SIAM J. OPTIM.
Vol. 27, No. 3, pp. 1554–1582

COMPLEXITY AND APPROXIMATION OF THE CONTINUOUS NETWORK DESIGN PROBLEM*

MARTIN GAIRING†, TOBIAS HARKS‡, AND MAX KLIMM§

Abstract. We revisit a classical problem in transportation, known as the (bilevel) *continuous network design problem*, CNDP for short. Given a graph for which the latency of each edge depends on the ratio of the edge flow and the capacity installed, the goal is to find an optimal investment in edge capacities so as to minimize the sum of the routing costs of the induced Wardrop equilibrium and the investment costs for installing the edge's capacities. While this problem is considered to be challenging in the literature, its complexity status was still unknown. We close this gap, showing that CNDP is strongly NP-hard and APX-hard, both on directed and undirected networks and even for instances with affine latencies. As for the approximation of the problem, we first provide a detailed analysis for a heuristic studied by Marcotte for the special case of *monomial* latency functions [P. Marcotte, *Math. Prog.*, 34 (1986), pp. 142–162]. We derive a closed form expression of its approximation guarantee for *arbitrary* sets of latency functions. We then propose a different approximation algorithm and show that it has the same approximation guarantee. Then, we prove that using the better of the two approximation algorithms results in a strictly improved approximation guarantee for which we derive a closed form expression. For affine latencies, for example, this best-of-two approach achieves an approximation factor of $49/41 \approx 1.195$, which improves on the factor of $5/4$ that has been shown before by Marcotte.

Key words. bilevel optimization, optimization under equilibrium constraints, network design, Wardrop equilibrium, computational complexity, approximation algorithms

AMS subject classifications. 90C33, 90C35

DOI. 10.1137/15M1016461

1. Introduction.

Since the seminal works of Pigou [25] and Wardrop [31], the impact of selfish behavior in congested transportation networks has been investigated intensively. Wardrop stated in his first principle the following notion of an equilibrium: "The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route." For a formalization of this principle in the context of selfish routing, consider a directed graph with latency functions on the edges and a set of origin-destination pairs, called *commodities*. Every commodity has a *demand* associated with it, which specifies the amount of flow that needs to be sent from the respective origin to the respective destination. The *latency* that a flow particle experiences when traversing an edge depends on the edge flow and is determined by a nondecreasing latency function. In practice, latency functions are calibrated to reflect edge specific parameters such as street length and capacity. One of the most prominent and popular classes of functions used in actual traffic models are the ones put forward by the Bureau of Public Roads (BPR) [30]; BPR-type latency

*Received by the editors April 10, 2015; accepted for publication (in revised form) March 6, 2017; published electronically August 3, 2017. A preliminary extended abstract of this paper appeared in the proceedings of the 17th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems 2014, [15]. This work was supported by EPSRC grants EP/J019399/1 and EP/L011018/1.

http://www.siam.org/journals/siopt/27-3/M101646.html

Funding: The research of the third author was carried out in the framework of MATHEON, supported by Einstein Foundation Berlin, B-M15.

†Department of Computer Science, University of Liverpool, United Kingdom (gairing@liverpool.ac.uk).

‡Maastricht University, The Netherlands (t.harks@maastrichtuniversity.nl).

§Institut für Mathematik, Technische Universität Berlin, Germany (max.klimm@hu-berlin.de).

functions are of the form $S_e(f_e) = t_e\big(1 + b_e(f_e/z_e)^4\big)$, where f_e is the edge flow, t_e represents the free-flow travel time, $b_e > 0$ is an edge-specific bias, and z_e represents the street capacity. A *Wardrop equilibrium* is a multicommodity flow, in which each commodity sends flow only along paths with minimum latency from its origin to its destination.

It is well known that Wardrop equilibria can be inefficient in the sense that they do not minimize the total travel time in the network [13]. The worst-case ratio of the total travel time of a Wardrop equilibrium and an optimal flow is known as the *price of anarchy*, and exact values for the price of anarchy of selfish routing are known for specific classes of latency functions [26, 28]. A prominent example of this inefficiency can be observed in the Braess paradox [7], where improving the network infrastructure by adding street capacity results in a Wardrop equilibrium with strictly higher total travel time. This nonmonotonic behavior illustrates that designing networks for efficient Wardrop equilibria is an important and nontrivial issue.

In this paper, we revisit a variant of the *continuous network design problem* (CNDP for short) previously studied by Marcotte [23]: We are given a graph for which the latency of each edge depends on the ratio of the edge flow and the capacity installed, and the goal is to find an optimal investment in edge capacities so as to minimize the sum of the routing cost of the induced Wardrop equilibrium and the investment cost. Note that throughout this paper we assume that latency functions are *edge-separable*, that is, for fixed capacities the latency of every edge depends only on its edge flow.

From a mathematical perspective, CNDP is a bilevel optimization problem (see, e.g., [8, 21] for an overview), where in the upper level the edge capacities are determined and, given these capacities, in the lower level the flow settles into a Wardrop equilibrium in which, for each commodity, only shortest paths are used. It is well known that the equilibrium condition in the lower level can be formulated as a minimization problem [3] thus turning CNDP into a bilevel optimization problem. Clearly, the lower level reaction depends on the first level decision because altering the capacity investment on a subset of edges may result in revised route choices by the users.

CNDP (including different variants) has been intensively studied since the late 1960s (cf. [1, 10, 12, 22]) and several heuristic approaches have been proposed since then; see Yang et al. [32] for a comprehensive survey. Most of the proposed heuristics are numerical in nature and involve iterative computations of relaxations of the problem (for instance the iterative optimization and assignment algorithm [24] and augmented Lagrangian methods or linearizations of the objective in the leader and follower problem) and no worst-case approximation guarantee is provided. An exception is the work of Marcotte [23], who considered several algorithms based on solutions of associated convex optimization problems which can be solved within arbitrary precision in polynomial time [16]. For *monomial* latency functions, he derives worst-case bounds for his heuristics; e.g., for affine latency functions he obtains a 5/4-approximation. For general monomial latency functions plus a constant (including BPR-type latency functions) he obtains a 2-approximation.

1.1. Our results and used techniques. Although CNDP and their variants have been studied by many researchers for more than 40 years, to the best of our knowledge, the computational complexity status of CNDP is still unknown. We settle this question by showing that CNDP is strongly NP-hard and APX-hard, both on directed and undirected networks and even for instances with affine latencies of the form

$S_e(f_e/z_e) = \alpha_e + \beta_e(f_e/z_e)$, $\alpha_e, \beta_e \geq 0$. For the proof of the NP-hardness, we reduce from 3-SAT. The reduction has the property that, in the case that the underlying instance of 3-SAT has a solution, the cost of an optimal solution is equal to the minimal cost of a relaxation of the problem, in which the equilibrium conditions are relaxed. The key challenge of the hardness proof is to obtain a lower bound on the optimal solution when the underlying 3-SAT instance has no solution. To this end, we relax the equilibrium conditions only partially, which enables us to bound the cost of an optimal solution from below by solving an associated constrained quadratic optimization problem. With a slightly adapted construction and a more detailed analysis, we can even prove APX-hardness of the problem. Here, we reduce from a symmetric variant of MAX-3-SAT, in which all literals occur exactly twice. This problem is known to be NP-hard to approximate by any factor better than 1016/1015; see Berman et al. [4, 5]. While all our hardness proofs rely on instances with an arbitrary number of commodities and respective sinks, a result by Marcotte [23] implies that, for instances in which all commodities share a common sink, CNDP can be solved to optimality in polynomial time.

In light of the hardness of CNDP for general network topologies, we focus on approximation algorithms. We first consider a polynomial time algorithm proposed by Marcotte [23]. This algorithm, which we call BRINGTOEQUILIBRIUM, first computes a relaxation of CNDP by removing the equilibrium conditions. Then, it reduces the edge capacities of each edge individually such that the flow computed in the relaxation becomes a Wardrop equilibrium. We give a novel closed form expression of the performance of this algorithm with respect to the set $\mathcal{S}$ of allowed latency functions. Specifically, we show that this algorithm is a $(1 + \mu(\mathcal{S}))$-approximation, where

$$\mu(\mathcal{S}) = \sup_{S\in\mathcal{S}} \sup_{x\geq 0: S(x)>0} \max_{\gamma\in[0,1]} \left\{ \gamma \left(1 - \frac{S(\gamma x)}{S(x)} \right) \right\}.$$

The value $\mu(\mathcal{S})$ has been used before by Correa et al. [9] and Roughgarden [26] in the context of price-of-anarchy bounds for selfish routing. Specifically, they showed that the routing cost of a Wardrop equilibrium is at most a factor of $\frac{1}{1-\mu(\mathcal{S})}$ away from the routing cost of a system optimum. The anarchy value of $\frac{1}{1-\mu(\mathcal{S})}$ appears naturally in the context of approximation algorithms for the continuous network design problem as the following (trivial) algorithm has an approximation guarantee of $\frac{1}{1-\mu(\mathcal{S})}$: first compute a relaxation of CNDP by removing the equilibrium conditions, then fix the capacities of the relaxation and compute a corresponding Wardrop equilibrium. It is worth noting that, since $1 + \mu(S) < \frac{1}{1-\mu(S)}$ for all sets of latency functions $\mathcal{S}$ with $\mu(S) > 0$, the algorithm of Marcotte has a better approximation guarantee than the trivial algorithm.

For the special case that $\mathcal{S}$ is the set of polynomials with non-negative coefficients and maximal degree Δ, we derive exactly the approximation guarantees that Marcotte obtained for monomials. As an outcome of our more general analysis, we further derive that this algorithm is a 2-approximation for general convex latency functions and a 5/4-approximation for latency functions that are both concave and semi-convex.

We then propose a new algorithm which we call SCALEUNIFORMLY. This algorithm first computes an optimal solution of the relaxation as before and then *uniformly* scales the capacities with a certain parameter $\lambda(\mathcal{S})$ that depends on the class of allowable latency functions $\mathcal{S}$. Based on well-known techniques using variational inequalities (Correa et al. [9] and Roughgarden [26]), we prove that this algorithm also yields a $(1 + \mu(\mathcal{S}))$-approximation.

TABLE 1

Approximation guarantees of the algorithms BRINGTOEQUILIBRIUM*,* SCALEUNIFORMLY*, and the best of the two for convex latency functions, concave latency functions, and sets of polynomials with non-negative coefficients depending on the maximal degree* Δ*. The approximation guarantees stated for convex latency functions even hold for sets of semi-convex latency functions as in Assumption* 2.1*. For* BRINGTOEQUILIBRIUM*, the approximation guarantees marked with* ($\star$) *have been obtained before in* [23]*.*

Functions	Approximation guarantees: BRINGTOEQUILIBRIUM / SCALEUNIFORMLY	Approximation guarantees: Better of the two
concave	$5/4 = 1.25$	$49/41 \approx 1.195$
convex	2	$9/5 = 1.8$
polynomials Δ		
0	1	1
1/4	$3381/3125 \approx 1.082$	≈ 1.064
1/3	$283/256 \approx 1.105$	≈ 1.083
1/2	$31/27 \approx 1.148$	$1849/1657 \approx 1.116$
1	$5/4 = 1.25\ ^\star$	$49/41 \approx 1.195$
2	$1 + \frac{2}{9}\sqrt{3} \approx 1.385^\star$	$\frac{311}{479} + \frac{180}{479}\sqrt{3} \approx 1.300$
3	$1 + \frac{3}{16}\sqrt[3]{4^2} \approx 1.472^\star$	≈ 1.369
4	$1 + \frac{4}{25}\sqrt[4]{5^3} \approx 1.535^\star$	≈ 1.418
∞	$2\ ^\star$	$9/5 = 1.8$

As our main result regarding approximation algorithms, we show that using the better of the two solutions returned by BRINGTOEQUILIBRIUM and SCALEUNIFORMLY yields strictly better approximation guarantees. We give a closed form expression for the new approximation guarantee as a function of $\mathcal{S}$. We demonstrate the applicability of this general bound by showing that the best-of-two algorithm achieves a $9/5$-approximation for $\mathcal{S}$ containing arbitrary convex latencies. For affine latencies it achieves a $49/41 \approx 1.195$-approximation, improving on the $5/4$ of Marcotte. An overview of our results compared to those of Marcotte can be found in Table 1.

1.2. Further related work. Quoting [32], CNDP has been recognized to be "one of the most difficult and challenging problems in transport" and there are numerous works approaching this problem. In light of the substantial literature on heuristics for CNDP, we refer the reader to the surveys [8, 14, 22, 32].

While, to the best of our knowledge prior to this work, the complexity status of CNDP was open, there have been several papers on the complexity of the *discrete (bilevel) network design problem*, DNDP for short; see [20, 27]. Given a network with edge latency functions and traffic demands, a basic variant of DNDP is to decide which edges should be removed from the network to obtain a Wardrop equilibrium in the resulting subnetwork with minimum total travel time. This variant is motivated by the classical Braess paradox, where removing an edge from the network may improve the travel time of the new Wardrop equilibrium. Roughgarden [27] showed that DNDP is strongly NP-hard and that there is no $(\lfloor n/2 \rfloor - \epsilon)$-approximation algorithm (unless $\mathsf{P} = \mathsf{NP}$), even for single-commodity instances. He further showed that for single-commodity instances the trivial algorithm of not removing any edge from the graph is essentially best possible and achieves a $\lfloor n/2 \rfloor$-approximation. For affine latency functions, the trivial algorithm gives a $4/3$-approximation (even for general networks) and this is also shown to be best possible. These results in comparison to ours highlight interesting differences. While DNDP is not approximable by any constant for convex latencies, for CNDP we give a $9/5$-approximation. Moreover, all hardness results for DNDP already hold for single-commodity instances, while in that case CDNP is solvable in polynomial time.

A more general variant of CNDP arises when initial edge capacities are given and a budget must be distributed among the edges to improve the resulting equilibrium; see, e.g., [12, 24, 32]. Very recently and independently of this work, Bhaskar et al. [6] studied the complexity of this problem. Among other results they show that the problem is weakly NP-hard in single-commodity networks that consist of parallel links in series. This again stands in contrast to the polynomial-time algorithm for single-commodity networks for the variant of CDNP considered in this paper. A series of works considered simplified network topologies (e.g., parallel links) or special latency functions (e.g., $M/M/1$ latency functions); see [6, 17, 18, 19].

