

# How Much Can Taxes Help Selfish Routing?\*

Richard Cole<sup>†</sup>      Yevgeniy Dodis<sup>‡</sup>      Tim Roughgarden<sup>§</sup>

October 30, 2003

## Abstract

We study economic incentives for influencing selfish behavior in networks. We consider a model of selfish routing in which the latency experienced by network traffic on an edge of the network is a function of the edge congestion, and network users are assumed to selfishly route traffic on minimum-latency paths. The quality of a routing of traffic is historically measured by the sum of all travel times, also called the *total latency*.

It is well known that the outcome of selfish routing (a *Nash equilibrium*) does not minimize the total latency, and that *marginal cost pricing*—charging each network user for the congestion effects caused by its presence—eliminates the inefficiency of selfish routing. However, the principle of marginal cost pricing assumes that taxes cause no disutility to network users; this is appropriate only when collected taxes can be feasibly returned (directly or indirectly) to the users. If this assumption does not hold and we wish to minimize the total user disutility (latency plus taxes paid)—the total *cost*—how should we price the network edges? Intuition may suggest that taxes can never improve the cost of a Nash equilibrium, but the famous *Braess's Paradox* shows this intuition to be incorrect.

We consider strategies for pricing network edges to reduce the cost of a Nash equilibrium. Since levying a sufficiently large tax on an edge effectively removes it from the network, our study generalizes previous work on designing networks for selfish users [29]. In this paper, we prove the following results.

- In a large class of networks—including all networks with linear latency functions—marginal cost taxes do not improve the cost of a Nash equilibrium.
- The best-possible benefit from arbitrary taxes does not exceed that of edge removal. In *every* network with linear latency functions, the benefit of taxes cannot

---

\*A preliminary version of this paper appeared in the Proceedings of the 4th ACM Conference on Electronic Commerce, June 2003.

<sup>†</sup>Department of Computer Science, New York University, 251 Mercer Street, New York, NY 10012. Supported in part by NSF grant CCR0105678. Email: cole@cs.nyu.edu.

<sup>‡</sup>Department of Computer Science, New York University, 251 Mercer Street, New York, NY 10012. Supported by in part an NSF CAREER Award. Email: dodis@cs.nyu.edu.

<sup>§</sup>Computer Science Division, UC Berkeley, Berkeley, CA 94720. Supported by an NSF Postdoctoral Fellowship. Most of this research was performed while the author was visiting NYU and supported by ONR grant N00014-98-1-0589 as a postdoctoral researcher at Cornell University. Email: timr@cs.berkeley.edu.

exceed that removing edges. There are networks with nonlinear latency functions, however, in which taxes are radically more powerful than edge removal.

- Strong inapproximability results known for network design [29] carry over to the problem of computing optimal taxes, proving this problem computationally intractable.

# 1 Introduction

## Selfish Routing and Marginal Cost Pricing

We study economic incentives for influencing selfish behavior in networks. We focus on a simple model of *selfish routing*, defined by Wardrop [36] and first studied in the theoretical computer science literature by Roughgarden and Tardos [33]. In this model, we are given a directed network in which each edge possesses a latency function describing the common delay experienced by all traffic on the edge as a function of the edge congestion. There is a fixed amount of traffic that wishes to travel from a source vertex  $s$  to a sink vertex  $t$ ; as in most earlier works, we assume that the traffic comprises a very large population of users, so that the actions of a single individual have negligible effect on network congestion. The quality of an assignment of traffic to  $s$ - $t$  paths is historically measured by the resulting sum of all travel times—the *total latency*. We assume that each network user, when left to its own devices, acts selfishly and routes itself on a minimum-latency path, given the network congestion due to the other users. In general such a “selfish” assignment of traffic to paths (a *Nash equilibrium*) does not minimize the total latency; put differently, the outcome of selfish behavior can be improved upon with coordination.

The inefficiency of selfish routing motivates the introduction of economic incentives to ensure that selfish behavior results in a socially desirable routing of traffic. An ancient idea—discussed informally as early as 1920 [27]—is to use *marginal cost pricing*. The principle of marginal cost pricing asserts that on each edge, each network user on the edge should pay a tax equal to the additional delay its presence causes for the other users on the edge. Assuming that all network users choose routes to minimize the sum of the latency experienced and the taxes paid, this principle ensures that the resulting Nash equilibrium achieves the minimum-possible total latency [3]. Briefly, the inefficiency of selfish routing can always be eradicated by pricing network edges appropriately.

The following observation motivates our work: the principle of marginal cost pricing is single-minded in its pursuit of a minimum-latency flow, and ignores the disutility to network users due to (possibly very large) taxes. This assumption is only appropriate when collected taxes can be feasibly returned (directly or indirectly) to the network users, for example by refunding taxes equally to all users (a “lump-sum refund”). In this paper, we are interested in settings where this assumption is not reasonable. For example, refunding the collected taxes to network users could be logistically or economically infeasible, or taxes could represent quantities of a non-monetary, non-refundable good such as time delays.

If we wish to minimize the total user disutility—latency plus taxes paid, a quantity we call the *cost*—rather than merely the total latency, how should we price the network edges?

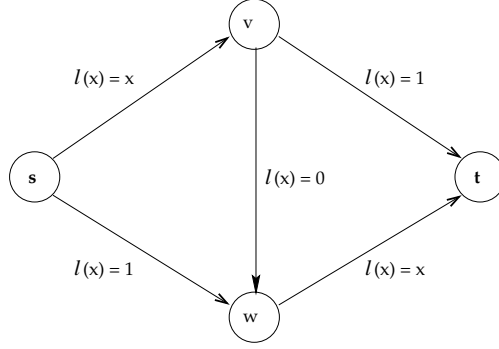


Figure 1: Braess's Paradox

## Braess's Paradox and the Power of Taxes

Intuition may suggest that taxes, which can only increase the disutility incurred by users along an  $s$ - $t$  path, can never improve the cost of a Nash equilibrium. A famous example that we describe next, called *Braess's Paradox* and shown in Figure 1, shows this intuition to be incorrect [5].

Each edge of the network in Figure 1 is labeled with its latency function, giving the delay incurred by traffic on the link as a function of the fraction of the traffic that uses the link. In the (unique) Nash equilibrium, all traffic uses route  $s \rightarrow v \rightarrow w \rightarrow t$  and experiences two units of latency. On the other hand, if half a unit of tax is levied on edge  $(v, w)$ , then in the Nash equilibrium half of the traffic uses each of the routes  $s \rightarrow v \rightarrow t$  and  $s \rightarrow w \rightarrow t$ . In this Nash equilibrium, everyone experiences latency  $\frac{3}{2}$  and no taxes are paid; this outcome has cost  $\frac{3}{2}$  and is clearly superior to the original Nash equilibrium in the absence of taxes.

It is illuminating to study the edge taxes that are dictated by the principle of marginal cost pricing for the network of Figure 1. As it turns out, these taxes are  $\frac{1}{2}$  on edges  $(s, v)$  and  $(w, t)$  and 0 on the other three edges. With these taxes, we obtain the same Nash equilibrium as in our previous solution with taxes (and thus all traffic experiences latency  $\frac{3}{2}$ ); however, all traffic must also pay  $\frac{1}{2}$  unit of tax. This solution thus has a cost of 2, and it seems evident that the previous solution should be preferred over marginal cost taxes in this example.

The potential power of taxes in decreasing the disutility experienced by selfish network users, together with the failure of marginal cost pricing to minimize this disutility, motivate the questions that we study in this paper:

- (1) Are marginal cost taxes ever a good idea for minimizing the cost of a Nash equilibrium?
- (2) A sufficiently large edge tax effectively removes the edge from the network, and the power of removing edges to improve a Nash equilibrium is well understood [29]. Taxation is thus at least as powerful as edge removal; when is it strictly more powerful?
- (3) Can we compute or approximate optimal taxes efficiently?

Problem Studied	Linear	General
Do marginal cost taxes ever help?	no	yes
Maximum benefit of taxes	4/3	$n/2$
Are taxes more powerful than network design?	no	yes
Approximability of optimal taxes	4/3	$[\omega(n^{1-\epsilon}), n/2]$

Table 1: Contributions of this paper. All results depend on the class of allowable edge latency functions. For simplicity, we list only our results for networks with linear latency functions and for networks with general (continuous, nondecreasing) latency functions. The benefit of taxes is measured by the ratio in Nash flow cost before and after taxes are levied. Approximability refers to the best-possible approximation ratio achievable by a polynomial-time algorithm, assuming  $P \neq NP$ .

## Our Results

Our contributions on these three questions are as follows.

- In all networks with linear latency functions, marginal cost taxes do not improve the cost of a Nash equilibrium.
- We prove that the maximum-possible benefit due to edge taxation is no more than what can be achieved, in the best case, via edge removal.
- For networks with linear latency functions, we show the stronger statement that in every such network—not merely worst-case examples—taxes cannot decrease the cost of a Nash equilibrium beyond what can be achieved by removing edges. By contrast, we show that in general there are networks in which taxes can radically improve over the best subgraph solution.
- We show that most of the very strong inapproximability results known for network design [29] carry over to the problem of computing optimal taxes. As a consequence, essentially the best-possible polynomial-time heuristic for the problem is to assign no taxes at all (assuming  $P \neq NP$ ).

These results are summarized in Table 1. Determining the maximum-possible benefit of taxes and the (in)approximability of the problem of computing optimal taxes requires only reasonably straightforward extensions of previous work on network design [29]; all other results of this paper require new constructions and proof approaches.

## Related Work

The model of selfish routing studied in this paper was first defined in the 1950’s by Wardrop [36], and has been extensively studied ever since. It has been generalized in multiple directions that we do not explore here (e.g., [8, 17, 18, 24, 28, 32]). For other network games motivated by the Internet, see e.g., [13, 25]. For further discussion and references on these topics, see [30].

