Department of Economics and Management
University of Brescia
Italy

WORKING PAPER

System Optimal Routing of Traffic Flows with User Constraints using Linear Programming

Enrico Angelelli
Valentina Morandi
Martin Savelsbergh
Grazia Speranza

WPDEM 5/2016

Via S. Faustino 74/b, 25122 Brescia – Italy
Tel 00390302988742 Fax 00390302988703
email: paqricerca.dem@unibs.it

System optimal routing of traffic flows with user constraints using linear programming

Enrico Angelelli (1) *Valentina Morandi* (2) *Martin Savelsbergh* (3) *Grazia Speranza* (4)

(1) *Department of Economics and Management, University of Brescia, Italy, enrico.angelelli@unibs.it*

(2) *Department of Economics and Management, University of Brescia, Italy, valentina.morandi1@unibs.it*

(3) *H. Milton Stewart School of Industrial & Systems Engineering, Georgia Tech Atlanta, USA,martin.savelsbergh@isye.gatech.edu*

(4) *Department of Economics and Management, University of Brescia, Italy, grazia.speranza@unibs.it*

Abstract

For static traffic assignment problems, it is well-known that (1) the total travel time in a user-equilibrium solution can be substantially higher than the total travel time in a system-optimum solution, and (2) the user-experienced travel time in a system-optimum solution can be substantially higher that the user-experienced travel time in a user-equilibrium solution. By seeking system optimal traffic flows subject to user constraints, a comprise solution can be obtained that balances system and user objectives. We propose a linear programming based approach to efficiently obtain a solution that effectively balances system and user objectives. A computational study reveals that solutions with near-optimal total travel times can be found in which most users experience travel times that are better than user-equilibrium travel times and few users experience travel times that are slightly worse than user-equilibrium travel times.

Keywords: Traffic assignment, congestion, latency function, constrained system optimum, linear programming, piecewise linear approximation

1 Introduction

Road congestion is one of the most pressing problems in urban areas, as it has negative economic, environmental, and health impacts. Since in most situations, road network infrastructure expansion is not feasible, because of a lack of available space or the huge costs, alternative solutions have to be explored. Route guidance is natural and popular choice. Many vehicles, nowadays, are equipped with navigational devices that can compute a shortest route (based on user preferences) from an origin to a destination. The

latest versions of these devices can also display current traffic conditions and use current traffic data on shortest route computations. Unfortunately, these devices do not consider the potential impact of the directions provided to a driver on the traffic system. As a consequence, the same route may be suggested to many users, which, if the recommended routes are followed by the users, may simply result in congestion occurring in other parts of the road network rather than a reduction in system-wide congestion or in avoiding an increase in system-wide congestion. Coordinating route guidance across the users of the system, and adherence to the suggested routes by the users, is needed to reduce congestion or, even better, to avoid congestion. In the (near) future, with the introduction of self-driving vehicles, this may become a reality. Our research focuses on optimization models to support these developments.

Route choice, or traffic assignment, concerns the selection of routes (alternatively called paths) between origins and destinations in transportation networks. It is the fourth step in the conventional transportation forecasting model, following trip generation, trip distribution, and mode choice. Traffic assignment problems are defined on a road network with an origin-destination (OD) matrix specifying demand, i.e, the number of vehicles per time unit that is expected to travel from each origin to each destination. Details on forecasting demand on road networks can be found in Sheffi (1985), Ben-Akiva and Lerman (1985), Florian and Hearn (1999) and de Dios Ortuzar and Willumsen (2011).

A distinction is made between static and dynamic traffic assignment problems. In a dynamic traffic assignment problem, origin-destination demand varies over time. Early studies on dynamic traffic assignment date back to Merchant and Nemhauser (1978). However, this class of traffic assignment problems is very challenging Jahn et al. (2005) and there is no generally accepted model or methodology Papageorgiou (1990); Mahmassani and Peeta (1995); Peeta and Ziliaskopoulos (2001); Ben-Akiva et al. (2012). In a static traffic assignment problem, on the other hand, origin-destination demand is assumed to be constant over time. Sheffi (1985) points out that this assumption is realistic during a rush hour period when traffic exhibits a steady-state behavior. This class of problems has been widely studied (see Sheffi (1985) and Patriksson (2015) for references) and is also the focus of our research.

In his seminal work, Wardrop (1952) introduced two, now famous, principles that describe the characteristics of the so-called user and system equilibrium. In the user equilibrium, the journey times along all routes used from an origin to a destination are equal and less than the journey time that would be experienced by a single user on any unused route, i.e., no user can lower his transportation cost through unilateral action. In the system equilibrium the average journey time is minimum, i.e., the sum of the journey times experienced

by all users is as small as possible.

To compute journey times, it is critical to be able to compute the experienced travel time on an arc, which, of course, depends on the traffic flow on the arc. To do so, a *latency function* $t_a(x)$ is used, which gives the experienced travel time on arc a given traffic flow x. The most popular latency functions are the Davidson's function $t_a(x) = t_a^{FF}(1 + \frac{\alpha x_a}{u'_a - x_a})$, where t_a^{FF} is the arc free-flow travel time and u'_a and α are tuning parameters, and the U.S. Bureau of Public Road function $t_a(x) = t_a^{FF}[1 + 0.15(\frac{x_a}{u_a})^4]$, where t_a^{FF} represents the arc free-flow travel time and u_a represents the maximum rate of vehicles that can enter an arc without experiencing substantial delays due to congestion.

Differences between user equilibrium and system optimum traffic assignments are extensively studied in literature (see, for example, Mahmassani and Peeta (1993)). The user equilibrium ensures fairness for users traveling between the same origin and destination. On the other hand, in a system optimum traffic assignment, some users may be assigned to paths that are much longer, in terms of distance or travel time, than paths assigned to other users traveling from the same origin to the same destination.

Controlling the unfairness that is inherent in a system optimum traffic assignment is critical when seeking an efficient and implementable traffic assignment. This has lead several researchers to explore system optimal routing of traffic flows with user constraints, in which only paths that ensure a certain level of fairness among users with the same origin and destination are considered. In order to generate a set of paths that is likely to result in an acceptable level of fairness, the concept of *normal length* of an arc is introduced, an a priori estimate of the traversal time of an arc. Natural choices for the normal length of an arc are its distance, its free-flow traversal time, or its traversal time in the user equilibrium. The distance and free-flow traversal time are normal length measures that do not depend on the demand in the network, while the traversal time in the user equilibrium does depend on the demand in the network. The set of paths generated for an origin-destination pair contains all paths that have a travel time that is within a certain percentage of a shortest travel time path evaluated using the normal length.

