s > 0 \).

Since \( |\mathcal{S}| < \infty \) and \( |A| < \infty \), there is a constant \( \kappa > 0 \) such that

\[
|\tau_s(0) - \tau_{s'}(0)| \leq \kappa,
\]

for any \( k = 1, \ldots, K \), any \( s, s' \in \mathcal{S}_k \), and thus

\[
\tau_s(f^*) \leq \tau_{s'}(f^*) \cdot \left( 1 + \frac{\beta \cdot \kappa}{(1 + \beta) \cdot \tau_s(f^*)} \right).
\]

It remains to bound \( \tau_{s'}(f^*) \) from below.

Observe that

\[
\max \{ f^*_{s'':} : s'' \in \mathcal{S}_k \} \geq \frac{d_k}{|\mathcal{S}|} \geq \frac{d_{\min}}{|\mathcal{S}|} \to \infty
\]

as \( d_{\min} \to \infty \), for any \( k = 1, \ldots, K \). For every \( k = 1, \ldots, K \), let \( s^*_k \in \mathcal{S}_k \) be the strategy such that

\[
f^*_k = \max \{ f^*_{s''} : s'' \in \mathcal{S}_k \} \in \Omega(d_{\min}).
\]

Then, for any \( k = 1, \ldots, K \),

\[
\tau_{s^*_k}(f^*) = \sum_{a \in A} r(a, s^*_k) \cdot \tau_a(f^*_a) = \sum_{a \in A} r(a, s^*_k) \cdot \left( \gamma_a \cdot \left( f^*_a \right)^\beta + \eta_a \right)
\]

\[
\geq \sum_{a \in A} r(a, s^*_k) \cdot \left( \sum_{l=1}^{K} \sum_{s \in \mathcal{S}_l} r(a, s) \cdot \left( f^*_s \right)^\beta + \eta_a \right) \in \Omega(d^\beta_{\min}).
\]

Therefore, for any \( k = 1, \ldots, K \), and any \( s' \in \mathcal{S}_k \), we obtain that

\[
\tau_{s'}(f^*) \geq \tau_{s^*_k}(f^*) + \beta \cdot \frac{\tau_{s'}(0) - \tau_{s^*_k}(0)}{1 + \beta} \in \Omega(d^\beta_{\min}).
\]

In summary, we obtain that

\[
\tau_s(f^*) \leq \tau_{s'}(f^*) \cdot \left( 1 + O\left( \frac{\beta \kappa}{\beta + 1} d^{-\beta}_{\min} \right) \right),
\]

(I.1)

for any \( k = 1, \ldots, K \), any \( s, s' \in \mathcal{S}_k \) with \( f_s > 0 \). This completes the proof.
J. Proof of Theorem 3.6

Proof of Theorem 3.6 Let \( f^* \) be an SO-profile, and \( \tilde{f} \) an NE-profile.

Note that the price functions \( \tau_a(\cdot) \) are convex. Because of the K.K.T. conditions, \( \tilde{f} \) is an optimal solution to the non-linear program (J.1).

\[
\begin{align*}
\text{minimize} & \quad \sum_{a \in A} \int_0^{f_a} \tau_a(t) dt \\
\text{s.t.} & \quad \sum_{s \in S_k} f_s = d_k, \text{ for all } k = 1, \ldots, K, \\
& \quad \sum_{k=1}^K \sum_{s \in S_k} r(a,s)f_s = f_a, \text{ for all } a \in A, \\
& \quad f_s \geq 0, \text{ for all } s \in S.
\end{align*}
\]

(J.1)

We then obtain that

\[
0 \geq \sum_{a \in A} \int_0^{f_a} \tau_a(t) dt - \sum_{a \in A} \int_0^{f_a^*} \tau_a(t) dt = \sum_{a \in A} \left[ \int_0^{f_a} \tau_a(t) dt - \int_0^{f_a^*} \tau_a(t) dt \right].
\]

(J.2)

