The impact of information-aware routing on road traffic
From case studies to game-theoretical analysis and simulations

by

Theophile Charles Prosper Cabannes

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alexandre M. Bayen, Chair
Professor Laurent El-Ghaoui
Professor Alexander Skabardonis
Professor Eric Goubault

Summer 2022

The impact of information-aware routing on road traffic
From case studies to game-theoretical analysis and simulations

Abstract

The impact of information-aware routing on road traffic
From case studies to game-theoretical analysis and simulations

by

Theophile Charles Prosper Cabannes

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Alexandre M. Bayen, Chair

During the 2010s decade, the increase of connectivity in the world led to the development of navigation applications that help vehicles to travel within the transportation network using real-time traffic information. *Information-aware routing* changed traffic patterns by spreading congestion in the network. Before information-aware routing, route choice was dictated by direction signs, themselves prescribed by city traffic plans. Consequently, with these new routing behaviors, some traffic plans have been outdated.

From a game theoretical point of view, by providing vehicles with the fastest path to reach their destination, information-aware routing suggestions direct the state of traffic toward a *Nash equilibrium*. The gap between the state of a game and a Nash equilibrium can be measured with the average deviation incentive. Using the restricting path choice model, we show that the average deviation incentive monotonically decreases when information-aware routing behaviors increase.

On the ground, if information-aware routing behaviors might (or might not) increase the overall transportation efficiency, some local roads receive more traffic than the one they are designed to sustain. In Los Angeles, CA, we measure a 3-fold increase in the flow of one I-210 off-ramp between 2014 and 2017. This is likely a consequence of the rise of information-aware routing behaviors due to an increase in navigational-app usage. Travel time data shows the travel time equalization between the main Eastbound I-210 route between Pasadena, CA, and Azusa, CA, and some alternate routes using local roads. The travel time equalization expresses that the state of traffic is a Nash equilibrium, which demonstrates the presence of information-aware routing behaviors. However, the severity of information-aware routing cannot be quantitatively assessed globally due to the lack of available floating-car data (trajectory data). Qualitatively, many cities and neighborhoods reported negative externalities of *cut-through traffic* due to information-aware routing: Los Angeles, CA recorded several crashes on the steep Baxter Street, Leonia, NJ reported a fatality on the Ford Lee Road,

and Fremont, CA faced the challenges of cut-through traffic on the Mission Boulevard. All around the world, cut-through traffic might induce higher travel times, delays, unreliable travel times for residents, noise, gas emissions, traffic accidents, decrease accessibility in affected neighborhoods, wrong directions, and infrastructure damage, among others. To mitigate these issues, many cities changed the design of their road network. Others used cap-and-trade techniques (e.g., access restriction for non-residents on the Ford Lee Road in Leonia, NJ), or Pigovian taxes (e.g., road pricing on Lombard Street in San Francisco, CA). The city of Fremont erected new stop signs, built speed bumps, and updated its traffic signal timing plans to decrease the Mission boulevard's attractivity.

To pick the best mitigation techniques to fight against cut-through traffic, we suggest using a digital twin of the city traffic. *Traffic microsimulators* can replicate, in the digital world, the movement of each vehicle within the road network. Because the challenges of running a traffic simulation are not apparent until one creates their own, we provide a blueprint to develop, calibrate and validate a traffic microsimulation. A traffic simulation of the Fremont, CA neighborhood affected by cut-through traffic due to information-aware routing is made open source, for anyone who would like to understand how information-aware routing might lead to cut-through traffic.

On the way, we realized that simulating the behaviors of each vehicle in the network is computationally expensive. Therefore, we proposed to cluster vehicles into a mean-field by developing a *mean-field routing game*. While macroscopic routing models already exist to estimate how, knowing the traffic demand, traffic evolves in the network, we envision that the mean-field routing game is the best tool to perform large-scale dynamic routing control. We also envision that large-scale dynamic routing control is enabled by navigational applications, with already existing features like *eco-routing* that have been launched in 2021 by *Google Maps*.

To my friends

Contents

List of Figures

List of Tables

Acknowledgments

"Je vous souhaite des rêves à
n'en plus finir et l'envie furieuse
d'en réaliser quelques uns."

Jacques Brel

I was introduced to Alexandre Bayen by Eric Goubault in late 2016. I first spoke with Alexandre Bayen in early 2017, trying to find a research internship outside of France, where I could apply computer science to urban planning. By then, I had not heard about U.C. Berkeley, and I was not able to say if San Francisco and California were on the West Coast or the East Coast of the U.S.. However, doing an internship with Alex Bayen seemed my best option at the time. Alex Bayen and this internship have changed my life. First, I was introduced to research. I discovered that both English and Mathematics are useful (I was more into physics and computer science in my undergraduate studies). Finally, Alex gave me the research subject that I addressed for more than five years, the negative impact of navigational applications on the quality of life in some residential areas. On the way, I found a job at Google research, still working on improving traffic "on the other side of the fence". Alex joined me at Google research, and it is a pleasure to continue working with him at Google. Besides influencing the beginning of my professional life, Alex has had a tremendous impact on my personal life. First, Alex enabled me to feel at home 10,000 km away from Paris: we speak the same language, went to the same school, and had the same beginning in professional life. Second, Alex introduced me to Anne-Sophie Bordry. Anne-Sophie, in addition to being a close friend, was the marriage officiant at my wedding with Marine Gasc (now Marine Cabannes I guess). Alex, I am inspired by your energy, your support, and your skills to promote our lab research. In addition, I am astonished by the trust you gave me. I am especially delighted that you let me coordinate the work of 16 GSIs for the EECS127/227AT class. The feeling I had when giving a 45-minute lecture about PCA that involves a (successful) live experiment in front of 400 students will always remain in my memory. Thank you, Alex!

Of the people I have had the chance to meet, only one person seems to have more energy than Alex: Marine, my wife. I first met Marine, the day I learned that I was admitted to the Ph.D. program of the EECS department at U.C. Berkeley. Marine, I am glad that you first believed that Berkeley was a prestigious bank. You thought that I was going to make a lot of money when I was going to move to the other side of the world. I also am glad that you thought that a thesis only takes 6 months, the time it took you to carry out your Master's thesis to become a physical therapist. Finally, I am glad that when asking your friends in France, they told you that a Ph.D. is done in 3 years (which is the case in France), while in the U.S. it usually takes 5 to 6 years. Your friends were not that wrong, due to COVID-19, I only spent two years in the U.S.. Every day, I am encouraged by your kindness, your smile, and your love. I am grateful to know you!

When it comes to kindness, I would like to thank Eric Goubault. Eric, I admire your cleverness joined with your humility. In addition to helping me to figure out some complex math equations, you have always been a stress-reliever. Your selflessness will continue to strike me for many years. I hope that I can follow your example, and keep my peace of mind when I am hit hard by the workload. Thank you, Eric, for all your support and your leading example!

Within the people I met during my Ph.D., I would first like to thank Neha Arora. Neha contacted me in November 2019 to speak about my work on simulating traffic in the city of Fremont, CA. It happens that I was in France for a week. Was it fate that led us to have our first meeting across a 9-hour time zone difference? Thank you, Neha, for your confidence in hosting me as an intern at Google. Thank you, again, for keeping the internship while the COVID-19 pandemic was changing the face of the world and the way we work. Thank you, for your inclusion that enabled me to work remotely for the spacetime team while being in France. Was it fate, again, that led you to suggest I join full-time Google Research the day when my flight back to the U.S. was canceled because the 3rd COVID-19 wave was rising all over the world?

During my Ph.D. at UC Berkeley, I had the chance to meet outstanding professors. First, I would like to thank Alexander Skabardonis, for his help to figure out what traffic engineering is. I am grateful to Laurent El-Ghaoui, who taught the EECS227B class, and has always helped me during the Ph.D.. Shankar Sastry has blown my mind during the EE291E - Hybrid system class lectures. Each EE291E lecture was a trip across time to understand how mathematicians create tools to answer simple problems that arise during our lives (how to represent the movement of an object, how to control it, how to model boundary conditions, etc.). In my opinion, Shankar represents what is missing in the French school of mathematics: explaining that mathematics is just a toolbox to model our world and that ultimately it is another world to discover. Alongside Shankar, I would like to thank Claire Tomlin. The introduction to linear-system was a great introduction to control. I am still amazed by the fact that in the discussion I had with Claire and Shankar, they have never said something that seemed irrelevant to me. I was blessed to meet Scott Shenker for 30 minutes to talk about game theory and the price of anarchy after a cold email. During this meeting, I discovered what sharpness means. I hope I will meet more people like Scott in my life, and I hope that French elites will learn from the U.S. what being accessible means. Thus far, I am speechless by the fact that Scott took 30 minutes of his time to speak to a first-year graduate student trying to find out how game theory can be useful in real-life. Finally, I would like to thank Gireeja Ranade for giving me her trust to teach the EECS127AT/227AT - optimization class, while she had a specific idea of how she wanted the class to look like.

Besides outstanding professors, I met remarkable students and researchers. First, my labmates were a second family to me. I have been involved as the "Chief Happiness Officer" of the lab to shape a new social life abroad. I will first acknowledge my friends Alben Bagabaldo, Yashar Farid, and Pariya for all the laughs and the good food shared with you. I feel that during my Ph.D., I visited Iran and the Philippines. I want to go there at some point. Then, I would like to thank my past roommates, Blaze Syka and Tom Sivertsen. We

had a lot of fun together! Hopefully, we will be able to meet at the Missouri Lounge again. I hope that BART did not pass by the house too often when Marine was visiting. I am grateful to the fun people group: Fangyu Wu, Jiayi Li, Marsalis Gibson, Zhe Fu, Askhan Yousefpour, Nathan Litche, Shuxia Tang, Joy Carpio, and Yiling You, and to the other labmates: Lucas Fischer, Vincent Figheria, Marco Sangiovanni-Vincentelli, Jessica Lazarus, Aboudy Kriechen, Alexander Keimer, Ayush Jain, Harry Dong, Junghwan Lee, Jon Davis, Eugene Vinitsky, Cathy Wu, Kathy Jang, Kanaad Parvate, Kevin Lin, and all the others.

Most of the thesis (especially chapter 5) would not have been possible without the help of the City of Fremont, and especially Noe Veloso and Daniel Miller. This would not have been possible without the help of Jane MacFarlane that helped me to work on traffic simulations. I am thankful to the Aimsun team and their support team that consistently helped the Fremont project team to understand the software, how to interact with it through scripts, and for the postgraduate license. Thanks to the Connected Corridor people, especially Anthony Patire and Francois Dion, who shared their knowledge about creating, calibrating, and validating an Aimsun microsimulation. Thanks to the France Berkeley Funds that funded the research collaboration between Eric Goubault and Alexandre Bayen Thanks to all the labmates that worked on Waze book that was largely re-used for chapter 2 and chapter 4 of the dissertation: Alex Keimer, Jessica Lazarus, Yashar Farid, Zhe Fu, Bingyu Zhao, Tania Veravelli, Pavan Yadevalli, Jiayi Li, Shuxia Tang, Arnaud De Guilhermier, Solene Olivier, Isabelle Zhou, and Henri Bataille. I thank all the Fremont project team: Alben, Alice Blondel (my cousin), Anson Tiong, Ayush, Carl Gan, Daniel Macuga, Daniel Zhang, Edson Romero, Jasper Lee, Jiayi, Jinheng Xu, Junghwan Lee, Jon Davis, Jose A La Torre, Lauren Zhou, Mengze Zhu, Michal Takac, Michael Zhang, Prakash Srivastava, Roham Ghotbi, Sayan Das, Shuli Yang, Trevor Wu, Xuan Su, Yanda Li, and Zixuan Yang. I was blessed to be able to collaborate with more than 20 people, and I hope I will be able to lead such a project again in the future. It was quite a journey to lead the creation of a microsimulation of the Fremont traffic, without having done any microsimulation before. Coordinating the work of many collaborators, while not truly knowing where we were going, was sometimes challenging. I hope that I was a great manager for this project.

Finally, I would like to thank the enablers: Helen Bassham and Shirley Salanio. Thank you so much for your kindness and your dedication. You are saving so much time for so many people. I think that, ultimately, the success of U.C. Berkeley is due to people like you!

At Google, I have met amazing people and coworkers. First, I am grateful to all the spacetime team: Yechen Li, Sanjay Ganapathy, Iveel Tsogsuren, Carolina Osorio, Alex Shashko, and Andrew Tomkins. I would like to also thank the collaborators in the Driving Sustainability team: Marc Nunkesser, Haizheng Zhang, Kevin Chen, and Ruben Lozano-Aguilera. It is stunning to see how much impact your team can have with the new eco-routing feature in Google Maps. Then, I am grateful to the mean-field game (MFG) team between DeepMind and Google Brain that welcomed me to France. Thank you for your help and your support, Romuald Elie and Olivier Pietquin. I appreciated the theoretical discussions, Mathieu Lauriere and Paul Muller. Many thanks for the code reviews, Marc Lanctot, Sergan Girgin, Julien Perolat, and Raphael Marinier. Besides the MFG team, thank

you to the policy team that also welcomed me to the Paris office: Olivier Esper, Hugues de Maupeou, Sylvain Beucher, Mathieu Vincens, and the others. Thanks to Omar Benjelloun, my mentor. The company is filled out with amazing people that I would like to thank.

Finally, I would like to thank my family for all their support. I would like to thank my father for giving me freedom in my personal life while giving me strong incentives in my professional life. I would like to thank my mother for giving me freedom in my professional life while giving me strong guidelines for my personal life. Thank you to my brothers, and their partners. I wish that at some point I will be able to discuss mathematics with my twin brother (who does very similar research as me). Thank you to my aunts, my uncles, and my cousins. Thank you to Meg, who introduced me to Linda, who kindly hosted me for 5 months at Berkeley in a hippy-like fashion. Thank you to my parents-in-law, and all the family of my wife. Thank you, Sylvie, Philippe, Viviane, and Didier, for letting me work in your house by the Omaha beach, where Americans saved the French freedom on June, 6th, 1945.

Finally, thanks to the reader, that went already that far in the dissertation!

Chapter 1

Introduction

"Savoir, penser, rêver. Tout est là."

Victor Hugo

1.1 Motivations, contribution, and outline

This thesis aims to understand the impact of information-aware routing behaviors (more specifically, fastest-path routing) on the transportation system, in order to enable novel traffic management strategies.

Chapter 2 offers an overview of the **fundamental concepts of traffic engineering** to grasp the dissertation content, while section 1.2 presents the necessary mathematical definitions. The transportation demand is spread to the transportation network based on the people's origin, destination, departure time, and mode of travel. Most of the vehicles using the road network choose their route to reach their destination. Some of them follow direction signs, and some of them have access to real-time information to uncover the route that minimizes their travel time. Assuming the travel demand is similar every day, all the vehicles might identify the route that minimizes their experienced travel time. In this case, no vehicles have incentives to change their route, and the state of traffic is in an equilibrium (also called Nash equilibrium).

Chapter 3 demonstrates that **information-aware routing behaviors steer road traffic to a Nash equilibrium**. The chapter introduces the **restricted path choice model** that depicts the impact of an increase in information-aware routing behaviors by dividing vehicles between the one that have access to the information (assumed to minimize their experienced travel time) and the others (assumed to follow the direction signs). Unintuitively, the Braess paradox establishes that information-aware routing behaviors might not improve the average delay in the network. Beside possible decrease in overall network metrics, chap-

ter 4 reveals that information-aware routing behaviors induce cut-traffic traffic on some local roads that cannot sustain the resulting traffic flow.

Because cities do not have efficient mitigation techniques against the negative externalities of cut-through traffic, chapter 5 investigates cut-through traffic in Fremont, CA using a microsimulation of the Mission San Jose road traffic that reproduces the movement of each vehicle within the network. As the creation, calibration, and validation of a traffic microsimulation is challenging, the chapter provides a blueprint for such a task, and the Fremont traffic microsimulation is made open-source.

Finally, chapter 6 introduces a game-theory dynamic-routing model to decrease the computation runtime of dynamic routing models. Mean-field games are used to solve the curse of dimensionality that occurs in dynamic games.

In this dissertation, we are mainly interested in short-term traffic management, and traffic operations. We will try to understand how routing behaviors might be managed to improve transportation efficiency given a traffic demand, assumed to be fixed. As a consequence, the thesis does not take into account mode shift (Pigou–Knight–Downs paradox [11]), departure time shift (rescheduling behavior [82]), or willingness to travel shift ("reduced demand" [235]) due to change in traffic implied by routing behavior changes.

1.2 Necessary background: definitions and notations

In this section, we present some elementary notions from mathematics, probabilities, traffic engineering, and game theory, to ensure that the reader has enough background to understand the dissertation. The reader can skip this section for now, and come back to it whenever some equations are not clear in the dissertation.

Primary mathematical notations

In chapter 3, in chapter 6, and in the few equations of chapter 5, we assume that the reader is familiar with the mathematical notions listed in [252].

Especially, in this dissertation:

- $\mathbb{N}$ is the set of natural numbers.
- $\mathbb{R}$ is the set of real numbers.
- $\mathbb{R}_+ = \{x \in \mathbb{R}, x \geq 0\}$ is the set of non-negative real numbers.
- $\mathbb{R}_{>0} = \{x \in \mathbb{R}, x > 0\}$ is the set of positive real numbers.
- For any natural number $n \in \mathbb{N}$, we denote $\mathbb{R}^n$ the set of the n-tuples of real numbers. This notation is also used for n-dimensional vector space over the field $\mathbb{R}$.
- For any finite set A, we denote its cardinal (i.e., number of elements) as $|A|$.

- For any set A and B, we denote A^B the set of functions which map each element of B to an element of A. If B is countable, A^B interprets as the set of sequences indexed by elements of B.
- For any finite set A of n elements, we consider, by abuse of notations, that $\mathbb{R}^A = \mathbb{R}^{|A|} = \mathbb{R}^n$.
- For any set A and B, we denote $A \times B$ the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$ (the Cartesian product).
- For any indexed family $(A_i)_{i \in I}$, we denote $\times_{i \in I} A_i$ the Cartesian product of the indexed family.
- For any subset A and B of a set E, we denote $A \cup B$ the union of A and B.
- For any indexed family $(A_i)_{i \in I}$ of subsets of a set E, we denote $\bigcup_{i \in I} A_i$ the union of the indexed family.
- For any set A and function $f \in \mathbb{R}^A$, we define $\min_{a \in A} f(a)$ as the minimum of the function f over A. If the minimum does not exist, the value of the min is set to infinity ∞. We define $\operatorname{argmin}_{a \in A} f(a)$ as the set of minimizers (possibly empty).

Notation 1.2.1 (Kronecker delta)**.** *For any predicate P, we denote* $\delta_P = \begin{cases} 1 \text{ if } P \\ 0 \text{ if } \neg P \end{cases}$.

In the dissertation, we might see tuples of real numbers as vectors, or matrices.

- For any vectors x and y, $x^\top y$ denotes their scalar product.
- For any positive natural numbers $p, n \in \mathbb{N}$, for any vector $x \in \mathbb{R}^n$, we denote $\|x\|_p$ the p-norm of the vector x.

Elementary probability theory definitions

Chapter 6 uses some concepts from the probability theory.

The set of probability measures on a space A is denoted $\mathcal{M}(A)$. Unless otherwise specified, random variables are denoted by capital letters. The expected value of a real-valued random variable X given the probability measure m is denoted $\mathbb{E}_m[X]$.

Fundamental traffic engineering definitions

In this dissertation, we are interested in modeling the movement of vehicles.

Definition 1.2.1 (Time)**.** *Time is represented as an interval $\mathcal{T} = [0, T]$ of $\mathbb{R}$.*

Definition 1.2.2 (Time buckets). *For convenience, time is continuously partition into time buckets $\theta \in \Theta$ in chapter 5.*

VThe vehicles travel across the road network on specific path and road sections. We formalize the notion of network and paths using graphs [100].

Definition 1.2.3 (Network, paths and demand). *The road network is defined as a finite strongly connected directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ where $\mathcal{V}$ is the set of vertices, and $\mathcal{L} \subset \mathcal{V} \times \mathcal{V}$ the set of links (links). For each origin $o \in \mathcal{V}$ and destination $d \in \mathcal{V}$, we denote $\mathcal{P}_{od}$ the set of feasible paths without cycles from o to d. We define the set of paths as $\mathcal{P} = \bigcup_{o,d \in \mathcal{V}} \mathcal{P}_{od}$.*

Notation 1.2.2 (Links, vertices, and paths). *Notations l and $\tilde{l}$ are used for links. Vertices are represented by o and d. Paths are represented by p and $\tilde{p}$.*

Across the dissertation, we assume that the number of vehicle $N_0 \in \mathbb{N}$ using the network, and their corresponding origin, destination, and departure time are known.

Definition 1.2.4 (Traffic demand). *The traffic demand consists of N_0-tuples of origin, destination, and departure time $(o, d, t) \in \mathcal{V} \times \mathcal{V} \times \mathcal{T}$.*

Traffic micro-models model each individual vehicle and their dynamics (as done in chapter 5). Solving analytically, or even computationally, a mathematical model with many vehicles that are modeled individually might be challenging. To decrease the complexity, vehicles can be aggregated into flow, using notions from fluid dynamics.

Remark 1.2.1 (Micro to macro: mean-field theory, non-atomic games, and mean-field games). *Aggregating a high-dimension of individual interacting agents (here vehicles that impact each others through their location in the road network) into macroscopic dynamics is known as non-atomic games in game theory [204]. Aggregating a high-dimension of individual stochastic dynamics into deterministic macroscopic dynamics is known as mean-field theory in physics [22]. Connecting both the game theory and the statistical physics community, non-atomic dynamic games with stochastic noise are referred to as mean-field games [139, 118]. By extension, in this dissertation, we coined any non-atomic dynamic game with the name mean-field game. In chapter 6,to improve the realism of the routing game models, we will define a dynamic routing game that we will solve with a mean-field game approximation.*

Traffic macro-models use traffic flow to model vehicles' dynamics. More specifically, vehicles are aggregated into path flows and link flows.

Definition 1.2.5 (Link flows). *For each link $l \in \mathcal{L}$ and each time bucket $\theta \in \Theta$, we define link flow $f_{l,\theta} \in \mathbb{R}_+$ as the number of vehicles entering or exiting the link l during θ.*

We denote $f_l = \sum_{\theta \in \Theta} f_{l,\theta}$ as the number of vehicles entering or exiting the link l during $\mathcal{T}$. We note the link flow vector $\mathbf{f} = (f_l)_{l \in \mathcal{L}}$.

Definition 1.2.6 (Path flows). *For each path $p \in \mathcal{P}$, we define the flow $h_p \in \mathbb{R}_+$ using the path p (path flows) as the number of vehicles traveling along the path during $\mathcal{T}$. We note the path flow vector $\mathbf{h} = (h_p)_{p \in \mathcal{P}}$.*

In 1964, the Bureau of Public Roads described in the *Traffic assignment manual for application with a large, high speed computer* [239] that the average travel time on a highway section over one hour can be derived from the average flow on the section, together with the section's capacity and its free flow travel time. Therefore, in many models (especially static macroscopic models), it is assumed that link travel time can be derived from link flow through travel time functions [181, 168, 208].

Definition 1.2.7 (Travel time function). *For each link $l \in \mathcal{L}$, we define the link travel time function $t_l \in \mathbb{R}_+^{\mathbb{R}_+}$, which gives the link travel time given the number of vehicles on the link: $t_l : x \to t_l(x)$. We assume t_l to be continuous. We define the travel time vector $\mathbf{t}(x) = (t_l(x))_{l \in \mathcal{L}}$.*

By abuse of notations, we will use both $\mathbf{t}$ as the mapping between instantaneous volume and link traversal time for vehicles entering the link in chapter 6, and as the mapping between link flow f_l and average link travel time over $\mathcal{T}$ in chapter 3.

Remark 1.2.2 (Travel time function). *During the remainder of this dissertation, travel time functions are assumed to be increasing as functions of the corresponding link volumes or link flows. More vehicles on a link increase the travel time of this specific link. The dissertation assumes that vehicles want to minimize their travel time. However, the travel time can be interpreted as any type of cost (for example, it can be interpreted as gas emission [127]).*

Definition 1.2.8 (Static traffic demand). *For convenience, we aggregate the traffic demand (definition 1.2.4) across time $\mathcal{T}$ into a static traffic demand $\boldsymbol{d} \in \mathbb{R}_+^{\mathcal{V} \times \mathcal{V}}$. For any origin $o \in \mathcal{V}$ and destination $d \in \mathcal{V}$, d_{od} denotes the number of vehicles traveling from the origin to the destination during $\mathcal{T}$.*

Definition 1.2.9 (Transportation analysis zones and centroids). *For convenience, vertices are partitioned across time into connected transportation analysis zones (TAZs) $c \in \mathcal{C}$, also known as centroids.*

Definition 1.2.10 (Timed origin-destination matrices). *We denote $d(o, d, \theta)$ the aggregation (i.e., sum) of the traffic demand across time $\theta \in \Theta$ and space $o, d \in \mathcal{C}$.*

By abuse of notations, depending on the context, o and d might denote vertices in chapter 3, links in chapter 6, or centroids in chapter 5.

Elementary game theory definitions

In chapter 3 and chapter 6, we model vehicles as players that choose their route such that they minimize their travel time using game theory [173]. The reader can refer to [165] and [24] for comprehensive textbooks on game theory.

In a game, it is assumed that players (e.g., vehicles) would like to play a strategy (e.g., choose a route) such that they minimize their costs (e.g., travel times).

Notation 1.2.3 (Player). *We denote the set of players $\mathcal{N} = \{1, 2, \dots, N\}$, where $N \in \mathbb{N}$ is the number of players. Players are represented by i and $j \in \mathcal{N}$.*

Notation 1.2.4 (Strategy). *Each player $i \in \mathcal{N}$ can play a strategy $\boldsymbol{\pi}^i$ from a set of possible strategies $\boldsymbol{\Pi}^i$.*

The tuple of each player's strategy is called a strategy profile $\underline{\boldsymbol{\pi}} = (\boldsymbol{\pi}^i)_i$. The set of strategy profiles is denoted $\underline{\boldsymbol{\Pi}} = \times_{i \in \mathcal{N}} \boldsymbol{\Pi}^i$.

For the sake of convenience, we might denote $\underline{\boldsymbol{\pi}} = (\boldsymbol{\pi}^i, \underline{\boldsymbol{\pi}}^{-i})$ where:

$$\underline{\boldsymbol{\pi}}^{-i} = (\boldsymbol{\pi}^1, \boldsymbol{\pi}^2, \dots, \boldsymbol{\pi}^{i-1}, \boldsymbol{\pi}^{i+1}, \dots, \boldsymbol{\pi}^N)$$

Because the cost a player experienced might depend on the other players' strategies, it is hard for a player to know which strategy to play without knowing the strategies of the other players. In traffic, one vehicle might not know which route will minimize its travel time without knowing how traffic will evolve in the network.

Notation 1.2.5 (Cost function). *For a player $i \in \mathcal{N}$, its cost is denoted J^i. The cost of a player is a function of the strategy profile: $J^i \in \mathbb{R}^{\underline{\boldsymbol{\Pi}}}$.*

In non-cooperative game theory, players are assumed to take the strategy that are estimated to minimize their experienced cost. In traffic engineering, one can model that the vehicles will take the routes that are predicted to be the fastest ones.

After the facts, a player might realize that their strategy did not minimize their cost. In this case, the player will have counterfactual regrets: they should have chosen another strategy.

Definition 1.2.11 (Counterfactual regret [55]). *The counterfactual regret ρ^i of a player $i \in \mathcal{N}$ is the difference between their cost given a strategy profile $\underline{\boldsymbol{\pi}}$ and the optimal cost they could have by unilaterally changing their strategy:*

$$\rho^i(\underline{\boldsymbol{\pi}}) = J^i(\underline{\boldsymbol{\pi}}) - \min_{\boldsymbol{\pi} \in \boldsymbol{\Pi}^i} J^i(\boldsymbol{\pi}, \underline{\boldsymbol{\pi}}^{-i})$$

We assume here that J^i and $\boldsymbol{\Pi}^i$ are such that the minimum exists.

