Department of Economics and Finance

Chair of Games and Strategies

Can Risk-Aversion Improve the Efficiency of Traffic Networks?

A comparison between risk-neutral and risk-averse routing games

Supervisor:
Prof.
Valerio Dose

Candidate:
Isabella Di Filippo
227341

Academic Year 2020/2021

“To see a World in a Grain of Sand
And a Heaven in a Wild Flower
Hold Infinity in the palm of your hand
And Eternity in a hour”

(Auguries of Innocence, William Blake)

Contents

Introduction

Life is full of social interactions among agents whose individual decisions determine the final aggregate social outcome. Sometimes these agents have the same objective and this leads to the alignment of individual and social incentives, but most of times individuals differ in their goals, and this may create inefficiencies at the aggregate level when we consider the welfare of the society as a whole.

Traffic networks are a typical example where the actions of a single agent have an impact on the utility of all other agents as well. Indeed, if we assume each agent wants to minimize the expected travel time of his route, then all agents will end up in choosing the same minimum-latency path, resulting in an equilibrium where traffic flows cause the congestion of the network and the degradation of the system's perfomance. On the other hand, when we set the aim of minimizing the total expected travel time incurred by *all* agents, the traffic flows that would follow will take into account the aggregate welfare of all agents and they would thus ensure the network is used in the most efficient and performing way.

In this dissertation, we focus on *non-atomic selfish routing games*, which model the way self-driven agents route traffic in a congested network, assuming each agent controls just a negligible portion of the overall traffic. We will focus mainly on the inefficiency of the system due to selfish agents, by comparing the social cost, in terms of delay, under equilibrium flows (also called *Wardrop equilibrium flows*) with the social cost of optimal flows, which are those arisen when we pursue the social objective function instead of the individual one.

The *Price of Anarchy* (PoA) is a measure of the inefficiency of a system's performance and captures the gap between social costs at the equilibrium level and social costs at the optimal level.

Whenever this gap is *'large'*, we may want some type of central intervention in order to align the incentives of the users with that of the system as a whole, closing in on the social optimum level and distribution of flows.

Furthermore, we will focus on how risk-aversion may impact and modify the equilibrium flows and the performance of the overall network.

Uncertainty about traffic conditions affects the decision-making of road users, causing them to add an extra buffer of time when planning trips in order to be sure to arrive at the destination in time. Uncertainty in the latency of a path may encode, for example , unfavorable weather conditions, accidents, traffic lights, parking supply, road work etc. In this context, drivers may be classified as *'risk-averse'*, since they may well trade-off mean with variance and take longer, yet more reliable and risk-minimizing routes. Another nice economic interpretation on the difference between a risk-neutral and a risk-averse user is that risk-neutral agents may be those pursuing a *'green'* objective function whose aim is to reduce long-term air pollutant emissions by vehicles, which they can achieve by minimizing their expected travel time.

In this dissertation, an analysis on how risk aversion of commuters affects traffic conditions is carried on. Uncertainty is incorporated in the traditional model of Wardrop equilibrium in non-atomic selfish routing games by adding on the edge cost functions the latency variance on that edge ($v_e(\cdot)$) times a fixed risk-aversion coefficient γ (homogenous among all agents), since we assume that users in routing games estimate the effect of uncertainty on their choices according to their respective attitudes towards risk.

To compare the social costs' gap between a risk-averse Wardrop equilibrium (RAWE) and a risk-neutral Wardrop equilibrium (RNWE), we introduce the Price of Risk Aversion of Lianeas, Evdokia Nikolova, and N. E. Stier-Moses (2019), defined as the ratio of the worst-case social cost of a RAWE to that of a RNWE, to quantify the inefficiency of the system's performance due to the agents' risk-aversion. We will see that there exists a structural upper bound to this ratio, and that it depends solely on the network's structure and topology, implying a limit to the additional degradation of the system due to risk-aversion.

After having assessed the inefficiency of a system when individuals agents are left to pursue their own objective, we discuss several ways of reducing, if not completely eliminating, such inefficiency. All of these solutions require their implementation by the so-called *social planner*, which is a kind of manager who, by imposing the right incentives and constraints, can lead self-driven agents to choose what it is best for them and for the society as a whole.

However, as we will see in the last part of this dissertation, not always a central management is needed to improve the performace of our traffic networks, and we will go through some examples where it is the risk-aversion of agents itself that, by increasing the perceived cost of some routes, can lead to the optimal flows of the network or at least do better in terms of total cost than risk-neutral Wardrop equilibrium flows.

What do we mean by inefficient outcome?

We have said that the outcome reached by self-minded agents in the routing game is most often inefficient, but why is this the case? And what do we mean by *'inefficient'*? An outcome of a game is inefficient when there is at least another outcome which would give players a higher payoff, or as in our case, where the cost suffered by agents would be lower, as it is in the optimal social outcome of the game.

To see why decisions made by rational agents may lead to inefficient outcomes, we will consider the case of the tragedy of the commons, developed by Garrett Hardin in 1968, which follows the logic of Nash equibrium concept and the purpose of the *Prisoner's Dilemma* example devised by Albert Tucker in the early 1950s. I've chosen the tragedy of commons example since it shares particularly nice policy implication with the case of congestion routing games of this dissertation. The *Tragedy of the Commons* example describes a situation where there are clearly benefits from cooperation but, since every player's decision is dictated by self-interest and each players has an incentive to *'free ride'* on the others, whatever the other player does, the equilibrium outcome is inefficient and suboptimal.

Common goods are those goods which are *non-excludable* (*i.e.* nobody can be prevented from using them) but nevertheless *rival in consumption*, which means that if I use the resource, my action affects the utility you can derive from the same resource, or put in more economic terms, the marginal cost of another person using the resource is positive. Suppose for example the following strategic game represented by Table 1.

Definition 0.1 (Strategic game). A *strategic game* is a model of interacting decision-makers which we are going to call *agents*. Each agent has a set of possible *actions* constituting their *action profiles* or *strategies* and they are called to select one of them. Each player select the best strategy for them according to their *preferences* over their set of actions, represented by payoff functions having only an ordinal significance.

Interaction in the model is captured by allowing each player to be affected by the actions of the other players and to take them into account when choosing the best strategy for themselves. In our example, the common resource is pasture and we have two agents, say Ben and Tommy, who are herdsmen willing to take advantage of the common resource available and would derive a positive utility by letting their cows grazing on the pasture. However, it is clear that if the pasture gets overcrowded, then it would become unsuitable for grazing and the resource would deplete. The fascinating yet paradoxical result of this interaction game is that the individual pursuit of what seems rational and profitable for each player leads to a collectively self-defeating result which is suboptimal for everybody!

In the matrix, the row player is Ben and the column player is Tommy. Each of them has two strategies: they can either add a cow to the common pasture or don't add it. The payoff they derive from each action are written in parentheses in the matrix (the payoff on the left refers to Ben, the one on the right to Tommy) and they are identified by the following relation $c > a > d > b$. We assume our herdsmen are rational and aim at maximizing their utility, so that their preference relations is $c \succ a \succ d \succ b$.

		Tommy	
		Add	Don't Add
Ben	Add	(d, d)	(c, b)
	Don't add	(b, c)	(a, a)

Table 1: The Tragedy of the Commons

Let's analyze the behaviour of Ben. If Ben believes Tommy is going to add another cow to the pasture (*i.e.* Tommy selects the first column), then Ben would like to add his cow too (since $d \succ b$); likewise, even when Tommy decides not to add another cow (*i.e.* by selecting the second column), Ben would be incentivized to let his own in (since $c \succ a$) and he would be irrational not to do so since the pasture is not crowded by the hypothetical additional cow of Tommy! Since Tommy is an herdsman just as alike as Ben, he would reason in the same way, leading to the equilibrium outcome of the game (d, d). Looking at the matrix, however, we see that if they both had chosen the strategy *don't add*, they would both have been better off, with a utility of $a \succ d$ each. Why is that so? Because when thinking about their own strategy, they didn't take into account the fact that the common pasture was a finite resource and sooner or later it would become overgrazed and would deplete due to the overcrowding caused by their cows. Indeed, even when Tommy decided to add another cow to the common pasture, Ben thought it was still profitable for him to add his own cow too, as long as there was still room for it.

Actually, neither Ben or Tommy is stupid: they were just so focused on their own utility that they couldn't see that, by reasoning the way they did, they were letting the pasture getting more and more crowded, till actually damaging their own herds. It's like neither Ben nor Tommy thought they were going to be *that guy*: the guy that comes and because of him the common pasture gets overgrazed. None of them thought he would be the last guy to make the pasture 'sold-out'.

What we are going to see in our selfish-network routing games is exactly this: it's like when selecting a route people are always going to think it is not because of them that the route will become congested and hence decide not to take into account the additional 'burden' (in terms of delay) they cause to other players; but as everybody reasons in this way the route *will* actually become overcrowded and drivers would be worse off than if they had thought about the additional cost they imposed on their peers when choosing the route they took. That's when the social planner steps in: his role is to make people see where their actions are taking them to, and guiding them in their choice as until when they would have the ability to select on their own the choice that would lead them to reach the optimal outcome.

Structure

In Chapter 1, we are going to present the model of nonatomic selfish routing games along with their Wardrop equilibrium, which we will refer subsequently to as *'Risk Neutral Wardrop Equilibrium'* (RNWE) in order to distinguish it from the *'Risk Averse Wardrop Equilibrium'* (RAWE) of Chapter 2. In particular, we are going to see that in a RNWE, all paths used will have equal and minimum-possible cost (in terms of delay) for the indivual agent (Beckmann et al. (1956)). We will then describe the objective function of minimizing total expected delay and we will characterize the optimal flows necessary for achieving this goal. A basic example of nonatomic routing games will be presented (*i.e.* the Pigou example) in order to make clear the difference between an equilibrium flow and an optimal one. Finally, we will introduce the *Price of Anarhcy* (PoA) as a measure of the inefficiency of the equilibria of the network and we will show the result of Roughgarden and Tardos (2002) on the tightness of the Pigou bound, a bound on the PoA of a network which is determined by the set of allowable cost functions and does not depend neither on the number of commodities of the network nor on its topology. Namely, we will see that for affine latency functions, the PoA is exactly $\frac{4}{3}$.

In Chapter 2, we incorporate risk aversion in the decision-making process of network's users and we present the related risk model along with its Risk-Averse Wardrop Equilibrium (RAWE). We then introduce the measure of the inefficiency of the system's performance due only to risk-aversion given by Lianeas, Evdokia Nikolova, and N. E. Stier-Moses (2019) : the *Price of Risk Aversion* or PRA. We will then show their result on the existence of a structural upper bound to the PRA and its corresponding structural lower bound for instances with a single origin-destination that does not depend on the class of latency functions but just on the topology of the network. In particular, Corollary 3.1 and Corollary 4.1 imply the tightness of such a bound for the set of instances with graphs with up to n vertices, where n is a power of 2.

In Chapter 3, we will focus on the problem of inefficiency in selfish routing games from a mechanism design point of view, and we present several ways of dealing with the PoA and solutions which provide the right incentives for self-driven traffic agents in order to naturally induce the optimal flow on a network. We will introduce two traditional methods for improving inefficiency on a network: marginal cost pricing or Pigouvian taxation, and augmentation of the capacity of a network. However, whether marginal cost pricing, through the imposition of tolls along the links of a network, allows network's users to internalize the marginal cost they impose on the others when using the network and ensures the optimality of the equilibrium flows thus ensued, it presents some problems for its feasibility. Indeed, the success of marginal cost pricing depends on the knowledge of the demand on the network, but in reality we do not know exactly such a demand or it may consists of stochastic information and not deterministic one. Hence, in this chapter we present two recently-developed system of tolls that do not depend on the knowledge of the network's demand. The first one relies on the paper of Colini-Baldeschi, Klimm, and Scarsini (2018) on the existence of tolls ensuring optimal flows which are nonetheless independent on the travel demand (the so-called *Demand Independent Optimum Tolls* or DIOTs). The second is the method developed by Xu and Sumalee (2015), in which an alternative marginal cost pricing called SN-MCP (*Stochastic Network - Marginal Cost Pricing*) that takes into account stochastic travel times and link flows is presented. As we will see, taking into account the stochasticity of the network in computing edge tolls is fundamental, otherwise the tolls necessary to reach the optimal flow would be underestimated, the improvements on expected total travel time would be lower than the ones under SN-MCP, and furthermore a stochastic environment is the only one available when working with observational data.

Finally, in Chapter 4 we analysed some examples where introducing risk-aversion in the network may improve its efficiency and lead to a reduction in total costs incurred by all agents. The first section focuses on Pigou's network examples, whereas the second one deals with Braess networks. In some examples (*i.e.* Example 1, 3, 5, 7, 8), we used marginal cost pricing in order to determine the amount of variance which will induce the optimal flow on the network, treating it as the *'edge toll'* we introduced in Chapter 3; whereas in other examples, we assumed certain variance and latency functions and then carried out a PRA or PoA analysis, depending on the cases. In particular, we are going to see less realistic instances (*e.g* Example 1); instances where the variance function is both optimal and plausible (*e.g.* Examples 3, 5, 7, 8); instances where total costs under a RNWE and total costs under a RAWE are equal (*e.g.* Examples 2, 9); and, most interestingly, instances where, for certain demand intervals and variance functions, risk-aversion actually improves the performance of the network (*e.g.* Examples 6, 10 and 11). This last category of examples is particularly important since it shows that sometimes being risk-averse and concerned about the variance along a path, may induce a flow distribution more favorable in terms of traffic performance and avoid congesting minimum-latency routes which may not be the optimal ones when trying to minimize total costs.

Related Work

In this work, we consider how risk-averse users can influence and alter the performance of traffic networks with respect to the traditional risk-neutral nonatomic selfish routing games first modeled by Wardrop (1952). The mathematics of the traditional model was first formalized and proved by Beckmann et al. (1956).

In Chapter 1, we used as main reference the book of Noam Nisan, Tim Roughgarden, Tardos Éva, et al. (2007). *Algorithmic Game Theory.* Cambridge University Press, where the main theorems and literature on routing games and inefficiency of equilibria are presented. The Pigou's example was first informally discussed by Pigou (1920). The concept of Price of Anarchy was first introduced by Koutsoupias and Papadimitriou (1999) and given its current term by Papadimitriou (2001). The Price of Anarchy in nonatomic selfish routing games has been extensively studied by Roughgarden and Tardos (2002) and Roughgarden (2006), who also proved the tightness of the Pigou bound for instances having affine latency functions. The Pigou bound itself is from Roughgarden (2003).

In Chapter 2, the literature on stochastic selfish-routing games with risk-averse players has been used. In particular, we refer to the works of Ordóñez and N. Stier-Moses (2010) and E. Nikolova and N. Stier-Moses (2014a) for risk-averse equilibrium flows and to the works of E. Nikolova and N. Stier-Moses (2014b) and Lianeas, Evdokia Nikolova, and N. E. Stier-Moses (2019), who extended the notion of PoA to the case of stochastic delays with risk-averse users through the Price of Risk Aversion (PRA), *i.e.* a measure of how much the inefficiency of the system's performance can be attribituted to the agents' risk-aversion, and computed structural upper and lower bounds to the PRA for arbitrary graphs with a single origin-destination pair and homogenous risk-averse players.

In Chapter 3, where we introduce ways to induce the optimal flow and deal with the PoA, we referred to the literature on marginal cost pricing (Pigou 1920) and capacity augmentation (Roughgarden 2006) for the traditional methods; whereas for marginal cost pricing with uncertain travel time and demand functions, we used the recent paper of Colini-Baldeschi, Klimm, and Scarsini (2018) on Demand-Independent Optimal Tolls (DIOTs) and the works of Xu and Sumalee (2015), Yang, Meng, and Lee (2004) and Li (2002) on trial-and-error implementation of stochastic marginal cost pricing.

The examples in Chapter 4 are inspired by the work of Fotakis, Kalimeris, and Lianeas (2020), which investigates how and the extent to which risk-aversion of players can be exploited to lower the PoA of nonatomic selfish-routing games with respect to the risk-neutral case. Similar research on the subject has been carried out by Yekkehkhany and Nagi (2020), who numerically studied the impact of risk-aversion on the PoA of atomic selfish-routing games and showed that the Braess paradox may not apply in a risk-aversion setting and the PoA may actually improve in both Braess and Pigou networks.

Chapter 1

The Traditional Model

1.1 Preliminaries

1.1.1 Selfish Routing Networks

Selfish routing games are non-cooperative games showing how self-interested users route traffic through a congested network. As we are going to see in Section 1.3, selfish routing games generally lead to inefficient outcomes, since users aim at maximizing their own utility and do not take into account the additional cost (*i.e.* delay) they impose on other users. The model of routing game we will focus on is ***nonatomic selfish routing***, where there is a very large number of users and each user controls a negligible portion of the overall traffic, so that the actions of an individual user have no effect on the network congestion, even though the joint actions of several players do affect other players.

A flow network is described by a directed graph $G = (V, E)$, with vertex set V and edge set E, and a set $\mathcal{I}$ of $(s_1, t_1), \ldots, (s_k, t_k)$ of source-sink vertex pairs, also called *commodities* or O-D pairs.
For a multicommodity network G, let $\mathcal{P}_i$ denote the $s_i - t_i$ paths of a network and $\mathcal{P}$ the union $\cup_{i=1}^{k} \mathcal{P}_i$. A *flow* f in G describes the routes chosen by players and it is identified by a non-negative vector indexed by $\mathcal{P}$ and whose coordinates belong to $\mathbb{R}$. For a flow f and a path $P \in \mathcal{P}_i$, we interpret f_P as the amount of traffic of commodity i that chooses path P to go from s_i to t_i. A flow f induces a *flow on edges* $\{f_e\}_{e \in E}$, where $f_e = \sum_{P \in \mathcal{P} : e \in P} f_P$ denotes the total amount of flow using edge e. We denote by r the non-negative vector of traffic rates and we say a flow f is feasible for a vector r if it routes all of the traffic: for each $i \in \{1, 2, \ldots, k\}$, $\sum_{P \in \mathcal{P}_i} f_P = r_i$.

We assign to each edge $e \in E$ of a network a *cost function* $c_e : \mathbb{R}^+ \mapsto \mathbb{R}^+$, which we assume continuous and non-decreasing and which allows us to to represent the cost in terms of time (*delay*) incurred by traffic that traverses edge e, as a function of the edge congestion f_e.

We describe such a routing game with a triple (G, r, c), also called *instance*.

1.1.2 Wardrop Equilibrium

In order to find the solution of nonatomic selfish routing games, or their *nonatomic equilibrium flow*, we will use the concept of Wardrop equilibrium defined below.

Definition 1.1. Let f be a feasible flow for the instance (G, r, c). The flow f is a *Wardrop equilibrium* if, for every commodity $i \in \{1, 2, \ldots, k\}$ and every pair $P, \tilde{P} \in \mathcal{P}_i$ of $s_i - t_i$ paths with $f_{P>0}$,

$$c_P(f) \leq c_{\tilde{P}}(f)$$

that is, all paths in use in a Wardrop equilibrium f have minimum-possibile cost. Furthermore, all paths of a given commodity used in a Wardrop equilibrium have equale cost.