Since \( \tau_a(x) = \gamma_a \cdot x^\beta + \eta_a \) for all \( a \in A \), we obtain for any feasible profile \( h = (h_s)_{s \in S} \) that

\[
\sum_{a \in A} \int_0^{h_a} \tau_a(t) dt = \sum_{a \in A} \frac{1}{\beta + 1} \tau_a(h_a) \cdot h_a + \frac{\beta}{\beta + 1} \sum_{a \in A} \tau_a(0) \cdot h_a
\]

\[
= \sum_{a \in A} \frac{1}{\beta + 1} \tau_a(h_a) \cdot h_a + \frac{\beta}{\beta + 1} \sum_{k=1}^K \sum_{s \in S_k} \tau_s(0) \cdot h_s
\]

\[
= \frac{1}{\beta + 1} C(h) + \frac{\beta}{\beta + 1} \sum_{k=1}^K \sum_{s \in S_k} \tau_s(0) \cdot h_s.
\]

(J.3)

Plugging (J.3) into (J.2) with \( h = f^* \) and \( h = \tilde{f} \) respectively yields

\[
0 \geq \frac{1}{\beta + 1} \left( C(\tilde{f}) - C(f^*) \right) + \frac{\beta}{\beta + 1} \left[ \sum_{k=1}^K \sum_{s \in S_k} \tau_s(0) \tilde{f}_s - \sum_{k=1}^K \sum_{s \in S_k} \tau_s(0) f^*_s \right],
\]

which in turn implies that

\[
1 \leq \text{PoA} \leq 1 + \frac{\beta}{C(f^*)} \left| \sum_{k=1}^K \sum_{s \in S_k} \tau_s(0) \tilde{f}_s - \sum_{k=1}^K \sum_{s \in S_k} \tau_s(0) f^*_s \right|.
\]

Now, we will show that

\[
\frac{\beta}{C(f^*)} \left| \sum_{k=1}^K \sum_{s \in S_k} \tau_s(0) \tilde{f}_s - \sum_{k=1}^K \sum_{s \in S_k} \tau_s(0) f^*_s \right| \in O(T^{-\beta}),
\]

(J.4)
which implies that \( \text{PoA} = 1 + O(T^{-\beta}) \).

We define
\[
L_{\text{max}} = \max\{\tau_s(0) : s \in S\}, \quad \text{and} \quad L_{\text{min}} = \min\{\tau_s(0) : s \in S\}.
\]
Obviously, \( L_{\text{max}} \) and \( L_{\text{min}} \) are constants independent of \( T = \sum_{k=1}^{K} d_k \), and
\[
\beta \left| \sum_{k=1}^{K} \sum_{s \in S_k} \tau_s(0) \tilde{f}_s - \sum_{k=1}^{K} \sum_{s \in S_k} \tau_s(0) f_s^* \right| \leq \beta (L_{\text{max}} - L_{\text{min}}) \sum_{k=1}^{K} d_k. \tag{J.5}
\]

Note that there is some \( s^* \in S \) such that \( f_s^* \geq \frac{\sum_{k=1}^{K} d_k}{|S|} = \frac{T}{|S|} \in \Omega(T) \). Note also that \( C(f^*) \geq f_s^* \tau_s^*(f^*) = f_s^* \cdot \sum_{a \in A} r(a, s^*) \tau_a(f_s^*) \). Since we assume (2.1), we obtain that
\[
\sum_{a \in A} r(a, s^*) \tau_a(f_s^*) \geq \gamma_{\text{min}} \cdot (f_s^*)^\beta + \eta_{\text{min}} \geq \gamma_{\text{min}} \frac{r_{\text{min}}^\beta \cdot T^\beta}{|S|^\beta} + \eta_{\text{min}},
\]
where \( \gamma_{\text{min}} = \min\{\gamma_a : a \in A\} > 0 \), \( \eta_{\text{min}} = \min\{\eta_a : a \in A\} > 0 \), \( r_{\text{min}} = \min\{r(a, s) : a \in A, s \in S \text{ with } r(a, s) > 0\} > 0 \). Observe also the fact that \( f_a^* \geq r(a, s^*) f_s^* \) for any \( a \in A \) with \( r(a, s^*) > 0 \). Therefore, we obtain that
\[
C(f^*) \geq \gamma_{\text{min}} \frac{r_{\text{min}}^\beta \cdot T^\beta}{|S|^\beta + 1}. \tag{J.6}
\]

