

Balancing Fairness and Efficiency in Traffic Routing via Interpolated Traffic Assignment

Devansh Jalota¹, Kiril Solovey², Matthew Tsao³, Stephen Zoepf⁴, and Marco Pavone¹

September 2023

Abstract

System optimum (SO) routing, wherein the total travel time of all users is minimized, is a holy grail for transportation authorities. However, SO routing may discriminate against users who incur much larger travel times than others to achieve high system efficiency, i.e., low total travel times. To address the inherent unfairness of SO routing, we study the β -fair SO problem whose goal is to minimize the total travel time while guaranteeing a $\beta \geq 1$ level of unfairness, which specifies the maximum possible ratio between the travel times of different users with shared origins and destinations.

To obtain feasible solutions to the β -fair SO problem while achieving high system efficiency, we develop a new convex program, the Interpolated Traffic Assignment Problem (I-TAP), which interpolates between a fairness-promoting and an efficiency-promoting traffic-assignment objective. We evaluate the efficacy of I-TAP through theoretical bounds on the total system travel time and level of unfairness in terms of its interpolation parameter, as well as present a numerical comparison between I-TAP and a state-of-the-art algorithm on a range of transportation networks. The numerical results indicate that our approach is faster by several orders of magnitude as compared to the benchmark algorithm, while achieving higher system efficiency for all desirable levels of unfairness. We further leverage the structure of I-TAP to develop two pricing mechanisms to collectively enforce the I-TAP solution in the presence of selfish homogeneous and heterogeneous users, respectively, that independently choose routes to minimize their own travel costs. We mention that this is the first study of pricing in the context of fair routing for general road networks (as opposed to, e.g., parallel road networks).

1 Introduction

Traffic congestion has soared in major urban centres across the world, leading to widespread environmental pollution and huge economic losses. In the US alone, almost 90 billion US dollars of economic losses are incurred every year, with commuters losing hundreds of hours due to traffic congestion [1]. A contributing factor to increasing road traffic levels is the often sub-optimal route selection by users due to the lack of centralized traffic control [2, 3]. In particular, *selfish routing*, wherein users choose routes to minimize their travel times, results in a user equilibrium (UE) traffic pattern that is often far from the system optimum (SO) [4, 5]. To cope with the efficiency loss due to the selfishness of users, several methods including the control of a fraction of compliant users [6] and marginal cost tolls, where users pay for the externalities they impose on others, have been used to enforce the SO solution as a UE [7, 8].

While determining SO tolls is of fundamental theoretical importance, it is of limited practical interest [9] since SO traffic patterns are often unfair with some users incurring much larger travel times than others. This discrepancy among user travel times is referred to as *unfairness*, which, more formally, is the maximum possible ratio across all origin-destination (O-D) pairs of the experienced travel time of a given user to the travel time of the fastest user between the same O-D pair. The unfairness of the SO solution can be quite high in real-world transportation networks, since users may spend nearly twice as much time as others travelling

¹ Stanford University, USA; {djalota, pavone}@stanford.edu.

² Technion - Israel Institute of Technology, Israel; kirilsol@stanford.edu.

³ Lyft Inc., USA; mattxtsao@gmail.com.

⁴ Massachusetts Institute of Technology, USA; szoepf@mac.com.

between the same O-D pair [10]. Moreover, a theoretical analysis established that the SO solution can even have unbounded unfairness [11].

The lack of consideration of user-specific travel times in the global SO problem has led to the design of methods that aim to achieve a balance between the total travel time of a traffic assignment and the level of fairness that it provides. In a seminal work, Jahn et al. [10] introduced the Constrained System Optimum (CSO) to reduce the unfairness of traffic flows by bounding the ratio of the normal length of a path of a given user to the normal length of the shortest path for the same O-D pair. Here, *normal length* is any metric for an edge that is fixed *a priori* and is independent of the traffic flow, e.g., edge length or free-flow travel time. While many approaches to solve the CSO problem have been developed [12, 13, 14], they suffer from the inherent limitation that the level of experienced unfairness in terms of user travel times can be much higher than the bound on the ratio of normal lengths that the CSO is guaranteed to satisfy. In addition to this drawback, the algorithmic approaches to solve the CSO problem are often computationally prohibitive and do not provide theoretical guarantees in terms of the resulting solution fairness and efficiency. Furthermore, it is unclear how to develop a pricing scheme to enforce such proposed traffic assignments in practice.

In this work, we study a problem analogous to CSO that differs in the problem’s unfairness constraints. In particular, we explicitly consider the experienced unfairness in terms of user travel times as in [15], which, arguably, is a more accurate representation of user constraints as it accounts for costs that vary according to a traffic assignment. Our work further addresses the algorithmic concerns of existing approaches to solve fairness-constrained traffic routing problems by developing (i) a computationally-efficient approach that trades off efficiency and fairness in traffic routing, (ii) theoretical bounds to quantify the performance of our algorithm, and (iii) a pricing mechanism to enforce the resulting traffic assignment.

Contributions. We study the β -fair System Optimum (β -SO) problem, which involves minimizing the total travel time of users subject to unfairness constraints, where a $\beta \geq 1$ bound on unfairness specifies the maximum allowable ratio between the travel times of different users with shared origins and destinations.

We develop a simple yet effective approach for β -SO that involves solving a new convex program, the interpolated traffic assignment problem (I-TAP). I-TAP interpolates between the fair UE and efficient SO objectives to achieve a solution that is simultaneously fair and efficient. This allows us to approximate the β -SO problem as an unconstrained traffic assignment problem, which can be solved quickly. We further present theoretical bounds on the total system travel time and unfairness level in terms of the interpolation parameter of I-TAP and establish several sensitivity results regarding the continuity of the optimal edge flows and total travel times in the interpolation parameter.

We then exploit the structure of I-TAP to develop two pricing schemes which enforce users to selfishly select the flows satisfying the β bound on unfairness computed through our approach. In particular, for homogeneous users with the same value of time we develop a natural marginal-cost pricing scheme. For heterogeneous users, we exploit a linear-programming method [16] to derive prices that enforce the optimal flows computed through I-TAP. We mention that our work is the first to study road pricing in connection with fair routing for general road networks as opposed to, e.g., parallel networks.

We then evaluate the performance of our approach on real-world transportation networks for the above (Section 6.2) and other valid notions of intra-O-D pair unfairness (Section 6.3). In our work, intra-O-D pair unfairness measures involve computing the level of unfairness for each O-D pair by comparing the travel times of users traveling between that O-D pair and then computing the maximum unfairness across the O-D pairs. The numerical results indicate significant computational savings as well as superior performance for all desirable levels of unfairness β , as compared to the algorithm in [10]. Moreover, our results demonstrate that our approach can reduce the level of unfairness by 50% while increasing the total travel time by at most 2%, which indicates that a huge gain in user fairness can be achieved for a small loss in efficiency, making our approach a desirable option for use in route guidance systems.

Finally, we present a discussion regarding the consideration of unfairness measures that go beyond comparing the travel times of users traveling between a single O-D pair. In particular, we present several examples highlighting the importance of studying the interplay between different measures of unfairness, e.g., comparing user travel times across O-D pairs, which would be instrumental in informing transportation policy surrounding fair traffic routing. In presenting these examples, we also describe how our proposed unfairness notion can be modified to account for some fairness concerns arising in these settings.

This paper is organized as follows. Section 2 reviews related literature. We introduce in Section 3

the β -SO problem and metrics to evaluate the fairness and efficiency of a feasible traffic assignment. We introduce the I-TAP method and discuss its properties in Section 4, and develop pricing schemes to enforce the computed traffic assignments as a UE in Section 5. We evaluate the performance of the I-TAP method through numerical experiments in Section 6 for both the aforementioned unfairness measure as well as additional intra-O-D pair unfairness measures. Finally, we present a discussion regarding the consideration of unfairness notions that compare users’ travel times across different O-D pairs in Section 7 and provide directions for future work in Section 8.

2 Related Work

The trade-off between system efficiency and user fairness has been widely studied in applications including resource allocation, reducing the bias of machine-learning algorithms, and influence maximization. While different notions of fairness have been proposed, the level of fairness is typically controlled through the problem’s objective or constraints. For instance, fairness parameters that trade-off the level of fairness in the objective can be tuned to investigate the loss in system efficiency in the context of influence maximization [17] and resource allocation [18] problems. On the other hand, fairness parameters that bound the degree of allowable inequality between different user groups through the problem’s constraints have been proposed to reduce bias towards disadvantaged groups [19]. For example, group-based fairness notions [20] have been studied to reduce the bias of machine-learning algorithms, and diversity constraints have been introduced to ensure that the benefits of social interventions are fairly distributed [21].

In the context of traffic routing, several traffic assignment formulations have been proposed to achieve a balance between multiple performance criteria [22, 23], with a particular focus on fairness considerations in traffic routing [10]. Since Jahn et al. [10] introduced the CSO problem, there have been both theoretical studies [24] as well as the development of heuristic approaches to solve the NP-hard CSO problem. For instance, [10] proposed a Frank-Wolfe based heuristic, wherein the solution to the linearized CSO problem is obtained by solving a constrained shortest-path problem, while another work [25] developed a second-order cone programming technique. Several subsequent approaches for CSO have considered linear relaxations of the original problem [12, 13, 14]. Each of these approaches bounds the level of unfairness in terms of normal lengths of paths by restricting the set of eligible paths on which users can travel to those that meet a specified level of normal unfairness. However, the experienced unfairness in terms of the travel times on the restricted path set may be much higher than the specified level of normal unfairness, which is an *a priori* fixed quantity.

This inherent drawback of the CSO problem in limiting the experienced unfairness in terms of user travel times was overcome by [15], which proposed two Mixed Integer Non-Linear Programming models to capture traffic-dependent notions of unfairness. Their approach to solve these models relies on a linearization heuristic for the edge travel-time functions, which are in general *non-linear*. Achieving a high level of accuracy of the linear relaxations in approximating the true travel-time functions, however, requires solving a large MILP which is computationally *expensive*. Unlike [15], our I-TAP method is computationally *inexpensive*, while directly accounting for non-linear travel-time functions.

A further limitation of the existing methods for fair traffic routing is that there are limited results in providing pricing schemes to induce selfish users to collectively form the proposed traffic patterns, e.g., those satisfying a certain bound on unfairness. For instance, [26] provides tolling mechanisms to enforce fairness constrained flows which applies only to the simplified model of a parallel network. In more general networks, [27] proposes an auction-based bidding mechanism for users to be assigned to precomputed paths. However, this approach cannot be applied as-is to our setting as users are unconstrained with respect to a specific path set. Thus, we develop road tolling mechanisms to induce selfish users to collectively form traffic assignments guaranteeing a specified level of unfairness.

3 Model and Problem Definition

We model the road network as a directed graph $G = (V, E)$, with the vertex and edge sets V and E , respectively. Each edge $e \in E$ has a normal length η_e , which represents a fixed quantity such as the physical length of the corresponding road segment, and a flow-dependent travel-time function $t_e : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{> 0}$, which

maps x_e , the rate of traffic on edge e , to the travel time $t_e(x_e)$. As is standard in the traffic routing literature, we assume that the function t_e , for each $e \in E$, is differentiable, convex, locally Lipschitz continuous, and monotonically increasing.

Users make trips between a set of O-D pairs, and we model users with the same origin and destination as one commodity, where K is the set of all commodities. Each commodity $k \in K$ has a demand rate $d_k > 0$, which represents the amount of flow to be routed on a set of directed paths \mathcal{P}_k between its origin and destination. Here \mathcal{P}_k is the set of all possible paths between the origin and destination corresponding to commodity k . The edge flow of each commodity k is given by $\mathbf{x}^k = \{x_e^k\}_{e \in E}$, while the aggregate edge flow of all commodities is denoted as $\mathbf{x} := \{x_e\}_{e \in E}$. For an edge flow $\mathbf{x} := \{x_e\}_{e \in E}$ and a path $P \in \mathcal{P} = \cup_{k \in K} \mathcal{P}_k$, the amount of flow routed on the path is denoted as \mathbf{x}_P , where the vector of path flows $\mathbf{f} = \{\mathbf{x}_P : P \in \mathcal{P}\}$. Then, the travel time on path P is $t_P(\mathbf{x}) = \sum_{e \in P} t_e(x_e)$, while $\eta_P = \sum_{e \in P} \eta_e$ is its normal length.

We assume users are selfish and thus choose paths that minimize their total travel cost that is a linear function of tolls and travel time. For a value of time parameter $v > 0$, and a vector of edge prices (or tolls) $\boldsymbol{\tau} = \{\tau_e\}_{e \in E}$, the travel cost on a given path P under the traffic assignment \mathbf{x} is given by $C_P(\mathbf{x}, \boldsymbol{\tau}) = vt_P(\mathbf{x}) + \sum_{e \in P} \tau_e$.

3.1 Traffic Assignment

In this work we will consider several variants of the traffic assignment problem (TAP). The goal of the SO traffic assignment problem (SO-TAP) is to route users to minimize the total system travel time. This behavior is captured in the following convex program:

Definition 1 (Program for SO-TAP [4]).

$$\min_{\mathbf{f}} \quad h^{SO}(\mathbf{x}) := \sum_{e \in E} x_e t_e(x_e), \quad (1a)$$

$$\text{s.t.} \quad \sum_{k \in K} \sum_{P \in \mathcal{P}_k: e \in P} \mathbf{x}_P = x_e, \quad \forall e \in E, \quad (1b)$$

$$\sum_{P \in \mathcal{P}_k} \mathbf{x}_P = d_k, \quad \forall k \in K, \quad (1c)$$

$$\mathbf{x}_P \geq 0, \quad \forall P \in \mathcal{P}, \quad (1d)$$

with edge flow Constraints (1b), demand Constraints (1c), and non-negativity Constraints (1d).

We mention that the total travel time objective is only a function of the aggregate edge flow \mathbf{x} , which is related to the path flow \mathbf{f} through Constraint (1b). Note for any given path flow \mathbf{f} that both the edge flow \mathbf{x} and the commodity-specific edge flows \mathbf{x}^k for each commodity $k \in K$ are uniquely defined.

Closely related to SO-TAP is the UE traffic assignment problem (UE-TAP) that emerges from the selfish behavior of users, where each user strives to minimize its own travel time, without regard to the effect that it has on the overall travel time of all the users in the system. This behavior is captured through the following convex program:

Definition 2 (Program for UE-TAP [4]).

$$\min_{\mathbf{f}} \quad h^{UE}(\mathbf{x}) := \sum_{e \in E} \int_0^{x_e} t_e(y) dy, \quad (2a)$$

$$\text{s.t.} \quad (1b) - (1d). \quad (2b)$$

While the integral objective used to define UE-TAP has not found a clear economic or behavioral interpretation within the transportation and game-theory communities [4], the optimal solution of UE-TAP corresponds to an equilibrium, which can be seen through the KKT conditions of this optimization problem. That is, UE-TAP provides a polynomial time computable method to determine the user equilibrium flows. A defining property of the UE solution is that it is fair for all users since the travel time of all the flow that is routed between the same O-D pair is equal. In contrast, at the SO solution the sum of the travel time and marginal cost of travel is the same for all users travelling between the same O-D pair. Thus, marginal

cost pricing is used to induce selfish users to collectively form the SO traffic pattern. While the number of constraints, which depend on the path sets \mathcal{P}_k , can be exponential in the size of the transportation network, both SO-TAP and UE-TAP are efficiently computable since they can be formulated without explicitly enumerating all the path level flows and constraints [4].

3.2 Fairness and Efficiency Metrics

We evaluate the quality of any given traffic assignment \mathbf{x} , which satisfies Constraints (1b)-(1d), using two metrics, namely: (i) efficiency, which is of importance to a traffic authority as well as all users collectively, and (ii) fairness, which is of direct importance to each user.

We evaluate the efficiency of a traffic assignment by comparing its total travel time to that of the SO edge flow \mathbf{x}^{SO} . Recalling that $h^{SO}(\mathbf{x})$ denotes the total travel time of the edge flow \mathbf{x} , the *inefficiency ratio* of \mathbf{x} is

$$\rho(\mathbf{x}) := \frac{h^{SO}(\mathbf{x})}{h^{SO}(\mathbf{x}^{SO})}. \quad (3)$$

Note that for the UE solution \mathbf{x}^{UE} , the inefficiency ratio is the Price of Anarchy (PoA) [28], which we denote as $\bar{\rho}$.