To the best of our knowledge, the first attempt at computing a constrained system optimum is due to Jahn et al. (2000). Their model seeks to minimize the total (experienced) travel time, uses the Davidson latency function, and restricts the set of paths considered by using the geographic distance as the normal length of an arc. The solution method is based on the Frank-Wolfe algorithm. In a subsequent paper, the latency function is replaced by the one provided by the U.S. Bureau of Public Road and the traversal time in the user equilibrium is used as the normal length of an arc Jahn et al. (2005). The computational results,

on data for seven real road networks, show that total travel time close to the system optimum total travel time can be achieved with travel times experienced by users close to the travel times experienced by users in the user equilibrium traffic assignment. A more theoretical assessment of the efficiency and fairness of constrained system optimum traffic assignments is presented in Schulz and Stier-Moses (2006). Recently, Lujak et al. (2015) investigated a model in which the weighted geometric mean of the experienced travel times of the used paths is minimized. A generic latency function is used and the set of paths considered is restricted by using the free-flow traversal time as the normal length of an arc. The solution method is a multi-agent negotiation model. Correa et al. (2007) study the problem of minimizing the maximum experienced travel time on any used path in road networks with congestion, i.e., road networks with flow-dependent arc traversal times. They compare the resulting traffic assignment with the system optimum and user equilibrium traffic assignments in terms of the average experienced travel time on any used path (which is equivalent to the total experienced travel time) and the unfairness, i.e., the maximum ratio of the largest experienced travel time and the smallest experience travel time for an origin-destination pair. Note that in this setting there are no restrictions on the set of paths considered. Angelelli et al. (2016) present a linear programming model in which user inconvenience is minimized while keeping the network non-congested, if possible, or at its minimum congestion level, otherwise. They assume a constant, flow-independent, latency function and measure user inconvenience as the ratio of the travel time of the user's path and the fastest travel time from the user's origin to the user's destination. The network is considered non-congested if the flow on every arc is less than or equal to the maximum rate of vehicles that can enter the arc without experiencing substantial delays. The computed user inconvenience is accurate only when the network can be kept non-congested. The set of paths considered is restricted by using the free-flow traversal time as the normal length of an arc. The advantage of their proposed model is that it can be solved with any (commercial) linear programming solver. The disadvantage is that it assumes constant, flow-independent, arc traversal times.

In this paper, a linear programming model is presented, thus retaining the desirable property that it can be solved efficiently by any (commercial) linear programming solver, that accommodates a flow-dependent latency function, thereby removing the unwanted limitation that user inconvenience is only computed accurately when the system is non-congested. The model seeks to minimize the total (experienced) travel time, using the latency function provided by the U.S. Bureau of Public Road, and restricts the set of paths considered by using the free-flow traversal time as the normal length of an arc. A piecewise linear approximation of the flow-dependent latency function ensures the linearity of the model. Thus, the model assigns paths

to users so as to minimize the total travel time experienced, but uses only paths that are expected to have a limited inconvenience. An extensive computational study demonstrates that solutions with near-optimal total travel times can be found in which most users experience travel times that are better than user-equilibrium travel times and few users experience travel times that are slightly worse than user-equilibrium travel times.

The remainder of the paper is organized as follows. In Section 2, we introduce and discuss the relevant traffic assignment models. In Section 3, we introduce the solution methods developed to solve these traffic assignments models. In Section 4, we discuss the results of an extensive study of the constrained system optimum model with a piecewise linear approximation of the latency function. Finally, in Section 5, we present some concluding remarks.

2 Traffic assignment models

In this section, we present the mathematical models that form the basis of this study. We consider a directed network $G = (V,A)$, where V represents the set of vertices and $A \subseteq V \times V$ represents the set of arcs. Each arc $(i,j) \in A$ represents a road segment on which vehicles can travel. A latency function $t_{ij}(x)$, is defined for each arc $(i,j) \in A$ representing the arc traversal time when the rate of vehicles entering the arc is x. The free-flow traversal time of an arc $(i,j) \in A$ is denoted by t_{ij}^{FF} and represents the traversal time experienced when no other vehicles enter the arc $t_{ij}^{FF} = t_{ij}(0)$. In addition, there is a set $C \subseteq V \times V$ of commodities, i.e., represented by origin-destination (OD) pairs, where each $c \in C$ has an origin $O_c \in V$, a destination $D_c \in V$, and a demand rate d_c, representing the number of vehicles per time unit that travel from O_c to D_c. The set of all the possible paths between origin O_c and destination D_c is denoted by K_c. The indicator a_{ij}^{kc} takes value 1 if path $k \in K_c$ contains arc $(i,j) \in A$ and takes value 0 otherwise. The decision variables y_c^k represent the flow of commodity $c \in C$ routed on path $k \in K_c$. The variables x_{ij} represent the (total) flow on arc $(i,j) \in A$.

2.1 The user equilibrium

The user equilibrium model (UE MODEL), as formulated in Beckmann et al. (1956), is the following:

$$
\begin{aligned}
\min \quad & \sum_{(i,j)\in A} \int_0^{x_{ij}} t_{ij}(\omega)\, d\omega & & \\
& x_{ij} = \sum_{c\in C} \sum_{k\in K_c} a_{ij}^{kc} y_{ck} & \forall (i,j) \in A & \quad (1) \\
& d_c = \sum_{k\in K_c} y_{ck} & \forall c \in C & \quad (2) \\
& x_{ij} \geq 0 & \forall (i,j) \in A & \quad (3) \\
& y_{ck} \geq 0 & \forall c \in C \quad \forall k \in K_c. & \quad (4)
\end{aligned}
$$

Constraints (1) set the flow on an arc as the sum of the flow on each path passing through the arc. Constraints (2) ensure that the demand d_c of commodity $c \in C$ is routed on paths in K_c. Finally, constraints (3) - (4) define the domains of the decision variables. As pointed out in Sheffi (1985), the objective function has no direct interpretation but is constructed in such a way that the solution to the model is equivalent to the user equilibrium.

2.2 The system optimum

The system optimum model (PATH-BASED SO MODEL), as formulated in Beckmann et al. (1956), is the following:

$$
\begin{aligned}
\min \quad & \sum_{(i,j)\in A} t_{ij}(x_{ij}) x_{ij} \\
& (1) - (4).
\end{aligned}
$$

The objective function minimizes the total travel time experienced by the users while the constraints are the same as in the user equilibrium model formulation. The equivalent arc-based formulation (ARC-BASED

SO MODEL) is the following:

$$\min \quad \sum_{(i,j)\in A} t_{ij}(x_{ij})x_{ij}$$

$$x_{ij} = \sum_{c\in C} x^c_{ij} \qquad \forall (i,j) \in A \tag{5}$$

$$\sum_{j\in V} x^c_{ij} - \sum_{j\in V} x^c_{ji} = 0 \qquad \forall c \in C \quad \forall i \in V, i \neq O_c, i \neq D_c \tag{6}$$

$$\sum_{j\in V} x^c_{ij} - \sum_{j\in V} x^c_{ji} = d_c \qquad \forall c \in C \quad i = O_c \tag{7}$$

$$\sum_{j\in V} x^c_{ij} - \sum_{j\in V} x^c_{ji} = -d_c \qquad \forall c \in C \quad i = D_c \tag{8}$$

$$x^c_{ij} \geq 0 \quad \forall c \in C \qquad \forall (i,j) \in A. \tag{9}$$

Variables x^c_{ij} represent the flow of commodity c on arc (i,j), Constraints (5) impose that the total flow on arc (i,j) is the sum of the flows of all commodities $c \in C$ passing through it, and constraints (6) - (8) guarantee flow conservation and that demand is completely routed. Finally, constraints (9) define the domain of the variables x^c_{ij}.

Note that the user equilibrium is a feasible solution to the system optimum model, and, therefore, the total travel time experienced in a system optimum is less than or equal to the total travel time experienced in a user equilibrium. The difference between the total travel times of the two traffic assignments is commonly referred to as the *price of anarchy* and reflects how selfish routing (user equilibrium) affects the total travel time when compared to a system-wide coordinated traffic assignment (system optimum).