With (J.5), (J.6), and the fact that \( T = \sum_{k=1}^{K} d_k \), we finally obtain that
\[
\frac{\beta \left| \sum_{k=1}^{K} \sum_{s \in S_k} \tau_s(0) \tilde{f}_s - \sum_{k=1}^{K} \sum_{s \in S_k} \tau_s(0) f_s^* \right|}{C(f^*)} \in O(T^{-\beta}),
\]
which completes the proof. \( \square \)

K. Proof of Theorem 3.7

Proof of Theorem 3.7 a) We will only prove a) for the sequence \( \{\tilde{f}(n)\}_{n \in \mathbb{N}} \) of NE-profiles. An almost identical argument applies to the sequence \( \{f^*(n)\}_{n \in \mathbb{N}} \) of SO-profiles. We first claim that the distribution vector \( \tilde{f}(n) \) of NE-profiles converges to a limit distribution \( \tilde{f} = (\tilde{f}_s)_{s \in S} \).

Claim 1. There exists a limit distribution \( \tilde{f} = (\tilde{f}_s)_{s \in S} \) such that
\[
\tilde{f}(n) \to \tilde{f} \quad \text{as} \quad n \to +\infty.
\]

Proof of Claim 1 We only need to prove the convergence of \( \tilde{f}_s(n) \) for an arbitrarily fixed \( s \in S \). Since the sequence \( \{\tilde{f}_s(n)\}_{n \in \mathbb{N}} \) is bounded, there exist two infinite subsequences \( \{n_i\}_{i \in \mathbb{N}} \) and \( \{m_i\}_{i \in \mathbb{N}} \) such that
\[
\tilde{f}_s = \lim_{i \to \infty} \tilde{f}_s(n_i) = \lim_{n \to \infty} \tilde{f}_s(n) \leq \lim_{n \to \infty} \tilde{f}_s(n) = \lim_{i \to \infty} \tilde{f}_s(m_i) = \tilde{f}_s.
\]
For the subsequence \( \{n_i\}_{i \in \mathbb{N}} \), there exists an infinite subsequence \( \{n_{ij}\}_{j \in \mathbb{N}} \), such that the \( \tilde{f}_s(n_{ij}) \) converges to a distribution vector \( \tilde{f}_s \) again because of the boundedness. Therefore, we assume,
w.o.l.g., that \( \tilde{f}_s^{(n)} \) itself converges to \( f \), and \( \hat{f}_s^{(m)} \) converges to a distribution vector \( \mathcal{F} \) respectively, otherwise we can take subsequences fulfilling this assumption.

Now we consider the two limit games with scaling sequences \( g_i = (T(d^{(n_i)}))^\beta \) and \( h_i = (T(d^{(m_i)}))^\beta \) respectively. Note that these two limit games coincide under scalings \( (T(d^{(n_i)}))^\beta \) and \( (T(d^{(m_i)}))^\beta \), respectively, since they have the same demand vector \( d \), the same resource set \( A \), and also the same travel time functions \( l_a(f_a) = \gamma_a \cdot f_a^\beta \) for all \( a \in A \). Since the \( l_a(f_a) \) are strictly convex, the common limit game has a unique NE-profile. We have shown in the proof of Theorem 3.2 that if the game is scalable and the distribution sequence corresponding to NE-profiles converges, then the limit of the distribution sequence is an NE-profile of the limit game. Therefore, both \( f \) and \( \mathcal{F} \) are NE-profiles of the limit game, which, by the uniqueness argument, implies that \( f = \mathcal{F} \). Then \( \mathcal{F}_s = f_s \). Therefore \( f_s^{(n)} \) converges.