To evaluate the fairness of a traffic assignment, we first introduce the notion of a *positive path* from [29].

Definition 3 (Positive Path). For any path flow \mathbf{f} with corresponding commodity-specific edge flows \mathbf{x}^k , a path $P \in \mathcal{P}_k$ is positive for a commodity $k \in K$ if for all edges $e \in P$, x_e^k is strictly positive. The set of all positive paths for a flow \mathbf{f} and commodity k is denoted as $\mathcal{P}_k^+(\mathbf{f}) = \{P : P \in \mathcal{P}_k, x_e^k > 0, \text{ for all } e \in P\}$.

The importance of the notion of a positive path is that the path decomposition of the commodity-specific edge flows \mathbf{x}^k may be non-unique; however the set of positive paths is always uniquely defined for such edge flows. That is, for commodity specific edge flows \mathbf{x}^k the set of positive paths for any two path decompositions \mathbf{f}_1 and \mathbf{f}_2 are equal, i.e., $\mathcal{P}_k^+(\mathbf{f}_1) = \mathcal{P}_k^+(\mathbf{f}_2)$.

We evaluate the fairness of a traffic flow \mathbf{f} with an edge decomposition \mathbf{x} through its corresponding unfairness U , which is defined as the maximum ratio across all O-D pairs of (i) the travel time on the slowest, i.e., highest travel time, positive path to (ii) the travel time on the fastest positive path between the same O-D pair:

$$U(\mathbf{f}) := \max_{k \in K} \max_{Q, R \in \mathcal{P}_k^+} \frac{t_Q(\mathbf{x})}{t_R(\mathbf{x})}.$$

That is, $U(\mathbf{f})$ returns the maximum possible ratio of travel times on positive paths across all commodities with respect to the path flow \mathbf{f} . As a result, the unfairness U is a number between one and infinity, and a traffic assignment has a high level of fairness if its unfairness is close to one while it has a low level of fairness if the corresponding unfairness is much larger than one. A discussion on numerically computing the above defined notion of unfairness is presented in Section 6. In contrast, other valid notions of unfairness could also be considered. For instance, for a given path flow decomposition \mathbf{f} with a corresponding edge flow \mathbf{x} , the unfairness $\tilde{U}(\cdot)$ of the path flows can be evaluated as the maximum ratio between the travel times of any two users travelling between the same O-D pair, i.e., $\tilde{U}(\mathbf{f}) = \max_{k \in K} \max_{Q, R \in \mathcal{P}_k : \mathbf{x}_Q, \mathbf{x}_R > 0} \frac{t_Q(\mathbf{x})}{t_R(\mathbf{x})}$. Note here that we only consider a ratio of travel times on paths with strictly positive flow for the path decomposition \mathbf{f} rather than the ratio of travel times for all positive paths. We defer a detailed treatment of path-based unfairness measures to Section 6.3 and highlight here some key features of the positive path based unfairness measure U .

The unfairness measure $U(\mathbf{f})$ can be efficiently computed and has the benefit that it applies to all possible path decompositions of the commodity specific edge flows \mathbf{x}^k . As a result, in the context of a single O-D pair travel demand, the unfairness measure $U(\mathbf{f})$ has the benefit that it is a property of the unique edge flow \mathbf{x} and is relevant when users are not constrained to a specific path decomposition, as happens in practice. In contrast, path decomposition specific unfairness measures, e.g., $\tilde{U}(\mathbf{f})$, though more interpretable from the perspective of practitioners, are likely to be more sensitive to the method used to compute the path decomposition. Furthermore, we note that the positive path based unfairness notion serves as an upper bound on the ratio of travel times for any two users travelling between the same O-D pair for the path flow

\mathbf{f} , i.e., $\tilde{U}(\mathbf{f}) \leq U(\mathbf{f})$ for all \mathbf{f} . As a result, our theoretical bounds on unfairness obtained for the positive path-based unfairness notion will naturally extend to path decomposition specific unfairness measures such as $\tilde{U}(\mathbf{f})$. Thus, in the rest of this paper we focus on the positive path-based unfairness measure and, for numerical comparison, we present other path decomposition specific unfairness measures, e.g., $\tilde{U}(\mathbf{f})$, in Appendix 6.3.

3.3 Toy Network Example

To illustrate the fairness and efficiency properties of the two optimization problems, SO-TAP and UE-TAP, studied in this work, we present a toy example of a two-edge Pigou network, as depicted in Figure 1. In particular, consider a demand of one that needs to be routed from the origin v_1 to the destination v_2 , with two edges (e_1 and e_2) connecting the origin to the destination. Observe that if the travel time functions on the two edges are given by $t_1(x_1) = 1$ and $t_2(x_2) = x_2$, then under the UE-TAP solution all users will be routed on edge two, while the SO-TAP solution that minimizes the total travel time will route 0.5 units of flow on both edges. The level of unfairness and the total travel time of the two traffic assignments are presented in the following table, which indicates that the UE-TAP solution is fair but inefficient while the SO-TAP solution is efficient but unfair.



Figure 1: A two-edge Pigou network example to illustrate the fairness and efficiency properties of the SO-TAP and UE-TAP optimization problems. For a demand of one, the UE-TAP solution routes all the flow on edge e_2 resulting in a fair solution but a total travel time of one. On the other hand, the SO-TAP solution routes 0.5 units of flow on both edges, resulting in an efficient solution with the minimum total travel time, but an unfair solution with an unfairness of two.

3.4 β -Fair System Optimum

Our focus in this work is in solving the following problem where we impose an upper bound on the maximum allowable level of unfairness in the network. In particular, for an unfairness parameter $\beta \in [1, \infty)$, any feasible path flow \mathbf{f} must satisfy $U(\mathbf{f}) \leq \beta$. To trade-off between user fairness and system efficiency, we consider the following β -fair System Optimum (β -SO) problem.

Definition 4 (Program for β -Fair System Optimum).

$$\min_{\mathbf{f}} \sum_{e \in E} x_e t_e(x_e), \quad (4a)$$

$$\text{s.t.} \quad (1b) - (1d), \quad (4b)$$

$$U(\mathbf{f}) \leq \beta. \quad (4c)$$

Note that without the unfairness Constraints (4c) (or when $\beta = \infty$), the above problem exactly coincides with SO-TAP. Furthermore, the β -SO problem is always feasible for any $\beta \in [1, \infty)$, since a solution to UE-TAP exists and achieves an unfairness of $\beta = 1$.

We also note that the difference between the β -SO and CSO problems is in the unfairness Constraints (4c). While the β -SO problem explicitly imposes an upper limit on the ratio of travel times on positive paths, the CSO problem imposes normal unfairness constraints for each path $P \in \mathcal{P}_k$ and any commodity $k \in K$ of the form $\eta_P \leq \phi \min_{P^* \in \mathcal{P}_k} \eta_{P^*}$ for some normal unfairness parameter $\phi \geq 1$. That is, the CSO problem minimizes the total travel time subject to flow conservation constraints over the set of paths with a normal unfairness level of at most ϕ . The authors of [10] use normal unfairness, which is a fixed quantity, as a

proxy to limit the maximum possible ratio of user travel times. Note that the experienced user travel times accounts for costs that vary according to a traffic assignment unlike normal unfairness.

The optimal solution of the β -SO problem corresponds to the highest achievable system efficiency whilst meeting unfairness constraints. However, solving β -SO directly is generally intractable as the unfairness Constraints (4c) are non-convex if the travel time function is non-linear. Moreover, since the unfairness metric studied in this work accounts for user costs that vary according to a traffic assignment, unlike normal unfairness that is an apriori fixed quantity, the NP-hardness of the CSO problem [10] suggests the computational hardness of β -SO [29, 15].

Finally, we mention that we consider a setting wherein the travel demand is time invariant and fractional user flows are allowed, as is standard in the traffic routing literature. Also, for notational simplicity, we consider for now a model where all users are homogeneous, i.e., they have an identical value of time v , and present an extension of our pricing result to the setting of heterogeneous users in Section 5.2. Furthermore, we consider a setting when users’ travel costs are a linear function of their travel time and tolls, a commonly used modeling assumption in the traffic routing literature [16, 30].

4 A Method for β -Fair System Optimum

In this section, we develop a computationally-efficient method for solving β -SO with edge-based unfairness constraints, to achieve a traffic assignment with a low total travel time, whose level of unfairness is at most β . In particular, we propose a new formulation of TAP, which we term interpolated TAP (or I-TAP), wherein the objective function linearly interpolates between the objectives of UE-TAP and SO-TAP. Our main insight is that the UE solution achieves a high level of fairness, whereas the SO solution achieves a low total travel time, and we wish to get the best of both worlds—high level of fairness at a low total travel time.

In this section, we describe the I-TAP method and evaluate its efficacy for the β -SO problem by addressing three key concerns regarding the solution efficiency, feasibility and computational tractability. In particular, we establish a relationship between I-TAP and β -SO through theoretical bounds on the inefficiency ratio (Section 4.2) and its optimality for two-edge Pigou networks (Section 4.3). We also establish the feasibility of I-TAP for β -SO by finding the range of values of the interpolation parameter such that the unfairness of the optimal I-TAP solution is guaranteed to be less than β (Section 4.2). Finally, we present an equivalence between I-TAP and UE-TAP to show that I-TAP can be computed efficiently (Section 4.4). These results indicate that we can approximate β -SO as an unconstrained traffic assignment problem that can be solved quickly. We also mention that we perform a sensitivity analysis to establish the continuity of the optimal traffic assignment and its total travel time in the interpolation parameter of I-TAP in Section 4.5.

4.1 Interpolated Traffic Assignment

We provide a formal definition of interpolated TAP:

Definition 5 (I-TAP). For a convex combination parameter $\alpha \in [0, 1]$, the interpolated traffic assignment problem, denoted as I-TAP $_{\alpha}$, is given by:

$$\min_{\mathbf{f}} h^{I\alpha}(\mathbf{x}) := \alpha h^{SO}(\mathbf{x}) + (1 - \alpha)h^{UE}(\mathbf{x}), \tag{5a}$$

$$\text{s.t.} \quad (1b) - (1d). \tag{5b}$$

A few comments are in order. First, it is clear that I-TAP $_0$ and I-TAP $_1$ correspond to UE-TAP and SO-TAP, respectively. Next, under the assumption that the travel time functions are strictly convex, we observe that for any $\alpha \in [0, 1]$ the program I-TAP $_{\alpha}$ has a strictly convex objective with linear constraints, and so I-TAP $_{\alpha}$ is a convex optimization problem with a unique edge flow solution $\mathbf{x}(\alpha)$.

For numerical implementation purposes, we propose a *dense sampling* procedure to compute a solution for β -SO with a low total travel time while guaranteeing a β bound on unfairness.

Algorithm for Computation of Optimal Interpolation Parameter To compute a good solution for β -SO, we evaluate the optimal solution $\mathbf{f}(\alpha)$ of I-TAP $_{\alpha}$ (with corresponding edge flows $\mathbf{x}(\alpha)$) for α taken

from a finite set $\mathcal{A}_s := \{0, s, 2s, \dots, 1\}$ for some step size $s \in (0, 1)$. That is, in $O(\frac{1}{s})$ computations of I-TAP, we can return the path flow $\mathbf{f}(\alpha^*)$ (with unique edge decomposition $\mathbf{x}(\alpha^*)$), for some $\alpha^* \in \mathcal{A}_s$, with the lowest total travel time that is at most β -unfair, i.e., $U(\mathbf{f}(\alpha^*)) \leq \beta$, and the value $h^{SO}(\mathbf{x}(\alpha^*))$ is minimized.

We observe experimentally (Section 6.2) that this method of computing the I-TAP solution for a finite set of convex combination parameters achieves a good solution for β -SO in terms of fairness and total travel time. We note here that our approach also naturally extends to other unfairness notions wherein the user equilibrium achieves the highest possible level of fairness, while the system optimum achieves the lowest total travel times (see Appendix 6.3). Finally, we restrict α to lie in the finite set \mathcal{A}_s since the exact functional form of the optimal solution $\mathbf{f}(\alpha)$ (with edge flow $\mathbf{x}(\alpha)$), and thus the unfairness $U(\mathbf{f}(\alpha))$ and the total travel time $h^{SO}(\mathbf{f}(\alpha))$ functions, in α is not directly known, though we show that $\mathbf{x}(\alpha)$ and $h^{SO}(\mathbf{x}(\alpha))$ are continuous in α in Section 4.5.

We also test (Section 6) an alternative approach to I-TAP, which instead of taking a convex combination of the SO-TAP and UE-TAP objectives, interpolates between their solutions. That is, we first compute the optimal solutions of UE-TAP (\mathbf{x}^{UE}) and SO-TAP (\mathbf{x}^{SO}), and return the value $(1 - \gamma)\mathbf{x}^{UE} + \gamma\mathbf{x}^{SO}$ for $\gamma \in [0, 1]$. While this Interpolated Solution (I-Solution) method only requires two traffic assignment computations as compared to $O(\frac{1}{s})$ computations of the I-TAP method, it leads to poor performance in comparison with the I-TAP method (see Section 6) and does not induce a natural marginal-cost pricing scheme, as I-TAP does (see Section 5). Thus, we focus our attention on I-TAP for this and the next sections.

4.2 Solution Efficiency and Fairness of I-TAP

In this section, we study the influence of the convex combination parameter α of I-TAP on the efficiency and fairness of the optimal solution $\mathbf{f}(\alpha)$ (and edge flow $\mathbf{x}(\alpha)$). In particular, we characterize (i) an upper bound on the inefficiency ratio as we vary α , and (ii) a range of values of α that are guaranteed to achieve a specified level of unfairness β for any optimal solution $\mathbf{f}(\alpha)$.

We first evaluate the performance of I-TAP by establishing an upper bound on the inefficiency ratio as a function of α . Theorem 1 shows that the upper bound of the inefficiency ratio $\rho(\mathbf{x}(\alpha))$ of the optimal traffic assignment of I-TAP $_\alpha$ is the minimum between two terms: (i) the Price of Anarchy (PoA) $\bar{\rho}$, and (ii) a more elaborate bound that is monotonically non-increasing in α .

Theorem 1 (I-TAP Solution Efficiency). *For any $\alpha \in (0, 1)$, let $\mathbf{x}(\alpha)$ be the optimal solution to I-TAP $_\alpha$. Then, the inefficiency ratio*

$$\rho(\mathbf{x}(\alpha)) \leq \min \left\{ \bar{\rho}, 1 + \frac{1 - \alpha}{\alpha} \cdot \frac{h^{UE}(\mathbf{x}(1)) - h^{UE}(\mathbf{x}(0))}{h^{SO}(\mathbf{x}(1))} \right\}.$$

Proof. To prove this result, we show that (i) $\rho(\mathbf{x}(\alpha)) \leq \bar{\rho}$ and (ii) $\rho(\mathbf{x}(\alpha)) \leq 1 + \frac{1 - \alpha}{\alpha} \frac{h^{UE}(\mathbf{x}(1)) - h^{UE}(\mathbf{x}(0))}{h^{SO}(\mathbf{x}(1))}$.

We first prove (i) by establishing that $h^{SO}(\mathbf{x}(\alpha)) \leq h^{SO}(\mathbf{x}(0))$ for all $\alpha \in [0, 1]$. Since $\mathbf{x}(\alpha)$ is the optimal solution to I-TAP $_\alpha$

$$\begin{aligned} h^{I_\alpha}(\mathbf{x}(\alpha)) &= \alpha h^{SO}(\mathbf{x}(\alpha)) + (1 - \alpha)h^{UE}(\mathbf{x}(\alpha)) \\ &\leq h^{I_\alpha}(\mathbf{x}(0)) \\ &= \alpha h^{SO}(\mathbf{x}(0)) + (1 - \alpha)h^{UE}(\mathbf{x}(0)). \end{aligned}$$

Next from the UE-TAP objective, it is the case that

$$h^{UE}(\mathbf{x}(0)) \leq h^{UE}(\mathbf{x}(\alpha)).$$

From these two inequalities, it follows that $h^{SO}(\mathbf{x}(\alpha)) \leq h^{SO}(\mathbf{x}(0))$ for all $\alpha \in [0, 1]$. Dividing both sides of the inequality by the minimum possible system travel time $h^{SO}(\mathbf{x}(1))$ establishes that the inefficiency ratio $\rho(\mathbf{x}(\alpha)) \leq \bar{\rho}$ proving claim (i).

