# When Game Theory Meets Transportation (Communication) Network

Qian Xie

***Abstract*—Due to the selfish behaviors and congestion effects existing in transportation and communication network. Transportation researchers, game theorists and computer scientists are devoted to formulating such problems, analyzing the equilibrium, providing solutions and designing strategies. While there is substantial amount of work making contributions to combine game theory with transportation setting, few literature reviews gives summary on them. In this literature review, we introduce the history of capturing congestion effect and strategic interaction in mathematical models and present details of three influential papers.**

**Keywords**: Congestion game, potential game, Wardrop equilibrium, selfish routing, Price of Anarchy.

## I. Introduction

### A. Overview

Game theory provides us with a mathematical framework for modelling strategic interactions among rational decision-makers when individual actions jointly determine the outcome. It has wide applications in economics, social science, biology, computer science and system science. The illustrative examples in textbooks and motivating examples in literature are largely from economics and society, such as Prisoner's Dilemma, Cornout's model of duopoly, market of lemons.

In recent years, game theory receives more and more attention by scientists in engineering fields. An exciting example is the emergence of algorithmic game theory. This area lies in the intersection of game theory and computer science with the goal of analyzing and designing algorithms in strategic environments. In 2012, three seminal papers laying foundation of growth in Algorithmic Game Theory" were recognized by Gödel Prize committee. Two of them introduced and developed the concept of "*Price of Anarchy*" [1,2], measuring the degradation of system efficiency due to the selfish behavior of its agents. Details will be presented in the following subsections of this literature review. Moreover, Roughgarden and Tardos (2012)[2] applied the idea of "Price of Anarchy" to the analysis of selfish routing in congested traffic network. This becomes the most classic example of the application in transportation.

Back to 1952, when John Nash just published his far-reaching work [3,4], Wardrop independently formed the idea similar to nooncoperative game and Nash equilibrium in his road traffic research,[5] where travelers correspond to players whose utilities are based on journey times. By stating two principles, Wardrop described different notions of equilibrium: *user equilibrium* (selfish equilibrium, decentralized decision-making) and *system optimum* (social equilibrium, joint decision-making). The difference between them can be observed under congestion effect. We can say this idea also coincides with Price of Anarchy by looking at the fall in efficiency from system optimum to user equilibrium.

It was Beckmann, McGurie and Winsten (1956)[6] who formulated Wardrop equilibrium as mathematical model. In their model, an average cost function is defined to represent link performance (journey times on roads) in terms of flows on it. Such functions can capture the congestion effect by assuming monotonicity (increasing with flows). Then Dafermos and Sparrow (1969)[7] further developed this topic to trip (route) assignment problem. At that time, the formulations are mainly based on operations research (optimization, mathematical programming).

On the other track, game theorists started noticing the congestion effects and network equilibrium. Rosenthal (1973) [8,9] considered a class of noncooperative game including the game of travelers on road networks. This class of game was later called "*congestion game*" since the relationship between pure strategy and cost (payoff) possesses the characteristic of congestion.

Since then, more and more researchers connect game theory with transportation, especially transportation network. Realizing that congestion game introduced by Rosenthal only assumes atomic players and numeric payoffs, Roughgarden and Tardos (2002)[2] developed versions of nonatomic players and functional payoffs. This paper breaks the barrier between game theory and transportation. There is a large growth of works lying in the intersection of game theory and transportation.

Roughgarden also mentioned the similarities between transportation network and communication network. He stated that their models and results can also be applied to communication network. However, Acemoglu, Johari and Ozdaglar (2007) pointed out the limitations

This literature review is for ECE-GY 6263 Game Theory project. The instructor is professor Quanyan Zhu.

Q. Xie is with the Department of Civil and Urban Engineering, New York University Tandon School of Engineering, Brooklyn, NY, USA. (email: qianxie@nyu.edu)

of routing models in communication network due to the existence of subnetworks. They created a term "*partially optimal routing*" describing the mix of selfish routing and optimal routing. With an adaptation of the routing model, they studied how partially optimal routing affects the network efficiency. Actually, before them, Roughgarden in his paper[10] also addressed such mix, but from another perspective – algorithmic mechanism design (referring to another 2012 Gödel Prize work[11]). By applying Stackelberg game model, he investigated the strategy of the system operator who centrally controls a fraction of traffic and aspires to optimize the overall traffic when selfish behaviors still exist.

Only a few literature reviews [12,13,14] about the interaction between game theory and transportation, not to mention the review of games related transportation network and communication network. The goal of this review is to help readers go over the development of the game models capturing the congestion effects, understand the appealing results as well as attract attentions of learners and researchers in both fields.

### B. Motivating examples

*1) Santa Fe (El Farol) bar problem:* Based on a bar in Santa Fe, New Mexico, Arthur (1994)[15] created a problem as follows:

- Each evening, people in the neighborhood simultaneously and independently decides whether go to the bar or stay at home.
- Since the bar has limited space, the visitors do not enjoy a good time at bar if it is too crowded.
- If less than 60% of the population go to the bar, they have a better experience than staying at home.
- If more than 60% of the population go to the bar, they would rather to stay at home.

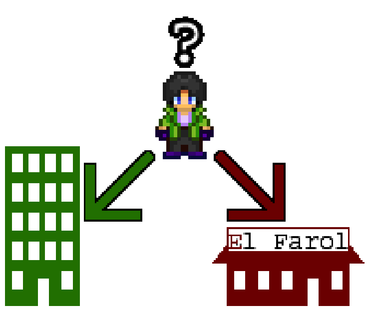



Fig. 1: Santa Fe (El Farol) bar problem

In,[16] this problem is generalized to the problem of sharing resources of limited capacity. The interaction among agents (players) is through the joint use of the resources.

In the field of transportation, congestion results from the interaction among drivers sharing roads is widely studied. Such problem can also be viewed as resource sharing problem and share similarity with Sante Fe (El Farol) bar problem.

*2) Pigou's example:* Consider the following setting described by Pigou (1920) [17,18]:

- A suburb and a nearby train station (denoted by $s$ and $t$, respectively) are connected by two non-interfering highways.
- A fixed number of drivers wish to commute from $s$ to $t$ at roughly the same time.
- One is short but narrow with the delay depending on congestion; the other is wide enough to accommodate all of the traffic without any crowding.
- The **latency functions** $l(\cdot)$ describe the latency or travel time experienced by drivers on a road as a function of the fraction of the overall traffic using that road.
- All drivers aim to minimize the time taken to drive from $s$ to $t$.

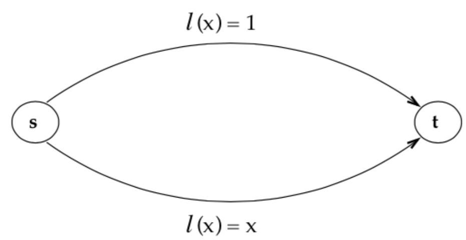



Fig. 2: Pigou's example

First let's look at drivers' selfish behaviors, any driver traveling on the top road (who experience one hour of latency) would envy those drivers on the bottom road (experiencing less than one hour of latency when not all traffic is on this route) and prefer to go to the other route. In the end, all traffic shift to the bottom road and the average travel time is **one hour**.

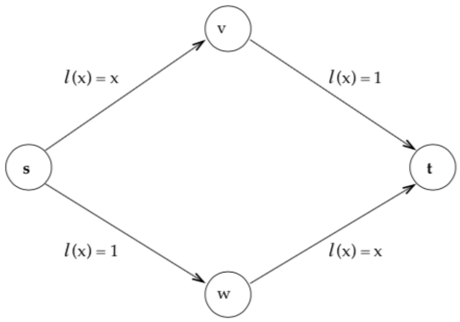
Now consider the case that there is a system operator who has the power of centralized control. By assigning half of the traffic to each of the two routes, half of the drivers can enjoy lighter traffic conditions and spend only 30 minutes on road. The average travel time can drop **from 60 to 45 minutes**.

This example implies that selfish behaviors do not guarantee socially optimal outcome. In other words, there are conflicts between individual interests and collective interests. Just as Adam Smith's famous quote, with system operator's "*invisible hand*", there can be a gain of *social welfare*. It is natural to come up with the question:

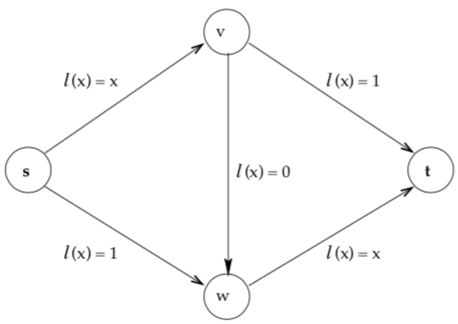
*what is the gap between selfish outcome and optimal outcome?*

*3) Baress's paradox:* Next, we show a less intuitive example discovered by Baress (1968) [19,20]:

- Again begin with a suburb $s$, a train station $t$, and a fixed number of drivers who wish to commute from $s$ to $t$.
- Initially, two non-interfering routes from $s$ to $t$, each comprising one long wide road and one short narrow road.
- In an effort to alleviate these unacceptable delays, a very short and very wide highway with constant latency function $l(x) = 0$ is built to join the midpoints of the two existing routes.