Let \( L(d) > 0 \) denote the social cost of the unique NE-profile in the limit game with demand vector \( d \). Note that \( L(d) \) depends only the demand vector \( d \). Since \( \{d^{(n)}\}_{n \in \mathbb{N}} \) and \( \{\tilde{f}^{(n)}\}_{n \in \mathbb{N}} \) converge and the limit game is in WDG, we obtain from the proof of Theorem 3.2 that

\[
\lim_{n \to \infty} \frac{C(\tilde{f}^{(n)})}{(T(d^{(n)}))^{\beta+1}} = L(d) > 0.
\]

b) Note that \( \{f^{*(n)}\}_{n \in \mathbb{N}} \) also converges as \( n \to +\infty \). This can be proved by a similar argument to that for Claim 1. Since the limit game has strictly convex travel time functions \( l_a(f_a) = f_a^\beta \) and is in WDG, we obtain that both \( \tilde{f}^{(n)} \) and \( f^{*(n)} \) converge to the same limit distribution \( f \) that is the unique NE-profile (also the unique SO-profile) of the limit game. Altogether, this proves Theorem 3.7. \( \Box \)

Acknowledgments

The authors would like to thank Mr. C. Hansknecht from TU Braunschweig for his help in installing the package cmcf!

References

Bingham NH, Goldie CM, Teugels JL (1987) Regular variation (Cambridge University Press).

Cole R, Dodis Y, Roughgarden T (2003) Pricing network edges for heterogeneous selfish users. Proceedings of the Annual ACM Symposium on the Theory of Computing 521–530.

Colini-Baldeschi R, Cominetti R, Mertikopoulos P, Scarsini M (2017) On the asymptotic behavior of the price of anarchy: Is selfish routing bad in highly congested networks? ArXiv:1703.00927v1 [cs.GT].

w.o.l.g., that \( \tilde{f}_s^{(n)} \) itself converges to \( f \), and \( \hat{f}_s^{(m)} \) converges to a distribution vector \( \mathcal{F} \) respectively, otherwise we can take subsequences fulfilling this assumption.

Now we consider the two limit games with scaling sequences \( g_i = (T(d^{(n_i)}))^\beta \) and \( h_i = (T(d^{(m_i)}))^\beta \) respectively. Note that these two limit games coincide under scalings \( (T(d^{(n_i)}))^\beta \) and \( (T(d^{(m_i)}))^\beta \), respectively, since they have the same demand vector \( d \), the same resource set \( A \), and also the same travel time functions \( l_a(f_a) = \gamma_a \cdot f_a^\beta \) for all \( a \in A \). Since the \( l_a(f_a) \) are strictly convex, the common limit game has a unique NE-profile. We have shown in the proof of Theorem 3.2 that if the game is scalable and the distribution sequence corresponding to NE-profiles converges, then the limit of the distribution sequence is an NE-profile of the limit game. Therefore, both \( f \) and \( \mathcal{F} \) are NE-profiles of the limit game, which, by the uniqueness argument, implies that \( f = \mathcal{F} \). Then \( \mathcal{F}_s = f_s \). Therefore \( f_s^{(n)} \) converges.

Let \( L(d) > 0 \) denote the social cost of the unique NE-profile in the limit game with demand vector \( d \). Note that \( L(d) \) depends only the demand vector \( d \). Since \( \{d^{(n)}\}_{n \in \mathbb{N}} \) and \( \{\tilde{f}^{(n)}\}_{n \in \mathbb{N}} \) converge and the limit game is in WDG, we obtain from the proof of Theorem 3.2 that

\[
\lim_{n \to \infty} \frac{C(\tilde{f}^{(n)})}{(T(d^{(n)}))^{\beta+1}} = L(d) > 0.
\]

b) Note that \( \{f^{*(n)}\}_{n \in \mathbb{N}} \) also converges as \( n \to +\infty \). This can be proved by a similar argument to that for Claim 1. Since the limit game has strictly convex travel time functions \( l_a(f_a) = f_a^\beta \) and is in WDG, we obtain that both \( \tilde{f}^{(n)} \) and \( f^{*(n)} \) converge to the same limit distribution \( f \) that is the unique NE-profile (also the unique SO-profile) of the limit game. Altogether, this proves Theorem 3.7. \( \Box \)

Acknowledgments

The authors would like to thank Mr. C. Hansknecht from TU Braunschweig for his help in installing the package cmcf!