We make no attempt at a detailed survey of previous work on the optimal pricing of resources, and instead refer the reader to [6, 16] for recent collections of surveys on pricing by the transportation science literature, and to [23] for an overview on approaches to optimal pricing in welfare economics. The transportation science community has long realized that marginal cost taxes may cause users to pay more than is necessary (see e.g. [4, 15]), but has apparently neglected the idea of incorporating the amount of taxes paid into the definition of social welfare. The more general problem of using taxation to optimize the sum of user utilities has been extensively studied by economists (see e.g. [23]), although not with the same emphasis on worst- and best-case performance guarantees and on computational efficiency that characterizes our work here. In both the economics and networking literature there are numerous papers on pricing networks (see e.g. [20, 21]), but this body of work has not studied the model of selfish routing considered in this paper. Finally, our approach is similar in spirit to work on “frugal mechanisms” pioneered by Archer and Tardos [1, 2] and studied further in [11, 22]. These papers concern mechanisms (such as auctions) that solve an optimization problem in an incentive-compatible way, but also make use of only moderate incentives.

The paper closest to the present work is that of Roughgarden on designing networks for selfish users [29]. Indeed, the central questions of [29], which concern the maximum-possible benefit from and the algorithmic complexity of removing edges from a network with selfish routing, can be viewed as a special case of some of the problems considered here, with edge taxes restricted to be either 0 or  $+\infty$ . Our work in this paper extends some of the results from [29] to the setting of more general taxes, and in addition tackles several problems that have not been previously considered.

## Organization

In Section 2 we formally define our traffic routing model and review useful results from past works. In Section 3 we ask if marginal cost pricing can improve the cost of selfish routing, and we resolve this question in the negative for networks with linear latency functions. In Sections 4 and 5 we ask if the power of taxes exceeds that of edge removal. In Section 4 we adapt previous work on network design [29] to show that the best-possible improvement due to taxes cannot exceed that of edge removal. In Section 5 we show that taxes never improve over the best solution achievable by removing edges in networks with linear latency functions, but can be radically more powerful in general networks. Finally, in Section 6 we ask whether optimal taxes can be efficiently computed. We answer this question in the negative by extending known inapproximability results for network design [29] to the problem of computing optimal taxes.

## 2 Preliminaries

### 2.1 The Model

We follow the notation and conventions of Roughgarden and Tardos [33]. We study a single-commodity flow network, described by a directed graph  $G = (V, E)$  with source vertex  $s$  and sink vertex  $t$ . We allow parallel edges but have no use for self-loops. We denote the

set of simple  $s$ - $t$  paths by  $\mathcal{P}$ , and we assume that this set is nonempty. A *flow* is a function  $f : \mathcal{P} \rightarrow \mathcal{R}^+$ ; for a fixed flow  $f$  we define  $f_e = \sum_{P:e \in P} f_P$  as the amount of traffic using edge  $e$  en route from  $s$  to  $t$ . With respect to a finite and positive *traffic rate*  $r$ , a flow  $f$  is said to be *feasible* if  $\sum_{P \in \mathcal{P}} f_P = r$ .

The network  $G$  suffers from congestion effects; to model this, we assume each edge  $e$  possesses a nonnegative, continuous, nondecreasing *latency function*  $\ell_e$  that describes the delay incurred by traffic on  $e$  as a function of the edge congestion  $f_e$ . The latency of a path  $P$  in  $G$  with respect to a flow  $f$  is then given by  $\ell_P(f) = \sum_{e \in P} \ell_e(f_e)$ . The quality of a flow is historically measured by its *total latency*  $L(f)$ , defined by

$$L(f) \equiv \sum_{P \in \mathcal{P}} \ell_P(f) f_P = \sum_{e \in E} \ell_e(f_e) f_e,$$

with the equality proved by writing  $\ell_P(f)$  as a sum over edges and reversing the order of summation. We will call a flow minimizing  $L(\cdot)$  *optimal* or *minimum-latency*. An optimal flow always exists, as the space of all flows is compact and  $L(\cdot)$  is assumed continuous.

Finally, we will allow a set of nonnegative *taxes*  $\{\tau_e\}_{e \in E}$  to be placed on the edges of a network  $G$ . We will write  $\tau_P = \sum_{e \in P} \tau_e$  for the total taxes on a path  $P$ . The *cost*  $C(f, \tau)$  of a flow  $f$  in a network with taxes  $\tau$  is the total disutility caused to network users, accounting for disutility due to both latency and taxes:

$$C(f, \tau) \equiv \sum_{P \in \mathcal{P}} [\ell_P(f) + \tau_P] f_P = \sum_{e \in E} [\ell_e(f_e) + \tau_e] f_e.$$

The functions  $L(\cdot)$  and  $C(\cdot, \tau)$  coincide if and only if  $\tau = 0$ .<sup>1</sup> We will call a triple  $(G, r, \ell)$  an *instance*, and will use the notation  $(G, r, \ell + \tau)$  to denote an instance in which taxes  $\tau$  have been levied on the edges of  $G$ .

## 2.2 Flows at Nash Equilibrium

We assume that noncooperative behavior results in a Nash equilibrium—a “stable point” in which no traffic has an incentive to unilaterally alter its strategy (i.e., its route from  $s$  to  $t$ ). We also assume that all agents seek to minimize the sum of the latency experienced and the tax paid.<sup>2</sup> We therefore expect that, in a flow at Nash equilibrium, all traffic is routed on paths with minimum-possible latency plus tax. Formally, we have the following definition.

**Definition 2.1** A flow  $f$  feasible for  $(G, r, \ell + \tau)$  is *at Nash equilibrium* or is a *Nash flow* if for all  $P_1, P_2 \in \mathcal{P}$  with  $f_{P_1} > 0$ ,

$$\ell_{P_1}(f) + \tau_{P_1} \leq \ell_{P_2}(f) + \tau_{P_2}.$$

---

<sup>1</sup>In several previous papers on selfish routing (without taxes), such as [31, 33], the terms total latency and cost were used synonymously.

<sup>2</sup>This assumption, while classical, is obviously quite strong. In general, we expect different agents to have different objective functions and to trade off time and money in different ways. This objection raises several interesting issues that are the subject of our ongoing work [7].

We will also make use of the following alternative definition of a Nash flow, which is a simple consequence of the fact that such a flow routes traffic only on shortest paths with respect to latencies plus taxes.

**Proposition 2.2** ([29]) *Let  $f$  be a flow feasible for  $(G, r, \ell + \tau)$ . For a vertex  $v$  in  $G$ , let  $d(v)$  denote the length, with respect to edge lengths  $\ell_e(f_e) + \tau_e$ , of a shortest  $s$ - $v$  path in  $G$ . Then*

$$d(w) - d(v) \leq \ell_e(f_e) + \tau_e$$

for all edges  $e = (v, w)$ , and  $f$  is at Nash equilibrium if and only if equality holds whenever  $f_e > 0$ .

An immediate consequence of Proposition 2.2 is that the property of being at Nash equilibrium is a property only of the flow vector on edges  $\{f_e\}$  induced by a flow  $f$ , rather than on the particular path decomposition.

**Corollary 2.3** *Suppose  $f, \tilde{f}$  are flows for  $(G, r, \ell + \tau)$  with  $f_e = \tilde{f}_e$  for all edges  $e$ . Then  $f$  is at Nash equilibrium if and only if  $\tilde{f}$  is at Nash equilibrium.*

Flows at Nash equilibrium enjoy several nice properties. We first consider the existence and uniqueness of Nash flows.

**Proposition 2.4** ([3, 10, 29, 33]) *Every instance  $(G, r, \ell + \tau)$  admits a directed acyclic flow at Nash equilibrium.*

**Proposition 2.5** ([3, 10, 33]) *If  $f, \tilde{f}$  are flows at Nash equilibrium for  $(G, r, \ell + \tau)$ , then:*

- (a)  $\ell_e(f_e) = \ell_e(\tilde{f}_e)$  for all edges  $e$ ;
- (b)  $C(f, \tau) = C(\tilde{f}, \tau)$ .

The next proposition, which follows easily from Definition 2.1 and the definition of cost, states that all paths used by a Nash flow have the same combined latency and tax and that the cost of a Nash flow is therefore expressible in a very simple form.

**Proposition 2.6** *If  $f$  is at Nash equilibrium for  $(G, r, \ell + \tau)$ , then there is a constant  $c \geq 0$  with  $\ell_P(f) + \tau_P = c$  whenever  $f_P > 0$ . Moreover,  $C(f, \tau) = r \cdot c$ .*

We adopt the notation  $c(G, r, \ell + \tau)$  for the value of the constant  $c$  in Proposition 2.6 for instance  $(G, r, \ell + \tau)$ , which is well defined by Propositions 2.4 and 2.5. Our final proposition states that, for fixed  $G, \ell$ , and  $\tau$ , the value of  $c(G, r, \ell + \tau)$  is nondecreasing in  $r$ . This fact was first proved by Hall [14] and we will use it in Section 3. For a combinatorial proof of the proposition, see Lin et al. [19].

**Proposition 2.7** ([14]) *The value  $c(G, r, \ell + \tau)$  is nondecreasing in  $r$ .*

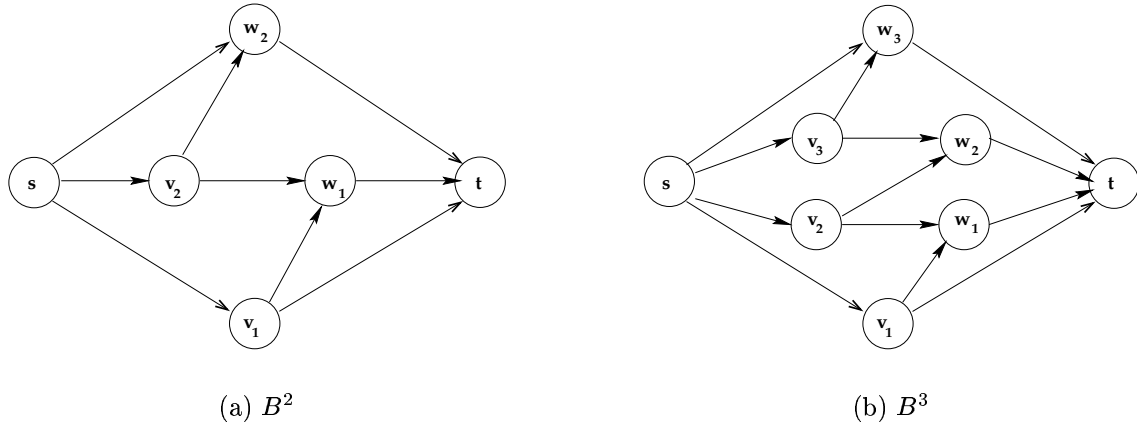


Figure 2: The second and third Braess graphs

### 2.3 The Braess Graphs

In this subsection we review the “Braess graphs”. These networks were first defined by Roughgarden [29] to show that removing edges from a network can decrease the total latency of a Nash flow by an unbounded amount. Since levying taxes can be viewed as a natural generalization of removing edges, these networks will also play an important role in this paper.