Now, to prove claim (ii), we note that $h^{I_\alpha}(\mathbf{x}(\alpha)) \leq h^{I_\alpha}(\mathbf{x}(1))$ since $\mathbf{x}(\alpha)$ is an optimal solution to I-TAP $_\alpha$. It thus follows that

$$\alpha h^{SO}(\mathbf{x}(\alpha)) + (1 - \alpha)h^{UE}(\mathbf{x}(\alpha)) \leq \alpha h^{SO}(\mathbf{x}(1)) + (1 - \alpha)h^{UE}(\mathbf{x}(1)).$$

Dividing this expression by $\alpha h^{SO}(\mathbf{x}(1))$ and rearranging gives us that

$$\begin{aligned} \frac{h^{SO}(\mathbf{x}(\alpha))}{h^{SO}(\mathbf{x}(1))} &\leq 1 + \frac{1-\alpha}{\alpha} \cdot \frac{h^{UE}(\mathbf{x}(1)) - h^{UE}(\mathbf{x}(\alpha))}{h^{SO}(\mathbf{x}(1))} \\ &\leq 1 + \frac{1-\alpha}{\alpha} \cdot \frac{h^{UE}(\mathbf{x}(1)) - h^{UE}(\mathbf{x}(0))}{h^{SO}(\mathbf{x}(1))}, \end{aligned}$$

where the second inequality follows from the fact that $h^{UE}(\mathbf{x}(0)) \leq h^{UE}(\mathbf{x}(\alpha))$ for any $\alpha \in [0, 1]$. This proves claim (ii). \square

Theorem 1 establishes that, even in the worst case, the ratio between the total travel time of the edge flow $\mathbf{x}(\alpha)$ of I-TAP $_{\alpha}$ and that of the system optimal solution is at most the PoA. This result is not guaranteed to hold for other state-of-the-art CSO algorithms, which may output traffic assignments with much larger total travel times than the UE solution. For instance, in Section 6 we show through numerical experiments that the algorithm proposed by Jahn et al. [10] achieves total travel times much higher than that of the user equilibrium solution for certain ranges of unfairness. Further, Theorem 1 shows that the upper bound on the inefficiency ratio becomes closer to one as the objective $h^{I\alpha}$ gets closer to h^{SO} .

The bound on the inefficiency ratio obtained from Theorem 1 is depicted in Figure 2. Note that in order to obtain the value of the convex combination parameter α^* at which the two upper bounds on the inefficiency ratio are equal we need to impose the constraint $\bar{\rho} = \frac{1-\alpha^*}{\alpha^*} \cdot \frac{h^{UE}(\mathbf{x}(1)) - h^{UE}(\mathbf{x}(0))}{h^{SO}(\mathbf{x}(1))}$, which yields that $\alpha^* = \frac{h^{UE}(\mathbf{x}(1)) - h^{UE}(\mathbf{x}(0))}{h^{SO}(\mathbf{x}(0)) + h^{UE}(\mathbf{x}(1)) - h^{UE}(\mathbf{x}(0))}$.

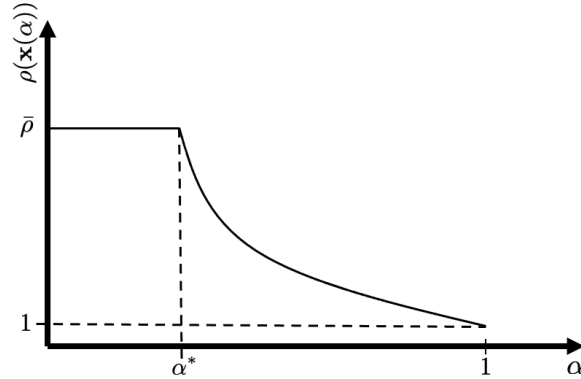


Figure 2: Upper bound on the inefficiency ratio of the optimal solution $\mathbf{x}(\alpha)$ of the problem I-TAP $_{\alpha}$ for $\alpha \in [0, 1]$

Having established a worst case performance guarantee for I-TAP in terms of the inefficiency ratio, we now establish a range of values of α at which any optimal solution $\mathbf{f}(\alpha)$ of I-TAP $_{\alpha}$ is guaranteed to attain a β bound on unfairness. In particular, we specialize the following result to polynomial travel time functions, e.g., the commonly used BPR function [4].

Theorem 2 (Feasibility of I-TAP for β -SO). *Suppose that the largest degree of the polynomial travel time functions $t_e(x_e)$ is m for some $e \in E$. Then, the unfairness of any optimal solution $\mathbf{f}(\alpha)$ of I-TAP $_{\alpha}$ is upper bounded by β , i.e., $U(\mathbf{f}(\alpha)) \leq \beta$, for any $\alpha \leq \frac{\beta-1}{m}$.*

The proof of Theorem 2 leverages the fact that at an equilibrium flow, under a given vector of tolls, the travel cost on all positive paths is equal for all users in any commodity k , as was established in [29, Lemma 2]. We provide a proof of this claim here for completeness.

Lemma 1 (Equality of Travel Costs on Positive Paths [29]). *Let the cost on an edge be given by $c_e(x_e, \tau_e) = t_e(x_e) + \tau_e$, where the value-of-time v of users is normalized to one. Then, under any equilibrium flow induced by the edge tolls τ , the total travel cost of any two positive paths P, Q are equal, i.e., $\sum_{e \in P} c_e(x_e, \alpha) = \sum_{e \in Q} c_e(x_e, \alpha)$.*

Proof. To prove this claim, we consider a network with travel time functions $c_e(x_e, \tau_e) = t_e(x_e) + \tau_e$ for all $e \in E$, which is valid since the travel time function is polynomial and so $c_e(x_e, \tau_e)$ is still differentiable, convex and monotonically increasing in x_e . Next, the equilibrium solution under a vector of tolls $\boldsymbol{\tau}$ is given by the UE-TAP with objective $\sum_{e \in E} \int_0^{x_e} c_e(y, \alpha) dy$, where $c_e(x_e, \tau_e) = t_e(x_e) + \tau_e$. By the first order necessary and sufficient conditions of this UE-TAP, for any two positive flow paths P, Q between the same O-D pair k , it holds that $\sum_{e \in P} c_e(x_e, \alpha) = \sum_{e \in Q} c_e(x_e, \alpha) = \mu_k$ for some μ_k . We now show that any positive path R between the same O-D pair that is not a path with positive flow, i.e., $\mathbf{x}_R = 0$, also has the same travel cost as any path P with positive flow, i.e., $\sum_{e \in R} c_e(x_e, \alpha) = \sum_{e \in P} c_e(x_e, \alpha) = \mu_k$.

Suppose by contradiction that there is a path R for a commodity k whose travel cost is not equal to μ_k . From the equilibrium condition of UE-TAP it must hold that $\sum_{e \in R} c_e(x_e, \tau_e) > \mu_k$. Next, let e_1, \dots, e_r denote the edges in path R in the order of traversal, and for each edge $e_i \in R$, consider a path used by commodity k using that edge, i.e., a path $P_i = R_i^s - e_i - R_i^t$ where R_i^s is the component of that path preceding edge e_i and R_i^t is the component of that path following edge e_i . Observe that the sum of the costs of these used paths P_i is given by $\sum_{i=1}^r \sum_{e \in P_i} c_e(x_e, \tau_e) = r\mu_k$. Next, considering the paths P'_1, \dots, P'_{r-1} , where $P'_1 = R_{i+1}^s - R_i^t$, we obtain that

$$r\mu_k = \sum_{i=1}^r \sum_{e \in P_i} c_e(x_e, \tau_e) = \sum_{i=1}^{r-1} \sum_{e \in P'_i} c_e(x_e, \tau_e) + \sum_{e \in R} c_e(x_e, \tau_e).$$

Since it holds from the KKT conditions of UE-TAP that $\sum_{e \in P'_i} c_e(x_e, \tau_e) \geq \mu_k$ for each $i \in \{1, \dots, r-1\}$, the above equality implies that $\sum_{e \in R} c_e(x_e, \tau_e) \leq \mu_k$, a contradiction. Thus, we must have that the total cost on all positive paths is exactly equal to μ_k , thereby proving our claim. \square

We now leverage Lemma 1 to complete the proof of Theorem 2.

Proof. Without loss of generality, we normalize the value of time v to 1 and for notational simplicity denote $x_e = x_e(\alpha)$ and $\mathbf{f} = \mathbf{f}(\alpha)$. We now establish this result in two steps. First, we show that if $c_e(x_e, \alpha) = t_e(x_e) + \alpha x_e t'_e(x_e) \in [t_e(x_e), \beta t_e(x_e)]$ for all edges e then the unfairness satisfies $U(\mathbf{f}) \leq \beta$. Next, we show for $\alpha \leq \frac{\beta-1}{m}$ that the cost satisfies $c_e(x_e, \alpha) \in [t_e(x_e), \beta t_e(x_e)]$, which, together with the first claim, implies the result.

To prove the first claim, we consider a network with travel time functions $c_e(x_e, \alpha)$ for all $e \in E$, which is valid since the travel time function is polynomial and so $c_e(x_e, \alpha)$ is still differentiable, convex and monotonically increasing in x_e . Next, from Lemma 1 it holds that the total cost on any two positive paths P and Q are also equal, i.e., $\sum_{e \in P} c_e(x_e, \alpha) = \sum_{e \in Q} c_e(x_e, \alpha)$ since at an equilibrium flow the cost of any two used paths are equal. This result implies for any commodity k and any two positive paths P, Q that

$$\begin{aligned} \sum_{e \in P} t_e(x_e) &\leq \sum_{e \in P} c_e(x_e, \alpha) = \sum_{e \in Q} c_e(x_e, \alpha) \\ &\leq \beta \sum_{e \in Q} t_e(x_e), \end{aligned}$$

i.e., the ratio of the travel times on positive paths can never exceed β . Thus, $U(\mathbf{f}) \leq \beta$.

Next, to prove the second claim, we note that since the travel time function is a polynomial function of degree m , we can let $t_e(x_e) = \sum_{i=0}^m \gamma_i x_e^i$. Then, we have:

$$\begin{aligned} x_e t'_e(x_e) &= x_e \sum_{i=1}^m i \gamma_i x_e^{i-1} = \sum_{i=1}^m i \gamma_i x_e^i \\ &\leq m \sum_{i=1}^m \gamma_i x_e^i \leq m \sum_{i=0}^m \gamma_i x_e^i \\ &= m t_e(x_e). \end{aligned}$$

Note that we used the fact that $\gamma_0 \geq 0$ in the second inequality, which follows since $t_e(x_e) \geq 0$ for all $x_e \in \mathbb{R}_{\geq 0}$. From the above inequalities, we note that $\alpha x_e t'_e(x_e) \leq \alpha m t_e(x_e)$. Now, if we set $\alpha m t_e(x_e) \leq (\beta - 1) t_e(x_e)$, then we have for any $\alpha \leq \frac{\beta-1}{m}$ that the cost $c_e(x_e, \alpha) \in [t_e(x_e), \beta t_e(x_e)]$ and thus the resulting flow \mathbf{f} has an unfairness of at most β . \square

Theorem 2 establishes a relation between the convex combination parameter α and the level of unfairness of any optimal path flow $\mathbf{f}(\alpha)$ when the edge travel-time functions are polynomial.

We can further show that the bound obtained in Theorem 2 is in fact tight by demonstrating an instance such that for any $\alpha > \frac{\beta-1}{m}$ the unfairness of the solution $\mathbf{f}(\alpha)$ of I-TAP $_{\alpha}$ is strictly greater than β .

Lemma 2 (Tightness of Unfairness Bound). *Suppose $\mathbf{f}(\alpha)$ is an optimal solution to I-TAP $_{\alpha}$ for any $\alpha \in [0, 1]$. Then, there exists a two-edge parallel network with polynomial travel time functions of degree at most m such that for any $\alpha > \frac{\beta-1}{m}$, the unfairness $U(\mathbf{f}(\alpha)) > \beta$.*

Proof. Consider a demand of one in a two edge Pigou network, with two nodes—an origin and a destination node—connected by two parallel edges e_1 and e_2 . For edge e_1 , let $t_1(x_1) = 1 + \epsilon x$ for some small $\epsilon > 0$, and let $t_2(x_2) = x_2^m$ for some $m \in \mathbb{N}$. Then, the first order necessary and sufficient KKT conditions of I-TAP $_{\alpha}$ for any $\alpha \in [0, 1]$ imply that

$$t_2(x_2) + \alpha x_2 t_2'(x_2) = t_1(x_1) + \alpha x_1 t_1'(x_1).$$

Substituting $t_2(x_2) = x_2^m$ and $t_1(x_1) = 1 + \epsilon x_1$ gives the following ratio between the travel times on the two links:

$$\frac{t_1(x_1)}{t_2(x_2)} = \frac{1 + \epsilon x_1}{x_2^m} - \frac{\alpha \epsilon x_1}{x_2^m} = 1 + m\alpha - \frac{\alpha \epsilon x_1}{x_2^m}.$$

By the unfairness constraint that $\frac{t_1(x_1)}{t_2(x_2)} \leq \beta$, it follows that if $\alpha > \frac{\beta-1}{m}$, then

$$\frac{t_1(x_1)}{t_2(x_2)} = 1 + m\alpha - \frac{\alpha \epsilon x_1}{x_2^m} > 1 + m\frac{\beta-1}{m} - \frac{\alpha \epsilon x_1}{x_2^m} = \beta - \frac{\alpha \epsilon x_1}{x_2^m}.$$

Taking $\epsilon \rightarrow 0$, we have that $\frac{t_1(x_1)}{t_2(x_2)} \geq \beta$ for any $\alpha > \frac{\beta-1}{m}$. Thus, the unfairness of the optimal flow $\mathbf{f}(\alpha)$, which is identical to the edge flow $\mathbf{x}(\alpha)$ in a parallel network setting, is at least β . \square

Together, Theorem 2 and Lemma 2 imply that a β level of unfairness can be guaranteed using I-TAP on all traffic networks only when $\alpha \leq \frac{\beta-1}{m}$, where m is the maximum degree of the polynomial corresponding to the travel time functions for each edge $e \in E$.

4.3 Optimality of I-TAP

In this section, we show that I-TAP exactly computes the minimum total travel time solution for any desired level of unfairness β in any two edge Pigou network. That is, there is some convex combination parameter α^* for which the solution of I-TAP $_{\alpha^*}$ is also a solution to the β -SO problem for any two edge Pigou network.

Lemma 3 (Optimality of I-TAP). *Consider a two edge Pigou network where the optimal solution of the β -SO problem is \mathbf{x}_{β}^* for any $\beta \in [1, \infty)$. Then, there exists a convex combination parameter α^* such that $\mathbf{x}(\alpha^*) = \mathbf{x}_{\beta}^*$, i.e., the solution of I-TAP $_{\alpha^*}$ and the optimal solution of the β -SO problem coincide.*

Proof. To prove the claim, we first note that in a two edge Pigou network the sum of the traffic flows on the two edges must add up to the traffic demand D . Thus, we can reduce the problem of determining a two-dimensional vector of edge flows to a single dimensional problem of determining the flow x_1 on the first edge, with the flow on the other edge given by $D - x_1$ for a traffic demand D . With slight abuse of notation, we denote $h^{SO}(x_1)$, $h^{UE}(x_1)$ and $h^{I\alpha}(x_1)$ to denote the system optimum, user equilibrium and I-TAP objectives, respectively.

We now complete the proof in two steps. First we show that the optimal solutions of I-TAP for any $\alpha \in [0, 1]$ and the β -SO problem lie between the user equilibrium and system optimum solutions, i.e., $x_1(\alpha), x_1^{\beta} \in [x_1(0), x_1(1)]$ for all $\alpha \in [0, 1]$ and $\beta \in [1, \infty)$, where we assume without loss of generality that $x_1(0) \leq x_1(1)$. Note here that we only compare the edge flows of I-TAP and β -SO since both the path and edge flows coincide for a two edge Pigou network. We then extend the result of Corollary 1 to show that the optimal solution $x_1(\alpha)$ is continuous in the closed interval $[0, 1]$. Note that both the above claims jointly

imply by the intermediate value theorem that for any $\beta \in [1, \infty)$ there exists some $\alpha^* \in [0, 1]$ such that $x_1(\alpha^*) = x_1^\beta$. We now proceed to prove the two claims.