2. Preliminaries. Let $G = (V, E)$ be a directed or undirected graph, V its set of vertices, and $E \subseteq V \times V$ its set of edges. We are given a set K of *commodities*, where each commodity $k \in K$ is associated with a triple $(s_k, t_k, d_k) \in V \times V \times \mathbb{R}_{>0}$, where $s_k \in V$ is the *source*, $t_k \in V$ the *sink*, and d_k the *demand* of commodity k. A multicommodity flow on G is a collection of non-negative flow vectors $(\boldsymbol{f}^k)_{k \in K}$ such that for each $k \in K$ the flow vector $\boldsymbol{f}^k = (f_e^k)_{e \in E}$ satisfies the flow conservation constraints:

$$\begin{aligned}
&\textstyle\sum_{u \in V:(s_k,u) \in E} f^k_{(s_k,u)} - \sum_{u \in V:(u,s_k) \in E} f^k_{(u,s_k)} = d_k,\\
&\textstyle\sum_{u \in V:(u,t_k) \in E} f^k_{(u,t_k)} - \sum_{u \in V:(t_k,u) \in E} f^k_{(t_k,u)} = d_k,\\
&\textstyle\sum_{u \in V:(u,w) \in E} f^k_{(u,w)} - \sum_{u \in V:(w,u) \in E} f^k_{(w,u)} = 0 \qquad \text{for all } w \in V \setminus \{s_k, t_k\}.
\end{aligned}$$

Whenever we write $\boldsymbol{f}$ without a superscript k for the commodity, we implicitly sum over all commodities, i.e., $f_e = \sum_{k \in K} f_e^k$ and $\boldsymbol{f} = (f_e)_{e \in E}$. We call f_e an *edge flow*. The set of all feasible edge flows will be denoted by $\mathcal{F}$.

The latency of each edge e depends on the installed capacity $z_e \geq 0$ and the edge flow f_e on e, and is given by a latency function $S_e : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ that maps f_e/z_e to a latency value $S_e(f_e/z_e)$, where we use the convention that $S_e(f_e/z_e) = \infty$ whenever $z_e = 0$. Throughout this paper, we assume that the set of allowable latency functions is restricted to some set $\mathcal{S}$ and we impose the following assumptions on $\mathcal{S}$.

Assumption 2.1. The set $\mathcal{S}$ of allowable latency functions is nonempty and only contains strictly increasing, unbounded, continuously differentiable, and semi-convex functions $S : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$. (A function S is called semi-convex if the function $x \mapsto x \cdot S(x)$ is convex.)

We note that Assumption 2.1 is slightly more general than requiring that all latency functions are strictly increasing and convex. For instance, the function $x \mapsto \sqrt{x}$ satisfies Assumption 2.1 although it is concave.

For a fixed vector of capacities $\boldsymbol{z} = (z_e)_{e \in E}$, the latency of each edge e only depends on the edge flow f_e. Under these conditions, there exists a Wardrop flow $\boldsymbol{f} = (f_e)_{e \in E}$, i.e., a flow in which each commodity only uses paths of minimal latency. It is well known (cf. [3, 11, 29]) that each Wardrop flow is a solution to the optimization problem

$$\min_{\boldsymbol{f} \in \mathcal{F}} \left\{ \sum_{e \in E} \int_0^{f_e} S_e(t/z_e)\, \mathrm{d}t \right\}, \tag{2.1}$$

and satisfies the *variational inequality*

$$\sum_{e\in E} S(f_e/z_e)(f_e - f'_e) \le 0 \tag{2.2}$$

for every feasible flow $\boldsymbol{f}' = (f'_e)_{e\in E} \in \mathcal{F}$. For a vector of capacities $\boldsymbol{z}$ we denote by $\mathcal{W}(\boldsymbol{z})$ the corresponding set of Wardrop flows $\boldsymbol{f}(\boldsymbol{z})$. Beckmann et al. [3] showed that Wardrop flows and optimum flows are related in the following way.

PROPOSITION 2.2 (Beckmann et al. [3]). *Let* $S^*_e(x) = (xS_e(x))' = S_e(x)+xS'_e(x)$ *be the* marginal cost function *of edge* $e \in E$. *Then* $\boldsymbol{f}^*$ *is an optimum flow with respect to the latency functions* $(S_e)_{e\in E}$ *if and only if it is a Wardrop flow with respect to* $(S^*_e)_{e\in E}$.

In the continuous network design problem (CNDP) the goal is to buy capacities z_e at a price per unit $\ell_e > 0$ so as to minimize the sum of the construction cost $C^Z(\boldsymbol{f},\boldsymbol{z}) = \sum_{e\in E} z_e\,\ell_e$ and the routing cost $C^R(\boldsymbol{f},\boldsymbol{z}) = \sum_{e\in E} S_e(f_e/z_e)\,f_e$ of a resulting Wardrop equilibrium $\boldsymbol{f}$. Observe that $C^R(\boldsymbol{f},\boldsymbol{z})$ is well defined as, by (2.1), it is the same for all Wardrop equilibria with respect to $\boldsymbol{z}$. Denote the combined cost by $C(\boldsymbol{f},\boldsymbol{z}) = C^R(\boldsymbol{f},\boldsymbol{z}) + C^Z(\boldsymbol{f},\boldsymbol{z})$.

DEFINITION 2.3 (Continuous network design problem (CNDP)). *Given a directed graph* $G = (V,E)$ *and, for each edge* e, *a latency function* S_e *and a construction cost* $\ell_e > 0$, *the continuous network design problem (CNDP) is to determine a non-negative capacity vector* $\boldsymbol{z} = (z_e)_{e\in E}$ *that minimizes*

$$\min_{\boldsymbol{z}\ge 0,\,\boldsymbol{f}\in\mathcal{W}(\boldsymbol{z})} \left\{ \sum_{e\in E} \Big(S_e(f_e/z_e)\; f_e + z_e\;\ell_e \Big) \right\}. \tag{CNDP}$$

Removing the condition that $\boldsymbol{f}$ is a Wardrop equilibrium in (CNDP), we obtain the following relaxation of the continuous network design problem:

$$\min_{\boldsymbol{z}\ge 0,\,\boldsymbol{f}\in\mathcal{F}} \left\{ \sum_{e\in E} \Big(S_e(f_e/z_e)\; f_e + z_e\;\ell_e \Big) \right\}. \tag{CNDP$'$}$$

Marcotte [23] showed that for *convex* and unbounded latency functions, the relaxed problem (CNDP$'$) can be solved efficiently by performing $|K|$ independent shortest path computations on the graph G, one for each commodity $k \in K$. The following proposition slightly generalizes his result to arbitrary, not necessarily convex latency functions that satisfy Assumption 2.1.

PROPOSITION 2.4 (Marcotte [23]). *The relaxation* (CNDP$'$) *can be solved by performing* $|K|$ *shortest path computations in polynomial time.*

Proof. The proof is based on Marcotte's proof idea for convex latency functions. We include it for completeness and mention the new arguments needed to cope with cost functions that satisfy Assumption 2.1 but are *not* convex.

First of all, by Assumption 2.1 it follows that the objective in (CNDP$'$) is convex, and thus first-order conditions are also sufficient for global optimality. As noted by Marcotte, we can first solve (CNDP$'$) in terms of capacities only leaving the flow values as parameters. We obtain as a necessary optimality condition that for every edge $e \in E$ with $z_e > 0$ we must have

$$S'_e(f_e/z_e)(f_e/z_e)^2 = l_e.$$

Provided that one can solve the equation

$$S'_e(x)x^2 = l_e \tag{2.3}$$

for every $e \in E$ beforehand, one can use the corresponding solutions $u_e \geq 0$ to derive a simpler reformulation of (CNDP$'$) leading to the following unsplittable flow problem:

$$\min_{\boldsymbol{f} \in \mathcal{F}} \left\{ \sum_{e \in E} \big(S_e(u_e) + \ell_e/u_e\big) f_e \right\}.$$

This problem can be solved by $|K|$ shortest path calculations.

For convex latency functions (as in [23]), (2.3) always admits a unique solution. To be able to deal with nonconvex functions S, we prove that under the weaker Assumption 2.1 we have $\lim_{x \downarrow 0} S'(x)x^2 = 0$ and $\lim_{x \uparrow +\infty} S'(x)x^2 = +\infty$. For the proof, see Lemma A.1 in the appendix. Then, using continuity of $S'(x)x^2$, the intermediate value theorem implies the existence of a solution to (2.3). ∎

It follows that for graphs with a single sink (or a single source), one can solve even (CNDP) by computing a shortest path tree.

Corollary 2.5 (Marcotte [23]). *In networks with only one sink vertex t (or source vertex s), the continuous network design problem (CNDP) can be solved in polynomial time.*

The proof relies on first solving (CNDP$'$) (as above) and then observing that the resulting shortest path tree (using that there is a single sink (or source) only) induces unique paths for every commodity with finite latency, thus inducing the optimal flow as a Wardrop equilibrium.

Remark 2.6. To speak about polynomial algorithms and hardness, we need to specify how the instances of CNDP, in particular the latency functions, are encoded; cf. [2, 16, 27]. While our hardness results hold even if all functions are linear and given by their (rational) coefficients, for our approximation algorithms we require that we can solve equations involving a latency function and its derivative, e.g., (2.3). Without this assumption, we still obtain the claimed approximation guarantees within arbitrary precision by polynomial time algorithms.

3. Hardness results. As the main results of this section, we show that CNDP is strongly NP-hard and APX-hard both on directed and undirected networks and even for affine latency functions. The main difficulty of the proofs is to obtain good lower bounds on the objective value of a solution.

3.1. NP-hardness. For ease of exposition, we first show that CNDP on directed networks is strongly NP-hard.

Theorem 3.1. *The continuous network design problem (CNDP) on directed networks is* NP*-hard in the strong sense, even if all latency functions are affine.*

Proof. To show the NP-hardness of the problem, we reduce from 3-Sat. Let ϕ be a Boolean formula in conjunctive normal form. We denote the set of variables and clauses of ϕ with $V(\phi)$ and $K(\phi)$, respectively, and set $\nu = |V(\phi)|$ and $\kappa = |K(\phi)|$. We denote by $L(\phi)$ the set of literals over $V(\phi)$, i.e., $L(\phi) = \{x_i \in V(\phi)\} \cup \{\bar{x}_i : x_i \in V(\phi)\}$. A solution of ϕ is a subset $A \subset L(\phi)$ such that $|A \cap \{x_i, \bar{x}_i\}| = 1$ and for all clauses $k = l_k \vee l'_k \vee l''_k \in K(\phi)$ we have $|A \cap \{l_k, l'_k, l''_k\}| \geq 1$. The computational problem 3-Sat is to decide for a given formula ϕ whether a solution exists.

We now explain the construction of a continuous network design problem based on ϕ that has the property that, for some $\epsilon \in (0, 1/8)$, an optimal solution has total cost less or equal to $(4 + \epsilon)\kappa + 2\kappa\nu$ if and only if ϕ has a solution; cf. Figure 1. Let $\epsilon \in (0, 1/8)$ be arbitrary. For each clause $k \in K(\phi)$, we introduce a *clause edge*

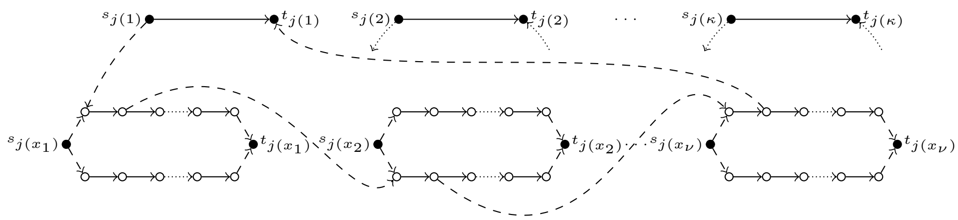

FIG. 1. *Hardness of the continuous network design problem. Clause* 1 *is equal to* $x_1 \vee \bar{x}_2 \vee x_\nu$. *Dashed edges have zero latency.*

$e(k)$ with latency function $S_{e(k)}(f_{e(k)}/z_{e(k)}) = 4 + f_{e(k)}/z_{e(k)}$ and construction cost $\ell_{e(k)} = (\epsilon/2)^2$. For each literal $l \in L(\phi)$ and each clause $k \in K(\phi)$, we introduce a *literal edge* $e(l,k)$ with latency function $S_{e(l,k)}(f_{e(l,k)}/z_{e(l,k)}) = f_{e(l,k)}/z_{e(l,k)}$ and construction cost $\ell_{e(l,k)} = 1$. We denote the set of clause edges and literal edges by E_K and E_L, respectively.

For each variable $x_i \in V(\phi)$, there is a *variable commodity* $j(x_i)$ with source $s_{j(x_i)}$, sink $t_{j(x_i)}$, and demand $d_{j(x_i)} = 1$. This commodity has two feasible paths: one path uses exclusively the literal edges $\{e(x_i,k) : k \in K(\phi)\}$ that correspond to the non-negated variable x_i, the other path uses exclusively the literal edges $\{e(\bar{x}_i,k) : k \in K(\phi)\}$ that correspond to the negated variable $\bar{x}_i$. In that way, each feasible path of the variable commodity $j(x_i)$ corresponds to a truth assignment of the variable x_i. For each clause $k = l_k \vee l'_k \vee l''_k$ with $l_k, l'_k, l''_k \in L(\phi)$, we introduce a *clause commodity* $j(k)$ with source $s_{j(k)}$, sink $t_{j(k)}$ and demand $d_{j(k)} = 1$. The clause commodity may either choose its corresponding clause edge $e(k)$ or the corresponding literal edges that occur in k, i.e., $e(l_k,k)$, $e(l'_k,k)$, and $e(l''_k,k)$. For notational convenience, we set $E_L(k) = \{e(l_k,k), e(l'_k,k), e(l''_k,k)\}$. We add some additional edges with latency 0 to obtain a network; see Figure 1 where these edges are dashed. In order to make sure that no further paths emerge, it is important that, for each variable commodity $j(x_i)$ and each literal $l \in \{x_i, \bar{x}_i\}$, the edges appear in the order $e(l,1), e(l,2), \ldots, e(l,\kappa)$. Note that although our construction uses edges with zero latency, the problem remains NP-hard, even if such edges are not allowed; see Remark 3.2 after this proof.