The last statement of the aforementioned definiton has been proved by Beckmann et al. (1956) who came up with important results on the existence and uniqueness of Wardrop equilibria which are stated in the following proposition:

Proposition 1.1. *Let (G, r, c) be an instance.*
(a) The instance (G, r, c) admits at least one Wardrop equilibrium.
(b) If f and $\tilde{f}$ are Wardrop equilibria for (G, r, c), then $c_e(f_e) = c_e(\tilde{f}_e)$ for every edge e.

Statement *(b)* implies that equilibrium flows are *essentially unique* in that all equilibrium flows of a nonatomic instance have the same cost.
Beckmann et al. have showed that the Wardrop equilibria of an instance (G, c, r) are precisely those flows which minimize the *potential function*:

$$\Phi(f) = \sum_{e \in E} \int_0^{f_e} c_e(x)dx \tag{1.1}$$

over all feasible flows for (G, r, c). Thus, since cost functions are continuous and the space of all flows is compact, Weierstrass's Theorem ensures the existence of a minimum and thus *(a)*, whereas since cost functions are non-decreasing, the function Φ is convex and statement *(b)* follows.

1.1.3 Defining an objective function

We consider here the utilitarian objective function of minimizing the total cost incurred by all the users of the network.
We define the cost incurred by a player choosing the path P with flow f as $c_P(f)$, and f_P as the amount of traffic experienced using the path P. Then, we define the cost of a flow as:

$$C(f) = \sum_{P \in \mathcal{P}} c_P(f) f_P \tag{1.2}$$

or, alternatively

$$C(f) = \sum_{e \in E} c_e(f_e) f_e \tag{1.3}$$

where we expanded $c_P(f)$ as $\sum_{e \in P} c_e f_e$.

1.1.4 Optimal flow

For an instance (G, r, c), we call a feasible flow optimal if it minimizes the cost over all feasible flows, *i.e.* if it minimizes (1.2). An alternative definiton of optimal flow can be given under certain hypotheses on the cost functions. We have the following proposition:

Proposition 1.2 (**Characterization of optimal flows**). *Let (G, c, r) be a non-atomic instance such that, for every edge e, the function $x \cdot c_e(x)$ is convex and continuously differentiable. Let $c_e^*(x_e) := (x \cdot c_e(x))'$ denote the marginal cost function of the edge e. Then f^* is an optimal flow for (G, r, c) if and only if, for every commodity $i \in \{1, 2, \ldots, k\}$ and every pair $P, \tilde{P} \in \mathcal{P}_i$ of $s_i - t_i$ paths with $f_P^* > 0$,*

$$c_P^*(f^*) \leq c_{\tilde{P}}^*(f^*).$$

The proof of this proposition follows directly from the first-order conditions of a convex optimization problem (in our case, our objective is to minimize the function representing the total expected cost over all the edges, *i.e.* (1.2)) and it gives us the following corollary:

Corollary 0.1 (**Equivalence of equilibrium and optimal flows**). *Let (G, r, c) be a nonatomic instance such that, for every edge e, the function $x \cdot c_e(x)$ is convex and continuously differentiable. Let c_e^* denote the marginal cost function of the edge e. Then f^* is an optimal flow for (G, r, c) if and only if it is an equilibrium flow for (G, r, c^*).*

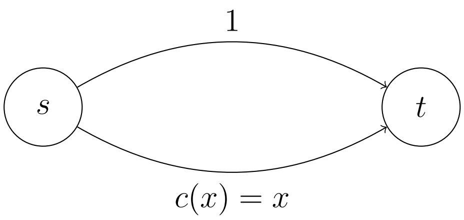
The definition of Wardrop equilibrium 1.1 and this corollary imply that also optimal flows can be defined in terms of equilibrium flow, but with respect to a different function than the minimization of an individual's travel time. Indeed, the optimal flow takes into account the increase in per-unit cost $c_e(x)$ which the additional flow of another user along that edge would impose, just as a selfish agent does when trying to minimize its total travel time; however, the optimal flow's calculation takes into account a second term $x \cdot c'_e(x)$ accounting for the increased congestion experienced by the flow already using that edge, whereas the flow at Wardrop equilibrium disregards this second *'conscientious'* term that takes care of the other users' experience too.

Proposition 1.3 (**Potential function for equilibrium flows).** *Let* (G, c, r) *be a nonatomic instance. A flow feasible for* (G, c, r) *is an equilibrium flow if and only if it is a global minimum for the potential function* Φ *in* (1.1).

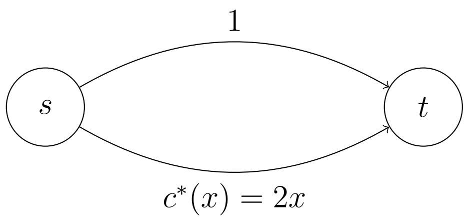
Indeed, corollary 0.1 can be used to build the potential function Φ we defined before in (1.1): the potential function is just the function from which equilibrium flows do arise as global minima and its derivative is the function with respect to whom equilibrium flows are defined as Wardrop equilibria.

1.2 Pigou's example

After the definitions given in the previous section, we are now ready to show an example of a selfish routing game with its Wardrop equilibria and optimal solutions. This example is named after Pigou (1920), who first came up with it.
In figure 1.1a two disjoint parallel edges connect a source s to a sink t. Each edge is labelled



(a) Pigou's example



(b) The optimal flow is the equilibrium flow of the same game with edges $c^*_e(x)$ of the objective function $x \cdot c_e(x)$

Figure 1.1

with a cost function $c(\cdot)$ which describes the cost in terms of travel time incurred by users of the edge, as a function of the traffic on that edge. There is $r = 1$ unit of traffic to be routed from s to t. The upper edge has a constant cost $c(x) = 1$, whereas the lower edge has a cost $c(x) = x$ that is increasing in the amount of traffic traversing it. Assuming all players are

homogenous in that they aim to minimize their travel time, we expect all users to choose the lower edge. Indeed, the lower edge is a dominant strategy in that it is superior to the upper route whenever some users choose not to take the lower route and it is never worse when it is fully congested. At equilibrium, all players choose the lower route and incur in a cost $c(1) = 1$ and it is easy to check that it is equal to the cost of the non-used path, so that no player has an incentive to deviate.
On the other hand, we can see that routing the traffic by splitting it equally between the upper and the lower route reduces the cost incurred by half of the players without making anybody worse-off: the half on the upper route incurs a cost of 1 as before, whereas the half assigned to the lower route has a cost of $c(x = \frac{1}{2}) = \frac{1}{2}$ and has reduced its cost by a half!
The flow where half of the traffic takes the upper edge and the other half the lower one, is indeed our optimal flow f^* which can be easily computed as the minimizer of the social cost function (1.2), which in our case can be written as $C(x) = 1 \cdot (1 - x) + x^2$ (where $1 - x$ is the flow sent to the upper edge and x the flow on the lower edge, such that $1 - x + x = 1$ and the demand of the network is satisfied) or as the equilibrium flow of the analoguous game with marginal cost functions $c^*(x) = 1$ on the upper edge and $c^*(x) = 2x$ on the lower edge as in our second definition resulting from 1.2 (see 1.1b).

Thus, we have just seen that when players act selfishly, meaning that they are just focused on minimizing their own travel time, they reach an equilibrium outcome which is not optimal from the point of view of the total cost $C(x)$, and it is in that sense inefficient.

1.3 Price of Anarchy

The outcome of rational behaviour by self-interested users (a non-cooperative equilibrium) in a game can often be inefficient, meaning that it is strictly Pareto inefficient from a *qualitative* perspective (*i.e.* it does not maximize social welfare) or that it does not maximize/minimize an objective function from a *quantitative* perspective. Inefficiency of equilibria is particularly evident in routing games which capture decision-making by multiple utility-maximizing agents in a networking context, internalizing the congestion externalities generated by self-interested agents. The price of anarchy (PoA) is one of the most used way of measuring the inefficiency of equilbria and it was first studied by Koutsoupias and Papadimitriou (1999) and given its current name by Papadimitriou (2001). It is defined as the ratio between the worst objective function value of an equilibrium of the game and that of an optimal outcome.
We will identify the price of anarchy in a routing games as follows:

Definition 1.2. The price of anarchy $\rho(G, r, c)$ of an instance (G, c, r) is

$$\rho(G, r, c) = \frac{C(f)}{C(f^*)}$$

where f is a Wardrop equilibrium and f^* is an optimal flow for (G, r, c). The *price of anarchy* $\rho(\mathcal{I})$ of a non-empty set $\mathcal{I}$ of instances is $\sup_{(G,c,r)\in\mathcal{I}} \rho(G, r, c)$.

This value is well defined unless there is a flow for a Wardrop equilibrium with zero cost; in this case, all Wardrop equilibria have zero cost, and we define the PoA of the instance to be 1. In general, a value cost close to 1 indicates that the given equilibrium outcome is approximately optimal.

Taking our previous example 1.1a and the cost function 1.2, we see that the average cost incurred in the Wardrop outcome is 1, whereas the average cost in the optimal outcome is $C(f^*) = (1 + \frac{1}{2}) \cdot \frac{1}{2} = \frac{3}{4}$. The *PoA* for the Pigou example is thus: $\rho(G, r, c) = \frac{1}{\frac{3}{4}} = \frac{4}{3}$.

1.3.1 Pigou bound

Roughgarden and Tardos (2002) and Roughgarden (2003) showed that the price of anarchy of selfish routing can be large when the cost functions are *'sufficiently non-linear'*. For example, suppose in Fig. 1.1a we replace the linear cost function of the lower edge with its non-linear variant $c(x) = x^p$, for p large in $\mathbb{N}$. In this case, the equilibrium flow still has a cost of 1; on the other hand, to compute the optimal flow, we need to equalize the marginal total expected travel time on the two edges (*i.e.* we want $1 - x = (p+1)\, x^p$, which gives the solution $x = (p+1)^{-\frac{1}{p}}$). Then, if we call $\epsilon = 1 - (p+1)^{-\frac{1}{p}}$, the optimal flow has cost $\epsilon + (1-\epsilon)^{p+1}$. We can clearly see that, when $p \to \infty \Rightarrow \epsilon \to 0$, implying the cost of the optimal flow approaches 0. Since in this instance the cost of the equilibrium flow is 1, the non-linearity of the cost function $c(x) = x^p$ implies that the PoA $\to \infty$ as p grows large or, as defined by the authors, the cost function is *'sufficiently non-linear'*.

However, the authors also showed that the PoA is controlled only by the set of allowable cost functions, and is independent of the number of commodities and the complexity of the allowable network topologies. In this context, the Pigou's example is particularly relevant in that it has been shown that the largest PoA among all networks with cost functions in the set $\mathcal{C}$ is achieved in a Pigou-like network.

Definition 1.3 (Pigou-like network)**.** A Pigou-like network (G, c, r) is characterized by a two-node, two-edge network and a traffic rate $r > 0$. Chosen a set of allowable cost functions $\mathcal{C}$, the upper edge has a constant cost function $c(r),\ c \in \mathcal{C}$ and the lower edge has a cost function $c(\cdot)$.

In such a network, the lower edge is always a dominant strategy and all the traffic will flow there, yielding a Wardrop equilibrium with total cost $c(r) \cdot r$. On the other hand, an optimal

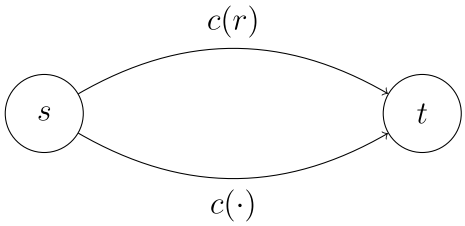


Figure 1.2: A Pigou-like network

flow won't route all the traffic on one edge, thus making it congested, but it would rather route x units of traffic on the lower edge and $(r-x)$ units on the upper edge, resulting in a total cost $[c(x)x+(r-x)c(r)]$. The Pigou-bound provides a natural lower bound on the PoA of instances with cost functions $c \in \mathcal{C}$.

Definition 1.4 (Pigou bound)**.** Let $\mathcal{C}$ be a non-empty set of cost functions. The *Pigou bound* $\alpha(c)$ for $\mathcal{C}$ is

$$\alpha(\mathcal{C}) = \sup_{c \in \mathcal{C}} \sup_{\substack{r \geq 0, \\ 0 < x \leq r}} \frac{r \cdot c(r)}{x \cdot c(x) + (r-x)c(r)},$$

with the numerator being the total cost at equilibrium flow, and the denominator the total cost at optimal flow.

The next proposition thus follows:

Proposition 1.4. *Let $\mathcal{C}$ be a set of cost functions that includes all of the constant cost functions, and let $\mathcal{I}$ denote the single-commodity instances with a two-node, two-link network and cost functions in $\mathcal{C}$. Then*

$$\rho(\mathcal{I}) \geq \alpha(\mathcal{C}).$$

In the next subsection it will be shown that the Pigou bound is indeed tight, and we will be able to find a matching upper bound on the PoA of selfish routing games in instances with cost functions in $\mathcal{C}$.

1.3.2 Tightness of Pigou bound

In order to show the tightness of the Pigou bound, we introduce the following characterization of a Wardrop equilibrium.

Proposition 1.5 (Variational inequality characterization). *Let f be a feasible flow for the nonatomic instance (G, r, c). The flow f is an equilibrium flow if and only if*

$$\sum_{e \in E} c_e(f_e) f_e \leq \sum_{e \in E} c_e(f_e) f_e^*$$

for every flow f^* feasible for (G, r, c).

This proposition follows from rewriting the inequality as a sum over paths $\sum_{P \in \mathcal{P}} c_P(f) f_P \leq \sum_{P \in \mathcal{P}} c_P(f) f_P^*$ and noticing that it always holds since by definition an equilibrium flow routes all traffic to the shortest paths, *i.e.* the paths that have minimal cost.

Theorem 1. *Let $\mathcal{C}$ be a set of cost functions and $\alpha(\mathcal{C})$ the Pigou bound for $\mathcal{C}$. If (G, c, r) is an instance with cost functions in $\mathcal{C}$, then*

$$\rho(G, r, c) \leq \alpha(\mathcal{C}).$$

The proof of the theorem follows by rearranging the equation in Definition 1.4 and taking $r = f_e$ and $x = f_e^*$, where f_e is the flow at equilibrium and f_e^* the optimal flow, then summing all over the edges and using the ***variational inequality characterization*** which ensures $\sum_{e \in E}(f_e^* - f_e)c_e(f_e) \geq 0$.

Proof. Let's use Definition 1.4. We instantiate parameters c, r, x for each edge $e \in E$ of the network, by taking $c \leftarrow c_e$, $r \leftarrow f_e$ and $x \leftarrow f_e^*$.
So,

$$\alpha(C) \geq \frac{f_e c_e(f_e)}{f_e^* c_e(f_e^*) + (f_e - f_e^*)c_e(f_e)} \tag{1.4}$$

$$f_e^* c_e(f_e^*) \geq \frac{1}{\alpha(C)} f_e c_e(f_e) + (f_e^* - f_e)c_e(f_e) \qquad \text{rearranging it}$$

$$C(f^*) \geq \frac{1}{\alpha(C)} C(f) + \sum_{e \in E}(f_e - f_e^*)c_e(f_e) \qquad \text{after summing over all the edges} \tag{1.5}$$

$$C(f^*) \geq \frac{C(f)}{\alpha(C)} \qquad \text{by using proposition 1.5} \tag{1.6}$$

or, equivalently

$$\frac{C(f)}{C(f^*)} \leq \alpha(C) \tag{1.7}$$

□

Proposition 1.4 and Theorem 1 imply the tightness of the Pigou bound and the following corollary:

Corollary 1.1. *Let $\mathcal{C}$ be a set of cost functions that includes all of the constant functions. Then the PoA of instances with cost functions in $\mathcal{C}$ is achieved by a small single-commodity instance with a two-node, two-edge network.*

This corollary states that the PoA is controlled *only* by the set of allowable cost functions, and it is independent on the number of commodities or on the network topology. Roughgarden and Tardos (2002) proved the PoA in networks with affine edge cost functions is precisely $\frac{4}{3}$.

1.3.3 PoA in networks with linear latency functions

By the example we have seen in Section 1.3, it is clear that the Pigou bound for the instances having affine edge cost functions is $\frac{4}{3}$. We will now prove the existence of a matching upper bound for these networks.

Definition 1.5 (Linear latency functions)**.** Let (G, r, c) be an instance of a non-atomic routing game. We say such a game has linear latency functions if the latency of each edge e is linear in the edge congestion, *i.e.* $c_e(x) = a_e x + b_e \; \forall e \in E$ and for some $a_e, b_e \geq 0$.

Before proceeding to the statement and proof of the theorem, we need the following two lemmas:

Lemma 1. *Let (G, r, c) having linear latency function and f be an equilibrium flow. Then,*

(a) the flow $f/2$ is optimal for $(G, r/2, c)$;

(b) the marginal cost of increasing the flow on a path P with respect to $f/2$ is equal to the cost of P with respect to f.

Proof. For part *(a)*, we know that if f is an equilibrium flow, then it must be $\sum_{e \in P} a_e f_e + b_e \leq \sum_{e \in P'} a_e f_e + b_e$ for $P, P' \in \mathcal{P}_i$ with $f_P > 0$. Since the optimal flow of the same instance requires $\sum_{e \in P} 2a_e f_e^* + b_e \leq \sum_{e \in P'} 2a_e f_e^* + b_e$ for $P, P' \in \mathcal{P}_i$ with $f_P^* > 0$, it follows that $f/2$ satisfies this second inequality for the instance $(G, r/2, c)$ and it is thus optimal for it. For part *(b)*, if $c_e(x) = a_e x_e + b_e$ then its marginal cost function is $c_e^*(x) = 2a_e x + b_e$, which implies $c_e^*(f_e/2) = c_e(f_e) \; \forall e \in E$, or equivalently, $c_P^*(f/2) = c_P(f)$. □

Lemma 2. *Let (G, r, c) be an instance with linear latency functions for which f^* is an optimal flow. Let $c_i^*(f^*)$ be the minimum marginal cost of increasing flow on an $s_i - t_i$ path with respect to f^*. Then for any $\delta > 0$, a feasible flow for the instance $(G, (1+\delta)r, c)$ has cost*

$$C(f) \geq C(f^*) + \delta \sum_{i=1}^{k} c_i^*(f^*) r_i.$$

This lemma states that the cost of increasing the amount of flow through a network is at least the marginal cost of increasing flow on any path with respect to the optimal flow.

Proof. Fix $\delta > 0$ and suppose f is feasible for $(G, (1+\delta)r, c)$. Note that $\forall e \in E$, $x_e \cdot c_e(x) = a_e x^2 + bx_e$ is a convex function, implying that

$$c_e(f_e)f_e \geq l_e(f_e^*)f_e^* + (f_e - f_e^*)l_e^*(f_e^*)$$

where the RHS is the tangent line approximation of the function $c_e(f_e)f_e$ at the point $f_e = f_e^*$ and the inequality comes from the characterization of convex and differentiable functions. Hence, it follows that

$$\begin{aligned}
C(f) &= \sum_{e\in E} c_e(f_e)f_e \\
&\geq \sum_{e\in E} c_e(f_e^*)f_e^* + \sum_{e\in E}(f_e - f_e^*)c_e^*(f_e^*) \\
&= C(f^*) + \sum_{i=1}^{k} \sum_{P\in\mathcal{P}_i} c_P^*(f^*)(f_P - f_P^*) \\
&\geq C(f^*) + \sum_{i=1}^{k} c_i^*(f^*) \sum_{P\in\mathcal{P}_i}(f_P - f_P^*) \quad \text{since } c_i^*(f^*) = c_P^*(f^*) \text{ unless } f_P^* = 0, \text{ where } c_i^*(f^*) < c_P^*(f^*) \\
&= C(f^*) + \delta \sum_{i=1}^{k} c_i^*(f^*) r_i. \quad \text{since the increase } f_P - f_P^* = \delta r_i
\end{aligned}$$

□

We are now ready to state the main theorem and provide a proof of it.