The  $k$ th Braess graph  $B^k = (V^k, E^k)$  is defined as follows: start with a set  $V^k = \{s, v_1, \dots, v_k, w_1, \dots, w_k, t\}$  of  $2k + 2$  vertices and define  $E^k$  by  $\{(s, v_i), (v_i, w_i), (w_i, t) : 1 \leq i \leq k\} \cup \{(v_i, w_{i-1}) : 2 \leq i \leq k\} \cup \{(v_1, t), (s, w_k)\}$  (see Figure 2). The graph  $B^1$  is the same as Braess’s original example (Figure 1).

Roughgarden [29] defined the following latency functions  $\ell^k$  on the graph  $B^k$ .

- (A) Edges of the form  $(v_i, w_i)$  are given the latency function  $\ell^k(x) = 0$ .
- (B) Edges of the form  $(v_i, w_{i-1}), (v_1, t), (s, w_k)$  are endowed with the latency function  $\ell^k(x) = 1$ .
- (C) Edges of the form  $(w_i, t)$  or  $(s, v_{k-i+1})$  receive a latency function  $\ell^k$  satisfying  $\ell^k(x) = 0$  for  $x \leq k/(k+1)$  and  $\ell^k(1) = i$ .

We will call edges of the form  $(v_i, w_i)$  *type A edges*, and so forth.

To describe flows in  $B^k$  in a convenient way, we require further notation. For  $i = 1, \dots, k$ , let  $P_i$  denote the path  $s \rightarrow v_i \rightarrow w_i \rightarrow t$ . For  $i = 2, \dots, k$ , let  $Q_i$  denote the path  $s \rightarrow v_i \rightarrow w_{i-1} \rightarrow t$ ; define  $Q_1$  to be the path  $s \rightarrow v_1 \rightarrow t$  and  $Q_{k+1}$  the path  $s \rightarrow w_k \rightarrow t$ .

Finally, we review the proof that the latency of a Nash flow in the instance  $(B^k, k, \ell^k)$  can be improved by a factor of  $k + 1$  by removing  $k$  edges from the network.

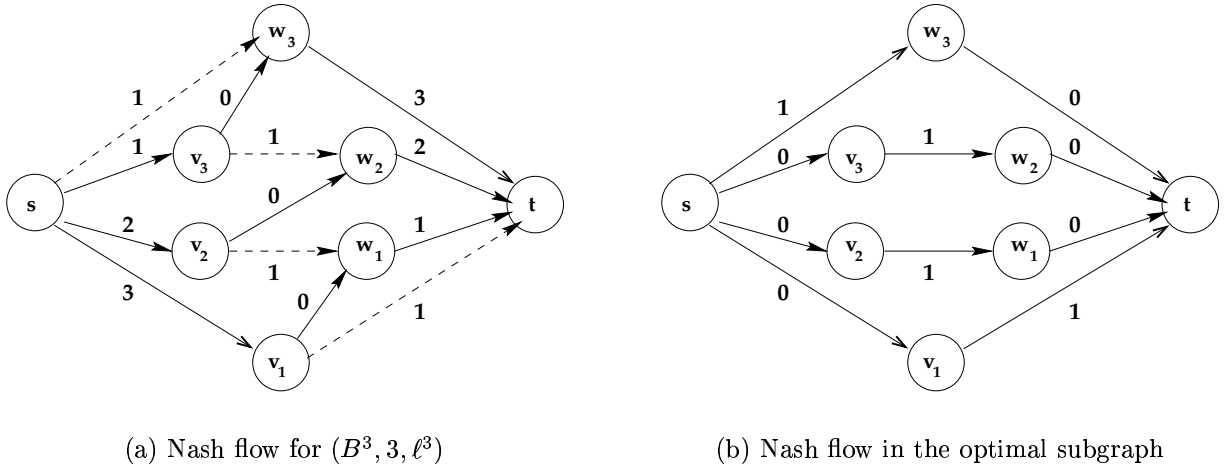


Figure 3: Proof of Proposition 2.8, when  $k = 3$ . Solid edges carry flow in the flow at Nash equilibrium, dashed edges do not. Edge latencies are with respect to flows at Nash equilibrium.

**Proposition 2.8** ([29]) *There is a subgraph  $H$  of  $B^k$  so that if  $f^*$  and  $f$  are Nash flows for  $(H, k, \ell^k)$  and  $(B^k, k, \ell^k)$ , respectively, then*

$$L(f^*) = \frac{L(f)}{k+1}.$$

*Proof:* Routing one unit of flow on each of  $P_1, \dots, P_k$  gives a flow at Nash equilibrium for  $(B^k, k, \ell^k)$  in which all traffic experiences latency  $k+1$  (see Figure 3(a)). On the other hand, if  $H$  is the subgraph obtained from  $B^k$  by deleting all type A edges, routing  $k/(k+1)$  units of flow on each of  $Q_1, \dots, Q_{k+1}$  yields a flow at Nash equilibrium for  $(H, k, \ell^k)$  with all traffic incurring latency 1 (see Figure 3(b)). ■

Since there are  $2k+2$  vertices in the  $k$ th Braess graph, we can rephrase Proposition 2.8 as follows.

**Corollary 2.9** ([29]) *Removing edges from a network with  $n$  vertices and continuous, non-decreasing latency functions can improve the total latency of a Nash flow by a factor of  $\lfloor n/2 \rfloor$ .*

We will also study networks in which the edge latency functions obey restrictions beyond continuity and monotonicity. We will focus on latency functions that are degree-bounded polynomials with nonnegative coefficients, though our techniques are also applicable to many other types of latency functions. Analogs of Proposition 2.8 and Corollary 2.9 for such latency functions again follow from the work of Roughgarden [29].

**Corollary 2.10** ([29]) *Removing edges from a network with latency functions that have degree at most  $p$  and nonnegative coefficients can improve the total latency of a Nash flow by an  $\Omega(p/\log p)$  factor.*

Corollary 2.10 can be proved with a modification of the construction used in Proposition 2.8. In fact, it suffices to consider the  $k$ th Braess graph with  $k \approx p/\log p$ . The traffic rate and latency functions are identical to those in the proof of Proposition 2.8, except that a type C edge of the form  $(w_i, t)$  or  $(s, v_{k-i+1})$  receives the latency function  $ix^p$ .

The constant implicit in the asymptotic lower bound of Corollary 2.10 is fairly small. A stronger lower bound for networks with linear latency functions follows from the network in Figure 1.

**Corollary 2.11** ([5, 29, 34]) *Removing edges from a network with linear latency functions can improve the total latency of a Nash flow by a factor of  $4/3$ .*

### 3 When Do Marginal Cost Taxes Help?

In this section, we study the cost of applying the principle of marginal cost pricing. In Subsection 3.1 we formalize marginal cost taxes and the classical guarantee that they induce the minimum-latency flow as a flow at Nash equilibrium. In Subsection 3.2, however, we show that marginal cost taxes cannot decrease the cost of a Nash flow in a network with linear latency functions.

#### 3.1 Marginal Cost Taxes

Recall that the principle of marginal cost pricing posits that each user should pay a tax equal to the additional delay other users experience because of its presence. Mathematically, this principle asserts that in a flow  $f$  feasible for instance  $(G, r, \ell)$ , the tax  $\tau_e$  assigned to edge  $e$  should be  $\tau_e = f_e \cdot \ell'_e(f_e)$  where  $\ell'_e$  denotes the derivative of  $\ell_e$  (assume for simplicity that the latency functions are differentiable). The term  $\ell'_e(f_e)$  corresponds to the marginal increase in latency caused by one user on the edge, and the term  $f_e$  is the amount of traffic that suffers from this increase. Marginal cost taxes come with the following guarantee, which is classical (see [3, 10, 30] for proofs).

**Proposition 3.1** *Let  $(G, r, \ell)$  be an instance with differentiable latency functions, admitting minimum-latency flow  $f^*$ . Let  $\tau_e = f_e^* \cdot \ell'_e(f_e^*)$  denote the marginal cost tax for edge  $e$  w.r.t.  $f^*$ . Then  $f^*$  is at Nash equilibrium for  $(G, r, \ell + \tau)$ .*

In words, marginal cost taxes induce the minimum-latency flow as a flow at Nash equilibrium.

#### 3.2 Networks with Linear Latency Functions

Proposition 3.1 gives an effective way to minimize the total latency of a Nash flow with edge taxes. But how effective are marginal cost taxes when we also account for the disutility to traffic due to taxes? Our next theorem identifies a reasonably large class of networks—networks in which all latency functions are linear, with the form  $\ell(x) = ax + b$ —in which marginal cost taxes are *guaranteed* to be unnecessary, if not detrimental. This theorem thus illustrates the dangers of marginal cost pricing when minimizing the total latency is not the sole goal.