First, we show that an optimal solution of the so-defined instance of the continuous network design problem P has total cost less or equal to $(4+\epsilon)\kappa + 2\kappa\nu$, if ϕ has a solution. To his end, let A be the set of literals that occur in a solution of ϕ. Then, a feasible solution of P is as follows. If, for a variable x_i, the solution A contains the *positive* literal x_i, we install capacity 1 for the corresponding *negative* literal edges $\{e(\bar{x}_i,k) : k \in K(\phi)\}$, and vice versa. Formally, we set

$$z_{(l,k)} = \begin{cases} 1 & \text{if } \neg l \in A, \\ 0 & \text{otherwise} \end{cases}$$

for all $l \in L(\phi)$, $k \in K(\phi)$. For each clause edge $e(k)$, $k \in K(\phi)$, we install capacity $2/\epsilon$. This particular capacity vector $\boldsymbol{z} = (z_e)_{e \in E}$ implies that each variable commodity $j(x_i)$ has a unique path of finite length, i.e., the path using the edges corresponding to the *negation* of the corresponding literal in A. Using that A is a solution of ϕ, we further obtain that for each clause commodity $j(k)$ at least one of the edges in $E_L(k)$ has capacity zero and, thus, infinite latency. This implies that the demand of each clause commodity $j(k)$ is routed along the corresponding clause edge $e(k)$. This Wardrop equilibrium is unique as each commodity has only a single path with finite latency.

For the total cost of this solution, we obtain

$$C(\boldsymbol{f},\boldsymbol{z}) = \sum_{e\in E_K}\left(\left(4+\frac{f_e}{z_e}\right)f_e + \left(\frac{\epsilon}{2}\right)^2 z_e\right) + \sum_{e\in E_L}\left(\frac{f_e^2}{z_e}+z_e\right) \tag{3.1}$$
$$= \sum_{e\in E_K}\left(\left(4+\frac{\epsilon}{2}\right)+\frac{\epsilon}{2}\right)+\frac{1}{2}\sum_{e\in E_L}(1+1)$$
$$= (4+\epsilon)\kappa + 2\kappa\nu.$$

Hence, an optimal solution has cost not larger than (3.1) if ϕ has a solution.

We proceed to prove that the total costs of an optimal solution are strictly larger than (3.1) if ϕ does *not* admit a solution. Let $\boldsymbol{z} = (z_e)_{e\in E}$ be the capacity vector of an optimal solution of P and let $\boldsymbol{f} = (f_e)_{e\in E}$ be a corresponding Wardrop flow. We distinguish two cases.

First case: $f_{e(k)} > 0$ for all $k \in K(\phi)$, i.e., each clause commodity $j(k)$ sends flow over the corresponding clause edge $e(k)$. Before we prove the thesis for this case, we need some additional notation. For the Wardrop flow f_e on edge $e \in E$, let f_e^V and f_e^K denote the flow on e that is due to the variable commodities and the clause commodities, respectively. We claim that there is a clause $\tilde{k} \in K(\phi)$, $\tilde{k} = l_{\tilde{k}} \vee l'_{\tilde{k}} \vee l''_{\tilde{k}}$ such that the flow of the variable commodities on each of the corresponding literal edges in $E(\tilde{k}) = \{e(l_{\tilde{k}},\tilde{k}), e(l'_{\tilde{k}},\tilde{k}), e(l''_{\tilde{k}},\tilde{k})\}$ is at least 1/2, i.e.,

$$f^V_{e(l_{\tilde{k}},\tilde{k})} \geq 1/2, \qquad f^V_{e(l'_{\tilde{k}},\tilde{k})} \geq 1/2, \qquad \text{and} \qquad f^V_{e(l''_{\tilde{k}},\tilde{k})} \geq 1/2. \tag{3.2}$$

For a contradiction, let us assume that for each clause $k = l_k \vee l'_k \vee l''_k$ there is a literal $l^*_k \in \{l_k, l'_k, l''_k\}$ such that $f^V_{e(l^*_k,k)} < 1/2$. As each variable commodity $j(x_i)$, $x_i \in V(\phi)$ splits its unit demand between the path consisting of the positive literal edges $\{e(x_i,k) : k \in K(\phi)\}$ and the path consisting of the negative literal edges $\{e(\bar{x}_i,k) : k \in K(\phi)\}$, at most one of these two paths is used with a flow strictly smaller than 1/2. Thus, the set of literals

$$A = \left\{l \in L(\phi) : f_e^V < 1/2 \text{ for all } e \in \{e(l,k) : k \in K(\phi)\}\right\}$$

contains for each variable at most one literal, i.e., $|A \cap \{x_i,\bar{x}_i\}| \leq 1$ for all $x_i \in V(\phi)$, and satisfies all clauses. This implies the existence of a solution $A^* \supseteq A$ of ϕ contradicting the assumption that no such solution exists. We conclude that there is a clause $\tilde{k}$ such that (3.2) holds.

We proceed to bound the total cost of a solution from below. As $\boldsymbol{f}$ is a Wardrop equilibrium, in which the clause commodity $j(\tilde{k})$ uses at least partially the clause edge $e(\tilde{k})$, we obtain

$$\sum_{e\in E(\tilde{k})}\frac{f_e}{z_e} \geq 4 + \frac{f_{e(\tilde{k})}}{z_{e(\tilde{k})}} > 4. \tag{3.3}$$

We bound the total cost of the solution $(\boldsymbol{f},\boldsymbol{z})$ by observing

$$C(\boldsymbol{f},\boldsymbol{z}) = \sum_{e\in E_L}\left(\frac{f_e^2}{z_e}+z_e\right) + \sum_{e\in E_K}\left(\left(4+\frac{f_e}{z_e}\right)f_e + \left(\frac{\epsilon}{2}\right)^2 z_e\right)$$
$$\geq \min_{\substack{x_e\geq 0, e\in E\\ \text{s.t. } \sum_{e\in E(\tilde{k})} f_e/x_e \geq 4}}\left\{\sum_{e\in E_L}\left(\frac{f_e^2}{x_e}+x_e\right)+\sum_{e\in E_K}\left(\left(4+\frac{f_e}{x_e}\right)f_e+\left(\frac{\epsilon}{2}\right)^2 x_e\right)\right\}.$$

Since the constraint $\sum_{e\in E(\tilde{k})} f_e/x_e \geq 4$ is independent of all variables x_e with $e \in E \setminus E(\tilde{k})$, we may optimize these variables independently and obtain

$$C(\boldsymbol{f},\boldsymbol{z}) \geq \sum_{e\in E_L\setminus E(\tilde{k})} \min\nolimits_{x_e\geq 0}\left\{\frac{f_e^2}{x_e}+x_e\right\} + \min\nolimits_{\substack{x_e\geq 0, e\in E(\tilde{k})\\ \text{s.t.} \sum_{e\in E(\tilde{k})} f_e/x_e\geq 4}} \left\{\sum_{e\in E(\tilde{k})}\left(\frac{f_e^2}{x_e}+x_e\right)\right\}$$
$$+ \sum_{e\in E_K} \min\nolimits_{x_e\geq 0}\left\{\left(4+\frac{f_e}{x_e}\right)f_e + \left(\frac{\epsilon}{2}\right)^2 x_e\right\}.$$

Calculating the respective minima for the edges in $E \setminus E(\tilde{k})$, we obtain

(3.4)
$$C(\boldsymbol{f},\boldsymbol{z}) \geq \sum_{e\in E_L\setminus E(\tilde{k})} 2f_e + \sum_{e\in E_K}(4+\epsilon)f_e + \min\nolimits_{\substack{x_e\geq 0, e\in E(\tilde{k})\\ \text{s.t.} \sum_{e\in E(\tilde{k})} f_e/x_e\geq 4}} \left\{\sum_{e\in E(\tilde{k})}\left(\frac{f_e^2}{x_e}+x_e\right)\right\}.$$

Each clause commodity $j(k)$ with $k \neq \tilde{k}$ can route its demand either over the clause edge $e(k)$ or over the three literal edges in $E_L(k)$. Every fraction of the demand routed over the clause edge contributes $4+\epsilon$ to the expression on the right-hand side of (3.4) while it contributes at least 6 when routed over the literal edges. Thus, the right-hand side of (3.4) is minimized when the clause commodities $j(k)$ with $k \neq \tilde{k}$ do not use the literal edges at all. We then obtain

$$C(\boldsymbol{f},\boldsymbol{z}) \geq (4+\epsilon)\left(\kappa - \frac{\sum_{e\in E(\tilde{k})} f_e^K}{3}\right) + 2\left(\kappa\nu - \sum_{e\in E(\tilde{k})} f_e^V\right)$$
$$+ \min\nolimits_{\substack{x_e\geq 0, e\in E(\tilde{k})\\ \text{s.t.} \sum_{e\in E(\tilde{k})} (f_e^K+f_e^V)/x_e\geq 4}} \left\{\sum_{e\in E(\tilde{k})}\left(\frac{(f_e^K+f_e^V)^2}{x_e}+x_e\right)\right\}.$$

Minimizing over f_e^V and f_e^K for $e \in E(\tilde{k})$ and using (3.2) and (3.3), we obtain as a lower bound

$$C(\boldsymbol{f},\boldsymbol{z}) \geq 2\kappa\nu + (4+\epsilon)\kappa + Q,$$

where Q is the solution to the constrained minimization problem

$$Q = \min\nolimits_{g_e,x_e,h} \sum_{e\in E(\tilde{k})}\left(\frac{(g_e+h)^2}{x_e}+x_e-2g_e\right) - (4+\epsilon)h$$
$$\text{s.t.} \quad \sum\nolimits_{e\in E(\tilde{k})} \frac{g_e+h}{x_e} \geq 4,$$
$$g_e \geq \frac{1}{2} \quad \text{for all } e\in E(\tilde{k}),$$

and the minimization is over $x_e \in \mathbb{R}_{>0}$, $g_e \in [0,1]$, $e \in E(\tilde{k})$, and $h \in [0,1]$.

It is straightforward to show that the optimal solution to the constraint optimization problem Q is equal to $Q = 1/8$ and is attained for $h = 0$, $g_e = 1/2$, and $x_e = 3/8$ for all $e \in E(\tilde{k})$; see Lemma B.1 in Appendix B for a formal proof. This implies that the total cost of a solution is not smaller than $2\kappa\nu + (4+\epsilon)\kappa + 1/8$, which concludes the first case.

Second case: There is a clause commodity $j(\tilde{k})$ that does not use its clause edge $e(\tilde{k})$, i.e., $f_{e(\tilde{k})} = 0$. As for the first case, we observe

$$\begin{aligned} C(\boldsymbol{f}, \boldsymbol{z}) &= \sum_{e \in E_L} \left(\frac{f_e^2}{z_e} + z_e \right) + \sum_{e \in E_K} \left(4 f_e + \frac{f_e^2}{z_e} + \left(\frac{\epsilon}{2}\right)^2 z_e \right) \\ &\geq \sum_{e \in E_L} 2 f_e + \sum_{e \in E_K} (4+\epsilon) f_e. \end{aligned}$$

Using that $j(\tilde{k})$ does not use its clause edge, we derive that the flow on the literal edges amounts to $\nu\kappa + 3$ and we obtain

$$\begin{aligned} C(\boldsymbol{f}, \boldsymbol{z}) &\geq 2(\kappa\nu + 3) + (4+\epsilon)(\kappa - 1) \\ &= 2\kappa\nu + (4+\epsilon)\kappa + 2 - \epsilon, \end{aligned}$$

which concludes the proof. □

In the following remark, we discuss that although the hardness proof of Theorem 3.1 used edges with zero latency, the hardness result continues to hold even if edges with zero latency are not allowed.

Remark 3.2. The continuous network design problem is NP-hard in the strong sense, even if no edges with zero latency are allowed.

Proof. Let M be an upper bound on the total cost of an optimal solution to a continuous network design problem constructed in the proof of Theorem 3.1 and let E_0 be the set of edges with zero latency. We replace each edge $e \in E_0$, $e = (s,t)$, $s,t \in V$ by an edge $e' = (s,t)$ with latency function $S_{e'}(f_{e'}/z_{e'}) = f_{e'}/z_{e'}$ and construction cost $\ell_{e'} = (\frac{\epsilon}{2M})^2$. For each new edge e', we introduce an additional commodity $i(e')$ with source $s_{i(e')} = s$, sink $t_{i(e')} = t$ and demand $d_{i(e')} = M$. To route the flow of commodity $i(e')$, each solution has to buy a sufficient capacity for the edge e'. For $z_{e'} = 2M^2/\epsilon$ the additional total cost on edge e' is ϵ. Thus, the routing cost and the total cost of the new edges can be made arbitrarily small. In conclusion, we can approximate the behavior of edges with zero latency within arbitrary precision by edges with unbounded latency functions. □

3.2. APX-hardness. With a similar construction as in the proof of Theorem 3.1, but a more detailed analysis, we can show that CNDP is in fact APX-hard. We reduce from a specific variant of MAX-3-SAT, which is NP-hard to approximate. The main technical difficulty remains to obtain good lower bounds on the optimal solution of CNDP. This turns out to be even more intricate for the APX hardness proof since we need essentially tight bounds on the costs of an optimal solution.

THEOREM 3.3. *The continuous network design problem (CNDP) on directed networks is APX-hard, even if all latency function are affine.*

Proof. An instance of 4-OCC-MAX-3-SAT is given by a Boolean formula ϕ in conjunctive normal form with the special property that each clause contains exactly three literals and each variable occurs exactly twice as a positive literal and exactly

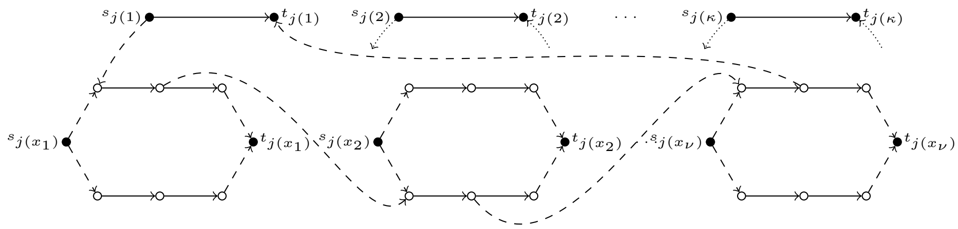

FIG. 2. *Network used to show the* APX*-hardness of the continuous network design problem. Clause* 1 *is equal to* $x_1 \vee \bar{x}_2 \vee x_\nu$. *Dashed edges have zero latency.*

twice as a negative literal. The problem to determine the maximal number of clauses that can be satisfied simultaneously is known to be NP-hard to approximate within a factor of $1016/1015 - \delta \approx 1.00099 - \delta$ for any $\delta > 0$; see Berman et al. [4, 5].