Theorem 2. *If (G, r, c) has linear latency functions, then $\rho(G, r, c) \leq \frac{4}{3}$.*

Proof. The proof of the theorem consists in sending through G an optimal flow for the instance $(G, r/2, c)$ and then augmenting this flow to one optimal for the instance (G, r, c).

$$\begin{aligned}
C(f^*) &\geq C(f/2) + \sum_{i=1}^{k} c_i^*(f/2)\frac{r_i}{2} \quad \text{taking } \delta = 1 \text{ in Lemma 2} \\
&= C(f/2) + \frac{1}{2}\sum_{i=1}^{k} c_i(f)r_i \quad \text{using part } (b) \text{ of Lemma 1} \\
&= C(f/2) + \frac{1}{2}C(f) \quad \text{since in equilibrium, all } s_i - t_i \text{ paths have equal latency } c_i \\
&\geq \frac{1}{4}\sum_e a_e f_e^2 + b_e f_e + \frac{1}{2}C(f) \quad \text{since } C(f/2) = \sum_e \frac{1}{4}a_e f_e^2 + \frac{1}{2}b_e f_e \\
&= \frac{1}{4}C(f) + \frac{1}{2}C(f) \\
&\Rightarrow \frac{C(f)}{C(f^*)} \leq \frac{4}{3}.
\end{aligned}$$

□

Thus, we have managed to find a matching upper bound to the Pigou bound for instances with affine cost functions, proving that the PoA for these networks is exactly $\frac{4}{3}$.

Chapter 2

The Risk Model

In this model, we incorporate risk and uncertainty in the decision-making process of the network's users through a *stochastic non-atomic routing game*. Indeed, in reality people may want to take into account the variability associated with the path they choose, ending up in selecting paths with higher expected travel time but lower variability in order to make sure to arrive at destination in time.

For example, imagine you have an exam that starts in a hour and you have just woken up: you would rather choose to take a longer but risk-free route that brings you to the destination in time for the exam roll-call than an apparently shorter route but with more risks related to traffic, accidents or roadwork.

In this context, drivers are risk-averse users who want to minimize the risk of uncertain delays, which may result due to averse weather conditions, road work, accidents or traffic lights.

If, as we have seen in the previos section, the PoA measures the degradation of a system's performance due to the misalignment of individual incentives and the socially optimal solution, the *Price of Risk Aversion* (PRA) measures the degradation of the system performance caused by variability and risk aversion. The PRA is the worst-case ratio of the social cost at a risk averse Wardrop equilibrium (RAWE) to that at a risk neutral Wardrop equilibrium (RNWE), which we will define in the following section.

2.1 Incorporating uncertainty in the traditional model

In order to incorporate uncertainty in the traditional model of section 1.1, we add to every edge of the network a random variable whose distribution depends on the edge flow. This random variable is called $\xi_e(f_e)$ and has mean $\mu = 0$ and standard deviation $\sigma_e(f_e)$ for arbitrary continuous functions $\sigma_e(\cdot)$. Now, the *mean-variance* objective for a path is given by

$$Q_p^\gamma(f) = l_p(f) + \gamma \upsilon_p(f)$$

where $l_p(\cdot)$ is the expected latency over the path, γ is a risk-aversion coefficient that depends on the users' preferences (which we assume to be homogenuous), and $\upsilon_p(\cdot)$ is the variance of travel time of the path. Using the variance as a proxy for the variability of a path leads to models which are *additive* (*i.e.* the cost of a path is the sum of the cost of its edges) and also have the *additive consistency property*, namely, a subpath of an optimal path is still optimal.

We consider a directed graph $G = (V, E)$ with the same characteristics as the one we presented in section 1.1. There is an aggregate demand of d units of flow and we consider a flow feasible when demand is satisfied, *i.e.* $\sum_{p\in\mathcal{P}} f_p = d$. Congestion is modeled with stochastic delay functions $l_e(f_e) + \xi_e(f_e)$, where $l_e(\cdot)$ is the flow-dependent expected delay over the edge e and $\xi_e(f_e)$ is a random variable representing some noise term on the delay. Latency functions l_e are assumed continuous and nondecreasing and the latency of a path p with respect to a flow f is defined as $l_p(f) := \sum_{e\in p} l_e(f_e)$. Random variables $\xi_e(f_e)$ are pairwise independent and have mean zero and standard deviation $\sigma_e(f_e)$ for arbitrary continuous functions $\sigma_e(\cdot)$. The variance along a path is defined as $\upsilon_p(f) = \sum_{e\in p} \sigma_e^2(f_e)$, and consequently the standard deviation along a path is $\sigma_p(f) = (\upsilon_p(f))^{1/2}$. It is also assumed the variance is never too large with respect to the mean such that $\upsilon_e(x_e)/l_e(x_e)$ is bounded from above by a constant κ for all $e \in E$ at the equilibrium flow $x_e \in \mathbb{R}_+$.

Now, we assume players are risk-averse and take into account the variability of the expected delay of their path when selecting routes. Indeed, players minimize the *mean-var* objective function mentioned before $Q_p^\gamma = l_p(f) + \gamma\upsilon_p(f)$, where γ is a constant term quantifying the risk aversion of the players, which we assume homogeneous. If $\gamma = 0$, then players are risk-neutral and we are back to the traditional model of Wardrop (1952). The instance of the problem is given by the tuple $(G, d, l, \upsilon, \gamma)$.

Definition 2.1 (Definition of *γ-equilibrium)***.** A *γ-equilibrium* of a stochastic nonatomic routing game is a flow f such that for every $p \in \mathcal{P}$ with positive flow, the path cost

$$Q_p^\gamma(f) \leq Q_q^\gamma(f)$$

for any other path $q \in \mathcal{P}$. For a fixed risk-aversion parameter γ, we refer to a γ-equilibrium as a RAWE, denoted by x. For $\gamma = 0$, we call it a 0-equilibrium or RNWE and denote it by z.

Since the variance is additive along edges, this represents a standard Wardrop equilibrium with respect to the modified edge cost functions $l_e(f_e) + \gamma\upsilon_e(f_e)$.

Definition 2.2 (Social cost function)**.** The *social cost* of a flow f is defined as the sum of the expected latencies of all players $C(f) := \sum_{p\in\mathcal{P}} f_p l_p(f) = \sum_{e\in E} f_e l_e(f_e)$.

In order to measure the degradation of the system performance due to risk-aversion, we introduce the measure of inefficiency given by Lianeas, Evdokia Nikolova, and N. E. Stier-Moses (2019) which this dissertation presents and to which all the following results relate to.

Definition 2.3 (Price of Risk Aversion)**.** Considering an instance family $\mathcal{F}$ of a routing game with uncertain delays, the price of risk aversion (PRA) associated with γ and κ is defined by:

$$\mathrm{PRA}(\mathcal{F},\gamma,\kappa) := \sup_{G,d,l,\upsilon}\left\{\frac{C(x)}{C(z)} : (G,d,l,\upsilon,\gamma)\in\mathcal{F}, \text{and } \upsilon(x)\leq\kappa l(x)\right\},$$

where x and z are the risk-averse and risk-neutral Wardrop equilibria of the instance, respectively.

2.2 Defining a structural upper bound

It can be proved that an upper bound of the PRA for a given instance with a single source-sink pair is $1+\gamma\kappa\eta$, which depends linearly on the degree of risk-aversion γ of players, on the upper bound κ of the var-to-mean ratio and on the parameter η, which is a positive integer that depends on the topology of the network. This bound is called *structural* in that it does not depend on the latency functions chosen but just on the structure of the network.

To prove this upper bound, we first need to state the four lemmas which lead to it. We assume that the aggregate demand $d = 1$.

Lemma 3. *Letting $p\in\mathcal{P}$ denote an arbitrary path, the social cost of a RAWE $C(x)$ is upper bounded by the path cost $Q_p^\gamma(x)$.Furthermore, if variance-to-mean ratio is bounded by κ at equilibrium, then $C(x)\leq(1+\gamma\kappa)\, l_p(x)$.*

The proof of this lemma follows from the fact that $l_q(x)+\gamma\upsilon_q(x)\leq l_p(x)+\gamma\upsilon_p(x)$ for all paths $q\in\mathcal{P}$ with $f_q > 0$.
In order to state the other three lemmas, it is necessary to introduce the notion of an alternating path.
An *alternating path* is constructed by two sets of edges contained in E:

$$A = \{e\in E|\, x_e\leq z_e\} \quad \text{and} \quad B = \{e\in E|\, z_e < x_e\},$$

and it is a $s-t$ path with forward edges in A and backward or reversed edges in B.

Definition 2.4 (**Alternating path**). A path $\pi := (e_1, \cdots, e_r)$ from s to t is alternating if reversing the direction of edges in B makes it a feasible $s-t$ path. π can also be written as $\pi = A_1 - B_1 - A_2 - \cdots - A_t - B_t - A_{t+1}$, where for every i each edge in $A_i \subseteq A$ is directed forward and every edge in $B_i \subseteq B$ is directed in the opposite direction.

Lemma 4. *For an instance $(G, d, l, \upsilon, \gamma)$, an alternating path exists.*

The existence of such an alternating path is guaranteed by the principle of flow conservation and the definitions of A and B given before and further details on the proof can be found in Lianeas, Evdokia Nikolova, and N. E. Stier-Moses 2019. Now, Lemma 1 can be extended to the case of alternating paths too.

Lemma 5. *Consider an arbitrary graph with general latencies and general variance functions satisfying that the var-to-mean ratio is bounded by κ at equilibrium. Letting π be an alternating path, then*

$$C(x) \leq (1 + \gamma\kappa) \sum_{e \in A \cap \pi} l_e(x_e) - \sum_{e \in B \cap \pi} l_e(x_e).$$

Proof. Let's consider the following s-t alternating path $\pi := A_1 - B_1 - A_2 - \cdots - A_{\eta-1} - B_{\eta-1} - A_\eta$. Here, by $A_i, B_i, etc.$ we denote the users' perceived cost on that edge (e.g. $A_i = \sum_{e \in A_i}(l_e(x_e) + \gamma\upsilon_e(x_e)$, for $A_i \subseteq A$). By definition each edge e in B_k carries flow $x_e > 0$ under a RAWE x, so there must exist an s-t path such that e belongs to it. Hence, there must be a flow-carrying path $C_k - B_k - D_k$, where C_k originates at the source s and D_k terminates at the sink t. We define $C_0 = D_\eta = \varnothing$.
By the equilibrium conditions, if $C_{k-1} - A_k$ is an alternative route between the same nodes connected by $C_k - B_k$, then it must be $C_k + B_k \leq C_{k-1} + A_k$ for all k. Indeed, we know that at equilibrium players choose the paths having the least cost in terms of delay, and since the path $C_k - B_k - D_k$ is chosen at equilibrium, then it must be that any other path connecting the same nodes must have a cost greater than or equal to its. Therefore,

$$C_k \leq C_{k-1} + A_k - B_k \leq \cdots \leq (A_1 + A_2 + \cdots + A_k) - (B_1 + B_2 + \cdots + B_k) \tag{2.1}$$

where the third inequality is obtained by using the first inequality for every k and recalling that $C_0 = \varnothing$.

Similarly, since $A_{k+1} - D_{k+1}$ is an alternative route for the nodes connected by $B_k - D_k$, we must have $B_k + D_k \leq A_{k+1} + D_{k+1}$, and therefore

$$D_k \leq A_{k+1} + D_{k+1} - B_k \leq \cdots \leq (A_{k+1} + A_{k+2} + \cdots + A_\eta - (B_k + B_{k+1} + \cdots + B_{\eta+1}) \tag{2.2}$$

where we used the first inequality for every k and the fact that $D_\eta = \varnothing$.

Then, for path $q = C_k - B_k - D_k$ for any k, we have the following:

$$\begin{aligned}
C(x) &= \sum_p x_p\, l_p(x) \\
&\leq \sum_p x_p(l_q(x) + \gamma\upsilon_q(x) - \gamma\upsilon_p(x)) && \text{since if } x_p > 0\text{, then } l_p(x) + \gamma\upsilon_p(x) \leq l_q(x) + \gamma\upsilon_q(x) \\
&\leq C_k + B_k + D_k && \text{after neglecting } -\gamma\upsilon_p(x) \text{ and since } \sum_p x_p = d = 1 \\
&\leq (A_1 + \cdots + A_\eta) - (B_1 + \cdots + B_{\eta-1}) && \text{using the results from (2.1) and (2.2)} \\
&\leq \sum_{i=1}^{\eta} \sum_{e\in A_i} (l_e(x_e) + \gamma\upsilon_e(x_e)) - \sum_{i=1}^{\eta-1} \sum_{e\in B_i} l_e(x_e) && \text{neglecting variances in negative term} \\
&\leq (1+\gamma\kappa) \sum_{e\in A\cap\pi} l_e(x_e) - \sum_{e\in B\cap\pi} l_e(x_e) && \text{applying the var-to-mean bound } \kappa.
\end{aligned}$$

□

The last lemma leading to the statement of the main theorem consists of finding a lower bound for the social cost of a RNWE z.

Lemma 6. *Letting π be an alternating path, the social cost of an RNWE z satisfies $C(z) \geq \sum_{e\in A\cap\pi} l_e(z_e) - \sum_{e\in B\cap\pi} l_e(z_e)$.*

Proof. The proof of this lemma is akin to the one used for Lemma 6. Since for any $e \in A_k$ we have $z_e > 0$ by definition, there must be a subpath $C_{k-1} - A_k$ used under z. If $C_k - B_k$ is an alternative route between the same endpoints, then it must be that $l_{C_{k-1}}(z) + l_{A_k}(z) \leq l_{C_k}(z) + l_{B_k}(z)$. Thus, it follows

$$l_{C_k}(z) \geq l_{C_{k-1}}(z) + l_{A_k}(z) - l_{B_k}(z)$$

which, after summing up for all k up to $\eta - 1$ and recalling that $C_0 = \varnothing$, yields $l_{C_{\eta-1}}(z) \geq \sum_{k=1}^{\eta-1}(l_{A_k}(z) - l_{B_k}(z))$. This last inequality proves Lemma 4 since $C(z) = l_{C_{\eta-1}}(z) + l_{A_\eta}(z)$. □

Now, we have all the lemmas we need in order to prove the theorem giving the structural upper bound of a PRA.

Theorem 3 (A Structural Upper Bound). *Consider a general instance with var-to-mean ratio at equilibrium less or equal to κ. Let π be an alternating path, then PRA $\leq 1+\gamma\kappa\eta$, where η identifies the number of disjoint forward subpaths in π,* i.e. *all $A_i \subseteq A$.*

Proof. The statement of the theorem follows by using the previous lemmas as such:

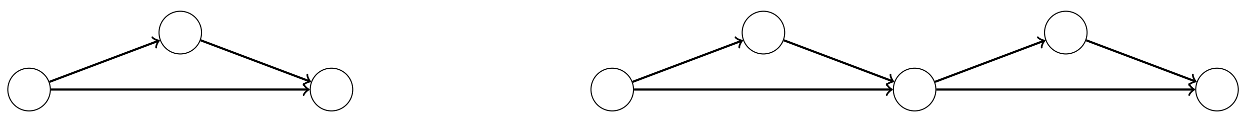
$$
\begin{aligned}
C(x) &\leq (1+\gamma\kappa) \sum_{e\in A\cap\pi} l_e(x_e) - \sum_{e\in B\cap\pi} l_e(x_e) && \text{by Lemma 3} \\
&\leq (1+\gamma\kappa) \sum_{e\in A\cap\pi} l_e(z_e) - \sum_{e\in B\cap\pi} l_e(z_e) && \text{given that in } A\ x_e \leq z_e \text{ and in } B\ z_e < x_e \\
&\leq C(z) + \gamma\kappa \sum_{e\in A\cap\pi} l_e(z_e) && \text{by Lemma 4} \\
&\leq C(z) + \gamma\kappa\eta C(z) = (1+\gamma\kappa\eta)C(z). && \text{since } d=1 \text{ and hence } l_e(z) \leq C(z)\ \forall e \in A\cap\pi
\end{aligned}
$$

□

Since the maximum number η of alternating forward subpaths is $\lceil(n-1)/2\rceil$, where n is the number of nodes in a given instance, the following corollary ensues:

Corollary 3.1. *The PRA in a general graph is upper bounded by* $1+\gamma\kappa\lceil(n-1)/2\rceil$.

Definition 2.5 (Series-parallel graphs)**.** Series-parallel graphs are graphs with two distinguished vertices (*i.e.* the source and the sink), also called *terminals*, formed by a sequence of series and parallel composition operations.
Let's consider a pair of two-terminal graphs (TTGs), X and Y.
The *series composition* consists in merging the sink of X with the source of Y, where the source of X becomes the source of the new graph and the sink of Y its sink.
The *parallel composition* consists in merging the sources of X and Y to create the source of the new graph, and merging the sinks of X and Y to create its sink.



(a) (b)

Figure 2.1: A series composition (2.1b) made by two copies of 2.1a.

(a) (b)

Figure 2.2: A parallel composition (2.2b) made by two copies of 2.2a.

Since in all series-parallel graphs the alternating path π must contain only forward edges in A, it follows that:

Corollary 3.2. *The PRA among all series-parallel instances is exactly* $1 + \gamma\kappa$*.*

2.3 A matching structural lower bound

The structural upper bounds provided in the previous section have corresponding matching lower bounds which are going to be proven in this chapter and which make the strucutral upper bounds of Theorem 3 and Corollary 3.1 tight, respectively.

Theorem 4. *For every positive integer i and every demand pair $r_A^i,\ r_N^i \in \mathbb{R}_{>0}$ such that $2^i r_A^i > (2^i - 1) r_N^i$, there exists a graph instance $G^i(r_A^i,\ r_N^i)$ with 2^{i+1} nodes that satisfies the following two properties:*

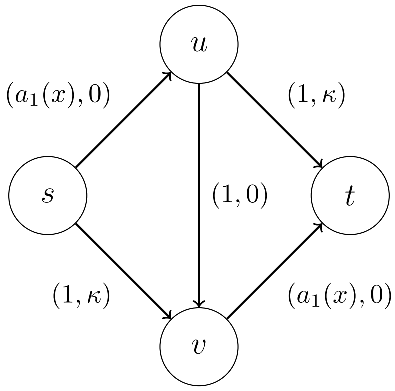
- *If r_A^i risk-averse players are routed through $G^i(r_A^i,\ r_N^i)$, then the mean-variance cost along used paths under the RAWE x, as well as the expected latency along used paths, is $1+2^i\gamma\kappa$. The social cost is thus $C(x) = (1 + 2^i\gamma\kappa)r_A^i$.*
- *If r_N^i risk-neutral players are routed through $G^i(r_A^i,\ r_N^i)$, then the expected latency along used paths under the RNWE z is 1. The social cost is thus $C(z) = r_N^i$.*

Proof. The proof is by induction on i. First, the base case $i = 1$ is proved, then in the inductive step it is shown that if the result of the theorem holds for the case $i-1$, then it must be satified for the case i too. The proof constructs the base case $i = 1$ as a Braess graph G^1 with 2^2 nodes and every graph G^i is built through two copies of G^{i-1} connected in a Braess-line fashion as shown in Figure 2.3.