(a) Initial Network




(b) Augmented Network

Fig. 3: Braess's paradox

First by symmetry, it is obvious that the traffic would split equally between two routes, incurring **1.5 hours = 90 minutes** of latency.

Now let's discuss how drivers would react to the addition of the "super" road. Each driver would figure out that following the new route $s \to v \to w \to t$ can save 30 minutes of travel time. Without coordination, all of them would simultaneously deviate from their previous route. However, in such scenario, the latency on the road $s \to v$ and $w \to t$ raises to one hour, leading to a longer total travel time of **two hours**. Despite this, neither of the two alternative routes $s \to v \to t$ and $s \to w \to t$ (also takes two hours) is superior and thus no driver has an incentive to reroute.

This example is counter-intuitive. Well-intentioned though the network designer may be, the attempts to reduce delay are failed and even produce the adverse effect. However, it is not easy for the network designer to predict the outcome of the selfish behaviors.

The Baress's paradox gives us a lesson that *more is less*. The network structure can play an important role in the traffic condition. While in the past, the network design is largely based on experience and surveys, it can be inspiring if we view it as an algorithmic problem.

### C. Outline

The following sections are organized as follows. A preliminary section is included to introduce some game theory concepts that will be frequently quoted in this review. Then we move on to three main parts each introducing a paradigm: congestion game, routing game and partially optimal routing game. In the last section, I will summarize the contribution and limitation of several seminal works in the last section and provide some future directions.

### D. This literature review

The subsection introduces how this literature review goes.

*1) Highlights only:* For the most part, I will present three papers on potential game, selfish routing and partially optimal routing [21,22,23].Other related work are also discussed at some length. Considering the topic of this literature review, I heavily use the language of transportation and examples from this field. Moreover, I put more weight on the intersection of game theory and transportation (communication) network and as a trade-off, stuffs like approximation, hardness, pricing and profit incentives are omitted.

*2) Self-consistent:* This review assumes relatively few prerequisites. Considering different definitions and models used in different works can create a barrier to readers, I try to unify the notations and introduce the model inheritance.

*3) Reader-friendly:* To help understanding, a variety of motivating and illustrative examples are provided, most of them have associated figures. Besides, I try to simplify the notations and give high-level idea of the proofs (some of them are also figure-based). For those elementary results (cornerstones), I devote some length to the details of their proofs.

## II. PRELIMINARY

### A. Strategic form games

A *strategic (form) game* is a model for a game in which all players (participants) act simultaneously without knowledge of other players' actions.

*Definition 1:* A **strategic game** is a tuple $(\mathcal{N}, ,S, u)$, where

- $\mathcal{N}$ is a finite set of **players**, $\mathcal{N} = \{1, 2, \cdots, n\}$.
- $\mathcal{S}_i$ is the set of available actions for player $i$
  - $s_i \in S_i$ is an **action** for player $i$;
  - $s_{-i} = (s_j)_{j \neq i}$ is a vector of actions for all players except $i$;
  - $(s_i, s_{-i}) \in \mathcal{S}$ is an **action profile**, or **outcome**;
  - $\mathcal{S} = \prod_{i=1}^{n} \mathcal{S}_i = \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_n$ is a set of all actions;
  - $\mathcal{S}_{-i} = \prod_{j \neq i} \mathcal{S}_i$ is a set of all actions for all players except $i$.
- $u_i : \mathcal{S} \to \mathbb{R}$ is the **payoff (utility) function** of player $i$.

A player's *strategy* determines the action that the player will take at any stage of the game. A *strategy profile* is a set of strategies for all players which fully specifies all actions in a game. The definition is similar to 1 by replacing *action* with *strategy*.

A *pure strategy* determines the move a player will make for any situation they could face.

A *mixed strategy* is an assignment of a probability to each pure strategy that directs a player to randomly choose a pure strategy. Since probabilities are continuous, there are infinitely many mixed strategies available to a player.

*Remark 1:* The term *action* and *pure-strategy* are sometimes used interchangeably.

The *best response* is the strategy (or strategies) which produces the most favorable outcome for a player, given other players' strategies.

*Definition 2:* The **best response** is a mapping $B_i(s_{-i}) : \mathcal{S}_{-i} \to \mathcal{S}_i$ such that

$$B_i(s_{-i}) \in \arg \max_{s_i \in \mathcal{S}_i} u_i(s_i, s_{-i}).$$

### B. Nash equilibrium

Nash (1950-1951) [3,4] proposed the solution concept *Nash equilibrium* to analyze the outcome of the strategic interaction of several decision makers.

*Definition 3:* A **(pure strategy) Nash Equilibrium** of a strategic game is a strategy profile $\mathbf{s} \in \mathcal{S}$ such that for all $i \in \mathcal{N}$,

$$u_i(\mathbf{s}_i, \mathbf{s}_{-i}) \geq u_i(s_i, \mathbf{s}_{-i})$$

for all $s_i \in \mathcal{S}_i$.

*Remark 2:* No player has the incentive to deviate from Nash equilibrium given the strategies of the other players.

*Proposition 1:* An action profile $\mathbf{s}$ is a Nash equilibrium if and only if

$$\mathbf{s} \in B_i(\mathbf{s}_{-i})$$

for all $i \in \mathcal{N}$.

### C. Fictitious play

*Fictitious play* is a learning process introduced by Brown (1951).[24] The idea is that players form beliefs about opponents (other players) and optimize their action with respect to these beliefs in a myopic manner, that is, maximize current payoff without considering the future payoff.

*Definition 4:* In a **fictitious play**, at time $t$, player $i$ forms the empirical frequency distribution of his/her opponent $j$'s past play according to the number of times player $j$ is observed to play action $\tilde{s}_j$ before time $t$:

$$\mu_j^t = \frac{1}{t} \sum_{\tau=0}^{t-1} I(s_j^\tau = \tilde{s}_j),$$

where $I$ is the indicator function.

Player $i$ then chooses his/her action to maximize his/her payoff, that is,

$$s_i^t \in \arg \max_{s_i \in \mathcal{S}_i} u_i(s_i, \mu_{-i}^t),$$

where $\mu_{-i}^t = \prod_{j \neq i} \mu_j^t$ for all $t$.

*Remark 3:* Players only need to know their own utility function.

*Proposition 2:* Proposition: Suppose a fictitious play sequence $\{s_t\}$ converges to $\sigma$ in the time-average sense. Then $\sigma$ is a Nash equilibrium of the game.

## III. CONGESTION GAME

### A. Definition

Rosenthal (1973)[9] proposed a class of games that models the competition among players over a finite set of resources and the payoff (cost) depends on the number of players choosing the same source. Two examples (applications) are provided in his paper, one considers traffic flows and the other is about firm production. Such game was later named *congestion game*. Monderer and Shapley (1996)[21] pointed out the significance of congestion game in economics.

*Definition 5:* A **congestion game** is a tuple $(\mathcal{N}, \mathcal{R}, \mathcal{S}, c)$, where

- $\mathcal{N}$ is a set of **players**;
- $\mathcal{R}$ is a set of **resources**;
- $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_n$ is a set of **actions**, where $\mathcal{S}_i \subset \mathcal{R}$ is the set of resources that player $i$ can take;
- $c = (c_1, \cdots, c_r)$, where $c_j : \mathbb{N} \to \mathbb{R}$ is a **cost function** for resource $j \in R$.

If $k$ players are using resource $j$, then the **cost** to each player is $c_j(k)$.

Given an action profile $s = (s_i, s_{-i})$, the utility function is defined as

$$u_i(s) = -\sum_{j \in S_i} c_j(k_j)$$

where $k_j = \chi_j(s)$ is the number of players using resource $j$ under action $s$.

For transportation scenario (more specifically, traffic assignment problem), drivers seeking minimal cost path are players and the shared roads correspond to resources. For each player, his/her action set is the set of paths for him/her to take between the origin and the destination. Besides, the cost can be defined as travel time, congestion price, etc.

### B. Existence of pure-strategy Nash equilibrium

Rosenthal (1973)[9] proved the following theorem in his pioneering work.

*Theorem 1:* Every congestion game has a pure-strategy Nash equilibrium.

*Proof.* Each player $i$ in the congestion game strives to solve an optimization problem as follows:

$$\max \quad P(s) = \sum_{j \in \cup \mathcal{S}_i} \Big( \sum_{k=1}^{k_j} c_j(k) \Big)$$

in which the potential function $P : S \to \mathbb{R}$ is exactly an approximation of the integral over the cost functions (see Fig 4).