We begin by establishing that $x_1(\alpha) \in [x_1(0), x_1(1)]$ for any $\alpha \in [0, 1]$. In particular, we will show that for any point not in the interval $[x_1(0), x_1(1)]$ that we can find another feasible point with a lower I-TAP objective value. To see this, we proceed by contradiction. Fix some $\alpha \in [0, 1]$ and suppose that $x_1(\alpha) < x_1(0)$. Then, we observe by the optimality of $x_1(0)$ for the user equilibrium traffic assignment problem that $h^{UE}(x_1(\alpha)) > h^{UE}(x_1(0))$ and by the strict convexity of h^{SO} that $h^{SO}(x_1(\alpha)) > h^{SO}(x_1(0))$. Together, both the strict inequalities imply that $h^{I\alpha}(x_1(\alpha)) > h^{I\alpha}(x_1(0))$, a contradiction. Through an almost identical argument, we can show that $x_1(\alpha) > x_1(1)$ is also not possible. Thus, we have shown that $x_1(\alpha) \in [x_1(0), x_1(1)]$ for any $\alpha \in [0, 1]$.

Next, we show that $x_1^\beta \in [x_1(0), x_1(1)]$ for any $\beta \in [1, \infty)$. In particular, we will show that for any point not in the interval $[x_1(0), x_1(1)]$ that we can find another feasible point with a lower system optimum objective value. To see this, we again proceed by contradiction. First suppose that $x_1^\beta < x_1(0)$. In this case, note that $h^{SO}(x_1^\beta) > h^{SO}(x_1(0))$ by the strict convexity of h^{SO} . Since $x_1(0)$ is a feasible solution to the β -SO problem for any $\beta \in [1, \infty)$ this implies that $x_1^\beta < x_1(0)$ cannot be an optimal solution to the β -SO problem. This implies that $x_1^\beta \geq x_1(0)$. Next, suppose by contradiction that $x_1^\beta > x_1(1)$. If $x_1^\beta = D$, then $h^{SO}(x_1^\beta) = Dt_1(D) > Dt_1(x_1(0)) = x_1(0)t_1(x_1(0)) + (D - x_1(0))t_2(x_2(0)) = h^{SO}(x_1(0))$ since $x_1(0) < x_1^\beta = D$. Thus, $x_1(0)$ achieves a lower total travel time and is feasible for any $\beta \in [1, \infty)$, and so $x_1^\beta = D$ cannot be a solution to the β -SO problem. Next, suppose that $x_1(1) < x_1^\beta < D$. In this case, we have that $\frac{t_1(x_1^\beta)}{t_2(D - x_1^\beta)} > \max\{\frac{t_1(x_1(1))}{t_2(D - x_1(1))}, 1\}$, and thus the ratio of travel times is strictly greater than that under the system optimal traffic assignment. Thus, we have that for any $\beta \in [1, \infty)$ it must be the case that $x_1^\beta \leq x_1(1)$, thereby proving our claim that $x_1^\beta \in [x_1(0), x_1(1)]$.

We now prove the second claim that $x_1(\alpha)$ is continuous in the closed interval $[0, 1]$. First observe by Corollary 1 that $x_1(\alpha)$ is continuous in the open interval $(0, 1)$. Thus, we just need to prove the continuity of $x_1(\alpha)$ for $\alpha = 0, 1$. To do so, we introduce a problem instance specific constant $C(G, D, \{t_e\}_{e=1}^2)$ that denotes the system optimum objective value when $x_1 = x_2 = D$. We note that $C(G, D, \{t_e\}_{e=1}^2)$ is an upper bound on the system optimal objective for any feasible traffic assignment and it is a constant that depends on the graph G , the demand D and the travel time functions. We prove the continuity for $\alpha = 0$ in what follows. First, observe that

$$\begin{aligned} h^{I\alpha}(x_1(0)) - h^{I\alpha}(x_1(\alpha)) &= \alpha (h^{SO}(x_1(0)) - h^{SO}(x_1(\alpha))) \\ &\quad + (1 - \alpha) (h^{UE}(x_1(0)) - h^{UE}(x_1(\alpha))), \\ &< \alpha (h^{SO}(x_1(0)) - h^{SO}(x_1(\alpha))), \\ &\leq \alpha C(G, D, \{t_e\}_{e=1}^2), \end{aligned}$$

where the strict inequality follows since $h^{UE}(x_1(0)) < h^{UE}(x_1(\alpha))$, and the last inequality follows by the boundedness of h^{SO} due to the finite demand D . Finally, by the strict convexity of the objective $h^{I\alpha}(\cdot)$ it must follow that $x_1(\alpha) \rightarrow x_1(0)$ as $\alpha \rightarrow 0$ since $h^{I\alpha}(x_1(0)) \rightarrow h^{I\alpha}(x_1(\alpha))$ as $\alpha \rightarrow 0$ by the above analysis. Thus, we have continuity of $x_1(\alpha)$ for $\alpha = 0$. Through an analogous analysis we can also obtain continuity for $\alpha = 1$.

Finally, we note that both the above claims jointly imply by the intermediate value theorem that for any $\beta \in [1, \infty)$ there exists some $\alpha^* \in [0, 1]$ such that $x_1(\alpha^*) = x_1^\beta$, thereby proving our result. \square

We mention that Lemma 3 compares only the edge flows of I-TAP and β -SO since the path and edge flows coincide for a two edge Pigou network. We also note that while the optimality for a Pigou network may appear restrictive, such networks are of both theoretical [7, 2] and practical significance [31].

4.4 Computational Tractability of I-TAP

Having established that we can solve I-TAP to obtain an approximate solution to β -SO, we now establish that I-TAP can be computed efficiently due to its equivalence to a parametric UE-TAP program.

Observation 1 (UE Equivalency of I-TAP). *For any $\alpha \in [0, 1]$, I-TAP $_\alpha$ reduces to UE-TAP with objective function $\sum_{e \in E} \int_0^{x_e} c_e(y, \alpha) dy$, where $c_e(y, \alpha) = t_e(y) + \alpha y t'_e(y)$.*

To see Observation 1, note that by the fundamental theorem of calculus, $h^{SO}(\mathbf{x})$ can be written as

$$h^{SO}(\mathbf{x}) = \sum_{e \in E} x_e t_e(x_e) = \sum_{e \in E} \int_0^{x_e} t_e(y) + y t'_e(y) dy.$$

Then, taking a convex combination of the SO-TAP and UE-TAP objectives, it is clear that

$$h^{I\alpha}(\mathbf{x}) = \sum_{e \in E} \int_0^{x_e} t_e(y) + \alpha y t'_e(y) dy.$$

Note that for each $\alpha \in [0, 1]$, the differentiability, monotonicity, and convexity of many typical travel time functions t_e , e.g., any polynomial function such as the BPR function [4], imply that the corresponding properties hold for the cost functions $c_e(x_e, \alpha)$ in x_e . For numerical implementation, the equivalency of I-TAP $_\alpha$ and UE-TAP implies that I-TAP $_\alpha$ inherits the useful property that the linearization step of the Frank-Wolfe algorithm [4], when applied to I-TAP $_\alpha$, corresponds to solving multiple unconstrained shortest path queries. The latter motivates the highly efficient approach which we employ in Section 6 to solve I-TAP $_\alpha$.

4.5 Sensitivity Analysis of I-TAP

To obtain a solution that is simultaneously of low cost while keeping within a β bound of unfairness, it is instructive to study the sensitivity of the optimal edge flow solution $\mathbf{x}(\alpha)$ of I-TAP $_\alpha$ in the convex combination parameter α . Specifically, performing such a sensitivity analysis is important since it allows us to characterize the continuity of the SO-TAP objective $h^{SO}(\mathbf{x}(\alpha))$, which we are looking to minimize in the β -SO problem. In this section, we leverage the equivalency of I-TAP and UE-TAP to establish continuity properties of the optimal edge flow solution $\mathbf{x}(\alpha)$ of I-TAP $_\alpha$ and the SO-TAP objective $h^{SO}(\mathbf{x}(\alpha))$.

We first establish the Lipschitz continuity of the optimal edge flow solution $\mathbf{x}(\alpha)$ of I-TAP $_\alpha$ in the convex combination parameter α . This result follows from the UE-TAP reformulation of I-TAP as in Observation 1 and a direct application of the Lipschitz continuity of the optimal solution of a parametric UE-TAP [32, Theorem 8.6a].

Corollary 1. (*Lipschitz Continuity of $\mathbf{x}(\alpha)$*) *For any $\alpha \in (0, 1)$, the edge flow solution $\mathbf{x}(\alpha)$ of the problem I-TAP $_\alpha$ is Lipschitz continuous in α .*

Corollary 1 implies a desirable property that small changes in α will only result in small changes in the optimal edge flows of I-TAP $_\alpha$. We now use Corollary 2 to establish the Lipschitz continuity of the SO-TAP objective.

Corollary 2. (*Lipschitz Continuity of h^{SO}*) *For any $\alpha \in (0, 1)$, let $\mathbf{x}(\alpha)$ be the optimal edge flow solution to I-TAP $_\alpha$. Then, $h^{SO}(\mathbf{x}(\alpha))$ is Lipschitz continuous in α .*

Proof. We prove the Lipschitz continuity of $h^{SO}(\mathbf{x}(\alpha))$ in α through the observation that the composition of Lipschitz continuous functions is Lipschitz continuous. First, observe that $\mathbf{x}(\alpha)$ is Lipschitz continuous by Corollary 1. Next, the SO-TAP objective $h^{SO}(\mathbf{x}) = \sum_{e \in E} x_e t_e(x_e)$ is continuous in its argument \mathbf{x} . Furthermore, $h^{SO}(\mathbf{x})$ is locally Lipschitz over a bounded set since the SO-TAP has a finite derivative as long as \mathbf{x} is bounded. Since both $h^{SO}(\mathbf{x})$ and $\mathbf{x}(\alpha)$ are Lipschitz continuous, it follows that $h^{SO}(\mathbf{x}(\alpha))$ is Lipschitz continuous. \square

While Corollary 2 establishes the continuity of $h^{SO}(\mathbf{x}(\alpha))$ in α , this relation is not necessarily monotone, which we show through numerical experiments in Section 6. We also note that, unlike the SO-TAP objective, the unfairness function $U(\mathbf{f}(\alpha))$ is discontinuous in α and we refer to Section 6 for an explanation of the discontinuity of the unfairness function in the interpolation parameter α .

5 Pricing to Implement Flows

In the previous section we presented a method for computing a solution to β -SO that keeps within a β bound of unfairness and strives to minimize the total travel time. In this section, we leverage the structure of I-TAP to develop pricing mechanisms to collectively enforce the I-TAP solution in the presence of selfish users that independently choose routes to minimize their own travel costs. That is, the prices are set such that the travel cost for users in the same commodity is equivalent, ensuring the formation of an equilibrium. We first consider the case of homogeneous users and show that I-TAP results in a natural marginal-cost pricing scheme. Then, we leverage a linear programming methodology of [16] to set road prices to enforce the flows computed through the I-TAP method for the setting of heterogeneous users.

In this section, for the ease of exposition, we focus our discussion on inducing the optimal edge flow $\mathbf{x}(\alpha)$ of I-TAP $_{\alpha}$. We mention that our approach can naturally be extended to enforcing optimal path flows $\mathbf{f}(\alpha)$ that satisfy a given level of unfairness. In particular, we can consider a setting wherein users are recommended to use a specified path set, e.g., by traffic navigational applications, as given by $\mathbf{f}(\alpha)$ and the tolls set are such that no user will have an incentive to deviate from their recommended paths. Finally, we also mention by the result of Theorem 2 that focusing on the edge flow solution $\mathbf{x}(\alpha)$ is without loss of generality for certain ranges of α since the unfairness bound for any optimal path flow solution $\mathbf{f}(\alpha)$ is guaranteed to be satisfied.

5.1 Homogeneous Pricing via Marginal Cost

In the setting where all users have the same value of time v , the structure of I-TAP $_{\alpha}$ yields an interpolated variant of marginal-cost pricing to induce selfish users to collectively form the optimal edge flow $\mathbf{x}(\alpha)$ of I-TAP $_{\alpha}$. This result is a direct consequence of the equivalence between I-TAP and UE-TAP from Observation 1.

Lemma 4 (Prices to Implement Flows). *Suppose that the edge flow $\mathbf{x}(\alpha)$ is a solution to I-TAP $_{\alpha}$ for some $\alpha \in [0, 1]$. Then $\mathbf{x}(\alpha)$ can be enforced as a UE by setting the prices as $\tau_e = \alpha x_e(\alpha) t'_e(x_e(\alpha))$ for each $e \in E$.*

Proof. Without loss of generality, normalize v to 1 and for notational convenience, denote $x_e = x_e(\alpha)$. To prove this result, note from Observation 1 that I-TAP $_{\alpha}$ is equivalent to UE-TAP with objective $\sum_{e \in E} \int_0^{x_e} t_e(y) + \alpha y t'_e(y) dy$. From the first-order necessary and sufficient KKT conditions [4] of this UE-TAP, for any two paths $P, Q \in \mathcal{P}_k$ with positive flow for a commodity $k \in K$, it must be that $\sum_{e \in P} (t_e(x_e) + \alpha x_e t'_e(x_e)) = \sum_{e \in Q} (t_e(x_e) + \alpha x_e t'_e(x_e))$. Thus, if the prices on each edge are set as $\tau_e = \alpha x_e t'_e(x_e)$, then all users in each commodity incur the same travel cost when using any two paths $P, Q \in \mathcal{P}_k$, establishing that \mathbf{x} is a UE. \square

Note from Lemma 4 that the edge prices are equal to α multiplied by the marginal cost of users on the edges. For this pricing scheme, Lemma 4 guarantees that selfish users will be induced to collectively form the solutions satisfying a specified bound on unfairness obtained through the I-TAP method.

Remark 1. We note that road tolling is not the only mechanism that can be used to induce selfish users to collectively form the flows computed through I-TAP. One of the notable mechanisms beyond road tolling to enforce the SO solution as a UE is that of tradable credit schemes, wherein users can trade initially issued credits freely in a competitive market and spend these credits to use roads with predetermined credit charges [33]. As with the marginal cost pricing scheme in Lemma 4, the equivalence between I-TAP and UE-TAP for any $\alpha \in [0, 1]$ enables us to derive a tradable credit scheme. In particular, following a similar line of reasoning to that used by Yang and Wang [33, Proposition 5], it can be shown that the optimal edge flow $\mathbf{x}(\alpha)$ of I-TAP $_{\alpha}$ for any $\alpha \in [0, 1]$ can be enforced as a user equilibrium through a tradable credit scheme.

5.2 Heterogeneous Pricing via Dual Multipliers

The pricing mechanism in Section 5.1 is inapplicable to the heterogeneous user setting as it would require unrealistically imposing different prices for users with different values of time for the same edges. In this section, we consider heterogeneous users and leverage a linear-programming method [16] to provide a necessary and sufficient condition to induce selfish users to collectively form the edge flow $\mathbf{x}(\alpha)$. We further establish

that $\mathbf{x}(\alpha)$ satisfies this condition for each $\alpha \in [0, 1]$. That is, appropriate tolls can be placed on the roads of the network to induce heterogeneous selfish users to collectively form the equilibrium edge flow $\mathbf{x}(\alpha)$.

Before presenting the pricing scheme, we first extend the notion of a commodity to a heterogeneous user setting. In particular, each user belongs to a commodity $k \in K$ when making a trip on a set of paths \mathcal{P}_k between the same O-D pair and has the value of time $v_k > 0$. Then, under a vector of edge prices $\boldsymbol{\tau} = \{\tau_e\}_{e \in E}$ the travel cost that users in commodity k incur on a given path $P \in \mathcal{P}_k$ under the traffic assignment \mathbf{x} is given by $C_P(\mathbf{x}, \boldsymbol{\tau}) = \sum_{e \in P} (v_k t_e(x_e) + \tau_e)$. Note that more than one commodity may make trips between the same O-D pair, and a user equilibrium forms when the travel cost for all users in a particular commodity is equal. We further note that we maintain the unfairness notion presented in the work thus far even for heterogeneous users. That is, irrespective of the value of time of two users travelling between the same O-D pair, the maximum possible ratio between their travel times can be no more than β .

We now leverage the following result to provide a necessary and sufficient condition that the optimal edge flow $\mathbf{x}(\alpha)$ of I-TAP $_\alpha$ for any $\alpha \in [0, 1]$ must satisfy for it to be enforceable as a UE through road pricing.