2.3 The constrained system optimum

Like the system optimum model, the constrained system optimum model (C-SO MODEL) assigns paths to commodities so as to minimize the total travel time experienced by users. However, the set of paths considered is restricted as follows. We define the path *inconvenience* to be the relative percentage difference between the normal length of the path and the shortest normal length path from the same origin to the same destination. For a given percentage γ, the set of paths K^γ_c for commodity $c \in C$ contains all the paths with an inconvenience not larger than γ. We refer to γ as the *maximum inconvenience*. Recall that the normal length

of an arc (and thus a path) can be defined in a number of different ways. The C-SO MODEL formulation is the following:

$$\min \quad \sum_{(i,j)\in A} t_{ij}(x_{ij})x_{ij}$$

$$x_{ij} = \sum_{c\in C}\sum_{k\in K_c^\gamma} a_{ij}^{kc} y_{ck} \qquad \forall (i,j)\in A \tag{10}$$

$$d_c = \sum_{k\in K_c^\gamma} y_{ck} \qquad \forall c\in C \tag{11}$$

$$x_{ij} \geq 0 \qquad \forall (i,j)\in A \tag{12}$$

$$y_{ck} \geq 0 \qquad \forall c\in C \quad \forall k\in K_c^\gamma. \tag{13}$$

The only difference with PATH-BASED SO MODEL is that the set of path K_c^γ for commodity $c \in C$ is a subset of K_c.

3 Solution methods

In this section, we present the solution methods for the models that form the basis of this study.

3.1 Computing the user equilibrium

There are several ways to compute the UE on a network. In Sheffi (1985) and Patriksson (2015) heuristic methods, such as the capacity restraint method and the incremental assignment method, as well as exact methods are proposed. We implemented the following convex combination method to obtain the UE:

Step 0: Perform an "all-or-nothing" assignment based on the free-flow traversal times $t_{ij}^0 = t_{ij}^{FF}$, $(i,j)\in A$, i.e., route the entire demand of a commodity on the fastest path with respect to t_{ij}^{FF}. The resulting flows are represented by $\{x_{ij}^1\}$. Set $n = 1$.

Step 1: Update the current traversal times based on the flows $\{x_{ij}^n\}$ and the latency function, i.e., $t_{ij}^n = t_{ij}(x_{ij}^n)$, $(i,j)\in A$.

Step 2: Perform an "all-or-nothing" assignment based on the current traversal times t_{ij}^n. The resulting flows

are denoted by $\{y_{ij}^n\}$. This gives a descent direction.

Step 3: Perform a line search, i.e., find the δ_n that solves

$$\begin{aligned} \min \quad & \sum_{(ij)\in A} \int_0^{x_{ij}^n+\delta_n(y_{ij}^n-x_{ij}^n)} t_{ij}(\omega)\, d\omega \\ & 0 \le \delta_n \le 1 \\ & \delta_n \ge 0. \end{aligned}$$

Since we have assumed that the arc latency function is $t_{ij}(x_{ij}) = t_{ij}^{FF}[1+0.15(\frac{x_{ij}}{u_{ij}})^4]$, the resulting one variable optimization problem is

$$\begin{aligned} \min \quad & \sum_{(ij)\in A} \{t_{ij}^{FF}[x_{ij}^n+\delta_n(y_{ij}^n-x_{ij}^n)] + \frac{0.15}{5u_{ij}^4}[x_{ij}^n+\delta_n(y_{ij}^n-x_{ij}^n)]^5\} \\ & 0 \le \delta_n \le 1 \\ & \delta_n \ge 0, \end{aligned}$$

which can be solved easily (note that t_{ij}^{FF}, x_{ij}^n and y_{ij}^n are fixed).

Step 4: Compute flows $\{x_{ij}^{n+1}\}$, i.e., $x_{ij}^{n+1} = x_{ij}^n + \delta_n(y_{ij}^n - x_{ij}^n)$, $(i,j) \in A$.

Step 5: Convergence test. Let l_c^n be the travel time for commodity c at the n-th iteration. If $\sum_{c\in C} \frac{|l_c^n - l_c^{n-1}|}{l_c^n} \le \kappa$, then STOP, otherwise set $n = n+1$ and return to Step 1.

3.2 Piecewise linear approximation

Since we want to take advantage of the enormous power of modern (commercial) linear programming solvers, we will replace the objective function in the system optimum models, when it is nonlinear, with a piecewise linear approximation. More specifically, we replace the terms $F_{ij}(x_{ij}) = t_{ij}(x_{ij})x_{ij}$ in the objective function of ARC-BASED SO MODEL and C-SO MODEL by a piecewise linear convex function on $[0, U_{ij}]$, where U_{ij} is a natural upper bound on the admissible flow x_{ij} through arc (i,j).

Let

$$B = \{b_{ij}^0 = 0, b_{ij}^1, ..., b_{ij}^{n-1}, b_{ij}^n = U_{ij}\}$$

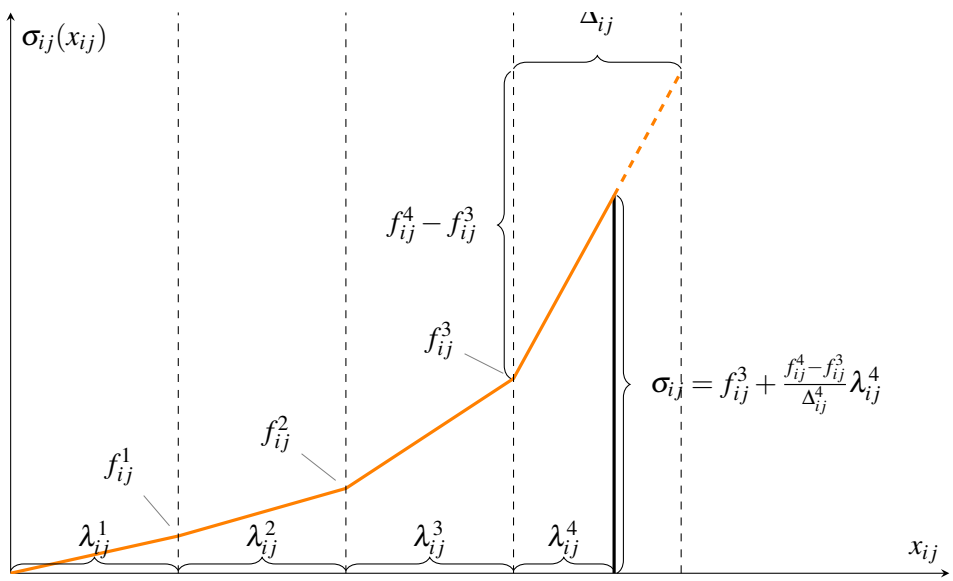


Figure 1. $\sigma_{ij}(x_{ij})$ construction

be a set of breakpoints that partition the domain $[0, U_{ij}]$ of function F_{ij} into n intervals, and let

$$F = \{f_{ij}^0 = F_{ij}(0) = 0, f_{ij}^1 = F_{ij}(b_{ij}^1), ..., f_{ij}^n = F_{ij}(U_{ij})\}$$

be the set of corresponding function values. Furthermore, let $\Delta_{ij}^h = b_{ij}^h - b_{ij}^{h-1}$ for $h = 1, \ldots, n$ be the length of the h-th interval. The piecewise linear approximation $\sigma_{ij}(x)$ of $F_{ij}(x)$ is given by (see Figure 1)

$$F_{ij}(x) \approx \sigma_{ij}(x) = \begin{cases} f_{ij}^0 + \frac{f_{ij}^1 - f_{ij}^0}{\Delta_{ij}^1}(x - b_{ij}^0) & x \in [b_{ij}^0, b_{ij}^1] \\ f_{ij}^1 + \frac{f_{ij}^2 - f_{ij}^1}{\Delta_{ij}^2}(x - b_{ij}^1) & x \in (b_{ij}^1, b_{ij}^2] \\ \ldots\ldots & \ldots. \\ f_{ij}^{n-1} + \frac{f_{ij}^n - f_{ij}^{n-1}}{\Delta_{ij}^n}(x - b_{ij}^{n-1}) & x \in (b_{ij}^{n-1}, b_{ij}^n]. \end{cases}$$

When the latency functions are convex, as is the case with the Davidson and the US Bureau of Public Road latency functions, then $F_{ij}(x)$ and its approximation $\sigma_{ij}(x)$ are convex as well. The convexity of $\sigma_{ij}(x)$ functions allows them to be easily expressed in a linear programming model.

