Fair network design problem: an application to EV charging station capacity expansion

Nagisa Sugishita[1], Ismail Sevim[1], Margarida Carvalho[1], Amira Dems[2], and Ribal Atallah[2]

[1]CIRRELT and Département d'informatique et de recherche opérationnelle, Université de Montréal

[2]Institut de Recherche d'Hydro-Quebec

Abstract

This study addresses the bilevel discrete network design problem (DNDP) with congestion, with special emphasis on fairness. The upper-level decision-maker (the network designer) selects a set of arcs to add to an existing transportation network, while the lower-level decision-makers (drivers) respond by choosing routes that minimize their individual travel times, resulting in user equilibrium. Most existing works in the literature primarily focus on minimizing the total travel time of all drivers and extending their solution approaches to other metrics remains a major research challenge. In this paper, we consider objectives related to individual travel times and fairness. To optimize such metrics, we propose a novel single-level reformulation of the DNDP based on strong duality of the lower-level problem.

To evaluate the performance of our method, we conduct numerical experiments on academic DNDP instances. We further demonstrate the practical relevance of our method through a case study on electric vehicle (EV) charging station capacity expansion. In this context, we introduce a fairness-based metric, the cost of sustainability, to quantify the inefficiency caused by EV adoption relative to a scenario where no charging is required. We then optimize expansion decisions to improve this metric. Experiments on a real-world road network in Quebec, incorporating existing public charging infrastructure, highlight the effectiveness and flexibility of our approach.

Acknowledgments: This research was supported by Hydro-Québec, NSERC Collaborative Research and Development Grant CRDPJ 536757 - 19, and the FRQ-IVADO Research Chair in Data Science for Combinatorial Game Theory.

Keywords: Network design, Bilevel programming, User equilibrium, Fairness

1 Introduction

The network design problem (NDP) is a fundamental challenge in transportation and infrastructure planning, with widespread applications in urban mobility and logistics. In this problem, a network designer first determines how to modify an existing transportation network, after which drivers respond by selecting routes that minimize their individual travel times, leading to the so-called Wardrop user equilibrium. The effect of congestion is modeled by using arc travel time functions that depend on their traffic volumes. Under certain conditions on the travel functions, the equilibrium computation can be formulated as a convex optimization problem (Beckmann et al. [1955]). Exploiting this formulation, the NDP can be seen as a bilevel programming problem where the leader represents the network designer and the follower represents the drivers using the transportation network. In most existing literature, the leader aims to minimize the total travel time of all drivers, disregarding other metrics such as fairness or individual travel times.

The NDP is typically classified into three categories: the discrete network design problem (DNDP), which involves binary decisions, such as the addition of new arcs; the continuous network design problem (CNDP), which involves continuous decisions, such as the capacity expansion of arcs; and the mixed network design problem (MNDP), which combines both the CNDP and the DNDP [Magnanti and Wong, 1984]. In this work, we focus on the DNDP.

Various solution methods for the DNDP have been studied over the past decades. However, most studies assume that the objective of the upper-level problem is the total travel time, and the developed solution methods are specifically tailored to this objective function. In some contexts, other metrics beyond total travel time may be of interest to the network designer. One such metric is fairness. Consider a road network where all drivers use conventional cars of the same specifications. The user equilibrium can be considered fair because, at equilibrium, two drivers with the same origin and destination experience the same travel time (Patriksson [2015]). However, when comparing different setups, drivers may perceive some degree of unfairness. For example, if there is no congestion on the road (e.g., during an off-peak period), drivers can travel in a shorter time. This difference is referred to as free-flow unfairness by Jahn et al. [2005].

We can also consider the discrepancy between different modes of transportation. As an example, suppose we are given a road network with electric vehicles (EVs) and conventional cars. Typically, an EV requires a non-negligible amount of time to charge the battery. Furthermore, often charging stations are scarce and EV drivers need to deviate from their preferred routes to visit a charging station. Therefore, EV drivers may spend extra time visiting and using charging stations along their trips, which would not be necessary if they used conventional cars. This extra time spent by EV drivers can be seen as a metric to indicate the inconvenience/inefficiency experienced by the EV drivers. The charging station operator may be interested in finding the charging station locations accounting for such a metric.

A significant challenge arises when incorporating alternative objectives, such as fairness, into existing exact solution methods for the DNDP [Farvaresh and Sepehri, 2013, Rey and Levin, 2025]. For example, as we will demonstrate in this study, the aforementioned fairness metrics are nonconvex (with respect to the flow variables, the variables used in the exact solution methods in the literature). As a result, the formulation becomes a nonconvex mixed-integer nonlinear program (MINLP), which is challenging to solve to global optimality.

Contributions. In this paper, we develop a novel modeling framework for the DNDP that facilitates the incorporation of a broad class of objective functions. The proposed framework relies on a single-level reformulation of the DNDP derived via strong duality. In particular, we leverage the duality theory of monotropic programming, originally developed by Rockafellar Rockafellar [1981]. To the best of our knowledge, this is the first study in the bilevel programming literature to employ monotropic programming duality. The resulting model is a convex mixed-integer nonlinear programming (MINLP) formulation of the DNDP and is amenable to solution techniques developed for convex MINLPs. Furthermore, the proposed reformulation is sufficiently flexible to accommodate alternative objective criteria beyond total travel time, including measures such as free-flow unfairness.

To showcase the use of our approach, we consider an application of the DNDP for the capacity expansion problem of EV charging stations. As concerns over climate change grow, the transportation sector faces increasing pressure to enhance its environmental sustainability. This pressure is particularly acute given the sector's substantial share of greenhouse gas emissions. For instance, recent figures show it is responsible for 28% of total emissions in the United States [U.S. Environmental Protection Agency, 2024], approximately 25% in the European Union [European Environment Agency, 2024], and 28% in Canada [Statistics Canada, 2023]. EVs are expected to play a key role in the transition toward sustainability. To promote EV adoption, it is crucial to ensure reliable access to public fast-charging stations through effective deployment, a priority reflected, for example, in the Government of Canada's commitment to help fund over 84,500 new charging stations by 2029 [Natural Resources Canada, 2024]. Henceforth, we will refer to fast-charging stations, also known as level 3 charging stations, simply as charging stations. The DNDP provides a framework for modeling the optimal sizing and placement of charging stations by considering the behaviors and interactions of EV drivers. We apply our formulation to optimize the total travel time and other inefficiency metrics (formally defined later) related to EV usage. We provide a case study using the transportation network of the province of Quebec. Our computational results showcase the sensitivity of the solution to the objective, highlighting the importance of appropriately choosing the objective for the upper-level problem.

Paper Structure. The paper is organized as follows. Section 2 provides a literature review. In Section 3, we outline formulations of the traffic assignment problem, the lower-level problem of the DNDP. Section 4 presents the DNDP and our single-level reformulation. We provide numerical experiments using test instances of the DNDP in Section 5. In Section 6, the capacity expansion problem of EV charging stations is formulated as the DNDP, and we provide numerical experiments based on the Quebec road network. We conclude the paper in Section 7.

2 Literature Review

The NDP involves the modification or installation of transportation network components, such as arcs. In this work, we primarily focus on the DNDP, the NDP with discrete decisions. Below, we first review related solution methods for the DNDP. We then review the studies that incorporate fairness metrics into the DNDP. We conclude this section by examining models related to the siting and sizing of EV charging stations that inform the context of our case study.

Network Design Problem: Solution Methods. Methods for solving the NDP can be classified into three main categories: branch-and-bound algorithms, single-level reformulations, and linear approximation methods.

The first category, branch-and-bound algorithms, represents one of the most commonly used exact approaches for solving the DNDP. Leblanc [1975] pointed out that the high-point relaxation (HPR) can be used to compute a dual bound. By relaxing the upper-level budget constraint, a further relaxation of the HPR was obtained, enabling faster computation of the dual bound. The relaxation was used within a branch-and-bound algorithm to solve the DNDP exactly. To solve the HPR efficiently, Farvaresh and Sepehri [2013] employed an outer-approximation method, while Bagloee et al. [2017] applied Benders' decomposition. More recently, Rey and Levin [2025] proposed a hybrid approach that combines outer-approximation with column generation.

Another approach is to transform the bilevel DNDP into a single-level formulation. Gao et al. [2005] proposed an approach to transform the bilevel DNDP into a single-level problem using the support function. Later, Farvaresh and Sepehri [2013] identified a flaw in the approach and demonstrated that it may yield suboptimal solutions. Luathep et al. [2011] formulated the MNDP as a single-level problem by enforcing user equilibrium through a variational inequality. The resulting formulation is a nonconvex MINLP, where nonlinear terms were approximated using piecewise-linear functions. Wang et al. [2015] developed an outer-approximation method to handle the variational inequality and achieve a globally optimal solution. Their approach was tested only on the instance presented by Gao et al. [2005]. The value function of the lower-level problem can be used to derive a single-level reformulation as well. For instance, Meng et al. [2001] used the value function to derive a nonconvex single-level formulation of the CDNP and developed a locally convergent algorithm based on the augmented Lagrangian method. The value function was also used by Rey and Levin [2025] to derive a valid inequality.

Several linear approximation methods have also been proposed. To solve the DNDP approximately, Farvaresh and Sepehri [2011] replaced the lower-level problem with the Karush–Kuhn–Tucker (KKT) conditions. This single-level reformulation yields a non-convex MINLP, and they linearized the nonlinear terms, transforming the problem into a single-level mixed-integer linear programming (MILP) formulation. Fontaine and Minner [2014a] linearized the travel time functions and approximated the lower-level problem using linear programming, which was then reformulated as a single-level MILP. They solved the resulting problem using Benders' decomposition. While these methods are not exact due to the piecewise-linear approximations, numerical experiments suggest they can produce high-quality solutions. For a computational study of linear approximation methods, see Rey [2020]. We also note that similar linearization approaches have been applied to the CNDP. For example, Wang and Lo [2010] addressed the CNDP by replacing the lower-level problem with the KKT conditions and linearizing the nonlinear terms.

In the general bilevel programming literature, a common approach for problems with a linear lower-level program is to derive a single-level reformulation using strong duality (Dempe [2002]). Within the NDP literature, the only known application of this approach is due to Fontaine and Minner [2014b], who first linearized the nonlinear terms in the lower-level problem and approximated it by a linear program. In contrast, the present work applies strong duality directly to the lower-level problem without resorting to linearization, thereby avoiding approximation error and yielding a reformulation equivalent to the original DNDP. This is achieved by leveraging the duality theory of monotropic programming, developed by Rockafellar [1981]. To the best of our knowledge, this represents the first use of monotropic programming duality within the bilevel programming literature.

Network Design Problem: Models based on Fairness. All of the methods reviewed above assume that the objective is to minimize total travel time. However, in many real-world applications, alternative performance metrics, such as fairness, are essential for producing socially acceptable solutions.

More broadly, there is a growing body of work in operations research that incorporates fairness considerations into optimization models. For example, Liu and Salari [2024] studied a facility location problem (FLP) with congestion, which can be viewed as a special case of the NDP on a bipartite network, and proposed a model that minimizes the worst expected total travel time experienced by users. Aboolian et al. [2022] examined a related variant in which cost-effective facility locations are sought subject to an upper bound on the worst user travel time. Digehsara et al. [2025] analyzed a fair FLP in which users follow a rank-based choice model without congestion under demand uncertainty, while Ljubić et al. [2024] developed decomposition algorithms for optimizing fairness-related metrics in the FLP.

In contrast, fairness considerations remain relatively limited in the NDP literature. Early work by Friesz et al. [1993] investigated a multiobjective version of the NDP that simultaneously minimized total user transportation costs, travel distance, construction costs, and residential displacement due to new infrastructure. Similarly, Meng and Yang [2002] studied a multiobjective NDP accounting for changes in individual travel times induced by network expansions. Both studies aggregated multiple objectives using weighted-sum formulations and applied simulated annealing to solve the resulting problems. Unlike these approaches, the present work develops an exact solution approach capable of accommodating fairness-related objectives.

Sizing and Placement of EV Charging Stations. Initially, the literature on EV (or alternative fuel vehicle) infrastructure planning focused on simple models that did not consider congestion. One of the most influential models in this literature is the Flow Refueling Location Problem (FRLP), introduced by Kuby and Lim [2005]. In their model, each demand is associated with an origin, a destination, and a flow volume, and is considered served if there are enough charging stations to enable drivers to complete their journeys without running out of battery. In the FRLP, drivers are assumed to follow the shortest path, which is determined based on travel time without considering congestion. Later, Kim and Kuby [2012] extended the model to allow drivers to deviate from the shortest paths. More specifically, given a predetermined deviation threshold, a driver is considered served if they can travel from their origin to their destination along a path whose length does not exceed their shortest path by more than the allowed deviation. Arslan et al. [2019] and Göpfert and Bock [2019] independently proposed branch-and-cut algorithms to solve this variant of the FRLP. Again, this model is based on travel distance or travel time, without accounting for congestion.