Lemma 5. *(Condition for Flow Enforceability [16, Theorem 3.1]) Suppose that the flow \mathbf{x} satisfies the feasibility Constraints (1b)-(1d). Further, consider the linear program with the variables d_P^k , which represents the flow of commodity k routed on path $P \in \mathcal{P}_k$, where \mathcal{P}_k denotes the set of all possible paths for commodity k :*

$$\min_{\substack{d_P^k, \forall P \in \mathcal{P}_k, \\ \forall k \in K}} \sum_{k \in K} v_k \sum_{P \in \mathcal{P}_k} t_P(\mathbf{x}) d_P^k, \quad (6a)$$

$$\text{s.t.} \quad \sum_{P \in \mathcal{P}_k} d_P^k = d_k, \quad \forall k \in K, \quad (6b)$$

$$d_P^k \geq 0, \quad \forall P \in \mathcal{P}_k, k \in K, \quad (6c)$$

$$\sum_{k \in K} \sum_{P \in \mathcal{P}_k: e \in P} d_P^k \leq x_e, \quad \forall e \in E, \quad (6d)$$

with demand Constraints (6b), non-negativity Constraints (6c) and capacity Constraints (6d). Then \mathbf{x} can be enforced as a user equilibrium if and only if Constraint (6d) is met with equality for each edge $e \in E$ at the optimal solution of the linear Program (6a)-(6d).

In particular, if $\mathbf{x}(\alpha)$ satisfies the above necessary and sufficient condition then it can be enforced as a user equilibrium through edge prices set based on the dual variables of the capacity constraints of the above linear program. We now show that $\mathbf{x}(\alpha)$ satisfies the necessary and sufficient condition in Lemma 5.

Lemma 6. *[Heterogeneous User Flow Enforceability] Suppose that the edge flow $\mathbf{x}(\alpha)$ is a solution for I-TAP $_\alpha$ for some $\alpha \in [0, 1]$. Then for the heterogeneous user setting, $\mathbf{x}(\alpha)$ can be enforced as a user equilibrium.*

Proof. To prove this result, from Lemma 5 it suffices to show that Constraints (6d) are met with equality at the optimal solution of the linear Program (6a)-(6d) for the flow $\mathbf{x}(\alpha)$. We now suppose by contradiction that the optimal solution $\tilde{\mathbf{x}}$ to the linear Program (6a)-(6d) for the edge flow $\mathbf{x}(\alpha)$ is such that for at least one edge $e^* \in E$ the Constraint (6d) is met with a strict inequality, i.e., $\tilde{x}_{e^*} < x_{e^*}(\alpha)$. Since the flow $\tilde{\mathbf{x}}$ is optimal for this linear program it follows that $\tilde{\mathbf{x}}_P \geq 0$ for all paths $P \in \mathcal{P}_k$ for all commodities $k \in K$, and that $\sum_{P \in \mathcal{P}_k} \tilde{\mathbf{x}}_P = d_k$ for each commodity $k \in K$ by the constraints of the linear program. Note here that the demand constraints satisfy $\sum_{P \in \mathcal{P}_k} \tilde{\mathbf{x}}_P = d_k$ and the non-negativity constraints $\tilde{\mathbf{x}}_P \geq 0$ imply that the corresponding demand and non-negativity constraints for I-TAP $_\alpha$ are also satisfied for the flow $\tilde{\mathbf{x}}$. By the edge decomposition of path flows it must further hold that $\sum_{k \in K} \sum_{P \in \mathcal{P}_k: e \in P} \tilde{\mathbf{x}}_P = \tilde{x}_e$. Thus, we observe that the flow $\tilde{\mathbf{x}}$ is a feasible solution to I-TAP $_\alpha$.

Next, from the capacity constraint of the linear program it follows that for each $e \in E$ that $\tilde{x}_e \leq x_e(\alpha)$ and that for at least one edge e^* that $\tilde{x}_{e^*} < x_{e^*}(\alpha)$ by assumption. Then, by the monotonicity of the I-TAP $_\alpha$ objective in the edge flows, we observe that $h^{I_\alpha}(\tilde{\mathbf{x}}) < h^{I_\alpha}(\mathbf{x}(\alpha))$, implying that $\mathbf{x}(\alpha)$ is not an optimal solution to I-TAP $_\alpha$, a contradiction. Thus, the flow $\tilde{\mathbf{x}}$ cannot exist, proving our claim that any optimal solution to the Program (6a)-(6d) for the edge flow $\mathbf{x}(\alpha)$ must meet the capacity constraint with equality. \square

Lemma 6 implies that even when users are heterogeneous the edge flow $\mathbf{x}(\alpha)$ can be enforced as an equilibrium flow using tolls set through the dual variables of the capacity constraints of a linear program.

Remark 2. We note that in the proof of Lemma 6, all that we required was that the objective function of I-TAP $_{\alpha}$ is monotonically increasing in the flow on each edge of the network. This suggests a more general sufficient condition for enforcing flows as a user equilibrium. In particular, any flow that satisfies the flow conservation constraints and is the solution of a convex program with some convex objective $f(\mathbf{x})$ that is monotonically increasing in x_e for each $e \in E$ can be induced as a user equilibrium. Note that the I-TAP $_{\alpha}$ objective is a special case of such a function $f(\mathbf{x})$.

6 Numerical Experiments

We now evaluate the performance of our I-TAP method for β -SO on several real-world transportation networks. The results of our experiments not only characterize the behavior of I-TAP but also highlight that, compared to the algorithm in [10], our approach has much smaller runtimes while achieving lower total travel times for most levels β of unfairness. In the following, we describe the implementation details of the I-TAP method and the unfairness metric as well as the data-sets we use. We further present the corresponding results to evaluate the performance of our approach on both the unfairness measure presented in Section 3.2 and additional intra-O-D pair unfairness measures.

6.1 Implementation Details and Data Sets

We tested our I-TAP method using a single-thread C++ implementation of the Conjugate Frank-Wolfe algorithm [34], which we made publicly available along with our data sets and results ([frank-wolfe-traffic](#), [fair-routing](#)). Our implementation is based on a previous repository which was developed for [35]. While there is a rich literature on algorithm design to solve the traffic assignment problem [36, 37], we decided to use the Frank-Wolfe algorithm which was shown recently to be superior in terms of running time [35]. For shortest-path computation in the all-or-nothing routine, we used the LEMON Graph Library [38]. Within the same framework we implemented the solution method for CSO that was presented in [10], where we used `r_c_shortest_paths` within the Boost C++ Libraries [39] for constrained shortest-path search.

To compute the resulting unfairness level for a given path flow \mathbf{f} from those approaches we first recover a path-based solution by recording the paths computed for each commodity k in every Frank-Wolfe iteration, and discarding paths whose relative weight in the final solution is negligible. Denote by \mathcal{P}_k^+ the resulting collection of paths for a given commodity k with strictly positive flow. We also maintain for each path $P \in \mathcal{P}_k^+$ the volume of flow used by the path for this commodity. Then we recover for each commodity k the edge-based solution by computing for each edge $e \in E$ the total flow resulting from the paths \mathcal{P}_k^+ . Then we discard edges from the graph whose flow is 0 for the commodity k , which yields the DAG $G_k = (V_k, E_k)$, with vertices V_k and edges E_k .

Finally, to compute the unfairness level for this commodity, we compute the shortest and longest paths from origin to destination over the graph $G_k = (V_k, E_k)$, where a weight for a given edge $e \in E_k$ is set to be $t_e(x_e)$, i.e., the travel time on the edge given the flows of all the commodities combined with respect to the edge flow solution \mathbf{x} corresponding to the path flow \mathbf{f} . To compute the shortest path over G_k we simply run a Dijkstra search, whose running time is $\Theta(|E_k| + |V_k| \log |V_k|)$. Although for general graphs computing the longest path is NP-hard, for the case of a DAG, we can compute it for the same running time as Dijkstra by negating the edge weights, i.e., using the weights $-t_e(x_e)$ and then finding the shortest path [40].

All results were obtained using a commodity laptop equipped with 2.80GHz 4-core i7-7600U CPU, and 16GB of RAM, running 64bit Ubuntu 20.04 OS. We ran the Frank-Wolfe algorithm for 100 iterations on each data-set, both for I-TAP and the method in [10]. We mention that this number of iterations generally allows the Frank-Wolfe algorithm to achieve a relative error of at most 10^{-5} when searching for UE and SO solutions over larger scenarios of the traffic assignment problem [34].

Table 1 shows the six instances we use for our study, which were obtained from [41]. We use the BPR travel time function [4], defined as

$$t_e(x_e) = \xi_e \left(1 + a \left(\frac{x_e}{\kappa_e} \right)^b \right), \quad (7)$$

Region Name	attributes			runtime (sec.)	
	$ V $	$ E $	$ K $	Jahn et al.	I-TAP
Sioux Falls (SF)	24	76	528	20.0	0.03
Anaheim (A)	416	914	1406	74.0	0.33
Massachusetts (M)	74	258	1113	24.3	0.09
Tiergarten (T)	361	766	644	18.2	0.20
Friedrichshain (F)	224	523	506	19.8	0.12
Prenzlauerberg (P)	352	749	1406	74.4	0.32

Table 1: Problem instance attributes and computation time. For each instance we report the number of vertices $|V|$, edges $|E|$, and OD pairs $|K|$. In addition, we report the computation time of each instance for the previous method of Jahn et al. [10] and our I-TAP method using 100 iterations of the Frank-Wolfe algorithm.

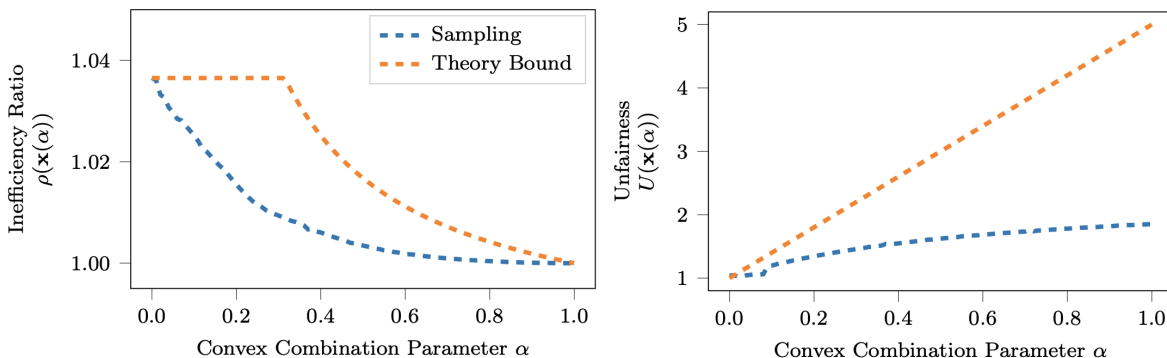


Figure 3: Comparison between the inefficiency ratio of the solution $\mathbf{x}(\alpha)$ of I-TAP on the Prenzlauerberg data-set and the theoretical bound on the inefficiency ratio obtained in Theorem 1 (left). The comparison between the theoretical upper bound of the level of unfairness for any convex combination parameter α and the unfairness level of the solution $\mathbf{x}(\alpha)$ of I-TAP on the Prenzlauerberg data-set is shown on the right. The values of the convex combination parameter were chosen at increments of 0.01.

where a, b are constants, ξ_e is the free-flow travel time on edge e , and κ_e is the capacity of edge e , which is the number of users beyond which the travel time on the edge rapidly increases. Since the constants $a = 0.15$ and $b = 4$ are typically chosen, we use these constants for the numerical experiments.

6.2 Results

Assessment of Theoretical Upper Bounds. We first assess the theoretical upper bounds on the inefficiency ratio and level of unfairness that were obtained in Section 4.2 with respect to the convex combination parameter α . The latter is obtained using a dense sampling method with increments of 0.01. We present the results for the Prenzlauerberg data-set and note that the results and the following discussion extend to other problem instances in Table 1 as well.

Figure 3 (left) depicts both (i) the change in the inefficiency ratio of the solution $\mathbf{x}(\alpha)$ using dense sampling and (ii) the theoretical upper bound of the inefficiency ratio (Theorem 1). As expected, the dense sampling procedure results in an inefficiency ratio that is below the theoretical upper bound for every value of α .

The comparison between the theoretically guaranteed level of unfairness for every value of α as obtained in Theorem 2 and the actual unfairness level of the I-TAP method is depicted in Figure 3 (right). Since the BPR travel time function we used in this work has a degree of four, we have that for any value of α , we can guarantee a level of unfairness of $4\alpha + 1$ by Theorem 2. Figure 3 (right) suggests that the theoretical upper bound is even more conservative for the case of unfairness as there is an even larger gap between the actual solution and the theoretical bound.

These findings further highlight the efficacy of the I-TAP approach for practical applications.

Behavior of I-TAP. For each of the transportation networks in Table 1, we now study the relationship

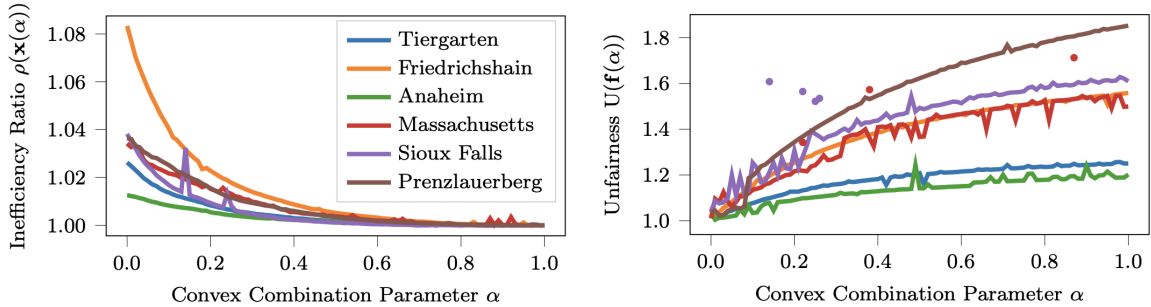


Figure 4: Variation in the inefficiency ratio (left) and the level of unfairness (right) of the optimal solution $\mathbf{x}(\alpha)$ of I-TAP $_{\alpha}$ with the parameter $\alpha \in [0, 1]$ for six different transportation networks from Table 1. The values of the convex combination parameter were chosen at increments of 0.01.

between the convex combination parameter α and the (i) total travel time, and (ii) unfairness. To review these relationships, we consider α to lie in the set \mathcal{A}_s with $s = 0.01$ increments.

Figure 4 (left) shows the relationship between the inefficiency ratio and α . Note that when $\alpha = 1$, the inefficiency ratio is one, since the interpolated objective is the SO-TAP objective, and when $\alpha = 0$, the inefficiency ratio is the Price of Anarchy (PoA), since the interpolated objective is the UE-TAP objective. As shown on the left in Figure 4, the inefficiency ratio is always between one and the PoA for each of the transportation networks, which corroborates the bound on the total travel time for any convex combination parameter, as obtained in Theorem 1. Furthermore, the inefficiency ratio varies continuously in α , which further aligns with the continuity of the total travel time function h^{SO} in α , as is characterized in Section 4.5. The jumps in the inefficiency ratio that can be observed for certain values of α for Sioux Falls aligns with the continuity result since the relative magnitude of the jumps is small. In particular, the change in the total travel time is less than 2% for a 1% change in the value of α .

The relationship between unfairness and α is depicted on the right in Figure 4. For readability of this figure, we marked outliers as points where large changes in the unfairness occur for small changes in α . The large jumps in the unfairness function $U(\mathbf{f}(\alpha))$ for some of the data sets, e.g., Sioux Falls and Anaheim, at particular values of α stem from the fact that the optimal solutions $\mathbf{f}(\alpha_1)$ and $\mathbf{f}(\alpha_2)$ for convex combination parameters α_1 and α_2 that are arbitrarily close may not have the same set of positive paths. In particular, it may happen for some convex combination parameter α_1 that path P is a positive path, but for some $\epsilon > 0$ and any α_2 such that $\|\alpha_1 - \alpha_2\|_2 \leq \epsilon$ we have that P is not a positive path. Since the travel time on path P could be very different from the travel times on other paths, the unfairness function is discontinuous in α .

Finally, for each transportation network the general trend of a decrease in the inefficiency ratio and an increase in the unfairness with an increase in α suggests that decreasing the total travel time comes at the cost of an increase in the unfairness and vice versa.