This results in the following LIN-SO MODEL:

$$
\begin{aligned}
\min \quad & \sum_{(ij)\in A} \sigma_{ij} \\
& x_{ij} = \sum_{h=1}^{n} \lambda_{ij}^{h} \qquad \forall (i,j) \in A && (14)\\
& \sigma_{ij} = \sum_{h=1}^{n} \frac{f_{ij}^{h} - f_{ij}^{h-1}}{\Delta_{ij}^{h}} \lambda_{ij}^{h} \qquad \forall (i,j) \in A && (15)\\
& 0 \le \lambda_{ij}^{h} \le \Delta_{ij}^{h} \qquad \forall (i,j) \in A \quad \forall h = 1,\ldots,n && (16)\\
& x_{ij} = \sum_{c\in C} x_{ij}^{c} \qquad \forall (i,j) \in A && (17)\\
& \sum_{j\in V} x_{ij}^{c} - \sum_{j\in V} x_{ji}^{c} = 0 \qquad \forall c \in C \quad \forall i \in V, i \neq O_c, i \neq D_c && (18)\\
& \sum_{j\in V} x_{ij}^{c} - \sum_{j\in V} x_{ji}^{c} = d_c \qquad \forall c \in C \quad i = O_c && (19)\\
& \sum_{j\in V} x_{ij}^{c} - \sum_{j\in V} x_{ji}^{c} = -d_c \qquad \forall c \in C \quad i = D_c && (20)\\
& x_{ij}^{c} \ge 0 \quad \forall c \in C \qquad \forall (i,j) \in A \\
& x_{ij} \ge 0 \quad \forall (i,j) \in A.
\end{aligned}
$$

Variable x_{ij}^{c} represents the flow of commodity c on arc (i,j) and variable x_{ij} represents to total flow on arc (i,j). They are linked through constraints (14). Auxiliary variables λ_{ij}^{h} represent the fraction of flow x_{ij} assigned to interval $[b_{ij}^{h-1}, b_{ij}^{h}]$. Constraints (14) ensure that the entire flow x_{ij} is assigned. The convexity of $F_{ij}(x)$, the objective function, and constraints (15) guarantee that in an optimal solution the flow x_{ij} is assigned to intervals in order of increasing indices, i.e., $\lambda_{ij}^{h} > 0$ if and only if $\lambda_{ij}^{h-1} = \Delta_{ij}^{h-1}$ for $h = 2,\ldots,n$. Constraints (18) - (20) guarantee flow conservation and that the entire demand is routed.

The piecewise linear approximation can be also applied to the C-SO MODEL and results in the following

LIN-C-SO MODEL:

$$
\begin{aligned}
\min \quad & \sum_{(ij)\in A} \sigma_{ij} & & \\
& x_{ij} = \sum_{h=1}^{n} \lambda_{ij}^{h} & \forall (i,j) \in A & \quad (21)\\
& \sigma_{ij} = \sum_{h=1}^{n} \frac{f_{ij}^{h} - f_{ij}^{h-1}}{\Delta_{ij}^{h}} \lambda_{ij}^{h} & \forall (i,j) \in A & \quad (22)\\
& 0 \le \lambda_{ij}^{h} \le \Delta_{ij}^{h} & \forall (i,j) \in A \quad \forall h = 1,\dots,n & \quad (23)\\
& x_{ij} = \sum_{c\in C} \sum_{k\in K_c^{\gamma}} a_{ij}^{kc} y_{ck} & \forall (i,j) \in A & \quad (24)\\
& d_c = \sum_{k\in K_c^{\gamma}} y_{ck} & \forall c \in C & \quad (25)\\
& x_{ij} \ge 0 & \forall (i,j) \in A & \\
& y_{ck} \ge 0 & \forall c \in C \quad \forall k \in K_c^{\gamma}. &
\end{aligned}
$$

As before, variables x_{ij} and y_{ck} represent the flow on arc (i,j) and the flow on the k-th path of commodity $c \in C$, respectively, and constraints (24) and (24) ensure that the flow on arc (i,j) is set correctly and that the entire demand for a commodity is routed, respectively. Constraints (21)-(23) play the same role as the corresponding constraints in LIN-SO MODEL. For convenience, the notation used in LIN-C-SO MODEL is summarized in Table 1.

4 Computational results

A large and diverse set of instances has been used in a computational study to assess the performance of the LIN-C-SO MODEL. The instances differ, among others, in terms of geography and demand pattern. Details of the instances can be found in Angelelli et al. (2016) and the instances are available at `http://or-brescia.unibs.it/instances`. For each instance, we generate a traffic assignment using a restricted set of paths for maximum inconvenience values, γ, ranging from 0% to 35% in increments of 1% (i.e., 36 traffic assignments). The linear programs are solved using CPLEX 12.6.0. The experiments were conducted on a Windows 64-bit computer with Intel Xeon processor E5-1650, 3.50 GHz, and 16 GB Ram. For all experiments, we have used the latency function proposed by the U.S. Bureau of Public Roads,

Table 1. Notation used in LIN-C-SO MODEL

Constrained system optimum notation	
Sets	
V	set of vertices
A	set of arcs
C	set of OD pairs
K_c	set of all possible paths for $c \in C$
K_c^γ	set of eligible paths for $c \in C$ with maximum inconvenience γ
Parameters	
u_{ij}	parameter representing the maximum rate of vehicles that can enter the arc $(i,j) \in A$ without experiencing substantial delays due to congestion
t_{ij}^{FF}	free-flow travel time of arc $(i,j) \in A$
a_{ij}^{kc}	1 if path $k \in K_c^\gamma$ contains arc $(i,j) \in A$, 0 otherwise
d_c	flow rate for OD pair $c \in C$
b_{ij}^h	h-th breakpoint related to the arc (i,j)
f_{ij}^h	value of the function $F(x_{ij})$ in breakpoint b_{ij}^h
n	number of intervals
Δ_{ij}^h	interval size
Decision variables	
y_{ck}	flow rate of OD pair $c \in C$ routed on path $k \in K_c^\gamma$
x_{ij}	total flow rate entering arc $(i,j) \in A$: $x_{ij} = \sum_{c \in C} \sum_{k \in K_c^\gamma} a_{ij}^{kc} y_{ck}$
Auxiliary decision variables	
σ_{ij}	total travel time multiplied by x_{ij} evaluated using the piecewise function
λ_{ij}	amount of vehicles in the $h-th$ interval of piecewise function of arc (i,j)

i.e., $t_{ij} = t_{ij}^{FF}(1+0.15(\frac{x_{ij}}{u_{ij}})^4)$. To define a set of restricted paths, we use the free-flow travel time of an arc as its normal length.