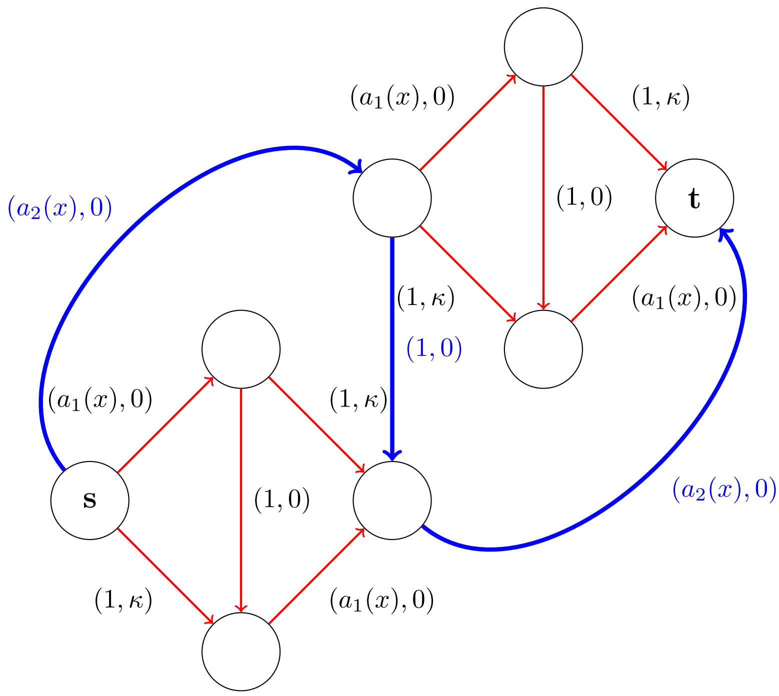
Base case

Let's consider the base case $i = 1$ and the graph 2.3a. If we assume that the requirements of the theorem are satisfied, then if we take any demand pair r_A^i, r_N^i such that $r_A^1 > r_N^1/2$ and we establish a mean latency function $a_1(x)$ such that $a_1(x)$ is strictly increasing for $x \geq r_N^1/2$ and such that $a_1(\frac{r_N^1}{2}) = 0$ and $a_1(r_A^1) = \gamma\kappa$, then:

- The flow x under the RAWE routes r_A^1 players along the zigzag path, causing the upper left and lower right edge to have each a cost $a_1(r_A^1) = \gamma\kappa$ and thus a total cost for the whole path of $1 + 2\gamma\kappa$ and a social cost $C(x) = (1 + 2\gamma\kappa)r_A^1$;
- The flow z under the RNWE splits the r_N^1 players between the top and bottom paths, yielding a total cost for each player of $a_1(r^1/2) + 1 = 1$ and a social cost $C(z) = r_N^1$.



(a) Braess graph for the instance $G^1(r_A^1, r_N^1)$ of the base case $i = 1$.



(b) The instance G^i for $i = 2$ constructed using to copies of G^1 (in red). Blue edges show how the two copies are connected to form G^2.

Figure 2.3

The proof of the base case is thus complete.

Inductive step on i

Now, assume we have an instance satifying the properties of the theorem for $i-1$ and we want to construct the instance for the case i. Let's set $r_A^{i-1} = (2^i r_A^i - r_N^i)/2^{i+1}$ and $r_N^{i-1} = r_N^i/2$, so that the requirement of the theorem that $r_A^{i-1} > ((2^{i-1}-1)/2^{i-1})r_N^{i-1}$ is satisfied. Let's construct a graph $G^i(r_A^i, r_N^i)$ from two copies of the graph G^{i-1} connected in a Braess-like fashion as shown in 2.3b.
Let's define the mean latency function $a_i(x)$ as follows: it is strictly increasing for $x \geq r_N^i/2$, $a_i(r_N^i/2) = 0$ and $a_i(r_A^i/2 + r_N^i/2^{i+1}) = 2^{i-1}\gamma\kappa$. Then we have the following:

- Under the RAWE x, the r_A^i players are routed as follows: $r_A^i/2 - r_N^i/2^{i+1}$ take the upper path, $r_N^i/2^i$ take the zigzag path and $r_A^i/2 - r_N^i/2^{i+1}$ the lower path. Hence, the upper left and lower right edge will each have a mean-variance cost, as well as expected latency, of $a_i(r_A^i/2 - r_N^i/2^{i+1} + r_N^i/2^i) = a_i(r_A^i/2 + r_N^i/2^{i+1}) = 2^{i-1}\gamma\kappa$. By induction, the flow inside each of the copies $G^{i-1}(r_A^{i-1}, r_N^{i-1})$ under the RAWE has a perceived cost for players of $1 + 2^{i-1}\gamma\kappa$. Hence, the path cost that players perceive in $G^i(r_A^i, r_N^i)$ under the RAWE is $1 + 2^{i-1}\gamma\kappa + 2^{i-1}\gamma\kappa = 1 + 2^i\gamma\kappa$, proving the first property of the theorem.
- On the other hand, under the RNWE z for the instance i, the r_N^i players are split equally between the top and bottom paths, yielding a perceived path cost for players of $a_i(r_N^i/2) = 0$ in either path and, by induction, passing through either of both copies $G^{i-1}(r_A^{i-1}, r_N^{i-1})$ has a cost of 1, thus implying $C(z) = r_N^i/2 \cdot 1 + r_N^i/2 \cdot 1 = r_N^i$, proving the second property of the theorem.

□

Using the results of this theorem, it is possible to find the matching lower bound to the PRA upper bound of the previous section.

Corollary 4.1. *If we choose any $n_0 \in \mathbb{N}$, there exists an instance on a graph G with $n \geq n_0$ vertices such that $\frac{C(x)}{C(z)} \geq 1 + \gamma\kappa\lceil(n-1)/2\rceil$.*

Proof. Let's apply Theorem 4 taking $r_A^i = r_N^i = d$ and $i = \{\min j \in \mathbb{N} : n_0 \leq 2^j\}$ to get instance $G^i(r_A^i, r_N^i)$. Then, by the two properties of the theorem we have

$$\frac{C(x)}{C(z)} = (1 + 2^i\gamma\kappa) = 1 + \gamma\kappa\frac{n}{2},$$

where the second equality follows from the fact that the instance G^i has $n = 2^{i+1}$ vertices by construction. □

This lower bound and the upper bound of Corollary 3.1, show the tightness of the PRA for the set of instances with graphs with up to n vertices when n is a power of 2, since in that case $\lceil (n-1)/2 \rceil = n/2$ always holds.

Theorem 5. *The upper bound for the PRA shown in Corollary 3.1 (*i.e. $1+\gamma\kappa\eta$*) and the lower bound shown in Corollary 4.1 coincide for graphs of size n equal to a power of 2. Otherwise, the gap between them is less than 2.*

Indeed, considering an instance with $2^{i+1} = 2\eta$ vertices, the only alternating path has exactly $2^i = \eta$ disjoint forward subpaths. Thus, $2\eta = n$ and Corollary 4.1 is verified.

Chapter 3

How to deal with the PoA

Mechanism design studies ways in which given desired objectives can be reached in a game where individual players act rationally and motivated solely by self-interest. Mechanism disegn has also been called *'inverse game theory'*, since instead of studying the outcomes of a given interaction game, it starts from given desidered outcomes and it goes backwards trying to design a mechanism in the game that would lead to those outcomes. When the desidered objective is to maximize social welfare, as a social planner would like to pursue, incentives are used that lead to the alignment of the individual's interest with the societal one.

In our selfish-routing game, we have seen that the aim of road users of minimizing their travel time often leads to the inefficiency of the system, since total social cost fails to be minimized. From a management point of view, this may call the need for centralized control routing traffic, through the actions of a social planner. However, with the right incentives, the role of this third actor may be avoided and the efficiency goal reached nonetheless.

Since as we have seen in Subsection 1.3.1, the inefficiency of the system, measured by the PoA, can be very large when cost functions are highly non-linear (Roughgarden and Tardos 2002), mechanisms have been developed in order to induce an optimal flow in a selfish-routing game without implying the need of a centralized control. The most well-known techniques to reduce the inefficiency of selfish-routing games in the non-atomic instances are *i)* imposing taxes on edges through marginal cost pricing and *ii)* augmenting the capacity of the network. However, when it comes to the first technique, problems may arise when the exact demand on the network is not known or when we introduce stochastic elements in our model. Thus, after viewing the classical model of marginal cost pricing in the next section, we present two other methods for imposing tolls inducing the optimal flow when we don't have th exact knowledge of the demand function on a network and when we are no more in a deterministic environment. Finally, in the last section we discuss the effects of an increase in the capacity of the network and how it compares to the original game.

3.1 Edge taxes and marginal cost pricing

As we have seen in Section 1.1, the failure of the equilibrium flow with respect to the optimal one is that players do not take into account the additional cost they induce on the other players when selecting a specific path and as such act *'selfishly'* and do not optimize the social cost function (1.2).
With marginal cost pricing, each player is charged for the additional cost they impose on other users of the edge they decide to take, thus incorporating this *'altruistic'* factor in their calculation to minimize the expected latency of their path.

The game with edge taxes is nothing but the same as Section 1.2 with modified cost functions on its edges. We assume that cost functions c_e on edges are differentiable and we impose a non-negative tax τ_e on each edge defined as $\tau_e := f_e \cdot c\,'_e(f_e)$, where $c\,'_e$ denotes the derivative of c_e. The term $c\,'_e(f_e)$ is the marginal increase in cost caused by the additional user of edge e, and f_e is the amount of traffic affected by such an increase. An equilibrium flow of such a game is the equilibrium flow of the instance (G, r, c^τ) where $c_e^\tau = c_e(x) + \tau_e, \forall x \geq 0$ is a shifted version of the original cost function c_e.
It is easy to see that, in such an instance, players selfishly minimize the sum of the edge costs and edge taxes, that corresponds exactly to the First Order Condition of the social objective function we defined in (1.2). Indeed, the equilibrium flow for the instance (G, r, c^τ) arise as the minimizer of the potential function $\Phi(f) = \sum_{e\in E} \int_0^{f_e} c_e(x) + \tau_e \, dx$, which is exactly equal to the social objective function $C(f) = \sum_{e\in E} c_e(f_e) f_e$.

Theorem 6. *Let (G, r, c) be a nonatomic instance such that, for every edge e, the function $x \cdot c_e(x)$ is convex and continuously differentiable. Let f^* be an optimal flow for (G, r, c) and let $\tau_e = f_e^* \cdot c_e'(f_e^*)$ denote the marginal cost tax for edge e with respect to f^*. Then f^* is an equilibrium flow for $(G, r, c + \tau)$.*

Even though this solution to deal with traffic congestion may be very appealing, it has three noteworthy drawbacks: first, it implicitly assumes network users tradeoff cost (in terms of time) and taxes in the same way, which might not be the case; secondly, for certain classes of latency functions derivatives can be very large and not credible or feasible in a real context; and finally, it relies on the exact knowledge of the demand function and origin-destination demand, which may not be known in reality and which may actually consist of stochastic information.

3.2 Optimal tolls with uncertain travel time and demand

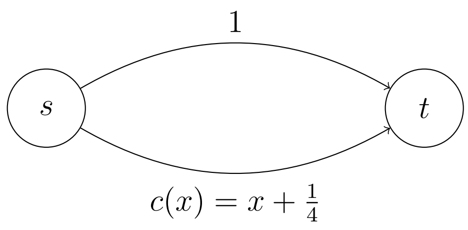
In this section, we will focus on systems of tolls developed so as to work in contexts where the exact demand on the network is not known or where we have stochastic traffic flow information. Namely, the first subsection deals with the existence of demand-independent optimal tolls (DIOTs) and relies on the paper of Colini-Baldeschi, Klimm, and Scarsini (2018); in the second subsection, a method that computes the marginal cost tolls necessary to ensure the convergence of the tolls and flows to the system optimum when link flows and travel times are stochastic is presented, based on the findings of Xu and Sumalee (2015).

3.2.1 Demand-Independent Optimal Tolls

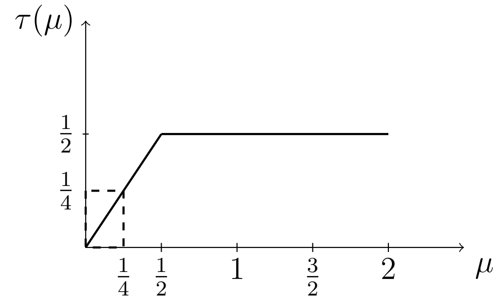
As it was seen in the previous subsection, the classic approach of marginal cost pricing, also called *first-best road pricing*, ensures the convergence of the network system to the social optimum whenever a toll equal to the difference between the marginal social cost and the marginal private cost at the system optimum flow is imposed on each edge (*i.e.* $\tau_e = f_e^* \cdot c'(f_e^*)$). However, the knowledge of the latter relies on the exact knowledge of the travel demand.

For example, if we consider the Pigou example of 1.2, the marginal cost function on the lower edge according to the demand μ is $\min\{\mu, \frac{1}{2}\}$, as Fig. 3.1b shows. The upper edge has no toll since $c'(\cdot) = 0$. Now, suppose that the central planner predicts a travel demand of $\mu = \frac{1}{4}$ from s to t and thus imposes a toll $\tau = \mu$ on the lower edge. However, the actual demand μ^{act} turns out to be 1. Wardrop equilibrium wants that $\frac{1}{4}$ of the demand will take the upper edge, and $\frac{3}{4}$ of the demand will take the lower edge (the new game is shown in Fig. 3.1a). From such an equilibrium, it follows a total cost $C(f) = \frac{1}{4} \cdot 1 + \frac{3}{4} \cdot \frac{3}{4} = \frac{13}{16}$, whereas the optimal flow when $\mu = 1$ is actually that the demand splits equally between the upper and lower edge, which can be achieved by imposing a toll $\tau = \frac{1}{2}$ on the lower edge, and has a total cost $C(f^*) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{8} < \frac{13}{16}$.

Colini-Baldeschi, Klimm, and Scarsini tackled the research question of finding tolls ensuring the induction of the optimal flow which are independent of the travel demand μ (the so-called *demand-independent optimum tolls* or DIOTs). They came up with the existence of DIOTs whenever the edge cost functions are of a BPR-type (*i.e.* a type of functions typically used to model traffic network behaviour) and they provided both a sufficient and necessary condition for them to exist. Furthermore, they also ensured the existence of non-negative DIOTs for directed acyclic multi-graphs with BPR cost functions and the existence of DIOTs subject to a budget constraint of non-negativity which is important for public policy implications. In what follows, the sufficient and necessary conditions for the existence of DIOTs when we have



(a) Routing game with toll $\mu = \frac{1}{4}$ imposed on the lower edge; $\mu^{act} = 1$.



(b) Marginal cost toll function for the lower edge: $\tau(\mu) = \mu$ if $\mu \in [0, \frac{1}{2}]$; $\tau(\mu) = \frac{1}{2}\ \forall \mu > \frac{1}{2}$, or $\tau(\mu) = \min\{\mu, \frac{1}{2}\}$.

Figure 3.1

BPR-type cost functions are presented, along with their proofs.

DIOTs with BPR-type cost functions - A sufficient condition

Definition 3.1 (BPR cost function)**.** A BPR cost function of degree β is an element of the following set:

$$\mathcal{C}_{BPR}(\beta) = \{c : \mathbb{R}_+ \to \mathbb{R}_+ : c(x) = t_c + a_c x^\beta \ \forall x \geq 0 \ , a \ , t \in \mathbb{R}_+\} \ ,$$

where x is the traffic flow along the edge, $t \geq 0$ is the free-flow travel time and a_c denotes the practical capacity of the edge considered, whereas β is a parameter fitted to the model.

The acronym BPR stands for Bureau of Public Roads, since they were put forward by the U.S. Bureau of Public Roads as link performance functions for their use in traffic models.

For the network model, the same model as Section 1.1 will be used, with the only adding of $\tilde{\mathcal{P}}^i = \{p \in \mathcal{P}^i : \tilde{\boldsymbol{f}}_{\boldsymbol{p}}(\boldsymbol{\mu}) > 0\}$ for some demand vector $\boldsymbol{\mu} = (\mu^i)_{i \in \mathcal{I}}$ and corresponding social optimum $\tilde{\boldsymbol{f}}(\boldsymbol{\mu})$ to denote the set of paths eventually used in an optimum flow for some demand vector $\boldsymbol{\mu}$ and for an O-D pair $i \in \mathcal{I}$.

Definition 3.2 (Equilibrium with tolls)**.** Let's consider the modified edge cost function with tolls $c_e^\tau(x_e) := c_e(x_e) + \tau_e$. Define $\mathcal{T}^\tau := (G, \mathcal{I}, \boldsymbol{c}^{\boldsymbol{\tau}})$, where $\boldsymbol{\tau}$ is the toll vector $\boldsymbol{\tau} = (\tau_e)_{e \in E} \in \mathbb{R}^E$. Then, f^τ is an equilibrium flow of $\mathcal{T}^\tau$ if, for all $i \in \mathcal{I}$, we have that

$$c_p^\tau(\boldsymbol{f}^{\boldsymbol{\tau}}) \leq c_{p'}^\tau(\boldsymbol{f}^{\boldsymbol{\tau}})$$

for all $p, p' \in \mathcal{P}^i$ such that $f_p^\tau > 0$.

Definition 3.3 (Demand-independent optimal tolls)**.** Let $\mathcal{T} = (G,\ \mathcal{I},\ c)$. A toll vector $\boldsymbol{\tau} \in \mathbb{R}^E$ is called demand-independent optimal toll (DIOT) for $\mathcal{T}$ if for every demand vector $\boldsymbol{\mu} \in \mathbb{R}_+^{\mathcal{I}}$, every corresponding equilibrium with tolls $\boldsymbol{f}^{\boldsymbol{\tau}}(\boldsymbol{\mu}) \in Eq(\mathcal{T}^{\tau})$ is optimal for $\mathcal{T}$, that is,

$$C\left(\boldsymbol{f}^{\boldsymbol{\tau}}(\boldsymbol{\mu})\right) = \sum_{p\in\mathcal{P}} f_p^{\boldsymbol{\tau}}(\boldsymbol{\mu}) c_p(\boldsymbol{f}^{\boldsymbol{\tau}}(\boldsymbol{\mu})) \leq C(\boldsymbol{f}(\boldsymbol{\mu})) = \sum_{p\in\mathcal{P}} f_p c_p(\boldsymbol{f}^{\boldsymbol{\tau}}(\boldsymbol{\mu})) \text{ for all } \boldsymbol{f}(\boldsymbol{\mu}) \in \mathcal{F}(\boldsymbol{\mu}),$$

where $\mathcal{F}(\boldsymbol{\mu})$ denotes the set of feasible flows for $\boldsymbol{\mu}$ and is defined as $\mathcal{F}(\boldsymbol{\mu}) = \{\boldsymbol{f} \in \mathbb{R}_+^{\mathcal{P}} : \sum_{p\in\mathcal{P}^i} f_p = \mu^i \quad \forall i \in \mathcal{I}\}$.