To show the APX-hardness of CNDP, we use the same construction and notation as in the proof of Theorem 3.1 but make use of the special property of ϕ that each literal occurs in exactly two clauses. Specifically, for each literal $l \in L(\phi)$, we only introduce two literal edges $e(l, k_l)$ and $e(l, k_l')$ for some $k_l, k_l' \in K(\phi)$. As a consequence each path of a variable commodity contains exactly two literal edges; see Figure 2.

We claim that the so-defined instance of CNDP has a solution with total cost in the interval $\left[7\kappa + \frac{1}{8}\tilde{\kappa}, (7+\epsilon)\kappa + (\frac{1}{4} + \frac{\epsilon}{2})\tilde{\kappa}\right]$ if and only if the minimum number of unsatisfied clauses is $\tilde{\kappa}$.

We proceed to prove that an optimal solution has total cost not larger than $(7+\epsilon)\kappa + (\frac{1}{4} + \frac{\epsilon}{2})\tilde{\kappa}$ if ϕ has a solution A that violates $\tilde{\kappa}$ clauses only. The proof of this thesis is similar to the one in Theorem 3.1, with the additional complication that we have to take special care of the nonsatisfied clauses. Let $A \subset L(\phi)$ with $|A \cap \{x_i, \bar{x}_i\}| = 1$ for all $x_i \in V(\phi)$ be an assignment of the variables of ϕ that satisfies exactly $\kappa - \tilde{\kappa}$ clauses and let $\tilde{K} \subseteq K(\phi)$ with $\tilde{\kappa} = |\tilde{K}|$ be the set of clauses that are not satisfied by A. Consider the tuple $(\boldsymbol{f}, \boldsymbol{z})$ defined as

$$z_{e(l,k)} = \begin{cases} 1 & \text{if } k \notin \tilde{K}, \neg l \in A, \\ \left(\dfrac{4}{3} + \dfrac{\epsilon}{6}\right)^{-1} & \text{if } k \in \tilde{K}, \neg l \in A, \\ 0 & \text{otherwise,} \end{cases} \qquad f_{e(l,k)} = \begin{cases} 1 & \text{if } \neg l \in A, \\ 0 & \text{otherwise,} \end{cases}$$

for all $l \in L(\phi), k \in \{k_l, k_l'\}$ and

$$z_{e(k)} = \frac{2}{\epsilon}, \qquad f_{e(k)} = 1,$$

for all $k \in K(\phi)$.

To show that the tuple $(\boldsymbol{f}, \boldsymbol{z})$ is a solution to CNDP, it suffices to prove that $\boldsymbol{f}$ is a Wardrop equilibrium for the latency functions defined by $\boldsymbol{z}$. For every variable commodity $j(x_i)$, $x_i \in V(\phi)$ and every clause commodity $j(k)$ that corresponds to a satisfied clause $k \in K(\phi) \setminus \tilde{K}$, it is straightforward to check that there is only one path with finite latency. Next, consider an arbitrary clause commodity $j(k)$ that corresponds to a nonsatisfied clause $k \in \tilde{K}$, $k = l_k \vee l_k' \vee l_k''$. Such a clause uses the clause edge $e(k)$ with latency $4 + \epsilon/2$. On the other hand, the corresponding literal edges $e(l_k, k)$, $e(l_k', k)$, and $e(l_k'', k)$ all have capacity $\left(\frac{4}{3} + \frac{\epsilon}{6}\right)^{-1}$ and carry one unit of flow of the variable commodities. Thus, their latencies sum up to $4 + \frac{\epsilon}{2}$, implying that clause commodity $j(k)$ is in equilibrium.

We proceed to calculate the total cost of the solution $(\boldsymbol{f}, \boldsymbol{z})$. Every literal edge that corresponds to a satisfied clause and the negation of a literal in A has capacity 1 and flow 1 and thus causes a total cost of 2. In contrast to this, each literal edge that corresponds to a violated clause and the negation of a literal in A has capacity $\left(\frac{4}{3} + \frac{\epsilon}{6}\right)^{-1}$ and flow 1 and, thus, causes a total cost of

$$\frac{4}{3} + \frac{\epsilon}{6} + \frac{1}{\frac{4}{3} + \frac{\epsilon}{6}} \leq \frac{25}{12} + \frac{\epsilon}{6}.$$

Further, each clause edge has capacity $2/\epsilon$ and is used by 1 unit of flow and, thus, contributes $4 + \epsilon$ to the total cost. We calculate

$$C(\boldsymbol{f}, \boldsymbol{z}) \leq (4 + \epsilon)\kappa + \sum_{e \in E_L} 2 f_e + \sum_{k \in \tilde{K}} \sum_{e \in E_L(k)} \left(\frac{25}{12} + \frac{\epsilon}{6} - 2\right) f_e.$$

Using $\sum_{e \in E_L} f_e = \frac{3}{2}\kappa$, we obtain

$$\begin{aligned} C(\boldsymbol{f}, \boldsymbol{z}) &\leq (4 + \epsilon)\kappa + 3\kappa + 3\tilde{\kappa}\left(\frac{1}{12} + \frac{\epsilon}{6}\right) \\ &= (7 + \epsilon)\kappa + \left(\frac{1}{4} + \frac{\epsilon}{2}\right)\tilde{\kappa}. \end{aligned}$$

We proceed to prove that an optimal solution $(\boldsymbol{f}, \boldsymbol{z})$ of CNDP has total cost not smaller than $7\kappa + \frac{1}{4}\tilde{\kappa}$ if each solution A of ϕ violates at least $\tilde{\kappa}$ clauses. For an edge flow $\boldsymbol{f}$, let $\boldsymbol{f}^V$ and $\boldsymbol{f}^K$ denote the edge flow that is due to the variable commodities and clause commodities, respectively. Using that for all edges e we have $f_e = 0$ if and only if $z_e = 0$, we bound $C(\boldsymbol{f}, \boldsymbol{z})$ by

$$\begin{aligned} C(\boldsymbol{f}, \boldsymbol{z}) &= \sum_{\substack{e \in E_L \\ f_e > 0}} \left(\frac{f_e^2}{z_e} + z_e\right) + \sum_{\substack{e \in E_K \\ f_e > 0}} \left(\left(4 + \frac{f_e}{z_e}\right) f_e + \left(\frac{\epsilon}{2}\right)^2 z_e\right). \\ &\geq \sum_{\substack{e \in E_L \\ f_e > 0}} \left(\frac{f_e^2}{z_e} + z_e\right) + 4 \sum_{\substack{e \in E_K \\ f_e > 0}} f_e. \end{aligned}$$

Each clause commodity has unit demand and uses either three literal edges in E_L or one clause edge in E_K. This implies $\sum_{e \in E_K} f_e^K + \sum_{e \in E_L} f_e^K/3 = \sum_{e \in E_K} f_e + \sum_{e \in E_L} f_e^K/3 = \kappa$. We then obtain

$$\begin{aligned} C(\boldsymbol{f}, \boldsymbol{z}) &\geq \sum_{\substack{e \in E_L \\ f_e > 0}} \left(\frac{(f_e^V + f_e^K)^2}{z_e} + z_e\right) + 4\left(\kappa - \frac{1}{3}\sum_{\substack{e \in E_L \\ f_e > 0}} f_e^K\right) \\ &= 4\kappa + \sum_{\substack{e \in E_L \\ f_e > 0}} \left(\frac{(f_e^V + f_e^K)^2}{z_e} + z_e - \frac{4}{3} f_e^K\right). \end{aligned} \tag{3.5}$$

Since $\sum_{e \in E_L} f_e^V = \frac{3}{2}\kappa$, we obtain

$$C(\boldsymbol{f}, \boldsymbol{z}) \geq 7\kappa + \sum_{\substack{e \in E_L \\ f_e > 0}} \left(\frac{(f_e^V + f_e^K)^2}{z_e} + z_e - 2 f_e^V - \frac{4}{3} f_e^K\right)$$

$$= 7\kappa + \sum_{k \in K(\phi)} \sum_{\substack{e \in E_L(k) \\ f_e > 0}} \left(\frac{(f_e^V + f_e^K)^2}{z_e} + z_e - 2 f_e^V - \frac{4}{3} f_e^K\right). \tag{3.6}$$

To obtain a lower bound on $C(\boldsymbol{f}, \boldsymbol{z})$ we bound for each clause $k \in K(\phi)$ the term $\sum_{e \in E_L(k): f_e > 0} \big((f_e^V + f_e^K)^2/z_e + z_e - 2f_e^V - 4f_e^K/3\big)$ by the solution of a constrained minimization problem for which the original solution $(\boldsymbol{f}, \boldsymbol{z})$ is feasible and has the same function value. To formulate the constraints of the minimization problem, for a clause $k \in K(\phi)$ we denote by $m_k(\boldsymbol{f}^V) = \min_{e \in E_L(k)} f_e^V$ the minimum flow of the variable commodities over the edges in $E_L(k)$. Consider the constrained minimzation problem

$$Q_k = \inf_{g_e, x_e, h} \sum_{\substack{e \in E_L(k)\\ f_e > 0}} \left(\frac{(g_e + h)^2}{x_e} + x_e - 2g_e - \frac{4}{3}h \right) \tag{3.7a}$$

$$\text{s.t.} \sum_{e \in E_L(k)} \frac{g_e + h}{x_e} \geq 4 \qquad \text{if } h < 1 \text{ and } f_e > 0 \text{ for all } e \in E_L(k), \tag{3.7b}$$

$$g_e \geq m_k(\boldsymbol{f}^V) \qquad \text{for all } e \in E_L(k) : f_e > 0, \tag{3.7c}$$

$$x_e > 0 \qquad \text{for all } e \in E_L(k) : f_e > 0, \tag{3.7d}$$

$$g_e + h > 0 \qquad \text{for all } e \in E_L(k) : f_e > 0, \tag{3.7e}$$

$$g_e \in [0, 1] \qquad \text{for all } e \in E_L(k) : f_e > 0, \tag{3.7f}$$

$$h \in [0, 1]. \tag{3.7g}$$

We claim that Q_k is a lower bound on the contribution of every clause k to (3.6), i.e.,

$$Q_k \leq \sum_{e \in E_L(k): f_e > 0} \left(\left(f_e^V + f_e^K\right)^2 / z_e + z_e - 2f_e^V - 4f_e^K/3 \right).$$

To see this claim, note that $f_e^K = f_{e'}^K$ for all $e, e' \in E_L(k)$ and substitute $f_e^V = g_e$, $f_e^K = h$, and $z_e = x_e$ for all $e \in E_L(k)$. Then, the objective (3.7a) of Q_k is equal to the contribution of clause k to (3.6). In addition, $(\boldsymbol{f}, \boldsymbol{z})$ clearly satisfies the constraints (3.7c) to (3.7g). Moreover, since $\boldsymbol{f}$ is a Wardrop equilibrium for the vector of capacities $\boldsymbol{z}$, the sum of the latencies of the edges in $E_L(k)$ is at least 4 in the case that a nonzero fraction of the flow of the clause commodity $j(k)$ is routed over the corresponding clause edge. With the substitution $f_e^K = h$ this corresponds to $h < 1$. The equilibrium constraint is trivially satisfied if $f_e = 0$ for one of the edges in $E_L(k)$ since in this case the optimality of $(\boldsymbol{f}, \boldsymbol{z})$ also implies that $z_e = 0$ so that the latency of that edge is infinite. This fact is expressed in (3.7b). We conclude that $(\boldsymbol{f}, \boldsymbol{z})$ satisfies all conditions of Q_k, $k \in K(\phi)$ so that $7\kappa + \sum_{k \in K(\phi)} Q_k$ gives a lower bound on $C(\boldsymbol{f}, \boldsymbol{z})$.