4.1 Piecewise linear approximation accuracy

In the computational experiments with LIN-SO MODEL and LIN-S-CO MODEL, we use a piecewise linear approximation of the objective function

$$F_{ij}(x_{ij}) = \sum_{(i,j)\in A} t_{ij}^{FF}(1+0.15(\frac{x_{ij}}{u_{ij}})^4)x_{ij}.$$

The piecewise linear approximation is obtained by dividing the domain $[0, U_{ij}]$ of the flow variable x_{ij} associated with arc (i, j) into $n = 1000$ equal length intervals. In Figure 2, we show $F_{ij}(x_{ij})$ and its piecewise linear approximation $\sigma_{ij}(x_{ij})$ for the latency function with parameters $t_{ij}^{FF} = 0.2$ hours and $u_{ij} = 2500$ vehicles/hour. We see that the two functions almost overlap - to enhance the visual effect, we have magnified parts of the figure using a 20x scaling factor. In fact, the maximum relative difference, i.e., $\frac{\sigma_{ij}(x_{ij})-F_{ij}(x_{ij})}{F_{ij}(x_{ij})}$, is always less than 0.5%. High-quality approximations can also be obtained with a smaller number of intervals, e.g., $n \geq 50$. However, as the solution times with the chosen number of intervals was acceptable, we did not explore the choice of the number of intervals in much detail.

4.2 Statistics

We collect and compute the statistics shown in Table 2. Note that the total travel time of a traffic assignment is always evaluated using the non-linear latency function $t_{ij}^{FF}(1+0.15(\frac{x_{ij}}{u_{ij}})^4)$.

4.3 Results for the LIN-C-SO model

We analyze the traffic assignments produced by the LIN-S-CO MODEL for maximum inconvenience values ranging from 0% to 35% in increments of 1% and averaged over the 80 instances. Recall that inconvenience and maximum inconvenience are defined in terms of the normal length of arcs and used only to define a restricted set of paths. The reported free-flow inconvenience and UE inconvenience are defined using the actual experienced travel times.

Travel time	
$\Theta_{\text{LIN-C-SO}}$	Experienced total travel time for the traffic assignment produced by the LIN-C-SO model.
$\Theta_{\text{LIN-SO}}$	Experienced total travel time for the traffic assignment produced by the LIN-SO model.
Θ_{UE}	Experienced total travel time for the traffic assignment produced by the UE algorithm.
User experience	
I_{FF}	Free-flow inconvenience, i.e., the average of the relative difference between the experienced travel time on the path and the travel time on the fastest path under free-flow conditions for the associated OD pair over all used paths weighted by the path flow.
I_{UE}	UE inconvenience, i.e., the average of the relative difference between the experienced travel time on the path and the travel time experienced in the user equilibrium for the associated OD pair over all used paths weighted by the path flow.
Arc utilization	
x_{ij}/u_{ij} classes	The fraction of arcs falling in each of the following six classes: • A ($0\% < x_{ij}/u_{ij} \leq 20\%$) • B ($20\% < x_{ij}/u_{ij} \leq 40\%$) • C ($40\% < x_{ij}/u_{ij} \leq 60\%$) • D ($60\% < x_{ij}/u_{ij} \leq 80\%$) • E ($80\% < x_{ij}/u_{ij} \leq 100\%$) • F ($x_{ij}/u_{ij} > 100\%$)
OD paths	
Selected paths	Selected paths in the optimal solution of the LIN-C-SO model: • average number of selected paths per OD pair, • maximum number of selected paths per OD pair.

Table 2. Statistics

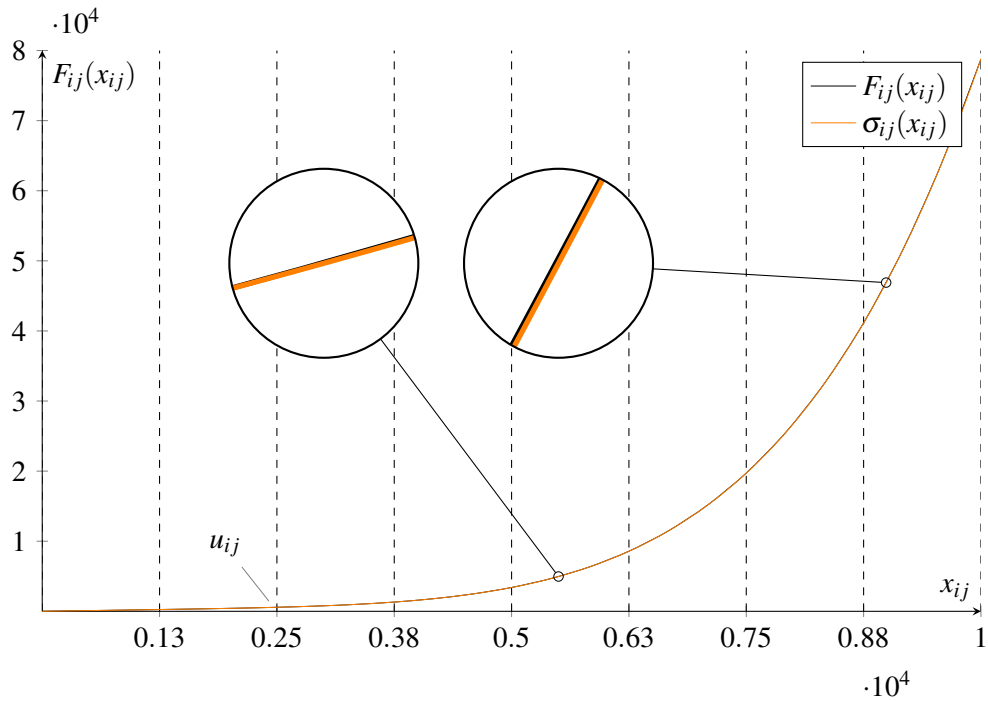


Figure 2. $F_{ij}(x_{ij})$ and $\sigma_{ij}(x_{ij})$ curves

Figure 3 shows the (average) total travel time, $\Theta_{\text{LIN-C-SO}}$, as a function of the maximum inconvenience, γ. As a reference, we have included, in the form of two horizontal lines, the (average) total travel time of the user equilibrium assignment, Θ_{UE}, and the (average) total travel time of the system optimum traffic assignment, $\Theta_{\text{LIN-SO}}$. We see that when the maximum inconvenience is greater than 10%, then $\Theta_{\text{LIN-C-SO}}$ is almost equal to $\Theta_{\text{LIN-SO}}$. Furthermore, we see that only when the maximum inconvenience, γ, is very small. i.e., 0, 1, or 2%, we have $\Theta_{UE} \leq \Theta_{\text{LIN-C-SO}}$. Already with a maximum inconvenience of 3%, there are enough paths in the set of restricted paths used in the LIN-S-CO MODEL to achieve a smaller (average) total travel time than in the UE. This is a manifestation of the "price of anarchy", i.e., individual decision making (UE equilibrium model) results in worse performance than coordinated decision making (SO and C-SO models).