**Theorem 3.2** *Let  $(G, r, \ell)$  be an instance with linear latency functions admitting minimum-latency flow  $f^*$  and Nash flow  $f$ . Let  $\tau$  denote the marginal cost taxes with respect to  $f^*$ . Then,*

$$L(f) = C(f, 0) \leq C(f^*, \tau).$$

*Proof:* Let  $(G, r, \ell)$  be an instance with linear latency functions, admitting minimum-latency flow  $f^*$  and Nash flow  $f$ . Write  $\ell_e(x) = a_e x + b_e$  for each edge  $e$ . The principle of marginal cost pricing dictates that  $\tau_e = f_e^* \cdot \ell'_e(f_e^*) = a_e f_e^*$ .

Define the modified latency function  $\ell_e^*$  by  $\ell_e^*(x) = 2a_e x + b_e$ . The functions  $\ell + \tau$  and  $\ell^*$  are not identically equal, but the identity  $\ell_e^*(f_e^*) = 2a_e f_e^* + b_e = \ell_e(f_e^*) + \tau_e$  holds for every edge  $e$ . It follows that  $f^*$  is at Nash equilibrium not only for the instance  $(G, r, \ell + \tau)$  (as guaranteed by Proposition 3.1), but also for the instance  $(G, r, \ell^*)$ . Moreover, in the notation of Proposition 2.6,  $c(G, r, \ell^*) = c(G, r, \ell + \tau)$ .

We next claim that  $f/2$  is at Nash equilibrium for  $(G, r/2, \ell^*)$  with  $c(G, r/2, \ell^*) = c(G, r, \ell)$ .<sup>3</sup> To see why, we note that since  $\ell_e(x) = a_e x + b_e$  and  $\ell_e^*(x) = 2a_e x + b_e$ , edge and path latencies with respect to  $f/2$  in  $(G, r/2, \ell^*)$  and with respect to  $f$  in  $(G, r, \ell)$  are identical. That  $f$  is a Nash flow for  $(G, r, \ell)$  then implies that  $f/2$  is at Nash equilibrium for  $(G, r/2, \ell^*)$ , with  $c(G, r/2, \ell^*) = c(G, r, \ell)$ .

Combining the two equalities of the previous paragraphs with an application of Proposition 2.7, we obtain

$$c(G, r, \ell) = c(G, r/2, \ell^*) \leq c(G, r, \ell^*) = c(G, r, \ell + \tau).$$

The theorem now follows from Proposition 2.6. ■

**Remark 3.3** Theorem 3.2 concerns only networks with linear latency functions, but it is easily extended to networks in which, for some fixed  $p \geq 0$ , every edge  $e$  has a latency function of the form  $a_e x^p + b_e$  with  $a_e, b_e \geq 0$ .

**Remark 3.4** There are networks with nonlinear latency functions in which marginal cost taxes can decrease the cost of a Nash flow. For example, it is possible to define such an example using the network of Figure 1. The latency functions employed are non-convex, and similar to step functions. The details are somewhat tedious and we omit them. We do not view this example as a positive result for marginal cost taxes (as the example is contrived), but rather as justification for restricting the network latency functions in Theorem 3.2 to achieve a somewhat sweeping condemnation of marginal cost taxes for minimizing the cost of a Nash flow. We leave open the question of whether the negative result of Theorem 3.2 can be proved under significantly weaker assumptions on the latency functions.

## 4 How Powerful Are Arbitrary Taxes?

In this section we study the following question: how much can the cost of a Nash flow improve after levying taxes on the edges? As we will see, a precise answer to this question follows easily from previous work on network design with selfish routing [29].

---

<sup>3</sup>This claim was in essence proved in [33].

The maximum-possible benefit from taxes will depend crucially on the type of network latency functions that we allow. This dependence is characteristic of much of the work on selfish routing [30]. Indeed, we have already seen a glimpse of such a dependence in the previous section, where marginal cost pricing can decrease the cost of a Nash flow, but not in networks with linear latency functions.

We will give two types of upper bounds on the maximum-possible reduction in cost due to taxes, depending on the generality of the network latency functions. The first two upper bounds are corollaries of previous work bounding the *price of anarchy* [26], defined as the largest possible ratio between the total latency of a Nash flow and of a minimum-latency flow. The price of anarchy is a function of the class of allowable latency functions, and this dependence is by now well understood. For example, the following statements are known (see e.g. [31] for further examples).

**Proposition 4.1** ([33]) *The price of anarchy in networks with linear latency functions is  $4/3$ .*

**Proposition 4.2** ([31]) *The price of anarchy in networks with latency functions that are polynomials with degree at most  $p$  and nonnegative coefficients is asymptotically  $\Theta(\frac{p}{\log p})$  as  $p \rightarrow \infty$ .*

Upper bounds on the price of anarchy translate directly to upper bounds on the largest decrease in cost achievable with taxes: at best, taxes replace the Nash flow in the original network with the minimum-latency flow, while causing no additional disutility to network users. We therefore have the following corollaries.

**Corollary 4.3** *Let  $(G, r, \ell)$  be an instance with linear latency functions and  $\tau$  a tax on edges. Let  $f$  and  $f^\tau$  be Nash flows for  $(G, r, \ell)$  and  $(G, r, \ell + \tau)$ , respectively. Then  $L(f) \leq \frac{4}{3} \cdot C(f^\tau, \tau)$ .*

**Corollary 4.4** *For some constant  $c_1 > 0$  the following statement holds for  $p \geq 2$ . If  $(G, r, \ell)$  is an instance with polynomial latency functions with degree at most  $p$  and nonnegative coefficients,  $\tau$  is a tax on edges, and  $f$  and  $f^\tau$  are Nash flows for  $(G, r, \ell)$  and  $(G, r, \ell + \tau)$ , respectively, then  $L(f) \leq c_1 \frac{p}{\log p} \cdot C(f^\tau, \tau)$ .*

In Subsection 2.3 we reviewed the Braess graphs, in which deleting edges can dramatically improve the total latency of a Nash flow. As we have noted, sufficiently large taxes can simulate edge deletions, so these examples translate directly into the present setting. In particular, Corollary 2.11 implies that the bound of Corollary 4.3 is the best possible, and Corollary 2.10 demonstrates that Corollary 4.4 is sharp up to a constant factor.

In networks with general latency functions, the price of anarchy is unbounded, even after restricting attention to two-node, two-link networks [33]. However, while a bounded price of anarchy is *sufficient* for a bound on the largest-possible benefit of taxes, it is not *necessary*. Indeed, with no assumptions whatsoever on the network latency functions (other than continuity and monotonicity), we can still obtain an upper bound that is a function of the network size.

**Theorem 4.5** *Let  $(G, r, \ell)$  be an instance with  $n$  vertices and  $\tau$  a tax on edges. Let  $f$  and  $f^\tau$  be Nash flows for  $(G, r, \ell)$  and  $(G, r, \ell + \tau)$ , respectively. Then*

$$L(f) \leq \left\lfloor \frac{n}{2} \right\rfloor \cdot C(f^\tau, \tau).$$

The proof of Theorem 4.5 is a straightforward extension of an argument from [29, Theorem 4.1] proving the weaker statement that deleting edges from a network can improve the total latency of a Nash flow by at most an  $\lfloor n/2 \rfloor$  factor. We include a proof sketch in Appendix A for completeness. Corollary 2.9 implies that the bound of Theorem 4.5 is the best possible.

## 5 Are Taxes More Powerful Than Edge Removal?

In Section 4, we saw that there is a strong connection between the power of taxation and the power of edge removal. Specifically, we found that for several natural classes of networks, the maximum-possible reduction in cost achievable by levying taxes on edges is the same as that by removing edges from the network. We have therefore resolved in the negative, for these classes of networks, the “best-case” version of the following question: are taxes strictly more powerful than edge removal? However, we have not resolved whether taxes can *ever* improve upon the best solution obtainable by removing edges. We study this issue in this section. In Subsection 5.1 we show that in *every* network with linear latency functions, taxes cannot improve the Nash flow more than the removal of some set of edges. By contrast, in Subsection 5.2 we show that in general networks, taxes can improve the Nash flow far beyond what is achievable by merely deleting edges from the network.

### 5.1 The Power of Edge Removal for Networks with Linear Latency Functions

#### 5.1.1 Overview

In this subsection we consider only networks with linear latency functions (of the form  $\ell(x) = ax + b$ ), and show that taxes are never more powerful than removing edges for these networks. To state this formally, we will say that a set  $\tau$  of taxes for the instance  $(G, r, \ell)$  is  $0/\infty$  if, for some Nash flow  $f^\tau$  for  $(G, r, \ell + \tau)$ ,  $\tau_e = 0$  or  $f_e^\tau = 0$  for each edge  $e$ . We note that  $0/\infty$  taxes are no more powerful than removing edges from the network, since if  $\tau$  is  $0/\infty$  then  $c(G, r, \ell + \tau) = c(H, r, \ell)$ , where  $H$  is the subgraph of  $G$  consisting of the edges with zero tax.

It will be convenient to abuse notation and, with respect to an instance  $(G, r, \ell)$ , to write  $C(\tau)$  for  $C(f^\tau, \tau)$ , where  $f^\tau$  is at Nash equilibrium for  $(G, r, \ell + \tau)$ . We note that  $C(\tau)$  is well defined by Proposition 2.5(b). A tax  $\tau$  is *optimal* for an instance  $(G, r, \ell)$  if it minimizes  $C(\tau)$  over all nonnegative tax vectors  $\tau$ .

**Theorem 5.1** *An instance with linear latency functions admits an optimal set of taxes that is  $0/\infty$ .*

The proof of Theorem 5.1 is somewhat involved. To avoid considering an arbitrary network with linear latency functions, our proof argues by contradiction and studies a minimal counterexample. As we will see, minimal counterexamples possess several convenient properties that facilitate the argument.