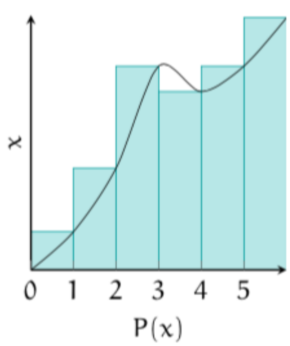



Fig. 4: Potential function

Let $\mathbf{s} \in S$ be the solution. Now suppose that this is not an equilibrium, which means for some $i \in \mathcal{N}$ and $t_i \in \mathcal{S}_i$,

$$u_i(t_i, \mathbf{s}_{-i}) > u_i(\mathbf{s}).$$

Then, the potential (objective value) under the new action

$$\begin{aligned}
P(t_i, \mathbf{s}_{-i}) = & \sum_{j \in \cup S_i} \Big( \sum_{k=1}^{\chi_j(t_i, \mathbf{s}_{-i})} c_j(k) \Big) \\
= & \sum_{j \in \cup S_i} \Big( \sum_{k=1}^{\chi_j(\mathbf{s})} c_j(k) \Big) - \sum_{j \in t_i \setminus s_i} c_j(\chi_j(s) + 1) \\
& + \sum_{j \in s_i \setminus t_i} c_j(\chi_j(s)) \\
= & P(\mathbf{s}) - u_i(t_i, \mathbf{s}_{-i}) + u_i(\mathbf{s}) \\
> & P(\mathbf{s})
\end{aligned}$$

is higher than the one under the optimal solution. A contradiction! ■

The key step in this proof is the construction of a potential function $P$. Monderer and Shapley (1996)[21] later formally defined the potential function and developed the idea to a wider class of game called potential game.

### C. Potential Game

*Definition 6:* A game $G = (\mathcal{N}, \mathcal{S}, u)$ is a **potential game** if there exists a function $\Phi : \mathcal{S} \to \mathbb{R}$ if for all $i$ and all $s_{-i} \to \mathcal{S}_{-i}$,

$$u_i(s_i, s_{-i}) - u_i(s_i', s_{-i}) = \Phi(s_i, s_{-i}) - \Phi(s_i', s_{-i}),$$

for all $s_i, s_i' \in \mathcal{S}_i$.

The function $\Phi$ is called a **potential function** for the game $G$.

Besides the concept of (exact) potential function (game) introduced here, Monderer and Shapley (1996)[21] also defined *ordinal potential* and *weighted potential*. The details are omitted in this review.

They also found the relationship between congestion game and potential game.

*Theorem 2:* Every congestion game is a potential game.

*Proof.* Verify that the following is a potential function for the congestion game

$$\Phi(s) = \sum_{j \in \cup S_i} \Big( \sum_{k=1}^{k_j} c_j(k) \Big).$$

■

*Remark 4:* The high level idea of this proof is: since utility functions are linear combinations of cost functions, most of the terms in the expansion can be canceled out when taking the difference between the potential values of two similar action profiles.

*Theorem 3:* Every finite potential game is isomorphic to a congestion game (see Figure 5).

The proof is provided in their Appendix B.[21]

Here is an example given by them to illustrate such equivalence.

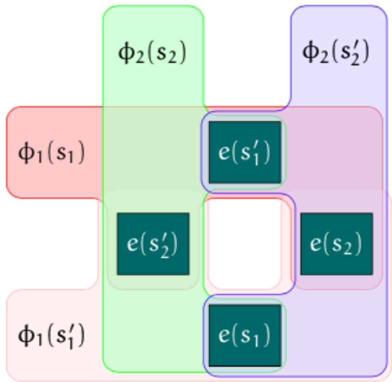



Fig. 5: Isomorphism between finite potential game and congestion game

$$\begin{tikzcd}
A \arrow[r, "c_1(1){,}c_1(2)"] \arrow[d, "c_3(1){,}c_3(2)"'] & B \arrow[d, "c_2(1){,}c_2(2)"] \\
D \arrow[r, "c_4(1){,}c_4(2)"'] & C
\end{tikzcd}$$

Fig. 6: A congestion model

*Example 1:* In a congestion model (see Fig 6), two drivers with OD pairs (A, C) and (B, D) have to share four roads AB, BC, AD, DC (numbered as 1, 2, 3, 4). The associated congestion game $CG$ is

$$\begin{pmatrix} (c_1(2)+c_2(1),c_1(2)+c_3(1)) & (c_2(2)+c_1(1),c_2(2)+c_4(1)) \\ (c_3(2)+c_4(1),c_3(2)+c_1(1)) & (c_4(2)+c_3(1),c_4(2)+c_2(1)) \end{pmatrix}$$

and the equivalent potential game $P$ is:

$$\begin{pmatrix} c_1(1)+c_1(2)+c_2(1)+c_3(1) & c_2(1)+c_2(2)+c_1(1)+c_4(1) \\ c_3(1)+c_3(2)+c_4(1)+c_1(1) & c_4(1)+c_4(2)+c_3(1)+c_2(1) \end{pmatrix}$$

### D. *Computation of equilibrium in congestion game*

Showing the existence of pure-strategy Nash equilibrium is not enough, Monderer and Shapley (1996) also studied the computation of the equilibrium in finite potential games, or equivalently, congestion games.

*Theorem 4:* A simple procedure (**MyopicBestResponse**) is guaranteed to find a pure-strategy Nash equilibrium of a congestion game.

*Proof.* Consider the following procedure:

- Start with an arbitrary action profile $s$
- While there exists a player $i$ for whom $s_i$ is not a best response to $s_{-i}$
  - $s'_i \leftarrow$ some best response by $i$ to $s_{-i}$
  - $s \leftarrow (s'_i, s_{-i})$
- Return $s$

*Theorem 5:* The **MyopicBestResponse** procedure is guaranteed to find a pure-strategy Nash equilibrium of a congestion game.

*Proof.* Following the idea of the proof for Theorem 1. At step of the while loop, $P(s)$ strictly increases. The reason follows from the best response argument $u_i(s'_i, s_{-i}) > u_i(s_i, s_{-i})$, and the definition of a potential function, $P(s'_i, s_{-i}) > P(s_i, s_{-i})$. Since there are only a finite number of action profiles, the algorithm must terminate and reaches a local minima. At this point, no player can achieve any improvement, and we reach a Nash equilibrium.

Noting that the myopic learning process is based on the one-sided better response dynamic, they also obtained a stronger learning result by considering the all players' belief updates:

*Theorem 6:* Every congestion game has the **Fictitious Play** property.

## IV. Routing Game

Inspired by the class of congestion game, Roughgarden (2002)[22] formally defined a routing model that allows fractional flows where in original congestion game the players are atomic.

### A. *Routing model*

Consider a directed network $G = (V, E)$ with vertex set $V$, edge set $E$, and $k$ origin-destination vertex pairs $(s_1, t_1), \cdots, (s_k, t_k)$. Parallel edges between vertices are allowed but there are no self-loops. The phrases *vertices* and *edges* are more commonly used in graph theory and graph algorithm, while in fields of operations research, transportation and computer networks, researchers favor *nodes* and *links* to refer the same thing.

For each player $i$, the set of paths between $s_i$ and $t_i$ is denoted as $\mathcal{P}_i$ ($P_i \neq \emptyset$). Let $\mathcal{P} = \cup_i \mathcal{P}_i$.

Researchers always use the term **commodities** for origin-destination pairs and paths. Specifically, commodity $i$ refers to an origin-destination pair $(s_i, t_i)$ and the set of paths $\mathcal{P}_i$.

*Remark 5:* Such networks have a unified name as *multicommodity* network [25,26].

*Definition 7:* A **flow** is a function $f : \mathcal{P} \to \mathbb{R}^+$. For a fixed flow $f$ and a link $e \in E$, $f_e = \sum_{P \in \mathcal{P}: e \in P} f_P$ is the total amount of flow on link $e$.

The flow of a particular commodity $i$ restricted to $\mathcal{P}_i$ can be written as $f^i$ and the flow of commodity $i$ on link $e$ is $f^i_e$.

*Definition 8:* A finite and positive **traffic rate** $r_i$ is associated with each origin-destination pair $(s_i, t_i)$, denoting the amount of flow from origin $s_i$ to destination $t_i$.

A flow $f$ is said to be **feasible** if for every $i$,

$$\sum_{P \in \mathcal{P}_i} f_P = r_i.$$

*Definition 9:* Each link $e \in E$ has a non-negative, continuous, differentiable and non-decreasing **latency function** $l_e(\cdot)$ capturing the congestion effects.

The **latency** of a path $P$ with respect to a flow $f$ is defined as the sum of the latency of the edges in the path:

$$l_P(f) = \sum_{e \in P} l_e(f_e).$$

*Definition 10:* A **routing instance** is a tuple $R = (G, r, l)$.

*Definition 11:* The **cost** $C(f)$ of a flow $f$ is the total latency incurred by $f$:

$$C(f) = \sum_{p \in \mathcal{P}} l_P(f) f_P = \sum_{e \in E} l_e(f_e) f_e.$$

With respect to an instance $(G, r, l)$, a feasible flow minimizing $C(f)$ is said to be **optimal** or **minimum-latency**.