Recently, there has been growing interest in modeling EV charging station infrastructure planning while considering congestion and user interaction. Typically, such models are formulated as bilevel programming problems, where the upper-level problem corresponds to the charging station operator and the lower-level problem models the behavior of EV drivers. The upper-level problem seeks to minimize costs, emissions, or travel times, while the lower-level problem captures driver behavior, including route choices, waiting times at charging stations, and vehicle range constraints. For example, Kınay et al. [2023] used logic-based Benders decomposition to minimize the installation costs of charging stations. Zhang et al. [2023] considered a bilevel programming problem in which the upper level is a multi-objective optimization problem aimed at minimizing both total costs and total service time. However, both Kınay et al. [2023] and Zhang et al. [2023] assumed that drivers do not take congestion into account when choosing their routes.

Zheng et al. [2017], Bao and Xie [2021], and Tran et al. [2021] employed the DNDP to model EV infrastructure planning. In these studies, the upper-level decision-maker determines the locations of charging stations, while the lower-level problem is formulated as a user equilibrium model that accounts for road congestion. Jing et al. [2017] formulated a bilevel DNDP problem in which the lower-level models drivers' behavior as a stochastic user equilibrium. These studies focus on road congestion rather than congestion at charging stations and use heuristics to find feasible solutions efficiently. Mirheli and Hajibabai [2023] developed a model to jointly optimize charging station locations and pricing. In their model, drivers choose routes based on travel time and cost, but congestion effects are not considered. He et al. [2018] considered a user equilibrium model that includes both travel time and charging time, where the latter is defined as a function of the required energy. However, congestion or waiting time at charging stations is not considered. In our case study, we explicitly model congestion at charging stations, optimize various metrics, including those related to fairness, by expanding stations' capacities, and develop a convex MINLP formulation.

3 User Equilibrium

In this section, we review various concepts related to the user equilibria, which we will use to define the DNDP in Section 4. First, in Section 3.1, we define some notations and assumptions. Then, in Section 3.2, we present the user equilibrium models. Lastly, in Section 3.3, we discuss the free-flow unfairness, which will be used as an objective of the DNDP in Section 4. A table of notation can be found in Appendix A.

3.1 Notation and Assumption

Let $G = (V, A)$ be a directed graph. The set A contains all the *existing arcs* as well as *candidate arcs* that can be built. For each arc $a \in A$, we associate a binary variable x_a to indicate its *availability*: The value of x_a is 1 if arc a is available and 0 otherwise. We write the set of feasible values of x as $X \subset \{0,1\}^A$. For example, X can model a budget constraint. If arc $a \in A$ already exists, we have $x_a = 1$ for any $x \in X$.

Denote by K the index set of the origin-destination (OD) pairs of nodes in the network. For each $k \in K$, we denote its *origin*, *destination* and *demand volume* by o_k, d_k and e_k, respectively. Furthermore, for each $a \in A$, let θ_a be the *travel time function*, the function that maps the link flow on a (total traffic volume on a) to its travel time.

Throughout this paper, we assume the following:

Assumption 1. *a) For each demand $k \in K$, there is a path from o_k to d_k that only uses existing arcs.*

b) For each arc $a \in A$, the travel time function is given by

$$\theta_a(s) = f_a + g_a s^p, \tag{1}$$

where $f_a \geq 0$ and $g_a \geq 0$ are non-negative constants, and $p \geq 1$ is a positive integer. The function is defined to be zero for $s < 0$, i.e., $\theta_a(s) = 0$ when $s < 0$.

These assumptions are not restrictive. If the road network is already well-connected, all the OD pairs are traversable even without building any new arcs. We also note that the collection of travel time functions modeled by (1) includes the Bureau of Public Roads (BPR) function (Patriksson [2015]) as well as a constant travel time. All the test instances we use in our numerical experiments (Sections 5 and 6) satisfy Assumption 1.

Many formulations of the bilevel DNDP (see, for example, Farvaresh and Sepehri [2013], Wang et al. [2013]) model the flow of each demand $k \in K$ along each arc $a \in A$. We use

$y \in \mathbb{R}^{K \times A}$ to denote the variable representing individual flows. When the underlying graph is large, the flow vector y involves a large number of variables. Interestingly, the bilevel DNDP can be modeled using aggregated flow variables, which are obtained by summing the flows of all OD pairs that share the same origin. This aggregation does not affect the set of feasible flows in the sense that there is a one-to-one mapping between the set of feasible flows and the set of feasible aggregated flows. The aggregation reduces the number of variables, at the cost of a weaker formulation of the bilevel DNDP. In particular, the continuous relaxation of the reformulation provides a weaker dual bound when the flow aggregation is used. As a result, whether this aggregation leads to a computational speed-up depends on the specific instance. This technique is well known in the (single-level) multicommodity flow literature; see the discussion in Chouman et al. [2017] and references therein. To the best of our knowledge, however, flow aggregation has not been exploited in the literature on the bilevel DNDP under user equilibrium. Our preliminary experiments showed that aggregating the flow variables reduced solution times (see Appendix B). Therefore, in this paper, we consider formulations where the flow variables are aggregated. To this end, let $O = \{v \in V : \exists k \in K, o_k = v\}$ denote the set of all origins. For each $o \in O$, define $K_o = \{k \in K : o_k = o\}$ as the set of demands originating from o. We define the aggregated flow over each arc a and origin o as $z_{oa} = \sum_{k \in K_o} y_{ka}$. In the remainder of the paper, we primarily use the aggregated flow variable z, except in a few cases where the disaggregated flow y is necessary. Adapting the arguments to use the disaggregated flow y is straightforward.

3.2 Wardrop Equilibrium Conditions

In this section, we assume $x \in X$ is given and present how user equilibrium is modeled. Let $G(x) = (V, A(x))$ be the subgraph of G only containing available arcs: $A(x) = \{a \in A : x_a = 1\}$.

To describe a set of feasible flows, define for each $k \in K$ and $v \in V$

$$e'_{kv} = \begin{cases} -e_k, & \text{if } v = o_k, \\ e_k, & \text{if } v = d_k, \\ 0, & \text{otherwise,} \end{cases}$$

and the aggregated variant for each $o \in O$ and $v \in V$ by $e_{ov} = \sum_{k \in K_o} e'_{kv}$. Then, a feasible flow must satisfy the flow conservation constraint

$$\sum_{a \in A_v^+(x)} z_{oa} - \sum_{a \in A_v^-(x)} z_{oa} = e_{ov}, \quad \forall o \in O, v \in V,$$

where $A_v^+(x)$ and $A_v^-(x)$ are subsets of $A(x)$ corresponding to incoming and outgoing arcs of v.

We also assume each driver is self-driven and chooses the route so that their travel time is minimized. This is referred to as Wardrop's second principle. It is shown by Beckmann et al.

[1955] that the user equilibrium is given as the optimal solution to the following problem:

$$\phi(x) = \min_{z} \sum_{a \in A(x)} \int_0^{\sum_{o \in O} z_{oa}} \theta_a(s) ds \tag{2a}$$

$$\text{s.t.} \sum_{a \in A_v^+(x)} z_{oa} - \sum_{a \in A_v^-(x)} z_{oa} = e_{ov}, \quad \forall o \in O, v \in V, \tag{2b}$$

$$z \geq 0. \tag{2c}$$

We use $Z(x)$ to denote the optimal solution set of (2). The KKT conditions of (2) are given by

$$\sum_{a \in A_v^+(x)} z_{oa} - \sum_{a \in A_v^-(x)} z_{oa} = e_{ov}, \qquad \forall o \in O, v \in V, \tag{3a}$$

$$\theta_a\left(\sum_{o \in O} z_{oa}\right) - \pi_{oh} + \pi_{ot} \geq 0, \qquad \forall o \in O, a = (t,h) \in A(x), \tag{3b}$$

$$\left(\theta_a\left(\sum_{o \in O} z_{oa}\right) - \pi_{oh} + \pi_{ot}\right) z_{oa} = 0, \quad \forall o \in O, a = (t,h) \in A(x), \tag{3c}$$

$$z \geq 0. \tag{3d}$$

The next proposition states the dual of (2). See Appendix C for the proof.

Proposition 1. *Suppose Assumption 1 holds. Then, the Lagrangian dual of* (2) *is given by*

$$\max_{\pi,\eta} \sum_{k \in K} e_k(\pi_{o_k d_k} - \pi_{o_k o_k}) - \sum_{a \in A(x)} \frac{p}{\overline{g}_a^{1/p}(p+1)} \eta_a^{\frac{p+1}{p}} \tag{4}$$

$$\begin{aligned} s.t.\ & \eta_a \geq \pi_{oh} - \pi_{ot} - f_a, && \forall o \in O, a = (t,h) \in A(x), \\ & \eta_a = 0, && \forall a \in A(x) : g_a = 0, \\ & \eta \geq 0, \end{aligned}$$

where $\overline{g}_a = g_a$ *if* $g_a > 0$ *and otherwise* $\overline{g}_a = 1$.

This dual formulation is exploited in the literature. For example, Fukushima [1984] develops a solution method based on this dual formulation. We denote the set of optimal values of π for formulation (4) by $\Pi(x)$.

Formulations (2) and (4) exhibit a special structure: all constraints are linear, and the objective function is separable and convex. Such problems are known as monotropic programming problems and have been extensively studied by Rockafellar [1981]. Due to this special structure, monotropic programming enjoys favorable duality properties under milder conditions than those required in general convex programming.

Proposition 2. *Suppose Assumption 1 holds.*

a) For any $x \in X$, both the primal problem (2) *and the dual problem* (4) *are feasible;*

b) strong duality holds; and

c) the primal and dual variables are optimal if and only if they satisfy the KKT conditions (3)*.*

See Rockafellar [1981] for the proofs.

3.3 Free-Flow Unfairness

In this section, we briefly review the concept called *free-flow unfairness*. We will use this to define fairness-oriented objective of the DNDP in Section 4.

For each OD pair, at equilibrium, drivers may use multiple paths. However, the travel time of each used path must be the same. If we use the disaggregated flow variable y, it is given by

$$\sum_{a \in A} y_{ka}\ \theta_a \left(\sum_{k \in K} y_{ka} \right) / e_k.$$

Note that the numerator is the total travel time experienced by drivers of OD pair k. This travel time must be larger than the shortest travel time drivers would experience, assuming there is no congestion on the roads. This difference is called the free-flow unfairness by Jahn et al. [2005]. To be more specific, let y be the equilibrium flow. The free-flow unfairness of OD pair $k \in K$ is

$$\zeta_k = \left(\sum_{a \in A} y_{ka}\ \theta_a \left(\sum_{k \in K} y_{ka} \right) / e_k - \nu_k \right) / \nu_k, \tag{5}$$

where ν_k is the shortest travel time on the graph with only the existing arcs, assuming no congestion on roads. Note that in general the right-hand side of (5) is nonconvex in y.

Interestingly, there is a "dual" representation of this quantity. As discussed by Patriksson [2015], for any $x \in X$, $k \in K$ and $\pi \in \Pi(x)$, $\pi_{o_k d_k} - \pi_{o_k o_k}$ is the travel time experienced by each individual of OD pair k to travel from o_k to d_k under the congestion at the user equilibrium. Therefore, the free-flow unfairness can be written as

$$\zeta_k = (\pi_{o_k d_k} - \pi_{o_k o_k} - \nu_k)/\nu_k. \tag{6}$$

In this work, we focus on free-flow unfairness as a running example. However, the argument can be easily adapted to accommodate any metric that is linear in individual travel time. For instance, Meng and Yang [2002] considered improvements in individual travel time, which can be modeled by replacing ν_k in (6) with the travel time at equilibrium in the current network.

4 Methodology

In this section, we present the DNDP and our single-level reformulation. In Section 4.1, we introduce the bilevel formulation of the DNDP under the typical objective of minimizing total travel time, as commonly considered in the literature. Next, in Section 4.2, we present an approach based on the value function and discuss its limitations when applied to objectives such as free-flow fairness. Finally, Section 4.3 introduces our single-level reformulation, which is derived using strong duality, and highlights its flexibility to integrate fairness objectives.

4.1 Bilevel formulation of DNDP

This section presents the bilevel formulation of the DNDP under the objective of minimizing total travel time, a common setting in the literature (see, e.g., Leblanc [1975] and Farvaresh and Sepehri [2013]). Recall that X denotes the set of feasible upper-level decisions x, and $Z(x)$ represents the set of equilibrium flows given x. The DNDP with total travel time as the objective can be formulated as

$$\min_{x,z} \sum_{a\in A}\left(\sum_{o\in O} z_{oa}\right)\ \theta_a\left(\sum_{o\in O} z_{oa}\right) \tag{7a}$$

$$\text{s.t. } x\in X, z\in Z(x). \tag{7b}$$

Objective (7a) minimizes the total travel time, while constraint (7b) ensures that the arc construction decision x is feasible and that z corresponds to an equilibrium flow.