Now, we will proceed with the statement of the *sufficient condition* for BPR cost functions routing games to admit a DIOT.

Theorem 7. *Let $\mathcal{T} = (G,\ \mathcal{I},\ c)$ and $c_e \in \mathcal{C}_{BPR}(\beta)$ for all $e \in E$. Let $\boldsymbol{\tau}$ be a toll vector such that*

$$\sum_{e\in p}\left(\tau_e + \frac{\beta}{\beta+1} t_e\right) \leq \sum_{e\in p'}\left(\tau_e + \frac{\beta}{\beta+1} t_e\right)$$

$\forall i \in \mathcal{I}$ and $\forall\, p \in \tilde{\mathcal{P}}^i$ and $\forall\, p' \in \mathcal{P}^i$. Then, $\boldsymbol{\tau}$ is a DIOT.

Proof. Take a demand vector $\boldsymbol{\mu} \in \mathbb{R}_+^{\mathcal{I}}$, a corresponding optimum flow $\tilde{\boldsymbol{f}}(\boldsymbol{\mu})$ in $\mathcal{T}$ and denote by $\tilde{x}_e = \sum_{p\in\mathcal{P}:e\in p} \tilde{f}_p(\boldsymbol{\mu})$ the total load imposed on edge e under $\tilde{\boldsymbol{f}}(\boldsymbol{\mu})$. Then, for all $i \in \mathcal{I}$ and all $p, p' \in \mathcal{P}^i$ with $\tilde{f}_p(\boldsymbol{\mu}) > 0$,

$$\sum_{e\in p} c_e(\tilde{x}_e) + c_e'(\tilde{x}_e)\tilde{x}_e \leq \sum_{e\in p'} c_e(\tilde{x}_e) + c_e'(\tilde{x}_e)\tilde{x}_e,$$

since $\tilde{\boldsymbol{f}}(\boldsymbol{\mu})$ is a local optimum and this condition is implied by the equivalence of equilibrium and optimal flows (see Corollary 0.1).
This condition implies it is always possible to find a $\lambda(p,p') \geq 0$ such that

$$\lambda(p,p') + \sum_{e\in p} c_e(\tilde{x}_e) + c_e'(\tilde{x}_e)\tilde{x}_e = \sum_{e\in p'} c_e(\tilde{x}_e) + c_e'(\tilde{x}_e)\tilde{x}_e$$

or, equivalently,

$$\lambda(p,p') + \sum_{e\in p}\left((\beta+1)a_e\tilde{x}_e^{\beta} + t_e\right) = \sum_{e\in p'}\left((\beta+1)a_e\tilde{x}_e^{\beta} + t_e\right) \tag{3.1}$$

The next step consists in showing that when $\boldsymbol{\tau}$ satisfies the condition of Theorem 7, then $\tilde{\boldsymbol{f}}(\boldsymbol{\mu})$ is a Wardrop equilibrium of $\mathcal{T}^{\tau}$.

First, we know that

$$\sum_{e\in p}(c_e(\tilde{x}_e) + \tau_e) = \sum_{e\in p}(a_e\tilde{x}_e^{\beta} + t_e + \tau_e)$$

which, by adding and subtracting $\beta a_e \tilde{x}_e^\beta \quad \forall\, e \in p$ is equal to:

$$
\begin{aligned}
&= \sum_{e\in p} \left((\beta+1)a_e\tilde{x}_e^\beta + t_e\right) - \sum_{e\in p}(\beta a_e\tilde{x}_e^\beta - \tau_e) \\
&= \sum_{e\in p'} c_e(\tilde{x}_e) + c_e'(\tilde{x}_e)\tilde{x}_e - \sum_{e\in p}(\beta a_e\tilde{x}_e^\beta - \tau_e) - \lambda(p,p') && \text{by (3.1)}
\end{aligned}
$$

Finally, by adding and subtracting $\tau_e \quad \forall e \in p'$, we have:

$$
= \sum_{e\in p'}(c_e(\tilde{x}_e) + \tau_e) + \sum_{e\in p'}(\beta a_e\tilde{x}_e^\beta - \tau_e) - \sum_{e\in p}(\beta a_e\tilde{x}_e^\beta - \tau_e) - \lambda(p,p') \tag{3.2}
$$

Since by assumption $\boldsymbol{\tau}$ satisfies the conditions of the theorem, by bringing $\sum_{e\in p'}\tau_e$ to the LHS and $\sum_{e\in p}\frac{\beta}{\beta+1}t_e$ to the RHS, we have:

$$
\sum_{e\in p}\tau_e - \sum_{e\in p'}\tau_e \le -\frac{\beta}{\beta+1}\left(\sum_{e\in p} t_e - \sum_{e\in p'} t_e\right). \tag{3.3}
$$

Therefore,

$$
\begin{aligned}
\sum_{e\in p}(c_e(\tilde{x}_e) + \tau_e) &\le \sum_{e\in p'}\left(c_e(\tilde{x}_e) + \tau_e + \beta a_e\tilde{x}_e^\beta + \frac{\beta}{\beta+1}t_e\right) - \\
&\quad - \sum_{e\in p}\left(\beta a_e\tilde{x}_e^\beta + \frac{\beta}{\beta+1}t_e\right) - \lambda(p,p') && \text{by (3.2) and (3.3)} \\
&= \sum_{e\in p'}(c_e(\tilde{x}_e) + \tau_e) + \frac{\beta}{\beta+1}\lambda(p,p') - \lambda(p,p') && \text{by (3.1)} \\
&= \sum_{e\in p'}(c_e(\tilde{x}_e) + \tau_e) - \frac{1}{\beta+1}\lambda(p,p').
\end{aligned}
$$

Since $\lambda(p,p') \ge 0$, the statement of the theorem follows, *i.e.* it implies $\sum_{e\in p}(c_e(\tilde{x}_e) + \tau_e) - \sum_{e\in p'}(c_e(\tilde{x}_e) + \tau_e) \le 0$ which is what the Wardrop equilibrium of $\mathcal{T}^\tau$ implies for $\tilde{f}_p(\boldsymbol{\mu}) > 0$. $\square$

DIOTs with BPR-type cost functions - A necessary condition

The result of the previous theorem is that $\sum_{e\in p}\tau_e + \frac{\beta}{\beta+1}t_e$ must be the same for all paths $p \in \tilde{P}^i$ and all O-D pairs $i \in \mathcal{I}$. The next theorem shows that this is actually a *necessary* condition for a DIOT to exist.

Theorem 8. *Let $\mathcal{T} = (G,\ \mathcal{I},\ c)$ be a game with BPR cost functions. If $\boldsymbol{\tau}$ is a DIOT for $\mathcal{I}$, then*

$$
\sum_{e\in p}\left(\tau_e + \frac{\beta}{\beta+1}t_e\right) = \sum_{e\in p^*}\left(\tau_e + \frac{\beta}{\beta+1}t_e\right)
$$

for all $i \in \mathcal{I}$ and all $p, p^* \in \tilde{\mathcal{P}}^i$.

The proof of this theorem is based on the continuity of the path flow functions of a Wardrop equilibrium (see Hall (1978)) and the fact that $L^i(p),\ p \in \mathcal{P}^i$ are open and cover the compact [0,1], where $L^i(p) = \bigcup_{S \in 2^{\mathcal{P}^i}: p \in S} L^i(S)$ and $L^i(S) = \{\lambda \in [0,1] : S = S^i(\tilde{\boldsymbol{f}}(\boldsymbol{\mu}(\lambda))\}$ denotes the set of values of λ for which the optimal flow $\tilde{\boldsymbol{f}}(\boldsymbol{\mu}(\lambda))$ has a possible support set $S \in 2^{\mathcal{P}^i}$ for some function $\boldsymbol{\mu}(\lambda), \mu \in [0,1]$ parametrizing the travel demands on a convex combination of $\mu', \mu'' \in \mathbb{R}_+^{\mathcal{I}}$ where $\tilde{f}_{p'}(\mu') > 0$ and $\tilde{f}_{p''}(\mu'') > 0$ for some arbitrary $p', p'' \in \tilde{\mathcal{P}}^i$ (Colini-Baldeschi, Klimm, and Scarsini 2018, Proof of Theorem 4).

From both the sufficient and necessary condition for a DIOT shown respectively by Theorems 7 and 8, it is easy to see that the *trivial* DIOT $\hat{\boldsymbol{\tau}} = (\hat{\tau}_e)_{e \in E}$ where

$$\hat{\tau}_e = -\frac{\beta}{\beta+1} t_e$$

is a solution satisfying both conditions.

From a policy point of view, it means that the central planner needs to subsidize rather than to impose a toll in order to induce the optimal flow on a network. Colini-Baldeschi, Klimm, and Scarsini showed that whenever the routing game consists of a *directed acyclic multi-graph*, or DAMG, which means that there exists a linear ordering $\prec$ of the nodes in G such that, if $v \prec v'$, there is no path going from v' to v in G, then it is always possible to find a non-negative DIOT for $\mathcal{T}$, *i.e.* it is a sufficient condition for its existence.

However, even when a non-negative DIOT does not exist, our social planner may be available to use some of its money to improve traffic conditions, subject to a budget constraint $\sum_{e \in E} \tau_e x_e \geq 0$ for any feasible flow $\boldsymbol{f}$ so that he is sure of not losing money due to this policy. Theorem 9 (Colini-Baldeschi, Klimm, and Scarsini 2018) shows that whenever in the previous games there exists an ordering $\prec$ ensuring that for all $i \in \mathcal{I}$ we have $o_i \prec d_i$, then a DIOT $\boldsymbol{\tau}$ can be found which satisfies the budget constraint and the social planner alike.

3.2.2 Marginal cost pricing in stochastic transportation networks

We have seen that imposing a toll equal to the externality each road user imposes on the network may not be so straightforward if the exact demand of the O-D pair is not known. However, recent studies (see Li (2002) and Yang, Meng, and Lee (2004)) came up with a trial-and-error approach which allows to compute nearly-optimal tolls even when we have limited demand information, under the assumption of a deterministic traffic flow network.

What happens when we are in a stochastic environment? Do the tolls computed under marginal cost pricing for a deterministic network guarantee the convergence of the system to its optimal state even when link flows and travel times are stochastic?

The answer is no, and Xu and Sumalee (2015) showed why a different marginal cost pricing scheme is needed in these cases and how it is computed. Their finding is extremely relevant in that in reality all transportation networks contain stochastic information (*e.g.* the variability of weather conditions, the presence of traffic accidents or road-work...), causing the variability of link flows and travel times and, even without these factors, observational and sampling errors from data collection imply that deterministic link flows and travel time functions may not be found at all in real-life examples.

The authors found a way of implementing a trial-and-error procedure ensuring the convergence of the system to its optimum in a stochastic traffic environment. This alternative marginal cost pricing is called SN-MCP (Stochastic Network - Marginal Cost Pricing) from here thereon, in order to distinguish it from the classical Pigouvian marginal cost pricing previously discussed (see Section 3.1).

The marginal-cost pricing model revisited

As we have seen previously, the general principle underlying marginal cost pricing is charging travellers a toll equal to the difference between the marginal social and private cost they impose on an edge of the network, *i.e.* $\tau_e := f_e \cdot c'_e(f_e) \quad \forall e \in E$.

Now, since the travel demand uncertainty due to the stochastic environment must be included in the model, the aforementioned identity will need to be modified in order to take into account these factors.

Before doing so, we need to introduce a different notation from the one of our classical model seen in Subsection 1.1. Random variables will be referred to in upper-case letters, whereas mean values will be denoted by lower-case letters. As before, $\mathcal{I}$ is the set of all O-D pairs and E is the set of all links. Furthermore, the following letters are introduced: q_i stands for *travel demand for an O-D pair* $i \in \mathcal{I}$, f^i_k for *traffic flow on a path* $k \in \mathcal{P}_i$, v_e for *traffic flow on link* $e \in E$, t_e for *travel time on link* $e \in E$, and $TT = \sum_{e \in E} V_e T_e$ is the total travel time.

For example, if we take the notation for the traffic flow on a given link $e \in E$, we have the following three notations:

- V_e: traffic flow on link $e \in E$, where V is a random variable;
- v_e: mean traffic flow on link $e \in E$;
- $\boldsymbol{v} = \{v_e\}$: vector of mean link flow.

The Stochastic Network-Marginal Cost Pricing (SN-MCP)

When we compare a system's equilibrium and a system's optimum, and we use the equivalence between equilibrium and optimum flows seen in Corollary 0.1, we note as we have already mentioned that what travellers miss in their calcultation for optimizing their travel time is that their total travel cost is actually $\frac{\partial E[TT]}{\partial v_e}$ (the analoguous of computing $\frac{\partial x \cdot c(x)}{\partial x}$ for the deterministic transportation network case) , where $E[TT] = E[\sum_{e\in E} V_e T_e]$ is the expected total travel time, and not just t_e. That is, the extra load that they impose on the link and thus on the other travellers using it, will change the expected total travel time of the network as well. Thus, the marginal external cost, that is the gap between $\frac{\partial E[TT]}{\partial v_e}$ and t_e is what is ignored by self-driven travellers and what we need to impose as a toll in order to make them take it into account for their *'optimization problem'* and guiding them to the system's optimum.

In the paper we are referring to, the link travel time function that have been used are of polynomial form, *i.e.*

$$T_e = t_e(V_e) = \sum_{j=0}^{m} b_{je} V_e^j, \quad \forall e \in E$$

where we recall V_e is the random variable for traffic flow on link $e \in E$.

It follows that the optimal toll, or SN-MCP must be:

$$\begin{aligned} \text{SN-MCP} &= \frac{\partial E[TT]}{\partial v_e} - t_e \\ &= \sum_{j=0}^{m} b_{je} \left(\frac{\partial E[V_e^{j+1}]}{\partial v_e} - E[V_e^j] \right) \end{aligned} \tag{3.4}$$

This equation can be rewritten as follows:

$$\text{SN-MCP} = \sum_{j=0}^{m} b_{je} \left(\frac{\partial E[V_e^{j+1}]}{\partial v_e} - \frac{\partial (E[V_e^j]E[V_e])}{\partial v_e} + E[V_e]\frac{\partial E[V_e^j]}{\partial v_e} \right)$$

simply by using the product rule, which implies that $\frac{\partial (E[V_e^j]E[V_e])}{\partial v_e} = E[V_e]\frac{\partial E[V_e^j]}{\partial v_e} + E[V_e^j]\frac{\partial E[v_e]}{\partial v_e} = E[V_e]\frac{\partial E[V_e^j]}{\partial v_e} + E[V_e^j]$ (since $\frac{\partial E[v_e]}{\partial v_e} = \frac{\partial v_e}{\partial v_e} = 1$), and replacing it in (3.4). Finally, the last product of the equation yields

$$v_e \frac{\partial E[T_e]}{\partial v_e},$$

whereas the sum of the first two gives

$$\sum_{j=0}^{m} b_{je} \left(\frac{\partial E[V_e^{j+1}]}{\partial v_e} - \frac{\partial (E[V_e^j]E[V_e])}{\partial v_e} \right),$$

which is exactly the definition of

$$\sum_{j=0}^{m} b_{je} \frac{\partial \text{Cov}[V_e^j, V_e]}{\partial v_e}.$$

Hence, we have:

$$\text{SN-MCP} = v_e \frac{\partial E[T_e]}{\partial v_e} + \sum_{j=0}^{m} b_{je} \frac{\partial \text{Cov}[V_e^j, V_e]}{\partial v_e} \tag{3.5}$$

The first term of (3.5) is called the *average MCP* and it can be regarded as the stochastic interpretation of the classical MCP, where the deterministic link flows and travel times have been replaced by their respective random variables; on the other hand, the second term is related to the variability of travel demand and is what a deterministic MCP ignores and, as long as this term is positive, underestimates.

The authors showed that, even when the mean and variance of our stochastic travel demand are not known nor available, a trial-and-error method can be used which iteratively computes the marginal-cost toll based on observational stochastic link flows and ensures its convergence to the optimal link toll pattern. We report for completeness the steps of this method and we leave the proof of its convergence to the optimum to the reader, who can find it in subsection *5.1* of Xu and Sumalee (2015).

- *Step 0 (initialization).* Let $\{v_e^{(0)}, e \in E\}$ be an initial set of feasibile mean link flows. Set $k = 0$.
- *Step 1 (estimation of link tolls).* For each $e \in E$, compute $\tau_e^{(k)}$ by (3.5), that is

$$\tau_e^{(k)} = v_e^{(k)} \frac{\partial E[T_e^{(k)}]}{\partial v_e^{(k)}} + \sum_{j=0}^{m} b_{je} \frac{\partial \text{Cov}[(V_e^{(k)})^j, V_e^{(k)}]}{\partial v_e^{(k)}}$$

- *Step 2 (observation of link flows).* Impose the tolls obtained from the previous step on the network. Then after having observed and collected the data on link flows over a period a time, compute the mean and variance of such a set.
  Let $\{\bar{v}_e^{(k)}, e \in E\}$ denoted the estimated mean value.
- *Step 3 (check convergence).* If $\frac{\|\bar{\boldsymbol{v}}^{(k)} - \boldsymbol{v}^{(k)}\|}{\|\boldsymbol{v}^{(k)}\|} < \epsilon$, then stop. $\epsilon > 0$ is chosen arbitrarily and it is the convergence tolerance.
  Otherwise, go to Step 4.
- *Step 4 (updating link flows).* Set

$$v_e^{(k+1)} = v_e^{(k)} + \alpha^{(k)} (\bar{v}_e^{(k)} - v_e^{(k)}), e \in E,$$

  and k:= k+1. $\alpha^{(k)}$ denotes a sequence of predetermined step sizes and is usually set at $1/k$. Then, go back to Step 1 and start again.

Finally, the authors conducted some numerical tests and observed the following facts: the tolls computed through the SN-MCP (both the true and the estimated ones) are higher than the ones computed through the average and the original MCP; the improvements in terms of expected total travel time made by the latters are lower than those obtained with the SN-MCP, and actually, when the network has a high *variance-to-mean ratio* the original MCP increases rathen than decreases expected total travel time; for three different levels of *variance-to-mean ratios*, all trial-and-error processes converge to the true optimal tolls within 10 iterations; on the other hand, the trial-and-error procedure does not lead to such a convergence if we use the average or the original MCP in the process.

These tests showed that it is fundamental to take into account the uncertainty of a transportation network when we want to guide this system to its optimal state, otherwise we would end up underestimating the tolls necessary for its achievement and being locked up in a suboptimal state. Not only this, but the authors were also able to show that such tolls ensure convergence to the optimum even when the mean and variance of link flows are not known but can just be estimated through observational data, and that sampling error won't affect its convergence neither. This last fact is paramount since, as we have seen, in reality networks are indeed stochastic and knowing a priori the mean and variance of link flows is not an available option for the social planner in practice.

3.3 Capacity augmentation

An alternative to edge taxes and to the need of centralized control is to increase the capacity of a network.