Precisely, an instance  $(G, r, \ell)$  is a *counterexample* if it admits no optimal  $0/\infty$  tax. A counterexample is *minimal* if it has as few edges as any other counterexample. A tax  $\tau$  is *good* for  $(G, r, \ell)$  if  $C(\tau) < C(\hat{\tau})$  for all  $0/\infty$  taxes  $\hat{\tau}$ . If  $(G, r, \ell)$  admits a good tax, then it is clearly a counterexample. We are not yet claiming that a counterexample  $(G, r, \ell)$  admits an optimal tax; conceivably, the infimum in  $\inf\{C(\tau) : \tau \geq 0\}$  is not attained by any tax. However, since the set  $\{C(\tau) : \tau \text{ is } 0/\infty\}$  is finite, every counterexample admits a good tax.

To achieve our desired contradiction, we will also need a tax that is in some sense minimal. Formally, we will call a tax *minimal* for an instance if it is an optimal tax and minimizes  $\sum_{e \in E} \tau_e$  among all optimal taxes  $\tau$ . Once we establish that minimal taxes exist for minimal counterexamples, the proof of Theorem 5.1 will be relatively short. Our key lemma is thus the following.

**Lemma 5.2** *A minimal counterexample admits a minimal tax.*

### 5.1.2 Properties of Minimal Counterexamples

To prove Lemma 5.2, we will require several properties of a minimal counterexample. We prove these next.

**Lemma 5.3** *Let  $(G, r, \ell)$  be a minimal counterexample,  $\tau$  a good tax, and  $f^\tau$  at Nash equilibrium for  $(G, r, \ell + \tau)$ . Then  $f_e^\tau > 0$  for all edges  $e$  of  $G$ .*

*Proof:* If  $f_e^\tau = 0$  for some edge  $e$  of  $G$ , then deleting  $e$  from  $G$  yields a counterexample with fewer edges. ■

**Corollary 5.4** *If  $(G, r, \ell)$  is a minimal counterexample, then  $G$  is directed acyclic.*

*Proof:* The corollary follows immediately from Proposition 2.4 and Lemma 5.3. ■

**Lemma 5.5** *If  $(G, r, \ell)$  is a minimal counterexample,  $\tau$  is a tax, and  $f^\tau$  is a Nash flow for  $(G, r, \ell + \tau)$ , then  $f$  is a Nash flow for  $(G, r, \ell + \tau)$  if and only if  $f_e = f_e^\tau$  for all edges  $e$ .*

*Proof:* The “if” direction is Corollary 2.3. For the “only if” direction, suppose for contradiction that  $f$  is another Nash flow for  $(G, r, \ell + \tau)$ . Put  $z_e = f_e - f_e^\tau$ . Since  $f$  and  $f^\tau$  are flows at the same traffic rate,  $z$  is a (signed) circulation: for each vertex  $v$  with edges  $\delta^+(v)$  having tail  $v$  and edges  $\delta^-(v)$  having head  $v$ ,  $\sum_{e \in \delta^+(v)} z_e = \sum_{e \in \delta^-(v)} z_e$ . By Proposition 2.5(a),  $z_e$  is non-zero only when  $e$  has a constant latency function.

Let  $\hat{\tau}$  be a good tax for  $(G, r, \ell)$  and  $f^{\hat{\tau}}$  a Nash flow for  $(G, r, \ell + \hat{\tau})$ . Since  $f_e^{\hat{\tau}} > 0$  for all edges  $e$  by Lemma 5.3,  $\{f_e^{\hat{\tau}} \pm \lambda z_e\}$  is a nonnegative vector on edges for sufficiently small  $\lambda$ ; in this case, it corresponds (after a path decomposition) to a flow feasible for  $(G, r, \ell + \hat{\tau})$ . Moreover, since  $z_e$  is non-zero only when  $e$  possesses a constant latency function, Proposition 2.2 implies that this flow is at Nash equilibrium for  $(G, r, \ell + \hat{\tau})$ . Since we assumed that some  $z_e$  is non-zero, choosing  $\lambda$  appropriately yields a Nash flow  $\tilde{f}$  for  $(G, r, \ell + \hat{\tau})$  with  $\tilde{f}_e = 0$  for some edge  $e$ , which contradicts Lemma 5.3. ■

**Lemma 5.6** *Let  $(G, r, \ell)$  be a minimal counterexample,  $\tau$  a good tax, and  $f^\tau$  a Nash flow for  $(G, r, \ell + \tau)$ . Then there is a constant  $c^\tau$  such that*

$$\ell_P(f^\tau) + \tau_P = c^\tau$$

for all  $s$ - $t$  paths  $P$  of  $G$ .

*Proof:* By Lemma 5.3, we can choose a path decomposition  $f$  of  $\{f_e^\tau\}$  with  $f_P > 0$  for all paths  $P \in \mathcal{P}$ ; the lemma now follows from Corollary 2.3 and Proposition 2.6. ■

### 5.1.3 Proof of Lemma 5.2

Recall that our key lemma, Lemma 5.2, asserts that a minimal counterexample admits a minimal tax. Our proof of this result requires two technical lemmas. The first states that bounded taxes suffice to minimize the cost of a Nash flow in a minimal counterexample.

**Lemma 5.7** *Let  $(G, r, \ell)$  be a minimal counterexample and  $\tau$  a good tax. Let  $G$  have  $n$  vertices and define  $\ell_{max} = \max_{e \in E} \ell_e(r)$ . There is a tax  $\tilde{\tau}$  with  $\max_e \tilde{\tau}_e \leq n\ell_{max}$  and  $C(\tilde{\tau}) \leq C(\tau)$ .*

*Proof:* Let  $\tau$  be a good tax for  $(G, r, \ell)$ , with  $f^\tau$  at Nash equilibrium for  $(G, r, \ell + \tau)$ . We next show how to decrease taxes while leaving  $f^\tau$  at Nash equilibrium.

By Corollary 5.4,  $G$  is directed acyclic and we can therefore topologically sort the vertices of  $G$  so that all edges of  $G$  travel forward. Since  $f_e^\tau > 0$  on all edges of  $G$  (Lemma 5.3),  $s$  is the first vertex in the ordering and  $t$  is the last. Beginning with the penultimate vertex and proceeding backward in the vertex ordering, we perform the following operation for each vertex  $v \neq s$ : if  $\tau_v \geq 0$  is the minimum tax on any edge with tail  $v$ , subtract  $\tau_v$  from the tax on each edge with tail  $v$  and add  $\tau_v$  to the tax on each edge with head  $v$ . This operation does not affect the total tax on any  $s$ - $t$  path nor the cost of any flow and leaves  $f^\tau$  at Nash equilibrium. When the source  $s$  is reached, subtract  $\tau_s$  from the tax on all edges with tail  $s$ ;  $f^\tau$  remains at Nash equilibrium and its cost can only decrease. Call the new set of taxes  $\tilde{\tau}$ . We have already argued that  $C(\tilde{\tau}) \leq C(\tau)$ ; it remains to show that  $\max_e \tilde{\tau}_e \leq n\ell_{max}$ .

We first observe that the tax-reducing operations enforce the following property: every vertex other than  $t$  has an outgoing edge with zero  $\tilde{\tau}$ -tax. From this, it follows that an entire  $s$ - $t$  path, say  $P_0$ , has zero  $\tilde{\tau}$ -tax. Since  $C(\tilde{\tau}) \leq C(\tau)$ ,  $\tilde{\tau}$  is a good tax and hence Lemma 5.6 implies that

$$\ell_P(f^\tau) + \tilde{\tau}_P = \ell_{P_0}(f^\tau)$$

for all paths  $P \in \mathcal{P}$ . Since  $\ell_{P_0}(f^\tau) \leq n\ell_{max}$  and every edge of  $G$  lies on some  $s$ - $t$  path (by Lemma 5.3), no edge tax in  $\tilde{\tau}$  can exceed  $n\ell_{max}$ , and the proof is complete. ■

The second technical lemma asserts continuity of the map  $\tau \mapsto C(\tau)$ .

**Lemma 5.8** *Let  $(G, r, \ell)$  be a minimal counterexample. Then the corresponding map  $\tau \mapsto C(\tau)$  is continuous.*

*Proof:* Lemma 5.5 ensures that the map  $\tau \mapsto \{f_e^\tau\}$  is well defined. Lemma 5.5 and a result of Dafermos and Nagurney [9, Theorem 3.1] imply that this map is continuous. The lemma then follows easily. ■

We are finally prepared to prove Lemma 5.2.

*Proof of Lemma 5.2:* We first show the weaker statement that a minimal counterexample  $(G, r, \ell)$  admits an optimal tax—that the infimum  $\inf_{\tau \geq 0} C(\tau)$  is attained. Let  $B$  denote the taxes with all components bounded by  $n\ell_{max}$ , where  $n$  is the number of vertices of  $G$  and  $\ell_{max} = \max_{e \in E} \ell_e(r)$ . Since  $(G, r, \ell)$  is a counterexample, the infimum  $\inf_{\tau \geq 0} C(\tau)$  is approached by good taxes. Since  $(G, r, \ell)$  is minimal, Lemma 5.7 implies that this infimum is approached by good taxes in  $B$ —that is,  $\inf_{\tau \in B} C(\tau) = \inf_{\tau \geq 0} C(\tau)$ . Since  $B$  is a compact subset of  $\mathcal{R}^E$  and  $C$  is continuous by Lemma 5.8, the infimum in  $\inf_{\tau \in B} C(\tau)$ , and hence in  $\inf_{\tau \geq 0} C(\tau)$ , is attained.

Let  $O$  denote the (non-empty) set of optimal taxes for  $(G, r, \ell)$ . We will show that the infimum  $\inf_{\tau \in O} \sum_e \tau_e$  is attained. By Lemma 5.7, there is some optimal tax  $\tau \in O$  with  $\sum_{e \in E} \tau_e \leq mn\ell_{max}$  where  $m$  is the number of edges of  $G$ . Since  $O$  is the inverse image of a closed set under a continuous map (write  $O = C^{-1}(C(\tau))$  for some  $\tau \in O$ ), it is a closed subset of  $\mathcal{R}^E$ . Restricting  $O$  to taxes with sum of all components at most  $mn\ell_{max}$ , we obtain a nonempty compact subset  $S$  of optimal taxes. Since  $\tau \mapsto \sum_e \tau_e$  is a continuous function, it attains a minimum on  $S$ ; this must also be its minimum on  $O$ , and the proof is complete. ■

### 5.1.4 Proof of Theorem 5.1

With Lemma 5.2 in hand, we can now prove Theorem 5.1 by a perturbation argument.