*Example 2:* An example of the total latency cost is shown in Fig 7. In this routing instance, the directed network has six nodes and seven links, each link has a latency function with respect to flow on it. There are two commodities each with an origin-destination pair and two paths as options. Two units of flows (traffic rate of 2) traverse the blue path and two units of flows traverse the red path. The total latency cost is calculated as $2 \times 1 + 2 \times 3 = 8$.

*Proposition 3:* Optimal flow always exists because the space of all feasible flows is a compact set and the cost function is continuous.

## B. Flow at Wardrop equilibrium

In Roughgarden and Tardos's work [2,22], they study the loss of social welfare due to selfish, uncoordinated behaviors in networks. The "selfish" outcome coincides with the idea of Nash equilibrium.

Wardrop proposed two criterion in his influential paper,[5] which is later generally adopted by researchers as Wardrop's Pinciple [27,28]. A flow satisfying Wardrop's first principle is called Wardrop equilibrium.[29]

The first principle can be phrased in the language of routing model as: at Wardrop equilibrium, all flow travels on minimum-latency paths. In other words, all nonzero flow paths must have equal latency, say $L_i(f)$.

*Definition 12:* A feasible flow $f$ for instance $(G, r, l)$ is at **Wardrop equilibrium**, or **Nash flow**, if and only if for every $i \in \mathcal{N}$ and $P, P' \in \mathcal{P}_i$ with $f_P > 0$,

$$l_P(f) \leq l_{P'}(f).$$

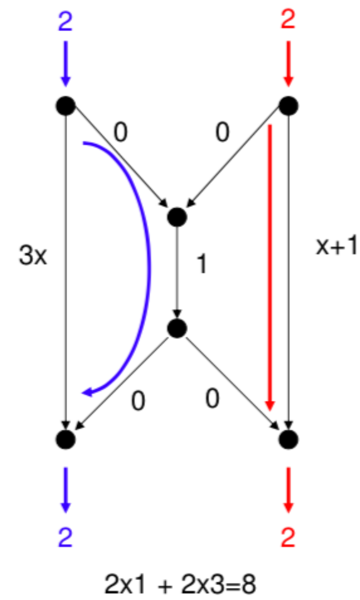



Fig. 7: Total latency cost

*Remark 6:* Similar to the no-deviation argument in Nash equilibrium, if a flow does not have this property, some traffic can improve its travel time by switching to a path with shorter travel time.

*Proposition 4:* If $f$ is a flow at Wardrop equilibrium for instance $(G, r, l)$, then the flow cost is

$$C(f) = \sum_{i \in \mathcal{N}} L_i(f) r_i.$$

*Assumption 1:* Each player in the network controls a negligible fraction of the overall traffic.

*Remark 7:* The definition of a flow at Wardrop equilibrium corresponds to an equilibrium in which each player chooses a single path of the network (a pure strategy), whereas in classical game theory a Nash equilibrium is defined via mixed strategies (with players of a game choosing probability distributions over pure strategies). With assumption 1, these two definitions are essentially equivalent.[29]

## C. Flow at system optimum

Wardrop's second principle introduced the idea of system optimum, that is, an assignment of traffic to paths minimizing the sum of all travel times (the total latency).

*Definition 13:* Flow at system optimum, or **optimal flow**, is the flow minimizing the total latency. The problem of finding a minimum cost (minimum latency) feasible flow in a multicommodity network can be modeled as the following *nonlinear program (NLP)*:

$$\begin{aligned}
\text{minimize} \quad & \sum_{e\in E} c_e(f_e) \\
\text{subject to} \quad & \sum_{P\in\mathcal{P}_i} f_P = r_i \qquad \forall i \in \mathcal{N} \\
& f_e = \sum_{P\in\mathcal{P}: e\in P} f_P \quad \forall e \in E \\
& f_P \geq 0, \qquad\qquad \forall P \in \mathcal{P}
\end{aligned}$$

Recalling that $c_e(f_e) = l_e(f_e)f_e$.

*Remark 8:* This formulation has an exponential number of variables, but it can be converted to an equivalent compact formulation (with decision variables only on edges and explicit flow conservation constraints at nodes) that requires only polynomially many variables and constraints.

Beckmann, Martin and Mcgurie (1956)[6] provided an interpretation of an optimal flow as a flow at Wardrop equilibrium with respect to a different set of edge latency functions called *marginal cost*.

*Definition 14:* The **marginal cost** of increasing flow on edge $e$ with differentiable latency function $l_e$ is denoted as

$$l_e^*(x) = \frac{d(x\cdot l_e(x))}{dx} = l_e(x) + x\cdot l_e'(x),$$

where $x\cdot l_e'(x)$ is called marginal external cost of congestion, or Pigovian tax.[17]

*Proposition 5:* Let $(G,r,l)$ be an instance with differentiable latency functions in which $x\cdot l_e(x)$ is a convex function for each edge $e$, with marginal functions $l^*$ defined above. Then a feasible flow $f$ is optimal if and only if it is at Wardrop equilibrium for the instance $(G,r,l^*)$.

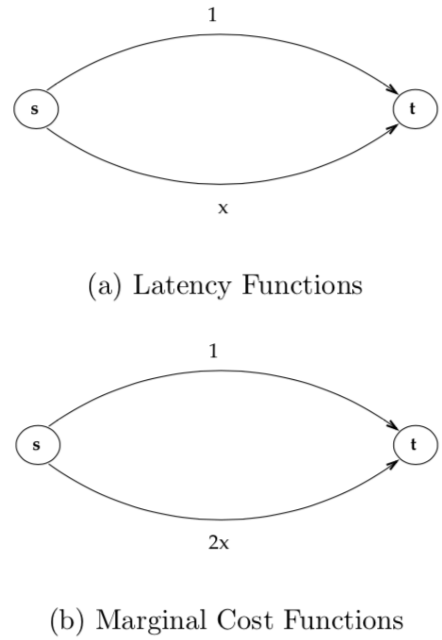
*Remark 9:* The optimal flow $f^*$ arises as a flow at Wardrop equilibrium with respect to latency function $l^*$. In this sense, the marginal cost function $l^*$ is actually the "optimal latency function".

*Example 3:* in Pigou's example shown in Subsection I-B2, the network has two nodes $s$ and $t$, two parallel edges with latency functions $l(x) = 1$ and $l(x) = x$, and 1 unit of traffic flow.

The latency of the two available paths from $s$ to $t$ is equaled when all flow routes on the bottom link. By Definition 12 we can obtain the flow at Nash equilibrium and by Proposition (or by inspection), the flow cost $C(f)$ is 1.

Given the latency functions, we can easily calculate the marginal cost functions: $l^*(x) = 1$ and $l^*(x) = 2x$ (see Figure 8). According to Proposition 5, when $x = 0.5$, that is, routing half of the traffic on each link can equalize the marginal costs on the two paths and induce a minimum latency flow $f^*$, giving the flow cost

$$C(f^*) = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot 1 = \frac{3}{4}.$$




(b) Marginal Cost Functions

Fig. 8: Marginal cost functions in Pigou's example

*Example 4:* Another example is the Baress's Paradox introduced in Subsection I-B3. Still consider one unit of flow from $s$ to $t$. After the addition of the zero-latency edge, there are three available paths. By Definition 12, the flow $f$ at Wardrop equilibrium would route all traffic on the path $s \to v \to w \to t$ and the latency of all three paths is equal to 2, as mentioned in Subsection I-B3. The flow cost is $C(f) = 2$.

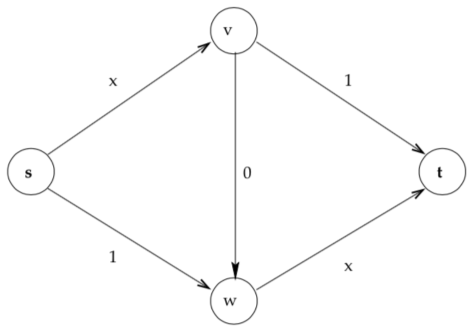
Replacing the latency functions with marginal cost functions (see Figure 9), the flow $f^*$ that routes half the traffic on each of the two two-hop paths can make the marginal costs of the three s-t paths equal, and is optimal by Proposition 5. The flow cost is calculated as

$$C(f^*) = \frac{1}{2}\cdot\left(\frac{1}{2}+1\right) + \frac{1}{2}\cdot\left(1+\frac{1}{2}\right) = \frac{3}{2}.$$

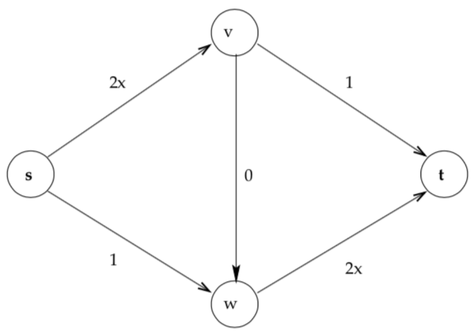
### D. Price of Anarchy

Recall the question raised in Subsection I-B2, given an arbitrary multicommodity flow network with congestion-dependent link latency, the gap between the total latency of a flow at Nash equilibrium and at system optimum (the minimum total latency) can be quantified as the *worst-case ratio* between them [22,30].