Remark 1.2.3 (Counterfactual regret, marginal regret, deviation incentives). *In the context of repeated games, counterfactual regret is sometimes referred to as marginal regret [55], or hindsight regret. To distinguish the concept of cumulative regret and marginal regret, the counterfactual regret (or marginal regret) is sometimes also referred to as deviation incentive [24]: if the game would happen again, then the player that did not play the optimal strategy to minimize their cost will have some incentive to deviate from their previous strategy. In the remaining of the dissertation, we will use the term* ***deviation incentive****.*

If the game is played again and again (e.g., all the vehicles have the same trips every day), then eventually the players will adapt their strategies to each other, until no one has any incentive to change their strategy (e.g., all the vehicles travel along the same route for their trip every day).

If no player has any incentive to change their strategies, then the game reaches an equilibrium: for every instance of the game, each player will play the same strategy as in the previous game instance. This equilibrium state is referred to as a Nash equilibrium.

Definition 1.2.12 (Nash equilibrium [24]). *A strategy profile $\underline{\boldsymbol{\pi}}$ is a Nash equilibrium, if, and only if, no player has a deviation incentive:*

$$\forall i \in \mathcal{N},\ \rho^i(\underline{\boldsymbol{\pi}}) = 0$$

Because players might not reach a state where no one has any incentive to deviate, the notion of approximate Nash equilibrium can be defined. More specifically, an ϵ-Nash equilibrium is such that no one has a deviation incentive higher than ϵ.

Definition 1.2.13 (ϵ-Nash equilibrium [24]). *For any $\epsilon \in \mathbb{R}$, a strategy profile $\underline{\boldsymbol{\pi}}$ is a ϵ-Nash equilibrium, if, and only if, no player has a deviation incentive higher than ϵ:*

$$\forall i \in \mathcal{N},\ \rho^i(\underline{\boldsymbol{\pi}}) \leq \epsilon$$

As an ϵ-Nash equilibrium considers the worst case for one player, we introduce the notion of average ϵ-user equilibrium that is better suited for non-atomic games (i.e., games with a continuum of players, as modeled in both chapter 3 and chapter 6):

Definition 1.2.14 (Average ϵ-Nash equilibrium). *For any $\epsilon \in \mathbb{R}$, a strategy profile $\underline{\boldsymbol{\pi}}$ is an average ϵ-Nash equilibrium, if and only if, the average deviation incentive is smaller than ϵ:*

$$\frac{1}{N} \sum_{i \in \mathcal{N}} \rho^i(\underline{\boldsymbol{\pi}}) \leq \epsilon$$

Remark 1.2.4 (Average ϵ-Nash equilibrium). *In an average ϵ-Nash equilibrium, an average "player" can expect to save ϵ by changing unilaterally their strategy. As highlighted in [35], this notion is similar to the definition of (ϵ, δ)-Nash equilibria in [88].*

Notation 1.2.6 (Average deviation incentive). *Given a strategy profile $\underline{\boldsymbol{\pi}} \in \underline{\boldsymbol{\Pi}}$, the average deviation incentive is denoted $\bar{\rho}(\underline{\boldsymbol{\pi}}) = \frac{1}{N} \sum_{i \in \mathcal{N}} \rho^i(\underline{\boldsymbol{\pi}})$.*

Remark 1.2.5 (Average deviation incentive). *Depending on the context, average deviation incentive might be referred to as average marginal regret [55, 47], average counterfactual regret [258], exploitability [149], hindsight routing, or relative gap to the Nash equilibrium (in traffic engineering, it is called relative gap to the user equilibrium [231]).*

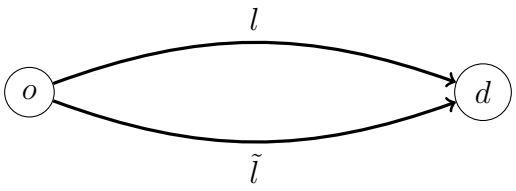


Figure 1.1: Benchmark network example to illustrate the average deviation incentive

Counterintuitively, a Nash equilibrium might not minimize the average cost of the players. This is known as the price of anarchy [173]. The strategy profile that minimizes the average cost of the players is called the social optimum.

Definition 1.2.15 (Social optimum [173])**.** *A strategy profile* $\underline{\boldsymbol{\pi}}$ *is a social optimum, if, and only if, the average player cost is minimized:*

$$\forall \underline{\boldsymbol{\pi}}' \in \underline{\boldsymbol{\Pi}}, \sum_{i \in \mathcal{N}} J^i(\underline{\boldsymbol{\pi}}) \leq \sum_{i \in \mathcal{N}} J^i(\underline{\boldsymbol{\pi}}')$$

Both chapter 3 and chapter 6 extends the classical definition of non-cooperative game theory with N players using non-atomic game (in chapter 3), and dynamic games and mean-field games (in chapter 6). Therefore, some notations might be overridden in the two chapters for the sake of convenience.

Traffic scenario considered in the dissertation

During the dissertation, we will consider several traffic scenarios to illustrate our statements. Two of them, the Pigou network and the Braess network, will be used across chapter 3 and chapter 6.

Pigou network

The Pigou network, shown in fig. 1.1, is formed by two parallel links l and $\tilde{l}$ joining the origin o to the destination d.

In this network, each link is a path from the origin to the destination and $\mathcal{P} = \{l, \tilde{l}\}$. The demand d_{od} and the cost functions t_l and $t_{\tilde{l}}$ will be defined when used in the dissertation. The Pigou network is derived from Arthur Pigou's work published in [184].

Braess network

The Braess network, shwon in fig. 1.2, is shaped by four nodes (A, B, C, and D). The network is derived from [40]. There are three paths (ABD, ACD and $ABCD$) from A to

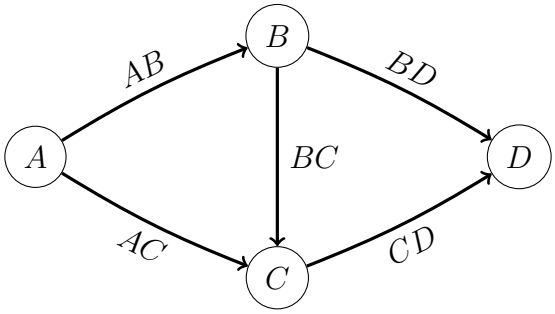


Figure 1.2: Braess network considered.

B. The demand $\boldsymbol{d}$ and the cost functions t_{AB}, t_{AC}, t_{BC}, t_{BD} and t_{CD} will be defined when used in the dissertation.

Alongside with these two networks, the traffic in Sioux Falls, ND, Fremont, CA, and Los Angeles, CA will be modeled, and the corresponding network will be introduced when used.

Part I

Traffic control in the age of information technology

Chapter 2

Traffic engineering perspective on routing behaviors

This chapter introduces key elementary transportation concepts to provide the reader with enough background on the topic of this dissertation: **understanding the impact of information-aware routing due to navigational apps on road traffic**. In section 2.1, transportation supply and transportation demand are explained. Then, transportation planning and transportation management are described in section 2.2. Finally, the topic of the dissertation is summarized in section 2.3.

2.1 Transportation demand and supply

Transportation systems are complex civil systems that facilitate the movement of people and goods throughout the world. The transportation system has a tremendous impact on the daily lives and well-being of both the users and nonusers of the system. Transportation system **users** benefit from accessing the transportation system as long as they can travel to and from desired locations in a timely, cost-effective, and convenient manner. In 2019, Texas A&M Transportation Institute estimates that the U.S. loses $166 billion per year due to the impact of congestion on fuel usage and productivity loss [205]. The average auto commuter spends 54 hours in congestion and wastes 21 gallons (79.5 liters) of fuel every year due to congestion for $1,010 in wasted time and fuel. If transportation engineers succeed in decreasing traffic congestion by even 1% in the U.S., then it would result in over $1.6 billion in annual savings.

Transportation engineering is the application of technology and scientific principles to the planning, functional design, operation, and management of facilities for any mode of transportation to provide for the safe, efficient, rapid, comfortable, convenient, economical, and ecological movement of people and goods [65]. At every level of granularity, from local, intra-neighborhood to international travel, transportation engineering can be appreciated through an economic lens as a market consisting of a **supply** (the road network) and a **de-**

mand (goods or travelers). Planners and engineers working for cities, public institutions, or consulting firms and transportation operators can adapt the supply to the transportation demand as theorized and recommended by Adam Smith in 1776 in *The Wealth of Nations* [209]. On the other hand, they can also attempt to influence the transportation demand. In 1798, Thomas Malthus predicted a "natural selection" between humans in case of higher demand than the supply available in *An Essay on the Principle of Population* [155]. In 1930, the adaptation of the demand to the available transportation supply was reported by the City of St. Louis [64]. In transportation, this phenomenon is called **induced demand**: if the supply increases, then the demand will increase because some people – constituting the **latent demand** – were unable to satisfy their demand before due to the higher cost of the supply.

An efficient transportation system is designed such that the transportation network **capacity** is equivalent to the travel demand. If the travel demand exceeds the network capacity, then the network will be critically congested leading to increased travel times, delays, increased gas emissions, increased transportation costs, etc. On the contrary, if the network is over-designed for the demand, then it is economically inefficient [222]. While transportation researchers provide tools to model, predict, estimate and analyze the impact of **transportation policies** on the resulting system, the implementation of any policy is political and should be seen like that – whether the policy supports *laissez-faire* as J.B. Colbert advocated in the 1680s, or if it supports that the supply should not pressure the demand and that the demand should control or manage the supply as advocated by Karl Marx in 1848 in *The Communist Manifesto* [159] and *Capital, A Critique of Political Economy* in 1867 and 1885 [158].

Influenced by Alfred Marshall – who defined mathematically the concept of supply and demand in 1890 in *Principle of Economics* [157] – Arthur Pigou introduced the notion of transportation supply and transportation demand in 1920 in *The Economics of Welfare* [184].

Transportation supply

The transportation supply results from the available **transportation infrastructures**, **facilities**, and their **management**. Transportation infrastructures and facilities include roads, train stations, airports, bus stations, public transit, and traffic control plans among others. Within a **transportation network** – formally, a geographic area where almost all the people using the transportation facilities are from and stay in – several public or private organizations might control or manage some part of the transportation supply. Public institutions are generally in charge of planning the transportation infrastructures and might delegate their design and their construction. Public and private transportation operators are in charge of the control and management of the transportation facilities.

This dissertation focuses mainly on the road transportation network. In the road transportation network, the roads are divided by type. In the U.S., traffic engineers distinguish three categories of roads, that are divided into seven subcategories [7, 232]: arterial roads – including Interstate, other freeways and expressways, other principal arterials, and minor

	Arterials				Collectors		Local roads
	Interstate	Other Freeways & Expressway	Other Principal Arterial	Minor Arterial	Major collector	Minor collector	
AADT (rural)	12,000 - 34,000	4,000 - 18,500	2,000 - 8,5000	1,500 - 6,000	300 - 2,600	150 - 1,110	15 - 400
AADT (urban)	35,000 - 129,000	13,000 - 55,000	7,000 - 27,000	3,000 - 14,000	1,100 - 6,300	1,100 - 6,300	80 - 700
VMT (rural) Extent for All States	20% - 38%	0% - 8%	14% - 30%	11% - 20%	12% - 23%	2% - 9%	8% - 23%
VMT (urban) Extent for All States	17% - 31%	0% - 17%	16% - 31%	14% -25%	5% -13%	5% - 13%	6% -25%

Table 2.1: Road types and their expected annual average daily traffic (AADT) and percentage of vehicle miles traveled (VMT) in the U.S. [232]. The expected AADT and VMT percentages depend on whether the road is in a rural or urban network. A rural network means that less than 75% of the network population is located in urban centers.

arterials –, collectors – including major collectors and minor collectors – and local roads. For each type of road, two notions of capacity are defined:

Definition 2.1.1 (Daily urban planning road capacity)**.** *The expected annual average daily traffic (AADT) [7, 8] is the daily vehicle count that a road is supposed to be able to sustain before facing traffic externalities that might lead to a significant decrease in residents' quality of life, high maintenance costs, and safety concerns. We refer to the expected AADT as the* ***daily urban planning capacity*** *in this dissertation. These capacities are reported in table 2.1.*

Definition 2.1.2 (Hourly road traffic capacity)**.** *The* ***critical density*** *[168] is the density of vehicles that a road can receive without being congested. Above this density, adding new vehicles to the road will lead to traffic congestion. We refer to the critical density as the* ***hourly road traffic capacity*** *in this dissertation. These capacities can be measured using the fundamental diagram of traffic flow [208] or derived based on the road geometry and other features [168]. They are usually used to define the link travel time functions (definition 1.2.7).*

Paradoxically, due to latent demand, creating new infrastructures might not decrease congestion in the network [64]. As an illustration, in Paris, the average daily time in transportation is about 1h30min per person [120]. People with higher accessibility to transportation facilities (i.e., shorter commute times) make more trips per day (i.e., do more leisure

trips and errands). Increasing the transportation supply in Paris might not improve the efficiency of the network (measured as the average delay time in congestion) but only increase the total number of trips in the network (or total vehicle-miles traveled). On the opposite, some road-rationing techniques, and other techniques (like parking-price management), are used by the city of Paris to decrease congestion by decreasing the modal share of the car [74]. Through this approach, the city of Paris does not only manage the supply, but implicitly manages the transportation demand.

Transportation demand

To implicitly manage the transportation demand, urban planners, transportation infrastructure designers, and transportation engineers are interested in understanding how it is influenced by the transportation supply. Understanding traveler choices might provide key elements to the question [222]. When one person travels, they will sequentially make 5 distinct choices that are expressed by the trip chain (fig. 2.1) [235]:

- Destination-choice; first, the traveler chooses their destination.
- Time-of-day choice; then, the traveler decides when they want to arrive or depart.
- Mode-choice; then, the traveler selects in which mode (e.g., public transportation, private car, carpool, bike, walk, etc.) they want to travel.
- Route-choice; after picking their destination, their departure time, and their private car to travel, the driver decides which route they will choose to arrive at their destination.
- Lane/facility use/choice; finally, once on the road and on a specific route, the driver chooses which lane they can travel on.

Using this interpretation of travelers' choices, the transportation demand can be divided into three different types of demand [235]:

Definition 2.1.3 (Travel demand)**.** *First, the* ***travel demand*** *is the total number of travelers that use the network from any origin to any destination, at any departure time (or arrival time).*

Definition 2.1.4 (Traffic demand)**.** *Second, the* ***road traffic demand*** *(definition 1.2.4) is the travel demand that travels using the road network. It is the total number of vehicles that use the network from any origin to any destination, at any departure time (or arrival time).*

Definition 2.1.5 (Facility demand)**.** *Third, the* ***road facility demand*** *is the number of vehicles on any route, road segment, or lane at any time of the day.*

Due to the difficulties of changing people's habits, influencing the traffic demand (destination choice or time-of-day choice) is mainly accomplished with long-term planning (i.e.,

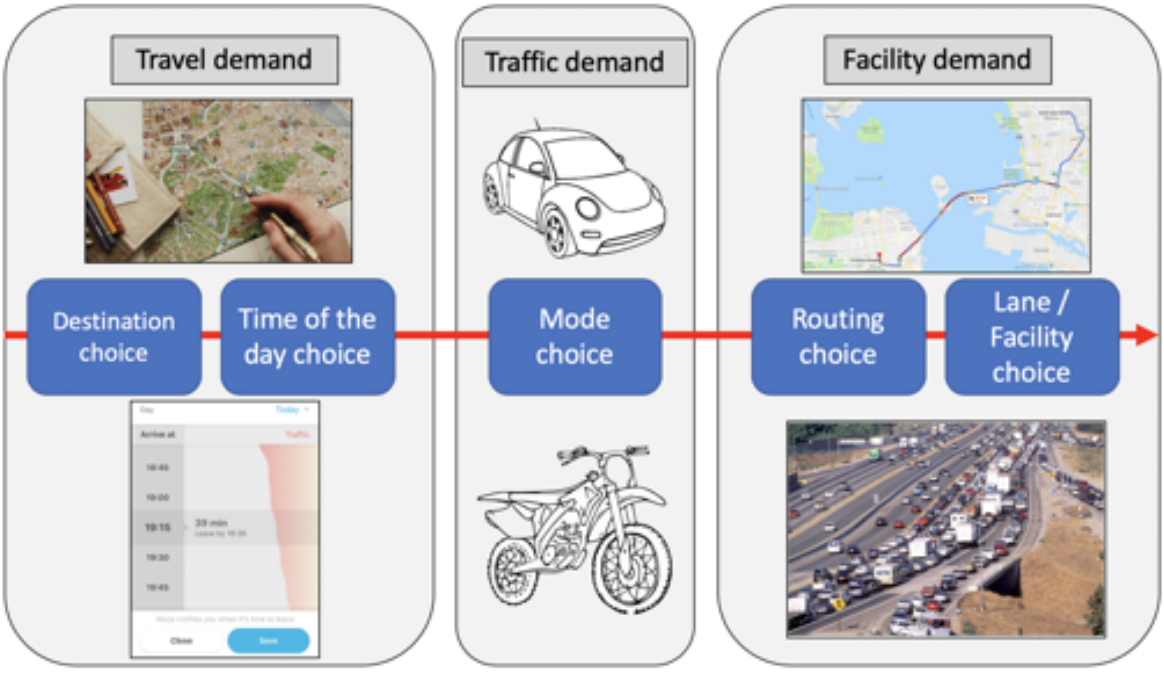


Figure 2.1: The trip chain. First, travelers choose their destination and time of departure. This creates the travel demand of the network. Then, they select their mode. The ones traveling by car establish the road traffic demand. Then, they decide their route and their lane, contributing to the facility demand of the route or the lane. The figure is adapted from [235, Figure 4].

transportation planning). Accordingly, short-term facility management (i.e., **transportation management** or **traffic operations**) mostly impacts the facility demand (facility choice). Both the mode choice and the route choice can be impacted through long-term planning or short-term traffic management.

2.2 Transportation planning and management

When an urban planner or a traffic engineer would like to improve the transportation network efficiency, they first need to define the **key performance indicators** (KPI) that they want to improve. Then, they should define the time frame they want to impact (i.e., long-term planning or short-term management). Finally, they should understand the tools they can use and predict their impact on the KPIs they want to optimize.

Transportation key performance indicators (KPIs)

We suggest dividing the KPIs of a transportation network into **measures of transportation effectiveness** [143, 235] and **measures of city efficiency**. In short-term planning and facilities management, the measures of effectiveness are monitored and forecasted by traffic engineers. In long-term planning, they are modeled and/or estimated to predict the impact of development projects on traffic. They are well-defined [143, 235, 36] and include:

- Travel metrics: including vehicle miles traveled (VMT), number of trips in the network, and average daily traffic on specific road sections.
- Travel time metrics: including vehicle hours traveled (VHT), mean waiting time at bus stops, mean travel time reliability [90], average marginal regrets (also known as the relative gap, or mean counterfactual regrets), mean bus delay, and travel times between predefined locations.
- Congestion metrics: including mean delay per vehicle-mile, mean travel time per vehicle-mile, mean time in queue, mean stopped time, mean speed in the network, and mean bus travel time.
- Efficiency metrics: including mean link vehicle-occupancy, mean bus-occupancy, mean vehicle-occupancy, and total time when bus capacity was exceeded.
- Sustainability metrics: including particles matters (PM) 2.5, PM 5, PM 10, O_3, Carbon, NO_2, SO_2, CO, CO_2, GHG, and volatile organic compounds emissions.
- Resiliency metrics: number of traffic signal phase failures, and average traffic incident management response time.

To the knowledge of the author – who met with traffic engineers of six different cities in the U.S. and France – the measures of city efficiency are more fuzzy because many cities react to issues that are voiced by residents and business owners, without quantitatively defining, measuring, or predicting the city efficiency. Some city engineers are skeptical that the city's efficiency can be described and measured only with indicators. Therefore, a lot of city efficiency ratings are subjective, based on intuitions, and depend on the political circumstances in the city. As stated before, we advise understanding any transportation policy as political, even though it might have been derived based on data or scientific models. We suggest a subclassification of possible city efficiency metrics into financial metrics, economic metrics, social metrics, environmental metrics, and operational metrics.

Financial metrics might include road maintenance cost and infrastructure damage, public transit subsidy cost, operational cost, operational revenue, investment, and institution debt.

Economic metrics might include city productivity, job accessibility [146], business accessibility [146], travel cost, fuel consumption and fuel cost, and battery consumption.

Social metrics might include safety (number of fatalities, number of accidents), air quality (volatile organic compound density), and health (number of death related to transportation

pollution), quality of life (average daily distribution of noise for specific segments), or transportation equity (including underserved community accessibility).

Environmental metrics can include total greenhouse gas emissions due to the transportation system (either instantaneous emissions, or life-cycle emissions). Operational metrics can include all the measures of transportation effectiveness and resiliency metrics (like average traffic incident management response).

With clear objectives in mind, a time frame should be defined to understand the set of tools that can be used to increase the system's efficiency. Long-term planning is referred to as transportation planning, while short-term planning and management are referred to as transportation management or operations.

Transportation planning

Transportation planning defines future policies, goals, investments, and spatial planning designs to prepare for future needs to move people and goods to destinations. Planning use cases include:

- Estimating the needs of transportation infrastructure when building a new mall.
- Design parking lots (for buses, cars, and bikes) when building new infrastructures (train stations, housing units, etc.).
- Changing a car-only road to a bike-only road to make the transportation network more sustainable.
- Implementing traffic-calming measures to decrease traffic on some streets.

Transportation planning is inherently related and interconnected to urban planning. As shown in fig. 2.2, land use determines the need for transport, and transport, in return, further determines spatial development.

First, the land use outlines the transportation infrastructure supplies. Second, it induces the travel demand, because the distribution of human activities in space requires people to move between the activities' locations. Through the people's decisions to make a trip towards a specific destination, at a specific time, with a given mode of transportation, the transportation demand induces the traffic and facility demands (fig. 2.1). Traffic features and individual dynamics, which defines the transportation system efficiency, follow from the traffic and facility demands being allocated to the transportation supply. The transportation efficiency results characterize land accessibility and attractiveness. Depending on the accessibility and attractiveness of each area, investors, developers, and planners decide on choosing appropriate locations for different human activities. Then, constructions – including building transportation infrastructure supplies – take place in the area. The distribution of land usage over the urban area determines the locations of human activities. After construction, people could update their decisions on where to live, to work, to do errands, to

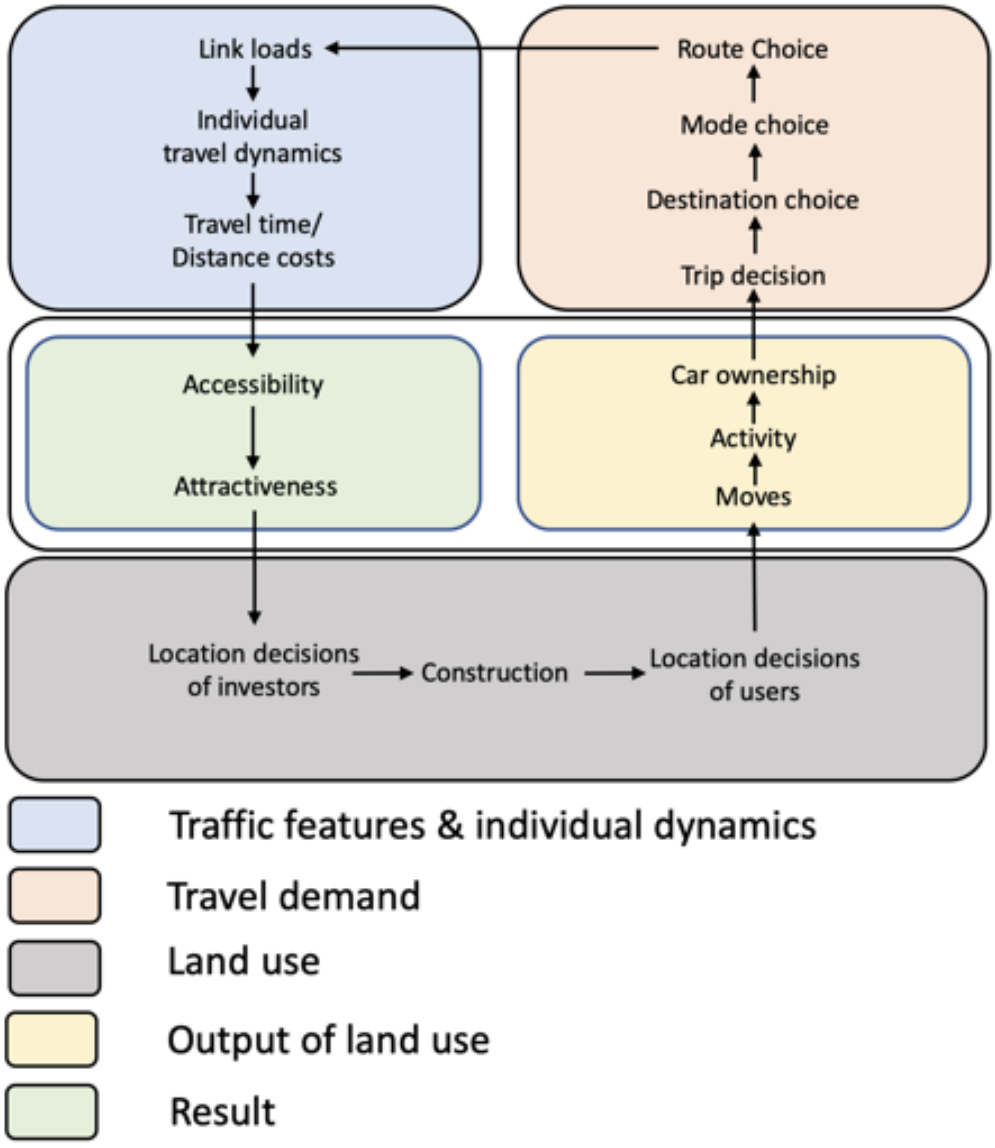


Figure 2.2: The land use cycle [249]. Land use determines the need for transport, and transport, in return, further determines spatial development. The figure is adapted from [249, Figure 1]

Figure 2.3: Active traffic and demand management (ATDM) can be divided into a four-step process to optimize the traffic system. First, the traffic state is estimated through traffic monitoring. Then, the performance of the system is assessed. Finally, strategies can be implemented. Before being implemented, their impact on the traffic state needs to be evaluated. Eventually, the best ones are implemented, and the traffic state is evaluated again. The figure is reproduced from [235, Figure 1].

get entertained, or to get an education. Depending on their needs for transportation, they can also decide to buy or to sell their car.

When the transportation supply resources become limited, the engineer can manage the transportation facilities to increase the network efficiency without creating costly new infrastructures (e.g., roads) or facilities (e.g., public transportation). For example, this might include updating the traffic signal timing plans [240].

Traffic management and operations

Traffic management (or traffic operations) is the part of transportation engineering that does not consider the planning of the transportation infrastructures [229]. Transportation operations use cases include:

- Mass evacuation of an area due to hazard.
- Incident management when a car crash obstructs a highway lane.
- Changing traffic signal plan operations.

Traffic management ([235]) can be divided into a four-step process to optimize the traffic system (see fig. 2.3) First, the traffic state is estimated through **traffic monitoring**. Then, the performance of the system is assessed using **transportation performance measures**. Finally, **traffic management strategies** can be implemented. Before being implemented, their impact on the traffic state needs to be evaluated. Eventually, a trade-off is done and the best ones are implemented, and the traffic state is evaluated again.

Traffic monitoring

To evaluate the state of traffic, field traffic data measurements are made with traffic detectors [129] (see fig. 2.4). Then the data are combined and analyzed inside a system that estimates and displays the state of traffic.

Transportation performance measures

To assess the performance of the transportation system, key performance indicators are defined and weighted together in a multi-objective function to optimize.

Traffic management strategies

Many engineering techniques can be used to improve the transportation-system performance without building new infrastructures. While the next subsection suggests a classification for any transportation engineering tools, the traffic management strategies can be subdivided into three types based on static supply change, dynamic supply change, and demand management:

- Road geometric strategies; without building new roads, traffic engineers might perform changes to the infrastructures to improve the transportation network. This might include markings, guard fences, road signs, speed bumps (humps or lumps), stop signs, lane restrictions, access eligibility (like high occupancy vehicle (HOV) and high-occupancy toll (HOT) lanes), and speed limits. To promote safe and livable streets, it might include complete street design (like in Los Angeles, CA [62]). More dynamic infrastructure changes might include dynamic lane assignment, dynamic lane restriction, temporal shoulder use, and managed lanes.