Next, we discuss the optimal solutions of (3.7). If $f_e > 0$ for all $e \in E_L(k)$, the optimal solution to this system (3.7) is equal to $Q_k = m_k(\boldsymbol{f}^V)/4$ and is attained for $f_e^V = m_k(\boldsymbol{f}^V)$ and $z_e = 3m_k(\boldsymbol{f}^V)/4$ for all $e \in E_k$, $k \in K(\phi)$; see Lemma B.1 in Appendix B. If, on the other hand, $f_e = 0$ for some $e \in E_L(k)$ we have $m_k(\boldsymbol{f}^V) = 0$ and constraint (3.7b) vanishes. We obtain

$$Q_k = \inf_{\substack{g_e \in [0,1]\\ x_e \geq 0\\ h\ \in [0,1]}} \sum_{\substack{e \in E_L(k)\\ f_e > 0}} \left(\frac{(g_e + h)^2}{x_e} + x_e - 2g_e - \frac{4}{3}h \right),$$

where with some abuse of notation we let $(g_e + h)^2/x_e = 0$ if $g_e + h = x_e = 0$ and $(g_e + h)^2/x_e = \infty$ if $g_e + h > 0$ and $x_e = 0$. First-order conditions on x_e imply that Q_k is minimized when $x_e = g_e + h$. Thus,

$$Q_k = \inf_{\substack{g_e \in [0,1] \\ h \in [0,1]}} \sum_{\substack{e \in E_L(k) \\ f_e > 0}} \left(2(g_e + h) - 2g_e - \frac{4}{3}h \right)$$
$$= \min_{h \in [0,1]} \sum_{\substack{e \in E_L(k) \\ f_e > 0}} \left(\frac{2}{3}h \right) = 0 = m_k(\boldsymbol{f}^V)/4.$$

Overall, we have shown that Q_k is at least $m_k(\boldsymbol{f}^V)/4$. This implies

$$C(\boldsymbol{f}, \boldsymbol{z}) \geq 7\kappa + \sum_{k \in K(\phi)} \frac{m_k(\boldsymbol{f}^V)}{4}.$$

To finish the proof it suffices to show that $\sum_{k \in K(\phi)} m_k(\boldsymbol{f}^V) \geq \tilde{\kappa}$ for each feasible flow of the variable commodities $\boldsymbol{f}^V$. To this end, let $\boldsymbol{f}^V$ be a flow minimizing $\sum_{k \in K(\phi)} m_k(\boldsymbol{f}^V)$. We claim that it is without loss of generality to assume that $\boldsymbol{f}^V$ is integral. To see this claim, suppose that the flow for all variable commodities except $j(x_i)$ is fixed and consider the variable commodity $j(x_i)$. Let p denote the fraction of the flow sent over the path consisting of the positive literal edges $e(x_i, k_{x_i})$ and $e(x_i, k'_{x_i})$. By definition, only the clauses k_{x_i} and k'_{x_i} contain the literal x_i and only the clauses $k_{\bar{x}_i}$ and $k'_{\bar{x}_i}$ contain the literal $\bar{x}_i$. We calculate the contribution of these four clauses to $\sum_{k \in K(\phi)} m_k(\boldsymbol{f}^v)$ as follows:

$$\sum\nolimits_{k \in \{k_{x_i}, k'_{x_i}, k_{\bar{x}_i}, k'_{\bar{x}_i}\}} m_k(\boldsymbol{f}^V) = \sum\nolimits_{k \in \{k_{x_i}, k'_{x_i}\}} \min\Big\{p, \min_{e \in E_L(k) \setminus e(x_i,k)} \big\{f_e^V\big\}\Big\}$$
$$+ \sum\nolimits_{k \in \{k_{\bar{x}_i}, k'_{\bar{x}_i}\}} \min\Big\{1-p, \min_{e \in E_L(k) \setminus e(\bar{x}_i,k)} \big\{f_e^V\big\}\Big\}.$$

For a fixed flow $\boldsymbol{f}^V$ on the literal edges not involving variable x_i, being the sum of concave functions, this expression is concave in p. Hence, the minimum is attained for either $p = 0$ or $p = 1$. Put differently, for any flow of the other variable commodities, the expression $\sum_{k \in K(\phi)} m_k(\boldsymbol{f}^V)$ is minimized when variable commodity $j(x_i)$ routes all of its demand on one path. Iterating this argument for all variable commodities, we conclude that it is without loss of generality to assume that $\boldsymbol{f}^V$ is integral.

For an integral flow $\boldsymbol{f}^V$ of the variable commodities, consider the assignment $A = \{l \in L(\phi) : f_{e(l,k_l)} = 0\}$. This assignment satisfies at most $\kappa - \tilde{\kappa}$ clauses. For any clause k satisfied by A we obtain $m_k(\boldsymbol{f}^V) = 0$. As the flow is integral, any clause k not satisfied by A has $m_k(\boldsymbol{f}^V) = 1$. We conclude that $\sum_{k \in K(\phi)} m_k(\boldsymbol{f}^V) \geq \tilde{\kappa}$ for all feasible flows of the variable commodities $\boldsymbol{f}^V$.

Plugging everything together, we obtain that the total cost of an optimal solution to CNDP lies in the range

$$\left[7\kappa + \frac{1}{4}\tilde{\kappa},\, (7+\epsilon)\kappa + \left(\frac{1}{4} + \frac{\epsilon}{2}\right)\tilde{\kappa}\right], \tag{3.8}$$

if $\kappa - \tilde{\kappa}$ is the maximal number of satisfiable clauses.

Berman et al. [4, 5] construct a family of symmetric instances of 4-Occ-Max-3-Sat with $\kappa = 1016n$, $n \in \mathbb{N}$ that has the property that for any $\delta \in (0, 1/2)$ it is NP-hard to distinguish between the systems where $(1016 - \delta)n$ clauses can be satisfied and systems where at most $(1015 + \delta)n$ clauses can be satisfied. Using (3.8), the corresponding instances of CNDP have the property that they have total cost at most $(7+\epsilon)1016n + \delta n(\frac{1}{4}+\frac{\epsilon}{2})$, if at least $(1016-\delta)n$ clauses can be satisfied, and total cost at least $7 \cdot 1016n + \frac{1}{4}(1-\delta)n$, if at most $(1015+\delta)n$ clauses can be satisfied. As we let ϵ and δ go to zero, we derive that it is NP-hard to approximate CNDP by any factor better than $7112.25/7112 \approx 1.000035$. This proves the APX-hardness of the problem. ∎

3.3. Hardness for undirected networks. With a similar construction, we can also show APX-hardness for CNDP on undirected networks.

THEOREM 3.4. *The continuous network design problem on undirected networks is APX-hard, even if all latency functions are affine.*

Proof. We closely mimic the proof of Theorem 3.3 and only sketch how to adjust it to the undirected case; see Figure 3. Compared to the directed case, we introduce some auxiliary edges with high latency in order to prevent the commodities from taking undesired paths.

Type one edges are the edges adjacent to $s_{j(x_i)}$ or $t_{j(x_i)}$ for $x_i \in V(\phi)$; their latency is set to $S_e(f_e/z_e) = 50$.

Type two edges connect $e(l, k_l)$ and $e(l, k'_l)$ for $l \in L(\phi)$; their latency is set to $S_e(f_e/z_e) = 100$.

Type three edges connect $s_{j(k)}$ or $t_{j(k)}$ for $k \in K(\phi)$ to a variable gadget; their latency is set to $S_e(f_e/z_e) = 0$.

Type four edges are the edges between two literal edges that correspond to the same clause, but different variables; their latency is set to $S_e(f_e/z_e) = 20$.

Type five edges are the additional edges that connect $s_{j(k)}$ with the clause edge $e(k)$ for $k \in K(\phi)$; their latency is set to $S_e(f_e/z_e) = 40$.

The theorem is proven showing that the total cost of an optimal solution to CNDP lies in the range $\left[197\kappa + \frac{1}{4}\tilde{\kappa},\ (197+\epsilon)\kappa + \left(\frac{1}{4}+\frac{\epsilon}{2}\right)\tilde{\kappa}\right]$ if $\kappa - \tilde{\kappa}$ is the maximum number of satisfied clauses.

To prove the upper bound on the cost, we proceed as in the proof of Theorem 3.3. We fix an assignment A of the variables that satisfies $\kappa - \tilde{\kappa}$ clauses and route all clause commodities along the clause edges and all variable edges along the negation of the assignment of the variable in A. It is left to argue that the so-defined flow is a Wardrop equilibrium. Since the auxiliary edges have nonzero latency, compared to the solution in the directed case, the latency cost of each clause commodity increased by the latency of a type-five edge, i.e., 40, and the latency cost of each variable commodity increased by twice the latency of a type-one edge plus the latency of a

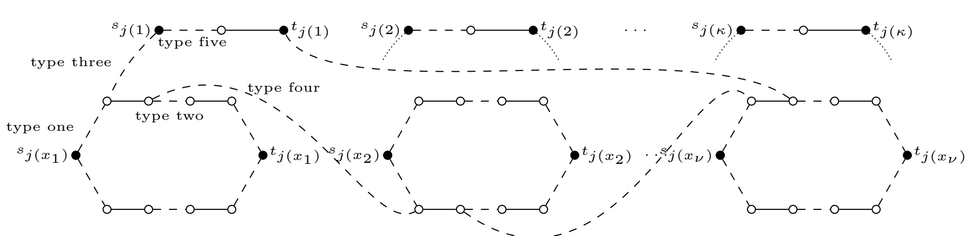

FIG. 3. *Network used to show the APX-hardness of the continuous network design problem on undirected graphs. Clause 1 is equal to $x_1 \vee \bar{x}_2 \vee x_\nu$. The clause edges (straight edges in the upper part of the graph) and the literal edges (straight edges in the lower part of the graph) are connected via different auxiliary edges (dashed). The auxiliary edges have different constant latencies dependent on their type as specified in the proof of Theorem 3.4.*

type-two edge, i.e., 200. Thus, the total cost increased by $40\kappa + \frac{3}{4} \cdot 200\kappa = 190\kappa$ giving a total cost of $(197 + \epsilon)\kappa + (1/4 + \epsilon/2)\tilde{\kappa}$. It is left to argue that this solution still constitutes a Wardrop equilibrium although all edges can now be used in both directions. To this end, note that each clause commodity uses its clause edge and experiences a total latency of $44 + \epsilon/2$, and thus has no incentive to use a path with a type-one or type-two edge. The only alternative path uses two type-three edges, two type-four edges, and the three corresponding literal edges where the latencies of the latter are summing up to $4 + \epsilon/2$, as in the proof in Theorem 3.3. We conclude that the clause commodities are in equilibrium.

For a variable commodity $j(x_i)$ we claim that the sum of the latencies of all auxiliary edges in a path is at least 200. To see this, note that each such path contains at least two type-one edges with latency 50 each. Further, for each literal $l \in \{x_i, \bar{x}_i\}$, the type-one edges adjacent to $s_{j(x_i)}$ and $t_{j(x_i)}$ are connected to literal edges $e(l, k_l)$ and $e(l', k_l')$ that correspond to different clauses. The distance between two literal edges $e(l, k_l)$ and $e(l', k_{l'})$ that correspond to different clauses is at least 100 because at least one type-two and two consecutive type-one edges have to be crossed. Using this observation, one can show that no variable commodity has an incentive to deviate.

For the lower bound, we argue as follows. If no variable commodity uses a type-three edge or a type-four edge, then each variable commodity has to split its flow between the path corresponding to the positive and the negative literal, respectively, and the lower bound can be proven analogously to the directed case.

So we are left with the cases that a variable commodity uses a type-three edge or a type-four edge. Let us first assume that we have an optimal solution, in which a variable commodity uses a type-four edge. We may assume without loss of generality that every literal edge that carries flow has a latency of at most 5, because we could decrease the total cost by increasing the capacity on these edges, otherwise. (However, we may not decrease the latency below $4 + \epsilon/2$ because this might give an incentive to the clause commodities to use these edges as well.) Every path available to a variable commodity uses at least two type-one edges as these edges are adjacent to the source and target of each variable commodity. As argued above, every feasible path uses at least either two additional type-one edges or one type-two edge. Using that the variable commodity also uses a type-four edge, this implies that the latency of the variable commodity is at least $200 + 20$. However, it would also be feasible to route that variable commodity along the path corresponding to the positive literal, say, while installing an additional capacity of $1/5$ on the two literal edges of the positive literal, resulting in a total cost of $200 + 10 + 2/5 < 220$. This low capacity would not prevent any of the clause commodities from using their clause edge and has a lower total cost. Thus, we may conclude that no variable commodity use a type-four edge. As any path of a variable commodity that uses a type-three edge also uses a type-five edge with latency 40, we may conclude that no variable commodity uses such an edge as well. ∎

For all our hardness results, we use instances with different source and sink vertices per commodity. In contrast, CNDP can be solved efficiently for instances in which commodities share a common sink (or source); see Corollary 2.5.

4. Approximation. Given the APX-hardness of the problem, we study the approximation of CNDP. We first provide a detailed analysis of the approximation guarantees of two different approximation algorithms. Then, as the arguably most interesting result of this section, we provide an improved approximation guarantee for taking the better of the two algorithms. The approximation guarantees proven in this section depend on the set $\mathcal{S}$ of allowable cost functions and are in fact closely related to the *anarchy value* value $\alpha(\mathcal{S})$ introduced by Roughgarden [26] and Correa et al. [9]. Intuitively, the anarchy value of a set of latency functions $\mathcal{S}$ is the worst case ratio between the routing cost of a Wardrop equilibrium and that of a system optimum of an instance in which all latency functions are contained in $\mathcal{S}$. Roughgarden [26] and Correa et al. [9] show that $\alpha(\mathcal{S}) = \frac{1}{1-\mu(\mathcal{S})}$, where

$$\mu(\mathcal{S}) = \sup_{S\in\mathcal{S}} \sup_{x\geq 0: S(x)>0} \max_{\gamma\in[0,1]} \left\{ \gamma \left(1 - \frac{S(\gamma x)}{S(x)} \right) \right\}. \tag{4.1}$$

To simplify notation, we assume throughout this section that all latency functions are strictly positive. This allows us to drop the condition that $S(x) > 0$ in (4.1). The following lemma gives an alternative representation of $\mu(S)$.

LEMMA 4.1. *For a latency function S, the following values coincide:*

1. $\sup_{x\geq 0} \max_{\gamma\in[0,1]} \{\gamma(1 - \frac{S(\gamma x)}{S(x)})\}$,
2. $\sup_{x\geq 0} \{\gamma \frac{S'(x)x}{S(x)+S'(x)x} : \gamma \in [0,1]$ *with* $S(x/\gamma) = S(x) + xS'(x)\}$.

Proof. For any cost function S, the expression $\sup_{x\geq 0} \max_{\gamma\in[0,1]} \{\gamma(1-\frac{S(\gamma x)}{S(x)})\}$ is non-negative and strictly positive for $\gamma \in (0,1)$ since S is strictly increasing. Moreover, the function $\gamma(1 - \frac{S(\gamma x)}{S(x)})$ is continuous in γ, thus, by the extreme value theorem of Weierstrass a maximum will be attained on the compact interval $[0,1]$. We conclude that the inner maximum is attained for some $\gamma \in (0,1)$. Hence, γ satisfies the first-order optimality conditions

$$0 = \left(1 - \frac{S(\gamma x)}{S(x)}\right) - \gamma x \cdot \frac{S'(\gamma x)}{S(x)} \qquad \Leftrightarrow \qquad S(x) = S(\gamma x) + \gamma x\, S'(\gamma x)$$

By substituting $y = \gamma x$, we obtain

$$\begin{aligned}
&\sup_{x\geq 0} \max_{\gamma\in[0,1]} \left\{ \gamma \left(1 - \frac{S(\gamma x)}{S(x)} \right) \right\} \\
&\quad = \sup_{y\geq 0} \left\{ \gamma \left(1 - \frac{S(y)}{S(y/\gamma)} \right) : \gamma \in [0,1] \text{ with } S(y/\gamma) = S(y) + yS'(y) \right\} \\
&\quad = \sup_{y\geq 0} \left\{ \gamma \cdot \frac{S'(y)\,y}{S(y) + S'(y)\,y} : \gamma \in [0,1] \text{ with } S(y/\gamma) = S(y) + yS'(y) \right\},
\end{aligned}$$

which proves the lemma. ∎

4.1. Two approximation algorithms. In this section, we present two approximation algorithms for CNDP. The first algorithm that we call BRINGTOEQUILIBRIUM (cf. Algorithm 1) was already proposed by Marcotte [23, section 4.3] and analyzed for monomial latency functions. Our contribution is a more general analysis of BRINGTOEQUILIBRIUM that works for arbitrary sets of latency functions $\mathcal{S}$, requiring only Assumption 2.1. The second algorithm, that we call SCALEUNIFORMLY (cf. Algorithm 2), is a new algorithm that we introduce in this paper.