Worse-case ratio has a fancy name *Price of Anarchy* that is borrowed from Koutsoupias and Papadimitriou (1999) [1,31] as a measure of lack of centralized coordination and a bound for the inefficiency of Nash equilibrium. They use a simple load balancing model, which is somewhat related to the resource sharing problem mentioned




(a) Latency Functions




(b) Marginal Cost Functions

Fig. 9: Marginal cost functions in Braess's paradox

in Subsection I-B1. In the load balancing model, a finite number of agents share a set of parallel links, and each of them chooses a probability distribution on the set of links (specifying the probability that the user would route all the flow on a given link). Each agent wishes to minimize the expected congestion he/she will experience, while the global objective is to minimize the expected load on the most congested edge. Then the worst-case Nash equilibrium is compared to a globally optimal choice of distributions. They obtained a tight analysis of this worst-case ratio in two-node, two-link networks and partial results for two-node networks with three or more parallel links.

Roughgarden and Tardos have a complete, successive analysis on general link latency functions.

### E. Selfish Routing

They first provided results for the simplest nontrivial networks with linear latency functions of the form $l(x) = ax + b$. Such networks have already been extensively studied [32,33,34,35] in the field of Transportation and Operations Research.

*Theorem 7:* In any multicommodity flow network with linear latency functions, the total latency of a flow at Nash equilibrium is at most $\frac{4}{3}$ times that of a minimum-latency flow.

For an instance $(G, r, l)$ in which each link latency function $l_e$ is of the form $l_e(x) = a_e x + b_e$, if $f$ and $f^*$ are the flow at Wardrop equilibrum and system optimum respectively, then the ratio (denoted as $\rho$) can be bounded by

$$\rho(G, r, l) = \frac{C(f)}{C(f^*)} \leq \frac{4}{3}.$$

*Proof.* By Proposition 4,

$$C(f) = \sum_{i \in \mathcal{N}} L_i(f) r_i.$$

Then they try to create an optimal flow based on $f$ for the instance $(G, r, l)$ via a two-step process (see Figure 11). In the first step, a flow optimal for the instance $(G, r/2, l)$ is sent through $G$ and in the second step this flow is augmented to one optimal for $(G, r, l)$. The first flow has cost at least $\frac{1}{4}C(f)$ and that the augmentation has cost at least $\frac{1}{2}C(f)$.

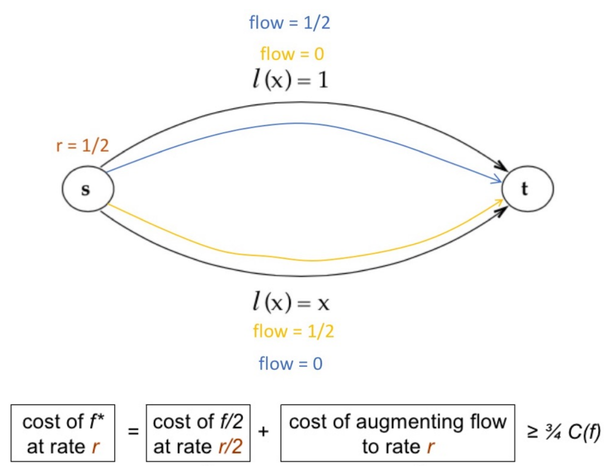



Fig. 10: Proof sketch of Theorem 7

For the first step, from Lemma 1 we know that the sent flow should be **half of the Nash flow**, the lower bound of this flow cost is

$$\begin{aligned} C(f/2) &= \sum_e \frac{1}{4} a_e f_e^2 + \frac{1}{2} b_e f_e \\ &\geq \frac{1}{4} \sum_e a_e f_e^2 + b_e f_e \\ &= \frac{1}{4} C(f) \end{aligned}$$

For the second step, marginal costs with respect to $f/2$ and latency with respect to $f$ coincide, that is, $L_i^*(f/2) = L_i(f)$ by 1. This establishes the necessary connection between the cost of augmenting $f/2$ to a flow feasible for

$(G, r, l)$ and the cost of a flow at Wardrop equilibrium. The augmentation cost is at least

$$\sum_{i \in \mathcal{N}} L_i^*(f/2)\frac{r_i}{2} = \frac{1}{2}\sum_{i \in \mathcal{N}} L_i(f) r_i = \frac{1}{2}C(f).$$

Summing the two parts, we have

$$C(f^*) \geq C(f/2) + \frac{1}{2}C(f) = \frac{3}{4}C(f).$$

■

*Remark 10:* This theorem gives an upper bound of the ratio. The lower bound of the ratio is already provided by Pigou's example and Baress's paradox where

$$\frac{C(f)}{C(f^*)} = \frac{4}{3}.$$

The analysis can be easily extended to network with nonlinear (polynomial) link latency functions like $l_e(x) = a_e x^p + b_e$ (where $a_e, b_e \geq 0$), the upper bound of the worst-case ratio is now

$$\begin{aligned} \rho(G, r, l) \leq & \frac{1}{1 - p \cdot (p+1)^{-(p+1)/p}} \\ = & \frac{(p+1)\sqrt[p+1]{p+1}}{(p+1)\sqrt[p+1]{p+1} - p} \\ = & \Theta\left(\frac{p}{\ln p}\right) \end{aligned}$$

whose limit goes to infinity, which means the ratio can be unbounded. This implies that *selfish routing can seriously degrade performance*.

The high-level idea is shown in Figure 11.

Next, they develop the techniques for networks with linear latency functions to obtain the price of anarchy with respect to other general classes of latency functions.

Noting that directly adapting the three-step approach in the proof of Theorem 7 to more general latency functions fails (even for quadratic functions) since there is no constant $c$ ($c = 2$ in the linear case) for which a scaled-down version $f/c$ of a Nash flow $f$ is optimal for the reduced traffic rates $r/c$.

To circumvent this problem, they view the proof of Theorem 7 from another perspective, that is, breaking up an optimal flow into two "pieces"and bounding each piece in terms of the cost of a Nash flow. In order to define the second piece of the optimal flow as an augmentation of the first and to lower bound its cost by means of marginal cost functions, the first piece is defined in a way to ensure that any augmentation with respect to it has large marginal cost. However, this requires scaling down a Nash flow $f$ by different factors on different links, thereby producing a "*pseudoflow*" (since the flow conservation constraints are violated).

In summary, the two-step process creating an optimal flow $f^*$ for instance $(G, r, l)$ proceeds as follows:

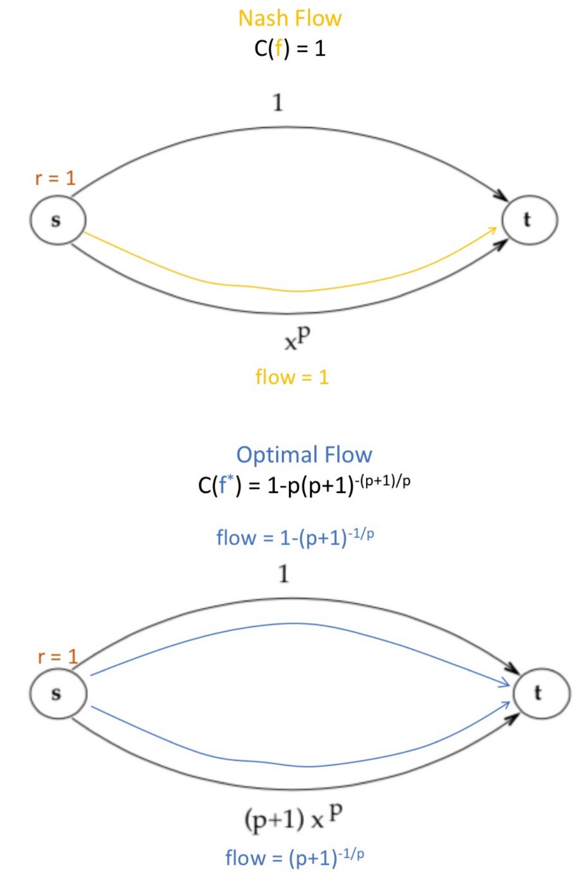



Fig. 11: Example of a network with polynomial latency functions

- Send a *pseudoflow* $\{\lambda_e f_e\}_{e \in E}$ such that
$$l_e^*(\lambda_e f_e) = l_e(f_e).$$
- Augment the *pseudoflow* to an optimal flow $f^*$.

The remain equstion is to lower bound the cost of an augmentation from the scaled-down pseudoflow to a flow feasible for the original instance.

*Example 5:* Consider the simple parallel-link network with latency functions $l(x) = 1$ and $l(x) = x^2$ (see Figure 12). The marginal cost function of the bottom link is

$$l^*(x) = x^2 + (2x) \cdot x = 3x^2.$$

Now we want

$$l^*(\lambda x) = l(x) \Leftrightarrow 3(\lambda x)^2 = x^2,$$

the scaling down factor should be $\lambda = 1/\sqrt{3}$.