As discussed in Section 3.2, the equilibrium set $Z(x)$ can be characterized as the solution set of an optimization problem. Therefore, formulation (7) is a bilevel program. In particular, it corresponds to the *optimistic* variant of the bilevel model (see Carvalho et al. [2025]). If multiple user equilibria exist, the optimistic formulation assumes that the most favorable one (i.e., the equilibrium with the smallest total travel time) is realized. In contrast, the *pessimistic* formulation assumes that the least favorable one (i.e., the equilibrium with the highest total travel time) will occur. To the best of our knowledge, the distinction between optimistic and pessimistic variants of the DNDP has not been explicitly discussed in the literature, and all existing works adopt the optimistic formulation (7). However, as shown by Roughgarden and Tardos [2000], in the case of multiple user equilibria, all equilibria result in the same total travel time. We provide the proof for completeness.

Proposition 3. *Suppose Assumption 1 holds. For any $x\in X$ and $z, z'\in Z(x)$, we have*

$$\sum_{a\in A}\left(\sum_{o\in O} z_{oa}\right)\ \theta_a\left(\sum_{o\in O} z_{oa}\right)=\sum_{a\in A}\left(\sum_{o\in O} z'_{oa}\right)\ \theta_a\left(\sum_{o\in O} z'_{oa}\right).$$

Proof. Pick any $x\in X$ and $\pi\in\Pi(x)$. For any $o\in O$ and $z\in Z(x)$, by summing (3c) for all

$a \in A(x)$, we obtain:

$$\sum_{a\in A(x)} z_{oa}\theta_a\left(\sum_{o'\in O} z_{o'a}\right) = \sum_{v\in V}\left(\sum_{a\in A_v^+(x)} z_{oa} - \sum_{a\in A_v^-(x)} z_{oa}\right)\pi_{ov} = \sum_{v\in V} e_{ov}\pi_{ov}.$$

Therefore, for any $z \in Z(x)$, the total travel time is

$$\sum_{a\in A}\left(\sum_{o\in O} z_{oa}\right)\ \theta_a\left(\sum_{o\in O} z_{oa}\right) = \sum_{a\in A(x)}\left(\sum_{o\in O} z_{oa}\right)\ \theta_a\left(\sum_{o\in O} z_{oa}\right) = \sum_{o\in O}\sum_{v\in V} e_{ov}\pi_{ov}.$$

This implies the desired result. □

In particular, under Assumption 1, there is no distinction between the optimistic and pessimistic formulations of the bilevel DNDP. As shown in Section 4.3.1, this property extends to any objective function that depends on individual travel times.

4.2 Single Level Reformulation based on Value Function

One of the most popular approaches for problem (7) is reformulating it to a single-level problem using the value function of the lower-level problem. For example, see Meng et al. [2001]. Using $\phi(x)$, the optimal objective value of (2), we obtain a single-level reformulation of (7) as

$$\min_{x,z} \sum_{a\in A}\left(\sum_{o\in O} z_{oa}\right)\theta_a\left(\sum_{o\in O} z_{oa}\right) \tag{8a}$$

$$\text{s.t.} \sum_{a\in A}\int_0^{\sum_{o\in O} z_{oa}} \theta_a(s)ds \le \phi(x), \tag{8b}$$

$$\sum_{a\in A_v^+} z_{oa} - \sum_{a\in A_v^-} z_{oa} = e_{ov}, \qquad \forall o \in O, v \in V, \tag{8c}$$

$$z_{oa} \le \sum_{k\in K_o} e_k x_a, \qquad \forall o \in O, a \in A, \tag{8d}$$

$$z \ge 0, x \in X. \tag{8e}$$

Objective (8a) seeks to minimize total travel time. Constraint (8b) ensures that z is optimal with respect to the lower-level objective, while constraints (8c)–(8e) ensure that z is a feasible flow. Leblanc [1975] showed that the value function $\phi(x)$ is non-increasing in x. As a result, the constraint involving the value function can be linearized as:

$$\min_{x,z} \sum_{a\in A} \left(\sum_{o\in O} z_{oa}\right) \theta_a \left(\sum_{o\in O} z_{oa}\right) \tag{9a}$$

$$\text{s.t.} \sum_{a\in A} \int_0^{\sum_{o\in O} z_{oa}} \theta_a(s)ds \leq \phi(x') + \hat{\phi}(x') \sum_{a\in A} x'_a(1-x_a), \quad \forall x' \in X, \tag{9b}$$

$$(8c) - (8e).$$

where $\hat{\phi}(x') = (\phi(0) - \phi(x'))$ for each $x' \in X$. Rey and Levin [2025] used a closely related cut.

Objective (9a) is convex (see Sheffi [1985]). Moreover, the left-hand side of constraint (9b) is the objective of the lower-level problem, which is convex in z. Therefore, formulation (9) is a convex MINLP, for which various solution methods have been proposed (see Bonami et al. [2012] for a survey). In our implementation, we adopt an outer-approximation method, as described in Appendix D. Note that formulation (9) contains a large number of constraints (9b). Nevertheless, as we demonstrate in our computational experiments in Section 5, these constraints can be handled lazily, allowing the formulation to be solved efficiently. In particular, only a small subset of elements in X (i.e., constraints (9b)) was required to determine the optimal solution.

One of the limitations of formulation (9) is its lack of flexibility. For instance, suppose we aim to minimize the sum of free-flow unfairness. As shown in (5), this can be expressed as

$$\min \sum_{k\in K} \left(\sum_{a\in A} y_{ka}\theta_a \left(\sum_{k'\in K} y_{k'a}\right) / e_k - \nu_k\right) / \nu_k$$
$$\text{s.t. } x \in X, y \in Y(x),$$

where $Y(x)$ denotes the set of disaggregated equilibrium flows for each $x \in X$. It is possible to replace the constraint $y \in Y(x)$ with an equivalent convex representation, as in (9). However, in this case, the objective function is nonconvex. As a result, the corresponding formulation becomes a nonconvex MINLP, which, unlike convex MINLPs, is significantly more challenging to solve to global optimality.

4.3 Single Level Reformulation based on Strong Duality

In this section, we propose our single-level reformulation of problem (7) based on strong duality. The resulting formulation is a convex MINLP and thus solvable by existing optimization techniques. By Proposition 1, $Z(x)$ is given as the projection of $Z'(x)$ and we obtain the following formulation of the DNDP:

$$\min_{x,z,\pi,\eta} \sum_{a\in A} \left(\sum_{o\in O} z_{oa}\right) \theta_a \left(\sum_{o\in O} z_{oa}\right) \tag{10a}$$

$$\text{s.t.} \sum_{a\in A} \int_0^{\sum_{o\in O} z_{oa}} \theta_a(s)ds$$

$$\leq \sum_{k\in K} e_k(\pi_{o_k d_k} - \pi_{o_k o_k}) - \sum_{a\in A} \frac{p}{\bar{g}_a^{1/p}(p+1)} \eta_a^{\frac{p+1}{p}}, \tag{10b}$$

$$\sum_{a\in A_v^+} z_{oa} - \sum_{a\in A_v^-} z_{oa} = e_{ov}, \qquad \forall o \in O, v \in V, \tag{10c}$$

$$z_{oa} \leq \sum_{k\in K_o} e_k x_a, \qquad \forall o \in O, a \in A, \tag{10d}$$

$$\eta_a \geq \pi_{oh} - \pi_{ot} - f_a - M_{oa}(1 - x_a), \qquad \forall o \in O, a = (t,h) \in A, \tag{10e}$$

$$\eta_a = 0, \qquad \forall a \in A : g_a = 0, \tag{10f}$$

$$z \geq 0, \eta \geq 0, x \in X, \tag{10g}$$

where $M \in \mathbb{R}^{O\times A}$ is a constant such that for any $x \in X$ there exists $(\pi, \eta) \in \Pi(x)$ satisfying

$$\pi_{oh} - \pi_{ot} - f_a \leq M_{oa}, \qquad \forall o \in O, a = (t,h) \in A \setminus A(x).$$

We examine approaches to compute valid values of M in Section 4.3.2. Constraint (10b) enforces strong duality, ensuring that the primal and dual variables are optimal for their respective lower-level problems. Constraints (10c), (10d) and (10g) ensure that z is feasible for the lower-level primal problem, while constraints (10e) - (10g) ensure that the feasibility of π and η to the lower-level dual problem. According to Sheffi [1985], the objective and the left-hand side of constraint (10b) are convex in z. Moreover, it is straightforward to verify that the right-hand side of constraint (10b) is concave in π and η. Thus, formulation (10) is a convex MINLP, which can be solved with an outer-approximation method (see Appendix D).

4.3.1 Optimization of Other Objectives

Formulation (10) contains more variables than formulation (9). These extra variables can be used to model different objective functions or constraints. For example, in light of (6), we can modify the DNDP (10) to optimize the total free-flow unfairness as

$$\min_{x,z,\pi,\eta} \sum_{k\in K} (\pi_{o_k d_k} - \pi_{o_k o_k} - \nu_k)/\nu_k \tag{11}$$

$$\text{s.t. (10b)} - \text{(10g)},$$

and the worst free-flow unfairness as

$$\begin{aligned}\min_{x,z,\pi,\eta,\xi}\ & \xi && (12)\\ \text{s.t. } & \xi \geq (\pi_{o_k d_k} - \pi_{o_k o_k} - \nu_k)/\nu_k, \quad \forall k \in K,\\ & (10b) - (10g).\end{aligned}$$

We note that formulations (11) and (12) are convex MINLP.

Theorem 2.5 in Patriksson [2015] states that at equilibrium, individual travel times are unique. As a result, both the total free-flow unfairness and the worst-case free-flow unfairness are the same across all user equilibria. As in Section 4.1, we conclude that there is no distinction between the optimistic and pessimistic formulations of problems (11) and (12).

4.3.2 Computing Big-M

This subsection presents our approach to compute a valid value of big-M. This can be done by bounding the difference of optimal dual values in $\Pi(x)$. The following proposition gives a bound for a specific $x \in X$, $k \in K$ and a pair of nodes (t, h).

Proposition 4. *Let $x \in X$, $o \in O$, $t, h \in V$ and $\delta > 0$. Let $z \geq 0$ be such that for each $o' \in O$ and $v \in V$, we have*

$$\sum_{a \in A_v^+(x)} z_{o'a} - \sum_{a \in A_v^-(x)} z_{o'a} = \begin{cases} e_{o'v} - \delta, & \text{if } o' = o \text{ and } v = t,\\ e_{o'v} + \delta, & \text{if } o' = o \text{ and } v = h,\\ e_{o'v}, & \text{otherwise,}\end{cases}$$

Then, under Assumption 1, for any $\pi \in \Pi(x)$, we have

$$\pi_{oh} - \pi_{ot} \leq \frac{1}{\delta}\left(\sum_{a \in A(x)} \int_0^{\sum_{o' \in O} z_{o'a}} \theta_a(s) ds - \phi(x)\right).$$

Proof. Let L be the Lagrangian of (2) (Appendix C). By strong duality, for any $\pi \in \Pi(x)$,

$$\begin{aligned}\phi(x) &= \min_{z' \geq 0} L(z', \pi)\\ &\leq L(z, \pi)\\ &= \sum_{a \in A(x)} \int_0^{\sum_{o' \in O} z_{o'a}} \theta_a(s) ds - \sum_{o' \in O}\sum_{v \in V} \pi_{o'v}\left(\sum_{a \in A_v^+(x)} z_{o'a} - \sum_{a \in A_v^-(x)} z_{o'a} - e_{o'v}\right)\\ &= \sum_{a \in A(x)} \int_0^{\sum_{o' \in O} z_{o'a}} \theta_a(s) ds - \delta\left(\pi_{oh} - \pi_{ot}\right).\end{aligned}$$

By rearranging the terms, we obtain the desired result. □

Let $\underline{G} = (V, \underline{A})$ be the graph consisting of the existing arcs, i.e., $\underline{A} = \cap_{x \in X} A(x)$. A flow in $\overline{Z}_\delta$ above can be obtained by combining a flow on $\underline{G}$ satisfying the demand and a δ-unit of flow on $\underline{G}$ from t to h, in the sense described in the next proposition.

Proposition 5. *Let $t, h \in V$ and $\delta > 0$. Let $z \geq 0$ be any flow on graph $\underline{G}$ satisfying:*

$$\sum_{a \in \underline{A}_v^+} z_{oa} - \sum_{a \in \underline{A}_v^-} z_{oa} = e_{ov}, \quad \forall o \in O, v \in V.$$

Furthermore, let $z' \geq 0$ be any flow on graph $\underline{G}$ such that for each $v \in V$

$$\sum_{a \in \underline{A}_v^+} z'_a - \sum_{a \in \underline{A}_v^-} z'_a = \begin{cases} -\delta, & \text{if } v = t, \\ \delta, & \text{if } v = h, \\ 0, & \text{otherwise.} \end{cases}$$

Then, under Assumption 1, for any $x \in X$, $o \in O$ and $\pi \in \Pi(x)$, we have

$$\pi_{oh} - \pi_{ot} \leq \frac{1}{\delta} \left(\sum_{a \in \underline{A}} \int_0^{z'_a + \sum_{o' \in O} z_{o'a}} \theta_a(s) ds - \underline{\phi} \right),$$

where $\underline{\phi} = \min_{x \in X} \phi(x)$.