References

Bingham NH, Goldie CM, Teugels JL (1987) Regular variation (Cambridge University Press).

Cole R, Dodis Y, Roughgarden T (2003) Pricing network edges for heterogeneous selfish users. Proceedings of the Annual ACM Symposium on the Theory of Computing 521–530.

Colini-Baldeschi R, Cominetti R, Mertikopoulos P, Scarsini M (2017) On the asymptotic behavior of the price of anarchy: Is selfish routing bad in highly congested networks? ArXiv:1703.00927v1 [cs.GT].
Colini-Baldeschi R, Cominetti R, Scarsini M (2016) On the price of anarchy of highly congested nonatomic network games. *International Symposium on Algorithmic Game Theory*, 117–128 (Springer, Lecture Notes in Computer Science 9928).

Correa JR, Schulz AS, Stier-Moses NE (2005) On the inefficiency of equilibria in congestion games. extended abstract. *Integer Programming and Combinatorial Optimization, International IPCO Conference, Berlin, Germany, June 8-10, 2005, Proceedings*, 167–181 (Springer, Lecture Notes in Computer Science 3509).

Dijkstra EW (1959) A note on two problems in connexion with graphs. *Numerische Mathematik* 1:269–271.

Fleischer L, Jain K, Mahdian M (2004) Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games. *IEEE Symposium on Foundations of Computer Science, 2004. Proceedings*, 277–285.

Fukushima M (1984) A modified Frank-Wolfe algorithm for solving the traffic assignment problem. *Transportation Research Part B Methodological* 18(2):169–177.

Harks T, Kleinert I, Klimm M, Möhring RH (2015) Computing network tolls with support constraints. *Networks* 65(3):262–285.

Jahn O, Möhring RH, Schulz AS, Stier-Moses NE (2005) System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations Research* 53(4):600–616.

Nisan N, Roughgarden T, Tardos E, Vaz VV (2007) *Algorithmic game theory* (Cambridge University Press).

Bureau of Public Roads (1964) *Traffic assignment manual* (U.S. Department of Commerce, Urban Planning Division, Washington, D.C.).

O’Hare SJ, Connors RD, Watling DP (2016) Mechanisms that govern how the price of anarchy varies with travel demand. *Transportation Research Part B Methodological* 84:55–80.

Papadimitriou C (2001) Algorithms, games, and the internet. *International Colloquium on Automata, Languages, and Programming*, 1–3 (Springer, Lecture Notes in Computer Science 2076).

Phang SY, Toh RS (2004) Road congestion pricing in singapore: 1975 to 2003. *Transportation Journal* 43(2):16–25.
Rosenthal RW (1973) A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory* 2(1):65–67.

Roughgarden T (2003) The price of anarchy is independent of the network topology. *Journal of Computer & System Sciences* 67(2):341–364.

Roughgarden T, Tardos E (2002) How bad is selfish routing? *Journal of the ACM* 49:236–259.

Roughgarden T, Tardos E (2004) Bounding the inefficiency of equilibria in nonatomic congestion games. *Games & Economic Behavior* 47(2):389–403.

Roughgarden T, Tardos E (2007) Introduction to the inefficiency of equilibria. *Algorithmic game theory*, 443–459 (Cambridge University Press, Cambridge).

Schmeidler D (1973) Equilibrium points of nonatomic games. *Journal of Statistical Physics* 7(4):295–300.

Smith MJ (1979) The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B Methodological* 13(4):295–304.

Wardrop JG (1952) Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II* 1:325–378.