*Proof of Theorem 5.1:* To derive a contradiction, let  $(G, r, \ell)$  be a minimal counterexample,  $\tau$  a minimal (and hence good) tax, and  $f^\tau$  a Nash flow for  $(G, r, \ell + \tau)$ . Since  $(G, r, \ell)$  is a counterexample,  $\sum_e \tau_e > 0$ . By Lemma 5.6, there is a constant  $c^\tau$  such that  $\ell_P(f^\tau) + \tau_P = c^\tau$  for all  $s$ - $t$  paths  $P$  of  $G$ .

Write  $\ell_e(x) = a_e x + b_e$  for each edge  $e$  of  $G$ . The equations

$$\sum_{e \in P} [a_e f_e + b_e + \tau_e] = c$$

for all  $P \in \mathcal{P}$ , together with the standard flow conservation constraints for  $f$ , form a system of equations linear in the  $m + 1$  variables  $\{f_e\}$  and  $c$  (where  $m$  is the number of edges in  $G$ ). By Lemma 5.5,  $(\{f_e^\tau\}, c^\tau)$  is the unique solution to this system, with flow nonnegativity constraints automatically satisfied (indeed, strictly by Lemma 5.3). Choosing  $m + 1$  linearly independent constraints with at least one constraint corresponding to a path with nonzero tax, there is a square linear system

$$A \begin{bmatrix} f \\ c \end{bmatrix} = d$$

for which  $\{f_e^\tau\}, c^\tau$  is the unique solution, namely  $A^{-1}d$ .

Let  $\tau_{\bar{e}} > 0$  be a positive edge tax appearing in this linear system (necessarily as part of the right-hand side  $d$ ). Consider perturbing this tax by subtracting a small number  $\epsilon > 0$ . This

translates to a perturbation of adding  $\epsilon$  to the right-hand side of all constraints corresponding to paths that include the edge  $\bar{e}$ . Since  $f_e^\tau > 0$  for all edges  $e$ , for perturbations sufficiently close to 0 the new Nash flow is given by  $(\{\bar{f}_e\}, \bar{c}) = A^{-1}\bar{d}$ , where  $\bar{d}$  is the old right-hand side  $d$  following the perturbation. Since  $\tau$  is an optimal tax minimizing  $\sum_e \tau_e$ , subtracting  $\epsilon > 0$  from  $\tau_{\bar{e}}$  must produce a non-optimal tax. By Proposition 2.6, it follows that  $c^\tau < \bar{c}$ . By linearity, however, the opposite perturbation of adding  $\epsilon$  to  $\tau_{\bar{e}}$  has the opposite effect, producing a tax  $\bar{\tau}$  inducing a solution  $(\{\bar{f}_e\}, \bar{c})$  with  $\bar{c} < c^\tau$ . This contradicts the optimality of  $\tau$ , and the proof is complete. ■

## 5.2 The Power of Taxes in General Networks

In the previous subsection we saw that, in networks with linear latency functions, taxes never improve over the best solution attainable by removing edges from the network. In this subsection we will demonstrate that this result does not extend to general networks: in fact, for each value of  $n \geq 2$ , there is an  $n$ -node network in which arbitrary nonnegative taxes can improve upon  $0/\infty$  taxes by an  $\lfloor n/2 \rfloor$  factor. With this result, we will have a good understanding of the relationship between taxes and edge removal in general networks. Briefly, removing edges can improve the Nash flow by an  $\lfloor n/2 \rfloor$  factor (Corollary 2.9), taxes can improve a Nash flow by an  $\lfloor n/2 \rfloor$  factor beyond what is achievable by removing edges (Theorem 5.9 below), but their combined power (equivalently, the power of taxes) cannot improve a Nash flow by more than an  $\lfloor n/2 \rfloor$  factor (Theorem 4.5).

**Theorem 5.9** *For each integer  $n \geq 2$ , there is an instance  $(G, r, \ell)$  with  $c(H, r, \ell) \geq \lfloor n/2 \rfloor$  for all subgraphs  $H$  of  $G$  but  $c(G, r, \ell + \tau) = 1$  for some tax  $\tau \geq 0$ .*

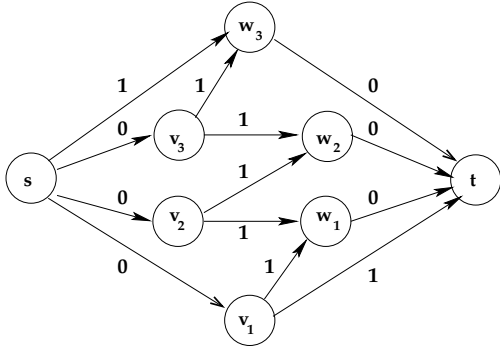
*Proof:* The construction is similar to that in an inapproximability result for network design [29, Theorem 4.3]. We can assume that  $n$  is even (for  $n$  odd, add an isolated vertex or subdivide an edge) and at least 4. We will work with the Braess graph  $B^k$  for which  $2k + 2 = n$  (see Subsection 2.3).

We define latency functions  $\ell^k$  on the graph  $B^k$  as follows.

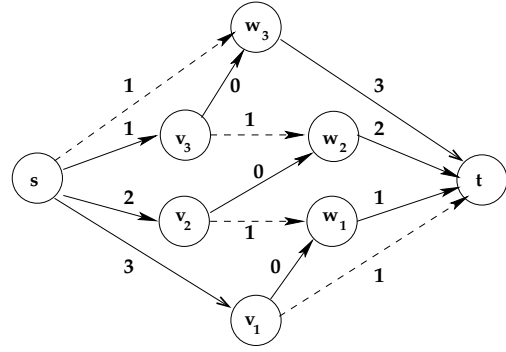
- (A) Edges of the form  $(v_i, w_i)$  receive the latency function  $\ell^k(x) = 0$ .
- (B) Edges of the form  $(v_i, w_{i-1}), (v_1, t), (s, w_k)$  receive a latency function  $\ell^k$  satisfying  $\ell^k(x) = 1$  for  $x \leq 1/(k+1)$  and  $\ell^k(x) = n/2$  for  $x \geq 1/(k+1) + \epsilon$ , where  $\epsilon > 0$  is a sufficiently small constant.
- (C) Edges of the form  $(w_i, t)$  or  $(s, v_{k-i+1})$  receive a latency function  $\ell^k$  satisfying  $\ell^k(x) = 0$  for  $x \leq 1 + 1/(k+1)$ ,  $\ell^k(1 + 1/k) = i$ , and  $\ell^k(x) = n/2$  for  $x \geq 1 + 1/k + \epsilon$ .

If a type B edge carries at least  $1/(k+1) + \epsilon$  units of flow or a type C edge carries at least  $1 + 1/k + \epsilon$  units of flow, we will say that the edge is *oversaturated*. A simple but important observation is that if a Nash flow oversaturates an edge in  $(H, r, \ell^k)$  for some subgraph  $H$  of  $B^k$ , then  $c(H, r, \ell^k) \geq n/2$ .

If we put a tax  $\tau$  of 1 on each edge of the form  $(v_i, w_i)$  and 0 elsewhere, then the following flow is at Nash equilibrium for  $(B^k, k+1, \ell^k + \tau)$ : 1 unit of flow is assigned to each of



(a) Latencies plus taxes in a Nash flow for  $(B^k, k + 1, \ell^k + \tau)$



(b) Latencies in a Nash flow for  $(B^k, k + 1, \ell^k)$

Figure 4: Proof of Theorem 5.9 when  $k = 3$ . Solid edges carry flow in the flow at Nash equilibrium, dashed edges do not. Edges are labeled with their cost (sum of latency and tax), where latencies are with respect to flows at Nash equilibrium.

$P_1, P_2, \dots, P_k$ , and  $\frac{1}{k+1}$  units of flow are assigned to each of  $Q_1, Q_2, \dots, Q_{k+1}$  (see Figure 4(a)). (See Subsection 2.3 for notation.) This Nash flow proves that  $c(B^k, k + 1, \ell^k + \tau) = 1$ .

To finish the proof, we must show that  $c(H, k + 1, \ell^k) \geq n/2$  for every subgraph  $H$  of  $B^k$ ; this requires a bit of case analysis. If  $H$  is all of  $B^k$ , then  $c(H, k + 1, \ell^k) = n/2$  because routing  $1 + 1/k$  units of traffic on each of  $P_1, P_2, \dots, P_k$  provides a flow at Nash equilibrium (see Figure 4(b)). This is equally true if  $H$  omits only edges eschewed by this Nash flow—the type B edges.

So suppose  $H$  omits some type C edge, and let  $\epsilon > 0$  be sufficiently small. If edge  $(s, v_i)$  is not in  $H$ , then any flow in  $H$  must oversaturate some edge incident to  $s$ . Thus  $c(H, k + 1, \ell^k) \geq n/2$  if  $H$  omits a type C edge incident to  $s$ . A symmetric argument applies to subgraphs  $H$  that omit a type C edge incident to  $t$ .

Finally, suppose  $H$  omits some type A arc, say  $(v_i, w_i)$ . The vertex  $v_i$  then has at most one outgoing edge in  $H$ , which must be a type B edge. If this edge is not oversaturated, then the edge  $(s, v_i)$  carries at most  $1/(k + 1) + \epsilon$  units of flow; as in the previous paragraph, this implies that some edge incident to  $s$  is oversaturated. In either case,  $c(H, k + 1, \ell^k) \geq n/2$  and the proof is complete. ■

## 6 The Complexity of Computing Optimal Taxes

In this section, we study the optimization problem of minimizing the cost of a Nash flow by taxing the network edges. We will see that the (in)approximability of this problem can be resolved by extending previous work on network design [29]. Unfortunately, we will discover that this optimization problem is intractable in a strong sense.