Their results are summarized in Table I, in which polynomial coefficients are assumed nonnegative. The parameters $u$ and $\sigma$ are the expectation and standard deviation of the associated queue service rate distribution. $R_{\max}$ denotes the maximum allowable amount of

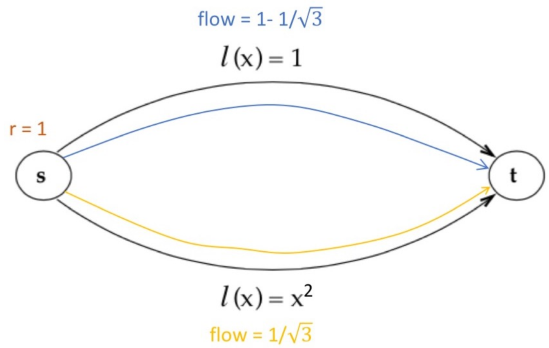



Fig. 12: Scaling-down factor of the pseudoflow

network traffic, and $u_{\min}$ denotes the minimum allowable link service rate (or capacity).

| Type of Link Latency Function | Mathematical Form | Price of Anarchy |
|---|---|---|
| Linear | $ax+b$ | $\frac{4}{3} \approx 1.333$ |
| Quadratic | $ax^2+bx+c$ | $\frac{3\sqrt{3}}{3\sqrt{3}-2} \approx 1.626$ |
| Cubic | $ax^3+bx^2+cx+d$ | $\frac{4\sqrt[3]{4}}{4\sqrt[3]{4}-3} \approx 1.896$ |
| Polynomial of degree $\leq p$ | $\sum_{i=0}^{P} a_i x^i$ | $\frac{(p+1)\sqrt[p+1]{p+1}}{(p+1)\sqrt[p+1]{p+1}-p} = \Theta\left(\frac{p}{\ln p}\right)$ |
| M/M/1 delay function | $(u-x)^{-1}$ | $\frac{1}{2}(1+\sqrt{\frac{u_{\min}}{u_{\min}-R_{\max}}})$ |
| M/G/1 delay function | $\frac{1}{u}+\frac{x(1+\sigma^2u^2)}{2u(u-x)}$ | very complicated |

TABLE I: The Price of Anarchy for common classes of link latency functions.

## V. Partially Optimal Routing

In many networks, there will be a mix of "selfishly controlled" and "centrally controlled" traffic—that is, the network is used by both selfish individuals and some central authority. For example, clients of a network may be charged at two different prices: clients paying the higher price are given access to the network and the ability to route their own traffic (presumably along a minimum-latency path), while clients paying only the "bargain rate" can use the network but have no control over how their traffic is routed (and thus this traffic qualifies as centrally controlled). Also, Korilis, Lazar and Orda (1997)[36] consider networks that allow a large customer to set up a so-called virtual private network of guaranteed and preassigned virtual paths for ongoing use,[37] and argue that the bandwidth needed for a *virtual private network* may be viewed as centrally controlled (with the paths chosen by the network manager) while individual users of the network continue to behave in a selfish and independent fashion.

### A. Stackelberg Routing

Roughgarden (2001)[10] investigated the following question: *given a network with centrally and selfishly controlled traffic, how should centrally controlled traffic be routed to induce "good" (selfish) behavior from the noncooperative users*? This indirect approach to controlling selfish behavior has several appealing aspects: no communication is required between network users and an algorithm, no notion of currency is needed (see the approach of algorithmic mechanism design [11,38,39]), no resources need to be added to or removed from the network, and the routing of centrally controlled traffic is often easily modified as the amount of traffic evolves over time.

He formulated this goal as an optimization problem via Stackelberg game in which the roles of different players are asymmetric. One player (responsible for routing the centrally controlled traffic and interested in minimizing total latency) acts as a *leader*, who holds his/her routing (strategy) fixed while all other players (the *followers*) react independently and selfishly to the leader's strategy, reaching a Nash equilibrium relative to it.

*Stackelberg game*, and the resulting solution concept of *Stackelberg equilibrium* have been well studied in the game theory literature (see [40,41] for introduction). Korilis, Lazar and Orda (1997)[42] focused on deriving necessary and sufficient conditions (on the number of selfish users, the fraction of the traffic that is centrally controlled, etc.) for the existence of a leader strategy inducing an optimal routing of all of the traffic; moreover, only one type of latency function is considered. By contrast, Roughgarden was interested in simple leader strategies that *always* induce optimal or **near-optimal** behavior from the network users for *any* set of latency functions.

After formulating the game, Roughgarden studied the problem of computing the optimal *Stackelberg strategy* in networks of parallel links, that is, a strategy for the leader that induces the followers to react in a way (at least approximately) minimizing the total latency in the network.

He proved that it is NP-hard to compute the optimal Stackelberg strategy in networks of parallel links and present simple strategies for such networks with provable performance guarantees. More precisely, he gave a simple algorithm that computes a leader strategy in a network of parallel links inducing an equilibrium with total latency no more than a constant times that of the minimum-latency flow; a simple variant on Pigou's example showed that no result of this type is possible in the absence of centrally controlled traffic and a Stackelberg strategy. He also proved stronger performance guarantees for networks of parallel links with linear latency functions.

*Definition 15:* A **Stackelberg instance** is a tuple $(G, r, l, \beta)$ referring to a routing instance $(G, r, l)$ with a parameter $\beta \in [0, 1]$ specifying the fraction of the network traffic that is centrally controlled.

*Definition 16:* A **Stackelberg strategy** for the Stackelberg instance $(G, r, l, \beta)$ is a flow feasible for $(G, \beta, l)$.

*Definition 17:* Let $f$ be a strategy for Stackelberg instance $(G, r, l, \beta)$, and define $\tilde{l}$ by

$$\tilde{l}_e(x) = l_e(f_e + x)$$

for link $e \in E$.

An **equilibrium induced by strategy** $f$ is a flow $g$ at Wardrop equilibrium for the instance $(G, (1-\beta)r, \tilde{l})$. We then say that $f+g$ is a **flow induced by** $f$ for $(G, r, l, \beta)$.

*Example 6:* Recall Pigou's example (Subsection I-B2). In the absence of centrally controlled traffic, a Nash flow incurs total latency $\frac{4}{3}$ times that of the optimal flow (see Remark 10). Suppose instead that half of the traffic is controlled by the network manager (that is, $\beta = 1$) and consider the strategy $f$ of routing all centrally controlled traffic on the top link (the link with latency function $l(x) = 1$). Then, as all remaining traffic will be routed on the lower link in the equilibrium induced by $f$, the flow induced by $f$ is precisely the minimum latency flow. Thus, in this particular instance, total latency can be minimized via a Stackelberg strategy.

Now slightly modify Pigou's example, in which we replace the latency function of the lower link with the latency function $l(x) = 2x$. The Nash flow puts half of the traffic on each edge for a cost of 1, while the optimal flow routes only $\frac{1}{4}$ of the traffic on the lower link, incurring a cost of $\frac{7}{8}$. On the other hand, if we again allow the network manager to route half of the traffic, we see that for *any* strategy $f$, each link has less than half traffic. The rest would spontaneously increase the traffic on both links to half of the traffic (see Figure 13). Whatever, the flow induced by $f$ is the flow at Wardrop equilibrium and hence is not optimal. In this example, there is no available strategy by which the network manager can improve network performance.

Three strategies are presented in Roughgarden's work, the first two are natural and sub-optimal while the third one is near-optimal. For simplicity, we introduce the three strategies under the network with linear functions and half of the traffic is centrally controlled ($\beta = \frac{1}{2}$).

- *Aloof* strategy. For an instance $(G, r, l, \frac{1}{2})$, if $f^*$ is the minimum-latency flow for instance $(G, \frac{r}{2}, l)$, then set $f = f^*$. This strategy can be interpreted as ignoring the existence of traffic that is not centrally controlled and choosing the strategy of minimum cost. However, it performs quite poorly in Pigou's example which route all flow on the bottom link and in the induced flow all traffic is routed on the bottom link.