- Traffic control strategies [235]; more dynamically, traffic engineers can manage dynamic infrastructure such as traffic lights. Traffic control strategies include traffic signal control, lane-capacity control, variable speed limit, variable message sign (see fig. 2.6), ramp metering, queue warning, speed harmonization, incident management, and dynamic HOT pricing.

- Traffic demand management strategies [235]; traffic engineers can also implement strategies to control the traffic demand (road geometric strategies and traffic control strategies are mainly about controlling the transportation supply). Traffic demand

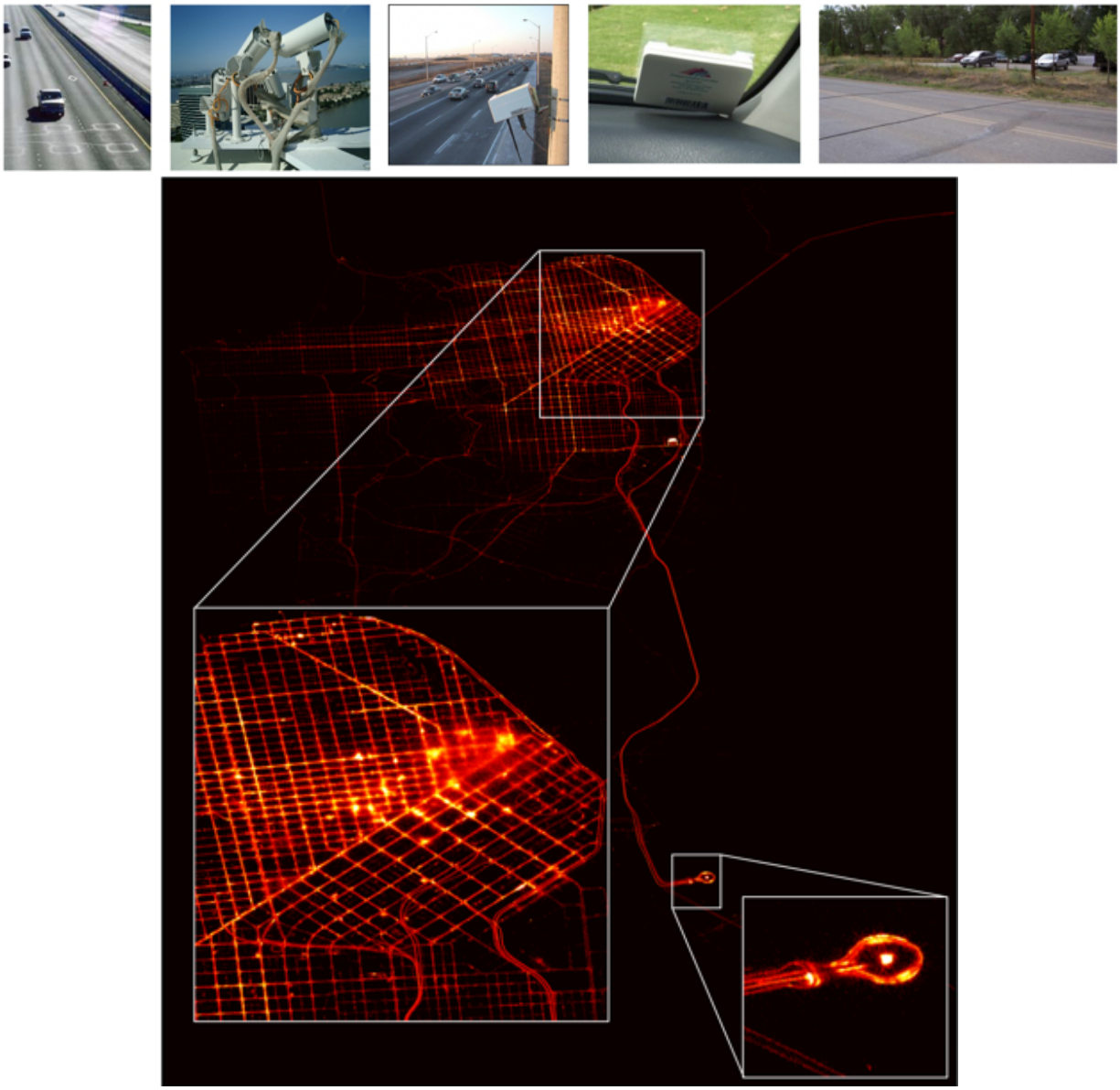


Figure 2.4: Sensing infrastructures and devices. On top: Classical sensing infrastructures [129] used in the pre-smartphone era (and still used as the backbone of traffic control by cities without access to mobile data in sufficient quantities. From right to left: loop detectors, CCTV cameras, radar, tolling RFID transponders used for traffic monitoring, and traffic counting tubes. At the bottom: Superposition of the GPS tracks of 500 vehicles sampled every 30 seconds throughout one day (yellow cab fleet of the city of San Francisco) circa 2009, for one day [180].

management includes cordon pricing, parking pricing, congestion pricing, road space rationing, carpool incentives, "work from home" programs, "alternative working hours" incentives, public transportation design, route guidance, and dynamic parking pricing. More dynamically, it might also include: dynamic fare reduction, dynamic HOV / managed lanes, dynamic pricing, dynamic ride-sharing, dynamic routing guidance, dynamic transit capacity assignment, on-demand transit (demand-responsive transport), predictive traveler information, and transfer connection protection.

Both the traffic control strategies and traffic demand management strategies can be:

- *Intelligent*, if they use advanced technologies like digital systems.
- *Active*, if they can change depending on the time of the day.
- *Adaptive*, if they can adapt to the traffic situation.

Implementation and trade-off

To improve the system performance, traffic management strategies should be implemented. Institutional actors can find the optimal strategies using control engineering. First, one needs to evaluate the current state of traffic with traffic monitoring. Second, one needs to assess the performance of the current state of traffic through performance analysis. Then, one needs to understand the impact of the tools that can be implemented on the traffic state using models or simulations. Once the impact of different strategies that can be implemented is known, then one should be implemented. This requires synergies between the different institutional actors that might have different objectives for the traffic state as they pay attention to different key performance indicators. Ideally, an objective function that uniquely quantifies the transportation-system performance should be defined between the different institutional actors. Then, the strategy that optimizes the objective function can be chosen.

Classification of the transportation planning and management tools

The different tools that urban planners and traffic engineers own to influence traveler choices can be classified based on their impact on the three demands (definition 2.1.3, definition 2.1.4, and definition 2.1.5).

The transportation planning tools can be divided between building or updating infrastructures and giving incentives to travelers [216]. The traffic management strategies can be clustered between the management of the road facilities, the management of the vehicle facilities, and the pricing and incentives policies [216]. Table 2.2 summarizes the different tools based on their type and their impact on the trip chain (fig. 2.1). As an illustration, fig. 2.5 presents all the traffic-calming tools that Fremont, CA has tried to mitigate the cut-through traffic in the Mission San Jose district.

Figure 2.5: Left: Traffic-calming strategies implemented by the city of Fremont to decrease traffic on local roads due to commuters' cut-through. Right: An illustration of the Fremont turn and access restriction during evening peak hours.

	Destination choice	Time of the day choice	Mode choice	Route choice	Lane/facility use/choice
Urban infrastructure	Transit oriented development Zoning		Complete street HOV lanes	Direction signs	Speed bump Signs and marking HOV lanes
Incentives	Work from home subsidy	Alternative working hours subsidy	Carpool Public transportation		
Road facilities management			Road space rationing Traffic signal priority	Turn restriction Access restriction Temporary direction signs Variable message signs Traffic police	Access eligibility Intersection management Ramp metering Reversible lanes Variable speed limits Lane specific signal
Vehicle facilities management	Navigational apps	Navigational apps Radio	Navigational apps Radio Ride-hailing regulations Demand responsive transport	Navigational apps Radio Fleet management	Navigational apps Radio
Incentives / pricing	Parking pricing Cordon pricing	Congestion pricing	Public transit subsidy Biking subsidy	HOT lanes Tolls Road user charge	HOT lanes Tolls
	Transportation planning			**Transportation management**	

Table 2.2: Classification of a non-exhaustive list of transportation planning tools and traffic management strategies based on their impact on the trip chain [235]. An effort of listing all the possible tools can be found in [216]. HOT = high-occupancy toll lane, HOV = high-occupancy vehicle lane.

Predicting the impact of planning tools and management strategies

To understand the impact of a city's candidate-transportation policy/strategy, transportation planners have three main options. First, they can perform case studies. For example, if the city of Fremont, CA would like to understand the impact of modifying the traffic signal timing plans on Mission Boulevard (a major street in the city) on cut-through traffic [225], the traffic engineers can attempt to extrapolate from what the city of Leonia did for Ford Lee Road [124] or what the city of Pleasanton did for Dublin Canyon Road [122]. If there are enough case studies, machine learning techniques can be used to make statistically significant predictions as to how a chosen policy or set of policies impacts a given state of traffic. Second, cities can use trial and error (often referred to as evidence-based practice or A/B experiments). The city of Fremont might try to change traffic signal plans, implement turn and access restrictions, activate ramp metering, and beyond [225], keeping only the policies that have the most desirable impact on the public. Third, a city can develop a digital twin (i.e., a model) of its road traffic network with which to try out and learn policies in a virtual

Figure 2.6: One of the first Variable Message Signs (VMS) on the New Jersey turnpike in the 1950s. Variable message signs were the first tools to manage route choice in real-time without the need of a traffic police agent, or temporary road signs. Nowadays, we envision that eventually variable message signs will be overridden by navigational apps or any other connected devices.

setting [160]. The city of Fremont might take this approach to avoid disturbing its citizens by frequently changing experimental traffic signal plans. Using digital twins is especially useful when tackling complex transportation challenges in which there is a lack of case studies, or it is infeasible to rapidly and affordably test potential solutions in the real world. Such cases include planning sustainable transportation systems where coordination between connected vehicles, transportation planners, and traffic managers is critical.

A blueprint to create, calibrate, and validate a large-scale traffic microsimulation is introduced in chapter 5.

2.3 Emerging routing behaviors in the 2010s decade

The evolution of the traffic information systems until 2020

To the author's knowledge, until World War II, traffic information related to routing had mostly been map information. In the 1950s, variable message signs (VMS) appeared [113], conveying some local information about traffic conditions (see fig. 2.6). The late 1960s saw the birth of electronic route guidance systems [193], in which mostly static routing information was provided to users (see fig. 2.7). In the 1975 the Highway Advisory Radio [80] was among the earliest services to provide live traffic information to its users. However, to the author's knowledge, for decades, only a few highways were instrumented by sensors.

Over the 2010 decade, thanks to the internet and crowdsourced GPS data, the information provided to drivers saw a boom. First, the connectivity between people tremendously increased due to the large expansion of the usage of the Internet allowed by the worldwide adoption of smartphones. In July 2019, 4.3 billion active internet users were counted [211]. In the U.S., in 2019, 95% of the population used the Internet monthly [210, slide 33]. Second,

Keeping Traffic in Hand

Tired of getting blind-sided by traffic jams?

A new, hand-held device from Berkeley-based Way To Go Corp. allows drivers to get instant information about traffic conditions on highways they plan to use.

Just tap two points on the gadget's map — say, San Francisco and San Jose — and a synthesized voice warns of traffic jams, accidents and other problems along the way.

The information is gathered by Metro Traffic Control, a company that also sells traffic information to radio stations, and is delivered by paging signals to the device within minutes of any new developments.

Way To Go is available at Cellular One outlets and Whole Earth Access stores at a hefty price — $199 for the hardware, plus $15 per month payable to Way To Go for the service.

Figure 2.7: Left: Onboard navigation unit prototype built by Honda in 1981 for route guidance. Right: Newspaper commercial for *Way to Go* aftermarket device circa 1990, one of the early prototypes of navigational apps built by UC Berkeley.

this connectivity led to the production of an extensive quantity of traffic data, often referred to as big data. In the fall of 2018, it was estimated that every smartphone used on average 7.0 GB of data every month [210, slide 178]. This number is 10 times higher in 2018 than in 2012 [210, slide 178]. Finally, the data can be processed because of the increase in the computational power available to traffic information companies due to cloud computing. As a consequence, the decade had seen the emergence of new technologies like demand-responsive transportation [153], ride-hailing services [66], and GPS-enable navigational apps (fig. 2.7), that enables traffic managers and operators to dynamically impact mode-choice and route choice. As an example, in 2022, route-choice can be managed through fastest-path routing guidance [251], eco-routing guidance [13, 104], or also mass evacuation routing guidance [170, 186]. In April 2018, Statista estimated that Google Maps receives 154.4 million monthly users in the U.S. [213]. Eventually, all vehicles in the road network will be connected, and traffic managers will be able to considerably improve the efficiency of the transportation system through intelligent and adaptive routing suggestions.

Classical route-choice models

Route-choice and direction signs

To increase the transportation system efficiency through economies of scale [209], the road is designed such that the traffic is aggregated to high-category roads where the speed and the capacity are designed to be high [7]. Any trip should begin on a low road category (i.e., a local road) where both the speed and the capacity are low. Along the trip, direction signs are designed such that the category of the roads used by a vehicle following the signs progressively increases and then decreases until the vehicle eventually reaches its destination (see fig. 2.8). Therefore, historically, vehicles followed direction signs that were designed to control the route choice [7].

Within the routing models used by engineers to plan the road network, the **all-or-nothing route assignment model** [181] assumes that vehicles follow the fastest path between their origin and destination under free-flow conditions. This model represents that vehicles follow the direction signs, given that the direction signs provide the fastest path between any origin to any destination if there is no congestion on the road.

Wardrop equilibrium

When congestion arises in the network, the direction signs might not indicate the fastest path to reach a destination anymore. Therefore, a second routing model – the **static traffic assignment model** (STA) [181] – considers that vehicles adapt their route based on congestion. More specially, the model assumes that vehicles choose the routes that minimize their experienced travel time (accounting for congestion created by other vehicles). Assuming the travel demand is similar every day, once all the vehicles have identified the route that minimizes their experienced travel time, no vehicles have incentives to change their route: the state of traffic is in an equilibrium (definition 1.2.12).

Answering Pigou's notion of social cost [184], Frank Knight first described the notion of traffic equilibrium in 1924 in *Some fallacies in the interpretation of social cost* [130]. In 1950, John Nash developed conditions to prove the existence of non-cooperative game equilibrium, now called Nash equilibrium [167] (definition 1.2.12). In 1952, the concept of Nash equilibrium was applied to the road transportation network by John Wardrop [247]. The traffic equilibrium is now called Wardrop equilibrium, user equilibrium, or Nash equilibrium. The Wardrop equilibrium is defined as a state where different vehicles traveling on several paths between the same origin and destination cannot have better travel time by being the only ones to change their path. In other words, the Wardrop equilibrium is such that vehicles have no deviation incentives (definition 1.2.11) to change their path. In the Wardrop equilibrium model, daily urban planning road capacities are not taken into account. Therefore, some trips might not follow the direction signs anymore. Indeed, some trips might use low-category roads between two highway roads (fig. 2.8).

The impact of information-aware routing behaviors on road traffic

Using a combination of the two classical routing models introduced in the previous subsection, chapter 3 models the impact of an increase in **information-aware routing** by dividing vehicles between the ones that have access to information (following the STA) and others (following the all-or-nothing assignment) using the **restricted path choice model**. It shows that the state of traffic converges toward a Nash equilibrium (i.e., the **average deviation incentive** decreases monotonically to 0) with an increase in information-aware routing behaviors.

Figure 2.8: July 14th, 2022 4:30 pm, traffic guidance on Google Maps within the Fremont, CA area. The request is done on July 13th and uses the Google Maps traffic prediction system. The first route suggested (in blue, orange, and red) uses a local road, then a collector road, then a minor arterial road, and then an interstate before using a minor arterial, a collector, and then a local road to reach its destination. The road categories of the road used by the trip increase and then decrease. However, in the second road suggested (in grey), this is not the case: a major arterial, a minor arterial, and a collector road are used between the same interstate road.

Chapter 3

Impact of information-aware routing behaviors on traffic from a game theory perspective

Navigational apps (like Waze, Google Maps, and Apple Maps) allegedly provide vehicles with the fastest path to reach their destinations [251]. Stepping back, at the system level, when everyone follows their fastest paths, the state of traffic becomes a Nash equilibrium or Wardrop equilibrium, where different vehicles travelling on several paths between the same origin and destination cannot have better travel time by being the only ones to change their path [247]. This translates to an equalization of the travel times of every path used between any specific origin and destination (fig. 3.1).

To model route-choice, transportation planners use the static traffic assignment (STA) [181, 161]. Used before app usage, this model assumes that vehicles choose the path that minimizes their experienced travel time. Therefore, the STA cannot directly model the impact of an increase in information-aware routing behaviors (i.e., navigation-app usage) on traffic. To account for the differences between app users and non-app users, the restricted path-choice model is introduced in section 3.1. In section 3.2, we show that an increase of information-aware routing behaviors steers the traffic to a Nash equilibrium (measured with the average deviation incentive). Experiments with a traffic microsimulation confirms the theoretical results, even with imperfect travel time estimation. In section 3.3, thought-experiments show that steering the state of traffic to a Nash equilibrium can either improve or decrease the network efficiency (as measured by the total vehicle-hours travelled in the network).

The content of this chapter is largely derived from [48, 45, 47].

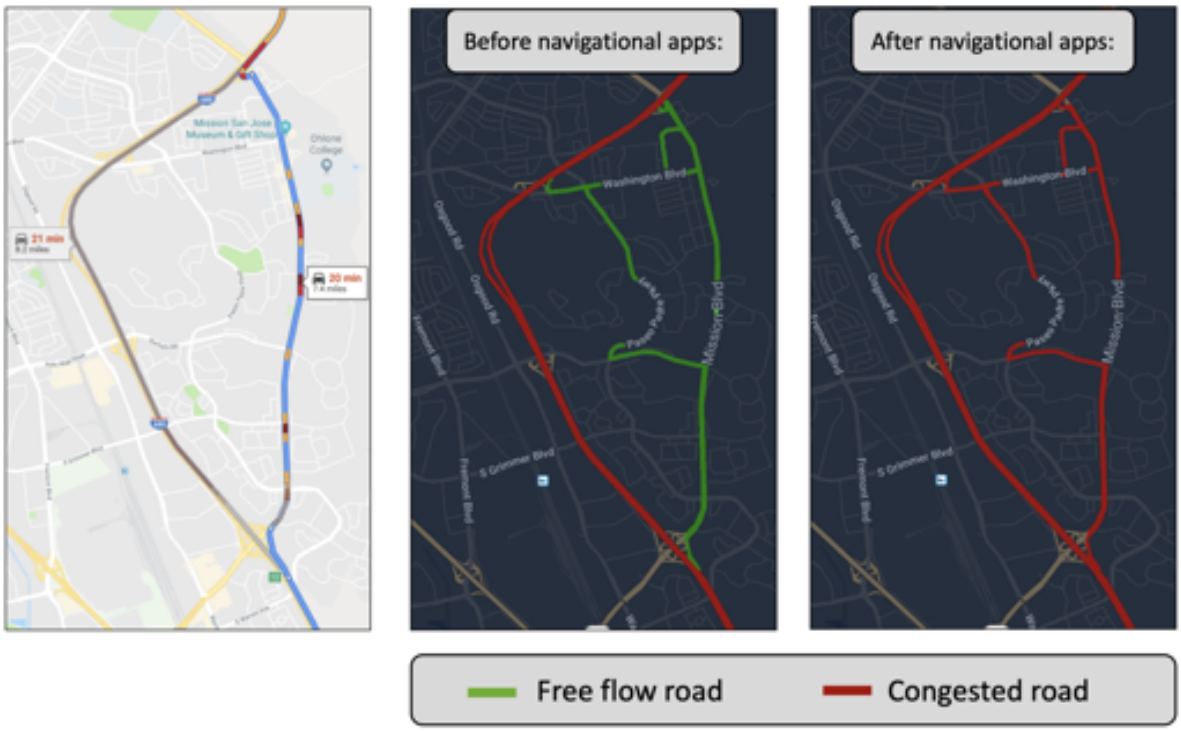


Figure 3.1: Illustrations of the impact of navigational apps on the Fremont, CA network. On the left: travel time equalization of the highway commute and of the short-cut commute routes (screenshot of Google Maps suggested directions to go north at 4pm on Monday July, 8$^{\text{th}}$ 2019). In the middle: before the apps. On the right: after the apps. The travel time equalization encodes the fact that navigational apps stir the traffic state toward a Nash equilibrium.

3.1 Static traffic assignment, navigational app usage, and average deviation incentive

This section aims to introduce the tools used in section 3.2 to understand the impact of an increase in information-aware routing behaviors (or navigational app usage) on the road traffic. First, the section describes the classical static traffic assignment (STA) model![181] that is also referred to as the static non-atomic routing game [173, Chapter 18]. Then, the section introduces the restricted path choice model, used to understand the impact of an increase in information-aware routing behaviors on road traffic. Finally, the section shows that the notion of average deviation incentive (or average marginal regret, average counterfactual regret, or exploitability) in game theory is identical to the notion of relative gap to the user equilibrium in transportation engineering.

Overview of the static traffic assignment, or non-atomic routing game

The static traffic assignment [181] is traditionally defined on a network $\mathcal{G}$ (definition 1.2.3). For each path $p \in \mathcal{P}$ of the network, the path flow $h_p \in \mathbb{R}_+$ of the path p is the flow of vehicles using this path p (definition 1.2.6). Assuming that the path flow allocation is static, the link flow $f_l \in \mathbb{R}_+$ on a link $l \in \mathcal{L}$ (definition 1.2.5) is the sum of the path flows of all paths using the link l. Then, for a given flow demand vector $\boldsymbol{d} \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ (definition 1.2.8) – which assigns for each origin $o \in \mathcal{V}$ and destination $d \in \mathcal{V}$ in the network, a flow demand $\mathrm{d}_{od} \in \mathbb{R}_+$ – we say that a path flow allocation $\mathbf{h} = (h_p)_p$ is feasible if for any origin destination pair (o, d), the demand between o and d is equal to the sum of the flows on the paths between o and d. Given a link flow allocation $\mathbf{f}$, we assume that the travel time $J(p, \mathbf{f})$ of each path p is the sum of the travel times t_l (definition 1.2.7) of every link l used by the path p. We assume that the travel time t_l of the link l is only a function of the link flow on the link f_l and of the characteristics of the link l (like length, speed limit, hourly road traffic capacity, etc.). A flow allocation is called a *user equilibrium* (definition 1.2.12) if all paths used between an origin o and a destination d have the same travel time for every od pair (Wardrop's first condition [247]). We specify this in the following.

Static Traffic Assignment framework [181]

This section appends notations and definitions to the ones presented in section 1.2.

To keep the traffic model simple, the static traffic assignment assumes static equilibrium conditions, and do not model any evolution of path flows and link flows over time.

Definition 3.1.1 (Incidence matrix)**.** *For each path $p \in \mathcal{P}$ (definition 1.2.3), we define the indicator $\delta_p = (\delta_{l\in p})_{l\in\mathcal{L}}$ (notation 1.2.1). This vector is called indicator of links included in the path p. We denote the incidence matrix $\Delta = (\delta_p)_{p\in\mathcal{P}}$*

Assumption 3.1.1 (Static equilibrium model)**.** *We assume static equilibrium conditions, i.e., $\forall p \in \mathcal{P}$, h_p is constant over time, and we have $\mathbf{f} = \Delta\mathbf{h}$ (with Δ the incidence matrix in definition 3.1.1).*

Remark 3.1.1 (Dynamic routing models)**.** *Inherently static models cannot model dynamic rerouting phenomenon, therefore chapter 5 introduces microsimulations where each individual is modeled. Because modeling each individual requires extensive computations, chapter 6 designs a dynamic routing game and solves it using mean-field game theory.*

In this static model, all the demand should be assigned to paths:

Definition 3.1.2 (Feasible assignment)**.** *Given a demand $\boldsymbol{d} \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ (definition 1.2.8), we define:*

- $\mathcal{H}_{\boldsymbol{d}} = \left\{ \mathbf{h} \in \mathbb{R}_+^{\mathcal{P}} \mid \forall (o,d) \in \mathcal{V}^2, \sum_{p\in\mathcal{P}_{od}} h_p = d_{od} \right\}$, *the set of feasible path flow allocations.*

- $\mathcal{F}_{\boldsymbol{d}} = \Delta\ \mathcal{H}_{\boldsymbol{d}} = \{\mathbf{f} \mid \exists\ \mathbf{h} \in \mathcal{H}_{\boldsymbol{d}},\ s.t.\ \mathbf{f} = \Delta\mathbf{h}\ \}$, *the set of feasible link flow allocations.*

We extend the link travel time (definition 1.2.7) into path travel time. In the static model, path travel time is the sum of the link travel time used by the path.

Definition 3.1.3 (Travel time function)**.** *For each path $p \in \mathcal{P}$, we define the travel time of the path p as $J(p, \mathbf{f}) = \sum_{l \in \mathcal{L}} t_l(f_l) \cdot \delta_p(l)$.*

Remark 3.1.2 (Path travel time)**.** *We sometimes refer to the travel time of a path as the cost of the path.*

The notation of the path travel time overrides notation 1.2.5, as vehicles will be modeled to choose the routes that minimize their travel time.

As stated in section 2.3, transportation planners assume that vehicles choose their path such that they minimize their travel time [247]. By interpreting each vehicle in the network as a player (notation 1.2.3), each route choice of a vehicle as a player's action (notation 1.2.4), and the travel time of a route as the cost of the action (notation 1.2.5), then the static traffic assignment can be understood as a game: the non-atomic routing game.

Definition 3.1.4 (Non-atomic routing game (chapter 18 of [173]))**.** *The set $(\mathcal{G}, \boldsymbol{d}, \mathbf{t})$ defines the non-atomic routing game [173, chapter 18].*

In a game, it is assumed that players (i.e., vehicles) would like to choose their actions (i.e., routes) such that they minimize their costs (i.e., travel times) (see section 1.2). The non-atomic routing game does not model individual vehicles. Instead, it models traffic flow. Therefore, the game should be understood as a game with a continuum of players (each representing one vehicle).

Remark 3.1.3 (Non-atomic routing games are potential games)**.** *Non-atomic routing games are games with an uncountable (or continuous) number of players. More specifically, they are potential games [164, 202]. Some notions defined in a classic non-cooperative game with finite number of players (see section 1.2) – like deviation incentives or regret (definition 1.2.11), ϵ-Nash equilibrium (definition 1.2.13), or average-ϵ-Nash equilibrium (definition 1.2.14) – can be extended to non-atomic games using measure-theoretical tools to deal with the continuum of players.*

Definition 3.1.5 (Deviation incentive)**.** *We extend the notion of deviation incentive (definition 1.2.11) to a vehicle. The deviation incentive of a vehicle is the difference between their travel time and the optimal travel time between their origin and destination. Because in the non-atomic routing game the number of vehicles is uncountable, we define the deviation incentive for a path p as:*

$$\rho_p(\mathbf{f}) = J(p, \mathbf{f}) - \min_{\tilde{p} \in \mathcal{P}_{od}} J(\tilde{p}, \mathbf{f})$$

Definition 3.1.6 (Average deviation incentive)**.** *Given a traffic demand* $\boldsymbol{d} \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ *and a flow allocation* $\mathbf{f} = \Delta\mathbf{h} \in \mathcal{F}_{\boldsymbol{d}}$*, we extend the notion of average deviation incentive (notation 1.2.6) to the state of traffic defined by the flow allocation* $\mathbf{f}$*:*

$$\bar{\rho}(\mathbf{f}) = \frac{1}{\|\boldsymbol{d}\|_1} \sum_{o,d\in\mathcal{V}} \sum_{p\in\mathcal{P}_{od}} h_p \cdot \rho_p(\mathbf{f})$$

We show in the next subsection that, given a traffic demand $\boldsymbol{d}$*, the average deviation incentive of a state of traffic is uniquely defined by* $\mathbf{f}$*.*

The extension of the notion of Nash equilibrium (definition 1.2.12) to the non-atomic routing game is sometimes called Wardrop equilibrium [247] or user equilibrium [181].