Theorem 9. *If f is an equilibrium flow for (G, r, c) and f^* is feasible for $(G, 2r, c)$, then $C(f) \leq C(f^*)$.*

Proof. Let f be the equilibrium flow for (G, r, c) and f^* the optimal flow for $(G, 2r, c)$. Let's define by L the common cost of equilibrium paths. Then, the cost of the equilibrium flow is $C(f) = \sum_{P \in \mathcal{P}} c_P(f) f_P = rL$, where the second equality comes from the fact that demand is satisifed (*i.e.* $\sum_{P \in \mathcal{P}} f_P = r$ and $c_P(f) = L$ for every path $P \in \mathcal{P}$ in equilibrium). On the other hand, the cost of an optimal flow with costant cost on each edge e equal to $c_e(f)$ is $\sum_{P \in \mathcal{P}} f^*_P c_P(f) \geq 2rL$ since in this case $\sum_{P \in \mathcal{P}} f^*_P = 2r$ by definition and $c_P(f) \geq L$ on every path with optimal flow.
Now, it is enough to show that

$$\sum_{e \in E} f^*_e c_e(f^*_e) \geq \sum_{e \in E} f^*_e c_e(f_e) - \sum_{e \in E} f_e c_e(f_e),$$

since it would be equivalent to prove that $C(f^*) \geq 2rL - rL = rL = C(f)$.

To do so, we show that for each $e \in E$,

$$f_e^*[c_e(f_e) - c_e(f_e^*)] \leq f_e c_e(f_e), \tag{3.6}$$

that is, we show that the total error we make in valuing the cost function under optimal flows when using $c_e(f_e)$ instead of $c_e(f_e^*)$ is at most the total cost $C(f)$ under equilibrium flow. Indeed, if $f_e^* \geq f_e$ then the left hand-side of the inequality is less or equal than 0; on the other hand, if $f_e^* < f_e$, then the left hand-side corresponds to the red-filled area in the graph 3.2, which is clearly less then the area of the grey-dashed rectangle corresponding to $f_e c_e(f_e)$. Thus, the inequality (3.3) is verified for every edge $e \in E$ and the proof is complete. □

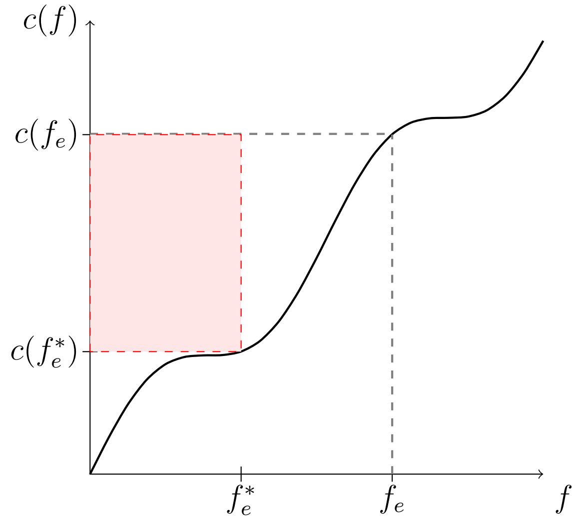


Figure 3.2: Graph of an hypothetical non-decreasing cost function showing the case where $f_e^* < f_e$.

This theorem has an analogous corollary which allows us to make the following inference: by improving the edge technology adding capacity in terms of a faster network, we can achieve an equal if not better result than using centralized control to achieve an optimal flow in the original game.

Corollary 9.1. *Let (G, r, c) be a nonatomic instance and define the modified cost function $\tilde{c}_e(x) = \frac{1}{2}c_e(x/2)$ for each edge e. Let $\tilde{f}$ be an equilibrium flow for $(G, r, \tilde{c})$ with cost $\tilde{C}(\tilde{f})$, and f^* a feasible flow for (G, r, c) with cost $C(f^*)$. Then $\tilde{C}(\tilde{f}) \leq C(f^*)$.*

Proof. Let f be an equilibrium flow for $(G, r/2, c)$ and f^* a flow feasible for (G, r, c). Then, by Theorem 9, $\sum_{e \in E} c_e(f_e)f_e \leq \sum_{e \in E} c_e(f_e^*)f_e^*$. Consider the flow $\tilde{f} = 2f$ a feasible flow for $(G, r, \tilde{c})$. Since $\tilde{c}_e(\tilde{f}_e) = \frac{1}{2}c_e(f_e)$ for each edge $e \in E$ and f is an equilibrium flow for (G, r, c), $\tilde{f}$ is an equilibrium flow for $(G, r, \tilde{c})$. Moreover,

$$\sum_{e \in E} \tilde{c}_e(\tilde{f}_e)\tilde{f}_e = \sum_{e \in E} \left(\frac{1}{2}c_e(f_e)\right)(2f_e) = \sum_{e \in E} c_e(f_e)(f_e),$$

thus showing that $\sum_{e\in E} \tilde{c}_e(\tilde{f}_e)\tilde{f}_e \leq \sum_{e\in E} c_e(f_e^*)f_e^*$. $\square$

If we consider the $M/M/1$ latency functions, where $c_e(x) = (u_e - x)^{-1}$ for $x \in [0, u_e)$, where u_e is the capacity on edge e, and we assume that $f_e < u_e$ on every edge e, then $\tilde{c}_e(x) = \frac{1}{2}(u_e - x/2) = 1/(2u_e - x)$. By Corollary 9.1, this means that to match and to outperform the job of a central planner or the benefit from central control, it is enough to double the capacity of every edge in the network.

Chapter 4

Possible efficiency improvements due to risk-aversion

In this chapter, we are going to present some traffic network examples where we analysed the risk-neutral, risk-averse and optimal flows, and we carried out a comparison among them, computing their PRA or, in certain cases, the PoA as well. As we are going to see, under certain assumptions and variance functions, introducing variance in the model and thereby computing the respective risk-aversion flows, results in a PRA ≤ 1, implying that sometimes risk-aversion may contribute positively to the system's efficiency and performance, and taking into account the variance along the edges may induce flows more close to the optimal ones than the ones obtained in a risk-neutral equilibrium.

4.1 Pigou networks

In the next examples, we modified the classic Pigou network of Section 1.2 by altering the latency function of the lower edge or by introducing variance on it.

The notation we have used is analogous to the one used in the previous chapters. Namely, we denote total demand as d, edge cost functions as $c(x)$, variance as $\upsilon(\cdot)$, the risk-aversion coefficient as γ, the var-to-mean ratio as κ and traffic flows as f. Furthermore, we will refer to the flow taking the lower edge as x, whereas the flow on the upper edge will be denoted by y. In particular, (x^{RNWE}, y^{RNWE}) refer to the risk-neutral flows under equilibrium, (x^{RAWE}, y^{RAWE}) to the risk-averse ones, and (x^{OPT}, y^{OPT}) to the optimal ones. Similarly, C(RNWE) are total costs under a risk-neutral Wardrop equilibrium, C(RAWE) are total costs under risk-averse Wardrop equilibrium and C(OPT) are total costs under optimal flows. If not specified otherwise, we assume the risk-aversion coefficient $\gamma = 1$.

Example 1

This is the basic Pigou example where $d = 1$ and the cost on the upper edge is 1, whereas the cost on the lower edge is $c(x) = x$. We know that in this case the optimal flow on the lower edge is $x^{OPT} = \frac{1}{2}$. In the risk-aversion model, we take the extreme case where $\gamma = 1$ and we consider graph 4.1. From marginal-cost pricing, we know that if we had a toll of $\frac{1}{2}$ on the lower edge, then the optimal flow would be reached. Hence, when $\upsilon(x) = \frac{1}{2}$ the cost perceived on the lower edge becomes $x + \frac{1}{2}$ and the optimal flow equals the RAWE flow. This implies that the variance-to-mean ratio $\kappa = \frac{1}{2} \cdot \frac{2}{1} = 1$ in equilibrium, which however may not be a realistic assumption.

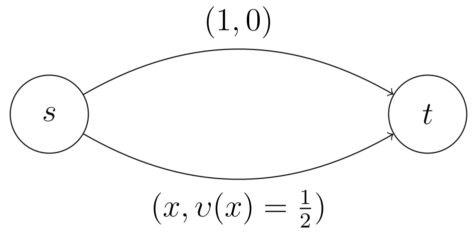


Figure 4.1: Example 1

Example 2

Now, we take the same example as before but we assume a variance described by an inverted U-shaped function $\upsilon(x) = -x^2 + x$. This assumption relies on the data found in Nezamuddin et al. (2009), where both for arterial and freeway roads, empirical data shows that the latency's variance, determined by the variance of the speed (which has a U-shaped function), is first increasing and then, after reaching a peak, decreases till stabilizing at a constant level for very high levels of traffic flow. In particular, we note that, if we take $0 < x \leq 1$, this variance function has a peak at $x = 0.5$ where $\upsilon(x) = 0.25$. This means that in terms of κ, the maximum variance-to-mean ratio is $\kappa = \frac{0.25}{0.5} = 0.5$, which may be a realistic scenario.

Here, we are going to analyse the example for a generic demand $0 < d \leq 1$.

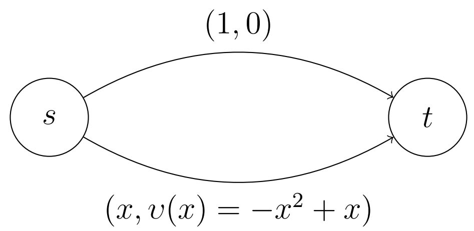


Figure 4.2: Example 2

We have the following flows:

- $x^{OPT} = d$, $y^{OPT} = 0$ for $0 < d \leq \frac{1}{2}$
- $x^{OPT} = \frac{1}{2}$, $y^{OPT} = d - \frac{1}{2}$ for $\frac{1}{2} < d \leq 1$
- $x^{RNWE} = d$, $y^{RNWE} = 0$ for $0 < d \leq 1$
- $x^{RAWE} = d$, $y^{RAWE} = 0$ for $0 < d \leq 1$

As we can notice, the flows under the RNWE and the RAWE coincide. Indeed, observe that the cost on the lower edge for $0 < d \leq 1$, when we take into account variance, is given by the function $c(x) = -x^2 + 2x \leq 1$, implying that the upper edge is not a competitive choice and it is always less preferred to the lower route by risk-averse users, or they're indifferent about it when the flow on the lower edge is $x = 1$. The variance function we have chosen thus implies that risk-averse users will always behave as risk-neutral ones and PRA $= 1$ for $0 < d \leq 1$.

In this case, we can just analyse the PoA for $0 < d \leq 1$.

We can split the total cost under optimal flows in two parts: for $0 < d \leq \frac{1}{2}$, we have C(OPT) $=$ d^2, whereas for $\frac{1}{2} < d \leq 1$, C(OPT) $= d - \frac{1}{4}$. On the other hand, the cost under either RNWE or RAWE is C(RNWE) $=$ C(RAWE) $= d^2$ for any level $0 < d \leq 1$.

The PoA thus becomes PoA $= \dfrac{d^2}{d^2} = 1$ for $0 < d \leq \frac{1}{2}$, and PoA $= \dfrac{d^2}{d - \frac{1}{4}}$ for $\frac{1}{2} < d \leq 1$, which is increasing for $\frac{1}{2} < d \leq 1$ and has as upper bound, for $d = 1$, the PoA of Roughgarden and Tardos (2002) for affine latency functions (*i.e.* PRA $= \frac{4}{3}$).

In this example, being risk-averse instead of risk-neutral does not translate into a worse PoA (*i.e.* a less efficient and performing network), since in equilibrium all risk-averse users would choose the lower edge knowing that it will always cost less than the upper one, even when they take into account its variance.

Example 3

Now, we analyse a Pigou example with general linear latency function $c(x) = \alpha x$ and variance $v(x) = \beta > 0$ on the lower edge. Demand $d = 1$. We assume $0 < \alpha \leq 1$, so the smaller α, the faster the edge. In this way, the lower edge costs always less than the upper one or it is equal to it when $\alpha = 1$. However, the lower edge is characterized by variance β, whereas the upper edge has a fixed cost $1 \geq \alpha x$ but zero variance, a property that our risk-averse users may well want to take into account.

The optimal flow of this network is $x^{OPT} = \frac{1}{2\alpha}, y^{OPT} = 0$ when $\alpha \leq \frac{1}{2}$, and $x^{OPT} = \frac{1}{2\alpha}, y^{OPT} = 1 - \frac{1}{2\alpha}$ when $\alpha > \frac{1}{2}$. From marginal cost pricing, we know that if we impose $\beta = \alpha x$

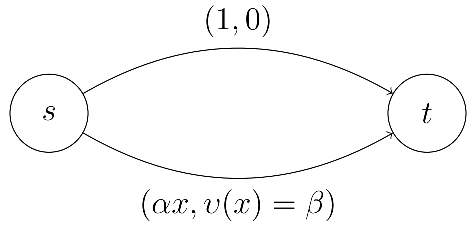


Figure 4.3: Example 3

then the optimal flow is naturally reached by road users. Hence, we are in the same scenario as Figure 4.1, with $\upsilon(x) = \beta = \alpha x$ and $\kappa = 1$ in equilibrium.

However, the aforementioned case assumes $\gamma = 1$. What happens if we take $0 < \gamma \leq 1$? We need to solve the following equation:

$$\gamma\beta = \alpha x$$

which implies that

$$\beta = \frac{1}{\gamma}\alpha x.$$

Hence, when the variance along the lower edge is exactly equal to the RHS of the above equation, the optimal flow would be naturally induced on this network. In this scenario, the var-to-mean ratio $\kappa = \dfrac{\beta}{\alpha x} = \dfrac{\frac{1}{\gamma}\alpha x}{\alpha x} = \dfrac{1}{\gamma} \geq 1$ for $0 < \gamma \leq 1$.

Example 4

Another interesting example is the one where we take the same network of the previous case, but we consider a generic demand $d \in \mathbb{R}^+$ and $\beta \in (0, 1)$. In this case, the optimal flow is $x^{OPT} = d$ on the lower edge and $y^{OPT} = 0$ on the upper one for $0 < d \leq \frac{1}{2\alpha}$, and $x^{OPT} = \frac{1}{2\alpha}$ and $y^{OPT} = d - \frac{1}{2\alpha}$ for $d \in (\frac{1}{2\alpha}, +\infty)$. For what concerns the risk-neutral equilibrium, we have:

- $x^{RNWE} = d$, $y^{RNWE} = 0$ for $0 < d \leq \frac{1}{\alpha}$
- $x^{RNWE} = \frac{1}{\alpha}$, $y^{RNWE} = \frac{\alpha d - 1}{\alpha}$ for $d \in (\frac{1}{\alpha}, +\infty)$

On the other hand, the RAWE solves the following system:

$$\begin{cases} \alpha x + \beta = 1 \\ x = d - y \end{cases}$$

Here, we have the following flows:

- For $0 < d \leq \frac{1-\beta}{\alpha}$:

  - $x^{RAWE} = d$ and $y^{RAWE} = 0$;

- For $d \in (\frac{1-\beta}{\alpha}, +\infty)$:

  - $x^{RAWE} = \frac{1-\beta}{\alpha}$ and $y^{RAWE} = d + \frac{\beta-1}{\alpha}$.

Again, we can split the demand function into three segments to compute the PRA:

- For $0 < d < \frac{1-\beta}{\alpha}$:

  - C(RNWE)$= \alpha d^2$, C(RAWE) $= \alpha d^2 \Rightarrow$ PRA $= 1$;

- For $\frac{1-\beta}{\alpha} < d \leq \frac{1}{\alpha}$:

  - C(RNWE) $= \alpha d^2$, C(RAWE) $= \frac{\alpha d+\beta^2-\beta}{\alpha} \Rightarrow$ PRA $= \frac{1}{\alpha d} + \frac{\beta^2-\beta}{(\alpha d)^2}$ which is decreasing for $0 < \beta < \frac{1}{2}$ and increasing for $\beta \in (\frac{1}{2}, 1)$. In particular, it reaches a minimum at $\beta = \frac{1}{2} \Rightarrow$ PRA $= \frac{1}{\alpha d}(1 - \frac{1}{4\alpha d})$. Furthermore, we know that whenever $\beta^2 - \beta + \alpha d(1 - \alpha d) < 0$ holds, then PRA < 1. If, on the other hand, we want to see how the PRA changes with respect to d, we have $\frac{\partial \text{PRA}}{\partial d} = -\frac{1}{\alpha d^2} - \frac{2(\beta^2-\beta)}{(\alpha d)^3}$, which, for $0 < \beta < 1$, is always decreasing, and it allows us to know that the PRA for this demand interval reaches its minimum at $d = \frac{1}{\alpha} \Rightarrow$ PRA $= 1 + \beta^2 - \beta < 1$.

- Finally, for $d \in (\frac{1}{\alpha}, +\infty)$:

  - C(RNWE)$= d$, C(RAWE) $= \frac{\alpha d+\beta^2-\beta}{\alpha} \Rightarrow$ PRA $= 1 + \frac{\beta^2-\beta}{\alpha d}$, with the interesting result that C(RAWE) $<$ C(RNWE) when $0 < \beta < 1$, and actually when $\beta = \frac{1}{2}$, C(OPT)$= d - \frac{1}{4\alpha}$ and C(RAWE) coincide. On the other hand, if we want to analyse how the PRA changes with respect to d in this demand interval, we have $\frac{\partial \text{PRA}}{\partial d} = \frac{(\beta^2-\beta)}{\alpha d^2}$, implying it is increasing in d.

Example 5

Another Pigou example it may be interesting to analyse is the one reported in Figure 4.4. We analyse the case when $1 \leq d < \frac{\beta}{\alpha}$. We have a lower edge which has always a lower latency than the upper route (*i.e.* $c(x) = \frac{x}{2}$) and has a variance function with undefined parameters α and β, with the only constraint α, β being non-negative numbers. We left these parameters undefined since we want to find the exact value of α and β at which the optimal flow can be self-enforced. By analysing the optimal flow of this instance, we have that when $d \geq 1$, the optimal flow routes exactly $x^{OPT} = 1$ on the lower edge and $y^{OPT} = d-1$ on the upper one, and the toll which would be needed to induce it in equilibrium is $\frac{x}{2}$. Furthermore, since we want our variance to be always positive for $x = 1$, and since the roots of the function $v(x) = -\alpha x^2 + \beta x$

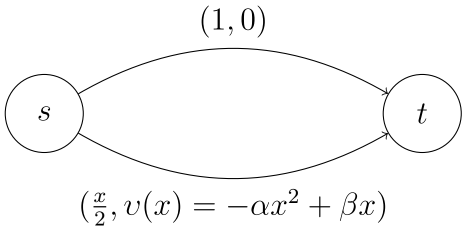


Figure 4.4: Example 5

are $x_1 = 0, x_2 = \frac{\beta}{\alpha}$, we also want $\frac{\beta}{\alpha} > 1 \Rightarrow \beta > \alpha$, which is the reason why we chose to analyse the demand interval $d \in [1, \frac{\beta}{\alpha})$.
The system we want to solve is thus:

$$\begin{cases} -\alpha x^2 + \beta x = \frac{x}{2} \\ x = 1 \end{cases}$$

whose solution is

$$-\alpha + \beta = \frac{1}{2}.$$

Hence, any such relation between α and β will do. If we take the simplest case, with $\alpha = \frac{1}{2}, \beta = 1$, then x is the solution of the equation:

$$-\frac{x^2}{2} + \frac{3x}{2} - 1 = 0,$$

who has roots $x_1 = 0$ and $x_2 = 1$, the second being exactly what we wanted to have.