Proof. Fix $x \in X$, $o \in O$ and define for each $o' \in O$, $a \in A$

$$\hat{z}_{o'a} = \begin{cases} z_{o'a}, & \text{if } o' \neq o, \\ z'_a + z_{o'a}, & \text{if } o' = o. \end{cases}$$

Flow $\hat{z}$ satisfies the condition in Proposition 4 and is feasible on $G(x)$. Thus, for any $\pi \in \Pi(x)$, we obtain

$$\begin{aligned} \pi_{oh} - \pi_{ot} &\leq \frac{1}{\delta} \left(\sum_{a \in A(x)} \int_0^{\sum_{o' \in O} \hat{z}_{o'a}} \theta_a(s) ds - \phi(x) \right) \\ &\leq \frac{1}{\delta} \left(\sum_{a \in \underline{A}} \int_0^{z'_a + \sum_{o' \in O} z'_{o'a}} \theta_a(s) ds - \underline{\phi} \right), \end{aligned}$$

as required. □

This proposition is useful since one can obtain a bound valid for multiple graphs. In our implementation, we compute the values of big-M as follows. First, we solve the traffic assignment (2) on G (i.e., the graph with all the existing and candidate arcs) to obtain a lower bound on $\underline{\phi}$: In light of the monotonicity of ϕ, $\underline{\phi} \leq \phi(x)$ for any $x \in X$. Then, we solve the traffic assignment (2) on $\underline{G}$ to obtain a flow z as well as travel times of the arcs under the congestion. Lastly, for each pair $t, h \in V$, we obtain the shortest path on $\underline{G}$ under the congestion induced by z. We consider δ unit of flow along the shortest path and use it as z'. Then, we invoke Proposition 5 with these z and z'. We choose the value of δ by grid search ($\delta \in \{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$).

5 Computational Performance

In this section, we study the performance of our single-level reformulation, formulation SD (10). We use formulation VF (9) as a baseline. We use two instances generated by Rey and Levin [2025]: SiouxFalls and Easter Massachusetts. Their statistics are shown in Table 1. For each transportation network, there are 10 test instances with different sets of candidate arcs. Let A_1 and A_2 be the sets of existing and candidate arcs, respectively. Then, set X is defined by

$$X = \left\{ x : \sum_{a \in A_2} c_a x_a \leq b,\ x_{a'} = 1,\ \forall a' \in A_1,\ x_{a'} \in \{0,1\},\ \forall a' \in A_2 \right\}, \tag{13}$$

where c_a is the construction cost of arc $a \in A_2$ and b is the budget. For more details on the instance generation, see Rey and Levin [2025]. All the methods are implemented in Python 3.9 using Gurobi 12.0, and run in single-thread mode. To compute the user equilibria, we use the column generation technique as described by Leventhal et al. [1973].

Table 1: Statistics of the test instances

Network	Nodes	Arcs	Candidate Arcs	Budget (%)	OD pairs
SiouxFalls	24	76	10	50	528
Easter Massachusetts	74	258	10	50	1113

5.1 DNDP to Optimize Fairness

We first evaluate the capability of formulation SD to optimize various metrics, as discussed in Section 4.3.1. To be more specific, we solve the DNDP to optimize the total travel time (10a), the total free-flow unfairness (11) and the DNDP to optimize the worst free-flow unfairness (12). Here, we do not consider the formulation based on the value function, since it cannot handle the objectives based on fairness. The comparison of the two formulations when the objective is the total travel time is given in Section 5.2.

Table 2 shows the number of test instances solved within the one-hour time limit (Solved), the average computational time in seconds (Time) and the average number of branch-and-bound nodes (# B&B Nodes). Compared to the optimization of the total travel time (10), the optimization of the total free-flow unfairness (11) tends to require more time on SiouxFalls, and more branch-and-bound nodes on both transportation networks. The optimization of the worst free-flow unfairness (12) requires even more computational time and branch-and-bound nodes than the optimization of the total free-flow unfairness (11) on average. There is one test instance for which the optimal solution was not found within one hour when the objective is the total free-flow unfairness, and four instances when the objective is the worst free-flow unfairness.

Table 2: Performance of formulation SD to optimize the total travel time (10), the total free-flow unfairness (11) and the worst free-flow unfairness (12)

Network	Objective	Solved	Time	# B&B Nodes
SiouxFalls	Travel time	10	248.2	294.5
	Total free-flow unfairness	10	337.4	640.5
	Worst free-flow unfairness	10	569.9	705.3
Easter Massachusetts	Travel time	10	1,646.1	166.5
	Total free-flow unfairness	9	1,987.5	472.4
	Worst free-flow unfairness	6	2,812.5	537.0

We speculate that the reason why optimizing unfairness seems harder than optimizing travel time is as follows: intuitively, we observe that the total travel time, the objective value of formulation (10), is similar to the objective value of the lower-level formulation (2). In fact, when there is no congestion ($g_a = 0$ for all $a \in A$), the two objective values coincide for any feasible flow. In this setup, constraints (10b), (10e) and (10f), along with dual variables π and η, are redundant in the sense that the optimal solution of formulation (10) does not change even if they are removed. Thus, these constraints are only necessary when congestion affects the system, and the "effective" formulation size might be smaller. These observations do not hold for the other objective functions used in formulations (11) and (12), as the upper- and lower-level objectives are no longer closely aligned, thereby accentuating the problem's bilevel nature and, consequently, its difficulty.

5.2 DNDP to Minimize the Total Travel Time

As we have seen in Section 5.1, the formulation based on strong duality is flexible, and it becomes particularly valuable when the objective is based on fairness. However, when the objective is the total travel time, the typical objective in the literature, the formulation based on the value function can be used to solve the DNDP too. Below, we observe the computational performance of formulation SD (10) and formulation VF (9) in this setting. In formulation VF (9), constraints (9b) are generated lazily: every time an integer solution is found, we compute the corresponding traffic assignment and check whether constraint (9b) is satisfied, adding it if necessary.

Table 3 shows the number of test instances solved within the one-hour time limit (Solved), the average computational time in seconds (Time), and the average number of branch-and-bound nodes (# B&B Nodes). For the test instances that are not solved within the time limit, the time is set to one hour, and the number of branch-and-bound nodes computed within this one-hour time limit are used. In the column labeled "Formulation", SD (standing for strong duality) and VF (value function) refer to formulations SD (10) and VF (9), respectively. In this experiment, all the test instances are solved within the one-hour time limit. Formulation SD (10) has a shorter average solution time on SiouxFalls, but on Easter Massachusetts, formulation VF (9) tends to be faster on average. Formulation VF (9) requires less branch-and-bound nodes than formulation SD (10), especially on Easter Massachusetts. Interestingly, both formulations require on average fewer branch-and-bound nodes on Easter Massachusetts, a larger transportation network.

Table 3: Performance of formulations SD (10) and VF (9) to solve the DNDP to minimize the total travel time

Network	Formulation	Solved	Time	# B&B Nodes
SiouxFalls	SD	10	248.2	294.5
	VF	10	614.2	252.6
Easter Massachusetts	SD	10	1,646.1	166.5
	VF	10	864.8	111.3

Especially on large test instances, formulation VF tends to be faster. For further analysis on the scalability of the two formulations, see Appendix E. Nonetheless, as we saw in Section 5.1, formulation SD (10) preserves convexity even when the objective involves fairness-related metrics, such as the free-flow unfairness. In contrast, modifying formulation VF (9) to optimize a fairness objective results in nonconvex MINLPs, for which proving global optimality is significantly more challenging. In this regard, the ability of the SD formulation to accommodate fairness objectives within a convex optimization framework constitutes a valuable contribution.

6 Case Study: EV Charging Station Capacity Expansion

In this section, we formulate the capacity expansion problem for EV charging stations as the DNDP. This formulation enables us to optimize charging station capacity expansion decisions while considering congestion at charging stations. The objective is to minimize either the total travel time or a specific fairness metric related to EV drivers.

In Section 6.1, we describe the setup of the problem. In this section, we also present the approach for modeling the behavior (i.e., the routing decision) of EV drivers as a traffic assignment problem, as well as a metric quantifying the inefficiency experienced by EV drivers. Section 6.2 presents the results of the numerical experiments using the Quebec road network.

6.1 Problem Setup

Suppose we have a highway road network with charging stations installed at some of the nodes. We are interested in EV drivers who travel long distances through this road network, represented as OD pairs. Specifically, we focus on the drivers who need to charge their EVs along their trips in order to reach their destinations. These EV drivers consider the time required to charge their vehicles and choose their routes in a way that minimizes their total travel time. Given the current road network with charging stations and their associated demands, our goal is to select a subset of charging stations for capacity expansion to improve a specified metric, such as the total travel time of EV drivers. We make the following assumptions:

Assumption 2.

a) All EVs have the same vehicle range.

b) Each OD pair is traversable, assuming the battery is fully charged at the origin and that intermediate charging at existing stations along the route is allowed.

c) The travel time on a road does not depend on the number of EV drivers using that road.

d) The time required to charge an EV at a charging station (including waiting time) can be modeled as in equation (1)*.*

The first assumption is standard in the literature (e.g., Kınay et al. [2023]). The assumption that travel time is independent of the number of EV drivers can be justified when the ratio of EVs to conventional vehicles is small. For instance, in 2023, EVs accounted for 4.4% of the total light vehicles in circulation in Quebec, highlighting that they are still relatively rare on the roads [Institut de la statistique du Québec, 2023]. When the share of EV drivers is not negligible, we can use the multi-class network equilibrium as developed by Chen et al. [2016]. Regarding the final assumption, there are situations in which alternative waiting-time functions, such as those obtained from queueing models, may be more appropriate. Under mild conditions, our strong duality approach extends to these more general functions. For clarity of exposition, however, we adopt the assumption stated above.

Traffic assignment with electric vehicles (EVs) is more complex than conventional traffic assignment due to the presence of charging requirements. In particular, a shortest path for a conventional vehicle may be infeasible for an EV because of its limited driving range, or suboptimal due to congestion and waiting times at charging stations. However, by employing the network transformation technique proposed by MirHassani and Ebrazi [2013], the EV traffic assignment problem can be converted into a standard user-equilibrium formulation, as described in Section 3, and the associated capacity expansion problem can be reduced to the conventional NDP. A detailed description of this transformation is provided in Appendix F.

To quantify the inefficiency introduced by the adoption of EVs, we define a fairness-based metric, termed the cost of sustainability, as the difference between the travel times experienced by EV drivers and conventional vehicle drivers. EV drivers may incur additional delays due to detours to charging stations and waiting times at these facilities, both of which depend on congestion levels at charging stations. The optimization of the total and worst-case cost of sustainability can be carried out by replacing ν_k with the travel time experienced by conventional vehicle drivers in (11) and (12).

6.2 Numerical Experiments with Quebec Road Network

Experimental Setup. To illustrate the applicability of our method, we run experiments on the road network in Quebec, including public EV charging stations. We use the 2021 Statistics Canada census data[1] to determine population centers. We consider the 20 population centers with the largest populations. We assume the vehicle range of an EV is 200 km and remove OD pairs whose distances are less than this vehicle range, resulting in 92 OD pairs. The demand volumes of the OD pairs are computed using the gravity model (Hodgson [1990]). We consider two demand scenarios: a low-demand scenario and a high-demand scenario, where the demand volume of each OD pair is increased by 50% from that of the low-demand scenario. The histogram of the OD demand volumes is presented in Appendix G.

The location data of charging stations is obtained from *Circuit électric*[2]. We use OpenStreetMap (OpenStreetMap contributors [2024]) with OSMnx (Boeing [2024]) to compute the distances between charging stations and population centers using the highway network in Quebec. Based on the computed distances, we construct the transformed network as discussed in Section 6.1. We consider 370 level-3 charging stations. The resulting network has a large number of arcs. To make the instance size moderate, we run the traffic assignment and compute the user equilibrium using the existing charging stations. We then remove the arcs with zero flows at equilibrium since they are unlikely to be used after capacity expansion. The statistics of the resulting graphs are shown in Table 4.

Table 4: Statistics of the graph before and after filtering arcs

Network	Nodes	Arcs	Candidate arcs	OD pairs
Original	924	65,888	370	92
Low Demand	924	354	87	92
Hight Demand	924	421	108	92

The time to charge an EV at a charging station (including waiting time) is modeled as $1/2 + (s/c)^4/100$, where s is the amount of demand served at this charging station and c is the number of outlets of this charging station.