Definition 3.1.7 (User equilibrium [181])**.** *For the traffic demand* $\boldsymbol{d}$*, a flow allocation* $\mathbf{f} = \Delta\mathbf{h} \in \mathcal{F}_{\boldsymbol{d}}$ *is a user equilibrium if and only if:*

$$\forall o, d \in \mathcal{V},\ \forall p \in \mathcal{P}_{od},\ h_p > 0 \implies \rho_p(\mathbf{f}) = 0 \tag{3.1}$$

The user equilibrium condition is also called the Wardrop's first condition [247].

Theorem 3.1.1 (Routing game Nash equilibrium is the STA user equilibrium [173])**.** *The user equilibrium is the Nash equilibrium of the non-atomic routing game* $(\mathcal{G}, \boldsymbol{d}, \mathbf{t})$ *[194, 164, 202, 173]. See Section 2.6.1 of [181] for a deeper discussion on the connection between traffic equilibria and network games.*

Similarly to the Nash equilibrium, the notion of ϵ-Nash equilibrium (definition 1.2.13), average-ϵ-Nash equilibrium (definition 1.2.14), and social optimum (definition 1.2.15) are extended to the traffic routing scenario.

Definition 3.1.8 (ϵ-user equilibrium)**.** *For the traffic demand* $\boldsymbol{d} \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$*, a flow allocation* $\mathbf{f} = \Delta\mathbf{h} \in \mathcal{F}_{\boldsymbol{d}}$ *is an* ϵ*-user equilibrium for* $\epsilon \in \mathbb{R}_{>0}$ *if and only if:*

$$\forall o, d \in \mathcal{V},\ \forall p \in \mathcal{P}_{od},\ h_p > 0 \implies \rho_p(\mathbf{f}) < \epsilon \tag{3.2}$$

Definition 3.1.9 (Average ϵ-user equilibrium [35])**.** *Let* $\epsilon \in \mathbb{R}_{>0}$*. For the traffic demand* $\boldsymbol{d} \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$*, a flow allocation* $\mathbf{f} = \Delta\mathbf{h} \in \mathcal{F}_{\boldsymbol{d}}$ *is an average* ϵ*-user equilibrium if and only if:*

$$\bar{\rho}(\mathbf{f}) < \epsilon \tag{3.3}$$

Remark 3.1.4 (Average ϵ-user equilibrium)**.** *In an average* ϵ*-Nash equilibrium, an average "vehicle" (infinitesimal fraction of traffic flow) can expect to save* ϵ *by changing unilaterally their path.*

Definition 3.1.10 (Social optimum)**.** *For the traffic demand* $\boldsymbol{d} \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$*, a flow allocation* $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$ *is a social optimum if and only if:*

$$\forall \mathbf{f}' \in \mathcal{F}_{\boldsymbol{d}}, \quad \mathbf{t}(\mathbf{f})^\top \mathbf{f} \leq \mathbf{t}(\mathbf{f}')^\top \mathbf{f}' \tag{3.4}$$

In traffic engineering, the social optimum is also known as the system-optimum [247].

Variational inequality and minimization problem formulation

For a static non-atomic routing game, Nash equilibrium exists, is unique, and is *easy to compute* as it can be expressed as the solution to a convex optimization problem, using a convex potential function (Rosenthal function) [194].

Definition 3.1.11 (Variational inequality [181]). *We define the variational inequality problem as finding an* $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$*, such that:*

$$\forall \mathbf{f}' \in \mathcal{F}_{\boldsymbol{d}},\ \mathbf{t}(\mathbf{f})^\top(\mathbf{f}' - \mathbf{f}) \geq 0 \tag{3.5}$$

Definition 3.1.12 (Minimization problem [194]). *We define the following optimization problem with the classical Rosenthal potential:*

$$\min_{\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}} \sum_{l \in \mathcal{L}} \int_0^{f_l} t_l(x)\ \mathrm{d}x \tag{3.6}$$

Property 3.1.1 (Interpretation and equivalences [194]). *If for all links* $l \in \mathcal{L}$*, the link travel time functions* $t_l(f_l)$ *are strictly increasing functions of the link flow* f_l *then the solution of the minimization problem* (3.6)*, the variational inequality* (3.5) *and the user equilibrium* (3.1) *are the same: (i.e,* (3.1) $\iff$ (3.5) $\iff$ (3.6)*).*

Modeling navigational app usage

In this subsection, we update the static traffic assignment model to model the impact of an increase of information-aware routing behaviors in the network.

In the static traffic assignment model (definition 3.1.12), vehicles are assumed to possess perfect information over the state of the network. Therefore, the STA model is useful to model information-aware routing (app usage). However, it cannot model how a vehicle with no access to information (non-app usage) might choose its route, and therefore cannot explain the impact of information-aware routing on traffic. To tackle this, the restricted path choice model separates vehicles into two populations: those who have access to information (app users) and those who do not (non-app users). In the restricted path choice model, non-app users are following the direction signs, while app users are modeled by the STA. This model is a way to reconcile both the all-or-nothing model and the static traffic assignment model (see section 2.3).

To make the restricted path choice model more generic, the non-app users are actually modeled to choose their path into a subset of possible paths.

Restricted path choice model

App users possess perfect knowledge of the path set $\mathcal{P}_{od}$ between every origin $o \in \mathcal{V}$ and destination $d \in \mathcal{V}$ (see definition 1.2.3). Non-app users route themselves on a non-empty subset $\mathcal{P}_{od}^{na}$ of the possible path $\mathcal{P}_{od}$ ($\mathcal{P}_{od}^{na} \subset \mathcal{P}_{od}$) between every origin $o \in \mathcal{V}$ and destination

$d \in \mathcal{V}$. The app users path flow vector is denoted by $\mathbf{h}^a$ and the non-app users path flow vector by $\mathbf{h}^{na}$ (see definition 3.1.1). We note $\mathbf{h} = \mathbf{h}^a + \mathbf{h}^{na}$. Let $\alpha \in [0,1]$ be the ratio (or percentage) of app users. If $\alpha = 1$, the restricted path choice model is equivalent to the static traffic assignment defined in definition 3.1.12 and [181]. In this case, every vehicle possesses perfect information, the user equilibrium is reached.

Definition 3.1.13 (User equilibrium of the restricted path choice model)**.** *The Wardrop's first condition for the restricted path choice model can be express as, there exist* $(\pi^a_{od})_{o,d\in\mathcal{V}}$, $(\pi^{na}_{od})_{o,d\in\mathcal{V}}$, $(h^a_p)_{p\in\mathcal{P}}$, *and* $(h^{na}_p)_{p\in\mathcal{P}^{na}}$ *such that:*

$$h^a_p \cdot (J(p,\Delta\mathbf{h}) - \pi^a_{od}) = 0 \qquad \forall o,d \in \mathcal{V},\ \forall p \in \mathcal{P}_{od} \tag{3.7}$$

$$h^{na}_p \cdot (J(p,\Delta\mathbf{h}) - \pi^{na}_{od}) = 0 \qquad \forall o,d \in \mathcal{V},\ \forall p \in \mathcal{P}^{na}_{od} \tag{3.8}$$

$$h^a_p \geq 0 \qquad \forall o,d \in \mathcal{V},\ \forall p \in \mathcal{P}_{od} \tag{3.9}$$

$$h^{na}_p \geq 0 \qquad \forall o,d \in \mathcal{V},\ \forall p \in \mathcal{P}^{na}_{od} \tag{3.10}$$

$$\pi_{od} \geq 0 \qquad \forall o,d \in \mathcal{V} \tag{3.11}$$

$$\pi^{na}_{od} \geq 0 \qquad \forall o,d \in \mathcal{V} \tag{3.12}$$

$$\sum_{p\in\mathcal{P}^{na}_{od}} h^{na}_p = (1-\alpha)\, d_{od} \qquad \forall o,d \in \mathcal{V} \tag{3.13}$$

$$\sum_{p\in\mathcal{P}_{od}} h^a_p = \alpha\, d_{od} \qquad \forall o,d \in \mathcal{V} \tag{3.14}$$

Theorem 3.1.2. *With the assumption of strictly increasing travel time functions (as in property 3.1.1), only a unique flow allocation satisfies the above Wardrop's condition ([39]).*

Proof.

Lemma 1. *The user equilibrium of the restricted path choice model solves the following optimization problem:*

$$\min_{(h^a_p)_{p\in\mathcal{P}},\ (h^{na}_p)_{p\in\mathcal{P}^{na}}} \sum_{l\in\mathcal{L}} \int_0^{f_l} t_l(x)\ \mathrm{d}x$$

$$\text{subject to } \mathbf{f} = \sum_{p\in\mathcal{P}} h^a_p \cdot \delta_p + \sum_{p\in\mathcal{P}^{na}} h^{na}_p \cdot \delta_p$$

$$(1-\alpha)\, d_{od} = \sum_{p\in\mathcal{P}^{na}_{od}} h^{na}_p \qquad \forall o,d \in \mathcal{V}$$

$$\alpha\, d_{od} = \sum_{p\in\mathcal{P}_{od}} h^a_p \qquad \forall o,d \in \mathcal{V}$$

Under the assumption of strictly increasing travel time functions, the optimization problem is a strictly convex problem, with a non-empty constraint set, therefore it admits a

unique solution [39]. Definition 3.1.13 is the KKT conditions of this optimization problem [181]. Therefore, the solution of the convex problem solves definition 3.1.13. This shows that definition 3.1.13 has a solution. Finally, the unicity of the solution can be shown by contradiction [48, Property 4.2.]. □

Notation 3.1.1 (Restricted path choice Wardrop equilibrium)**.** *Following assumption 3.1.1, definition 1.2.6, definition 3.1.1 and definition 3.1.2, we denote the Wardrop flow allocation* $\mathbf{f}^\star_\alpha = \Delta(\mathbf{h}^{a,\star} + \mathbf{h}^{na,\star})$.

Introducing the average deviation incentive

This section is focused on introducing a function to evaluate how far a traffic state is from a user equilibrium. We present the characteristics desired for this function. We first discuss several quantities one might consider using and explain their advantages and disadvantages. Then, we explain why the average deviation incentive (definition 3.1.6) best meets the target.

Evaluating the distance between a traffic state and a user equilibrium

An observed traffic state – defined as a feasible flow allocation $\mathbf{f}$ (definition 3.1.2) – could be neither a user equilibrium (definition 3.1.7) nor a social optimum (definition 3.1.10). It can be interesting to know how far from a user equilibrium this traffic state is. For example, it would constitute a way to know whether vehicles efficiently chose the route that minimizes their own travel time given the flow allocation $\mathbf{f}$. This will be particularly relevant to experimentally assess the impact of routing apps on traffic. To achieve this goal, we define a function $\mathcal{R}$ which takes as an input the traffic state $\mathbf{f}$ and returns a positive real value which quantifies how far this traffic state $\mathbf{f}$ is from a user equilibrium.

Formally, given a network $\mathcal{G}$ (definition 1.2.3), a traffic demand $\boldsymbol{d} \in \mathbb{R}_+^{\mathcal{V}\times\mathcal{V}}$ (definition 1.2.8), and a travel time vector function $\mathbf{t}$ (definition 1.2.7), $\mathcal{R}$ should satisfy the following properties:

i. $\mathcal{R} : \mathcal{F}_{\boldsymbol{d}} \mapsto \mathbb{R}_+$, is a function of a feasible flow allocation $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$ (definition 1.2.5) and returns a non-negative real value.

ii. $\mathcal{R}(\mathbf{f}) = 0 \iff \mathbf{f}$ is a user equilibrium (definition 3.1.7). The function should characterize all user equilibria.

iii. If $\mathcal{R}(\mathbf{f}) < \epsilon$ then $\mathbf{f}$ is at an average-ϵ-user equilibrium for $\epsilon \in \mathbb{R}_{>0}$ given (definition 3.1.9).

iv. For every $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$, $\mathcal{R}(\mathbf{f})$ is tractable, i.e., $\mathcal{R}(\mathbf{f})$ can be computed in polynomial time with respect to $|\mathcal{L}|$ and $|\mathcal{V}|$.

v. $\mathcal{R}$ is a continuous function of the flow allocation $\mathbf{f}$.

These properties should be satisfied for every network $\mathcal{G}$, traffic demand $\boldsymbol{d}$ and travel time vector function $\mathbf{t}$.

Remark 3.1.5 (Distance)**.** *As we want $\mathcal{R}$ to be only a function of the observed state of traffic $\mathbf{f}$, it cannot be defined as a mathematical distance. We present later why defining $\mathcal{R}(\mathbf{f}) = \min_{\mathbf{f}^\star \in \mathcal{F}^\star_{\boldsymbol{d}}} d(\mathbf{f}, \mathbf{f}^\star)$ – where $\mathcal{F}^\star_{\boldsymbol{d}}$ is the set of user equilibrium flow allocation and d is a metric like $d(x, y) = \|x - y\|_2$ – would not satisfy properties we want for $\mathcal{R}$.*

A benchmark example to give some context and refute possible Nash gap quantifier candidate functions

In this section, we present different functions that can be considered to quantify the gap between an observed state of traffic and a user equilibrium. We present a benchmark network to show that average deviation incentive is more appropriate to define $\mathcal{R}$ than the other candidates of functions.

The case considered. Let us consider the Pigou network in fig. 1.1. The network consists of two nodes (o and d) and two paths ($l_1 = l$ and $l_2 = \tilde{l}$). The cost of each path depends on the flow on this path (f_1 and f_2). The link travel time functions (definition 1.2.7) are: $t_1(f_1) = 1 + f_1$ and $t_2(f_2) = M + 2 + f_2$ with $M \in \mathbb{R}_+$ given. We consider a demand of $\mathrm{d}_{od} = 1$ between o and d. The user equilibrium flow can be directly determined from the Wardrop first principle: $f_1^\star = 1$ and $f_2^\star = 0$. We note $\mathbf{f}^\star$ the vector $\mathbf{f}^\star = (f_1^\star, f_2^\star)$. Now, imagine that we observe a traffic flow $\mathbf{f}'$ where $f_1' = 1 - \alpha$ and $f_2' = \alpha$ with $0 \leq \alpha \leq 1$ given. We seek to understand what kind of functions could serve as good candidates (in the sense of properties (i), (ii),(iii),(iv) and (v) defined previously) to measure how far the observed flow $\mathbf{f}'$ is from a Nash equilibrium (here $\mathbf{f}^\star$).

Inadequacy of a flow-based function. A first intuition approach might consist of comparing the link flows between the observed state $\mathbf{f}'$ and the Nash equilibrium flow $\mathbf{f}^\star$: $\mathcal{R}(\mathbf{f}') = \|\mathbf{f}^\star - \mathbf{f}'\|_2$. In the case considered, we have that $\mathcal{R}(\mathbf{f}') = \sqrt{2}\alpha$. The function $\mathcal{R}(\mathbf{f}')$ satisfies properties (i), (ii) and (iv) for the case we consider. But the function $\mathcal{R}(\mathbf{f}')$ does not satisfy the property (iii). Indeed, $\mathcal{R}(\mathbf{f}')$ does not depend on M (or on the differences on the cost of both paths). Therefore, we consider that a flow based function is not relevant to quantify a gap to Nash.

Inadequacy of a cost-based function. Another possible approach would be to consider the travel time function $\mathcal{R}(\mathbf{f}') = \|\mathbf{t}(\mathbf{f}^\star) - \mathbf{t}(\mathbf{f}')\|_2$. In the case considered, the travel time function is just a translation of the flow vector, so that we still have $\mathcal{R}(\mathbf{f}') = \sqrt{2}\alpha$. For the same reasons as the flow based function, we consider the cost based function inadequate.

Inadequacy of a hybrid-based function. A third approach is to consider norm on $\mathbf{f} \cdot \mathbf{t}(\mathbf{f})$: $\mathcal{R}(\mathbf{f}') = |\mathbf{f}^\star \cdot \mathbf{t}(\mathbf{f}^\star) - \mathbf{f}' \cdot \mathbf{t}(\mathbf{f}')|$ (like in [189]). But, as seen in definition 3.1.10, $\mathbf{f} \cdot \mathbf{t}(\mathbf{f})$ is the total travel time on the network (the sum of the travel times of all cumulated flows). So any function on $\mathbf{f} \cdot \mathbf{t}(\mathbf{f})$ will be related to the social optimality of the solution. This will not satisfy the fact that we want to measure a gap between a traffic state $\mathbf{f}$ and a user equilibrium. In the case considered, $\mathcal{R}(\mathbf{f}') = \alpha|M - 1 + 2\alpha|$. If $|M - 1 + 2\alpha| = 0$, $\mathbf{f}'$ is not a Nash equilibrium but $\mathcal{R}(\mathbf{f}') = 0$. So $\mathcal{R}(\mathbf{f}')$ does not satisfy the property (ii).

Inadequacy of a price of anarchy-like function. Considering an idea similar to the price of anarchy [195], one can define $\mathcal{R}(\mathbf{f}') = \frac{\mathbf{f}' \cdot \mathbf{t}(\mathbf{f}')}{\mathbf{f}^\star \cdot \mathbf{t}(\mathbf{f}^\star)}$ where $\mathbf{f}^\star$ is a user equilibrium. This is similar to the hybrid based function: $\frac{\mathbf{f}' \cdot \mathbf{t}(\mathbf{f}')}{\mathbf{f}^\star \cdot \mathbf{t}(\mathbf{f}^\star)} = 1 + \frac{1}{\mathbf{f}^\star \cdot \mathbf{t}(\mathbf{f}^\star)} \cdot (\mathbf{f}' \cdot \mathbf{t}(\mathbf{f}') - \mathbf{f}^\star \cdot \mathbf{t}(\mathbf{f}^\star))$. Therefore, this type of function is equivalent to a hybrid-based function and, thus, is inadequate for our purpose.

The worst "deviation incentive" of vehicles function. Another type of approach is to define $\mathcal{R}$ using the game theoretical framework. Using deviation incentives, we do not need to know every user equilibrium flow allocation $\mathbf{f}^\star$ to find out whether a state of traffic is a user equilibrium. A user equilibrium (definition 3.1.7) is defined as a state of traffic where nobody can achieve a better travel time by being the only one changing their route (Wardrop's first condition, see definition 3.1.7): i.e., no one has deviation incentive (definition 3.1.5). One idea is to define $\mathcal{R}(\mathbf{f}')$ as the worst deviation incentive of vehicles in the observed traffic assignment. This is equivalent to define $\mathcal{R}(\mathbf{f}')$ as the smallest ϵ such that $\mathbf{f}'$ is a ϵ-Nash equilibrium:

$$\mathcal{R}(\mathbf{f}') = \max_{\substack{p \in \mathcal{P} \\ h_p > 0}} \rho_p(\mathbf{f}')$$

As the deviation incentive is defined for a path flow allocation h_p and not a link flow allocation $\mathbf{f}$, property (i) is not satisfied. Assuming h_p is known, properties (i) and (ii) are satisfied. As $\frac{\|\cdot\|_1}{n} \leq \|\cdot\|_\infty$, property (iii) is also satisfied. Satisfying property (iv) is more complicated, as the number of paths of a network is generally an exponential function of the number of nodes $|\mathcal{V}|$ and the number of links $|\mathcal{L}|$.

Property (v) is not satisfied. For the considered case, we have that

$$\mathcal{R}(\mathbf{f}') = \begin{cases} M + 2\alpha & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha = 0 \end{cases}.$$

We see that $\mathcal{R}(\mathbf{f}')$ is not continuous in $\mathbf{f}'$. If $\alpha = 0$ then $\mathcal{R}(\mathbf{f}') = 0$. However, if $\alpha = 0^+$ then $\mathcal{R}(\mathbf{f}') = M$.

The average "deviation incentive" of vehicles function. We define $\mathcal{R}(\mathbf{f}')$ as the average deviation incentive (definition 3.1.6) of the state of traffic defined by $\mathbf{f}'$:

$$\mathcal{R}(\mathbf{f}') = \bar{\rho}(\mathbf{f}')$$

Based on the definition of the average-ϵ-user equilibrium (definition 3.1.9), $\mathcal{R}(\mathbf{f}')$ is the smallest ϵ such that $\mathbf{f}'$ is at average-ϵ-user equilibrium. We will use the average deviation incentive function $\bar{\rho}$ as $\mathcal{R}$ in the remainder of the chapter.

The next subsection shows that the function satisfies properties (i), (ii), (iii), (iv) and (v).

The average deviation incentive

In this section, we formulate some properties of the average deviation incentive, which quantifies how much time an average vehicle can expect to save by changing their path to the optimal one. In particular, the average deviation incentive satisfies the properties stated before.

Definition 3.1.14 (Best path, optimal flow pattern)**.** *Given a flow allocation $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$, which provides the cost vector $\mathbf{t}(\mathbf{f})$, we define:*

- *An optimal path between o and d, as $p^*_{od}(\mathbf{f}) \in \underset{p \in \mathcal{P}_{od}}{\operatorname{argmin}} \delta_p^\top \mathbf{t}(\mathbf{f}) = \mathcal{P}^*_{od}(\mathbf{f})$*
- *An all-or-nothing allocation $\mathbf{y}(\mathbf{f}) \in \mathcal{F}_{\boldsymbol{d}}$ based on the travel times at the flow $\mathbf{f}$, as $\mathbf{y}(\mathbf{f}) = \sum_{o,d \in \mathcal{V}} d_{od} \cdot \delta_{p^*_{od}(f_l)}$ for $\delta_{p^*_{od}(f_l)} \in \mathcal{P}^*_{od}(\mathbf{f})$. In this definition, the full od demand d_{od} is allocated to an optimal path (between o and d) computed with the current flow allocation $\mathbf{f}$.*

Definition 3.1.15 (Shortest travel time)**.** *From definition 3.1.14, any optimal path $p^*_{od}(\mathbf{f})$ is a shortest path (with respect to cost) between o and d given the cost on each link $\mathbf{t}(\mathbf{f})$. Therefore, given $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$, we can define the minimum travel time between any $o, d \in \mathcal{V}$ as:*

$$\pi_{od}(\mathbf{f}) = \mathbf{t}(\mathbf{f})^\top \delta_{p^*_{od}(f_l)}$$

Remark 3.1.6 (Existence and non-uniqueness)**.** *For all $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$, $p^*_{od}(\mathbf{f})$ and $\mathbf{y}(\mathbf{f})$ exist but might not be unique.*

We show that the average deviation incentive can be computed as the inner product of the travel time vector and the actual flow allocation minus the all-or-nothing flow allocation normalized with the total demand.

Property 3.1.2 (Average deviation incentive)**.** *The average deviation incentive of the flow pattern $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$ can be computed as follows:*

$$\bar{\rho}(\mathbf{f}) = \frac{1}{\|\boldsymbol{d}\|_1} \mathbf{t}(\mathbf{f})^\top (\mathbf{f} - \mathbf{y}(\mathbf{f})) \tag{3.15}$$

where $\|\boldsymbol{d}\|_1 = \sum_{o,d \in \mathcal{V}} d_{od}$ and $\mathbf{y}(\mathbf{f})$ is an all-or-nothing solution, as in definition 3.1.14.

Proof. Using $\mathbf{f} = \Delta\mathbf{h}$, we have:

$$\begin{aligned}
\mathbf{t}(\mathbf{f})^\top \mathbf{f} &= \sum_{(o,d)\in\mathcal{V}^2} \sum_{p\in\mathcal{P}_{od}} h_p \cdot J(p,\mathbf{f}) \\
\mathbf{t}(\mathbf{f})^\top \mathbf{y}(\mathbf{f}) &= \sum_{(o,d)\in\mathcal{V}^2} \mathbf{t}(\mathbf{f})^\top \delta_{p^*_{od}(f_l)} \cdot \mathrm{d}_{od} = \sum_{(o,d)\in\mathcal{V}^2} \pi_{od}(\mathbf{f}) \cdot \mathrm{d}_{od} \\
\mathbf{t}(\mathbf{f})^\top (\mathbf{f} - \mathbf{y}(\mathbf{f})) &= \sum_{(o,d)\in\mathcal{V}^2} \left(\left(\sum_{p\in\mathcal{P}_{od}} h_p \cdot J(p,\mathbf{f}) \right) - \mathrm{d}_{od} \cdot \pi_{od}(\mathbf{f}) \right) \\
\sum_{p\in\mathcal{P}_{od}} h_p = \mathrm{d}_{od} \implies \quad \mathbf{t}(\mathbf{f})^\top (\mathbf{f} - \mathbf{y}(\mathbf{f})) &= \sum_{(o,d)\in\mathcal{V}^2} \sum_{p\in\mathcal{P}_{od}} h_p \cdot (J(p,\mathbf{f}) - \pi_{od}(\mathbf{f})) \\
\frac{1}{\|\boldsymbol{d}\|_1} \cdot \mathbf{t}(\mathbf{f})^\top (\mathbf{f} - \mathbf{y}(\mathbf{f})) &= \bar{\rho}(\mathbf{f})
\end{aligned}$$

Note that this shows that $\bar{\rho}$ is defined even if $\mathbf{y}(\mathbf{f})$ is not unique. □

Remark 3.1.7 (Measuring the average deviation incentive)**.** *Because the average deviation incentive is a function of the link flow only, the link travel time and the traffic demand, it can be accessed with loop detectors and demand survey. Knowing the path flows is not required to measure the average deviation incentive. The function $\bar{\rho}$ satisfies the property (iv).*

Remark 3.1.8 (Interpretation of the average deviation incentive)**.** *Since $\rho_p(\mathbf{f})$ represents the time a vehicle on path $p \in \mathcal{P}_{od}$ could save by choosing the best path for their trip, $\bar{\rho}$ can be interpreted as the average time a vehicle could expect to save by changing unilaterally their path.*

Property 3.1.3 (The average deviation incentive is a positive real value and characterizes all user equilibria)**.** *As $p^*_{od}(\mathbf{f})$ is the fastest path between o and d, we have*

$$\frac{1}{\|\boldsymbol{d}\|_1} \cdot \mathbf{t}(\mathbf{f})^\top (\mathbf{f} - \mathbf{y}(\mathbf{f})) = \max_{\mathbf{x}\in\mathcal{F}_{\boldsymbol{d}}} \frac{1}{\|\boldsymbol{d}\|_1} \cdot \mathbf{t}(\mathbf{f})^\top (\mathbf{f} - \mathbf{x}) \geq 0 \tag{3.16}$$

thus:

$$\bar{\rho}(\mathbf{f}) = 0 \iff \forall \mathbf{f}' \in \mathcal{F}_{\boldsymbol{d}},\ \mathbf{t}(\mathbf{f})^\top (\mathbf{f}' - \mathbf{f}) \geq 0 \tag{3.17}$$

Equation (3.16) implies that $\forall \mathbf{f} \in \mathcal{F}_{\boldsymbol{d}},\ \bar{\rho}(\mathbf{f}) \geq 0$: $\bar{\rho} : \mathcal{F}_{\boldsymbol{d}} \mapsto \mathbb{R}_+$, is a function of a feasible flow allocation and returns a positive real value (property (i)).

Equation (3.17) – using the variational inequality definition of user equilibrium (definition 3.1.11) – provides that $\bar{\rho}(\mathbf{f}) = 0 \iff \mathbf{f}$ is a user equilibrium (property (ii)).

Property 3.1.4 (The average deviation incentive as a measure of vehicle efficiency)**.** *Given $\epsilon \in \mathbb{R}_{>0}$, from remark 3.1.8, it is straightforward that $\forall \mathbf{f} \in \mathcal{F}_{\boldsymbol{d}},\ \bar{\rho}(\mathbf{f}) \leq \epsilon \iff \mathbf{f}$ is an average-ϵ-Nash equilibrium (definition 3.1.9). This is property (iii).*

So, the average deviation incentive is a good way to characterize how close to a user equilibrium the state of traffic is.

A player is defined as efficient if they take one of the best routes between their origin and destination. Then $\bar{\rho}$ can be interpreted as a measure of the efficiency of the vehicles.

The closer $\bar{\rho}$ is to 0, the less inclined players are to change their paths. If $\bar{\rho} = 0$, the user equilibrium is reached.