By an  $\alpha$ -approximation algorithm for a minimization problem, we mean an algorithm

that runs in polynomial time and returns a solution no more than  $\alpha$  times as costly as an optimal solution. We will call the value  $\alpha$  the *approximation ratio* or *performance guarantee* of the algorithm.

The maximum-possible benefit achievable with taxes, as determined in Section 4, has immediate consequences for the performance guarantee of the *trivial algorithm*—the algorithm that assigns all edges zero tax. In particular, Corollary 4.3, Corollary 4.4, and Theorem 4.5 imply the following.

**Corollary 6.1** *The trivial algorithm is a:*

- (a)  $\frac{4}{3}$ -approximation algorithm for the problem of taxing edges to minimize the cost of a Nash flow in networks with linear latency functions
- (b)  $\Theta(p/\log p)$ -approximation algorithm for the problem of taxing edges to minimize the cost of a Nash flow in networks with latency functions that are polynomials with degree at most  $p$  and nonnegative coefficients
- (c)  $\lfloor n/2 \rfloor$ -approximation algorithm for the problem of taxing edges to minimize the cost of a Nash flow in  $n$ -node networks with arbitrary latency functions.

Roughgarden [29] gave several inapproximability results for the problem of minimizing the cost of a Nash flow with  $0/\infty$  taxes. Using the same constructions in [29] and somewhat more involved arguments, most of these inapproximability results can be extended to the problem of computing optimal taxes. For example, we have the following theorems.

**Theorem 6.2** *For every  $\epsilon > 0$ , there is no  $(\frac{4}{3} - \epsilon)$ -approximation algorithm for the problem of taxing edges to minimize the cost of a Nash flow in networks with linear latency functions (unless  $P = NP$ ).*

**Theorem 6.3** *There is no  $o(p/\log p)$ -approximation algorithm for the problem of taxing edges to minimize the cost of a Nash flow in networks with latency functions that are polynomials with degree at most  $p$  and nonnegative coefficients (unless  $P = NP$ ).*

**Theorem 6.4** *For every  $\epsilon > 0$ , there is no  $O(n^{1-\epsilon})$ -approximation algorithm for the problem of taxing edges to minimize the cost of a Nash flow in  $n$ -node networks with arbitrary latency functions (unless  $P = NP$ ).*

For completeness, we give a proof of Theorem 6.2 in Appendix B. Theorems 6.3 and 6.4 follow from the constructions in [29] by similar but more tedious arguments.

## 7 Directions for Further Research

We have undertaken the first study of using edge taxes to minimize the cost—latency plus taxes paid—of a Nash flow. While we have answered several basic questions, some obvious gaps in our results remain. We next list three of the more glaring ones.

- (Q1) Suppose we restrict the edge latency functions of a network beyond continuity and monotonicity, but allow nonlinear latency functions. When can marginal cost taxes improve the cost of a Nash flow? When can levying taxes decrease the cost of a Nash flow beyond what is achievable by removing edges?
- (Q2) It is known [29, Theorem 4.3] that the problem of minimizing the cost of a Nash flow in a general network with  $0/\infty$  taxes is hard to approximate to within an  $\lfloor n/2 \rfloor - \epsilon$  factor. Is the problem equally difficult with arbitrary taxes?
- (Q3) Which results of this paper remain true in multicommodity flow networks, with multiple sources and destinations?

It would also be interesting to study other approaches to simultaneously minimizing the total latency and the taxes paid by traffic, especially those for which non-trivial algorithmic results are possible (cf. Section 6). For example, one could attempt to minimize the total latency subject to obeying an upper bound on the taxes paid by the traffic.

## Acknowledgements

We thank four anonymous referees for their comments on an earlier draft of this paper, and David Parkes for suggesting that alternative formulations of the problem could elude the complexity results of Section 6.

## References

- [1] A. Archer and É. Tardos. Truthful mechanisms for one-parameter agents. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 482–491, 2001.
- [2] A. Archer and É. Tardos. Frugal path mechanisms. In *Proceedings of the 13th Annual Symposium on Discrete Algorithms*, pages 991–999, 2002.
- [3] M. Beckmann, C. B. McGuire, and C. B. Winsten. *Studies in the Economics of Transportation*. Yale University Press, 1956.
- [4] P. Bergendorff, D. W. Hearn, and M. V. Ramana. Congestion toll pricing of traffic networks. In P. M. Pardalos, D. W. Hearn, and W. W. Hager, editors, *Network Optimization*, pages 51–71. Springer-Verlag, 1997.
- [5] D. Braess. Über ein paradoxon der verkehrsplanung. *Unternehmensforschung*, 12:258–268, 1968.
- [6] K. J. Button and E. T. Verhoef. *Road Pricing, Traffic Congestion and the Environment*. Edward Elgar, 1998.

- [7] R. Cole, Y. Dodis, and T. Roughgarden. Pricing network edges for heterogeneous selfish users. In *Proceedings of the 35th Annual ACM Symposium on the Theory of Computing*, pages 521–530, 2003.
- [8] J. R. Correa, A. S. Schulz, and N. E. Stier Moses. Selfish routing in capacitated networks. Technical Report 4319-03, MIT Sloan Working Paper, 2003. Preliminary version in *SODA '03*.
- [9] S. C. Dafermos and A. Nagurney. Sensitivity analysis for the asymmetric network equilibrium problem. *Mathematical Programming*, 28:174–184, 1984.
- [10] S. C. Dafermos and F. T. Sparrow. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards, Series B*, 73B(2):91–118, 1969.
- [11] E. Elkind, A. Sahai, and K. Steiglitz. Frugality in path auctions. In *Proceedings of the 15th Annual Symposium on Discrete Algorithms*, 2004. To appear.
- [12] S. Fortune, J. Hopcroft, and J. Wyllie. The directed subgraph homeomorphism problem. *Theoretical Computer Science*, 10:111–121, 1980.
- [13] E. J. Friedman and S. Shenker. Learning and implementation on the Internet. Working paper, 1997.
- [14] M. A. Hall. Properties of the equilibrium state in transportation networks. *Transportation Science*, 12(3):208–216, 1978.
- [15] D. W. Hearn and M. V. Ramana. Solving congestion toll pricing models. In P. Marcotte and S. Nguyen, editors, *Equilibrium and Advanced Transportation Modeling*, pages 109–124. Kluwer Academic Publishers, 1998.
- [16] B. Johansson and L. Mattsson. *Road Pricing: Theory, Empirical Assessment and Policy*. Kluwer, 1995.
- [17] M. Kearns and Y. Mansour. Efficient Nash computation in large population games with bounded influence. In *18th Conference on Uncertainty in Artificial Intelligence*, pages 259–266, 2002.
- [18] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413, 1999.
- [19] H. Lin, T. Roughgarden, and É. Tardos. A stronger bound on Braess’s paradox. In *Proceedings of the 15th Annual Symposium on Discrete Algorithms*, 2004. To appear.
- [20] J. K. MacKie-Mason and H. R. Varian. Pricing congestible network resources. *IEEE Journal on Selected Areas in Communications*, 13(7):1141–1149, September 1995.
- [21] L. W. McKnight and J. P. Bailey. *Internet Economics*. MIT Press, 1998.

- [22] M. Mihail, C. H. Papadimitriou, and A. Saberi. On certain connectivity properties of the Internet topology. In *Proceedings of the 44th Annual Symposium on Foundations of Computer Science*, pages 28–35, 2003.
- [23] J. A. Mirrlees. The theory of optimal taxation. In K. J. Arrow and M. D. Intriligator, editors, *Handbook of Mathematical Economics, Volume III*, chapter 24, pages 1197–1249. North-Holland, 1986.
- [24] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14:124–143, 1996.
- [25] D. Monderer and M. Tennenholtz. Distributed games. *Games and Economic Behavior*, 28:55–72, 1999.
- [26] C. H. Papadimitriou. Algorithms, games, and the Internet. In *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing*, pages 749–753, 2001.
- [27] A. C. Pigou. *The Economics of Welfare*. Macmillan, 1920.
- [28] R. W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.
- [29] T. Roughgarden. Designing networks for selfish users is hard. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 472–481, 2001. Full version available from <http://www.cs.berkeley.edu/~timr>.
- [30] T. Roughgarden. *Selfish Routing*. PhD thesis, Cornell University, 2002.
- [31] T. Roughgarden. The price of anarchy is independent of the network topology. *Journal of Computer and System Sciences*, 67(2):341–364, 2003. Preliminary version in *STOC '02*.
- [32] T. Roughgarden and É. Tardos. Bounding the inefficiency of equilibria in nonatomic congestion games. To appear in *Games and Economic Behavior*.
- [33] T. Roughgarden and É. Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002. Preliminary version in *FOCS '00*.
- [34] L. Schulman. Personal communication. October 1999.
- [35] R. E. Tarjan. *Data Structures and Network Algorithms*. SIAM, 1983.
- [36] J. G. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil Engineers, Pt. II*, volume 1, pages 325–378, 1952.

## A Proof of Theorem 4.5

*Proof of Theorem 4.5:* Let  $(G, r, \ell)$  be an instance in which  $G$  has  $n$  vertices, and let  $\tau$  be a tax on the edges. Let  $f$  be a directed acyclic Nash flow for  $(G, r, \ell)$  (see Proposition 2.4), and  $f^*$  a Nash flow for  $(G, r, \ell + \tau)$ . Let  $d$  and  $d^*$  be the corresponding distance labels of Proposition 2.2. In the notation of Proposition 2.6, we need to prove that  $D \equiv d(t) = c(G, r, \ell)$  is at most  $\lfloor n/2 \rfloor$  times  $D^* \equiv d^*(t) = c(G, r, \ell + \tau)$ .

An ordering of the vertices of  $G$  is *good* if it satisfies the following two properties.

(P1) All  $f$ -flow travels forward in the ordering.

(P2) The  $d$ -values of vertices are non-decreasing in the ordering.