We consider the capacity expansion of each charging station by 1 unit (e.g., if we have a charging station of capacity $c = 2$, we may expand it to capacity $c = 3$). For each charging

[1] `https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/`
[2] `https://lecircuitelectrique.com/en/`

station (candidate arc a), we sample the cost of the capacity expansion (the construction cost c_a of the candidate arc) from a uniform distribution between 0.5 and 1.5. We use X as defined in equation (13).

We propose a simple greedy heuristic as a baseline. The method works as follows: We first evaluate the user equilibrium with the current set of charging stations in place. Next, we compute the "utility" of each charging station. This is defined as the ratio of the total demand served by the station to its capacity. We then sort the charging stations in decreasing order of their utility, from the most to the least used. Starting from the highest utility, we choose the stations until we reach a budget limit. After selecting the stations, we perform another traffic assignment to evaluate the objective value (e.g., total travel time). Thus, this greedy method requires two traffic assignment computations: one for evaluating the user equilibrium and another for evaluating the objective after selecting the charging stations.

Results and Discussion. First, we use formulations (11) and (12) to optimize the total and worst cost of sustainability, as described in Section 6.1. Recall that the cost of sustainability is defined as the extra travel time EV drivers experience compared to the drivers of conventional cars. Tables 5 and 6 show the upper bound[3] (UB), the improvement of the upper bound relative to the greedy method (Rel.), the lower bound (LB), the gap between upper and lower bound (Gap), and the number of branch-and-bound nodes (# B&B Nodes). We set the time limit to one hour. We observe that increasing demand makes the problem harder: the optimality gap tends to get worse and fewer branch-and-bound nodes are explored. The budget seems to have little impact on the performance of the method.

Next, we present the performances of formulations (10) and (9) to seek the capacity expansion decisions that minimize the total travel time of EV drivers. Table 7 shows the upper and lower bounds obtained after the one-hour time limit, among other statistics. The definition of each column is the same as in Table 5. In all the setups, formulations SD (10) and VF (9) provide tighter upper bounds (better solutions) than the greedy method. Formulation VF (9) tends to give tighter bounds than formulation SD (10), and the optimality gap is smaller in all the setups. For each formulation, when we use the high-demand scenario, the optimality gap tends to worsen, and the number of branch-and-bound nodes decreases. There are no clear trends in terms of methodologies' performance regarding budget variations. As long as the objective is the total travel time, formulation VF (9) seems to be more competitive than formulation SD (10) and the greedy method. Additional results can be found in Appendix G.

It is interesting to note that the worst cost of sustainability is harder to optimize than the total cost of sustainability, which is harder than the total travel time. On any setup, we explore the fewest branch-and-bound nodes and tend to have the worst optimality gap when optimizing the worst cost of sustainability. This behavior is consistent with the results obtained on the

[3]The best incumbent value.

Table 5: Performance of formulation SD (11) to solve the DNDP to minimize the total cost of sustainability

Budget	Demand	Formulation	UB	(Rel.)	LB	Gap	# B&B Nodes
5	Low	Greedy	173.37		-	-	-
		SD	169.77	(-2.08 %)	142.90	18.80 %	3074
	High	Greedy	462.55		-	-	-
		SD	453.59	(-1.94 %)	339.59	33.57 %	1755
10	Low	Greedy	163.79		-	-	-
		SD	157.51	(-3.83 %)	130.24	20.94 %	2720
	High	Greedy	426.30		-	-	-
		SD	404.35	(-5.15 %)	298.91	35.27 %	1618
20	Low	Greedy	145.36		-	-	-
		SD	144.08	(-0.88 %)	119.33	20.75 %	2897
	High	Greedy	370.35		-	-	-
		SD	360.21	(-2.74 %)	261.76	37.61 %	1881

Table 6: Performance of formulation SD (12) to solve the DNDP to minimize the worst cost of sustainability

Budget	Demand	Formulation	UB	(Rel.)	LB	Gap	# B&B Nodes
5	Low	Greedy	0.56		-	-	-
		SD	0.55	(-1.85 %)	0.44	26.44 %	840
	High	Greedy	1.02		-	-	-
		SD	1.02	(-0.36 %)	0.73	38.95 %	729
10	Low	Greedy	0.56		-	-	-
		SD	0.51	(-8.95 %)	0.40	28.14 %	889
	High	Greedy	0.98		-	-	-
		SD	0.91	(-6.93 %)	0.64	43.54 %	668
20	Low	Greedy	0.50		-	-	-
		SD	0.50	(0.00 %)	0.36	38.48 %	833
	High	Greedy	0.90		-	-	-
		SD	0.86	(-4.07 %)	0.53	62.79 %	736

Table 7: Performance of formulations SD (10) and VF (9) to solve the DNDP to minimize the total travel time

Budget	Demand	Formulation	UB	(Rel.)	LB	Gap	# B&B Nodes
5	Low	Greedy	1,424.72		-	-	-
		SD	1,414.09	(-0.75 %)	1,376.14	2.76 %	1713
		VF	1,406.94	(-1.25 %)	1,376.25	2.23 %	3724
	High	Greedy	2,657.46		-	-	-
		SD	2,605.22	(-1.97 %)	2,528.32	3.04 %	987
		VF	2,581.72	(-2.85 %)	2,529.71	2.06 %	2776
10	Low	Greedy	1,399.55		-	-	-
		SD	1,385.52	(-1.00 %)	1,350.34	2.60 %	1576
		VF	1,379.23	(-1.45 %)	1,350.28	2.14 %	4300
	High	Greedy	2,563.26		-	-	-
		SD	2,515.46	(-1.86 %)	2,431.67	3.45 %	1053
		VF	2,486.97	(-2.98 %)	2,432.16	2.25 %	3050
20	Low	Greedy	1,351.52		-	-	-
		SD	1,348.91	(-0.19 %)	1,319.55	2.22 %	1771
		VF	1,344.84	(-0.49 %)	1,319.57	1.91 %	3743
	High	Greedy	2,416.45		-	-	-
		SD	2,398.83	(-0.73 %)	2,313.38	3.69 %	1198
		VF	2,371.61	(-1.86 %)	2,313.93	2.49 %	2756

academic instances, which showed that incorporating fairness objectives strengthens the bilevel structure of the problem and increases computational difficulty.

Final remarks and insights. We emphasize the importance of formulations SD (11) and (12). The total travel time experienced by all EV drivers is one metric to measure the quality of service provided by the charging station operator. However, from each user's perspective, arguably, they are more interested in the time they personally experience, rather than the total travel time experienced by all EV drivers. In this regard, our formulations (11) and (12) optimize metrics relevant to each driver's experience. Furthermore, while we focused on the cost of sustainability in this experiment (i.e., the difference between an EV driver and a conventional car driver with the same origin and destination), it is straightforward to modify the formulation to consider the absolute (rather than relative to conventional car drivers) time or improvement of travel time experienced by EV drivers (see Section 3.3).

The greedy heuristic is computationally inexpensive and yields reasonable solutions, but it does not provide any guarantee of optimality. The proposed formulations, while requiring more computational effort, allow for rigorous assessment of solution quality through optimality gaps. Although exact optimality is harder to achieve for fairness objectives, our SD formulations still provide better solutions and help quantify trade-offs between efficiency and equity in network design. Indeed, as shown in the supplementary analyses in Appendix G, especially under small budgets, the topology of the solutions obtained when optimizing total travel time, total cost of sustainability, and worst cost of sustainability, can be fundamentally different from the solution for one of the objectives distanciating from the value obtained by others. This provides further insight into how decision-makers can use the proposed methods to develop prioritization strategies. For instance, Table 6 shows that improvements in the worst cost of sustainability tend to saturate after adding a small number of charging stations. Therefore, allocating a modest portion of the budget (approximately 10 stations) to improve this equity-focused metric may be sufficient. The remaining budget can then be directed toward minimizing total travel time, allowing decision-makers to balance equity and efficiency more effectively.

7 Conclusions and Future work

In this paper, we studied a single-level reformulation of the DNDP. Our approach was based on strong duality of the traffic assignment problem that characterizes user equilibrium. We also presented an alternative single-level reformulation using the value function, a widely used technique in the DNDP literature. Assuming the upper-level objective is the total travel time, both approaches lead to convex MINLPs, for which various solution methods exist.

To illustrate the utility of our reformulation, we applied it to fairness-driven objectives, specifically total and worst-case free-flow unfairness. We showed that the value function approach becomes difficult to adapt in these cases, as it results in nonconvex MINLPs. In contrast, our reformulation based on strong duality extends naturally to these objectives while preserving convexity. Through numerical experiments on benchmark transportation networks, we evaluated the performance of our approach. On smaller instances, our method outperformed the value function formulation in terms of runtime, whereas the latter demonstrated better scalability on larger instances. However, our method could also handle a broader class of fairness metrics, demonstrating its flexibility.

We further demonstrated the applicability of our reformulation to a real-world case study involving capacity expansion for EV charging stations. Modeling the problem as a DNDP, we compared our solution with that of a greedy heuristic and observed improvements in the objectives, including total travel time and the cost of sustainability, i.e., the inefficiency caused by EV adoption.

The reformulation based on strong duality assumes that the travel time function follows the BRP function. It also assumes that the OD pairs are connected, even in the absence of candidate arcs. It would be valuable to extend our method to handle more general instances with relaxed assumptions. Furthermore, in our case study on charging stations, we focused primarily on capacity expansion decisions. The siting of new charging stations can be modeled either with an SOS1 constraint or by extending the logic to compute a valid value of big-M, which is left as future research. Additionally, developing a faster, decomposition-based solution method to solve our reformulation more efficiently is of interest.

References

Robert Aboolian, Oded Berman, and Majid Karimi. Probabilistic set covering location problem in congested networks. *Transportation Science*, 56(2):528–542, 2022.

Okan Arslan, Oya Ekin Karaşan, A. Ridha Mahjoub, and Hande Yaman. A branch-and-cut algorithm for the alternative fuel refueling station location problem with routing. *Transportation Science*, 53(4):1107–1125, 2019. ISSN 0041-1655. doi: 10.1287/trsc.2018.0869.

Saeed Asadi Bagloee, Majid Sarvi, and Michael Patriksson. A hybrid branch-and-bound and benders decomposition algorithm for the network design problem. *Computer-Aided Civil and Infrastructure Engineering*, 32(4):319–343, 2017. doi: 10.1111/mice.12224.

Zhaoyao Bao and Chi Xie. Optimal station locations for en-route charging of electric vehicles in congested intercity networks: A new problem formulation and exact and approximate partitioning algorithms. *Transportation Research Part C: Emerging Technologies*, 133:103447, 2021.

Martin J. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation.* RAND Corporation, Santa Monica, CA, 1955.

Geoff Boeing. Modeling and analyzing urban networks and amenities with osmnx. Working paper, University of Southern California, Munich, 2024.

Pierre Bonami, Mustafa Kilinç, Jeff Linderoth, Sven Leyffer, and Jon Lee. Algorithms and software for convex mixed integer nonlinear programs. In *Mixed Integer Nonlinear Programming*, The IMA Volumes in Mathematics and its Applications, pages 1–39. Springer New York, New York, NY, 2012. ISBN 1461419263. doi: 10.1007/978-1-4614-1927-3_1.

Margarida Carvalho, Gabriele Dragotto, Andrea Lodi, Sriram Sankaranarayanan, et al. Integer programming games. *Foundations and Trends® in Optimization*, 7(4):264–391, 2025.

Zhibin Chen, Fang He, Lihui Zhang, and Yafeng Yin. Optimal deployment of autonomous vehicle lanes with endogenous market penetration. *Transportation research part C: emerging technologies*, 72:143–156, 2016.

Mervat Chouman, Teodor Gabriel Crainic, and Bernard Gendron. Commodity representations and cut-set-based inequalities for multicommodity capacitated fixed-charge network design. *Transportation Science*, 51(2):650–667, 2017. ISSN 0041-1655. doi: 10.1287/trsc.2015.0665.

Stephan Dempe. *Foundations of bilevel programming*. Springer, 2002.

Amin Ahmadi Digehsara, Menglei Ji, Amir Ardestani-Jaafari, and Hoda Bidkhori. Equity-driven facility location: A two-stage robust optimization approach. *Computers & Operations Research*, 176:106920, 2025.

European Environment Agency. Transport and mobility, 2024. URL `https://www.eea.europa.eu/en/topics/in-depth/transport-and-mobility`. Accessed: 2025-08-19.

Hamid Farvaresh and Mohammad Mehdi Sepehri. A single-level mixed integer linear formulation for a bi-level discrete network design problem. *Transportation Research Part E: Logistics and Transportation Review*, 47(5):623–640, 2011. doi: 10.1016/j.tre.2011.02.001.

Hamid Farvaresh and Mohammad Mehdi Sepehri. A branch and bound algorithm for bi-level discrete network design problem. *Networks and Spatial Economics*, 13(1):67–106, 2013. ISSN 1566-113X. doi: 10.1007/s11067-012-9173-3.