For both approximation algorithms, we first compute an optimum solution $(\boldsymbol{f}^*, \boldsymbol{z}^*)$ to a relaxation of CNDP without the equilibrium constraints; i.e., we compute a solution $(\boldsymbol{f}^*, \boldsymbol{z}^*)$ to the problem $\min_{\boldsymbol{z}\geq 0} \min_{\boldsymbol{f}\in\mathcal{F}} \sum_{e\in E} \big(S_e(f_e/z_e)\ f_e + z_e\ \ell_e\big)$ which can be

Algorithm 1. BringToEquilibrium.

Input: Graph $G = (V, E)$ with latency functions S_e and construction costs ℓ_e, $e \in E$.
Output: Capacity vector $\boldsymbol{z}$, corresponding Wardrop equilibrium $\boldsymbol{f}$.

$(\boldsymbol{f}^*, \boldsymbol{z}^*) \leftarrow$ solution to (CNDP′)
for all $e \in E$ **do**
  $x_e \leftarrow f_e^*/z_e^*$
  $\gamma_e \leftarrow$ solution to $S_e(x_e) + S_e'(x_e)x_e = S_e(x_e/\gamma_e)$
  $z_e \leftarrow \gamma_e z_e^*$
end for
return $(\boldsymbol{f}^*, \boldsymbol{z})$

Algorithm 2. ScaleUniformly.

Input: Graph $G = (V, E)$ with latency functions S_e and construction costs ℓ_e, $e \in E$.
Output: Capacity vector $\boldsymbol{z}$, corresponding Wardrop equilibrium $\boldsymbol{f}$.

1: $(\boldsymbol{f}^*, \boldsymbol{z}^*) \leftarrow$ solution to (CNDP′)
2: $p \leftarrow C^R(\boldsymbol{f}^*, \boldsymbol{z}^*)/C(\boldsymbol{f}^*, \boldsymbol{z}^*)$
3: $\lambda \leftarrow \mu(\mathcal{S}) + \sqrt{\mu(\mathcal{S})\frac{p}{1-p}}$
4: $\boldsymbol{f} \leftarrow$ Wardrop equilibrium with respect to $\lambda \boldsymbol{z}^*$
5: **return** $(\boldsymbol{f}, \lambda \boldsymbol{z}^*)$

done in polynomial time (Proposition 2.4). Then, in both algorithms, we reduce the capacity vector $\boldsymbol{z}^*$, and determine a Wardrop equilibrium for the new capacity vector. The algorithms differ in the way we adjust the capacity vector $\boldsymbol{z}^*$. While in BringToEquilibrium we reduce the edge capacities individually such that the optimum solution to the relaxation (CNDP′) is a Wardrop equilibrium, in ScaleUniformly, we scale all capacities uniformly by a factor λ (cf. lines 2–3) and compute a Wardrop equilibrium for the scaled capacities.

We first show that the approximation guarantee of BringToEquilibrium is at most $1+\mu(\mathcal{S})$. For the proof of this result we use the first-order optimality conditions for the vector of capacities $\boldsymbol{f}^*$ obtained as a solution to the relaxed problem (CNDP′) in combination with the variational inequalities technique used in the price of anarchy literature (e.g., Roughgarden [26] and Correa et al. [9]).

Theorem 4.2. *The approximation guarantee of* BringToEquilibrium *is at most* $1 + \mu(\mathcal{S})$.

Proof. Let $(\boldsymbol{f}^*, \boldsymbol{z}^*)$ be the relaxed solution computed in the first step of BringToEquilibrium. By the necessary Karush-Kuhn-Tucker optimality conditions, $(\boldsymbol{f}^*, \boldsymbol{z}^*)$ satisfies

$$\ell_e = S_e'(f_e^*/z_e^*)(f_e^*/z_e^*)^2 \qquad \text{for all } e \in E \text{ with } z_e^* > 0. \tag{4.2}$$

Eliminating ℓ_e in the statement of the relaxed problem (CNDP′) we obtain the following expression for the total cost of the relaxation:

$$C(\boldsymbol{f}^*, \boldsymbol{z}^*) = \sum_{e \in E} \Big(S_e(f_e^*/z_e^*) + S_e'(f_e^*/z_e^*)(f_e^*/z_e^*)\Big) f_e^*. \tag{4.3}$$

For each $e \in E$ let $x_e = f_e^*/z_e^*$ if $z_e^* > 0$, and $x_e = 0$, otherwise. We define a new vector of capacities $\boldsymbol{z}$ by $z_e = \gamma_e \cdot z_e^*, e \in E$, where $\gamma_e \in [0,1]$ is a solution to the equation

$$S_e(x_e) + S_e'(x_e)\,x_e = S_e(x_e/\gamma_e). \tag{4.4}$$

By Proposition 2.2, we are interested in bounding $C(\boldsymbol{f}^*, \boldsymbol{z})$. To this end, we calculate

$$\begin{aligned} C(\boldsymbol{f}^*, \boldsymbol{z}) &= \sum_{e\in E} \left(S_e(x_e/\gamma_e) f_e^* + \ell_e\, z_e\right) \\ &\overset{(4.4)}{=} \sum_{e\in E} \Big(\big(S_e(x_e) + S_e'(x_e)\,x_e\big) f_e^* + \gamma_e\, \ell_e\, z_e^*\Big) \\ &\overset{(4.2)}{=} \sum_{e\in E} \Big(\big(S_e(x_e) + S_e'(x_e)\,x_e\big) f_e^* + \gamma_e\, S_e'(x_e)\,x_e\, f_e^*\Big). \end{aligned} \tag{4.5}$$

By (4.1), (4.4), and Lemma 4.1, we have $\gamma_e\, S_e'(x_e)\,x_e \le \mu(\mathcal{S})\,(S_e(x_e) + S_e'(x_e)\,x_e)$. Combining this inequality with (4.5) gives

$$\begin{aligned} C(\boldsymbol{f}^*, \boldsymbol{z}) &\le (1+\mu(\mathcal{S})) \sum_{e\in E} \Big(\big(S_e(x_e) + S_e'(x_e)\,x_e\big) f_e^* \\ &\overset{(4.3)}{=} (1+\mu(\mathcal{S}))\, C(\boldsymbol{f}^*, \boldsymbol{z}^*), \end{aligned}$$

which proves the claim. ∎

We proceed by showing that ScaleUniformly achieves the same approximation guarantee of $1+\mu(\mathcal{S})$. Recall that ScaleUniformly first computes a relaxed solution $(\boldsymbol{f}^*, \boldsymbol{z}^*)$. Then, this relaxed solution is used to compute an optimal scaling factor $\lambda \le 1$ with which all capacities are scaled subsequently. The algorithm then returns the scaled capacity vector $\lambda \boldsymbol{z}^*$ together with a corresponding Wardrop equilibrium $\boldsymbol{f} \in \mathcal{W}(\lambda \boldsymbol{z}^*)$.

It is interesting to note that a (worse) approximation guarantee of 2 can be inferred directly from a bicriteria result of Roughgarden and Tardos [28], who showed that for any instance the routing cost of a Wardrop equilibrium is not worse than a system optimum that ships twice as much flow. This implies that for $\lambda = 1/2$ we have $C(\boldsymbol{f}, \lambda \boldsymbol{z}^*) \le 2C(\boldsymbol{f}^*, \boldsymbol{z}^*)$, as claimed.

For the proof of the following result, we take a different road that allows us to express the approximation guarantee of ScaleUniformly as a function of the parameter p defined as the fraction of the total cost $C(\boldsymbol{f}^*, \boldsymbol{z}^*)$ of the relaxed solution allotted to the routing costs $C^R(\boldsymbol{f}^*, \boldsymbol{z}^*)$. This is an important ingredient for the analysis of the best-of-two algorithm.

Theorem 4.3. *The approximation guarantee of* ScaleUniformly *is at most* $1+\mu(\mathcal{S})$.

Proof. The algorithm first computes an optimum solution $(\boldsymbol{f}^*, \boldsymbol{z}^*)$ of the relaxed problem (CNDP′). Then $p \in [0,1]$ is defined as the fraction of $C(\boldsymbol{f}^*, \boldsymbol{z}^*)$ that corresponds to the routing cost $C^R(\boldsymbol{f}^*, \boldsymbol{z}^*)$, i.e.,

$$C^R(\boldsymbol{f}^*, \boldsymbol{z}^*) = \sum_{e\in E} S_e(f_e^*/z_e^*)\, f_e^* = p\, C(\boldsymbol{f}^*, \boldsymbol{z}^*),$$

Now, we define $\lambda = \mu(\mathcal{S}) + \sqrt{\mu(\mathcal{S})\frac{p}{1-p}}$ and consider the capacity vector $\lambda \boldsymbol{z}^*$, in which the capacities of the optimal solution to the relaxation are scaled uniformly by λ. Finally, we compute a Wardrop equilibrium with respect to capacities $\lambda \boldsymbol{z}^*$. Let $\boldsymbol{f}$ the corresponding equilibrium flow. We now bound the routing and installation cost of $(\boldsymbol{f}, \lambda \boldsymbol{z}^*)$ separately. For the installation cost, we obtain

$$C^Z(\boldsymbol{f}, \lambda \boldsymbol{z}^*) = \sum_{e \in E} \lambda\, \ell_e\, z_e = \lambda(1-p)\, C(\boldsymbol{f}^*, \boldsymbol{z}),$$

and for the routing cost

$$\begin{aligned} C^R(\boldsymbol{f}, \lambda \boldsymbol{z}^*) &= \sum_{e \in E} S_e\left(\frac{f_e}{\lambda z_e^*}\right) f_e \le \sum_{e \in E} S_e\left(\frac{f_e}{\lambda z_e^*}\right) f_e^* \\ &= p\, C(\boldsymbol{f}^*, \boldsymbol{z}^*) + \sum_{e \in E}\left(S_e\left(\frac{f_e}{\lambda z_e^*}\right) f_e^* - S_e\left(\frac{f_e^*}{z_e^*}\right) f_e^*\right), \end{aligned} \tag{4.6}$$

where the first inequality uses the variational inequality (2.2). We proceed to bound $S_e(\frac{f_e}{\lambda z_e^*}) f_e^* - S_e(\frac{f_e^*}{z_e^*}) f_e^*$ in terms of the routing cost $S_e(\frac{f_e}{\lambda z_e^*}) f_e$ for that edge e. To this end, note that for each edge $e \in E$ with $f_e^* > 0$ we have

$$\begin{aligned} \frac{S_e\left(\frac{f_e}{\lambda z_e^*}\right) f_e^* - S_e\left(\frac{f_e^*}{z_e^*}\right) f_e^*}{S_e\left(\frac{f_e}{\lambda z_e^*}\right) f_e} &\le \sup\nolimits_{S \in \mathcal{S}} \sup\nolimits_{x,y \ge 0, z > 0} \frac{S\left(\frac{y}{\lambda z}\right) x - S(\frac{x}{z}) x}{S\left(\frac{y}{\lambda z}\right) y} \\ &= \sup\nolimits_{S \in \mathcal{S}} \sup\nolimits_{x,y \ge 0} \frac{S\left(\frac{y}{\lambda}\right) x - S(x) x}{S\left(\frac{y}{\lambda}\right) y} \\ &= \sup\nolimits_{S \in \mathcal{S}} \sup\nolimits_{x,y \ge 0} \frac{S(y) x - S(x) x}{S(y) \lambda y}. \end{aligned}$$

This implies $y \ge x$ and we may substitute $x = \gamma y$ with $\gamma \in [0,1]$. We then obtain for each edge $e \in E$ with $f_e^* > 0$ that

$$\begin{aligned} \frac{S_e\left(\frac{f_e}{\lambda z_e^*}\right) f_e^* - S_e\left(\frac{f_e^*}{z_e^*}\right) f_e^*}{S_e\left(\frac{f_e}{\lambda z_e^*}\right) f_e} &\le \sup\nolimits_{S \in \mathcal{S}} \sup\nolimits_{y \ge 0} \max\nolimits_{\gamma \in [0,1]} \frac{\gamma S(y) - \gamma S(\gamma y)}{\lambda S(y)} \\ &= \sup\nolimits_{S \in \mathcal{S}} \sup\nolimits_{y \ge 0} \max\nolimits_{\gamma \in [0,1]} \frac{\gamma}{\lambda}\left(1 - \frac{S(\gamma\, y)}{S(y)}\right) \\ &= \frac{\mu(\mathcal{S})}{\lambda}. \end{aligned} \tag{4.7}$$

Plugging (4.7) in (4.6), we obtain for the routing cost $C^R(\boldsymbol{f}, \lambda \boldsymbol{z}^*) \le p\, C(\boldsymbol{f}^*, \boldsymbol{z}^*) + \frac{\mu(\mathcal{S})}{\lambda}\, C^R(\boldsymbol{f}, \lambda \boldsymbol{z}^*)$ or, equivalently, $C^R(\boldsymbol{f}, \lambda \boldsymbol{z}^*) \le \frac{p}{1 - \mu(\mathcal{S})/\lambda} C(\boldsymbol{f}^*, \boldsymbol{z}^*)$. Thus, we can bound the total cost of the outcome of ScaleUniformly by

$$\begin{aligned} C(\boldsymbol{f}, \lambda \boldsymbol{z}^*) &= C^R(\boldsymbol{f}, \lambda \boldsymbol{z}^*) + C^Z(\boldsymbol{f}, \lambda \boldsymbol{z}^*) \\ &\le \frac{p}{1 - \mu(\mathcal{S})/\lambda} C(\boldsymbol{f}^*, \boldsymbol{z}^*) + \lambda(1-p)\, C(\boldsymbol{f}^*, \boldsymbol{z}^*) \\ &= \lambda\left(\frac{p}{\lambda - \mu(\mathcal{S})} + 1 - p\right) C(\boldsymbol{f}^*, \boldsymbol{z}^*). \end{aligned}$$

Since $\lambda = \mu(\mathcal{S}) + \sqrt{\mu(\mathcal{S}) \frac{p}{1-p}}$, we obtain

$$\frac{C(\boldsymbol{f}, \lambda \boldsymbol{z}^*)}{C(\boldsymbol{f}^*, \boldsymbol{z}^*)} \le p + 2\sqrt{p(1-p)\mu(\mathcal{S})} + \mu(\mathcal{S})(1-p) = \left(\sqrt{p} + \sqrt{\mu(\mathcal{S})(1-p)}\right)^2. \tag{4.8}$$

Elementary calculus shows that $\left(\sqrt{p} + \sqrt{\mu(\mathcal{S})(1-p)}\right)^2$ attains its maximum at $p = \frac{1}{1+\mu(\mathcal{S})}$. Substituting this value into (4.8) gives $C(\boldsymbol{f}, \lambda \boldsymbol{z}^*)/C(\boldsymbol{f}^*, \boldsymbol{z}^*) \le 1 + \mu(\mathcal{S})$ as claimed. ∎

For particular sets $\mathcal{S}$ of latency functions, we compute upper bounds on $\mu(\mathcal{S})$ in order to obtain an explicit upper bound on the approximation guarantees of BringToEquilibrium and ScaleUniformly.