Solution Quality Comparison. We now explore the efficiency-fairness tradeoff through a comparison of the Pareto frontier of the I-TAP method to the approach in [10], which is a benchmark solution for fair traffic routing, and the I-Solution method described in Section 4.1. To this end, we depict the Pareto frontier of the (i) I-TAP method for 0.01 and 0.05 increments of the parameter α , (ii) I-Solution method for 0.01 increments of the convex combination parameter γ , and (iii) Jahn et al.’s approach [10] for 0.05 increments of the normal unfairness parameter ϕ lying between one and two.

Figure 5 depicts the Pareto frontiers, i.e., the set of all Pareto efficient combinations of system efficiency and user fairness, for the six transportation networks in Table 1. In particular, observe that the Pareto frontiers of the I-TAP method are below that of the other two approaches for most values of unfairness. This observation indicates that the I-TAP method outperforms the other two approaches since the inefficiency ratio of the I-TAP solution is the lowest for most desired levels of unfairness. Only for the Sioux Falls and Prenzlauerberg data-sets, the algorithm in [10] achieved lower inefficiency ratios than both the I-TAP and I-Solution methods for higher values of unfairness, which, in practice, would be undesirable. Furthermore, note that, unlike the two convex-combination approaches, the solution of the algorithm in [10] can result in inefficiency ratios that are much greater than the PoA for low levels of unfairness. The I-TAP method

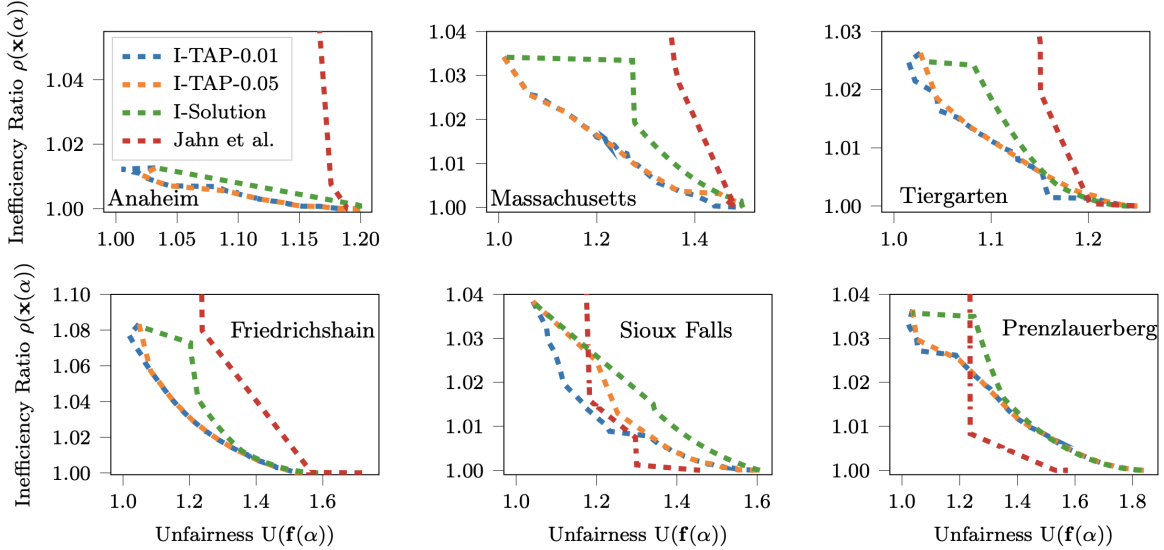


Figure 5: Pareto frontier depicting the trade-off between efficiency and fairness for the (i) I-TAP method with a step size $s = 0.01$, (ii) I-TAP method with $s = 0.05$, (iii) I-Solution method with $s = 0.01$, and (iv) Jahn et al.’s method [10] with $s = 0.05$.

outperforms the I-Solution method since the set of paths that users can traverse is not restricted to the union of the routes under the UE and SO solutions as is the case for the I-Solution method. In particular, there may be traffic assignments with lower total travel times that use paths not encapsulated by the restricted set of paths corresponding to the I-Solution method. Furthermore, while the PoA for each of the data-sets is quite low, some real-world transportation networks may have much higher PoA values (even as high as two) [5], which would make the trade-off between efficiency and fairness even more prominent.

Finally, the Pareto frontiers of the I-TAP method for the 0.01 and 0.05 increments of the convex combination parameter almost overlap each other for all the transportation networks other than Sioux Falls. Since Sioux Falls has a highly discontinuous unfairness function (cf. Figure 4), it is likely that the 0.05 increments of α values may not capture all the low total travel time solutions that keep within a β bound of unfairness that the 0.01 increments of α may be able to capture. For all the other datasets, the near equivalence of the Pareto frontiers for the two α discretizations suggests that a good performance of the I-TAP method can be achieved with coarse discretizations of the convex combination parameter set. Thus, we only need to compute a solution to the convex program I-TAP_α for relatively few values of α to characterize the Pareto frontier, implying the computational efficiency of the I-TAP method.

Runtime Comparison. We report in Table 1 the runtime of the Jahn et al. method [10] and our I-TAP method. For each instance we report the average runtime over the parameters ϕ and α for the competitor and our method, respectively. We observe that our approach is faster by at least three orders of magnitude. This is unsurprising since our method solves *unconstrained* shortest-path queries, which can be implemented in $O(|E| + |V| \log |V|)$ time, within each Frank-Wolfe iteration, whereas [10] solves *constrained* shortest-path queries which are known to be NP-hard. We do mention that a more efficient implementation of constrained shortest-path query can be achieved by directly implementing a label-correcting algorithm rather than using the `r_c_shortest_paths` routine from Boost, which is overly general for our setting and hence less efficient. Nevertheless, even with this improvement it would still be much slower than the unconstrained near-linear algorithm. Notice that both approaches can be sped up via parallel computation of shortest-path queries, and our method can be made even faster through modern heuristics for shortest-path queries in transportation networks, such as contraction hierarchies [42], as in [35].

Unfairness Measure	Formula ($\tilde{U}(\mathbf{f})$)
Envy Free	$\max_{k \in K} \max_{Q, R \in \mathcal{P}_k: \mathbf{x}_Q, \mathbf{x}_R > 0} \frac{t_Q(\mathbf{x})}{t_R(\mathbf{x})}$
Used Nash	$\max_{k \in K} \max_{R \in \mathcal{P}_k^+, Q \in \mathcal{P}_k: \mathbf{x}_Q > 0} \frac{t_Q(\mathbf{x})}{t_R(\mathbf{x})}$
Discrete Gini Coefficient	$\max_{k \in K} \frac{\sum_{P, Q \in \mathcal{P}_k} \mathbf{x}_Q \mathbf{x}_P t_Q(\mathbf{x}) - t_P(\mathbf{x}) }{2d_k \sum_{P \in \mathcal{P}_k} \mathbf{x}_P t_P(\mathbf{x})}$

Table 2: Path-based unfairness Measures that depend on the realization of a specific path flow decomposition \mathbf{f} of an edge flow \mathbf{x} .

6.3 Comparison on Additional Intra-O-D Pair Unfairness Measures

Thus far, we have considered an unfairness notion that compares the ratio of the travel times on positive paths. However, there have been many other unfairness notions that have been investigated in the fair traffic routing literature as well as in other economic applications to evaluate the distribution of a resource among users. In this section, we present a few other commonly-used unfairness measures used in the traffic routing and economics literature (Section 6.3.1) and a numerical study of these unfairness measures (Section 6.3.2) on the six instances in Table 1.

6.3.1 Intra-O-D Pair Unfairness Measures for Path Flows

In this section, we present several additional intra-O-D pair measures that can be used to evaluate the level of unfairness of a traffic assignment. While many unfairness measures for traffic routing have been proposed, most notions rely on a quantification of the discrepancy between user travel times. For instance, [43] measure unfairness through the maximum ratio of the travel times between two users for a given set of path flows \mathbf{f} of the edge flow \mathbf{x} . This definition of unfairness is termed as “envy-free” by [29] since one user can envy another’s path by a factor no more than β . An analogous notion to the envy-free definition is that of a “Used Nash Equilibrium”, wherein the level of unfairness is calculated through the maximum ratio between the experienced travel time for any given user to the travel time on any other positive path between the same O-D pair. Note that the “Used Nash Equilibrium” serves as an intermediary to the unfairness notion presented Section 3.2 and the envy-free notion. There are also several other definitions of max-min fairness proposed in the traffic routing literature, and we defer the interested reader to a discussion of some of these notions to [10].

Beyond the fairness notions considered in the traffic routing literature, there have also been other measures proposed to evaluate fairness or equity in the distribution of a given resource amongst users. One such popular measure is the Gini coefficient [44], which is used to evaluate the level of dispersion of wealth in society. Applying this idea in context of traffic routing, we can use the discrete Gini coefficient measure [45, 46] to evaluate the spread in the travel times of users travelling between the same O-D pair. We summarize the above mentioned unfairness measures in Table 2.

Note that some path decompositions may be more fair according to one metric as compared to another. For instance, consider the situation when a small group of users incur an exceedingly large travel time while the remaining users travelling between the same O-D pair incur a small travel time. In this case, the level of unfairness for the Envy Free and Used Nash measures will typically be high but the level of unfairness as measured according to the Gini measure will be quite low. As a result, each of these metrics has their own merit as well as limitations when applied to measuring the level of unfairness of specific path flow decompositions.

Finally, observe that each of the above defined notions of unfairness involve taking a maximum of the unfairness measure for each O-D pair. We note that different methods of aggregating across O-D pairs can also be used, e.g., averaging the unfairness measure. Furthermore, there is a key aspect of the above defined unfairness measures that makes the I-TAP a particularly natural approach for balancing efficiency and fairness. In particular, for each of the above defined unfairness notions, the user equilibrium achieves the highest possible level of fairness, while the system optimum achieves the lowest total travel times. Thus, interpolating between the UE and SO objectives can potentially achieve the best of both worlds, i.e., a low total travel time with a high level of fairness.

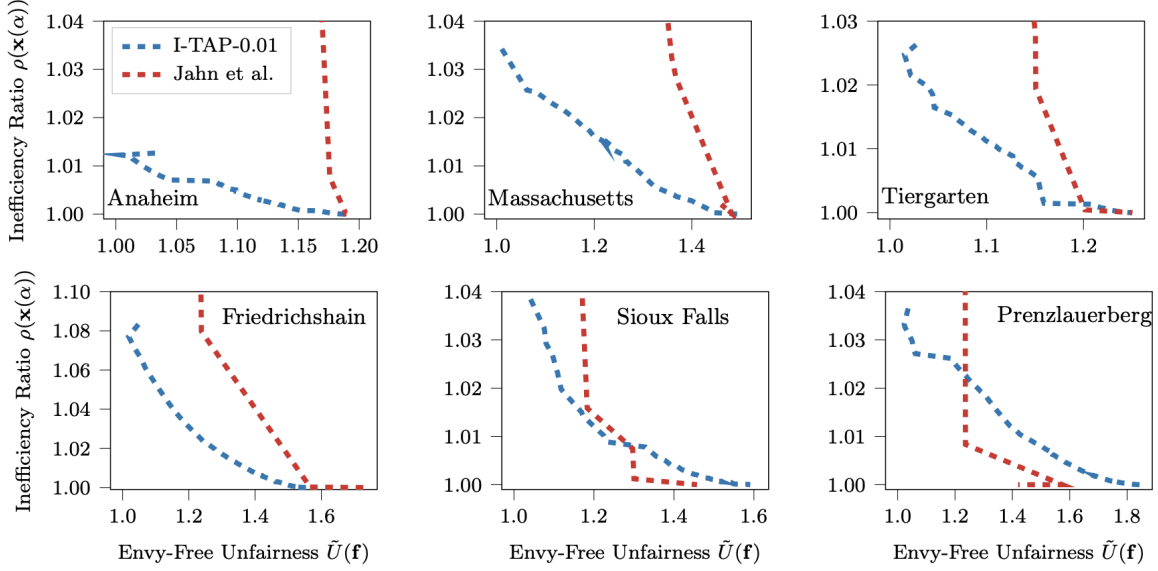


Figure 6: Pareto frontier depicting the trade-off between efficiency and “envy-freeness” for the (i) I-TAP method with a step size $s = 0.01$, and (ii) Jahn et al.’s method with $s = 0.05$.

6.3.2 Results

We now present some numerical results to demonstrate the performance of I-TAP with respect to the above defined path-based unfairness measures. In particular, we use the Frank Wolfe method to compute a specific path decomposition for I-TAP and use this to evaluate the level of unfairness based on the Envy-Free and discrete Gini coefficient metrics. Figures 6 and 7 depict the trade-off between system efficiency and unfairness for the two metrics, Envy-Freeness and the discrete Gini coefficient, respectively.

We note that the Pareto-frontier depicting the Envy-Free metric in Figure 6 is almost identical to the Pareto frontier of the edge-based unfairness metric presented in Section 3.2. This observation suggests that for many practical instances these two notions of unfairness are quite close to each other. We note this despite the fact that the Envy-Free notion involves the maximum ratio between experienced travel times for a given flow decomposition while the edge-based unfairness metric involves the maximum possible ratio between travel times on positive paths and is thus independent of the realized path decomposition. Since the Used Nash measure is an intermediary between the Envy-Free and the edge-based unfairness measure in Section 3.2, we note that it would also correspond to near identical Pareto frontiers for the six tested instances. Thus, we do not depict the Used Nash unfairness measure.

For the discrete Gini coefficient metric, we first note that at the user equilibrium solution the discrete Gini coefficient is zero, rather than having unfairness of one for the earlier explored unfairness measures. This is because all users travelling between the same O-D pair have equal travel times at the user equilibrium and thus the numerator of the discrete Gini coefficient is zero. Note that the further the discrete Gini coefficient is away from zero the more unfair the solution.

We observe from Figure 7 that, modulo numerical errors, the discrete Gini coefficient of the user equilibrium for all the scenarios is very close to zero. As with the earlier explored unfairness measures, we also observe that the I-TAP method outperforms the approach used in [10] even on the discrete Gini coefficient metric for low levels of desired Gini coefficients for all data-sets but Prenzlauerberg.

7 Discussion: Beyond Intra-O-D Pair Fairness

In this work, we investigated the β -SO problem to mitigate unfairness concerns resulting from system optimum routing. In particular, we investigated an unfairness notion (termed as *positive path unfairness*, for

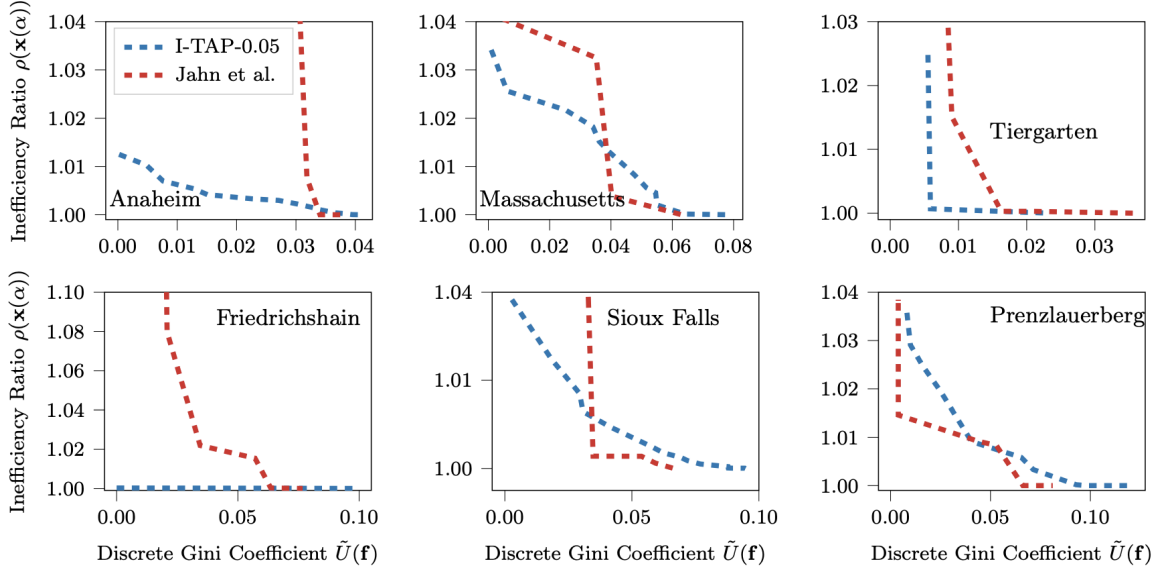


Figure 7: Pareto frontier depicting the trade-off between efficiency and discrete Gini coefficient for the (i) I-TAP method with a step size $s = 0.05$, and (ii) Jahn et al.'s method with $s = 0.05$.

succinctness, in this section) that compares the ratio of travel times of users on positive paths for users traveling between the same O-D pair. The premise behind using positive path unfairness is that it guarantees that the ratio of travel times of all users traveling between the same O-D pair is bounded. Unlike the system optimal solution wherein some users incur the burden of longer travel times in the pursuit of system efficiency, under positive path unfairness constraints, no user faces the burden of excessively long travel times compared to other users traveling between the same O-D pair. Since tolls are a mechanism to enforce desired flow patterns, a consequence of enforcing positive path unfairness constraints is that users with lower incomes, which typically face the burden of longer travel times, now incur travel times similar to those of users with higher incomes from the same O-D pair. In other words, enforcing positive path unfairness constraints helps mitigate the distributional consequences of road congestion pricing by ensuring that users in all income groups traveling between the same O-D pair incur similar travel times bounded by the factor β .