Pirmin Fontaine and Stefan Minner. Benders decomposition for discrete–continuous linear bilevel problems with application to traffic network design. *Transportation Research Part B: Methodological*, 70:163–172, 2014a. doi: 10.1016/j.trb.2014.09.007.

Pirmin Fontaine and Stefan Minner. Benders decomposition for discrete–continuous linear bilevel problems with application to traffic network design. *Transportation Research Part B: Methodological*, 70:163–172, 2014b.

Terry L Friesz, G Anandalingam, Nihal J Mehta, Keesung Nam, Samir J Shah, and Roger L Tobin. The multiobjective equilibrium network design problem revisited: A simulated annealing approach. *European Journal of Operational Research*, 65(1):44–57, 1993.

Masao Fukushima. On the dual approach to the traffic assignment problem. *Transportation Research Part B: Methodological*, 18(3):235–245, 1984. ISSN 0191-2615.

Ziyou Gao, Jianjun Wu, and Huijun Sun. Solution algorithm for the bi-level discrete network design problem. *Transportation Research Part B: Methodological*, 39(6):479–495, 2005.

Paul Göpfert and Stefan Bock. A branch&cut approach to recharging and refueling infrastructure planning. *European Journal of Operational Research*, 279(3):808–823, 2019. ISSN 0377-2217.

Jia He, Hai Yang, Tie-Qiao Tang, and Hai-Jun Huang. An optimal charging station location model with the consideration of electric vehicle's driving range. *Transportation Research Part C: Emerging Technologies*, 86:641–654, 2018.

M. John Hodgson. A flow-capturing location-allocation model. *Geographical Analysis*, 22(3): 270–279, 1990. doi: 10.1111/j.1538-4632.1990.tb00210.x.

Institut de la statistique du Québec. Véhicules électriques en circulation, 2023. URL `https://statistique.quebec.ca/vitrine/developpement-durable/strategie-2023-2028/economie-verte-et-responsable/vehicules-electriques-circulation?onglet=faits-saillants-et-graphiques`. Accessed: 2025-08-20.

Olaf Jahn, Rolf H Mohring, Andreas S Schulz, and Nicolas E Stier-Moses. System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations Research*, 53(4):600–616, 2005. ISSN 0030-364X. doi: 10.1287/opre.1040.0197.

Wentao Jing, Kun An, Mohsen Ramezani, and Inhi Kim. Location design of electric vehicle charging facilities: a path-distance constrained stochastic user equilibrium approach. *Journal of Advanced Transportation*, 2017(1):4252946, 2017. doi: 10.1155/2017/4252946.

Jong-Geun Kim and Michael Kuby. The deviation-flow refueling location model for optimizing a network of refueling stations. *International Journal of Hydrogen Energy*, 37(6):5406–5420, 2012. ISSN 0360-3199. doi: 10.1016/j.ijhydene.2011.08.108.

Ömer Burak Kınay, Fatma Gzara, and Sibel A Alumur. Charging station location and sizing for electric vehicles under congestion. *Transportation Science*, 57(6):1433–1451, 2023.

Michael Kuby and Seow Lim. The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences*, 39(2):125–145, 2005. ISSN 0038-0121.

Larry J Leblanc. An algorithm for the discrete network design problem. *Transportation Science*, 9(3):183–199, 1975. ISSN 0041-1655. doi: 10.1287/trsc.9.3.183.

T Leventhal, George Nemhauser, and Jr Trotter, Leslie Earl. A column generation algorithm for optimal traffic assignment. *Transportation Science*, 7(2):168–176, 1973. ISSN 0041-1655.

Sheng Liu and Nooshin Salari. Fairness in accessibility of public service facilities. *Operations Research Letters*, 55:107141, 2024.

Ivana Ljubić, Miguel A Pozo, Justo Puerto, and Alberto Torrejón. Benders decomposition for the discrete ordered median problem. *European Journal of Operational Research*, 317(3): 858–874, 2024.

Paramet Luathep, Agachai Sumalee, William HK Lam, Zhi-Chun Li, and Hong K Lo. Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach. *Transportation Research Part B: Methodological*, 45(5):808–827, 2011.

Thomas L Magnanti and Richard T Wong. Network design and transportation planning: Models and algorithms. *Transportation Science*, 18(1):1–55, 1984. doi: 10.1287/trsc.18.1.1.

Qiang Meng and Hai Yang. Benefit distribution and equity in road network design. *Transportation Research Part B: Methodological*, 36(1):19–35, 2002.

Qiang Meng, Hai Yang, and Michael GH Bell. An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. *Transportation Research Part B: Methodological*, 35(1):83–105, 2001.

Seyyed A MirHassani and Roozbeh Ebrazi. A flexible reformulation of the refueling station location problem. *Transportation Science*, 47(4):617–628, 2013. ISSN 0041-1655. doi: 10.1287/trsc.1120.0430.

Amir Mirheli and Leila Hajibabai. Charging network design and service pricing for electric vehicles with user-equilibrium decisions. *IEEE Transactions on Intelligent Transportation Systems*, 24(3):2888–2902, 2023.

Natural Resources Canada. Zero-emission vehicle infrastructure program (zevip), 2024. URL `https://www.nrcan.gc.ca/energy-efficiency/transportation-alternative-fuels/zero-emission-vehicle-infrastructure-program/21876`. Accessed: 2025-08-19.

OpenStreetMap contributors. OpenStreetMap. `https://www.openstreetmap.org`, 2024. Accessed: 2024-10-31.

Michael Patriksson. *The Traffic Assignment Problem: Models and Methods*. Dover Publications, Headquarters location, 2015. ISBN 978-0486787909.

David Rey. Computational benchmarking of exact methods for the bilevel discrete network design problem. *Transportation Research Procedia*, 47:11–18, 2020.

David Rey and Michael Levin. A branch-and-price-and-cut algorithm for discrete network design problems under traffic equilibrium. *Optimization Online*, 2025. URL `https://optimization-online.org/?p=28420`.

R. Tyrrell Rockafellar. Monotropic programming: descent algorithms and duality. In *Nonlinear programming 4*, pages 327–366. Elsevier, 1981.

Tim Roughgarden and Éva Tardos. How bad is selfish routing? In *41st Annual Symposium on Foundations of Computer Science*, pages 93–102, 2000. doi: 10.1109/SFCS.2000.892069.

Yosef Sheffi. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Hoboken, New Jersey, 1985.

Statistics Canada. Canadians' commutes: Still car-heavy, but some lighter footprints, 2023. URL `https://www.statcan.gc.ca/o1/en/plus/3798-canadians-commutes-still-car-heavy-some-lighter-footprints`. Accessed: 2025-08-19.

Cong Quoc Tran, Dong Ngoduy, Mehdi Keyvan-Ekbatani, and David Watling. A user equilibrium-based fast-charging location model considering heterogeneous vehicles in urban networks. *Transportmetrica A: Transport Science*, 17(4):439–461, 2021.

U.S. Environmental Protection Agency. Fast facts on transportation greenhouse gas emissions, 2024. URL `https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions`. Accessed: 2025-08-19.

David ZW Wang and Hong K Lo. Global optimum of the linearized network design problem with equilibrium flows. *Transportation Research Part B: Methodological*, 44(4):482–492, 2010.

David ZW Wang, Haoxiang Liu, and Wai Y Szeto. A novel discrete network design problem formulation and its global optimization solution algorithm. *Transportation Research Part E: Logistics and Transportation Review*, 79:213–230, 2015.

Shuaian Wang, Qiang Meng, and Hai Yang. Global optimization methods for the discrete network design problem. *Transportation Research Part B: Methodological*, 50:42–60, 2013.

Bo Zhang, Meng Zhao, and Xiangpei Hu. Location planning of electric vehicle charging station with users' preferences and waiting time: multi-objective bi-level programming model and HNSGA-II algorithm. *International Journal of Production Research*, 61(5):1394–1423, 2023.

Hong Zheng, Xiaozheng He, Yongfu Li, and Srinivas Peeta. Traffic equilibrium and charging facility locations for electric vehicles. *Networks and Spatial Economics*, 17(2):435–457, 2017.

<u>Online supplement</u>

Fair network design problem: an application to EV charging station capacity expansion

A Notation

Table 8 lists the symbols used in this paper.

Table 8: Symbols and their description

Symbol	Description
V	Set of nodes indexed by v
A	Set of arcs indexed by a
$A(x)$	Set of available arcs
$A_v^+(x)$	Set of available incoming arcs of v
$A_v^-(x)$	Set of available outgoing arcs of v
G	Graph (V, A)
K	Set of OD pairs indexed by k
X	Set of feasible network design decisions x
O	Set of nodes that are origins
K_o	Set of OD pairs whose origins are o
$Y(x)$	Set of user equilibrium flow given x
$\Pi(x)$	Optimal solution set of the dual lower-level problem given x
$Y'(x)$	Set of pairs of user equilibrium flow and dual values given x, $Y(x) \times \Pi(x)$
o_k	Origin of k
d_k	Destination of k
e_k	Demand volume of k
ν_k	Shortest travel time of k on the graph with only the existing arcs, without congestion on roads
ϕ	Value function of the lower-level problem
θ_a	Travel time function of a, defined as $\theta_a(s) = f_a + g_a s^p$
x_a	Binary variable that is 1 if and only if we build a
y_{ka}	Disaggregated flow on a corresponding to k
z_{oa}	Aggregated flow on a corresponding to demands with o
π, η	Variables in the dual lower-level problem
ζ_k	Free-flow unfairness of k

B Comparison of Aggregated and Disaggregated Models

In this section, we compare the performance of the disaggregated model (formulation with disaggregated flow variable y) and the aggregated model (10) (formulation with aggregated flow variable z).

The arguments in Sections 3 and 4 can be extended to use disaggregated flow variable y. For example, we obtain a formulation of the NDP corresponding to formulation (10) but with disaggregated flow variable y as follows:

$$
\begin{aligned}
\min_{x,y,\pi,\eta} \ & \sum_{a\in A}\left(\sum_{k\in K} y_{ka}\right)\theta_a\left(\sum_{k\in K} y_{ka}\right) && (14)\\
\text{s.t. } & \sum_{a\in A}\int_0^{\sum_{k\in K} y_{ka}} \theta_a(s)ds \\
& \quad \le \sum_{k\in K} e_k(\pi_{kd_k}-\pi_{ko_k}) - \sum_{a\in A}\frac{p}{g_a^{1/p}(p+1)}\eta_a^{\frac{p+1}{p}}, \\
& \sum_{a\in A_v^+} y_{ka} - \sum_{a\in A_v^-} y_{ka} = e'_{kv}, && \forall k\in K, v\in V,\\
& y_{ka} \le e_k x_a, && \forall k\in K, a\in A,\\
& \eta_a \ge \pi_{kh_a} - \pi_{kt_a} - f_a - M'_{ka}(1-x_a), && \forall k\in K, a=(t,h)\in A,\\
& \eta_a = 0, && a\in A: g_a = 0,\\
& y\ge 0, \eta\ge 0, x\in X,
\end{aligned}
$$

where M' is a sufficiently large constant, which can be computed using a routine similar to Proposition 5. Analogously, we can modify formulation (9) to use disaggregated flow variable y as

$$
\begin{aligned}
\min_{x,y} \ & \sum_{a\in A}\left(\sum_{k\in K} y_{ka}\right)\theta_a\left(\sum_{k\in K} y_{ka}\right) && (15)\\
\text{s.t. } & \sum_{a\in A}\int_0^{\sum_{k\in K} y_{ka}} \theta_a(s)ds \le \phi(x') + \hat{\phi}(x')\sum_{a\in A} x'_a(1-x_a), && \forall x'\in X,\\
& \sum_{a\in A_v^+} y_{ka} - \sum_{a\in A_v^-} y_{ka} = e'_{kv}, && \forall k\in K, v\in V,\\
& y_{ka} \le e_k x_a, && \forall k\in K, a\in A,\\
& y\ge 0, x\in X.
\end{aligned}
$$

For each of the 10 test instances using SiouxFalls transportation network, we measure the performances of the four formulations (10), (14), (9) and (15). Table 9 shows the number of test instances solved within the two-hour time limit and the average computational time. The disaggregated formulations are always worse than the aggregated correspondents on average, in terms of the number of instances solved within the timelimit and the average computational time.

Table 9: Performances of disaggregated and aggregated models

Formulation	Aggregation	Equation	Solved	Time
SD	No	(14)	1	3,428.6
	Yes	(10)	10	217.0
VF	No	(15)	6	2,305.8
	Yes	(9)	10	558.8

C Dual of Traffic Assignment Problem

In this section, we derive the dual traffic assignment problem (4). We start with a lemma.

Lemma 1. *Let $p \geq 1$, $a \geq 0$ and $b \in \mathbb{R}^n$. For $z \in \mathbb{R}^n_+$, define*

$$f(z) = \frac{a}{p+1}\left(\sum_{i=1}^{n} z_i\right)^{p+1} - \sum_{i=1}^{n} b_i z_i.$$

1. If $a > 0$, then

$$\min_{z\in\mathbb{R}^n_+} f(z) = -\frac{p}{a^{1/p}(p+1)}\Big([\max\{b_i : i = 1,\ldots,n\}]_+\Big)^{(p+1)/p},$$

where $[x]_+ = \max\{x, 0\}$.