We then obtain the following corollary of Theorem 4.2 and Theorem 4.3. The formal proof of this corollary is deferred to Appendix C.

Corollary 4.4. *For a set $\mathcal{S}$ of latency functions satisfying Assumption* 2.1, *the approximation guarantee of* BringToEquilibrium *and* ScaleUniformly *is at most*

(a) 2, *without further requirements on $\mathcal{S}$,*

(b) 5/4, *if $\mathcal{S}$ contains concave latencies only,*

(c) $1 + \frac{\Delta}{\Delta+1}\left(\frac{1}{\Delta+1}\right)^{1/\Delta}$, *if $\mathcal{S}$ contains only polynomials with non-negative coefficients and degree at most Δ, i.e., each $S \in \mathcal{S}$ is of the form $S(x) = \sum_{j\in J} a_j x^j$ for a finite set $J \subset \mathbb{R}_{\geq 0}$ with $\max J = \Delta$ and $a_j \geq 0$ for all j.*

4.2. Best-of-two approximation. In this section we show that although both algorithms, BringToEquilibrium and ScaleUniformly, achieve an approximation guarantee of $1 + \mu(\mathcal{S})$, by taking the better of the two algorithms we obtain a strictly better performance guarantee.

The key idea of the proof is to extend the analysis of the BringToEquilibrium algorithm in order to express its approximation guarantee as a function of the parameter p that measures the proportion of the routing cost in the total cost of a relaxed solution. This allows us to determine the worst-case p for which the approximation factor of both algorithm is minimized.

Our improved upper bound uses $\mu(\mathcal{S})$ as defined in (4.1) and a second parameter $\gamma(\mathcal{S})$, which is defined as follows:

$$\gamma(\mathcal{S}) = \sup_{S\in\mathcal{S}} \sup_{x>0} \sup\bigl\{\gamma \geq 0 : S(x/\gamma) = S(x) + xS'(x)\bigr\}. \tag{4.9}$$

Recall that by Lemma 4.1, $\mu(\mathcal{S}) = \sup_{S\in\mathcal{S}} \sup_{x\geq 0}\{\gamma \frac{S'(x)x}{S(x)+S'(x)x}\}$, where γ solves $S(x/\gamma) = S(x) + xS'(x)$. Intuitively, $\gamma(\mathcal{S})$ is an upper bound on the values of γ that satisfy this equation.

Theorem 4.5. *Taking the better solution of* BringToEquilibrium *and* ScaleUniformly *has an approximation guarantee of at most*

$$\frac{(\gamma(\mathcal{S}) + \mu(\mathcal{S}) + 1)^2}{(\gamma(\mathcal{S}) + \mu(\mathcal{S}) + 1)^2 - 4\mu(\mathcal{S})\gamma(\mathcal{S})} < 1 + \mu(\mathcal{S}).$$

Proof. Recall from (4.8) that the approximation guarantee of the algorithm ScaleUniformly is $\bigl(\sqrt{p} + \sqrt{\mu(\mathcal{S})(1-p)}\bigr)^2$, where $p = C^R(\boldsymbol{f}^*, \boldsymbol{z}^*)/C(\boldsymbol{f}^*, \boldsymbol{z}^*)$. We extend our analysis of BringToEquilibrium using this parameter p. With the notation in Theorem 4.2, by (4.5), BringToEquilibrium returns a feasible solution $(\boldsymbol{f}^*, \boldsymbol{z})$ with

$$C(\boldsymbol{f}^*, \boldsymbol{z}) = \sum_{e\in E}\Bigl(\bigl(S_e(x_e) + S_e'(x_e)\, x_e\bigr) f_e^* + \gamma_e\, S_e'(x_e)\, x_e\, f_e^*\Bigr),$$

where $x_e = f_e^*/z_e^*$. We obtain

$$\begin{aligned} C(\boldsymbol{f}^*, \boldsymbol{z}) &= p\, C(\boldsymbol{f}^*, \boldsymbol{z}^*) + \sum_{e\in E} S_e'(x_e)\, x_e\, f_e^*(1+\gamma_e) \\ &\overset{(4.9)}{\leq} p\, C(\boldsymbol{f}^*, \boldsymbol{z}^*) + (1+\gamma(\mathcal{S})) \sum_{e\in E} S_e'(x_e)\, x_e\, f_e^* \\ &= p\, C(\boldsymbol{f}^*, \boldsymbol{z}^*) + (1+\gamma(\mathcal{S}))(1-p)\, C(\boldsymbol{f}^*, \boldsymbol{z}^*) \\ &= \big(1+\gamma(\mathcal{S})(1-p)\big)\, C(\boldsymbol{f}^*, \boldsymbol{z}^*). \end{aligned}$$

Thus, by taking the best of the two heuristics, we obtain an approximation guarantee of

$$\max_{p\in(0,1)} \min\left\{1+\gamma(\mathcal{S})(1-p), \left(\sqrt{p} + \sqrt{\mu(\mathcal{S})(1-p)}\right)^2\right\}.$$

We claim that

$$\max_{p\in(0,1)} \min\left\{1+\gamma(1-p), \left(\sqrt{p} + \sqrt{\mu(1-p)}\right)^2\right\} = \frac{(\gamma+\mu+1)^2}{(\gamma+\mu+1)^2 - 4\mu\gamma} < 1+\mu \tag{4.10}$$

for all $\gamma, \mu \in (0, 1]$, which implies the result.

To see the claim, first observe that $1 + \gamma(1-p)$ is linearly decreasing in p. Elementary calculus shows that the expression $\big(\sqrt{p} + \sqrt{\mu(1-p)}\big)^2$ attains its maximum at $p = \hat{p} := \frac{1}{1+\mu}$, is increasing when $p < \hat{p}$, and decreasing afterwards. Further,

$$\left(\sqrt{\hat{p}} + \sqrt{\mu(1-\hat{p})}\right)^2 = 1+\mu$$

and

$$1+\gamma(1-\hat{p}) = 1+\mu\frac{\gamma}{1+\mu} < 1+\mu,$$

thus, the inequality in (4.10) follows.

Moreover, it follows that the maximum on the left-hand side of (4.10) is attained for the unique $p^* \in (0, \hat{p})$ such that $1 + \gamma(1-p^*) = (\sqrt{p^*} + \sqrt{\mu(1-p^*)})^2$. Thus, p^* is a solution to the equation

$$\begin{aligned} 0 &= -(1-p^*) - \gamma(1-p^*) + 2\sqrt{p^*(1-p^*)\mu} + \mu(1-p^*) \\ &= (1-p^*)\left(2\sqrt{\mu\frac{p^*}{1-p^*}} + \mu - \gamma - 1\right), \end{aligned}$$

and since $p^* < 1$

$$0 = 2\sqrt{\mu\frac{p^*}{1-p^*}} + \mu - \gamma(\mathcal{S}) - 1.$$

The unique solution to this equation is

$$p^* = \frac{(\gamma-\mu+1)^2}{(\gamma-\mu+1)^2 + 4\mu}. \tag{4.11}$$

Plugging this into the left-hand side of (4.10) gives

$$\frac{(\gamma+\mu+1)^2}{(\gamma+\mu+1)^2-4\mu\gamma},$$

which proves the identity in (4.10). ∎

It is not necessary to run both approximation algorithms to get this approximation guarantee. After computing the optimum solution to the relaxation (CNDP′), we can determine the value for $p = C^R(\boldsymbol{f}^*, \boldsymbol{z}^*)/C(\boldsymbol{f}^*, \boldsymbol{z}^*)$ and proceed with ScaleUniformly if $p \leq p^*$ (cf. (4.11)) and with BringToEquilibrium otherwise.

For particular sets $\mathcal{S}$ of latency functions, we evaluate $\mu(\mathcal{S})$ and $\gamma(\mathcal{S})$ and obtain the following corollary of Theorem 4.5. The formal proof of this corollary is deferred to Appendix C.

Corollary 4.6. *For a set $\mathcal{S}$ of latency functions satisfying Assumption* 2.1*, the approximation guarantee in Theorem* 4.5 *is at most*

(a) $9/5$*, without further requirements on* $\mathcal{S}$*,*
(b) $49/41 \approx 1.195$*, if* $\mathcal{S}$ *contains concave latencies only,*
(c) $1 + \frac{4\Delta(\Delta+1)}{2(2\Delta+1)(\Delta+1)^{1+1/\Delta}+(\Delta+1)^{2(1+1/\Delta)}+1}$*, if* $\mathcal{S}$ *contains only polynomials with nonnegative coefficients and degree at most* Δ*, i.e., each* $S \in \mathcal{S}$ *is of the form* $S(x) = \sum_{j\in J} a_j x^j$ *for a finite set* $J \subset \mathbb{R}_{\geq 0}$ *with* $\max J = \Delta$ *and* $a_j \geq 0$ *for all* j*.*

5. Conclusion. We reconsidered a variant of the continuous network design problem (CNDP) and established, to the best of our knowledge, the first hardness results for CNDP. Further, we provided a general approximation guarantee for an algorithm studied by Marcotte [23], which depends on the set of allowed cost functions. Interestingly, the approximation guarantee is related to the *anarchy value* of the set of cost functions. We then introduced a new approximation algorithm and showed that it achieves the same approximation guarantee as Marcotte's algorithm. Finally, we showed that the approximation guarantees can be improved by taking the best of both approximation algorithms.

There are several open questions that deserve further research. All results work so far for the case of edge-separable latency functions. The case of nonseparable latencies is—from the approximation point of view—not well understood. Designing the network infrastructure assuming dynamic equilibrium models seems another challenging area for further research. In particular, it would be interesting to design approximation algorithms for rather simple dynamic flow models (e.g., using simple graph topologies and deterministic queuing models).

Appendix A. Additional material for the proof of Proposition 2.4.

Lemma A.1. *Let* S *satisfy Assumption* 2.1*. Then, the following holds:*

1. $\lim_{x\downarrow 0} S'(x)x^2 = 0$,
2. $\lim_{x\uparrow +\infty} S'(x)x^2 = +\infty$.

Proof. We first observe that differentiability and semi-convexity of S implies for any $x, y \geq 0$

$$S(y)y \geq S(x)x + \big(S'(x)x + S(x)\big)(y-x).$$

For 1, take $y = 2x$. We obtain

$$\begin{aligned} S(2x)2x &\geq S(x)x + \big(S'(x)x + S(x)\big)x \\ &\Leftrightarrow S'(x)x^2 \leq (S(2x) - S(x))2x. \end{aligned}$$

Taking the limit $\lim_{x\downarrow 0}$ on both sides of the inequality proves the first claim.

For 2, take $y = 1$. We obtain

$$\begin{aligned} S(1) &\geq S(x)x - x\big(S'(x)x + S(x)\big) + S'(x)x + S(x) \\ &\Leftrightarrow S'(x)x^2 \geq S'(x)x + S(x) - S(1). \end{aligned}$$

Taking the limit $\lim_{x\uparrow+\infty}$ on both sides of the inequality and using that S is unbounded and strictly increasing proves the second claim. □

Appendix B. Additional material for the proofs of Theorem 3.1 and Theorem 3.3.

LEMMA B.1. *Let $E \subseteq \{1,2,3\}$ be nonempty, $m \in [0,1]$, and*

$$\begin{aligned} Q = \inf_{\substack{g_e\in[0,1],e\in E\\ x_e>0,e\in E\\ h\in[0,1]}} \sum_{e\in E}\left(\frac{(g_e+h)^2}{x_e} + x_e - 2g_e - \frac{4}{3}h\right) \\ \textit{s.t.} \quad \textstyle\sum_{e\in E}\frac{g_e+h}{x_e} &\geq 4 \quad \textit{if } h<1, \\ g_e &\geq m \quad \textit{for all } e\in E, \\ g_e + h &> 0 \quad \textit{for all } e\in E. \end{aligned}$$

Then, $Q = \frac{1}{4}m$, which is attained for $h = 0$ and $g_e = m, x_e = \frac{3}{4}m$ for all $e \in E$.

Proof. We show the proof only for the case $m > 0$. The case $m = 0$ follows from the case $m > 0$ by considering the limit $m \to 0$.

We consider the problem

$$\begin{aligned} Q' = \inf_{\substack{g_e\geq 0,e\in E\\ x_e>0,e\in E\\ h\in[0,1]}} \sum_{e\in E}\left(\frac{(g_e+h)^2}{x_e} + x_e - 2g_e - \frac{4}{3}h\right) \\ \text{s.t.} \quad \textstyle\sum_{e\in E}\frac{g_e+h}{x_e} &\geq 4 \quad \text{if } h<1, \\ g_e &\geq m \quad \text{for all } e\in E, \\ g_e + h &> 0 \quad \text{for all } e\in E. \end{aligned}$$

where the upper bounds on the variables g_e, $e \in E$ are dropped.