Property 3.1.5 (Continuity)**.** *The average deviation incentive $\bar{\rho}(\mathbf{f})$ is continuous with respect to $\mathbf{f}$. Property (v) is satisfied.*

Proof. We have $\bar{\rho}(\mathbf{f}) = \frac{1}{\|\boldsymbol{d}\|_1}\mathbf{t}(\mathbf{f})^\top(\mathbf{f} - \mathbf{y}(\mathbf{f})) = \frac{1}{\|\boldsymbol{d}\|_1}\left(\mathbf{t}(\mathbf{f})^\top\mathbf{f} - \min_{\tilde{\mathbf{f}}\in\mathcal{F}_{\boldsymbol{d}}}\mathbf{t}(\mathbf{f})^\top\tilde{\mathbf{f}}\right)$. Because $\mathbf{t}(\mathbf{f})$ is continuous with respect to $\mathbf{f}$ (definition 1.2.7), it suffices to show that $\min_{\tilde{\mathbf{f}}\in\mathcal{F}_{\boldsymbol{d}}}\mathbf{t}(\mathbf{f})^\top\tilde{\mathbf{f}}$ is continuous with respect to $\mathbf{f}$. This is a linear program (LP) ($\mathcal{F}_{\boldsymbol{d}}$ defined in definition 3.1.2 is a polytope). The optimal objective value of an LP is continuous with respect to perturbation on the objective function [39]. □

Remark 3.1.9 (Computational time)**.** *An optimal path p^*_{od} can be found in $\mathcal{O}(|\mathcal{L}|\cdot\log(|\mathcal{V}|))$ with Dijkstra's algorithm [78]. The average deviation incentive can be found in $\mathcal{O}(|\mathcal{L}|\cdot|\mathcal{V}|\log(|\mathcal{V}|))$ with a sequential application of $|\mathcal{V}|$ Dijkstra's algorithms. Therefore, the average deviation incentive satisfies property (iv).*

The average deviation incentive $\bar{\rho}$ can be computed with "local data" only, i.e. link cost, link flow and demand. Property 3.1.2 shows that we only need to compute the inner product of the travel time vector and the difference between the flow allocation and the all-or-nothing flow allocation given the current travel time vector. To compute the all-or-nothing flow allocation, only the demand and the current travel time vector are needed.

Remark 3.1.10 (Frank Wolfe's Algorithm [94, 35])**.** *The user equilibrium can be seen as the Nash equilibrium of the static non-atomic routing game (theorem 3.1.1). It has been shown that a no-regret learning algorithm in selfish routing converges to a user equilibrium of the system [35, 132, 173, Chapter 4].*

For solving the minimization problem (3.6), we can use the Frank Wolfe's algorithm [94], a projected gradient descent algorithm. This algorithm minimizes an approximation of $\bar{\rho}$ over each iteration of the algorithm. It is equivalent to a no-regret learning algorithm [35]:

At the termination of this algorithm, we have $\bar{\rho}(\mathbf{f}) \leq \epsilon$. Here, ϵ is an input parameter of the algorithm which represents the accuracy threshold on the value of $\mathbf{f}$.

In traffic engineering, the average deviation incentive is called the relative gap to the user equilibrium [231], and it is used as a termination condition for the Frank-Wolfe algorithm.

Algorithm 1: Frank Wolfe's algorithm: the average deviation incentive as a criterion of convergence

Data: $(\mathcal{V}, \mathcal{L})$, $\epsilon \in \mathbb{R}_{>0}$, $\boldsymbol{d} \in \mathbb{R}_+^{|\mathcal{V}|\times|\mathcal{V}|}$, $\mathbf{t} \in \mathbb{R}_+^{|\mathcal{L}|} \to \mathbb{R}_+^{|\mathcal{L}|}$
Set $k = 1$;
Take any $\mathbf{f}^k \in \mathcal{F}_{\boldsymbol{d}}$;
while $\bar{\rho}(\mathbf{f}^k) > \epsilon$ **do**
  $k = k + 1$;
  $\mathbf{f}^k = \mathbf{f}^{k-1} + \frac{1}{k} \cdot (\mathbf{y}(\mathbf{f}^{k-1}) - \mathbf{f}^{k-1})$;
end
Result: $\mathbf{f}^k$

3.2 Information-aware routing behaviors steer the state of traffic toward a Nash equilibrium

Sensitivity analysis of the equilibrium state of the restricted path choice model with respect to the app usage ratio proves that the average deviation incentive monotonically decreases to 0 with an increase of app usage.

Theoretical convergence of the restricted path choice model to Nash with the increase of app usage.

In the Nash equilibrium of the restricted-path choice model (definition 3.1.13), app users do not have "regrets" while non-app users "regret" to not know $\mathcal{P}_{od}$. Without loss of generalities, and for the sake of convenience, we assume in this section that $\|\boldsymbol{d}\|_1 = 1$. We can express the average deviation incentive associated with $\mathbf{f}_\alpha^\star$ (notation 3.1.1):

$$\bar{\rho}(\mathbf{f}_\alpha^\star) = \sum_{o,d\in\mathcal{V}} \sum_{p\in\mathcal{P}_{od}} h_p \rho_p(\mathbf{f}_\alpha^\star)$$

$$eq.\ (3.7), eq.\ (3.8) \implies \bar{\rho}(\mathbf{f}_\alpha^\star) = \sum_{o,d\in\mathcal{V}} \sum_{p\in\mathcal{P}_{od}} h_p^a \cdot (\pi_{od}(\mathbf{f}_\alpha^\star) - \pi_{od}(\mathbf{f}_\alpha^\star)) + h_p^{na} \cdot (\pi_{od}^{na}(\mathbf{f}_\alpha^\star) - \pi_{od}(\mathbf{f}_\alpha^\star))$$

$$\pi_{od}(\mathbf{f}_\alpha^\star) = \pi_{od}(\mathbf{f}_\alpha^\star) \implies \bar{\rho}(\mathbf{f}_\alpha^\star) = \sum_{o,d\in\mathcal{V}} \sum_{p\in\mathcal{P}_{od}} h_p^{na} \cdot (\pi_{od}^{na}(\mathbf{f}_\alpha^\star) - \pi_{od}(\mathbf{f}_\alpha^\star))$$

Then eq. (3.13) gives:

$$\bar{\rho}(\mathbf{f}_\alpha^\star) = (1-\alpha) \sum_{o,d\in\mathcal{V}} \mathrm{d}_{od} \cdot (\pi_{od}^{na}(\mathbf{f}_\alpha^\star) - \pi_{od}(\mathbf{f}_\alpha^\star)) \tag{3.18}$$

Similarly to satisfying property (v) of the average deviation incentive (continuity with respect to the link flow allocation), we are interested in the continuity of $\bar{\rho}(\mathbf{f}_\alpha^\star)$ with respect to α.

Theorem 3.2.1 (Continuity of the average deviation incentive with the ratio of app users)**.** *The average deviation incentive $\bar{\rho}(\mathbf{f}^\star_\alpha)$ is continuous as a function of α.*

Proof. This follows from the continuity of the average deviation incentive (see property 3.1.5) and of the continuity of $\mathbf{f}^\star_\alpha$ with respect to α. The continuity of $\mathbf{f}^\star_\alpha$ with respect to α is due to the convexity of the restricted path choice model, as shown in the proof of theorem 3.1.2. A more detailed proof is given in [110, Theorem 1]. □

Theorem 3.2.2 (Monotonicity and convergence to Nash)**.** *For α_1, α_2 such that $0 \leq \alpha_1 \leq \alpha_2 \leq 1$:*

$$\bar{\rho}(\mathbf{f}^\star_{\alpha_2}) \leq \bar{\rho}(\mathbf{f}^\star_{\alpha_1}) \text{ and } \lim_{\alpha_2 \to 1} \bar{\rho}(\mathbf{f}^\star_{\alpha_2}) = 0$$

Proof. First, given that $\bar{\rho}(\mathbf{f}^\star_\alpha)$ is continuous with α (theorem 3.2.1), $\bar{\rho}(\mathbf{f}^\star_{\alpha=1}) = 0$ (eq. (3.18)) gives that $\lim_{\alpha \to 1} \bar{\rho}(\mathbf{f}^\star_\alpha) = 0$.

Then, we can use the sensitivity analysis of the travel cost $\pi_{od}(\mathbf{f}^\star_\alpha)$ and $\pi^{na}_{od}(\mathbf{f}^\star_\alpha)$ with respect to α (as in [72, 166]). By using eq. (3.18), we have:

$$\begin{aligned}
\bar{\rho}(\mathbf{f}^\star_{\alpha_1}) - \bar{\rho}(\mathbf{f}^\star_{\alpha_2}) &= (1-\alpha_1) \sum_{o,d \in \mathcal{V}} \mathrm{d}_{od} \cdot \left(\pi^{na}_{od}(\mathbf{f}^\star_{\alpha_1}) - \pi_{od}(\mathbf{f}^\star_{\alpha_1})\right) \\
&\quad - (1-\alpha_2) \sum_{o,d \in \mathcal{V}} \mathrm{d}_{od} \cdot \left(\pi^{na}_{od}(\mathbf{f}^\star_{\alpha_2}) - \pi_{od}(\mathbf{f}^\star_{\alpha_2})\right) \\
&= (\alpha_2 - \alpha_1) \sum_{o,d \in \mathcal{V}} \mathrm{d}_{od} \cdot (\pi^{na}_{od}(\mathbf{f}^\star_{\alpha_1}) - \pi_{od}(\mathbf{f}^\star_{\alpha_1})) \\
&\quad + (1-\alpha_2) \sum_{o,d \in \mathcal{V}} \mathrm{d}_{od} \cdot \left((\pi^{na}_{od}(\mathbf{f}^\star_{\alpha_1}) - \pi^{na}_{od}(\mathbf{f}^\star_{\alpha_2})) - (\pi_{od}(\mathbf{f}^\star_{\alpha_1}) - \pi_{od}(\mathbf{f}^\star_{\alpha_2}))\right)
\end{aligned}$$

Because $\alpha_1 \leq \alpha_2$ and $\pi^{na}_{od}(\mathbf{f}^\star_{\alpha_1}) \geq \pi_{od}(\mathbf{f}^\star_{\alpha_1})$ then $(\alpha_2 - \alpha_1) \sum_{o,d \in \mathcal{V}} \mathrm{d}_{od} \cdot (\pi^{na}_{od}(\mathbf{f}^\star_{\alpha_1}) - \pi_{od}(\mathbf{f}^\star_{\alpha_1})) \geq 0$.

Using Dafermos sensitivity analysis of travel cost with respect to the demand [72, Theorem 4.2], we will show that $\sum_{o,d \in \mathcal{V}} \mathrm{d}_{od} \cdot \left((\pi^{na}_{od}(\mathbf{f}^\star_{\alpha_1}) - \pi^{na}_{od}(\mathbf{f}^\star_{\alpha_2})) - (\pi_{od}(\mathbf{f}^\star_{\alpha_1}) - \pi_{od}(\mathbf{f}^\star_{\alpha_2}))\right) \geq 0$. Since $(1-\alpha_2) \geq 0$, it will complete the proof.

Changing the problem (in definition 3.1.13) into a stationary traffic assignment problem by vectorizing it, we denote $\tilde{\pi}_{o,d} = (\pi_{od}(\mathbf{f}^\star_{\alpha_1}),\ \pi^{na}_{od}(\mathbf{f}^\star_{\alpha_1}))$, $\tilde{\mathrm{d}}_{o,d} = (\alpha_1\ \mathrm{d}_{od},\ (1-\alpha_1)\ \mathrm{d}_{od})$, and $\tilde{\pi}^\star_{o,d} = (\pi_{od}(\mathbf{f}^\star_{\alpha_2}),\ \pi^{na}_{od}(\mathbf{f}^\star_{\alpha_2}))$, $\tilde{\mathrm{d}}^\star_{o,d} = (\alpha_2\ \mathrm{d}_{od},\ (1-\alpha_2)\ \mathrm{d}_{od})$. This notation is inspired by Dafermos [72]. Then, the Dafermos sensitivity analysis of the travel cost with respect to the

demand [72, Theorem 4.2] gives:

$$\sum_{o,d\in\mathcal{V}} (\tilde{\pi}^{\star}_{o,d} - \tilde{\pi}_{o,d})^{\top}(\tilde{\mathrm{d}}^{\star}_{o,d} - \tilde{\mathrm{d}}_{o,d}) \geq 0$$

Going back to previous notations:

$$\sum_{o,d\in\mathcal{V}} (\pi_{od}(\mathbf{f}^{\star}_{\alpha_2}) - \pi_{od}(\mathbf{f}^{\star}_{\alpha_1}))^{\top}((\alpha_2-\alpha_1)\ \mathrm{d}_{o,d}) - (\pi^{na}_{od}(\mathbf{f}^{\star}_{\alpha_2}) - \pi^{na}_{od}(\mathbf{f}^{\star}_{\alpha_1}))^{\top}((\alpha_2-\alpha_1)\ \mathrm{d}_{o,d}) \geq 0$$

$$(\alpha_2-\alpha_1)\sum_{o,d\in\mathcal{V}} \mathrm{d}_{o,d}\cdot\left(((\pi_{od}(\mathbf{f}^{\star}_{\alpha_2}) - \pi_{od}(\mathbf{f}^{\star}_{\alpha_1})) - (\pi^{na}_{od}(\mathbf{f}^{\star}_{\alpha_2}) - \pi^{na}_{od}(\mathbf{f}^{\star}_{\alpha_1}))\right) \geq 0$$

$$\sum_{o,d\in\mathcal{V}} \mathrm{d}_{o,d}\cdot\left(((\pi^{na}_{od}(\mathbf{f}^{\star}_{\alpha_1}) - \pi^{na}_{od}(\mathbf{f}^{\star}_{\alpha_2})) - ((\pi_{od}(\mathbf{f}^{\star}_{\alpha_1}) - \pi_{od}(\mathbf{f}^{\star}_{\alpha_2}))\right) \geq 0$$

This shows the claim $\bar{\rho}(\mathbf{f}^{\star}_{\alpha_1}) \geq \bar{\rho}(\mathbf{f}^{\star}_{\alpha_2})$. □

The average deviation incentive decreases monotonically to zero when the ratio of app users increases uniformly using the restricted path choice model.

Simulations showing decrease of the average deviation incentive implied by an increase in app-usage

In this subsection, we compute the user equilibrium of the restricted path choice model for different ratios of app users/non-app users on two networks. First, on a benchmark network with three paths, we show that the average deviation incentive (definition 3.1.6) converges to 0 when the app usage increases. Then, on the Los Angeles, CA (L.A.) network, simulations show the same phenomenon with the restricted path choice model (definition 3.1.13). Because the restricted path choice model is a static model, we also simulate traffic with the Aimsun microsimulator [227], that models each vehicle independently. Calibrated simulations of the I-210 traffic in Los Angeles, CA confirm that an increase of information-aware routing behaviors steer the traffic to a Nash equilibrium.

Open source code to solve static traffic assignment used for the work. Transportation networks with hourly road traffic capacity (definition 2.1.2), free-flow travel times, travel demand, (definition 1.2.8) and the solver used for the static traffic assignment are open source [109, 134]. The link travel time functions (definition 1.2.7) are set to $t_l(f_l) = t^0_l\left(1 + 0.15\left(\frac{f_l}{c_l}\right)^4\right)$ where t^0_l is the free flow travel time on the link, c_l is its hourly road traffic capacity, and f_l is the flow on the link (definition 1.2.5), as suggested by [239]. Open Street Map [109], and OSMNX [37] can be used to model the road network.

App usage on a benchmark network

The first computation (shown in fig. 3.2) studies the impact of the number of app users on the traffic state. The highway hourly road traffic capacity is 6,000 veh./h. The arterial hourly road traffic capacities are fixed at 2,000 veh./h. The OD demand (definition 1.2.8) is set to 20,000 veh./h ($d_{od} = 20,000$). We split the demand between the two populations of vehicles: app users $\boldsymbol{d}_a$ and non-app users $\boldsymbol{d}_{na}$. We have $\boldsymbol{d}_a + \boldsymbol{d}_{na} = \boldsymbol{d}$ and $\boldsymbol{d}_a = \alpha\, \boldsymbol{d}$ with $\alpha \in [0, 1]$. We call α the ratio (or percentage) of app users. Using the restricted path choice model (definition 3.1.13), non-app users stay on the highway regardless of the traffic conditions.

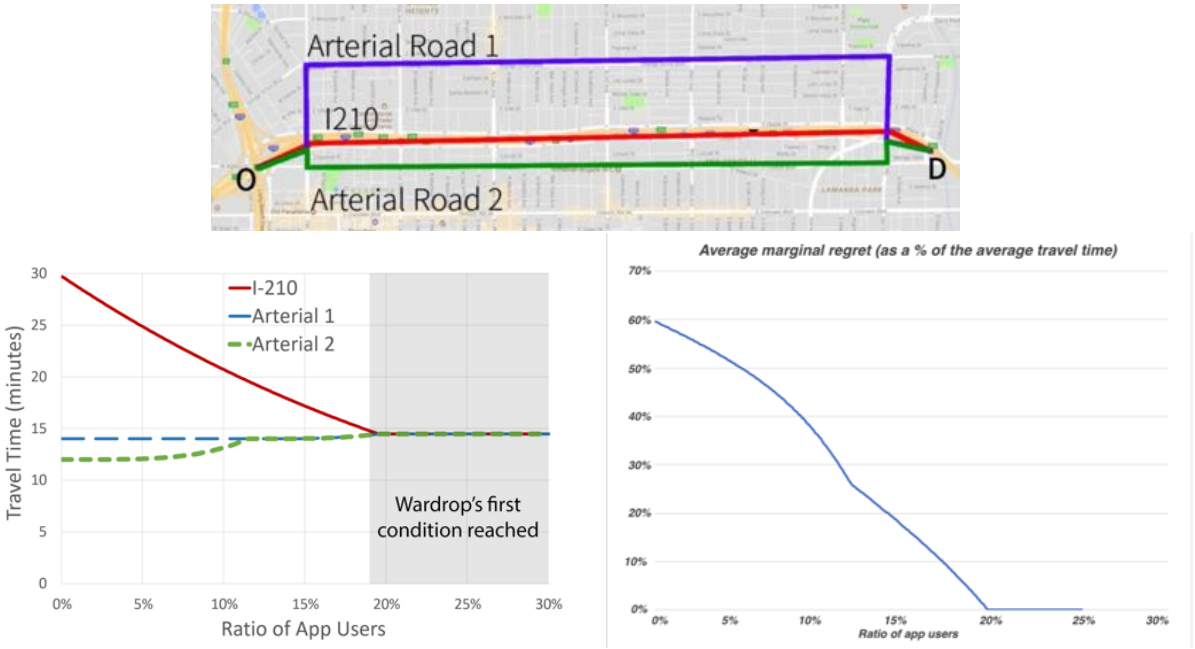


Figure 3.2: On top: Benchmark network. At the bottom: The path travel times (on the left) and the average deviation incentive (on the right) as a function of the percentage of app usage. On the benchmark network, the traffic converges to a user equilibrium state when app usage increase. Figure reproduced from [47, Figure 3].

As shown in fig. 3.2, with 0% routed users, the entire flow stays on the highway. As the ratio of app users increases, app users start using Arterial Road 2 (AR2), because it is faster than the congested highway. This transfer relieves the freeway, but increases congestion on AR2. When the travel time on AR2 becomes as high as the travel time on Arterial Road 1 (AR1), app users start taking AR1 as well. Travel times stop evolving when app usage reaches 18%, which corresponds to a travel time equalization phenomenon: in these conditions, no app-user can reroute to decrease their travel time and Wardrop's first condition is reached.

The average deviation incentive decreases with the increase of app usage. It reaches 0 when app usage reaches 18% at this point the user equilibrium condition is reached.

Remark 3.2.1 (Cut-through traffic and hignway efficiency)**.** *Interestingly, in 1989, Adolf D. May working with CalTrans and General Motors supported the development of navigational apps stating that apps will spread congestion over space, which will lead to a decrease of the highway facility demand and therefore improve the highway efficiency [71]. This is confirmed by this first benchmark model. However, the decrease of highway facility demand is at the price of an increase of some local road facility demand, that have a lower urban planning capacity (definition 2.1.1) than highways. Therefore, some challenges due to congestion on local roads arise from the new information-aware routing behaviors, as explained in chapter 4.*

App usage on the L.A. network

A second simulation is performed on the L.A. network (fig. 3.3). Simulations use the cognitive cost static model [224] with app user percentages ranging from 0% to 100%, with a 1% increment. For each of these simulations, traffic demand data is collected from the American Community Survey, composed of 96,077 *od* pair. The network is built from Open Street Map [109]. Traffic demand is set consistently at rush hour levels to find the effects of app usage when networks are congested.

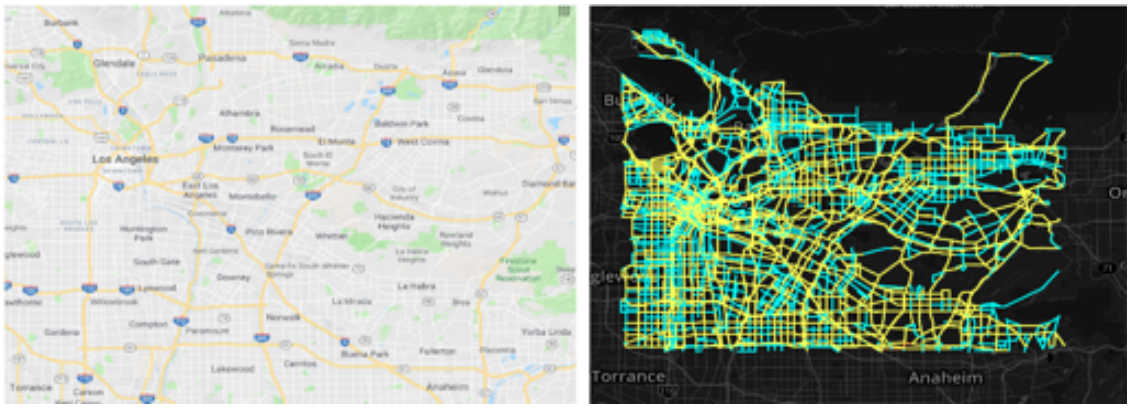


Figure 3.3: Los Angeles, CA (L.A.) network considered for experiment. On the left: a map of the L.A. basin, on the right: the graph we use to model the L.A. basin. Figure reproduced from [47, Figure 4].

The average deviation incentive decreases monotonically with the increase of navigational app usage (fig. 3.4). The fact that the decrease is monotonic tells that whatever the percentage of app users is, the traffic will be closer to a user equilibrium when app usage increases.

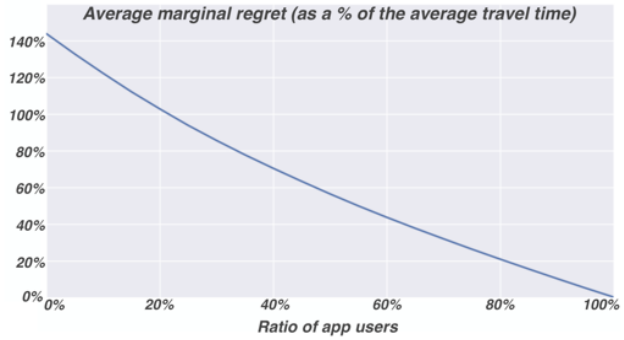


Figure 3.4: Average deviation incentive as a function of the percentage of app usage on the L.A. network. The average deviation incentive decreases monotonically to 0 when app usage increases. Figure reproduced from [47, Figure 5].

Remark 3.2.2. *For every simulation run and on every type of network, the average deviation incentive monotonically decreases with the increase of navigational app usage. This is not the case with the price of anarchy, which depends on the network configuration.*

We show the evolution of path flow for a particular *od* pair. This *od* pair has been chosen to be one of the *od* pairs with the highest demand. This particular *od* pair starts in slightly southeast of Compton and ends just north of Burbank. Figure 3.5 shows the top five paths taken for this *od* pair. Almost all of these paths take the SR 2 through Glendale. One takes the I-210 through Pasadena (the green one).

In the 0% to 35% app usage range, almost all app users take the green path, which is the fastest. But then, with 35% app usage, app users begin to take other paths, particularly the blue (Path 1) and red (Path 3) ones. 35% app usage is exactly when the travel time of path green, blue, and red equalize. App users always follow the fastest paths. Then, after 35% app usage, the travel time of the other paths fall below that of the green path and all app users leave the green path for other paths.

Remark 3.2.3 (Travel time evolution)**.** *It is important to see that here the travel time of these paths depends on other od pairs. Even after 35% app usage, when no rerouting occurs, the path travel time still varies, Mainly because vehicles from other od pairs still change their path while the ratio of app usage increases.*

Further demonstration through microsimulations

While the previous models present several desirable features such as being analytical, compact, and implementable at scale, they are idealized and static. To further connect this work with practice, we implemented the concepts embedded in these models into microsimulations (which integrate app usage at the individual vehicle level). This was completed using

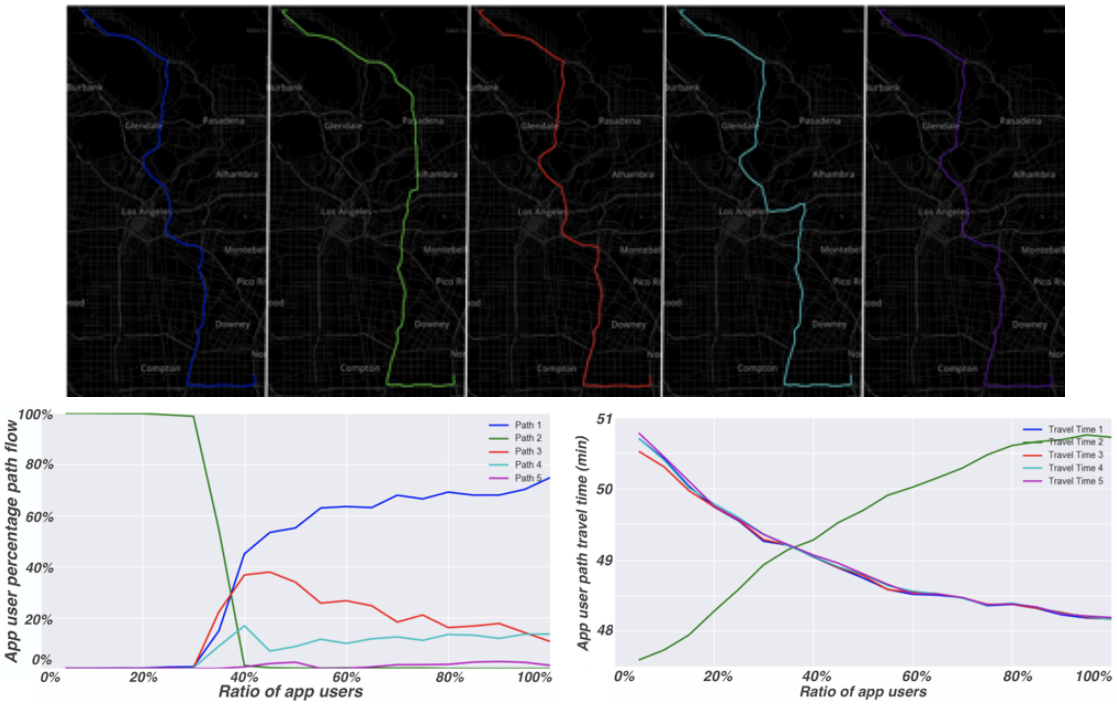


Figure 3.5: Impact of the increase of app usage on path choice and path travel time for a specific *od* pair with the increase of app usage. On top: The 5 main paths are used by app users. The blue path is the path used by non-app users. At the bottom right: the travel time of the 5 paths as a function of the app usage. At the bottom left: the percentage of app users on the 5 paths as a function of the overall app usage. When there are no app users, every vehicle uses the highway. The green side road is a shortcut for app users. When there are more than 35% of app users, the green path is not a shortcut anymore. This path gets congested because of other motorists that use this path for their trips. App users always use paths that have the smallest travel time. Figure reproduced from [47, Figure 6].