2. If $a = 0$, then

$$\min_{z\in\mathbb{R}^n_+} f(z) = \begin{cases} 0, & \text{if } b_i \leq a \text{ for all } i = 1,\ldots,n, \\ -\infty, & \text{otherwise.} \end{cases}$$

In particular, Lemma 1 implies that

$$\min_{z\geq 0} f(z) = \begin{cases} \max\limits_{\eta\geq 0}\left\{-\dfrac{p}{\bar{a}^{1/p}(p+1)}\eta^{(p+1)/p} : \eta \geq b_i,\ \forall i = 1,\ldots,n\right\}, & \text{if } a > 0, \\ \max\limits_{\eta\geq 0}\left\{-\dfrac{p}{\bar{a}^{1/p}(p+1)}\eta^{(p+1)/p} : \eta = 0, \eta \geq b_i,\ \forall i = 1,\ldots,n\right\}, & \text{otherwise,} \end{cases} \tag{16}$$

where $\bar{a} = a$ if $a > 0$ and otherwise $\bar{a} = 1$.

Proof. It is straightforward to show the case when $a = 0$. In the following, we assume $a > 0$.

Pick i' such that $b_{i'} \geq b_i$ for any $i = 1, \dots, n$. For every $z \in \mathbb{R}^n_+$, we have $f(z) \geq f(z')$, where

$$z'_i = \begin{cases} \sum_{j=1}^{n} z_j, & i = i', \\ 0, & i \neq i'. \end{cases}$$

It follows that

$$\min_{z \in \mathbb{R}^n_+} f(z) = \min_{u \in \mathbb{R}_+} g(u),$$

where

$$g(u) = \frac{a}{p+1} u^{p+1} - b_{i'} u.$$

By computing the derivative of g, we can obtain the minimizer as

$$\min_{u \in \mathbb{R}_+} g(u) = g\left(\left(\frac{[b_{i'}]_+}{a}\right)^{1/p}\right) = -\frac{p}{a^{1/p}(p+1)} \left([b_{i'}]_+\right)^{(p+1)/p}.$$

□

Now, we derive the dual of problem (2).

Proof of Proposition 1. Fix $x \in X$. Let π be the dual variable corresponding to constraint (2b). Then, the Lagrangian is

$$\begin{aligned} L(z, \pi) = &\sum_{a \in A(x)} \int_0^{\sum_{o \in O} z_{oa}} \theta_a(s) ds - \sum_{o \in O} \sum_{v \in V} \pi_{ov} \left(\sum_{a \in A_v^+(x)} z_{oa} - \sum_{a \in A_v^-(x)} z_{oa} - e_{ov} \right) \\ = &\sum_{a \in A(x)} \left(f_a \left(\sum_{o \in O} z_{oa} \right) + \frac{g_a}{p+1} \left(\sum_{o \in O} z_{oa} \right)^{p+1} \right) \\ &\qquad - \sum_{o \in O} \sum_{a=(t,h) \in A(x)} (\pi_{oh} - \pi_{ot}) z_{oa} + \sum_{k \in K} e_k (\pi_{o_k d_k} - \pi_{o_k o_k}) \\ = &\sum_{a=(t,h) \in A(x)} \left(\frac{g_a}{p+1} \left(\sum_{o \in O} z_{oa} \right)^{p+1} - \sum_{o \in O} (\pi_{oh} - \pi_{ot} - f_a) z_{oa} \right) + \sum_{k \in K} e_k (\pi_{o_k d_k} - \pi_{o_k o_k}) \\ = &\sum_{a \in A(x)} L_a(z_a, \pi_a) + \sum_{k \in K} e_k (\pi_{o_k d_k} - \pi_{o_k o_k}), \end{aligned}$$

where

$$L_a(z_a, \pi_a) = \frac{g_a}{p+1} \left(\sum_{o \in O} z_{oa} \right)^{p+1} - \sum_{o \in O} (\pi_{oh} - \pi_{ot} - f_a) z_{oa}.$$

The Lagrangian dual problem is

$$\max_{\pi} \min_{z\geq 0} L(z,\pi) = \max_{\pi} \left\{ \sum_{a\in A(x)} \min_{z\geq 0} L_a(z_a,\pi_a) + \sum_{k\in K} e_k(\pi_{o_k d_k} - \pi_{o_k o_k}) \right\}.$$

In light of Lemma 1 and equation (16), this problem is equivalent to problem (4). □

D Outer-approximation methods

In this section, we present the outer-approximation method used to solve formulation (9). The method to solve formulation (10) is similar.

Using the definition of the travel cost function (1), we can rewrite formulation (9) as

$$\begin{aligned}
\min_{x,z,\theta} \; & \sum_{a\in A} f_a\left(\sum_{o\in O} z_{oa}\right) + g_a\theta_a \\
\text{s.t. } & \sum_{a\in A} f_a\left(\sum_{o\in O} z_{oa}\right) + \frac{g_a}{p+1}\theta_a \leq \phi(x') + \hat{\phi}(x')\sum_{a\in A} x'_a(1-x_a), && \forall x' \in X, \\
& \sum_{a\in A_v^+} z_{oa} - \sum_{a\in A_v^-} z_{oa} = e_{ov}, && \forall o\in O, v\in V, \\
& z_{oa} \leq \sum_{k\in K_o} e_k x_a, && \forall o\in O, a\in A, \\
& \theta_a \geq \left(\sum_{o\in O} z_{oa}\right)^{p+1} \\
& z\geq 0, x\in X.
\end{aligned}$$

Let us define

$$\begin{aligned}
u(X',I') = \min_{x,z,u} \; & \sum_{a\in A} f_a\left(\sum_{o\in O} z_{oa}\right) + g_a\theta_a \\
\text{s.t. } & \sum_{a\in A} f_a\left(\sum_{o\in O} z_{oa}\right) + \frac{g_a}{p+1}\theta_a \leq \phi(x') + \hat{\phi}(x')\sum_{a\in A} x'_a(1-x_a), && \forall x' \in X', \\
& \sum_{a\in A_v^+} z_{oa} - \sum_{a\in A_v^-} z_{oa} = e_{ov}, && \forall o\in O, v\in V, \\
& z_{oa} \leq \sum_{k\in K_o} e_k x_a, && \forall o\in O, a\in A, \\
& \theta_a \geq \alpha^{p+1} + (p+1)\alpha^p\left(\sum_{o\in O} z_{oa} - \alpha\right), && \forall a\in A, \alpha\in I'_a, \\
& z\geq 0, x\in X.
\end{aligned}$$

Problem $u(X', I')$ is equivalent to problem (9) if $X' = X$ and $I'_a = \mathbb{R}$ for all $a \in A$. However, if $X' \subsetneq X$ or $I'_a \neq \mathbb{R}$ for some $a \in A$, $u(X', I')$ is a relaxation of problem (9). In our implementation, we let $X' = I'_a = \emptyset$ for all $a \in A$ and pass $u(X', I')$ to Gurobi. Every time Gurobi solves a branch-and-bound node or finds a candidate integer solution, we search for a violated constraint (up to a tolerance of 10^{-6}), and add it as a lazy constraint.

E Additional Analyses for Section 5

In this section, we conduct an additional experiment to evaluate the scalability of the proposed methods. The experiment is based on a grid-shaped transportation network in which nodes are arranged in a 4×4 grid, and each pair of adjacent nodes is connected by an edge. The travel time on each edge is defined as $\theta(s) = 1 + \alpha s^4$, where α is sampled uniformly from the interval $[1, 2]$. The number of candidate arcs, denoted by n, is varied from 8 to 20. For each instance, n existing arcs are selected at random, and for each selected arc, a parallel arc is added to the network as a candidate arc. The travel time of each candidate arc is defined as $\theta(s) = 1 + s^4$. Demand is generated by randomly sampling twenty OD pairs of nodes. For each value of n, we generate 20 independent instances.

Table 10 reports the average computational time, the number of instances solved within the two-hour time limit, the dual bound at the root node (expressed as a percentage gap from the optimal objective value), the average number of branch-and-bound nodes, and the average ratio of computational time to the number of branch-and-bound nodes. When the number of candidate arcs is small, formulation SD is generally faster than formulation VF. However, as the number of candidate arcs increases, formulation VF becomes more competitive. When $n = 20$, formulation VF solves all 20 instances, whereas formulation SD fails to solve 4 instances within the time limit.

Interestingly, formulation SD tends to provide tighter dual bounds than formulation VF on average, and this difference becomes more pronounced as the number of candidate arcs increases. Nevertheless, both the number of branch-and-bound nodes and the average time required to solve the LP relaxation at each node increase with the number of candidate arcs. This increase is substantially sharper for formulation SD than for formulation VF, resulting in superior scalability of formulation VF compared to formulation SD.

F Detailed Explanation on Electric Vehicle Traffic Assignment

In this section, we review the network transformation to model the traffic assignment of EVs.

Let $G' = (V', A')$ be a road network. Let $O, D \subset V'$ denote the sets of nodes used as origins and destinations, respectively, and let $C \subset V'$ represent the set of nodes with charging stations. By duplicating nodes if necessary, we assume that O, D, and C are mutually disjoint.

Table 10: Performance of formulations SD (10) and VF (9) to solve the DNDP to minimize the total travel time

# Candidates	Formulation	Time	Solved	Root Bound (%)	# B&B Nodes	Time per Node
8	SD	6.66	20	29.74	60.15	0.11
	VF	13.00	20	30.34	48.80	0.27
9	SD	16.32	20	32.69	101.50	0.16
	VF	29.41	20	36.34	85.00	0.34
10	SD	37.12	20	36.45	157.15	0.23
	VF	72.92	20	35.78	135.70	0.51
11	SD	67.25	20	35.93	246.10	0.25
	VF	153.80	20	40.60	218.05	0.69
12	SD	261.59	20	39.64	443.60	0.53
	VF	288.59	20	44.01	328.75	0.80
13	SD	570.50	20	38.54	579.35	0.81
	VF	531.92	20	44.05	506.40	1.02
14	SD	616.22	20	38.38	695.00	0.86
	VF	535.93	20	46.51	603.00	0.92
15	SD	886.94	20	43.00	869.55	0.92
	VF	575.34	20	48.97	684.40	0.85
16	SD	1,026.87	20	40.98	911.70	1.08
	VF	639.62	20	50.53	748.05	0.89
17	SD	2,668.80	18	41.22	1,273.80	1.71
	VF	661.33	20	54.53	925.30	0.75
18	SD	2,266.75	19	45.31	1,256.55	1.52
	VF	595.28	20	51.55	824.40	0.72
19	SD	2,637.27	18	43.39	1,368.50	1.65
	VF	616.91	20	60.39	884.40	0.70
20	SD	3,856.50	16	45.12	1,720.35	1.93
	VF	864.45	20	57.03	1,094.40	0.74

An EV driver must choose a path that can be completed without running out of battery. For example, consider graph G_1 in Figure 1. A filled node indicates the presence of a charging station, while an open node denotes the absence of one. Given that the vehicle range is 200 km, path $(1, 4, 5, 3, 8)$ is not feasible, as it is not possible to travel from node 1 to node 3 (nodes 4 and 5 do not have charging stations).

MirHassani and Ebrazi [2013] proposed a technique to transform the road network to facilitate handling the vehicle range. The transformed graph $G = (V, A)$ has $V = O \cup D \cup C$, and for each $(o, d) \in O \times D$, a path from o to d in graph G corresponds to a path in road network G' that respects the vehicle range (and vice versa). The construction of A is straightforward: for each $t, h \in V$, we have $(t, h) \in A$ if and only if the distance from t to h is less than or equal to the vehicle range of an EV, and $(t, h) \in O \times C$, $C \times C$, or $C \times D$. For example, consider again road network G_1 shown in Figure 1. If there is a single OD pair (i.e., $K = \{1\}$) from node 1 to node 8, and the vehicle range of an EV is 200 km, we obtain G_2 after this transformation. Path $p_1 = (1, 2, 6, 8)$ in graph G_2 corresponds to path $p'_1 = (1, 4, 2, 4, 6, 7, 8)$ in graph G_1. The intermediate nodes 2 and 6 in path p_1 correspond to the nodes to charge an EV.

We note that the cost (travel time) of a path in the transformed network consists of both arc costs (time spent on the road) and node costs (time spent at the charging stations). By introducing auxiliary nodes and arcs, the node costs can be represented as costs associated with the auxiliary arcs. For example, if the time to charge an EV at a charging station is given by $1/2 + s^4$, graph G_2 can be further transformed into graph G_3.

G Additional Analyses for Section 6

In this section, we provide additional analyses on the experiments in Section 6.