Figure 4a shows the (average) free-flow inconvenience, I_{FF}, as a function of the maximum inconvenience, γ. The dashed line helps to identify those values of the maximum inconvenience for which the experienced free-flow inconvenience is larger than the maximum inconvenience. We see that that happens for maximum inconvenience values less than or equal to 4%. In addition, we see that when the maximum inconvenience is greater than 10%, I_{FF} is less than or equal to 3%. As the average free-flow inconvenience

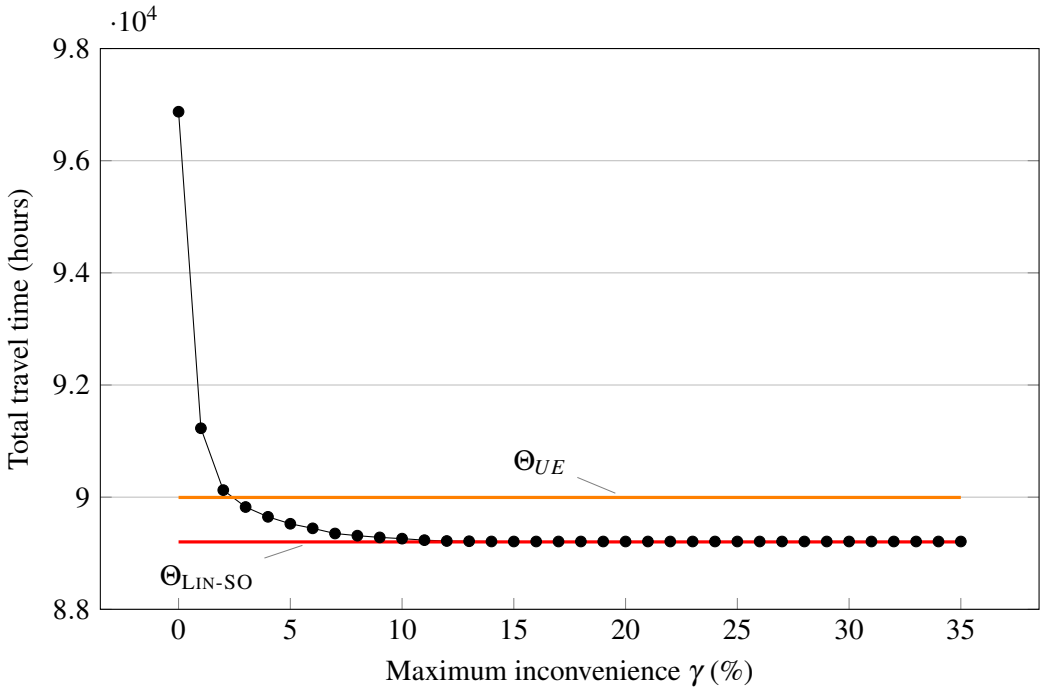
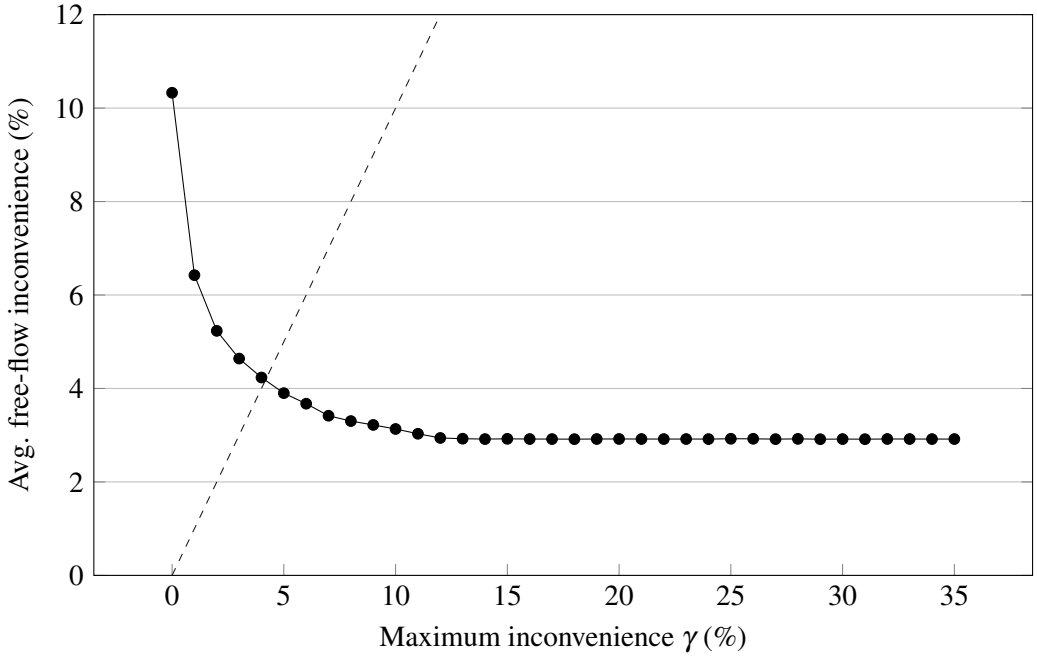


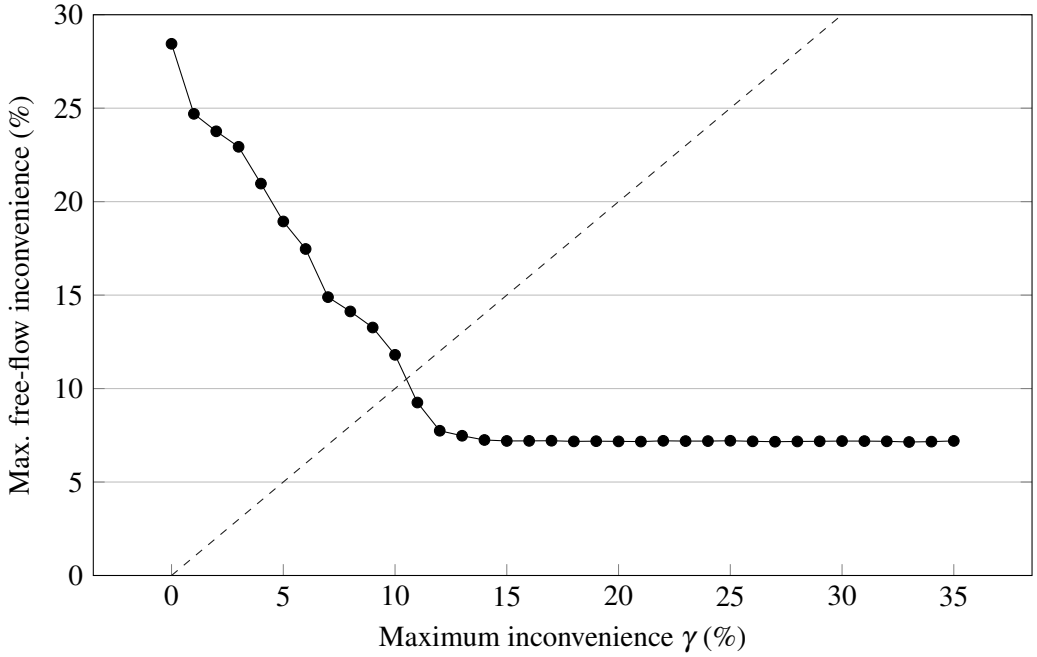
Figure 3. Total travel time, $\Theta_{\text{LIN-C-SO}}$ as a function of the maximum inconvenience, γ

conveys only part of what is happening, in Figure 4b, we show the maximum free-flow inconvenience over all instances as a function of the maximum inconvenience. As expected, it takes longer for the maximum experienced free-flow inconvenience to drop below the maximum inconvenience, i.e., only for maximum inconvenience values greater than or equal to 10%. We also see that maximum free-flow inconvenience stabilizes around 7% for maximum inconvenience values greater than or equal to 14%.