Let $(g_e)_{e\in E}, (x_e)_{e\in E}, h$ be an optimal solution to Q'. To show that the solution has the desired structure, we will gradually reduce the possible space of the solution.

First, let us consider the case $h < 1$. We claim that $h = 0$ in that case. Suppose not, for a contradiction. Then, the alternative solution $(g'_e)_{e\in E}$, $(x'_e)_{e\in E}$, h defined as $g'_e = g_e + h$, $x'_e = x_e$, and $h' = 0$ satisfies all constraints and strictly reduces the objective value, which is a contradiction to the optimality of the original solution. We substitute $y_e = g_e/x_e$ for all $e \in E$. We then obtain the problem

$$\begin{aligned} Q'' = \inf_{\substack{g_e>0,e\in E\\ y_e>0,e\in E}} \sum_{e\in E} g_e\Big(y_e + \frac{1}{y_e} - 2\Big) \\ \text{s.t.} \quad \textstyle\sum_{e\in E} y_e &\geq 4, \\ g_e &\geq m \quad \text{for all } e\in E. \end{aligned}$$

When $m > 0$, the problem Q'' is clearly minimized when $g_e = m$ for all $e \in E$. Since the term $\sum_{e\in E} y_e + \frac{1}{y_e} - 2$ is convex and the constraint $\sum_{e\in E} y_e = 4$ is linear, the optimal solution of Q'' is attained for $y_e = 4/3$ for all $e \in E$. After resubstitution, we derive that the optimal solution to Q' is $g_e = m$ and $x_e = \frac{3}{4}m$. As this solution is also feasible for Q, it also solves Q to optimality and we obtain $Q = \frac{1}{4}m$.

For the second case, let us assume that $h = 1$. Then, we obtain the minimization problem

$$Q = \inf_{\substack{g_e \in [m,1], e \in E \\ x_e > 0, e \in E}} \sum_{e \in E} \left(\frac{(g_e + 1)^2}{x_e} + x_e - 2g_e - \frac{4}{3} \right).$$

By first-order conditions, the optimal capacities are $x_e = g_e + 1$ for all $e \in E$. We then obtain

$$Q = \min_{g_e \in [m,1], e \in E} \sum_{e \in E} \left(2(g_e + 1) - 2g_e - \frac{4}{3} \right) = 2,$$

which is larger than $\frac{1}{4}m$ for all $m \in [0, 1]$ and the claimed result follows. ◻

Appendix C. Additional material for the proofs of Corollary 4.4 and Corollary 4.6.

For the proofs of Corollary 4.4 and Corollary 4.6, we give bounds on $\mu(\mathcal{S})$ and $\gamma(\mathcal{S})$ for the respective sets $\mathcal{S}$ of allowable latency functions. Theorem 4.2, Theorem 4.3, and Theorem 4.5 then give the claimed approximation guarantees.

Arbitrary latency functions. First, we consider case (a) of both corollaries, where $\mathcal{S}$ is a class of arbitrary non-negative and strictly increasing latencies. We observe that

$$\begin{aligned}
\mu(\mathcal{S}) &= \sup_{S \in \mathcal{S}: S(x) > 0} \sup_{x \geq 0} \max_{\gamma \in [0,1]} \gamma \left(1 - \frac{S(\gamma x)}{S(x)} \right) \leq 1, \\
\gamma(\mathcal{S}) &= \sup_{S \in \mathcal{S}} \sup_{x > 0} \sup\{\gamma \geq 0 : S(x/\gamma) = S(x) + xS'(x)\} \leq 1.
\end{aligned}$$

Now Corollary 4.4 (a) follows immediately and Corollary 4.6 (a) follows from the fact that

$$\frac{(\gamma(\mathcal{S}) + \mu(\mathcal{S}) + 1)^2}{(\gamma(\mathcal{S}) + \mu(\mathcal{S}) + 1)^2 - 4\mu(\mathcal{S})\gamma(\mathcal{S})} \tag{C.1}$$

is strictly increasing in $\gamma(\mathcal{S})$ and $\mu(\mathcal{S})$.

Concave latency functions. Next, consider case (b) of both corollaries, where $\mathcal{S}$ contains concave latencies only. Observe that

$$\begin{aligned}
\mu(\mathcal{S}) &= \sup_{S \in \mathcal{S}} \sup_{x \geq 0} \max_{\gamma \in [0,1]} \left\{ \gamma \left(1 - \frac{S(\gamma x)}{S(x)} \right) \right\} \\
&\leq \sup_{S \in \mathcal{S}} \sup_{x \geq 0} \max_{\gamma \in [0,1]} \left\{ \gamma \left(1 - \gamma - \frac{(1-\gamma)S(0)}{S(x)} \right) \right\} \\
&\leq \max_{\gamma \in [0,1]} \gamma(1 - \gamma) \\
&= 1/4,
\end{aligned}$$

where the first inequality uses the concavity of all functions $S \in \mathcal{S}$.

We next show that $\gamma(\mathcal{S}) \leq 1/2$. Let $S \in \mathcal{S}$ be arbitrary. Since S is concave, we get for any $x \geq 0, y \geq 0$ that

$$S(y) \leq S(x) + S'(x)(y - x).$$

Take $y = 2x$, which yields

$$S(2x) \leq S(x) + S'(x)x.$$

Thus, we get for any $\gamma \geq 0$ satisfying $S(x/\gamma) = S(x)+S'(x)x$ the following inequality:

$$S(x/\gamma) = S(x) + S'(x)x \geq S(2x).$$

Since S is strictly increasing we obtain $\gamma \leq 1/2$.

Polynomial latency functions. Finally, consider case (c) of both corollaries, where there is a finite set $J \subset \mathbb{R}_{\geq 0}$ of exponents with $\max J = \Delta$ such that $\mathcal{S}$ contains only latency functions of type $S(x) = \sum_{j \in J} a_j x^j$, with $a_j \geq 0$ for all j. Denote $\boldsymbol{a} = (a_j)_{j \in J}$. We calculate

$$\begin{aligned}
\mu(\mathcal{S}) &= \sup_{S \in \mathcal{S}} \sup_{x \geq 0} \max_{\gamma \in [0,1]} \left\{ \gamma \left(1 - \frac{S(\gamma x)}{S(x)} \right) \right\} \\
&= \sup_{\boldsymbol{a} \geq 0} \sup_{x \geq 0} \max_{\gamma \in [0,1]} \left\{ \gamma \left(1 - \frac{\sum_{j \in J} a_j \gamma^j x^j}{\sum_{j \in J} a_j x^j} \right) \right\} \\
&= \sup_{\boldsymbol{a} \geq 0} \sup_{x \geq 0} \max_{\gamma \in [0,1]} \left\{ \gamma \left(\frac{\sum_{j \in J} a_j x^j (1 - \gamma^j)}{\sum_{j \in J} a_j x^j} \right) \right\}.
\end{aligned}$$

As $(1 - \gamma^j)$ is increasing in j for every $\gamma \in (0, 1)$, it follows that the supremum over $\boldsymbol{a} \geq 0$ is attained if $a_\Delta > 0$ and $a_j = 0$ for all $j \in J \setminus \{\Delta\}$. We obtain

$$\begin{aligned}
\mu(\mathcal{S}) &= \max_{\gamma \in [0,1]} \gamma (1 - \gamma^\Delta) \\
&= \left(\frac{1}{\Delta + 1} \right)^{1/\Delta} \left(1 - \frac{1}{\Delta + 1} \right) \\
&= \left(\frac{1}{\Delta + 1} \right)^{1/\Delta} \left(\frac{\Delta}{\Delta + 1} \right).
\end{aligned}$$

which directly implies the statement of Corollary 4.4 (c).

We next show that $\gamma(\mathcal{S}) \leq \left(\frac{1}{\Delta+1}\right)^{1/\Delta}$. To this end we show that, for any latency function in $\mathcal{S}$ and $\gamma := \left(\frac{1}{\Delta+1}\right)^{1/\Delta}$, we have $S(x/\gamma) \leq S(x) + S'(x)x$ for any $x > 0$. Observe that

$$S(x) + S'(x)x = \sum_{j \in J} a_j x^j + \sum_{j \in J} j a_j x^j = \sum_{j \in J} (1 + j) \cdot a_j x^j$$

and

$$S(x/\gamma) = S(x \cdot (\Delta + 1)^{1/\Delta}) = \sum_{j \in J} (\Delta + 1)^{j/\Delta} a_j x^j.$$

Since $a_j \geq 0$ for all j, it suffices to show that $(\Delta + 1)^{j/\Delta} \leq 1 + j$ for all $j \in J$. This follows from the fact that $(\Delta + 1)^{j/\Delta}$ is convex in j, and we get equality for $j = 0$ and $j = \Delta$. It follows that $\gamma(\mathcal{S}) \leq \left(\frac{1}{\Delta+1}\right)^{1/\Delta}$.

Plugging these values in (C.1) and rearranging terms, we obtain the approximation guarantee claimed in Corollary 4.6 (c).

Acknowledgments. We would like to thank the two anonymous referees for their detailed and helpful suggestions that helped to improve the exposition of the paper.

REFERENCES

[1] M. Abdulaal and L. J. LeBlanc, *Continuous equilibrium network design models*, Transp. Res. Part B, 13 (1979), pp. 19–32.

[2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice-Hall, Upper Saddle River, NJ, 1993.

[3] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics and Transportation*, Yale University Press, New Haven, CT, 1956.

[4] P. Berman, M. Karpinski, and A. D. Scott, *Approximation hardness and satisfiability of bounded occurrence instances of SAT*, in Electronic Colloquium on Computational Complexity (ECCC), 2003, Computational Complexity Foundation, Report No. 27.

[5] P. Berman, M. Karpinski, and A. D. Scott, *Approximation hardness of short symmetric instances of MAX-3SAT*, in Electronic Colloquium on Computational Complexity (ECCC), 2003, Computational Complexity Foundation, Report No. 49.

[6] U. Bhaskar, K. Ligett, and L. J. Schulman, *Network improvement for equilibrium routing*, in Integer Programming and Combinatorial Optimization - 17th International Conference, IPCO 2014, Bonn, Germany, June 23–25, 2014, Springer, Berlin, 2014, pp. 138–149.

[7] D. Braess, *Über ein Paradoxon aus der Verkehrsplanung*, Unternehmensforschung, 12 (1968), pp. 258–268. (German.)

[8] B. Colson, P. Marcotte, and G. Savard, *An overview of bilevel optimization*, Ann. Oper. Res., 153 (2007), pp. 235–256.

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

[10] S. C. Dafermos, *Traffic assignment and resource allocation in transportation networks*, Ph.D. thesis, John Hopkins University, Baltimore, MD, 1968.

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

[12] G. B. Dantzig, R. P. Harvey, Z. F. Lansdowne, D. W. Robinson, and Steven F. Maier, *Formulating and solving the network design problem by decomposition*, Transp. Res. Part B, 13 (1979), pp. 5–17.

[13] P. Dubey, *Inefficiency of Nash equilibria*, Math. Oper. Res., 11 (1986), pp. 1–8.

[14] T. L. Friesz, *Transportation network equilibrium, design and aggregation: Key developments and research opportunities*, Transp. Res. Part A, 17 (1985), pp. 411–426.

[15] M. Gairing, T. Harks, and M. Klimm, *Complexity and approximation of the continuous network design problem*, in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014, September 4–6, 2014, Barcelona, Spain, 2014, Dagstuhl Publishing, Saarbrücken Wadern, Germany, pp. 226–241.

[16] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, 2nd ed., Algorithms and Combinatorics 2, Springer, Berlin, 1993.

[17] Y. A. Korilis, A. A. Lazar, and A. Orda, *Architecting noncooperative networks*, IEEE J. Sel. Area Commun., 13 (1995), pp. 1241–1251.

[18] Y. A. Korilis, A. A. Lazar, and A. Orda, *Avoiding the Braess paradox in noncooperative networks*, J. Appl. Probab., 36 (1999), pp. 211–222.

[19] L. Libman and A. Orda, *The designer's perspective to atomic noncooperative networks*, IEEE/ACM Trans. Networking, 7 (1999), pp. 875–884.

[20] H. Lin, T. Roughgarden, É. Tardos, and A. Walkover, *Stronger bounds on Braess's paradox and the maximum latency of selfish routing*, SIAM J. Comput., 25 (2011), pp. 1667–1686.

[21] Z. Luo, J. Pang, and D. Ralph, *Mathematical programs with equilibrium constraints*, Cambridge University Press, Cambridge, UK, 1996.

[22] T. L. Magnanti and R. T. Wong, *Network design and transportation planning: Models and algorithms*, Transp. Sci., 18 (1984), pp. 1–55.

[23] P. Marcotte, *Network design problem with congestion effects: A case of bilevel programming*, Math. Program., 34 (1986), pp. 142–162.

[24] P. Marcotte and G. Marquis, *Efficient implementation of heuristics for the continuous network design problem*, Ann. Oper. Res., 34 (1992), pp. 163–176.

[25] A. C. Pigou, *The Economics of Welfare*, Macmillan, London, UK, 1920.

[26] T. Roughgarden, *The price of anarchy is independent of the network topology*, J. Comput. Syst. Sci., 67 (2002), pp. 341–364.

[27] T. Roughgarden, *On the severity of Braess's paradox: Designing networks for selfish users is hard*, J. Comput. Syst. Sci., 72 (2006), pp. 922–953.

[28] T. Roughgarden and É. Tardos, *How bad is selfish routing?*, J. ACM, 49 (2002), pp. 236–259.

[29] M. J. Smith, *The existence, uniqueness and stability of traffic equilibria*, Transp. Res., 13 (1979), pp. 295–304.
[30] Traffic assignment manual, U. S. Bureau of Public Roads, U.S. Department of Commerce, Urban Planning Division, Washington, DC, 1964.
[31] J. G. Wardrop, *Some theoretical aspects of road traffic research*, Proc. Inst. Civil Eng., 1 (1952), pp. 325–378.
[32] H. Yang and M. G. H. Bell, *Models and algorithms for road network design: A review and some new developments*, Transport Rev., 18 (1998), pp. 257–278.