G.1 Histogram of Demands

Figure 2 shows the histogram of demand volumes. The histogram uses the low-demand scenario. However, the demand volumes for the high-demand scenario are obtained by scaling the low-demand volumes by 50%. Thus, the shape of the histogram for the high-demand scenario remains the same.

G.2 Comparison of Equilibria

Figures 3 and 4 show the total travel time and the total/worst costs of sustainability of the equilibria computed by formulation SD to optimize the total travel time (10), the total cost of sustainability (11) and the worst cost of sustainability (12). The color and shading pattern of each marker indicate the metric being optimized, while the shape of the marker corresponds to the budget: a square, triangle, and circle represents a budget of 5, 10, and 20, respectively.

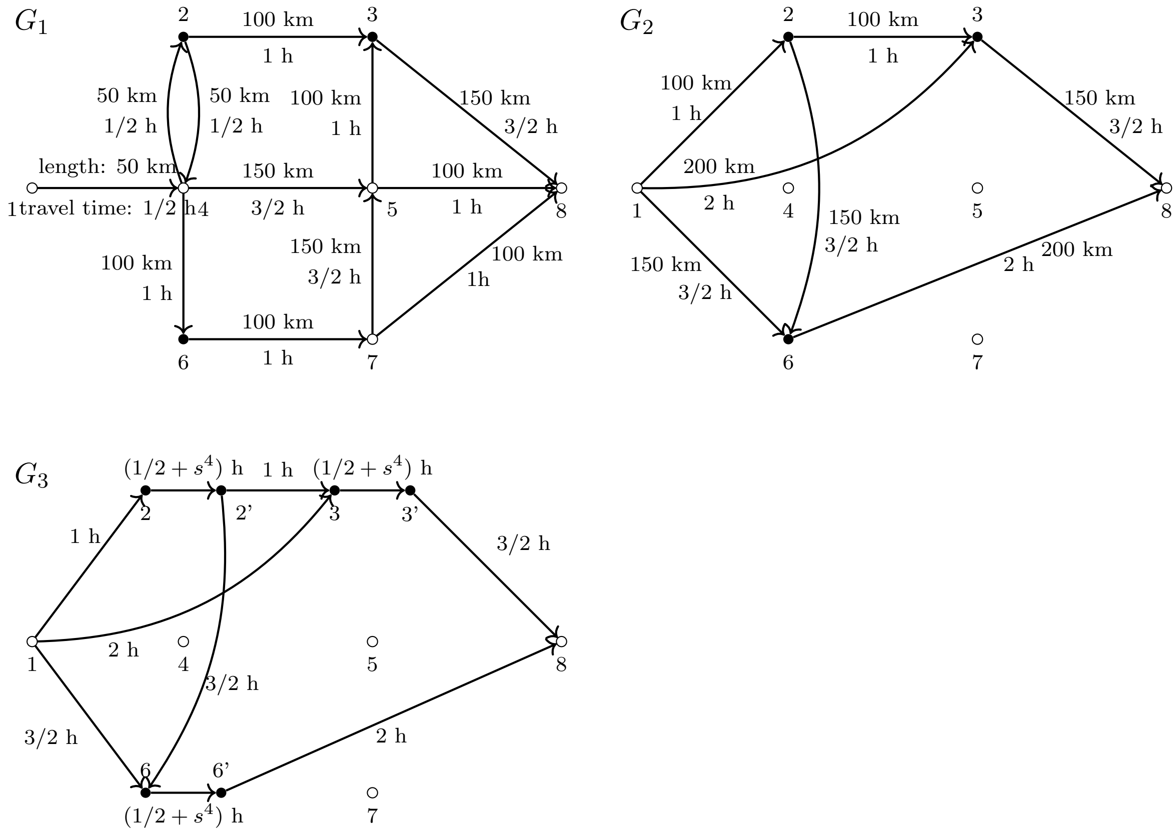

Figure 1: Graphs illustrating network transformation: Road network G_1 is transformed into the extended graph G_2, which is further converted into G_3, where the node costs in G_2 are represented as arc costs.

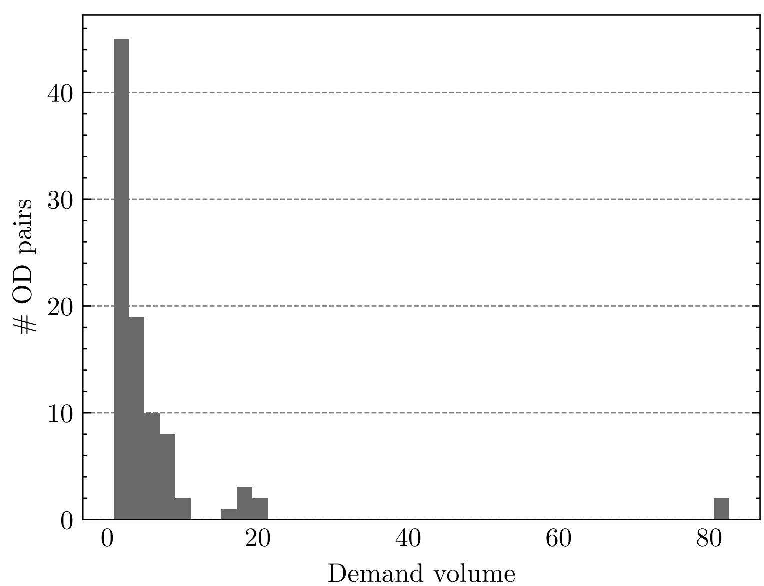

Figure 2: Histogram of the demand volumes in the low-demand scenario

For larger budgets, optimizing for one objective tends to yield near-optimal values for the other objectives as a byproduct. However, for smaller budgets, differences start to emerge, highlighting the need for decision-makers in these cases to carefully choose their priorities.

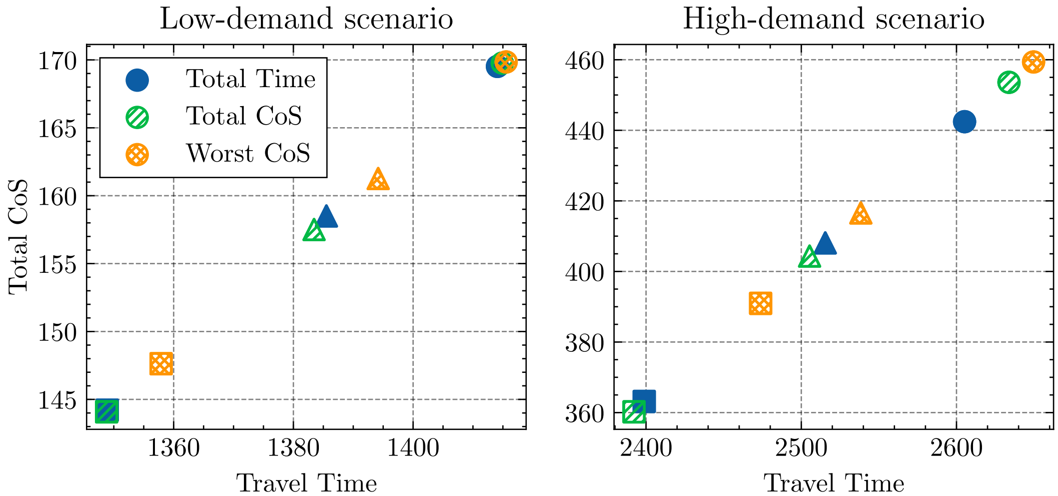

Figure 3: Total travel time and total cost of sustainability of the user equilibria computed at various equilibria

G.3 Comparison of Charging Station Locations for Capacity Expansion

Figures 5 and 6 show the selected charging stations for capacity expansion in the two demand scenarios, with a budget of 5. These figures are generated using the outputs of formulation SD, which optimizes the total travel time (10). The outputs differ between demand scenarios, indicating the sensitivity of the results to the demand data.

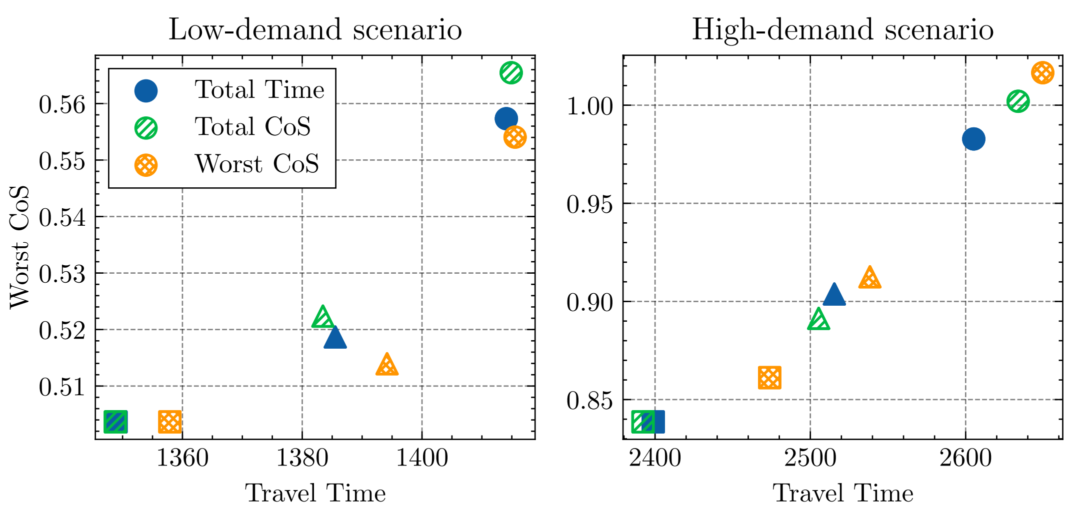

Figure 4: Total travel time and worst cost of sustainability of the user equilibria computed at various equilibria

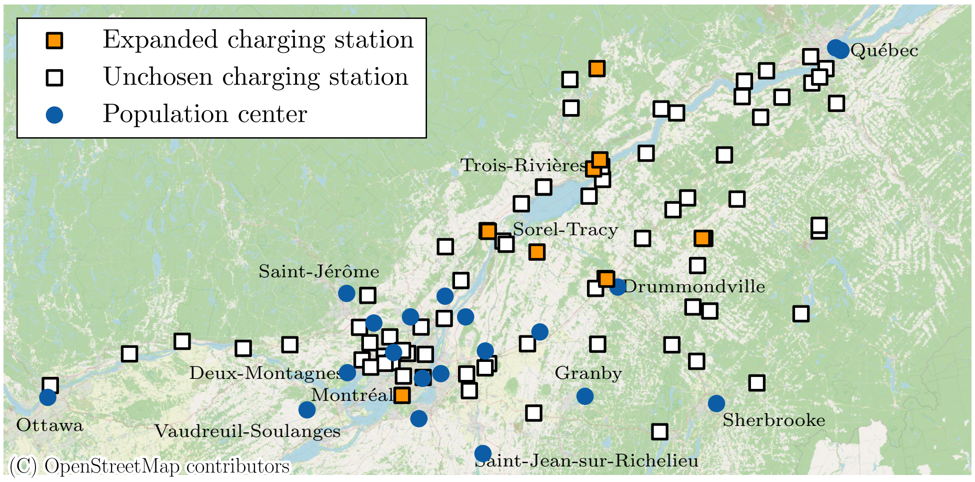

Figure 5: Selected charging stations in the low-demand scenario with a budget of 5

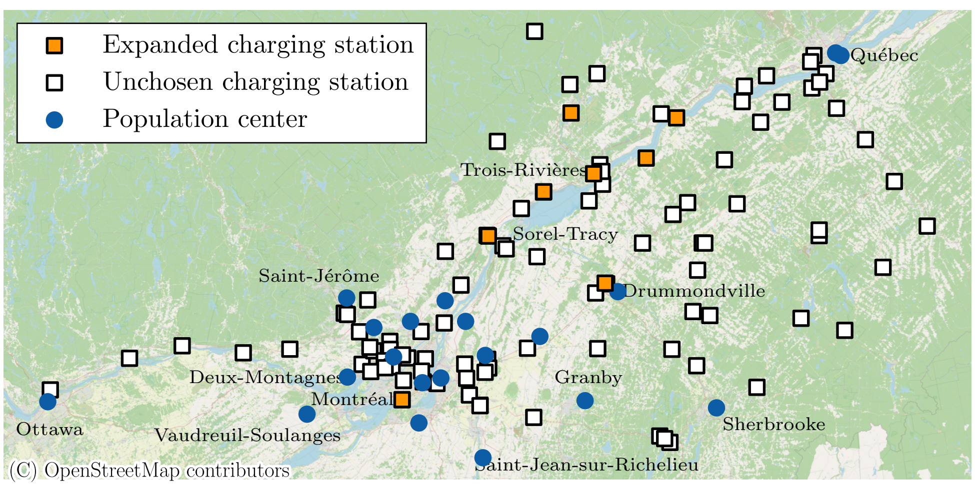

Figure 6: Selected charging stations in the high-demand scenario with a budget of 5