Example 6

Suppose we have a demand $0 < d \leq 1$ and the following Pigou example where the upper edge has a cost of $\frac{1}{2}$ and zero variance, whereas the lower edge has a cost $c(x) = x$ but variance $\upsilon(x) = -x^2 + x$. This modification of the cost on the upper edge implies that, for risk-averse users and for a demand $d \in (0, 1]$, the upper route is actually an option, since now the cost on the lower edge (variance included) $c(x) = -x^2 + 2x \leq \frac{1}{2}$ only for $x \in (0, 0.29]$.

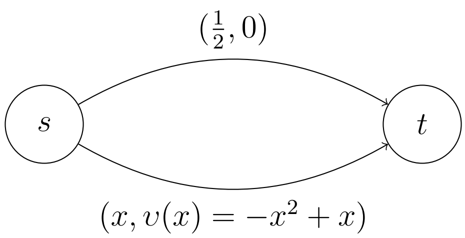


Figure 4.5: Example 6

In this case, assuming $\gamma = 1$ for RAWE, we have the following flows:

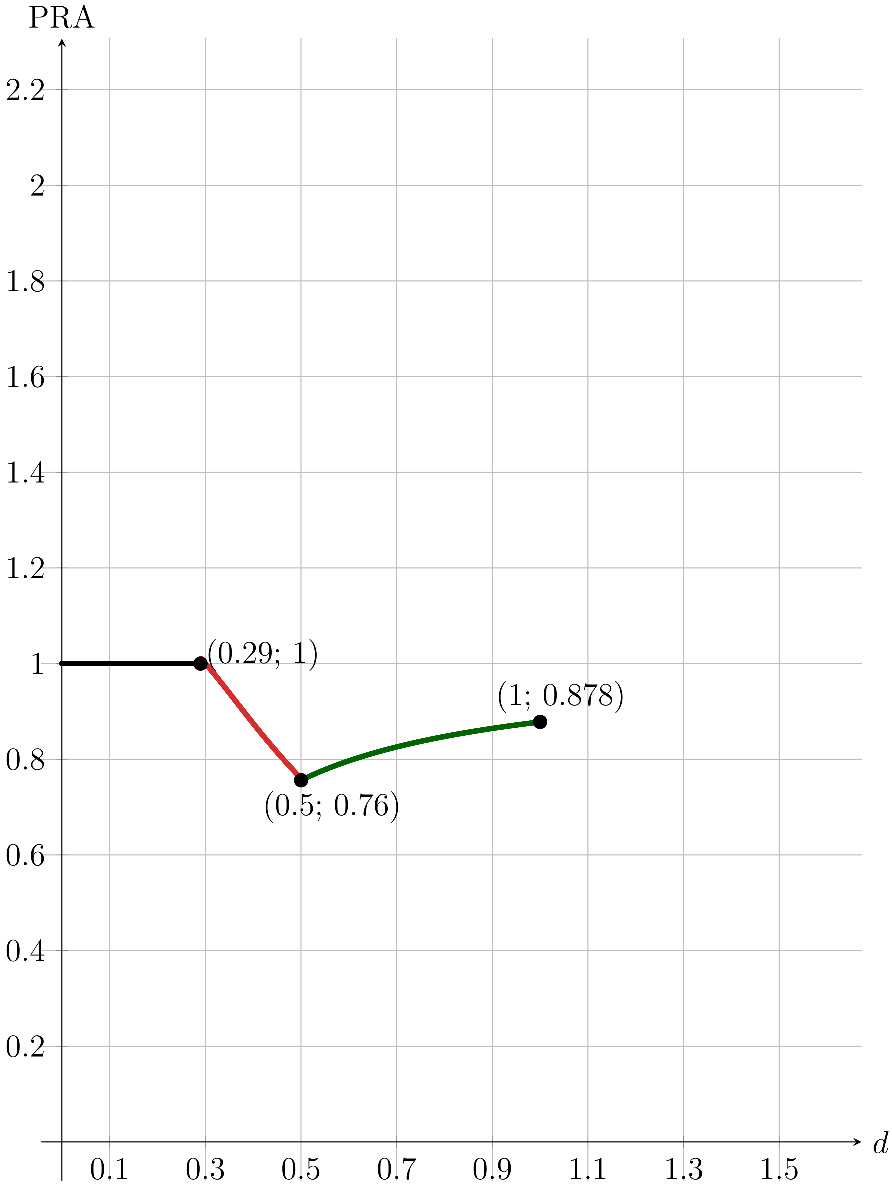


Figure 4.6: Example 6; PRA function for $0 < d \leq 1$.

- $x^{OPT} = d$, $y^{OPT} = 0$ for $0 < d \leq \frac{1}{4}$
- $x^{OPT} = \frac{1}{4}$, $y^{OPT} = d - \frac{1}{4}$ for $\frac{1}{4} < d \leq 1$
- $x^{RNWE} = d$, $y^{RNWE} = 0$ for $0 < d \leq \frac{1}{2}$
- $x^{RNWE} = \frac{1}{2}$, $y^{RNWE} = d - \frac{1}{2}$ for $\frac{1}{2} < d \leq 1$
- $x^{RAWE} = d$, $y^{RAWE} = 0$ for $0 < d \leq 0.29$
- $x^{RAWE} = 0.29$, $y^{RAWE} = d - 0.29$ for $0.29 < d \leq 1$.

In order to analyse the PRA, we can split the demand d into three segments:

- For $0 < d \leq 0.29$, we have:

– C(RNWE) $= d^2$, C(RAWE) $= d^2 \Rightarrow$ PRA $= 1$;

- For $0.29 < d \leq \frac{1}{2}$, we have:
  - C(RNWE)$= d^2$, C(RAWE) $= \frac{d}{2} - 0.061 \Rightarrow$ PRA $= \frac{d-0.12}{2d^2}$, which tends to 1 as $d \to 0.29$ and it is decreasing till reaching a minimum at $d = 0.5 \Rightarrow$ PRA $= 0.76$.
- Finally for $\frac{1}{2} < d \leq 1$, we have:
  - C(RNWE) $= \frac{d}{2}$, C(RAWE)$=\frac{d}{2} - 0.061 \Rightarrow$ PRA < 1 always and has a maximum at $d = 1$, where PRA $= 0.878$.

Hence, in this case being risk-averse improves the performance of the network, and the PRA is always less than or equal to 1 for $0 < d \leq 1$. In particular, if we compare C(OPT) and C(RAWE) for $0.29 < d \leq 1$, we can observe that C(OPT) $= (\frac{1}{4})^2 + (d - \frac{1}{4})\frac{1}{2} = \frac{1}{2}d - \frac{1}{16} \approx$ C(RAWE) $= \frac{1}{2}d - 0.061$.

Example 7

Now, let's see what happens when we modify the lower edge by adding a fixed cost $c > 0$.

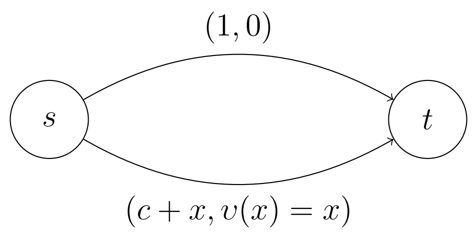


Figure 4.7: Example 7

In this case, we know that if we have a variance $\upsilon(x) = x$, then the optimal flow is induced. Is it plausible? Well, in this case our var-to-mean ratio would be $\kappa = \frac{x}{c+x}$, which is increasing in x but at a decreasing rate, which might be a realistic assumption, given that the variance cannot increase that much when we have very large flows since at a certain moment the capacity of the road would be reached and no additional traffic could be added anymore. In particular, $\lim_{x\to\infty} \dfrac{x}{c+x} = 1$.

Indeed, we have that:

- For $0 < d \leq \frac{1-c}{2}$, $x^{RAWE} = x^{OPT} = d \Rightarrow$ C(RAWE) $=$ C(OPT) $= cd + d^2$
- For $d \in (\frac{1}{2}, +\infty)$,$x^{RAWE} = x^{OPT} = \frac{1-c}{2} \Rightarrow$ C(RAWE) $=$ C(OPT) $= \frac{(-c^2+2c-1)}{4} + d$

In both cases, PRA ≤ 1, since C(RAWE) $=$ C(OPT). Indeed, one can verify that C(RNWE) $= cd+d^2$ when $0 < d \leq 1-c$ and C(RNWE) $= d$ for $d \in (1-c, +\infty)$. In the first case, C(RNWE) $=$ C(RAWE) $=$ C(OPT) $\Rightarrow$ PRA $= 1$, whereas in the second case, C(RNWE) $>$ C(RAWE) $=$ C(OPT), since for any $c \in \mathbb{R}^+$, we have $\frac{(-c^2+2c-1)}{4} \leq 0$.

Example 8

In this example, we consider a generic demand $d \in \mathbb{R}^+$, we modify the cost on the lower edge as $c(x) = \sqrt{x}$ and leave a general variance $\upsilon(x)$. In this case, the optimal flow is $(x^{OPT} = d, y^{OPT} = 0)$ for $0 < d \leq \frac{4}{9}$, and $(x^{OPT} = \frac{4}{9}, y^{OPT} = d - \frac{4}{9})$ for $d \in (\frac{4}{9}, +\infty)$. By marginal cost pricing, the variance needed to induce the optimal flow on the lower edge would be $\upsilon(x) = \frac{1}{2}\sqrt{d}$ in the first case, and $\frac{1}{3}$ in the second case, which are both plausible variance functions. Furthermore, in this case the var-to-mean ratio is always less than 1: in the first case, $\kappa = \frac{\frac{1}{2}\sqrt{d}}{\sqrt{d}} = \frac{1}{2}$, and in the second case $\kappa = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$ as well.

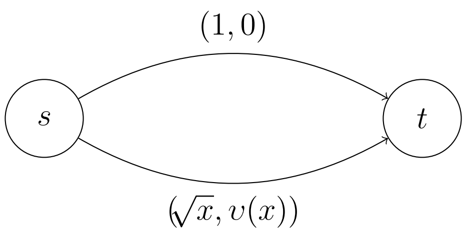


Figure 4.8: Example 8

Example 9

Lets' consider the following example with $\beta < \alpha$, with $\alpha, \beta \in (0, 1]$ and demand $d \in \mathbb{R}^+$.

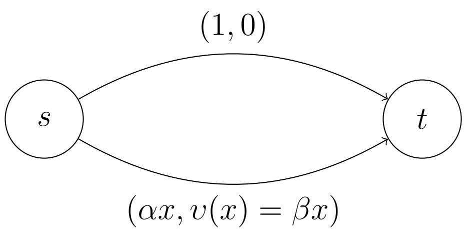


Figure 4.9: Example 9

We have the following flows:

- $x^{RAWE} = d$ and $y^{RAWE} = 0$ for $0 < d \leq \frac{1}{\alpha+\beta}$;

- $x^{RAWE} = \frac{1}{\alpha+\beta}$ and $y^{RAWE} = d - \frac{1}{\alpha+\beta}$ for $d > \frac{1}{\alpha+\beta}$;
- $x^{RNWE} = d$ and $y^{RNWE} = 0$ for $0 < d \leq \frac{1}{\alpha}$;
- $x^{RNWE} = \frac{1}{\alpha}$ and $y^{RNWE} = d - \frac{1}{\alpha}$ for $d > \frac{1}{\alpha}$;

The respective costs and PRA are:

- For $0 < d \leq \frac{1}{\alpha+\beta}$:
  - $\mathrm{C(RAWE)} = \alpha d^2, \mathrm{C(RNWE)} = \alpha d^2 \Rightarrow \mathrm{PRA} = 1$;
- For $\frac{1}{\alpha+\beta} < d \leq \frac{1}{\alpha}$:
  - $\mathrm{C(RAWE)} = d - \frac{\beta}{(\alpha+\beta)^2}, \mathrm{C(RNWE)} = \alpha d^2 \Rightarrow \mathrm{PRA} = \frac{d-\frac{\beta}{(\beta+\alpha)^2}}{\alpha d^2}$, which is always greater than 1 for that demand interval;
- Finally, for $d > \frac{1}{\alpha}$:
  - $\mathrm{C(RAWE)} = d - \frac{\beta}{(\alpha+\beta)^2}, \mathrm{C(RNWE)} = d - \frac{\alpha-1}{\alpha} \Rightarrow \mathrm{PRA} = \frac{d-\frac{\beta}{(\beta+\alpha)^2}}{d-\frac{\alpha-1}{\alpha}}$, which is never less than 1 for values of $\alpha > 1$.

Hence, we have found that for levels of demand $0 < d \leq \frac{1}{\alpha+\beta}$, the equilibrium under risk-aversion does not worsen the system's perfomance more than a risk-neutral equilibrium does.

Furthermore, if we compute the optimal flows for that same demand interval we have the following results:

- For $0 < d \leq \frac{1}{2\alpha}$, $x^{OPT} = d$ and $y^{OPT} = 0 \Rightarrow \mathrm{C(OPT)} = \alpha d^2$;
- For $\frac{1}{2\alpha} < d <\leq \frac{1}{\alpha}$,$x^{OPT} = \frac{1}{2\alpha}, y^{OPT} = d - \frac{1}{2\alpha} \Rightarrow \mathrm{C(OPT)} = d - \frac{1}{4\alpha}$.

Hence, we can observe that for $0 < d \leq \frac{1}{2\alpha}$, $\mathrm{C(RAWE)} = \mathrm{C(RNWE)} = \mathrm{C(OPT)} \Rightarrow \mathrm{PoA} = \mathrm{PRA} = 1$, whereas for the segment $\frac{1}{2\alpha} < d \leq \frac{1}{\alpha}$, we have that the PoA, using risk-neutral costs, is $\mathrm{PoA} = \frac{4(\alpha d)^2}{4\alpha d-1}$, which is increasing in d and it approaches 1 when $d \to \frac{1}{2\alpha}$, whereas it approaches $\frac{4}{3}$ when $d \to \frac{1}{\alpha}$.

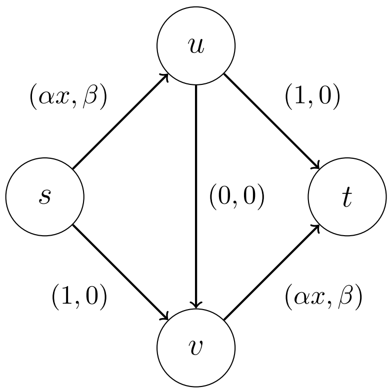
4.2 Braess networks

In the following two examples, we analysed Braess networks, modifying the latency and variance on the central edge $u-v$ (see Figure 4.11). Before moving on, however, we need to clarify some new notation concerning the paths in Figures 4.10 and 4.11: we refer to paths $s-u-t$ and $s-v-t$ as P_1 and P_2, respectively, and to the flows going through them as f_1 and f_2; the zig-zag path $s-u-v-t$ is denoted by P_3 and its flow is f_3.

Example 10

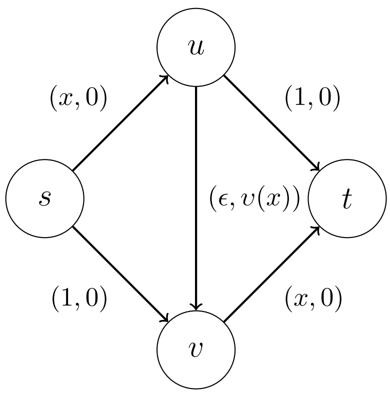
This is a classic Braess graph with modified edge functions $(\alpha x, \beta)$ for edges $s-u$ and $v-t$. Again, as in Example 1, if $\beta = \alpha x$, then the optimal flow is reached; however, as we've already pointed out, this would imply a var-to-mean ratio $\kappa = 1$, which is not a realistic scenario.

Figure 4.10: Example 10



Instead, we can consider the classic Braess network but this time we introduce variance, $v(x)$, just on the new edge $u-v$, who has now a mean latency of $\epsilon \approx 0$. We consider the instance where $d = 1$.

Figure 4.11: Braess network with variance on edge $u-v$



In this case, by marginal cost pricing, we need to have the following total cost on the zig-zag path in order to have the same costs we would have on edges $s-u-v-t$ in the optimal instance, when we take into account the marginal cost we impose on all the other agents as well:

$$x + \gamma v(x) + \epsilon + x = 4x + \epsilon$$

which implies that

$$v(x) = \frac{2x}{\gamma},$$

which when $\gamma = 1 \Rightarrow v(x) = 2x$. In more economic terms, this may be interpreted as a realistic scenario where the edge $u - v$ represents not a new road, but a highway exit which has been newly built and which allows drivers to take a faster route to reach their final destination. In this context, it is plausible that the latency of this new edge $u - v$ has a fixed cost very small that does not depend on the flow x (thus, $\epsilon \approx 0$), since you just need to get your ticket and pass the toll booth, but it can have a very high variance depending on the flow x going through it (thus, $v(x) = \frac{2x}{\gamma}$), since long and unexpected queues may form in Christmas or holiday season for example. Note, however, that in this way we do not modify the edges with latency functions and variances $(x, 0)$, implying that what we are going to obtain as our risk-averse flows won't be exactly the optimal ones.

Now, let's compute the RNWE, RAWE and optimal flows along with their total costs.

To compute the risk-neutral Wardrop equilibrium, assuming all three paths are going to be used, we solve the following system of equations:

$$\begin{cases} C_1(f_1, f_2, f_3) = f_1 + f_3 + 1 = \lambda \\ C_2(f_1, f_2, f_3) = 1 + f_2 + f_3 = \lambda \\ C_3(f_1, f_2, f_3) = f_1 + 2f_3 + \epsilon + f_2 = \lambda \\ f_1 + f_2 + f_3 = 1 \\ f_1, f_2, f_3 \geq 0. \end{cases}$$

The resulting flow vector is $(\boldsymbol{f}) = (f_1 = \epsilon, f_2 = \epsilon, f_3 = 1 - 2\epsilon)$.
To compute the total costs under RNWE, we need to plug-in our results in:

$$\text{C(RNWE)} = (f_1 + f_3)^2 + f_1 + f_2 + (f_2 + f_3)^2 + \epsilon f_3.$$

Obtaining, for $\epsilon \approx 0$, C(RNWE) $= 2$.

Now, we move to the computation of the risk-averse Wardrop equilibrium, assuming again full support and a variance $\epsilon(x) = \frac{2x}{\gamma}$ on the edge $u - v$:

$$\begin{cases} C_1(f_1, f_2, f_3) = f_1 + f_3 + 1 = \lambda \\ C_2(f_1, f_2, f_3) = 1 + f_2 + f_3 = \lambda \\ C_3(f_1, f_2, f_3) = f_1 + 2f_3 + \epsilon + \gamma\frac{2f_3}{\gamma} + f_2 = \lambda \\ f_1 + f_2 + f_3 = 1 \\ f_1, f_2, f_3 \geq 0. \end{cases}$$

For $\epsilon \approx 0$, the solutions are $(\boldsymbol{f}) = (f_1 = \frac{2}{5}, f_2 = \frac{2}{5}, f_3 = \frac{1}{5})$.
Total costs are:

$$\text{C(RAWE)} = (f_1 + f_3)^2 + f_1 + \epsilon f_3 + (f_2 + f_3)^2 + f_2,$$

which, for $\epsilon \approx 0$ and for our flow $\boldsymbol{f}$, yields $\text{C(RAWE)} = \frac{38}{25} \approx 1.52$.