The free-flow inconvenience compares the experienced travel time to the shortest free-flow travel time. A more relevant and insightful comparison, however, may be the comparison with the experienced travel time in the user equilibrium traffic assignment, i.e., I_{UE}. Figures 5a and 5b show the average value of I_{UE} and the maximum value of I_{UE}, respectively, as a function of the maximum inconvenience. We see that for a maximum inconvenience of 5%, the average I_{UE} becomes negative. This means that, on average, users are experiencing travel times that are shorter than the travel time experienced in the user equilibrium traffic assignment. In fact, for a maximum inconvenience of 5% the maximum I_{UE} is also low, about 7%. For higher maximum inconvenience values, we see even better performance, e.g., at 12% the average I_{UE} is -1% and the maximum I_{UE} is 0. That is, at 12%, we are able to reduce the average experienced travel time without increasing the maximum experience travel time (compared to the travel times experienced in the UE traffic



(a) Average I_{FF}



(b) Maximum I_{FF},

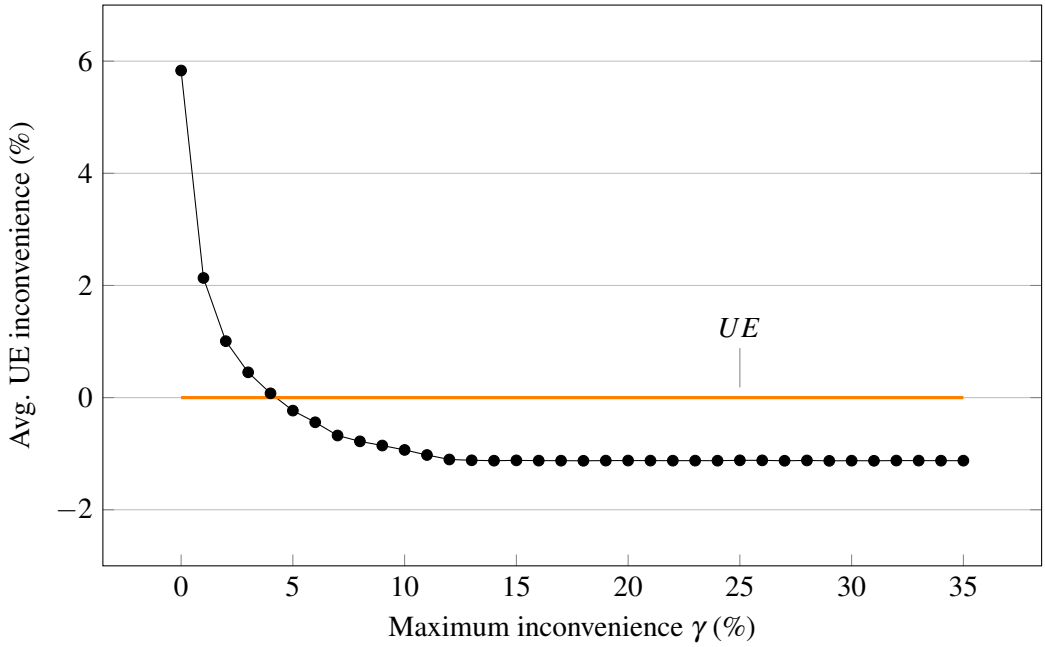
Figure 4. Free-flow inconvenience, I_{FF}

assignment).

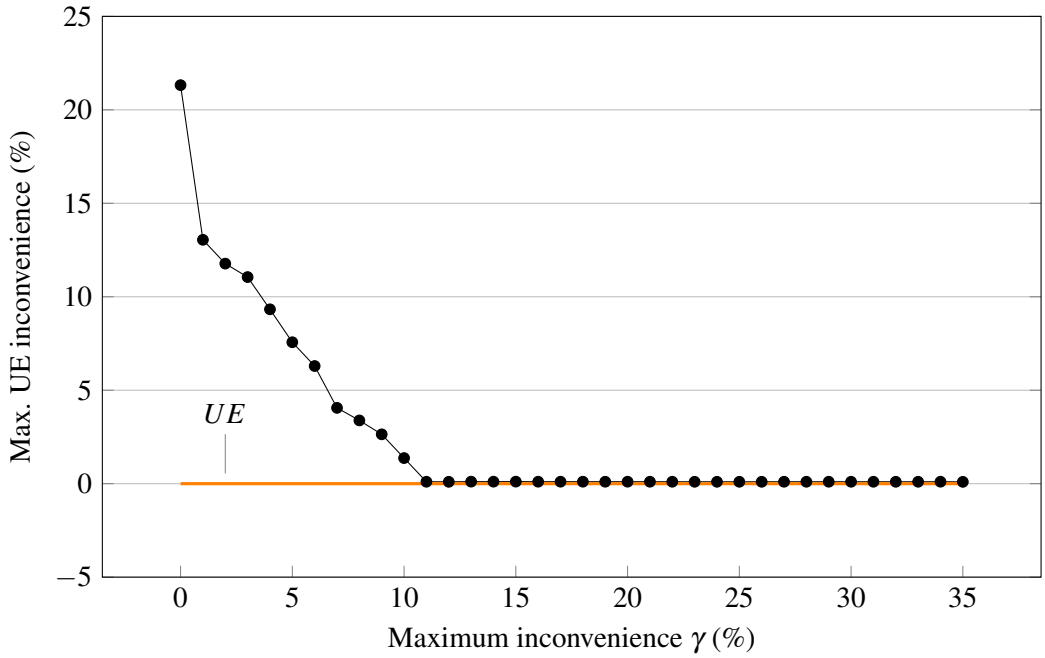
To provide more detail about the traffic assignment produced by the LIN C-SO MODEL, we present, in Figure 6, the average and maximum number of used paths for an OD pair, and, in Figure 7, information on the ratio of arc flow and arc capacity across the arcs in the network. Figure 6 shows that, on average, the number of used paths for an OD pair is close to one, which means that, on average, the demand of an OD pair is routed along a single path. However, it is also clear that in most settings the demand of a few OD pairs is routed along multiple paths, up to seven paths for $\gamma = 19\%$. Figure 7 shows the distribution of the ratio of arc flow and arc capacity across the arcs in the network. The arcs are divided into utilization classes A through F. Class A contains the least used arcs (with a ratio between 0 and 0.2). Intermediate classes contain arcs with moderate flows, in which any user experienced delays are still acceptable. Class F contains the most heavily used arcs with flows exceeding the practical capacity, which implies that on these arcs users will experience significant delays. As expected, we see that the fraction of heavily used arcs decreases when the maximum inconvenience increases so as to avoid significant delays and minimize total experienced travel time.

5 Final remarks and future research

The LIN-C-SO MODEL has an important advantage over previously proposed constrained system optimum models: it is a linear programming model and, hence, it is more likely that even very large instances can be solved efficiently. As such, it has larger potential to be of use in practical settings. However, the upfront enumeration of the sets of restricted paths from an origin to a destination may prove to be computationally prohibitive for very large instances, when the number of paths could easily reach several millions. A natural extension, therefore, will be developing a column generation approach for the LIN-C-SO MODEL, possibly using heuristics to dynamically identify beneficial paths. Furthermore, as pointed out in the introduction, the steady-state assumption, which allows us to work with static traffic assignments, may be reasonable during rush hour periods, but will not be appropriate for a period in which the demand varies over time. To accommodate varying demand over time, a time-dependent variant of the LIN-C-SO MODEL can be investigated.