There is at least one good ordering. To see why, first topologically sort the vertices of  $G$  according to the (directed acyclic) flow  $f$  to ensure property (P1). An ordered pair  $(u, v)$  of vertices is *bad* if  $d(v) < d(u)$  in spite of  $v$  following  $u$  in the ordering. Property (P2) is equivalent to the non-existence of bad vertex pairs. By Proposition 2.2,  $d$ -values cannot decrease across an edge carrying  $f$ -flow. Hence, if  $(u, v)$  is a bad pair, there is no path of  $f$ -flow-carrying edges from  $u$  to  $v$ . Thus, transposing  $u$  and  $v$  does not violate property (P1) and strictly decreases the number of bad vertex pairs. It follows that a good ordering must exist.

Place a good ordering on the vertices of  $G$  and label them  $v_0, v_1, \dots, v_{n-1}$  accordingly. Call an edge  $e$  of  $G$  *light* if  $f_e \leq f_e^*$  with  $f_e^* > 0$ , and *heavy* otherwise. We can finish the proof by establishing two claims (see also Figure 5).

- (1) If  $v_j$  precedes  $t$  in the good ordering, then there is a path of light edges beginning in  $\{v_0, v_1, \dots, v_j\}$  and terminating in  $\{t, v_{j+2}, v_{j+3}, \dots, v_{n-1}\}$ .
- (2) If there is a path of light edges from  $u$  to  $v$ , then  $d(v) \leq d(u) + D^*$ .

Since  $d(s) = 0$  and  $d$ -values are nondecreasing in the good ordering, applying these two claims inductively to the sets  $\{v_0, \dots, v_{2i}\}$  gives  $d(v_{2i}) \leq i \cdot D^*$  for  $v_{2i}$  equal to or preceding  $t$ . If  $t = v_{2i}$  for an integer  $i$ , the theorem follows immediately. If  $t = v_{2i+1}$ , then  $d(v_{2i}) \leq i \cdot D^*$  and the theorem follows from one further application of the two claims (to  $\{v_0, \dots, v_{2i}\}$ ).

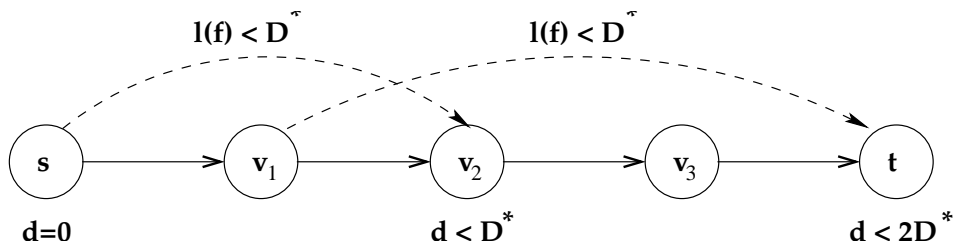


Figure 5: Proof of Theorem 4.5. If  $f$  is the flow sending one unit of flow on the four-hop path and  $f^*$  is the flow sending half a unit of flow on each of the other two paths, then the dashed edges are light.

To prove the first claim, let  $v_j$  precede  $t$  in the good ordering. By property (P1) of good orderings, no  $f$ -flow enters the  $s$ - $t$  cut  $S = \{v_0, \dots, v_j\}$ . Since the net  $f$ -flow and  $f^*$ -flow escaping any  $s$ - $t$  cut is precisely  $r$  (see e.g. [35, Lemma 8.1]), at least one light edge escapes  $S$ . If some such edge has its head in  $\{t, v_{j+2}, \dots, v_{n-1}\}$ , we are done. If not, all such light edges terminate at a vertex  $v_{j+1}$  that precedes  $t$  in the ordering. By the above argument, some light edge  $e$  escapes  $\{v_0, \dots, v_{j+1}\}$ . Since all light edges emanating from  $S$  end at  $v_{j+1}$ ,  $e$  begins at  $v_{j+1}$ . Thus, its concatenation with any light edge escaping  $S$  provides the desired path of light edges.

For the second claim, let  $P$  be a path of light edges from  $u$  to  $v$ . Since taxes are nonnegative,  $\ell_e(f_e) \leq \ell_e(f_e^*) + \tau_e$  for each edge in  $P$ . Since  $f_e^* > 0$  for all edges  $e$  of  $P$ , Proposition 2.2 implies that

$$d(v) - d(u) \leq \sum_{e \in P} \ell_e(f_e) \leq \sum_{e \in P} \ell_e(f_e^*) + \tau_e = d^*(v) - d^*(u).$$

Similarly,

$$0 \leq d^*(u) \leq d^*(v) \leq D^*$$

and hence  $d(v) - d(u) \leq d^*(v) - d^*(u) \leq D^*$ , which completes the proof of the claim and hence the theorem. ■

## B Proof of Theorem 6.2

*Proof of Theorem 6.2:* As in [29, Theorem 3.3], our reduction will proceed from the problem 2 DIRECTED DISJOINT PATHS (2DDP): given a directed graph  $G = (V, E)$  and distinct vertices  $s_1, s_2, t_1, t_2 \in V$ , are there  $s_i$ - $t_i$  paths  $P_i$  for  $i = 1, 2$  such that  $P_1$  and  $P_2$  are vertex-disjoint? Fortune et al. [12] showed that this problem is NP-complete. We will show how a  $(\frac{4}{3} - \epsilon)$ -approximation algorithm for minimizing the cost of a Nash flow with taxes in a network with linear latency functions can be used to distinguish “yes” and “no” instances of 2DDP in polynomial time.

Let  $\mathcal{I} = (G = (V, E), s_1, s_2, t_1, t_2)$  be an instance of 2DDP. Augment the vertex set  $V$  by a source  $s$ , a sink  $t$ , and directed edges  $(s, s_1), (s, s_2), (t_1, t), (t_2, t)$  (see Figure 6). Denote the new network by  $G' = (V', E')$ . We give the edges of  $E'$  the following latency functions: all edges of  $E$  receive the latency function  $\ell(x) = 0$ , edges  $(s, s_2)$  and  $(t_1, t)$  are given the latency function  $\ell(x) = x$ , and edges  $(s, s_1)$  and  $(t_2, t)$  are endowed with the latency function  $\ell(x) = 1$ .

Following the notation of Subsection 5.1, we will write  $C(\tau)$  for the cost of a Nash flow in  $(G', 1, \ell + \tau)$ . To complete the proof, we need only show the following two statements:

- (i) if  $\mathcal{I}$  is a “yes” instance of 2DDP, then there is a set  $\tau$  of taxes with  $C(\tau) = 3/2$ ; and
- (ii) if  $\mathcal{I}$  is a “no” instance, then  $C(\tau) \geq 2$  for all possible taxes  $\tau$ .

To prove (i), let  $P_1$  and  $P_2$  be vertex-disjoint  $s_1$ - $t_1$  and  $s_2$ - $t_2$  paths in  $G$ , respectively. Set the taxes  $\tau_e$  on edges  $e$  in  $E \setminus E'$ ,  $P_1$ , and  $P_2$  to zero. Set taxes  $\tau_e$  on other edges to be

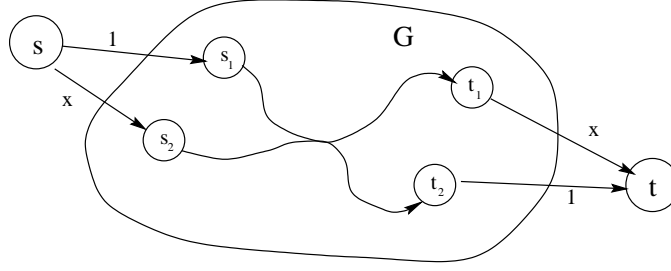


Figure 6: Proof of Theorem 6.2. In a “no” instance of 2DDP, existence of  $s_1-t_1$  and  $s_2-t_2$  paths implies the existence of an  $s_2-t_1$  path.

infinite (or sufficiently large). Then, routing half a unit of flow on each of  $P_1$  and  $P_2$  yields a flow at Nash equilibrium proving that  $C(\tau) = 3/2$ .

For (ii), consider taxes  $\tau$  and a Nash flow  $f^\tau$  for  $(G', 1, \ell + \tau)$ . Let  $d$  denote the distance labels of Proposition 2.2. Since there is one unit of traffic, (ii) is tantamount to showing that  $d(t) \geq 2$ .

First, if  $f^\tau$  routes all flow through a single  $s_i-t_j$  pair, then the latency of all flow paths is 2 and hence  $d(t) \geq 2$ . We can thus assume that  $f^\tau$  routes flow through at least two distinct  $s_i-t_j$  pairs. Similarly, we can assume that no flow path contains both  $s_1$  and  $t_2$ .

Now suppose that  $f^\tau$  routes flow on a path  $P_1$  containing  $s_1$  and  $t_1$  and also on a path  $P_2$  containing  $s_2$  and  $t_2$ . Since  $\mathcal{I}$  is a “no” instance of 2DDP,  $P_1$  and  $P_2$  share an internal vertex  $v$  (see Figure 6). Since  $P_1$  is a flow path of the Nash flow  $f^\tau$  that contains  $v$  and edge  $(s, s_1)$ , Proposition 2.2 implies that  $d(v) \geq 1$ . Similarly, since  $P_2$  is a flow path containing  $v$  and edge  $(t_2, t)$ , Proposition 2.2 implies that  $d(t) \geq d(v) + 1 \geq 2$ .

The final case arises when all flow paths contain a single  $s_i$  or a single  $t_i$  (but not both). If all flow paths contain a single  $s_i$ , then all flow uses edge  $(s, s_i)$ , so  $d(s_i) \geq 1$ . As in the previous paragraph, since there is a flow path of  $f^\tau$  containing  $s_i$  and edge  $(t_2, t)$ , Proposition 2.2 implies that  $d(t) \geq d(s_i) + 1 \geq 2$ . A similar argument proves that  $d(t) \geq 2$  if all flow paths of  $f^\tau$  contain a single  $t_i$ . ■