Furthermore, we know that the optimal flows solve the following system:

$$\begin{cases} C_1(f_1, f_2, f_3) = 2(f_1 + f_3) + 1 = \lambda \\ C_2(f_1, f_2, f_3) = 1 + 2(f_2 + f_3) = \lambda \\ C_3(f_1, f_2, f_3) = 2(f_1 + f_3) + \epsilon + 2(f_2 + f_3) = \lambda \\ f_1 + f_2 + f_3 = 1 \\ f_1, f_2, f_3 \geq 0. \end{cases}$$

Its resulting flows, for $\epsilon \approx 0$, are $(\boldsymbol{f}) = (f_1 = \frac{1}{2}, f_2 = \frac{1}{2}, f_3 = 0)$.
The final total costs are:

$$\text{C(OPT)} = f_1 + f_2 + f_1^2 + f_2^2 = \frac{3}{2}.$$

Hence, we have the following interesting result: even if the optimal flows never use the zig-zag path, the total flow x on the edges $s - u$ and $v - t$ is $\frac{1}{2}$ for the optimal instance and $x^{RAWE} = f_1 + f_3 = \frac{3}{5}$, for the risk-averse one, meaning that the risk-averse flow congests those edges just 0.1 more than the optimal flows. On the other hand, the risk-neutral flows only use the zig-zag path, implying that $x^{RNWE} = 1$, and those edges are fully congested. Hence, by introducing variance in this model, we avoid selfish drivers congesting the zig-zag path and induce a flow distribution more favorable for the optimal outcome. Indeed, as we have seen, $\text{C(RAWE)} \approx 1.52 \approx \text{C(OPT)} = 1.5$. In this case, $PRA < 1$, namely:

$$\text{PRA} = \frac{38}{25 \cdot 2} = \frac{19}{25} \approx 0.76.$$

Example 11

Let's consider again Figure 4.11 but under a generic demand $d \in \mathbb{R}^+$.

In order to compute the risk-neutral equilibrium flows for a generic demand d, we had to analyse all the possible combinations of supports. Namely, we ended up with three demand intervals with their respective equilibrium flows.

For $0 < d \leq 1$, the only path used in equilibrium is the zig-zag path P_3 and $(\boldsymbol{f}) = (f_1 =$

$f_2 = 0, f_3 = d)$. We obtained this result by solving the system:

$$\begin{cases} C_1(f_1, f_2, f_3) = f_3 + 1 > \lambda \\ C_2(f_1, f_2, f_3) = f_3 + 1 > \lambda \\ C_3(f_1, f_2, f_3) = 2f_3 + \epsilon = \lambda \\ f_1 + f_2 + f_3 = d \\ f_1, f_2 = 0, f_3 > 0. \end{cases}$$

In this system and the ones that follow, we computed equilibrium flows and costs for $\epsilon \approx 0$, thereby ignoring it in our computations.
Hence, total costs in this case are:

$$\mathrm{C(RNWE)} = 2f_3^2 = 2d^2.$$

For $1 < d < 2$, all three paths are used, with flow vector $(\boldsymbol{f}) = (f_1 = d-1, f_2 = d-1, f_3 = 2-d)$. The system we solved is:

$$\begin{cases} C_1(f_1, f_2, f_3) = f_1 + f_3 + 1 = \lambda \\ C_2(f_1, f_2, f_3) = f_2 + f_3 + 1 = \lambda \\ C_3(f_1, f_2, f_3) = f_1 + f_3 + f_2 + f_3 + \epsilon = \lambda \\ f_1 + f_2 + f_3 = d \\ f_1, f_2, f_3 > 0. \end{cases}$$

Total costs are:

$$\mathrm{C(RNWE)} = (f_1 + f_3)^2 + f_1 + f_2 + (f_2 + f_3)^2 = 2d.$$

Finally, the last demand interval is $d \geq 2$, for which only paths P_1 and P_2 are used, and hence $(\boldsymbol{f}) = (f_1 = f_2 = \frac{d}{2}, f_3 = 0)$. The system we solved is:

$$\begin{cases} C_1(f_1, f_2, f_3) = f_1 + 1 = \lambda \\ C_2(f_1, f_2, f_3) = f_2 1 = \lambda \\ C_3(f_1, f_2, f_3) = f_1 + f_2 + \epsilon > \lambda \\ f_1 + f_2 + f_3 = d \\ f_1, f_2 > 0, f_3 = 0. \end{cases}$$

Total costs are:

$$\mathrm{C(RNWE)} = \frac{d^2}{2} + d.$$

Now we move on to the risk-averse equilibrium.
Even in this case, we had to divide the total demand d into three intervals.

For $0 < d \leq \frac{1}{3}$, only the zig-zag path P_3 is used, with flow vector $(\boldsymbol{f}) = (f_1 = f_2 = 0, f_3 = d)$. The system solved is:

$$\begin{cases} C_1(f_1, f_2, f_3) = f_3 + 1 > \lambda \\ C_2(f_1, f_2, f_3) = f_3 1 > \lambda \\ C_3(f_1, f_2, f_3) = 4f_3 + \epsilon = \lambda \\ f_1 + f_2 + f_3 = d \\ f_1, f_2 = 0, f_3 > 0. \end{cases}$$

Total costs under this equilibrium are:

$$\mathrm{C(RAWE)} = 2f_3^2 = 2d^2.$$

For $\frac{1}{3} < d < 2$, all three paths are used. The flow vector is $(\boldsymbol{f}) = (f_1 = \frac{3}{5}d - \frac{1}{5}, f_2 = \frac{3}{5}d - \frac{1}{5}, f_3 = \frac{2}{5} - \frac{d}{5})$ and is obtained by solving:

$$\begin{cases} C_1(f_1, f_2, f_3) = f_1 + f_3 + 1 = \lambda \\ C_2(f_1, f_2, f_3) = f_2 + f_3 + 1 = \lambda \\ C_3(f_1, f_2, f_3) = f_1 + 4f_3 + \epsilon + f_2 = \lambda \\ f_1 + f_2 + f_3 = d \\ f_1, f_2, f_3 > 0. \end{cases}$$

Total costs are:

$$\mathrm{C(RAWE)} = (f_1 + f_3)^2 + f_1 + (f_2 + f_3)^2 + f_2 = \frac{8}{25}d^2 + \frac{38}{25}d - \frac{8}{25}.$$

Finally, the last interval is for $d \geq 2$, for which only P_1, P_2 are used, with flow vector $(\boldsymbol{f}) = (f_1 = f_2 = \frac{d}{2}, f_3 = 0)$. The system solved is:

$$\begin{cases} C_1(f_1, f_2, f_3) = f_1 + 1 = \lambda \\ C_2(f_1, f_2, f_3) = f_2 + 1 = \lambda \\ C_3(f_1, f_2, f_3) = f_1 + + \epsilon + f_2 > \lambda \\ f_1 + f_2 + f_3 = d \\ f_1, f_2 > 0, f_3 = 0. \end{cases}$$

Total costs are:

$$\mathrm{C(RAWE)} = \frac{d^2}{2} + d.$$

We now have all the elements we need to carry out our PRA analysis.

For $0 < d \leq \frac{1}{3}$, we have that

$$\text{C(RAWE)} = \text{C(RNWE)} = 2d^2 \Rightarrow \text{PRA} = 1.$$

For $\frac{1}{3} < d \leq 1$, we have:

$$\text{C(RAWE)} = \frac{8}{25}d^2 + \frac{38}{25}d - \frac{8}{25}, \text{C(RNWE)} = 2d^2 \Rightarrow \text{PRA} = \frac{\frac{8}{25}d^2 + \frac{38}{25}d - \frac{8}{25}}{2d^2}$$

which is increasing until $d \approx 0.42$, where it is reaches a maximum PRA ≈ 1.06, and then it starts decreasing till it reaches its minimum at $d = 1 \Rightarrow \text{PRA} = 0.76$, which is consistent with our previous example for $d = 1$.

For $1 < d < 2$, we have:

$$\text{C(RAWE)} = \frac{8}{25}d^2 + \frac{38}{25}d - \frac{8}{25}, \text{C(RNWE)} = 2d \Rightarrow \text{PRA} = \frac{\frac{8}{25}d^2 + \frac{38}{25}d - \frac{8}{25}}{2d}$$

which is increasing but always less than 1, indeed for $d \rightarrow 1 \Rightarrow \text{PRA} \rightarrow 0.76$, whereas for $d \rightarrow 2 \Rightarrow \text{PRA} \rightarrow 1$.

Finally, for $d \geq 2$, we have:

$$\text{C(RAWE)} = \text{C(RNWE)} = \frac{d^2}{2} + d \Rightarrow \text{PRA} = 1.$$

To conclude with, in this case introducing risk-aversion in the system does not degrade its performance. Actually, for certain demand intervals, it improves the system's performance and the flow it induces are better for minimizing total social costs rather than the ones computed under risk-neutral equilibrium.

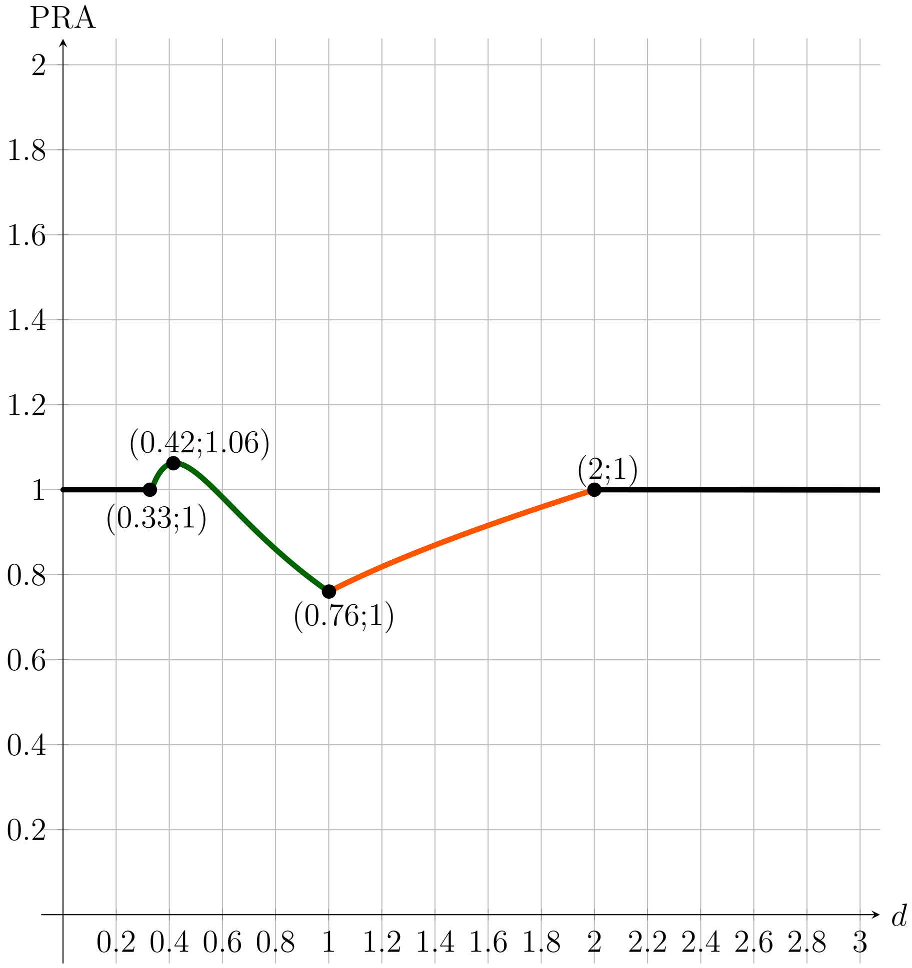


Figure 4.12: Example 11; PRA function for $d \in (0, +\infty)$.

Conclusions

Throughout this dissertation, we have seen how people may model their behaviour according to their level of risk-aversion. When drivers are risk-neutral, meaning they do not take into account the variance of their path, then they only care about minimizing the expected latency of their path; on the other hand, if they do are risk-averse and concerned about the variance along their path, their aim is to minimize the expected latency of their path *plus* the variance along it.

The first case has been extensively treated in Chapter 1, where we have characterized the equilibrium of the network with risk-neutral agents (Risk Neutral Wardrop Equilibrium or RNWE) and we have seen that all paths used in a Wardrop equilibrium have the same and minimum-possible cost. We have also introduced a measure of the inefficiency of such a network: the Price of Anarchy (PoA), which is the ratio of the worst social costs value under a RNWE to the optimal social costs value. We have shown that the PoA has an upper bound (the Pigou bound) which is independent of the number of commodities and of the topology of the network.

The risk-aversion case has been treated in Chapter 2, where we have seen that drivers end up choosing the paths that minimize their *mean-variance* objective function, and we called this equilibrium Risk Averse Wardrop Equilibrium (RAWE). To measure the extent to which risk-aversion may contribute to the degradation of the system's performance, we introduced the Price of Risk Aversion (PRA), which is the ratio of the worst social costs value under a RAWE to the worst social costs value under a RNWE. We have shown the proof of Lianeas, Evdokia Nikolova, and N. E. Stier-Moses (2019) showing that such PRA has an upper bound of $1 + \gamma\kappa\eta$, where γ is the risk-averse coefficient of drivers, κ is the upper bound of the var-to-mean ratio of the network, and η is a parameter depending on the network's topology and which defines the number of disjoint forward subpaths in the alternating path π presented in the chapter. We called this bound *structural*, since it does not depend on the class of latency functions chosen (on the contrary, the Pigou bound on the PoA of Chapter 1 does), but just on the network's structure and topology. Furthermore, we have seen that for graphs with n vertices equal to a power of 2, this upper bound has a matching lower bound, *i.e.* it is tight (see Theorem 5).

Finally, in the last two chapters we have focused on the mechanism design side of the efficiency of non-atomic selfish-routing games. In particular, we have presented the traditional method of marginal-cost pricing, where by imposing a toll along the edges equal to the difference between marginal social costs and marginal individual costs at the system's optimum flow (*i.e.* $\tau_e = f_e \cdot c_e'(f_e)$), the optimal flow is induced in the network. We also analysed the analogous stochastic case after the studies of Xu and Sumalee (2015), called Stochastic Network Marginal Cost Pricing (SN-MCP), where a second term accounting for the stochasticity of the network has to be added to determine the toll necessary to induce the optimal flow on the network (*i.e.* $\sum_{j=0}^{m} b_{je} \frac{\partial \mathrm{Cov}[V_e^j, V_e]}{\partial v_e}$, which accounts for the uncertainty of travel demand and differs as such from the deterministic and classical toll). We have also presented the study of Colini-Baldeschi, Klimm, and Scarsini (2018) on Demand Independent Optimal Tolls (DIOTs), which are tolls ensuring the optimal flow would be reached by selfish drivers, even when the exact demand and demand function on the network are not known. Also, we have presented another way to reach the optimal flow without imposing tolls along the edges of the network, which is by increasing the capacity of the network. Namely, by Corollary 9.1 we have seen that for $M/M/1$ latency functions, it is enough to double the capacity on every edge in the network to obtain a result in terms of efficiency at least as good as the optimal social costs which would be achieved through the job of a social planner imposing tolls on the network's edges.

On the other hand, in the last chapter we tried to come up with some examples and scenarios where it is the risk-aversion of drivers itself that may bring an improvement to the system's performance with respect to the risk-neutral equilibrium case. In particular, we have seen in Example 6 that, for $0 < d \leq 1$ and for a U-shaped variance function $v(x) = -x^2 + x$, the PRA is always less than or equal to 1 and total costs under RAWE are very close to total costs under the optimal flow, which implies it is possible to close in on the social optimum without imposing any edge toll. For the Braess networks, we have seen in Examples 10 and 11 that, by introducing a variance $v(x) = \frac{2x}{\gamma}$ on the central edge $u - v$ and with the exception of a tiny demand interval (*i.e.* $\frac{1}{3} < d \leq 0.42$), the PRA is less than or equal to 1, reaching a minimum of PRA$= 0.76$ at $d = 1$.

Bibliography

Beckmann, J. Martin et al. (1956). *Studies in the Economics of Transporation.* Yale University Press.

Colini-Baldeschi, Riccardo, Max Klimm, and Marco Scarsini (2018). "Demand-Independent Optimal Tolls". In: *45th International Colloquium on Automata, Languages and Programming.*

Fotakis, Dimitris, Dimitris Kalimeris, and Thanasis Lianeas (2020). "Improving Selfish Routing for Risk-Averse Players". In: *Theory of Computing Systems.*

Hall, Michael A. (1978). "Properties of the equilibrium state in transportation networks". In: *Transportation Science.*

Heap, Shaun P. H. and Yanis Varoufakis (2004). *Game Theory: A Critical Introduction.* Routledge.

Koutsoupias, E. and C. H. Papadimitriou (1999). "Worst-case equilibria". In: *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science.*

Li, M.Z.F. (2002). "The role of speed-flow relationship in congestion pricing implementation with an application to Singapore". In: *Transportation Research Part B: Methodological.*

Lianeas, Thanasis, Evdokia Nikolova, and Nicolas E. Stier-Moses (2019). "Risk-Averse Selfish Routing". In: *Mathematics of Operations Research.*

Nezamuddin, J. et al. (2009). *Statistical Patterns of Traffic Data and Sample Size Estimation.* URL: `http://transops.s3.amazonaws.com/uploaded_files/APPENDIX%20D%20-%20Analysis%20of%20Sample%20Size.pdf`.

Nikolova, E. and N. Stier-Moses (2014a). "A Mean-Risk Model for the Traffic Assignment Problem with Stochastic Travel Times". In: *Operation Research.*

— (2014b). *The Burden of Risk Aversion in Mean-Risk Selfish Routing.* arXiv: `1411.0059 [cs.GT]`.

Nisan, Noam, Tim Roughgarden, Tardos Éva, et al. (2007). *Algorithmic Game Theory.* Cambridge University Press.

Ordóñez, Fernando and N. Stier-Moses (2010). "Wardrop Equilibria with Risk-Averse Users". In: *Transportation Science.*

Osborne, Martin J. (2003). *An Introduction to Game Theory.* Oxford University Press.

Papadimitriou, C. H. (2001). “Algorithms, games, and the Internet”. In: *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing.*

Pigou, A. C. (1920). *The economics of welfare.*

Roughgarden, Tim (2003). “The Price of Anarchy is Independent of the Network Topology”. In: *Journal of Computer and System Sciences.*

— (2006). *Selfish Routing and the Price of Anarchy.* MIT Press.

Roughgarden, Tim and Éva Tardos (2002). “How Bad is Selfish Routing?” In: *Journal of the ACM.*

Wardrop, John G. (1952). “Some theoretical aspects of road traffic research”. In: *Proceedings of the Institute of Civil Engineers.*

Xu, Wei and Agachai Sumalee (2015). “Evolutionary Marginal Cost Pricing Schmeme Implementation Based on Stochastic Traffic Flow Information”. In: *Mathematical Problems in Engineering, Hindawi Publishing Corporation.*

Yang, H., Q. Meng, and D.H. Lee (2004). “Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions”. In: *Transportation Research Part B: Methodological.*

Yekkehkhany, Ali and Rakesh Nagi (2020). *Risk-Averse Equilibrium for Autonomous Vehicles in Stochastic Congestion Games.* arXiv: 2007.09771 [cs.GT].