(a) Average I_{UE},



(b) Maximum I_{UE},

Figure 5. UE inconvenience, I_{UE}

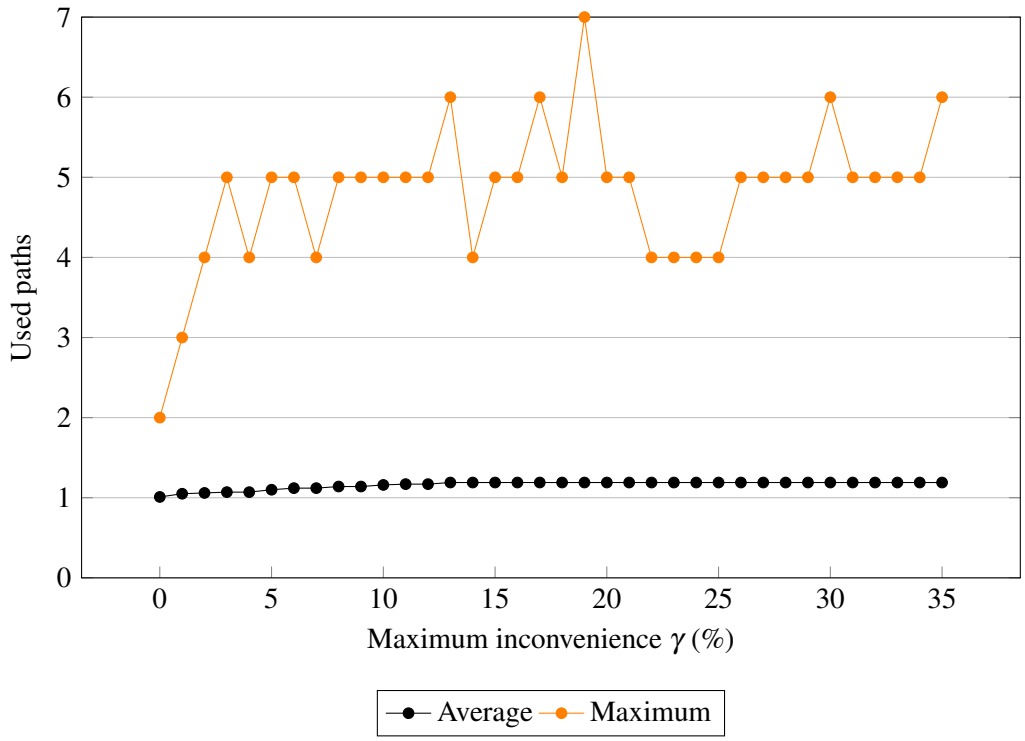


Figure 6. Average and maximum number of each OD pair used paths

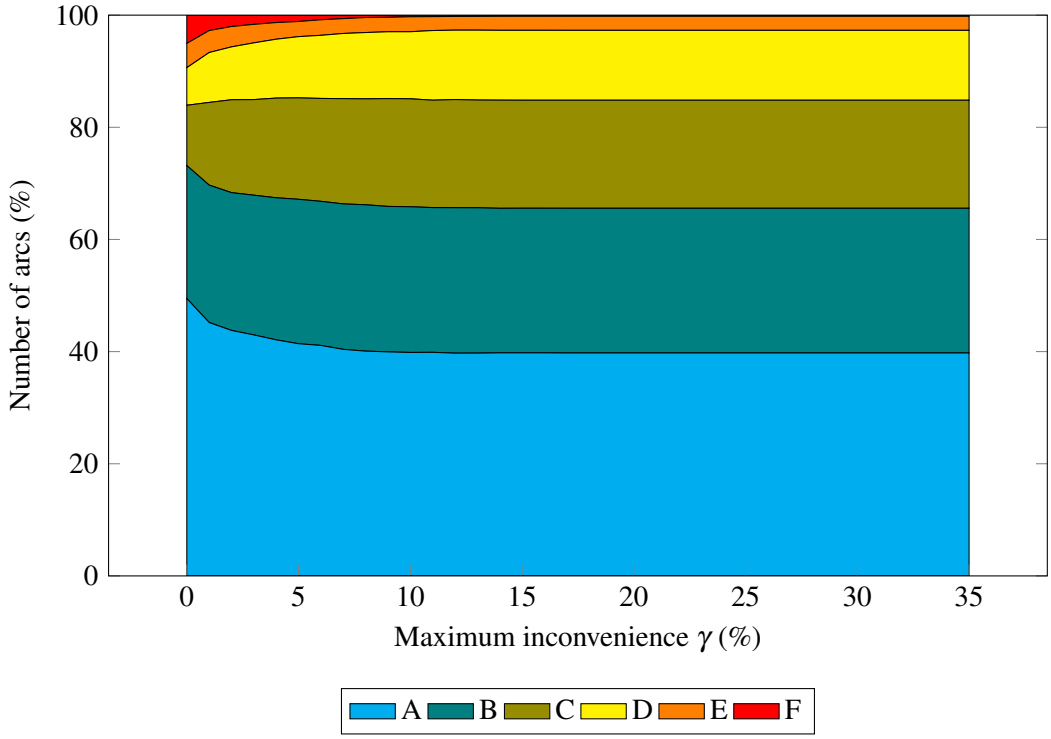


Figure 7. Arc utilization distribution

References

E. Angelelli, I. Arsik, V. Morandi, M. Savelsbergh, and M. Speranza. Proactive route guidance to avoid congestion. *Transportation Research Part B: Methodological*, 94:1 – 21, 2016. doi: http://dx.doi.org/10.1016/j.trb.2016.08.015.

M. Beckmann, C. McGuire, and C. B. Winsten. Studies in the economics of transportation. Technical report, RAND Corporation, 1956.

M. E. Ben-Akiva and S. R. Lerman. *Discrete choice analysis: theory and application to travel demand*. MIT press, 1985.

M. E. Ben-Akiva, S. Gao, Z. Wei, and Y. Wen. A dynamic traffic assignment model for highly congested urban networks. *Transportation Research Part C: Emerging Technologies*, 24:62–82, 2012.

J. R. Correa, A. S. Schulz, and N. E. Stier-Moses. Fast, fair, and efficient flows in networks. *Operations Research*, 55:215–225, 2007.

J. de Dios Ortuzar and L. G. Willumsen. *Modelling transport*. John Wiley & Sons, 2011.

M. Florian and D. Hearn. Network equilibrium and pricing. In *Handbook of Transportation Science*, pages 361–393. Springer, 1999.

O. Jahn, R. H. Möhring, and A. S. Schulz. Optimal routing of traffic flows with length restrictions in networks with congestion. In *Operations Research Proceedings 1999*, pages 437–442. Springer, 2000.

O. Jahn, R. H. Möhring, A. S. Schulz, and N. E. Stier-Moses. System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations Research*, 53:600–616, 2005.

M. Lujak, S. Giordani, and S. Ossowski. Route guidance: Bridging system and user optimization in traffic assignment. *Neurocomputing*, 151:449–460, 2015.

H. S. Mahmassani and S. Peeta. Network performance under system optimal and user equilibrium dynamic assignments: implications for advanced traveler information systems. *Transportation Research Record*, 1993.

H. S. Mahmassani and S. Peeta. System optimal dynamic assignment for electronic route guidance in a congested traffic network. *Urban traffic networks: dynamic flow modeling and control*, 1995.

D. K. Merchant and G. L. Nemhauser. A model and an algorithm for the dynamic traffic assignment problems. *Transportation science*, 12:183–199, 1978.

M. Papageorgiou. Dynamic modeling, assignment, and route guidance in traffic networks. *Transportation Research Part B: Methodological*, 24:471–495, 1990.

M. Patriksson. *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.

S. Peeta and A. K. Ziliaskopoulos. Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1:233–265, 2001.

A. S. Schulz and N. E. Stier-Moses. Efficiency and fairness of system-optimal routing with user constraints. *Networks*, 48:223–234, 2006.

Y. Sheffi. Urban transportation network. *Equilibrium analysis with mathematical programming methods, Prentice Hall*, 1985.

J. G. Wardrop. Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1:325–362, 1952.