While positive path unfairness, as with other unfairness measures studied in the literature [10, 11, 29], helps mitigate fairness concerns for users traveling between a given O-D pair, several other considerations must be taken into account when designing transportation policy that involves the implementation of such constraints in practice. In particular, the fairness of a traffic assignment can be gauged using several measures beyond positive path unfairness, where different unfairness notions may often conflict with each other [47]. For instance, achieving a low level of unfairness for users traveling between a given O-D pair may raise fairness concerns for users traveling between different O-D pairs. Thus, it is crucial to understand the interplay between multiple notions of unfairness to better inform transportation policy surrounding fair traffic routing.

While developing new algorithms and theoretical results for measures beyond intra-O-D pair unfairness is outside the scope of this work, in this section, we study several examples in which a fair assignment under the positive path unfairness notion may be unfair according to other notions of unfairness. Our purpose behind introducing these examples is to ensure that policymakers are aware of the pros and cons of each unfairness measure and express caution to the fact that no unfairness notion is likely to meet all of the fairness requirements of a policymaker without compromising much on system efficiency. In addition, our presented examples motivate further research in fair traffic routing under alternative notions of unfairness, particularly when considering fairness concerns across O-D pairs. In presenting these examples, we also describe how our proposed unfairness notion could be modified to account for the fairness concerns that arise in these settings and note that these are stylized constructions to demonstrate the limitations of our

proposed unfairness measure and that such stylized setups typically may not occur in practical traffic routing applications. In the following, we summarize our three examples and present their corresponding details in the remainder of this section.

Each of our examples considers the interplay between positive path unfairness and unfairness notions that compare the travel time outcomes of users across different O-D pairs. To this end, we focus on settings with multiple O-D pairs, as positive path unfairness already ensures that the ratio of user travel times for all users traveling between a given O-D pair is bounded. Our first two examples consider settings when the system-optimum solution corresponds to a traffic assignment with a positive path unfairness of one. However, these solutions correspond to users traveling between one of the O-D pairs facing the brunt of the longer travel times to achieve system efficiency. Our third example considers a setting wherein the system-optimum solution corresponds to an outcome with a positive path unfairness of greater than one. In this example, we show how imposing positive path unfairness constraints results in users from one O-D pair becoming much worse off to reduce the degree of unfairness for users belonging to another O-D pair. For this example, we show, however, that the degree to which the travel times of users between one O-D pair are increased depends on the characteristics of the transportation network and the corresponding O-D demands. Thus, there are other settings in which a decrease in the positive path unfairness of users from a given O-D pair may correspond to only a small increase in the travel time of users from the other O-D pair.

We reiterate that these chosen examples and constructions are stylized to demonstrate the limitations of positive path unfairness and, in general, unfairness measures that compare user travel times for a single O-D pair and that such stylized setups are not representative of practical traffic routing contexts. Furthermore, while these examples point towards the consideration of alternative notions of unfairness that consider the travel times of users across different O-D pairs, we note that considering these other unfairness notions can come at a significant loss of system efficiency. Since unfairness notions associated with a single O-D pair may often conflict with unfairness notions across O-D pairs, as elucidated through these examples, the onus is on policy-makers to prioritize certain notions of unfairness for their particular transportation network characteristics and associated O-D demands.

7.1 Example 1: Unfairness across O-D Pairs

Our first example considers a setting of two O-D pairs where the system-optimum traffic assignment achieves a positive path unfairness of one. The network structure and corresponding O-D pairs, given by (o_1, r_1) and (o_2, r_2) , for this example are depicted in Figure 8. In the following, we present O-D demands and edge travel time functions under which the system-optimum solution corresponds to all users between O-D pair (o_1, r_1) traveling on edge e_4 , which has a much longer travel time than that on the other path, i.e., the path $e_1 \rightarrow e_2 \rightarrow e_3$. After presenting this example, we discuss its implications and present a modification to our notion of positive path unfairness that applies to this example, along with its corresponding pros and cons.

Example 1. We consider the setting with a demand of $d_1 = 1$ between the O-D pair (o_1, r_1) , with origin o_1 and destination r_1 , and a demand of $d_2 = 10$ between the O-D pair (o_2, r_2) . We further consider the following travel time functions for the four edges of the traffic network: $t_1(x_1) = \epsilon$, $t_2(x_2) = \frac{x_2}{10}$, $t_3(x_3) = \epsilon$, and $t_4(x_4) = 2$, where $\epsilon > 0$ is a small constant. In this setting, we observe that the system-optimum assignment corresponds to the outcome when all users between the O-D pair (o_1, r_1) use edge e_4 and all users between (o_2, r_2) use edge e_2 , i.e., the system optimal flow is $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*) = (0, 10, 0, 1)$. For a proof of this claim, please see the following paragraph. Next, note that under this outcome, all users traveling between the same O-D pair use the same path and thus the positive path unfairness of the corresponding solution is also one. Despite the fact that the system-optimum solution corresponds to a positive path unfairness of one, we note that users traveling between O-D pair (o_1, r_1) have an alternate path, i.e., the path with edges $e_1 \rightarrow e_2 \rightarrow e_3$, on which they can travel that corresponds to a travel time of $1 + 2\epsilon < 2$ for small $\epsilon > 0$. Thus, while the system-optimum solution is fair corresponding to the notion of positive path unfairness, it may not be fair from the perspective of the users traveling between O-D pair (o_1, r_1) as they have an alternate path with approximately half their current experienced travel time.

Efficiency of Flow \mathbf{x}^ :* We show that the traffic assignment \mathbf{x}^* is indeed the assignment corresponding to the minimum total travel time in the system. To this end, first note that this assignment above corresponds to a total travel time of 12. Next, note that all users between O-D pair (o_2, r_2) must travel on edge e_2 as that is the only feasible path on which these users can travel. Then, suppose that for O-D pair (o_1, r_1) that

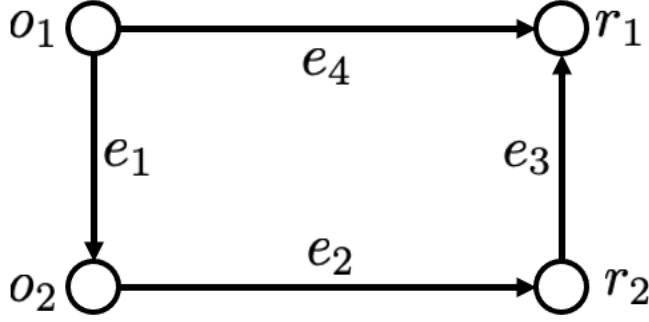


Figure 8: Traffic network with two O-D pairs and four edges to elucidate a setting when a system-optimum traffic assignment has a positive path unfairness of one but still may be perceived as unfair along other unfairness measures. The two O-D pairs are denoted by (o_1, r_1) and (o_2, r_2) , where users between the O-D pair (o_2, r_2) have one feasible path on which they can travel, i.e., edge e_2 , while users between the O-D pair (o_1, r_1) have two feasible paths on which they can travel, i.e., edge e_4 or the path $e_1 \rightarrow e_2 \rightarrow e_3$.

y users travel on edge e_4 and $1 - y$ users travel on the path $e_1 \rightarrow e_2 \rightarrow e_3$. In this case the total travel time of all users is given by

$$2y + \frac{10 + 1 - y}{10}(10 + 1 - y) + 2\epsilon(1 - y) = 12 + \frac{(1 - y)^2}{10} + 2\epsilon(1 - y) \geq 12$$

for any $y \in [0, 1]$. Hence, we have proven our claim that the assignment \mathbf{x}^* where all users between O-D pair (o_1, r_1) use the edge e_4 and all users between O-D pair (o_2, r_2) use the edge e_2 is in fact the system optimal traffic assignment.

Discussion of Example 1: The above example elucidates a setting when the system-optimum traffic assignment has a positive path unfairness of one but still may be unfair along other unfairness measures. In particular, the system-optimum solution may be unfair for users traveling between O-D pair (o_1, r_1) as they have an alternate path with approximately half their current experienced travel time. Observe that the presence of an additional path on which users between O-D pair (o_1, r_1) can travel meant that these users incurred the burden of the longer travel times. On the other hand, users traveling between the O-D pair (o_2, r_2) benefited from the shorter travel times as edge e_2 was the only feasible path on which they could travel to complete their trip.

In practice, users traveling between O-D pair (o_1, r_1) can potentially benefit by misreporting their true O-D pair to (o_2, r_2) and traverse the edges e_1 and e_3 on their own to complete the trip between their original O-D pair in a shorter time. To overcome such a concern, the positive path unfairness notion could be modified from considering the ratio of user travel times on positive paths between a given O-D pair to considering the ratio of the maximum travel time on any positive path to the minimum travel time on any path between each O-D pair. We note, however, that our proposed algorithmic approach and corresponding theoretical results do not directly extend to such alternative notions of unfairness and defer a deeper exploration of such unfairness measures to future research.

Finally, we note from our numerical experiments that the travel times of the shortest positive path between a given O-D pair and the shortest path typically coincide for the traffic networks in Table 1, further highlighting the stylized nature of the above-constructed example. Furthermore, we note that adopting a different unfairness notion for the above example would come at the cost of system efficiency, which suggests the importance of the objectives of a policymaker in terms of which unfairness notion they seek to prioritize relative to the efficiency of the system.

7.2 Example 2: Burden of Longer Travel Times can be Put on One Group of Users

As with our first example, our second example also considers a setting of two O-D pairs where the system-optimum traffic assignment achieves a positive path unfairness of one. However, unlike the previous example wherein users from one of the O-D pairs had just one feasible path, we now consider a setting when users from both O-D pairs have two feasible paths on which they can travel. The network structure and corresponding O-D pairs, given by (o_1, r_1) and (o_1, r_2) , for this example are depicted in Figure 9. In the following, we present O-D demands and edge travel time functions under which the system-optimum traffic assignment achieves a positive path unfairness of one and the burden of the longer travel times is placed on users traveling between one of the O-D pairs, e.g., the O-D pair with a greater share of lower-income users, and discuss its corresponding implications. We further discuss a modification to our unfairness notion that is applicable to this example and comment on how a small degree of randomness in the O-D demands of users can break the symmetry needed for the outcomes observed in this example that are unfair for users traveling between different O-D pairs.

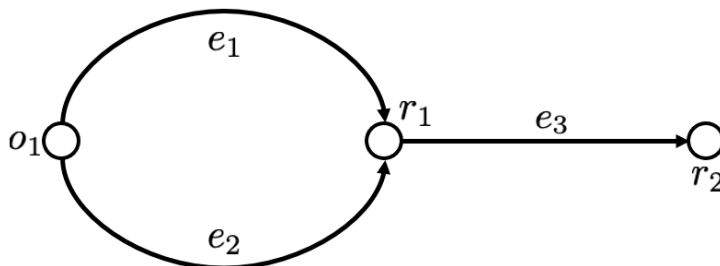


Figure 9: Traffic network with two O-D pairs and three edges to elucidate a setting when a system-optimum traffic assignment has a positive path unfairness of one and still may be perceived unfair when comparing the travel times of users across O-D pairs when using a subset of the network. The two O-D pairs are denoted by (o_1, r_1) and (o_1, r_2) , i.e., all users share the same origin.

Example 2. We consider the setting where there is a demand of $d_1 = \frac{3}{8}$ between the O-D pair (o_1, r_1) and a demand of $d_2 = \frac{5}{8}$ between the O-D pair (o_1, r_2) . Furthermore, let the travel time functions of the edges be such that $t_1(x_1) = 0.5 + x_1$, $t_2(x_2) = x_2$, and $t_3(x_3) = x_3$. In a practical context, the origin o_1 can represent the homes of users, r_1 can represent a location such as a grocery store, and r_2 can represent a destination such as an airport. Furthermore, the edges e_1 and e_2 can be interpreted as two alternate connector routes between the residential part of the city and the commercial part. For the purposes of this example, we assume that all users travelling to the destination r_1 are low-income users while those traveling to the destination r_2 have high-incomes.

Then, it can be verified using calculus that the system efficient traffic assignment corresponds to a solution where $\frac{3}{8}$ users use edge e_1 while $\frac{5}{8}$ users use edge e_2 . Note that one assignment to achieve this split of users across these two edges involves routing all the demand between O-D pair (o_1, r_1) on edge e_1 while routing all the demand between O-D pair (o_1, r_2) on the path $e_2 \rightarrow e_3$. Furthermore, note that such a solution can be enforced through prices in the setting when the values of time of the users scale with their incomes. The corresponding solution achieves a positive path unfairness of one as all users between each O-D pair only traverse one path. However, we observe that despite the fact that users between the two O-D pairs traverse a common set of nodes, i.e., travel between the nodes o_1 and r_1 , the travel times of users traveling between O-D pair (o_1, r_1) are greater than that of the users traveling between O-D pair (o_1, r_2) . As a result, we observe in this example that the burden of the longer travel times in traversing nodes o_1 to r_1 is placed on one group of users, i.e., those belonging to the lower-income section of society.

Discussion of Example 2: While one of our premises behind introducing unfairness constraints to route traffic was to decrease the burden of longer travel times on particular groups of users, i.e., those belonging to the lower-income sections, this example demonstrated that such concerns may not be completely mitigated

by enforcing positive path unfairness constraints if we compare the travel times of users across O-D pairs that traverse a common set of nodes. To this end, we note that the positive path unfairness notion we study can be modified in the above example to compare the ratio of user travel times on subgraphs corresponding to their shared nodes in the graph. In particular, such an unfairness notion involves redefining the true origins and destinations of users to shared origins and destinations to enable a comparison of the travel times of users across shared nodes in the graph. For instance, in the above example, the destination of the O-D pair (o_1, r_2) can be modified to r_1 to compute the unfairness of the solution. We note that such an unfairness notion can be useful for travel during the morning or evening rush hour commutes when a large proportion of users travel between a residential area and the city center, with all users using one of a few paths. While such an unfairness notion is well-defined for the graph structure and O-D demand in Example 2, we do note that extending it to general graphs may be challenging and potentially ill-defined in some cases.

We also note that in the above example, the system-optimum solution corresponded to a positive path unfairness of one, as there was a perfect split of user demands between the two O-D pairs. Note that if the demand $d_1 = \frac{3}{8} + \epsilon$ and $d_2 = \frac{5}{8} - \epsilon$ for any small non-zero ϵ , then the system-optimum solution would have corresponded to a positive path unfairness of strictly greater than one. In other words, the chosen demands in Example 2 correspond to an edge case with the system-optimum solution being deemed fair on the positive path unfairness measure while corresponding to an unfair outcome when comparing the travel times of users across O-D pairs. To this end, we note that even a small degree of randomness in the user demands between the two O-D pairs ensures that the positive path unfairness measure coincides with the unfairness of user travel times across O-D pairs in Example 2. Finally, note that reducing the positive path unfairness of the system-optimum traffic assignment with perturbed user demands will correspond to shifting some of the demand from edge e_2 to edge e_1 until the corresponding positive path unfairness requirement for single O-D pair travel is met, which will also correspond to reduced travel times for the lower income users.