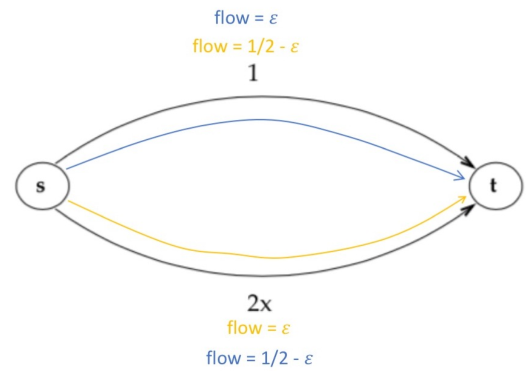



Fig. 13: Limitation of Stackelberg routing

- *Scale* strategy. If $f^*$ is the minimum-latency flow for $(G, r, l)$, then set $f = \frac{1}{2}f^*$, the scaled optimal flow. This strategy does not work well either. Again consider the two-node, two-link example with latency functions $l_1(x) = 1$ and $l_2(x) = 3x$ and traffic rate 1. The minimum-latency flow routes $\frac{2}{3}$ of the flow on the first link and the rest on the second link for a cost of $\frac{5}{6}$, hence the Scale strategy would route $\frac{1}{3}$ units of flow on the first link and $\frac{1}{6}$ units on the second link. All selfish traffic is then routed on the second link, inducing the flow with 1 units of flow on the first link and the rest on the second. Now the cost is 1, larger than $\frac{7}{8}$. It seems that the Scale strategy routed too much flow on the second link.
- The *Largest Latency First (LLF)* Strategy. A link $e$ is said to be *saturated* by a strategy $f$ if $f_e = f_e^*$. The LLF strategy saturates edges one-by-one in order from the largest latency with respect to $f^*$ to the smallest, until there is no centrally controlled traffic remaining. The running time is $O(|E|^2)$ when all link latency functions are linear.

### B. *Traffic engineering within subnetworks*

Most large-scale networks, such as Internet, consist of interconnected administrative domains that control traffic within their networks. The routing model proposed by Roughgarden and Tardos is not perfectly suitable for communication networks since the providers' role in routing traffic (that is, *traffic engineering*) is ignored. More precisely, there are conflicting interests between users and network providers. while users care about end-to-end performance, individual network providers seek to optimize their own objective. Such strategic interaction between *selfish across-domain routing* and *traffic engineering* (network providers redirect traffic within their administrative domain to achieve minimum total latency[43]) has been studied in several works [44,45,46,47,23].

Acemoglu, Johari and Ozdaglar (2007) first formally modeled the interaction as *partially optimal routing* and theoretically addressed the key question of whether partially optimal routing (traffic engineering) improves overall network performance.

Their model inherited the multicommodity network and the routing model. Note that the latency functions cannot only describe the congestion on roads, but also delay or probability of packet loss. Their innovation is allowing subsets of the links in the network ("subnetworks") to be independently owned and operated by different providers, and consider the possibility that these providers engage in traffic engineering and route traffic to minimize the total (or average) latency within their subnetworks.

The answer to the key question is "it depends". On the one hand, when the Braess' paradox occurs within the global network, partially optimal routing may have a higher total latency cost than the Wardrop equilibrium. On the other hand, there exists a special class of network structure that guarantee the improvement of network performance.

First we slightly modify the routing instance in IV as network routing instance here.

*Definition 18:* A network routing instance is a tuple $R = (V, E, P, s, t, r, l)$ where

- $(V, E)$ is the network we want to study
- $s, t \in V$ are distinguished source and destination nodes
- $P$ is the set of paths from $s$ to $t$
- $r$ and $l$ have the same destination with that in the routing instance

Giving routing instance $(V, E, P, s, t, r, l)$ and $(V_0, E_0, P_0, s_0, t_0, r, l)$ where $(V_0, E_0)$ is the subnetwork of $(V, E)$. Define a **new routing instance** $(V', E', P', s, t, r, l')$ as follows

- $V' = (V \backslash V_0 \cup \{s_0, t_0\}$
- $E' = (E \backslash E_0 \cup \{(s_0, t_0)\}$
- $l' = \{l_e\}_{e \in E \backslash E_0} \cup \{l_0\}$

Name $R'$ as the **equivalent POR instance** for $R$ with respect to $R_0$. The overall network flow in $R$ with partially optimal routing in $R_0$, $f^{POR}(R, R_0)$, is defined as:

$$f^{POR}(R, R_0) = f^{WE}(R'),$$

where $R'$ is the equivalent POR instance for $R$ with respect to $R_0$ and $f^{WE}$ is the flow at Wardrop equilibrium.

The following result can be viewed as generalized Baress's Paradox.

*Proposition 6:* Consider the routing instance $R = (V, E, P, s, t, r, l)$ and subnetwork routing instance $R_0 = (V_0, E_0, P_0, s_0, t_0, r, l)$. Assume that the POR paradox occurs in $R$ with respect to $R_0$. Then Braess' paradox occurs in $R$.

*Proof.* I only provide the high-level idea here. Uniformly lower the latency functions in the subnetwork $R_0$, such that the Wardrop effective latency of $R_0$ is given by $l_0$ (the effective latency of optimal routing within $R_0$).

*Corollary 1:* Given a routing instance $R$, if Braess' paradox does not occur in $R$, then partially optimal routing with respect to any subnetwork always improves the network performance.

Milchtaich (2006)[48] has shown that Braess' paradox does not occur in directed graphs where the underlying undirected graph has a *series-parallel* structure. Based on this, they discovered that for a network with *serial-parallel* links, partially optimal routing always improves the overall network performance.

They also adopted the concept of Price of Anarchy and discussed its application in the problem of partially optimal routing.

*Theorem 8:* Consider a routing instance $R \in \mathcal{R}^{aff}$ with affine latency functions and a subnetwork $R_0$ of $R$. Then,

$$\frac{C(f^{SO}(R))}{C(f^{POR}(R, R_0))} \geq \frac{3}{4}.$$

This result follows from Lemma 2 and the following proposition.

*Proposition 7:* Consider a routing instance $R \in \mathcal{R}^{conc}$ with concave latency functions and a subnetwork $R_0$ of $R$. Then,

$$\frac{C(f^{SO}(R))}{C(f^{WE}(R))} \geq \frac{3}{4}.$$

They prove this proposition using similar geometric proof as [49,50] (see Figure 14).

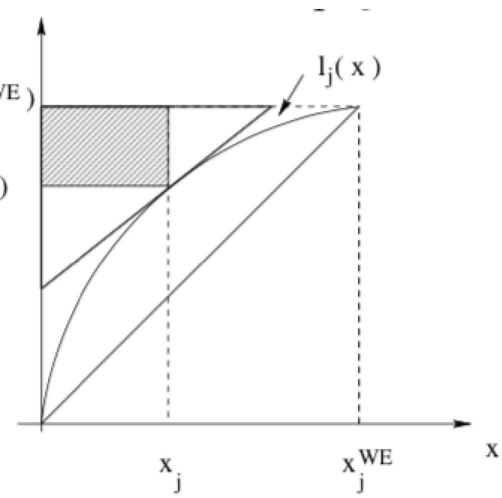



Fig. 14: Proof of proposition

## VI. Concluding Remarks

### A. Contributions and limitations

Rosenthal defined the congestion game and proved the existence of pure-strategy Nash equilibrium. However, he only assumed atomic players and numeric payoffs. Monderer and Shapley introduced potential functions and potential game. They also showed the isomorphism between congestion game and potential game.

Koutsoupias and Papadimitriou proposed the concept of "Price of Anarchy". Roughgarden and Tardos formulated routing game and applied the idea of Price of Anarchy to the analysis of selfish routing. They also adapted the algorithmic mechanism design framework of Nisan and Ronan to the study of system operator's optimal strategy. The six researchers (three ground-breaking works) contribute to the growth of algorithmic game theory.

Noting the limitation of routing game model, Acemoglu, Johari and Ozdaglar included the providers' roles in partially optimal routing and pricing. They also followed the definition of Price of Anarchy.

### B. Future directions

The interactions between selfish behavior and global optimization in networks are rather complex. There are many questions remain to be answered.

One is the extent of Baress's Paradox in typical networks. Although Roughgarden has provided a worst-case analysis and generalized Baress's Paradox for primally optimal routing has been discussed, there is still a lack of general and practical framework for network designers to reference.

Another direction is traffic control. Roughgarden's Stackelberg routing is a small step. What if *admission control* is permitted? What if there are attackers who have the power of spoofing data and manipulating routing instructions in order to achieve their selfish goals? What if the decisions of players are not made simultaneously or there are multiple stages of the decision-making process (each stage as a simultaneous game, Stackelberg game or even differential game)? How to capture the dynamic nature of interactions?

## APPENDICES

*Lemma 1:* Suppose $(G, r, l)$ has linear latency functions and $f$ is a flow at Wardrop equilibrium. Then,

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

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

*Proof.* From Definition 12 and Definition 9 we know that for every $i \in \mathcal{N}$ and $P, P' \in \mathcal{P}_i$ with $f_P > 0$,

$$\sum_{e \in P} a_e f_e + b_e \leq \sum_{e \in P'} a_e f_e + b_e.$$

It follows from the relationship between the latency function and the marginal cost function (see Definition 14) that $f/2$ is optimal for $(G, r/2, l)$ since

$$\sum_{e \in P} 2a_e(f_e/2) + b_e \leq \sum_{e \in P'} 2a_e(f_e/2) + b_e.$$

Using the same manner we can argue that $l_e^*(f_e/2) = l_e(f_e)$ for each link $e$ and hence with Definition 9, $l_P^*(f/2) = l_P(f)$ for each path $P$.