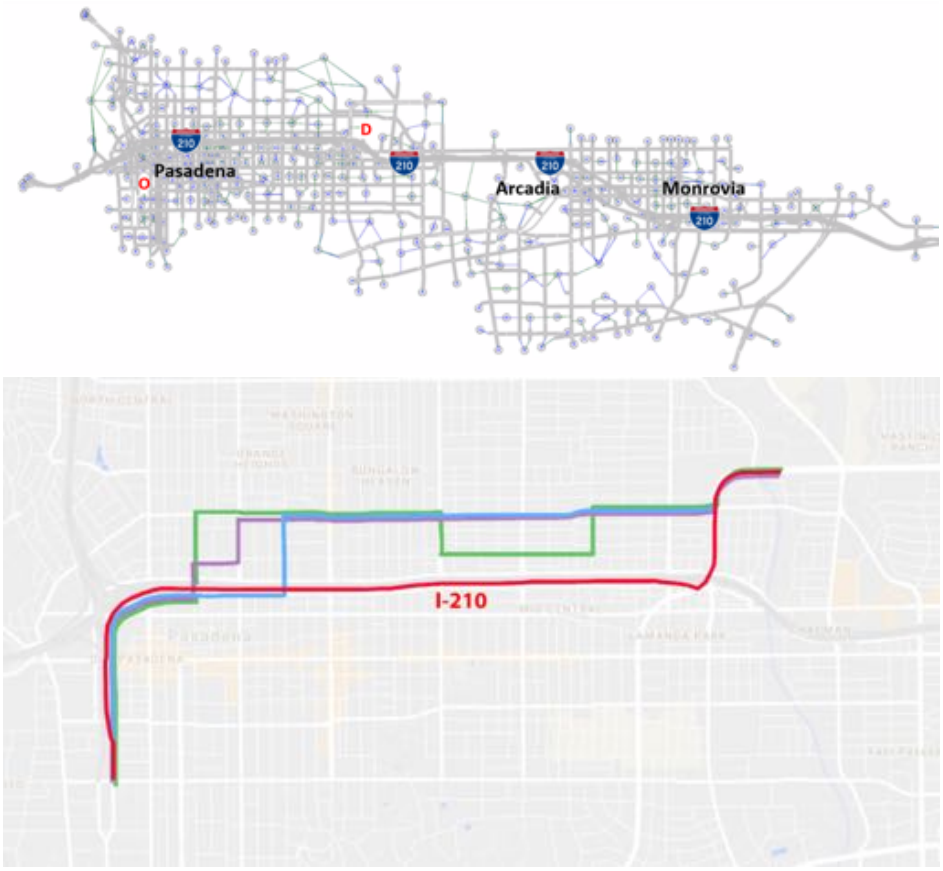


Figure 3.6: On top: The selected *od* pair for the microsimulation experiments. At the bottom: A selection of alternate paths that app users took instead of the I-210 and the I-210 route (taken by non-app users) shown in red. Figure reproduced from [48, Figure 7].

TSS' *Aimsun* [227] on the I-210 corridor segment between Pasadena and Monrovia (fig. 3.6). Aimsun uses a car following model to describe the movement of individual vehicles through a network (see chapter 5 for more explanations about traffic microsimulators). The Aimsun model of the I-210 is a calibrated corridor model [179]. Data from the California Department of Transportation (DOT) freeway loop sensors and city traffic studies are used to establish realistic OD demand. Traffic control plans from the California DOT, Arcadia, and Pasadena are incorporated into the model. The Connected Corridors project is a fundamental component of creating response plans for incident response and congestion mitigation in the I-210 corridor. As a result, the Aimsun model of the I-210 realistically simulates the evolution of traffic over the network.

We explicitly model the effect of information on routing behaviors by considering app users and non-app users. We assume that apps suggest to the app users the route that they

estimate to be the fastest one. We also assume that app users follow the recommendation of the apps. App users select their fastest paths based on the state of the network. They are allowed to dynamically reroute based on real-time information. We assume that the paths induced by the direction signs are designed to be the path obtained by solving the static user equilibrium. Therefore, assuming that non-app users mainly follow direction signs, they are routed based on the static traffic assignment. Non-app users are unable to change routes during the simulation. We expect to see non-app users spend more of their trip on the I-210, whereas app users may be encouraged to use alternate routes that make use of arterial roads to avoid highway congestion.

Since path costs (i.e., path travel times) are an essential component of app user behavior, these costs have to be updated frequently in order to guarantee that vehicles are routed based on up-to-date travel time information. A low frequency of cost updates (e.g., updating path costs every 20 minutes in a 60-minute simulation) will lead to undesired effects. For instance, assume that a path has low travel time (i.e., a low cost) due to low traffic flow. App users will start routing themselves onto this path, which will lead to the congestion of the path. However, the cost of the path is not updated (since the cost update frequency is low) so app users will continue to route onto the path, further worsening the congestion. To prevent such effects, we use a one-minute cost update frequency.

Definition 3.2.1 (Cost update frequency)**.** *The cost update frequency is the frequency at which the path costs are updated in the Aimsun simulation. In Aimsun, the cost update frequency is referred to as the routing cycle time interval size (or cycle time).*

The I-210 corridor is composed of over $4,000$ *od* pairs and more than $10,000$ links. There are four major highways in the I-210 corridor, namely the east/west-bound I-210, the north/south-bound I-605, the north/south-bound California 101, and the east/west-bound I-134. Background flow from a typical weekday morning peak hour is obtained from PeMS and city data collected for the Connected Corridors project [242]. During the peak hour, over $75,000$ vehicles enter the network hourly. As in the benchmark scenario, we fix the demand between *od* pairs and then perturb the percentage of app users between a single *od* pair as shown in fig. 3.6. We start with 10% app users and increase to 90% using 10% increments. We focus our analysis on a single *od* pair because the complexity of the network is high and therefore results are difficult to interpret when all *od* pairs are perturbed simultaneously.

The path flows and travel times of the main path (freeway) and alternative paths (shown in orange and green in fig. 3.6) converge as the number of app users between the specific *od* pair increases (fig. 3.7). The average deviation incentive shown in fig. 3.7 also decreases as the percentage of app users increases. The average deviation incentive increases slightly at 30%, 50%, 80%, and 90% app users. This phenomenon is caused by a discrepancy between predicted travel time and experienced travel time when many vehicles use the navigation applications. In the simulation, the suggestions given to the app users only take into account current and past information, it does not forecast future traffic conditions. As a result, large numbers of app users will reroute themselves onto alternative routes with low predicted costs

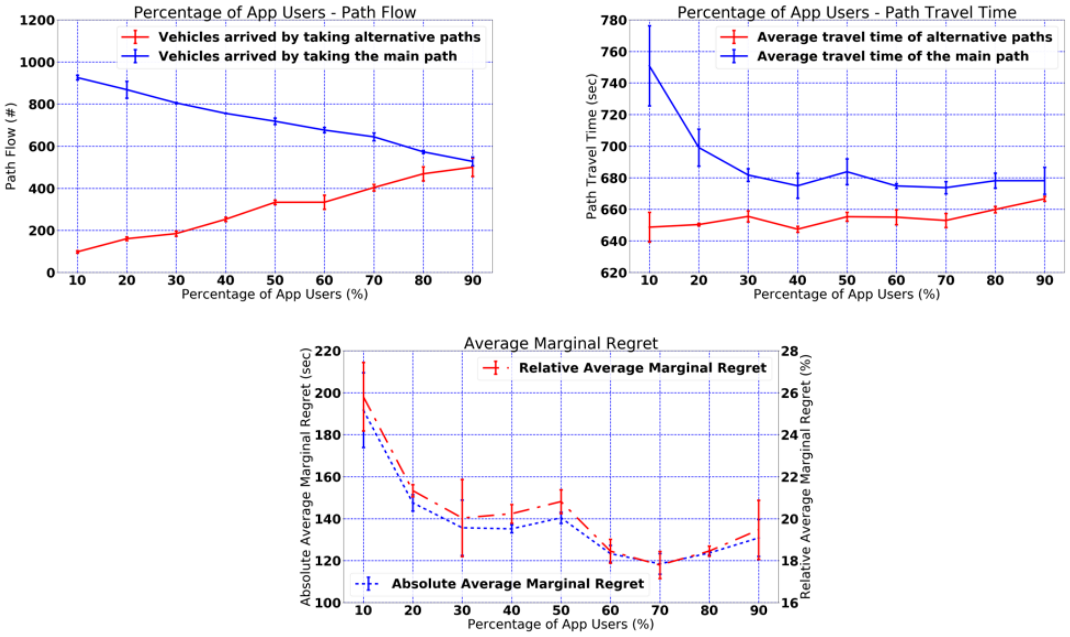


Figure 3.7: The I-210 network simulation results. On the top left: Path flow on the main path (freeway shown in red in fig. 3.6) and all alternative paths (shown in blue, purple, and green in fig. 3.6). On the top right: Path travel time convergence between the main freeway path and all alternative paths. At the bottom: Absolute and relative the average deviation incentive as percentage of app users in the network increases. The average deviation incentive shows that after 30% of navigational-app usage, the efficiency of the apps' predictions decreases if they do not take into account their own impact on the evolution of traffic. Figure reproduced from [45, Figure 3].

that will turn out to have high experienced cost, because too many are using them. In order words, **after 30% of navigational-app usage, the efficiency of the apps' predictions decreases if they do not take into account their own impact on the evolution of traffic**.

As a conclusion, **information-aware routing behaviors steer traffic to a Nash equilibrium**.

3.3 Case studies of information-aware routing behaviors

Section 3.2 shows that an increase in information-aware routing behaviors steer the state of traffic to a Nash equilibrium. However, it does not explain if it increases or decreases the

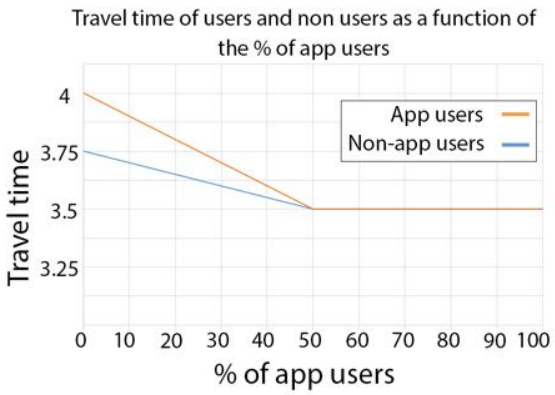


Figure 3.8: Impact of the increase of app usage on the first Braess network scenario: everybody gets a better travel time when the app usage increase.

overall network efficiency. In this section, we show that the answer of whether an increase in app usage increases or decrease the overall network efficiency cannot be answered: it is specific to each scenario considered.

On the Braess network [40], we show that an increase of app usage might decrease the individual travel time of each vehicle (both for app users and non-app users) in the network. But it might also increase the individual travel time of each vehicle. The phenomenon is known as the Braess paradox [40], a replica of the prisoner's dilemma [230] in routing games.

When information-aware routing make things better for everyone

This section shows on a toy example that the increase of app usage could lead to a situation better for everybody (even for the non-app users).

To this end, let us consider the restricted path choice model (definition 3.1.13) on the Braess network (fig. 1.2) with four nodes (A, B, C, D) and three paths (ABD, ACD and $ABCD$). The demand is set to 100 vehicles which want to go from A to D. Link travel time functions (definition 1.2.7) are $t_{AB}(x) = 1 + \frac{x}{100}$, $t_{AC}(x) = 2$, $t_{BC} = 0.25$, $t_{BD}(x)(x) = 2$ and $t_{CD} = 1 + \frac{x}{100}$. This is equivalent to the experiment of [69].

App users are routed on the shortest path between $ABCD$, ABD and ACD. Non-app users are assumed to not know the path AC and BD. Therefore, they are routed on the shortest path $ABCD$.

We observe that the travel times of both app users and non-app users decrease as a function of the percentage of app users (fig. 3.8).

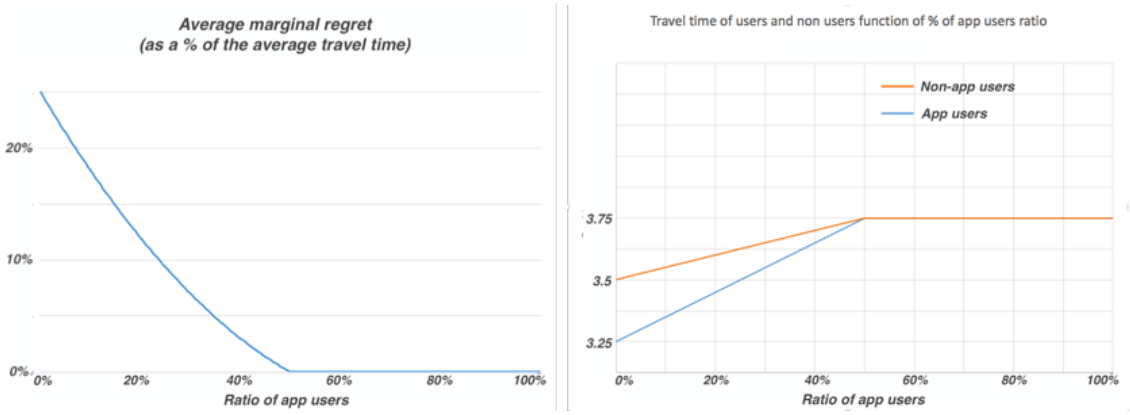


Figure 3.9: Impact of the increase of app usage on the second Braess network scenario: everybody gets a worse travel time when the app usage increases. On the left: the average deviation incentive as a function of the app usage. The average deviation incentive of the vehicles decreases monotonically when app usage increases. On the right: the travel time of app users (blue) and non-app users (orange) as a function of the app usage. The travel time of every traveler (non-app users and app users) increases when app usage increases. Figure reproduced from [47, Figure 8].

When information-aware routing is worse for everybody

In his book [184], Pigou introduced the notion of **selfish routing** through an example where the traffic equilibrium is not socially optimum (i.e., where spreading congestion over the network does not decrease the average delay in the network).

In this subsection, we provide a thought-experiment where the information-aware routing brings a modern version of William Forster Lloyd and Garrett Hardin's *Tragedy of the Commons* [148], where the cows have been replaced by vehicles, the grass has been replaced by road capacity, and the action of eating the grass is now embodied by the vehicles' route choice (see section 1.1.2 of [173]). Indeed, section 3.2 shows that the increase of app usage leads the traffic to converge to a Nash equilibrium. It is known that Nash equilibria traffic assignment can be worse than the socially optimal traffic assignment; this is the price of anarchy [167, 181, 195]. We claim that routing apps can reproduce the Braess paradox [40]. In the Braess paradox, adding a new road might increase total travel time in the network, even without an increase of the facility demand.

To this end, we consider the same setup as the previous subsection (fig. 1.2), but with a different knowledge of the network $\mathcal{P}^{na}$ by non-app users. App users are routed on the shortest path between $ABCD$, ABD and ACD. Non-app users are assumed to not know the link BC. Therefore, they are routed on the shortest path between ABD and ACD.

The increase of travel times of both app users and non-app users as a function of the percentage of app users in shown in fig. 3.9.

In this particular case, even if the usage of apps decreases the average deviation incentive (definition 3.1.6), the travel time of every vehicle increases with an increase of app usage. This toy example shows that the use of apps might not be beneficial for society. Indeed, the chapter shows that app usage leads the state of traffic to converge to a Nash equilibrium (using the average deviation incentive as quantifier) which might not be socially optimal.

Urban planning consideration

This chapter was focused on the impact of shortest-time routing on the road traffic network efficiency. It shows that the increase of information-aware routing steers traffic to a Nash equilibrium. This phenomenon might decrease or increase the total travel time in the road network. However, this chapter was mainly focusing on the hourly traffic road capacities (definition 2.1.2) and not the daily urban planning road capacities (definition 2.1.1). Ultimately, navigation apps influence the distribution of vehicles across space and time within the transportation system, without taking into account urban planning road capacities, or taking into account the differences between roads and streets. In doing so, routing guidance systems have a nontrivial impact on the magnitude and distribution of the external costs of road transport. The external costs (or externalities) might include a decrease of some traffic operational performance measures (like total vehicle-miles traveled), a decrease in travel time reliability, a possible decrease of network resiliency in case of misleading guidance suggestions, an increase in infrastructure damage, or even a decrease of a neighborhood quality of life, or an increase of safety concerns, among others. Chapter 4 studies the negative externalities of cut-through traffic induced by information-aware routing behaviors from the residential point of view.

Chapter 4

Cut-through traffic due to information-aware routing and residential anger

> "Apple Maps: Our artisanal cartographers hope you enjoy this pleasant journey. 28 min
>
> Google Maps: Our algorithm has determined an optimal path for the most efficient route given curent traffic conditions. 25 min
>
> Waze: Drive through this dude's living room. 17 min."
>
> *Anonymous tweet*

During the 2010s decade, information-aware routing had changed the state of traffic (see chapter 3), mainly by spreading congestion away from congested areas. In 1920, Arthur Pigou [184] showed that information-aware routing (or selfish-routing) can lead to a sub-optimal state of traffic due to the price of anarchy [195]. Besides a possible decrease in the transportation operational performance in the Pigou paradox thought experiment (section 3.3), information-aware routing provokes cut-through traffic on some local roads. Section 4.1 showed that both travel time equalization (Wardrop equilibrium condition, see definition 3.1.7) and the resulting cut-through traffic are observed from ground data in Los Angeles, CA. Regrettably, on some low-capacity local roads, the cut-through traffic due to information-aware routing leads to an excess of the daily urban planning road capacity (definition 2.1.1) by the facility demand (definition 2.1.5). As predicted by Thomas

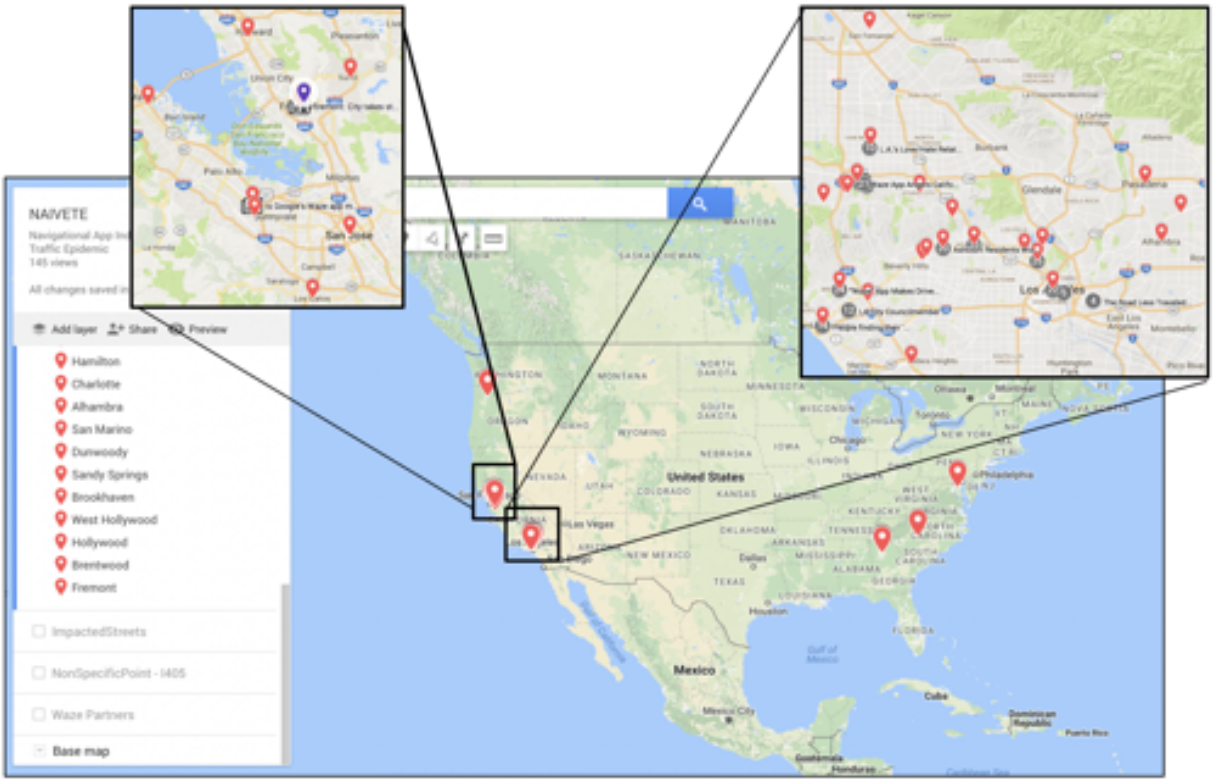


Figure 4.1: Sample of the occurrences of the negative externalities of the information-aware routing in the popular media in the US and specifically Northern and Southern California. The figure is adapted from [48, Figure 1].

Malthus [155], this leads to a negative impact on the system efficiency. Indeed, congestion incurs fixed, non-market, and indirect costs for society [147], which can lead to anger from residents negatively impacted by traffic (section 4.2). On the ground, in neighborhoods across the U.S. and abroad, residents alerted the public to their experiences. Countless reports in the media (see fig. 4.1) exhibited serious safety risks to related cut-through traffic on local roads induced by information-aware routing, in addition to degradation of public health caused by noise and air pollution, and congestion effects such as delay and poor travel time reliability. In the long term, these externalities can be mitigated by decreasing the traffic demand (definition 2.1.4) or increasing the facilities supply. But, in the short-term, or in urban areas where decreasing traffic demand or increasing the facility supply is challenging, the mitigation of the negative externalities due to information-aware routing can only be tackled by impacting the route choice of the vehicles. Unfortunately, as of 2022, few traffic-management tools can impact route choice; some of which are presented in section 4.3.

4.1 Measuring the cut-through traffic through induction-loop detectors

Due to information-aware routing, vehicles – that would usually stay on the main road for their trips – might use the fastest path to reach their destination. For example, fig. 4.2 shows that in Fremont, CA during peak hours, routes using side roads might be 10 minutes faster than routes using the congested highway. As a consequence, some vehicles take such shortcuts, changing traffic patterns and changing the facility demand. Eventually, it creates traffic on local roads in residential areas. If the new facility demand induced by cut-through traffic in local roads exceeds the daily urban planning road capacity, then negative externalities of traffic – such as degradation of public health caused by noise and air pollution, safety risks from the high chance of traffic collisions, and congestion – occur.

Travel time equalization showcases that information-aware routing behaviors occurs in Los Angeles, CA

In this section, a case study – borrowed from [48] – is performed along the eastbound I-210 corridor in the Los Angeles, CA basin. Because of the geography of the corridor, information-aware routing moves some traffic from the highway to routes parallel to the I-210. To study cut-through traffic due to information-aware routing behaviors, five routes – routinely suggested by Google Maps in 2017 – that lead from northwest Pasadena, CA to Azusa, CA (in northeast Los Angeles, CA) are considered for the study, one along I-210 and four alternative routes (see fig. 4.3).

Using travel time data from INRIX (fig. 4.3), we see that on March 10th, 2014, the travel time on the freeway increased from 3:30 pm to 4 pm until it is not beneficial to use the freeway anymore. At this time (4 pm), the travel times of all paths increased until 4:20 pm, and remained almost equal between 4 pm to 5 pm. This travel time equalization between the paths between 4 pm to 5 pm translates that the state of traffic is in a Nash equilibrium. This is expected if some vehicles in the network follow information-aware routing suggestions.

Likely, the state of traffic before 4 pm was already in a Nash equilibrium, even if the travel times on each path are not equal. Indeed, all the vehicles, which traveled from the origin to the destination shown in fig. 4.3, used the highway. Rerouting on local roads only happens when the traffic is highly congested.

This equalization of the travel times translates into the spread of freeways' congestion to parallel side-roads, as shown in chapter 3 in fig. 3.1. If the spread of congestion from highways to side-roads can be seen as an increase in the network efficiency (as stated by the public relations departments of the companies developing navigational apps [218, 248]), it was shown that this phenomenon might generate high flow on low-volume capacity roads [224, Figure 6].

Human-based knowledge (newspaper articles, resident complaints), tracking of the navigational apps route suggestions (fig. 4.2), and consulting case studies [177] have qualitatively

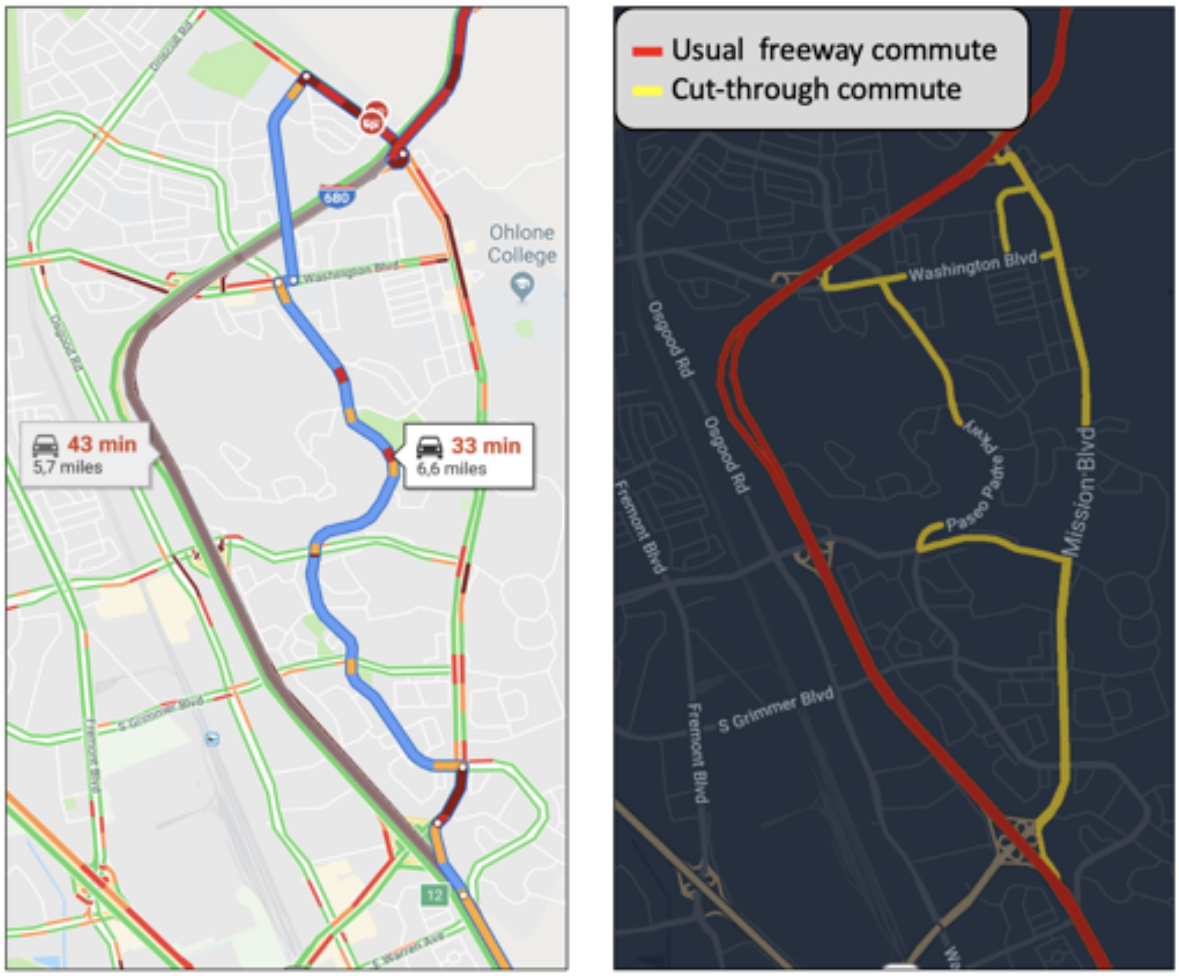
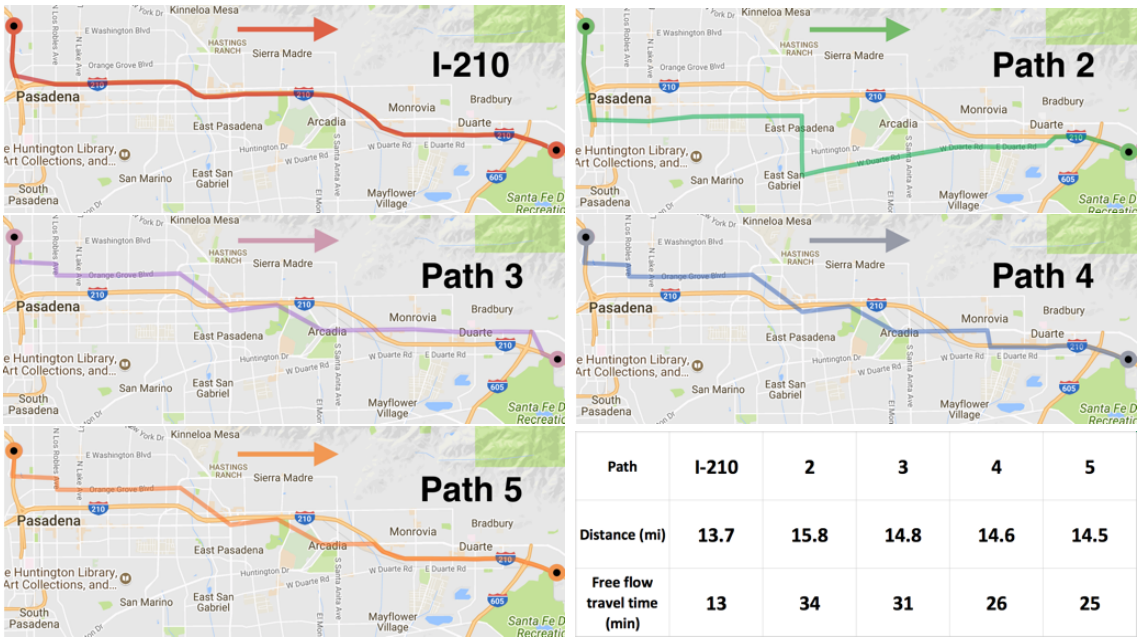


Figure 4.2: Cut-through traffic in Fremont, CA. On the left: Screenshot of directions suggested by Google Maps in July 2019 to go from the south to the north-east of the map. The distance between the west and the east side of the map is around 2 miles (3 km). On the right: Fremont, CA map showing the usual freeway commute and the cut-through commute routes.