7.3 Example 3: Enforcing Unfairness Constraints Can Worsen Outcomes for Users between Certain O-D Pairs

Thus far, we have considered two examples where the system-optimum solution corresponds to a setting with a positive path unfairness of one. In the following, we consider a setting when the positive path unfairness of the traffic assignment under the system-optimum solution is strictly larger than one. In this context, we show through an example that imposing positive path unfairness constraints can result in differential outcomes for users across different O-D pairs. In particular, we show that decreasing the ratio of user travel times for users traveling between a given O-D pair can result in an increased travel time for users traveling between the other O-D pair. In other words, the imposition of unfairness constraints to improve the fairness outcomes for users traveling between a given O-D pair can increase the travel times of users traveling between another O-D pair. The network structure and corresponding O-D pairs, given by (o_1, r_1) and (o_2, r_2) , for this example are depicted in Figure 10. In the following, we present O-D demands and edge travel time functions under which users belonging to one O-D pair incur a significant increase in their travel times to reduce the ratio of travel times of users belonging to the other O-D pair. We note, however, that our result is sensitive to the chosen travel time functions and O-D demands. To this end, we present an additional example of O-D demands and edge travel time functions to elucidate that imposing positive path unfairness constraints to reduce the ratio of travel times of users between one O-D pair may only have a small impact on user travel times for the other O-D pair. Our presented examples point towards the investigation of transportation network design and the modification of the road properties, e.g., edge travel time functions through changes in road capacities, under which certain notions of unfairness across O-D pairs can be guaranteed to hold.

Example 3. We consider the setting when a demand of 1.5 travels between the O-D pair (o_1, r_1) and a demand of 0.5 travels between the O-D pair (o_2, r_2) depicted in Figure 10. Furthermore, we consider the following travel time functions on the edges of the network: $t_1(x_1) = 2.75$, $t_2(x_2) = x_2^2$, $t_3(x_3) = x_3^2$, $t_4(x_4) = \epsilon$, and $t_5(x_5) = 5$, where $\epsilon > 0$ is a small constant.

For this setting, the system-optimum outcome corresponds to routing all the users traveling between O-D pair (o_2, r_2) on path $e_3 \rightarrow e_4$, 0.5 units of users between O-D pair (o_1, r_1) on the path $e_2 \rightarrow e_3$, and one unit of users on the path given by the edge e_1 . However, observe that the positive path unfairness of this solution for O-D pair (o_1, r_1) is 2.2, as the ratio of the travel time on edge e_1 to that under the path $e_2 \rightarrow e_3$ is 2.2.

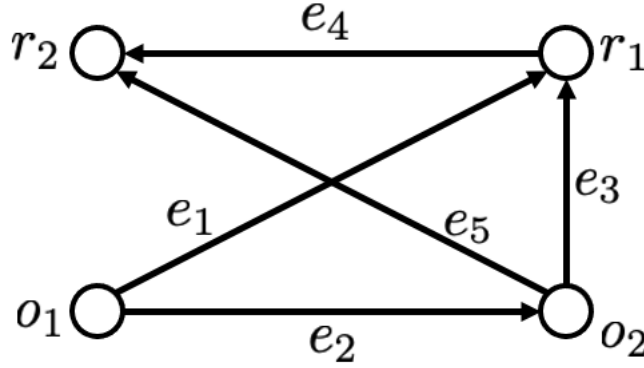


Figure 10: Traffic network with two O-D pairs and five edges to elucidate a setting when a system-optimum traffic assignment achieves a positive path unfairness of strictly greater than one. The two O-D pairs are given by (o_1, r_1) and (o_2, r_2) with two possible paths to travel between each O-D pair.

Next, if we impose an unfairness constraint to ensure a positive path unfairness of one, then some users from edge e_1 will need to be routed on the path $e_2 \rightarrow e_3$. In particular, if $\frac{1}{4}(\sqrt{21} - 1)$ users on edge are routed on the path $e_2 \rightarrow e_3$, then the total travel time of that path becomes 2.75 as now $\frac{1}{4}(\sqrt{21} - 1)$ users traverse edge e_2 while $0.5 + \frac{1}{4}(\sqrt{21} - 1)$ users traverse edge e_3 . In this outcome (which can be enforced under the setting of no tolls), the travel time of all users between O-D pair (o_1, r_1) is 2.75 while that of all users between O-D pair (o_2, r_2) is approximately 1.95.

Thus, we observe that relative to the system-optimum outcome, the above solution has a positive path unfairness of one as all users between the same O-D pair traverse the same path. However, users between O-D pair (o_2, r_2) , which were fair to begin with under the system optimal outcome incur a travel time increase by a factor of approximately 1.95 to reduce the positive path unfairness of the solution of O-D pair (o_1, r_1) . Thus, we observe that imposing positive path unfairness constraints can worsen outcomes for certain O-D pairs that were fair to begin with under the system-optimum outcome.

The above example elucidated a setting when the travel times of users traveling between one O-D pair can be much worse off to achieve a lower level of positive path unfairness for users traveling between another O-D pair. We now present another example wherein we modify the O-D demands and edge travel time functions in the above example to demonstrate that the travel times of users traveling between O-D pair (o_2, r_2) will only increase marginally while reducing the level of positive path unfairness of users traveling between O-D pair (o_1, r_1) .

Example 4. Consider the same graph as in Figure 10 with a demand of 1.5 between the O-D pair (o_1, r_1) and a demand of 0.5 between the O-D pair (o_2, r_2) . Furthermore, we consider the following travel time functions on the edges of the network: $t_1(x_1) = 4.1$, $t_2(x_2) = x_2^2$, $t_3(x_3) = 0.85 + 0.1x_3$, $t_4(x_4) = \epsilon$, and $t_5(x_5) = 1.15$, where $\epsilon > 0$ is a small constant.

Then, the system-optimum outcome corresponds to the setting when all the demand between (o_2, r_2) is routed on the path $e_3 \rightarrow e_4$ and one unit of demand for the O-D pair (o_1, r_1) is routed on the path $e_2 \rightarrow e_3$ while 0.5 units of demand is routed on the path with the edge e_1 . This corresponds to an outcome with a positive path unfairness of 2.05 as the ratio of the travel times of users between O-D pair (o_1, r_1) is $\frac{4.1}{2} = 2.05$. However, if we introduce unfairness constraints to ensure a positive path unfairness of one, then the maximum travel time incurred by users traveling between O-D pair (o_2, r_2) is 1.15. In other words, the travel time of users traveling between O-D pair (o_2, r_2) is at most 1.15 times their travel time prior to the imposition of the unfairness constraints. As a result, in such a setting, we observe that enforcing positive path unfairness constraints to equalize the travel times of users traveling between O-D pair (o_1, r_1) comes with little loss in terms of the increase in the travel times to users from O-D pair (o_2, r_2) .

The above examples elucidate that the sensitivity of the increase in travel times of users across O-D pairs due to the imposition of unfairness constraints hinges significantly on the specific O-D demands and edge

travel time functions. Note that increasing the travel times of users traveling between O-D pair (o_2, r_2) by a factor of almost two, as in Example 3, due to the imposition of positive path unfairness constraints to equalize the travel times of users between O-D pair (o_1, r_1) may be deemed as unacceptable while an increase in their travel times by a factor of at most 1.15, as in Example 4, may be acceptable. As a result, our presented examples point towards the investigation of transportation network design and the modification of the road properties, e.g., edge travel time functions, under which enforcing unfairness constraints with a given O-D pair will result in changes in user travel times across O-D pairs that are deemed “acceptable”.

8 Conclusion and Future Work

In this paper, we developed (i) a computationally efficient method for traffic routing that trades-off system efficiency and user fairness, and (ii) pricing schemes to enforce fair traffic assignments as a UE. We introduced the I-TAP method, which involves solving interpolated traffic assignment problems by taking a convex combination between the UE-TAP and SO-TAP objectives, to find an efficient feasible traffic assignment to the β -SO problem. We then established various solution properties of I-TAP, including computational tractability and solution continuity, and developed theoretical bounds on the inefficiency ratio and unfairness level in terms of the convex combination parameter of I-TAP. To enforce the traffic assignments outputted by the I-TAP method as a UE, we developed a marginal-cost pricing scheme when users are homogeneous, and a linear programming based pricing scheme when users are heterogeneous. We then presented numerical experiments to evaluate the performance of our approach on real-world transportation networks. The results indicated that our algorithm is not only very computationally efficient but also generally results in traffic assignments with lower total travel times for most levels of unfairness as compared to the algorithm in [10]. Finally, we presented a discussion regarding the consideration of unfairness notions that compare users’ travel times across different O-D pairs.

There are various directions for future research. First, it would be worthwhile to develop theoretical bounds for I-TAP to demonstrate its applicability to other notions of unfairness, some of which are studied in Section 6.3, and extend its optimality beyond two-edge Pigou networks. Next, it would be interesting to investigate fairness notions that compare user travel times across different O-D pairs (see Section 7). Finally, given the significant computational advantages of the I-TAP method, it would be interesting to study the generalizability of our approach when accounting for costs beyond the travel times of users, such as environmental pollution and user discomfort.

References

- [1] Sean Fleming. Traffic congestion cost the us economy nearly \$87 billion in 2018, 2019. Accessed February 19, 2021.
- [2] Tim Roughgarden and Éva Tardos. How bad is selfish routing? *J. ACM*, 49(2):236–259, March 2002.
- [3] Tim Roughgarden and Éva Tardos. Bounding the inefficiency of equilibria in nonatomic congestion games. *Games and Economic Behavior*, 47(2):389 – 403, 2004.
- [4] Yossi Sheffi. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, New Jersey, 1 edition, 1985.
- [5] Jing Zhang, Sepideh Pourazarm, Christos Cassandras, and Ioannis Paschalidis. The price of anarchy in transportation networks: Data-driven evaluation and reduction strategies. *Proceedings of the IEEE*, 106(4):538–553, 2018.
- [6] Guni Sharon, Michael Albert, Tarun Rambha, Stephen Boyles, and Peter Stone. Traffic optimization for a mixture of self-interested and compliant agents. In *ISAIM*, 2018.
- [7] Arthur Pigou. *Wealth and Welfare*. Macmillan and Co., London, 1 edition, 1912.
- [8] John Glen Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3):325–362, 1952.

- [9] Hai Yang and Hai-Jun Huang. *Mathematical and Economic Theory of Road Pricing*. Emerald Publishing, Oxford, 1 edition, 2005.
- [10] O. Jahn, R. Möhring, A. Schulz, and N. Stier-Moses. System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations Research*, 53(4):600–616, 2005.
- [11] Tim Roughgarden. How unfair is optimal routing? In *ACM-SIAM Symposium on Discrete Algorithms*, 2002.
- [12] Enrico Angelelli, Idil Arsik, Valentina Morandi, Martin Savelsbergh, and Maria Grazia Speranza. Proactive route guidance to avoid congestion. *Transportation Research Part B: Methodological*, 94:1–21, 2016.
- [13] Enrico Angelelli, Valentina Morandi, Martin Savelsbergh, and Maria Grazia Speranza. System optimal routing of traffic flows with user constraints using linear programming. *European Journal of Operational Research*, 2020.
- [14] Enrico Angelelli, Valentina Morandi, and Maria Grazia Speranza. Congestion avoiding heuristic path generation for the proactive route guidance. *Computers & Operations Research*, 99:234–248, 2018.
- [15] Enrico Angelelli, Valentina Morandi, and Maria Grazia Speranza. Minimizing the total travel time with limited unfairness in traffic networks. *Computers & Operations Research*, 123:105016, 2020.
- [16] Lisa Fleischer, Kamal Jain, and Mohammad Mahdian. Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games. In *IEEE Symposium on Foundations of Computer Science*, 2004.
- [17] Aida Rahmattalabi, Shahin Jabbari, Himabindu Lakkaraju, Phebe Vayanos, E. Rice, and Milind Tambe. Fair influence maximization: A welfare optimization approach. *CoRR*, abs/2006.07906, 2020.
- [18] Dimitris Bertsimas, Vivek Farias, and Nikolaos Trichakis. On the efficiency-fairness trade-off. *Management Science*, 58(12):2234–2250, 2012.
- [19] Ana-Andreea Stoica, Jessy Xinyi Han, and Augustin Chaintreau. Seeding network influence in biased networks and the benefits of diversity. In *Proceedings of The Web Conference 2020, WWW '20*, page 2089–2098, New York, NY, USA, 2020. Association for Computing Machinery.
- [20] Lily Hu and Yiling Chen. Fair classification and social welfare. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20*, page 535–545, New York, NY, USA, 2020. Association for Computing Machinery.
- [21] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. Group-fairness in influence maximization. *CoRR*, abs/1903.00967, 2019.
- [22] Robert Dial. Bicriterion traffic assignment: Basic theory and elementary algorithms. *Transportation Science*, 30(2):93–111, 1996.
- [23] Robert Dial. Bicriterion traffic assignment: Efficient algorithms plus examples. *Transportation Research Part B: Methodological*, 31(5):357–379, 1997.
- [24] Andreas Schulz and Nicolás Stier-Moses. Efficiency and fairness of system-optimal routing with user constraints. *Networks*, 48(4):223–234, 2006.
- [25] Vedat Bayram, Barbaros Tansel, and Hande Yaman. Compromising system and user interests in shelter location and evacuation planning. *Transportation Research Part B: Methodological*, 72:146–163, 2015.
- [26] Bryce Ferguson and Jason Marden. The impact of fairness on performance in congestion networks. In *2021 American Control Conference (ACC)*, pages 4521–4526, 2021.
- [27] Marin Lujak, Stefano Giordani, and S. Ossowski. Route guidance: Bridging system and user optimization in traffic assignment. *Neurocomputing*, 151, 09 2014.

- [28] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Theoretical Aspects of Computer Science*, 1999.
- [29] Soumya Basu, Ger Yang, Thanasis Lianeas, Evdokia Nikolova, and Yitao Chen. Reconciling selfish routing with social good. *CoRR*, abs/1707.00208, 2017.
- [30] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. Pricing network edges for heterogeneous selfish users. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 521–530, New York, NY, USA, 2003. Association for Computing Machinery.
- [31] Caltrans. Us 101 south, corridor system management plan, 2010, 2010. Accessed July 01, 2020.
- [32] Georg Still. Lectures on parametric optimization: An introduction. Optimization Online, March 2018.
- [33] Hai Yang and Xiaolei Wang. Managing network mobility with tradable credits. *Transportation Research Part B: Methodological*, 45(3):580–594, 2011.
- [34] Maria Mitradjieva and Per Olov Lindberg. The stiff is moving - conjugate direction frank-wolfe methods with applications to traffic assignment. *Transp. Sci.*, 47(2):280–293, 2013.
- [35] Valentin Buchhold, Peter Sanders, and Dorothea Wagner. Real-time traffic assignment using engineered customizable contraction hierarchies. *ACM J. Exp. Algorithmics*, 24, December 2019.
- [36] Hillel Bar-Gera. Origin-based algorithm for the traffic assignment problem. *Transportation Science*, 36(4):398–417, 2002.
- [37] Hillel Bar-Gera. Traffic assignment by paired alternative segments. *Transportation Research Part B: Methodological*, 44(8):1022–1046, 2010.
- [38] Balázs Dezső, Alpár Jüttner, and Péter Kovács. Lemon—an open source c++ graph template library. *Electronic Notes in Theoretical Computer Science*, 264(5):23–45, 2011.
- [39] Boost C++ Libraries. www.boost.org, 2010.
- [40] Robert Sedgewick and Kevin Wayne. *Algorithms, 4th Edition*. Addison-Wesley, Princeton University, 2011.
- [41] Transportation Networks for Research Core Team. Transportation networks for research. github.com/bstabler/TransportationNetworks, 2016. Accessed January 20, 2021.
- [42] R. Geisberger, P. Sanders, D. Schultes, and C. Vetter. Exact routing in large road networks using contraction hierarchies. *Transportation Science*, 46(3):388–404, 2012.
- [43] José Correa, Andreas Schulz, and Nicolás Stier-Moses. Fast, fair, and efficient flows in networks. *Operations Research*, 55(2):215–225, 2007.
- [44] Frank Farris. The gini index and measures of inequality. *The American Mathematical Monthly*, 117(10):851–864, 2010.
- [45] Di Wu, Yafeng Yin, Siriphong Lawphongpanich, and Hai Yang. Design of more equitable congestion pricing and tradable credit schemes for multimodal transportation networks. *Transportation Research Part B: Methodological*, 46(9):1273–1287, 2012.
- [46] Devansh Jalota, Kiril Solovey, Karthik Gopalakrishnan, Stephen Zoepf, Hamsa Balakrishnan, and Marco Pavone. When efficiency meets equity in congestion pricing and revenue refunding schemes. *CoRR*, abs/2106.10407, 2021.
- [47] Jon Kleinberg. Inherent trade-offs in algorithmic fairness. *SIGMETRICS Perform. Eval. Rev.*, 46(1):40, jun 2018.