*Lemma 2:* Assume that the latency functions $l_i$ of all the links in the subnetwork are nonnegative affine functions. Then, the effective latency of POR, $l_0(r_0)$, is a nonnegative concave function of $r_0$.

## REFERENCES


[1] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *Annual Symposium on Theoretical Aspects of Computer Science*. Springer, 1999, pp. 404–413.

[2] T. Roughgarden and É. Tardos, "How bad is selfish routing?" *Journal of the ACM (JACM)*, vol. 49, no. 2, pp. 236–259, 2002.

[3] J. F. Nash *et al.*, "Equilibrium points in n-person games," *Proceedings of the national academy of sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[4] J. Nash, "Non-cooperative games," *Annals of mathematics*, pp. 286–295, 1951.

[5] J. G. Wardrop, "Road paper. some theoretical aspects of road traffic research." *Proceedings of the institution of civil engineers*, vol. 1, no. 3, pp. 325–362, 1952.

[6] M. Beckmann, C. B. McGuire, and C. B. Winsten, "Studies in the economics of transportation," Tech. Rep., 1956.

[7] S. C. Dafermos and F. T. Sparrow, "The traffic assignment problem for a general network," *Journal of Research of the National Bureau of Standards B*, vol. 73, no. 2, pp. 91–118, 1969.

[8] R. W. Rosenthal, "The network equilibrium problem in integers," *Networks*, vol. 3, no. 1, pp. 53–59, 1973.

[9] Rosenthal and R. W, "A class of games possessing pure-strategy nash equilibria," *International Journal of Game Theory*, vol. 2, no. 1, pp. 65–67, 1973.

[10] T. Roughgarden, "Designing networks for selfish users is hard," in *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. IEEE, 2001, pp. 472–481.

[11] N. Nisan, "Algorithms for selfish agents," in *Annual Symposium on Theoretical Aspects of Computer Science*. Springer, 1999, pp. 1–15.

[12] C. Fisk, "Game theory and transportation systems modelling," *Transportation Research Part B: Methodological*, vol. 18, no. 4-5, pp. 301–313, 1984.

[13] Y. Hollander and J. N. Prashker, "The applicability of non-cooperative game theory in transport analysis," *Transportation*, vol. 33, no. 5, pp. 481–496, 2006.

[14] H. Zhang, Y. Su, L. Peng, and D. Yao, "A review of game theory applications in transportation analysis," in *2010 International Conference on Computer and Information Application*. IEEE, 2010, pp. 152–157.

[15] W. B. Arthur, "Inductive reasoning and bounded rationality," *The American economic review*, vol. 84, no. 2, pp. 406–411, 1994.

[16] B. Mishra, A. Greenwald, and R. Parikh, "The santa fe bar problem revisited: Theoretical and practical implications," in *Proceedings of the summer festival on game theory international conference, State University of New York at Stony Brook, July 20-24, 1998*, 1998.

[17] A. Pigou, *The economics of welfare*. Routledge, 2017.

[18] F. H. Knight, "Some fallacies in the interpretation of social cost," *The Quarterly Journal of Economics*, vol. 38, no. 4, pp. 582–606, 1924.

[19] D. Braess, "Über ein paradoxon aus der verkehrsplanung," *Unternehmensforschung*, vol. 12, no. 1, pp. 258–268, 1968.

[20] J. D. Murchland, "Braess's paradox of traffic flow," *Transportation Research*, vol. 4, no. 4, pp. 391–394, 1970.

[21] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[22] T. Roughgarden, "Selfish routing," CORNELL UNIV ITHACA NY DEPT OF COMPUTER SCIENCE, Tech. Rep., 2002.

[23] D. Acemoglu, R. Johari, and A. Ozdaglar, "Partially optimal routing," *IEEE Journal on selected areas in communications*, vol. 25, no. 6, pp. 1148–1160, 2007.

[24] G. W. Brown, "Some notes on computation of games solutions," RAND CORP SANTA MONICA CA, Tech. Rep., 1949.

[25] L. R. Ford Jr and D. R. Fulkerson, "A suggested computation for maximal multi-commodity network flows," *Management Science*, vol. 5, no. 1, pp. 97–101, 1958.

[26] J. Tomlin, "Minimum-cost multicommodity network flows," *Operations Research*, vol. 14, no. 1, pp. 45–51, 1966.

[27] R. Steinberg and W. I. Zangwill, "The prevalence of braess' paradox," *Transportation Science*, vol. 17, no. 3, pp. 301–318, 1983.

[28] R. Steinberg and R. E. Stone, "The prevalence of paradoxes in transportation equilibrium problems," *Transportation Science*, vol. 22, no. 4, pp. 231–241, 1988.

[29] A. Haurie and P. Marcotte, "On the relationship between nash—cournot and wardrop equilibria," *Networks*, vol. 15, no. 3, pp. 295–308, 1985.

[30] T. Roughgarden, *Selfish routing and the price of anarchy*. MIT press Cambridge, 2005, vol. 174.

[31] C. H. Papadimitriou, "Algorithms, games, and the internet," in *International Colloquium on Automata, Languages, and Programming*. Springer, 2001, pp. 1–3.

[32] M. Frank, "The braess paradox," *Mathematical Programming*, vol. 20, no. 1, pp. 283–302, 1981.

[33] Frank and Marguerite, *Cost Deceptive Links on Transportation Networks*. Systems Optimization Laboratory, Department of Operations Research, Stanford . . . , 1983.

[34] E. I. Pas and S. L. Principio, "Braess' paradox: Some new insights," *Transportation Research Part B: Methodological*, vol. 31, no. 3, pp. 265–276, 1997.

[35] C. M. Penchina, "Braess paradox: Maximum penalty in a minimal critical network," *Transportation Research Part A: Policy and Practice*, vol. 31, no. 5, pp. 379–388, 1997.

[36] Y. A. Korilis, A. A. Lazar, and A. Orda, "Capacity allocation under noncooperative routing," *IEEE Transactions on Automatic Control*, vol. 42, no. 3, pp. 309–325, 1997.

[37] K. P. Birman, "Building secure and reliable network applications," in *International Conference on Worldwide Computing and Its Applications*. Springer, 1997, pp. 15–28.

[38] N. Nisan and A. Ronen, "Algorithmic mechanism design," *Games and Economic behavior*, vol. 35, no. 1-2, pp. 166–196, 2001.

[39] A. Ronen, "Algorithms for rational agents," in *International Conference on Current Trends in Theory and Practice of Computer Science*. Springer, 2000, pp. 56–70.

[40] T. Basar and G. J. Olsder, *Dynamic noncooperative game theory*. Siam, 1999, vol. 23.

[41] D. Fudenberg and J. Tirole, "Game theory, 1991," *Cambridge, Massachusetts*, vol. 393, no. 12, p. 80, 1991.

[42] Y. A. Korilis, A. A. Lazar, and A. Orda, "Achieving network optima using stackelberg routing strategies," *IEEE/ACM transactions on networking*, vol. 5, no. 1, pp. 161–173, 1997.

[43] J. Rexford, "Route optimization in ip networks," in *Handbook of Optimization in Telecommunications*. Springer, 2006, pp. 679–700.

[44] L. Qiu, Y. R. Yang, Y. Zhang, and S. Shenker, "On selfish routing in internet-like environments," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*. ACM, 2003, pp. 151–162.

[45] H. Zhang, Y. Liu, W. Gong, and D. Towsley, "Understanding the interaction between overlay routing and traffic engineering," *University of Massachusetts CMPSCI, Technical Report*, pp. 04–63, 2004.

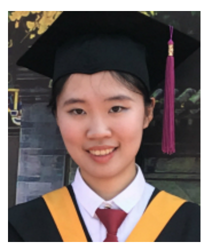
[46] Y. Liu, H. Zhang, W. Gong, and D. Towsley, "On the interaction between overlay routing and underlay routing," in *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, vol. 4. IEEE, 2005, pp. 2543–2553.

[47] F. Kelly and T. Voice, "Stability of end-to-end algorithms for joint routing and rate control," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 2, pp. 5–12, 2005.

[48] I. Milchtaich, "Network topology and the efficiency of equilibrium," *Games and Economic Behavior*, vol. 57, no. 2, pp. 321–346, 2006.

[49] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "On the inefficiency of equilibria in congestion games," in *International Conference on Integer Programming and Combinatorial Optimization*. Springer, 2005, pp. 167–181.

[50] Correa, J. R, A. S. Schulz, and N. E. Stier-Moses, "Selfish routing in capacitated networks," *Mathematics of Operations Research*, vol. 29, no. 4, pp. 961–976, 2004.



**Qian Xie** is a PhD student at the Department of Civil and Urban Engineering and the C2SMART University Transportation Center in the New York University Tandon School of Engineering. She received her B.Eng. degree in Computer Science from Tsinghua University in 2019. She is interested in studying resilience of civil infrastructures and designing secure network systems, using control-theoretic and game-theoretic approaches.