Path	I-210	2	3	4	5
Distance (mi)	13.7	15.8	14.8	14.6	14.5
Free flow travel time (min)	13	34	31	26	25

Figure 4.3: I-210 freeway section and four alternative arterial paths considered for the travel time analysis. These paths have been chosen among the routinely suggested routes provided by Google Maps for the experiment because they include primarily arterial roads. The bottom-right sub-figure shows the distance and the free flow travel time of each path. The figure is reproduced from [48, Figure 2].

demonstrated that app usage increases cut-through traffic. To quantify how much the increase in traffic flow is due to navigational app usage and how much is due to an increase in traffic demand, traffic data is needed. Without traffic data, one cannot quantify how substantial the cut-through problem is.

Road traffic data

Traffic engineers distinguish two types of traffic data [229]:

- Traffic counts are **cross-sectional data**. At a specific location, a detector counts the number of vehicles going through a given road section (see fig. 4.5). It provides traffic engineers with the traffic flow of the road section depending on the hours of the day (definition 1.2.5).
- The trajectories of every or a subset of vehicles that are inside the road network are **trajectory and floating-car data** (see fig. 4.5). It requires vehicle identification and tells to traffic engineers how the traffic flow moves inside the network (definition 1.2.6).

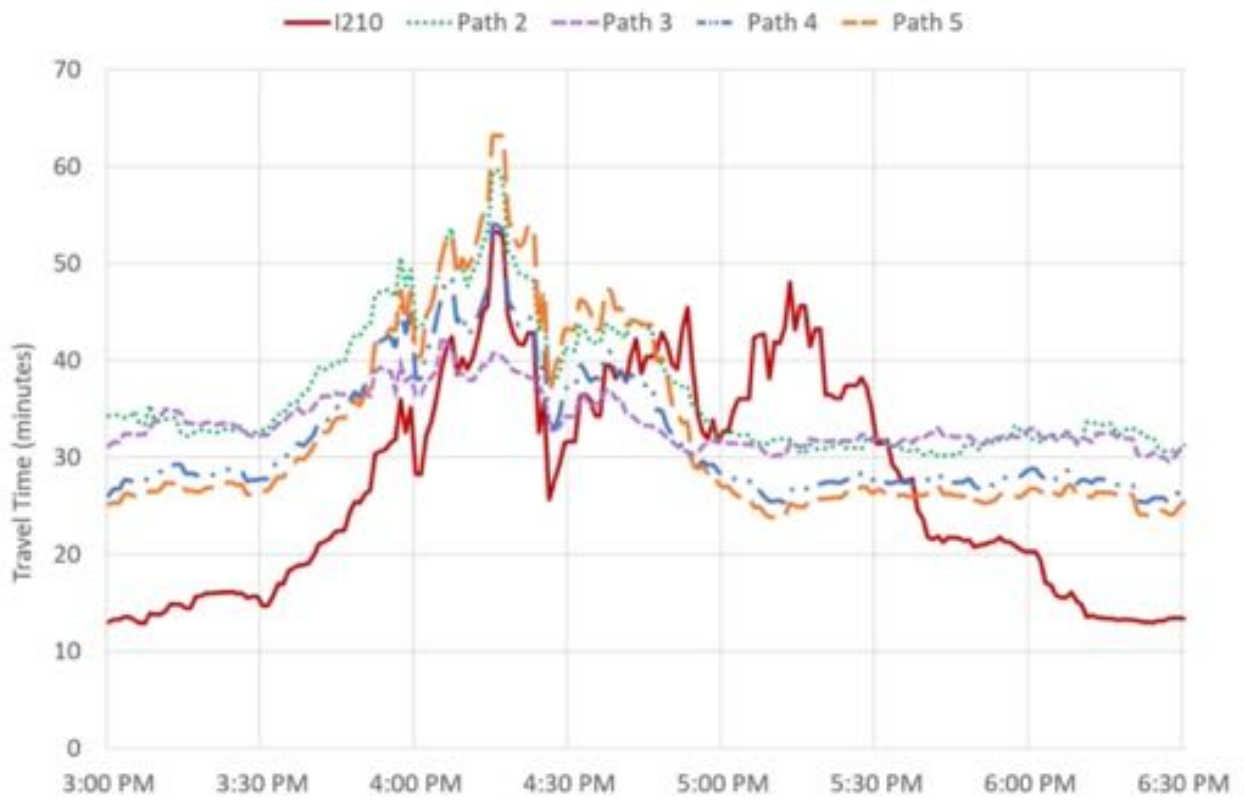
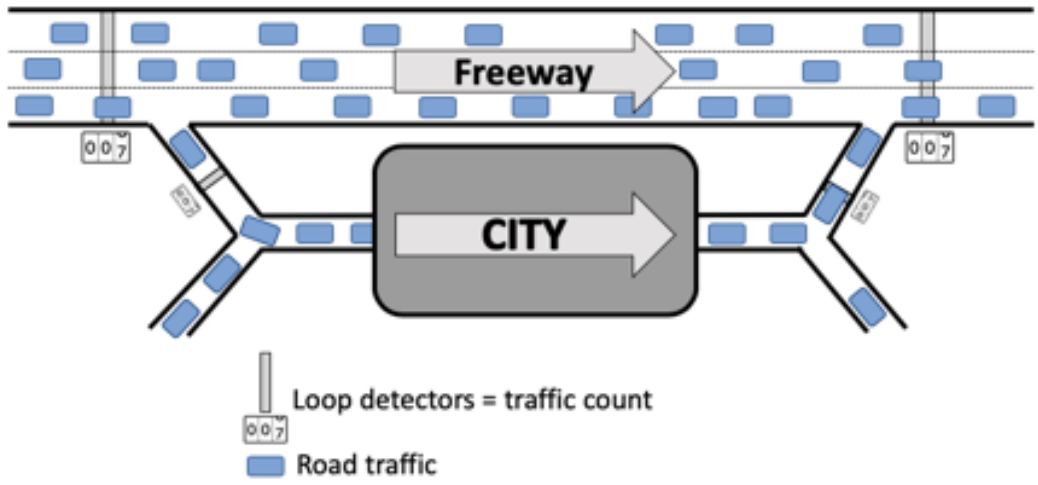
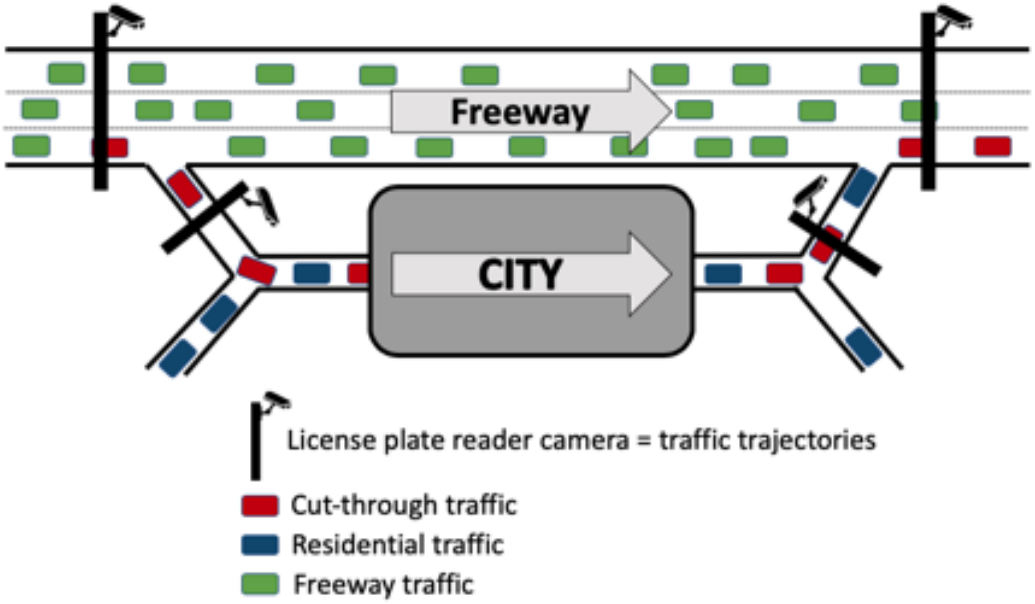


Figure 4.4: Evolution of travel times on the paths parallel to the I-210 considered (fig. 4.3) on March 10^{th}, 2014. At the beginning and end of the peak hour (3-6 PM), the travel time on the freeway is close to the free flow travel time, and the freeway is faster than the arterial routes. However, when the freeway travel time increases, arterial detours become beneficial alternatives. During high congestion, drivers can reroute themselves to arterial roads to reduce their travel time, leading to travel time equalization among parallel routes. The figure is reproduced from [48, Figure 3].



(a) Cross-sectional data



(b) Trajectory data

Figure 4.5: Schematic illustration of cross-sectional, floating-car, and trajectory data. (a) Cross-sectional data can be obtained through loop detectors. It counts the number of vehicles going through specific points. Aggregated in a system, the data can estimate the number of vehicles on any road section. However, it cannot explicitly show the existence of cut-through traffic, as it does not render the routes of the vehicles. (b) Trajectory data can be obtained through license plate reader cameras. It identifies every vehicle and derive its route. Combined with a system, it can estimate the number of vehicles on any route of the network. It can explicitly show the existence of cut-through traffic (i.e., the vehicles in red).

Cross-sectional data suggests that information-aware routing increases traffic on local roads near the I-210 in Los Angeles, CA

Cross-sectional data can be seen as traffic counts: it gives the traffic flow at a specific location at any time [229]. From a fluid mechanics point of view, cross-sectional data is called Eulerian data. It might also be called static data.

Cross-sectional data can be obtained from several types of sensors [129], including inductive Loops, show in fig. 2.4 magnetometers, magnetic sensors, microwave radars, shown in fig. 2.4, active infrared sensors or laser radars, passive infrared sensors, ultrasonic sensors, acoustic sensors, and video Image Processors, show in fig. 2.4.

Combined inside a system, cross-sectional data can indicate the traffic flow on almost every road section of the network (called link flow allocation). For example, current traffic flow on California highways is monitored by PeMS [243]. PeMS data, which consists of flow data from inductive loop sensors, are publicly available.

PeMS data analysis (fig. 4.6) shows that, during the evening peak hours, the median off-ramp flow during March weekdays on the Michillinda Avenue and Baldwin Avenue I-210 exits had seen a 1.5 and 3-fold increase respectively between 2013 and 2017. Both exits are located along the I-210 path shown in fig. 4.3, and they were commonly suggested by applications like Google Maps in 2017. While some of this increase can be explained by an increase in demand for Arcadia, it is unlikely that this can be explained solely by demand growth. Additionally, since these exits were frequently recommended by navigation-apps that had increased in popularity during the same period, the increased flow can be partially explained by app usage. Nevertheless, cross-sectional data does not reveal the cut-through caused by the drivers using exit 31.

Remark 4.1.1 (PeMS data quality)**.** *PeMS data is not exploitable for the Fremont, CA case, because the off-ramp loop detectors on the I-680 to Mission Street boulevard are almost broken (only 14% of the traffic is observed on the Off Ramp VDS 403255 – Mission Blvd off diag – I680-N from Tue 03/11/2014 00:00 to Mon 03/11/2019 11:59:59).*

In addition to observing the effect of navigational apps during peak hours of a single day (fig. 4.4), or examining off-ramp flow increase on the congested highway across the years, we examine INRIX speed data over the 2014 and 2015 years on the I-210 alternate paths (fig. 4.3). As the number of app users increases, it is expected that the travel time on arterial streets will increase as well, due to increased flow rerouting around congestion on the freeway. The average travel time on the paths was computed during peak hours for each week from January to June each year. As expected, in one year, the travel times along the alternative paths increased by roughly 20%, i.e., around five minutes (fig. 4.7).

Figure 4.7 also shows that the I-210 was always faster on average during a day than the arterial roads. This gives an additional hint that the rerouting phenomenon only happened during peak hours (as it can be similarly guessed from fig. 4.4).

The travel time on the I-210 path oscillated around 15 minutes over the two years, remaining roughly constant over time. This occurrence can be explained by *latent demand*

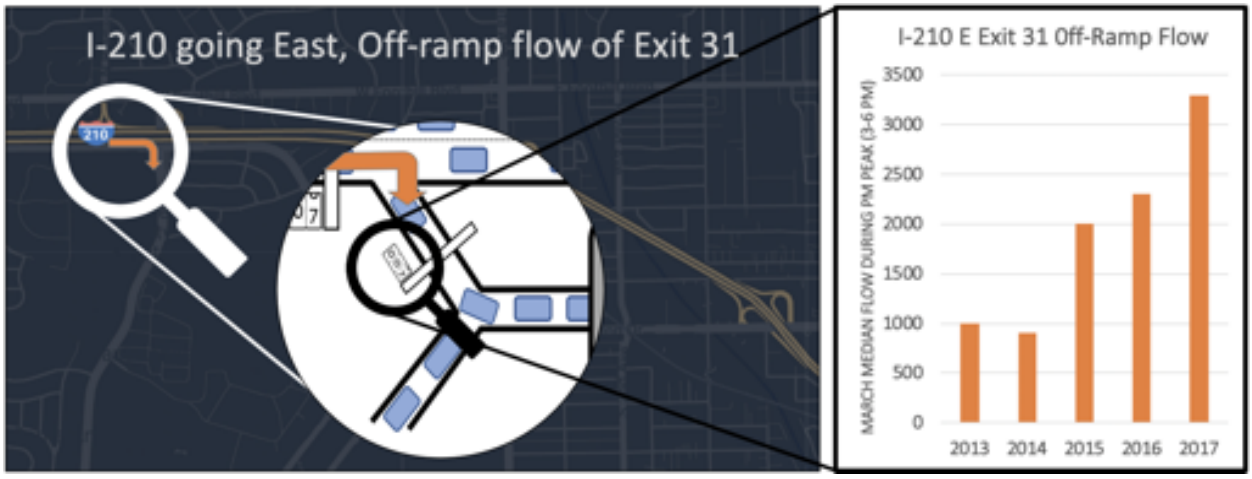


Figure 4.6: On the left: Location of the off-ramp traffic detector of exit 31 of the I-210 East in Arcadia, CA. On the right: Evolution of the median off-ramp flow from I-210 during March weekday peak hours between 2013 and 2017. The 3-fold increase in off-ramp flow that occurs between 2014 and 2017 coupled with the decrease in speed on parallel paths provides evidence in favor of app-induced arterial rerouting patterns. The figure is adapted from [48, Figure 4].

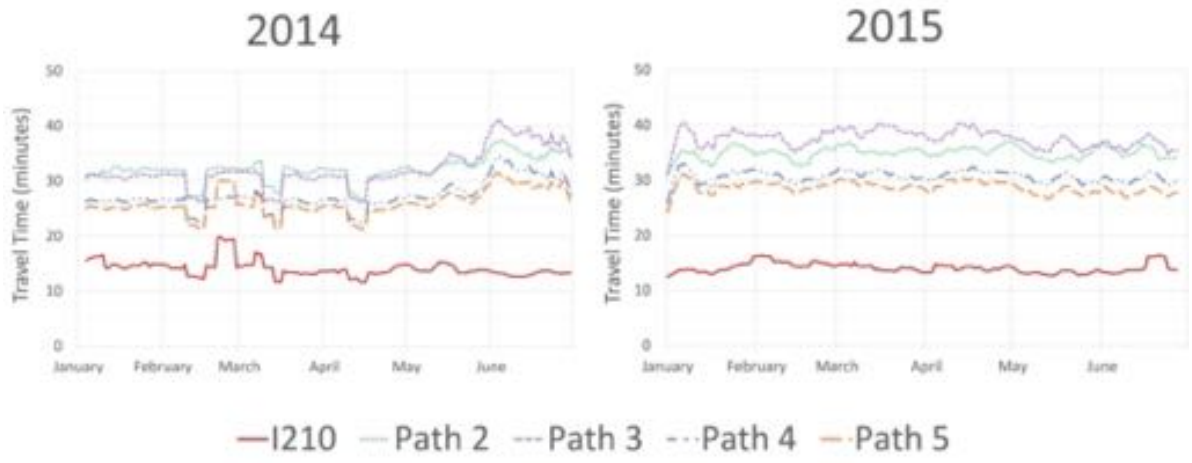


Figure 4.7: Evolution of the average travel time computed with INRIX data on the five paths during peak hours (4:30 to 5:30 PM) considered (fig. 4.3) in 2014 and 2015, for each week from January to June. While the travel time on the I-210 remains roughly constant over two years, alternative paths suffer a 20% increase in travel time. The observed drops in 2014 may be irregularities from data flaws. The figure is reproduced from [48, Figure 4].

or by the very low marginal cost of extracting vehicles from the freeway (when a typical freeway lane capacity is around 2000 veh./h). Either, all the vehicles that are removed due to information-aware routing behaviors are replaced by vehicles that were not traveling before because of congestion on the highway. Either, the amount of vehicles removed is too small when compared to the total number of vehicles using the highway to impact travel time on the highway.

If cross-sectional data gives significant clues (i.e., travel time equalization, and I-210 off-ramp 3-fold increase over 4 years) about the existence of cut-through traffic, it does not indicate where the people come from or go to (called path flow allocation) (see fig. 4.5). Therefore, the severity of cut-through traffic due to an increase in navigation-app usage and induced information-aware routing has never been globally quantified due to lack of publicly available trajectory data.

Floating-car data are needed to quantify the magnitude of cut-through traffic due to information-aware routing

Trajectory data is the trajectories of every vehicle that is inside the road network. Floating-car data is a subset of the trajectory data: it is the trajectories of some vehicles inside the road network. From a fluid mechanics point of view, trajectory data is called Lagrangian data. It might also be called dynamic data.

Current trajectory and floating-car data requires identifying the vehicles inside the network. It can be obtained from license plate readers, camera vehicles tracking with drones or fixed cameras (like Next Generation Simulation data [241]), GPS devices that give GPS position in real-time of vehicles, or GPS traces that give past GPS position of vehicles [229].

High penetration floating-car data can explicitly show the existence of the cut-through traffic caused by navigational app users. The data will show the cut-through used by the navigational app users, it will also reveal the number of people using the different cut-through routes. In Fremont, a traffic study [192] using floating car data showed that, during the 2019 afternoon rush hours, 10% of the traffic supposed to take the I-680 were using the Mission Boulevard or the Paseo Padre Parkway instead. In Pleasanton, a traffic study [123] showed that cut-through traffic accounted for 84% of the traffic on the Laurel Creek drive in 2016.

Currently, the author is not aware of any traffic control center or any city traffic engineers in the U.S. or in Europe that have access to such data. This is mainly because getting floating car data requires vehicle identification, which might be expensive and can raise privacy concerns.

However, smartphone apps record the location of people through smartphones. Therefore, apps have access to floating-car data (see fig. 4.8).

To improve the quality of traffic control strategies implemented by cities, cities might want to ask the navigational apps for floating car data from the app users. This might require new regulations or new partnerships between cities and navigational apps. France has already passed a regulation to ask navigational applications to share their traffic data [14]. However,

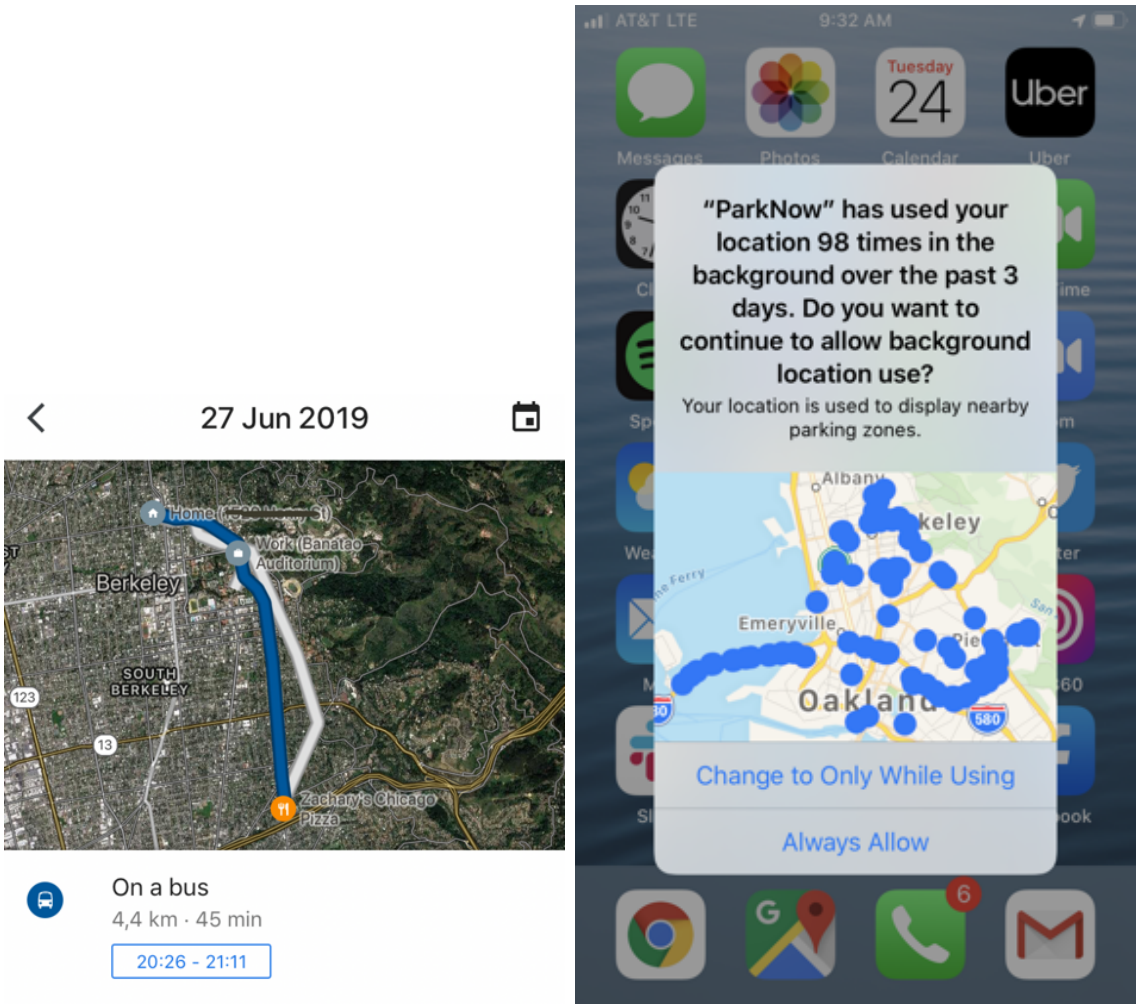


Figure 4.8: Google Maps location and transportation mode tracking. Screenshot from the author's iPhone in June 2019.

there are obvious privacy issues that might disable navigational applications to share any user data with any government entities. Also, a lot of navigational apps are trying to store fewer users' location data on their servers to increase users' privacy [103].

4.2 Materialization of the cut-through traffic at the residential level

As shown in section 4.1, one of the consequences of information-aware routing is that traffic congestion is spread onto local roads near congested highways. It becomes a challenge when some of these roads receive a higher traffic flow than they are designed to sustain (i.e., when the facility demand on some local roads exceeds their corresponding urban planning road capacity). It eventually turns into a nightmare for the residents living near these streets. Inherently, this is due to the fact that navigational-apps consider the hourly traffic road capacity when suggesting routes to their users without taking into account the daily urban planning road capacity. In other words, the shortest-path algorithm does not consider any other path characteristics than its travel time. For example, it does not include explicit knowledge about streets with a lot of pedestrian crossings, locations of elementary schools, or difficult intersection crossings.

The geometry of the roads is designed as a function of the expected flow that they would receive [7] based on their category (fig. 2.8). As shown in fig. 2.8, in the U.S., local roads are designed to receive less than 400 vehicles per day on average in rural areas and 700 in urban areas [8]. A countless number of complaints emerged against the negative externalities of cut-through traffic due to information-aware routing (see fig. 4.1).

Enumeration of some negative externalities due to cut-through traffic

On the one hand, some commuters save a few minutes due to the usage of apps. On the other hand, navigational apps create high traffic flows on local roads that lead to many negative externalities for the communities affected by cut-through traffic. These negative externalities can include higher travel times, delays, unreliable travel times for residents [90], noise, gas emissions, traffic accidents (fig. 4.9), decreased accessibility in affected neighborhoods, wrong directions [131], and infrastructure damage, etc.

Cut-through traffic changes the spatial distribution of noise levels and emissions in the neighborhood [174, 87]. From an individual's long-term perspective, congestion increases psychological stress [246, 174] and decreases people's leisure time, which consequently decreases people's physical activities. Many studies and media reports showed that road traffic systems have a strong association with the decreased mental well-being of commuters or residents who live close to the road: traffic's unpredictability is a significant source of stress,

and it can easily disrupt daily routines; the air and noise pollution produced by traffic further add to such frustration and anxiety [174].

Infrastructure damage. The cumulative load on the road depends on the number of vehicles using it and their weight. Infrastructure damage increases with the number of vehicles using the road [171, 128, 221]. Recognizing that current road users usually adopt various transportation modes including driving, cycling, and walking, policymakers adjust the component of transportation funding such that everyone bears the cost of transportation [226]. There are three major funding resources (cost of driving): (1) general taxes, such as income and sales tax ; (2) gas taxes; (3) revenue from tolls and fares [226, 34]. In the U.S., most of these revenues go to the states, therefore, the states mainly pay for the transportation infrastructures. Especially, the states are in charge of maintaining highways; the roads designed to be used by commuters. The states are not in charge of maintaining local roads. Because small cities cannot afford damages caused by heavy usage of their local roads, if the commuters use some local roads, then the state should subsidize the maintenance of these roads.

Safety concerns. Based on a survey of road safety performance, speeding is the number one road safety problem in many countries, often contributing to as many as one-third of fatal crashes and serving as an aggravating factor in most crashes [52]. In Leonia, NJ, one pedestrian died struck by a bus in 2014 on Ford Lee Road, a road with a lot of cut-through traffic [91].

Of 69 newspaper articles about negative externalities of information-aware routing, 39 of them discuss safety issues due to cut traffic that the author analyzed (a subset of the articles shown in fig. 4.1), 23 of them deal with congestion issues, and 16 report noise and quality of life issues.

Remark 4.2.1 (Feature-specific routing)**.** *Note that some navigational-app might take some road characteristics in their routing suggestions, like tolls, fuel consumption, or safety [187]. For example, in 2021, Google Maps launched eco-routing to account for fuel consumption [104, 13].*

The negative externalities due to cut-through traffic are all over the world

Cut-through traffic and the resulting negative externalities due to information-aware routing have been reported all around the world. To name a few cities, San Francisco, CA, Boston, MA, Medford, MA, Tel Aviv, Israel, Leonia, NJ, Los Angeles, CA, Fremont, CA, Bordeaux, France, Lyon, France, Meudon, France, have been vocal about these issues [152].

In early 2018 residents of Echo Park, Los Angeles County, CA began reporting excessive thru-traffic on Baxter St.; one of the steepest streets in America. Drivers unfamiliar with the 32% grade of Baxter St. are in danger of collisions with difficult-to-see oncoming traffic and

Figure 4.9: Bus crash because it was routed on a road too steep. Photos by Ingrid Peterson via Flickr.

brake failures. Many vehicles even spun out onto neighbors' gardens [215] (see fig. 4.9). The rerouting typically occurred during peak hours, when southbound vehicles on Alessandro St. were routed onto N. Alvarado St. to avoid the build-up of traffic caused by the merging of both the Glendale Freeway and Alessandro St. onto Glendale Blvd. The app users were directed to turn left onto Baxter St. from Alessandro St. to bypass the bottleneck by traveling down either N. Alvarado St. or Lake Shore Blvd. instead of Glendale Blvd. In May 2018, Echo Park converted the two blocks of Baxter St. on either side of N. Alvarado St. into disjoint one-way roadways to prohibit through traffic from driving uphill on Baxter St [54]. In Los Angeles, the Sherman Oaks neighborhood residents also reported high traffic due to cut-through [23].

In Leonia, NJ (9,200 residents), the traffic flow on Fort Lee Road increased from 4,000 to 14,000 vehicles per day between 2014 and 2017 according to Mayor Judah Zeigler [175]. If the traffic is uniformly distributed over 11 hours, it represents a traffic flow that **increases from 6 vehicles entering the road every minute to 24 vehicles entering the road every minute**! Fort Lee Road is an urban collector street. Therefore, it is designed to receive a daily traffic flow lower than 6,300 vehicles per day, which represents 9 vehicles

entering the road every minute [232] (see fig. 2.8). Consequently, safety concerns were raised by the residents [124]. Indeed, a pedestrian died in 2014 on Ford Lee Road after being struck by a bus [91].

In Lyon, France, as of 2022, vehicles coming from the south might use the Quai Jean-Jacques Rousseau to avoid congestion on highway A7. In Bordeaux, France, the Avenue de Courrejean is used by some vehicles wishing to short-cut the congested highway A62 coming from the south. In Meudon, France, the Rue d'Arthelon is used by some commuters reaching Paris from the southwest suburbs that do not take the congested N118 highway.

In Fremont, CA, faced with increased congestion due to altered commuter behaviors, Mission San Jose neighborhood residents encounter heavy traffic while commuting, but also when doing mundane tasks like picking children up from school or grocery shopping (see fig. 4.10). Mission San Jose is a neighborhood of the City of Fremont, CA, a suburb of the San Francisco Bay Area located in Alameda County. The community is situated at the southern entrance to the Sunol Grade, a mountain pass for Interstate 680 that links the major job centers of Silicon Valley to the southwest with bedroom communities to the northeast. Due to its position at the entrance to one of the most heavily-trafficked transportation corridors in the region, traffic next to the neighborhood along Interstate 680 has always been severe [4]. However, with the rise of real-time detours pioneered by GPS-navigational apps, more and more of this traffic has been routed onto local neighborhood streets over time [176]. Despite the traffic-calming measures implemented by the city of Fremont, the Mission San Jose neighborhood still suffers from heavy traffic. This problem has been exacerbated in recent years by an increase in job growth in Silicon Valley, which might have counterbalanced the improvement due to the city's policies [225]. With the continued increase in congestion, residents have started blaming the concern on city policies for allowing the development of new housing units and office parks nearby. The cut-through traffic challenges are induced both by information-aware routing and high traffic demand.

Aggravating factors of the negative externalities of cut-through traffic: travel and traffic demand

As stated before, the negative externalities of cut-through traffic are due to a facility demand (definition 2.1.5) that exceeds the daily urban planning road capacity of local roads in the affected areas. Because, the facility demand is at the end of the trip chain ((fig. 2.1), any increase in the travel demand (definition 2.1.3) or the traffic demand (definition 2.1.4) will aggravate cut-through traffic.

Increase of the travel demand

The travel demand is globally increasing due to several factors. First, the world population is increasing: 2,536 millions in 1950, 7,380 in 2015 [237]. So, as a consequence of having more people, they are more travelers. Secondly, people are more clustered in cities. Urbanization (percentage of people living in urban areas) is growing everywhere in the world: in 2020,

Figure 4.10: Congested Fremont, CA neighborhood near I-680. The picture was taken from a drone camera and published in [61, Page 1].

55% of the world's population lives in urban areas, and in 2050, the number is expected to be 68% [238]. As a consequence, the increase in travelers is focused on metropolitan areas, creating more travel demand for these specific areas. An indirect consequence of urbanization – that increases, even more, the travel demand – is the development of urban sprawl, leading to an increase in the average commute distance. For example, in France, the median commute distance has increased from 7 km in 1975 [223, Table 2] to 8 km in 1982 [223, Table 2], 13 km in 1999 [70, Table 6], and 15 km in 2013 [70, Table 6].

Fremont, CA is a striking example where the increase in travel demand has largely exacerbated cut-through traffic (fig. 4.11). Fremont, CA is located in the San Francisco Bay Area, to the east of Silicon Valley. Between 2014 and 2018, 152,000 new jobs have been created in Silicon Valley. However, only 28,000 housing units have been created in the Valley. Of the 152,000 new workers of the Silicon Valley, 86,000 live north or east of Fremont, CA, increasing the number of commuters between the East bay and the Silicon Valley. On the I-680 going north, a cut-through given by apps like Google Maps is Mission Boulevard, Fremont, CA (see fig. 4.2). In 2018, the traffic on Mission boulevard was so heavy that even small businesses have complained about the decrease in business accessibility.

Increase of the traffic demand

As a result of the increase in the travel demand, the traffic demand is increasing all over the world. However, the traffic demand is also increasing independently of the travel demand. Indeed, traffic demand increases faster than travel demand because of the mode-shift implied by **transportation network companies** (TNCs) and the increase in **on-demand delivery**.

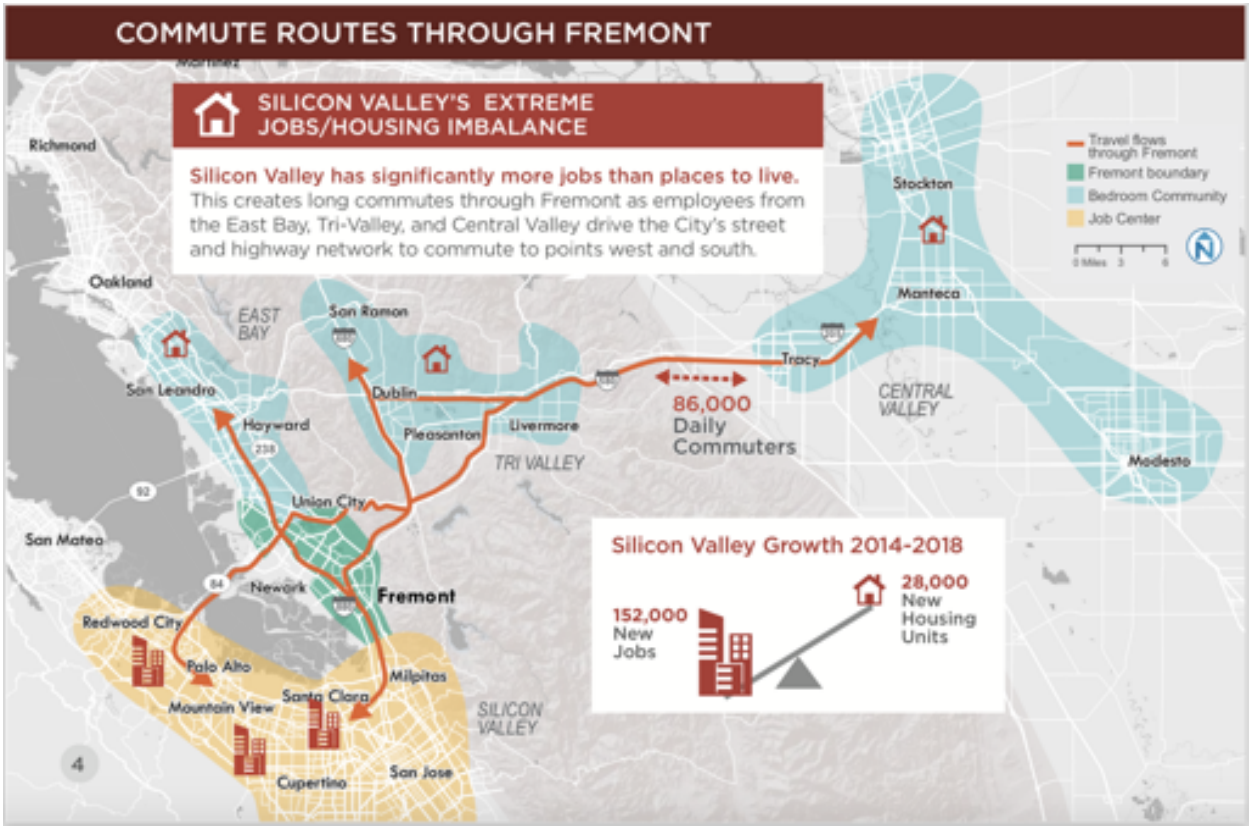


Figure 4.11: Schematic illustration of traffic created in Fremont, CA due to the demographic increase in the San Francisco Bay Area. Screenshot from March 2019 Fremont Mobility Action Plan [225, Page 4].

Traffic demand had considerably increased due to the sudden emergence of transportation network companies (TNCs) like Uber and Lyf [44]. Uber was created in March 2009. Ten years later, in 2019, Uber was estimated to have 110 million worldwide monthly users [212]. Statista states that 996.7 million people used TNCs in 2019 [212]. These new users – almost 1 billion new users in 10 years – often represent new cars on the roads. As reported by the Boston Metropolitan Area Planning Council, in 2018, 59% of trips made by TNCs are adding vehicles to the road network (see fig. 4.12 or [97, Figure 11]). Accordingly, **the worldwide increase of traffic demand due to TNCs can be approximately estimated to be 588 million annual riders** (56% of 996.7 million TNCs users). In San Francisco, CA, Lyft, and Uber were responsible for 13% of the combined Vehicle Miles Traveled inside the city in 2019 [20].

Concurrently to the increase of the traffic demand due to TNCs, another increase in the traffic demand is due to the development of **last-mile delivery**. According to the economic

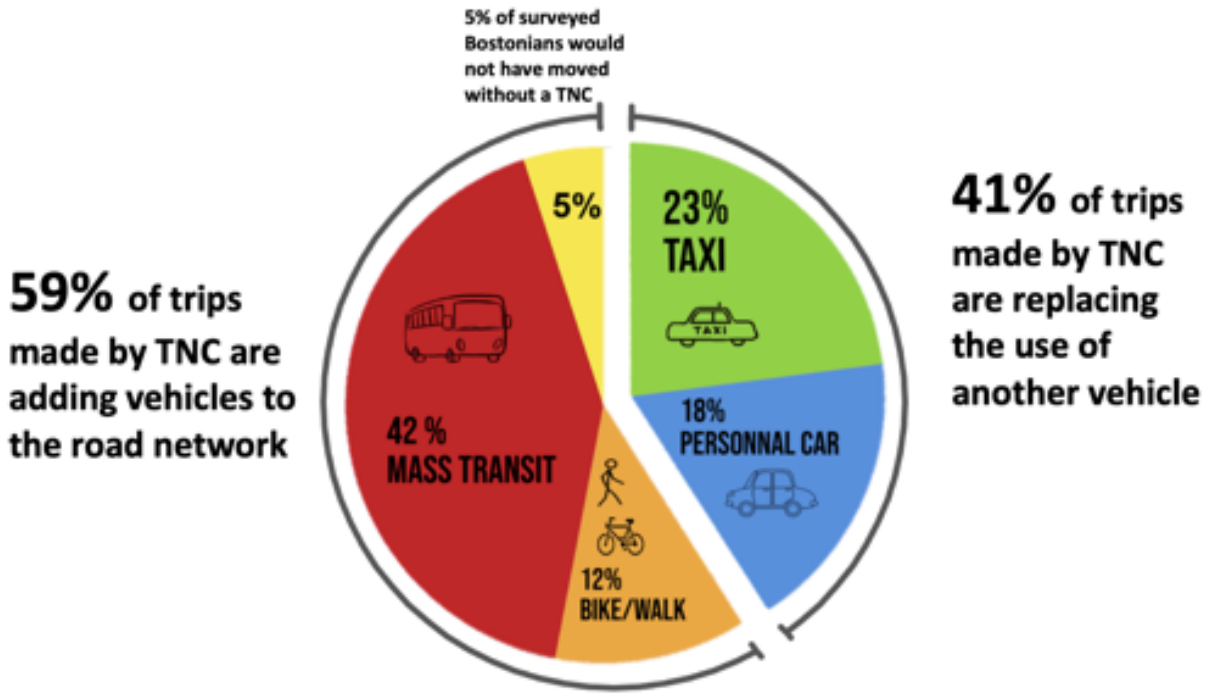


Figure 4.12: Mode of transportation replaced by transportation network companies (TNCs) like Uber and Lyft in Boston in February 2018 [97, Figure 11].

research department of the Federal Reserve Bank of St.Louis [10], e-commerce retail sales had heavily grown during the 2010s decades. Representing 1% of the total sales in 2000, it represented 10% of the total sales in 2019. Consequently, last-mile delivery had densely increased the number of trucks and small delivery trucks on the roads. Brookings's analysis of Federal Highway Administration data [42] shows that trucks – which used to drive 40% of the time on urban roadways (and 60% on highways) in 1966 – drove 60% of the time on urban roadways in 2014. This leads to an **8-fold increase in urban vehicle miles traveled by all trucks between 1966 and 2014**.

4.3 Cut-through traffic mitigation techniques

To mitigate the negative externalities of cut-through traffic, cities can use several approaches, including the ones listed in table 2.2. Externality is a concept from economics where the free market leads to some indirect costs to society. Interestingly, the concept was first introduced by Pigou in [184] in 1920, using the example of road traffic and cut-through traffic due to information-aware routing. Many answers to mitigate negative externalities can be found in

the economics literature. As stated in section 4.2, the negative externalities of cut-through traffic can also be alleviated by decreasing the traffic demand in the network.

Mitigation of negative externalities in economics

In the economics literature, several types of solutions to decrease the externalities of a market are suggested [60]:

- **Cap-and-trade**: assigning property rights for supply usage.
- **Pigovian tax**: taxing the market that generates the externalities.

In the specific case of the externalities due to navigational-app usage, if a regulator cannot directly control the navigational apps (which is the case in free economies), then the regulations and policies to mitigate the externalities cannot differentiate app users and non-app users. Adding a tax that is applied to everyone but should target a specific set of users is known as imperfect taxation [107].

Cap-and-trade

A first approach to mitigate cut-through traffic is to ban it. In this approach, the road can be restricted to non-resident traffic. In economics, this approach was theorized by Ronald Coase in [67]; property rights are assigned to the use of supply.

In the U.S., Leonia used this approach, NJ. On January 22nd, 2018, under the pressure of inhabitants of the city, the city closed 60 streets to thru-traffic during rush hours [145, 175]. Any commuter going through Ford Lee Road during the rush hours was fined by the local police. However, in the U.S. the roads are public. Therefore, on January 30^{th}, 2018, a resident from a nearby city filed a complaint accusing Leonia of "infringing on her access to public roads and violating the public's right to freedom of travel" [188]. The road was opened back in August 2018 after the invalidation of the restriction ordinances by the state Superior Court [133].

In France, some roads can be marked as residential-only [68]. Commuters do not have the right to use residential-only roads. However, because turning a local road into a residential-only road requires a lot of administrative work, and can only be done for some specific roads, the French Parliament decided to directly regulate the navigational apps in 2021 [15]. In Tel Aviv, Israel, some residents gathered to create a group that filed suit against one navigational app [256].

In economics, the cap-and-trade approach is relevant when tools exist to mitigate the externality, and when the externalities can be evaluated properly. If the externalities are not correctly evaluated, then the cap-and-trade approach might negatively affect the market without being aware of it. For example, in Leonia, NJ the restrictions on non-residents created issues for local businesses because it decreased the city's accessibility and, as a consequence, its attractivity [217].

When the system is complex, and accurately estimating the externalities of the market is arduous. Instead of putting hard constraints on the market, softer constraints can be put through the use of taxes.

Internalizing the externality with Pigovian taxes.

Another way to reduce the negative externalities of a market (or market failures) is to internalize them. More specifically, a regulator can tax the costs of the externalities to the market participants. Such taxes are called Pigovian taxes [203], as they were derived from Arthur Pigou's work [184]. For example, a sugar tax that discourages unhealthy diets and offsets the economic costs of obesity is a Pigovian tax.

In the case of information-aware routing, an obvious Pigovian tax consists of charging commuters for using a specific route. As an illustration, between Monterey, CA, and Carmel-by-the-Sea, CA, the 17-miles drive is tolled to keep quiet the residential area around it [233].

Lombard Street pricing. Another example is Lombard Street in San Francisco, CA. In the 2010s, according to a study led by the San Francisco transport authority, the "crooked street" welcomed over 2 million visitors per year [200]. During a busy day, the scenic section of Lombard Street (between Hyde St and Leavenworth St) was overcrowded by more than 8 thousand pedestrians and 2.5 thousand vehicles. This swarming congests the nearby areas, especially above the winding part (spilling back above Larkin St). Residents complained about the mess led by the visitors' spillover: indeed, according to a survey, the noise, the littering, and trespassing incidents were making that place less bearable than a common residential area [200]. To counter those issues, many propositions had been proposed. For instance, a short-term proposition (made by residents through a survey) was to penalize the tour companies that were contributing to the super-concentration of visitors and to be stricter on people's behavior (add more officers and raise the fines in that area). There was also a proposition of putting signs in different languages to dissuade visitors from using that road. A long-term solution was to implement cut-through pricing to regulate the number of comings and goings. In June 2014, the "crooked street" was completely closed for vehicles (except for residents) during the 6 most crowded hours of the day (12:00 PM to 6:00 PM). As a result, there was no more congestion above the problematical section of the street, and the pedestrians were more numerous [200]. Regulating the number of cars and remodeling the infrastructure seemed to be a solution to counter cut-through issues. To that extent, San Francisco's Board of Supervisors unanimously OKed state legislation on April 16th, 2019, that requires people who want to drive down the street to make a reservation and pay a fee ($5 or $10 depending on the day). The bill was accepted by both the Californian assembly and senate with a large majority [49]. However, the bill was later vetoed by Gov. Gavin Newsom to prevent social equity issues [106]: the road should be "available to all, regardless of their ability to pay".

Electronic Road Pricing (ERP) in Singapore. In Singapore, electronic road pricing enables the government to tax the usage of given roads during specific hours of the day [101]. Combined with license plate readers, the ERP system in Singapore can recover the routes of vehicles in the network. Therefore, this system is able to influence any vehicle's route choice through specific financial incentives. However, the author believes that due to potential privacy threats, it is unlikely that such systems will exist in the U.S. or Europe anytime soon.

Speed limit decrease as an imperfect Pigovian tax. When thinking about the travel time as a cost of travel, increasing travel time on some local roads can also be interpreted as a Pigovian tax. For example, the city of Fremont installed speed bumps in 2016 to slow down traffic and avoid cut-through in the Mission San Jose neighborhood [26]. The traffic management strategies presented before were selective: controlled access (like in Leonia, NJ), or road pricing for non-residents (like the 17-miles drive between Monterey, CA and Carmel, CA) only impacts the commuters. However, decreasing the speed limit increases the travel time for every vehicle, penalizing the residents as much as commuters. This uniform impact over the vehicles is due to imperfect tax differentiation [77]: it impacts every vehicle the same way, even if cut-through travelers have a higher negative externality on local roads than residential drivers.

The cost of imperfect differentiation in traffic congestion management

In this subsection, we introduce a benchmark example to understand the cost of imperfect differentiation when decreasing speed limit in order to calm traffic. Let consider the network shown in fig. 4.13.

The network is a benchmark for Fremont, CA, the link l_1 represent the I-680 highway, and both link l_2 and l_3 represents the Mission boulevard, a minor arterial urban road used as a cut-through by some commuters. We use the static traffic assignment (section 3.1) to represent the traffic allocation in the network. The commuter demand is set to $d_{od_1} = 20$, the resident demand is set to $d_{od_2} = 1$. There is no other demand.

We consider three routes $p_1 = od_1$, $p_2 = od_2d_1$, and $p_3 = od_2$. We do not consider any other route. All the resident demand is using the route p_3, and the commuter demand should be allocated between p_1 and p_2. We denote $h_1 = h_{p_1}$, $h_2 = h_{p_2}$, and $h_3 = h_{p_3}$

We consider that the city of Fremont can change the speed limit on the Mission boulevard section l_2. Therefore, the link travel time functions (definition 1.2.7) are given by:

$$\begin{aligned} t_1(\mathbf{h}) &= 8 + f_1 \\ t_2(\mathbf{h}) &= 2 + \boldsymbol{\pi} + 2f_2 \\ t_3(\mathbf{h}) &= 4 + 2f_3 \end{aligned}$$

Where $\boldsymbol{\pi} \in \mathbb{R}_+$ is a variable that Fremont can set to update the speed limit on l_2, and $\mathbf{f} = \Delta\mathbf{h}$.

Figure 4.13: The benchmark Fremont game network.

We denote $\gamma = \frac{h_1}{d_{od_1}}$ the percentage of commuters using the I-680 freeway. The user equilibrium condition (definition 3.1.7) gives that $\gamma = \min\{1, 0.8 + 0.1\boldsymbol{\pi}\}$.

We consider that the city of Fremont would like to set the speed limit (i.e., find $\boldsymbol{\pi}$) such that the travel time of the resident is minimized and that no commuters use local roads. A tradeoff needs to be made by the city when setting the speed limit: a high speed limit means fast travel time for the resident but high cut-through traffic, while a low speed limit can decrease cut-through traffic but penalizes the residents who drive. We model the cost that the city of Fremont minimizes as:

$$J_F(\mathbf{h}) = h_2^2 + 2t_2(\mathbf{h})$$

In this set up, $\boldsymbol{\pi}^\star = 5$ minimizes the cost of the city. In this case, $\gamma = 0.85$, $t_2 = 25$ and $t_1 = 15$. The results are presented and discussed in fig. 4.14.

The variable $\boldsymbol{\pi}^\star$ can be interpreted as a Pigovian tax. The commuters are suffering from the Pigovian tax. If $\boldsymbol{\pi} = 0$, then $\gamma = 0.85$ and the commuters' travel time would be $t_1 = 24$. Therefore, the commuters have a deadweight loss of 1 [114].

Ideally, the city could minimize the amount of cut-through traffic while maximizing the speed for the residents. This could be done with route-based pricing, for example. In this context, the usage of p_2 would be heavily taxed, while the usage of p_3 would not be taxed at all. However, in the scenario shown, the city cannot differentiate between the commuters and the residents when setting the speed limit on Mission boulevard (l_2). Decreasing the speed limit on Mission boulevard is a Pigovian tax with imperfect differentiation [77]. The optimal tax would be such that $\gamma = 1$ and $\boldsymbol{\pi} = 0$. In this case, the resident would have a travel time of $t_2 = 4$. Therefore, the cost of imperfect differentiation in this benchmark example is 11 [77].

Instead of increasing the cost of the market participants (the vehicles) that create externalities, a regulator can subsidize market participants that do not create externalities. For example, a city can subsidize public transportation to decrease the traffic demand.

Reducing the facility demand through road facilities management

Because the cut-through traffic is a facility demand issue, a city can manage its road facilities to decrease the facility demand. As of 2022, the author mainly found that decreasing the speed on cut-through roads (Pigovian taxes) and adding access restrictions (cap-and-trade) have been implemented to mitigate cut-through traffic.

To **decrease the speed of cut-through roads**, cities can:

- Set up **stop signs**, as done by Fremont, CA (see fig. 2.5) [121].
- Set up **speed bumps/humps/lumps**, as done by Passadena, CA [156] or partially by Fremont, CA [26].
- Reduce the **speed limits** [52]. This can be achieved by setting up low-speed alerts or low-speed zone around schools or senior apartments [214].

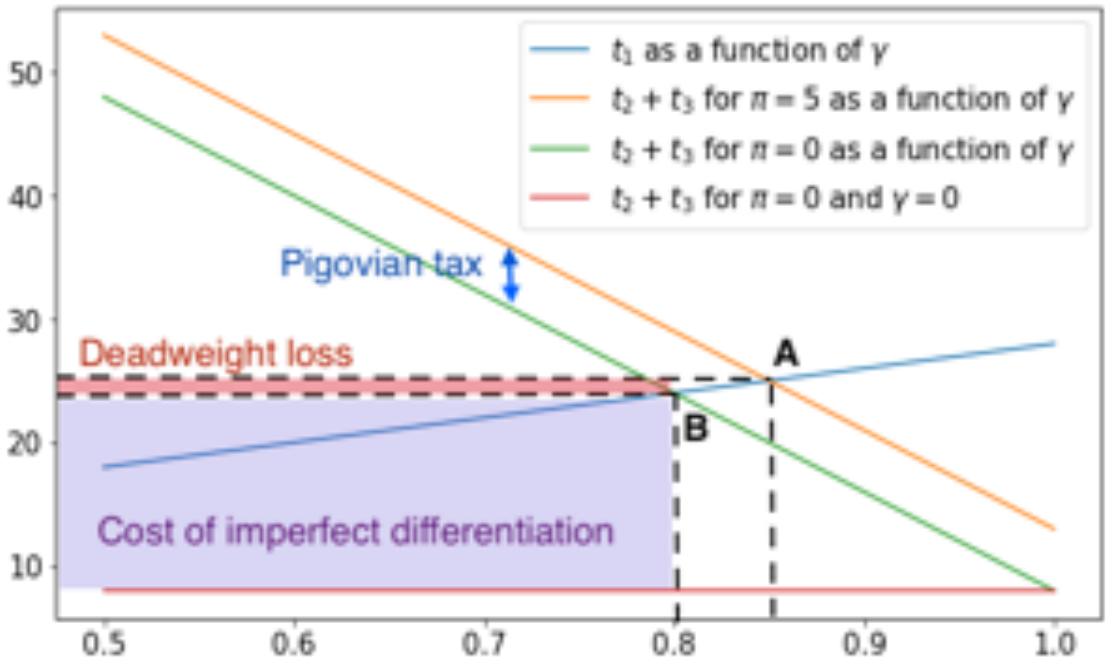


Figure 4.14: The Fremont game results. The cost of the path p_1 as a function of the percentage of commuters using the I-680 freeway γ is shown in blue. This cost does not depend on the action $\boldsymbol{\pi}$ of Fremont. The other curves represent the cost of the path p_2 for different scenarios. The user equilibrium condition is translated by the equalization of the travel time between p_1 and p_2. On the figure, this equalization happens at the intersections of the curves (points A and B). When the market is not regulated ($\boldsymbol{\pi} = 0$), the cost of p_2 is shown in green. In this case, the user equilibrium is located at the point B, where $\gamma = 0.8$ and $t_1 = 24$. When the city of Fremont best regulate the market ($\boldsymbol{\pi} = 5$), the cost of p_2 is shown in orange as a function of γ. In this case, the user equilibrium is such that $\gamma = 0.85$ and $t_1 = 25$. The optimal regulation would be such that $\gamma = 1$ and $\boldsymbol{\pi} = 0$. In this case, the cost of p_2 is independent to γ, and it is equal to 9 as shown in red on the figure. In economics, the difference in the cost of p_2 between the free market scenario and the scenario with regulation is referred to as the deadweight loss. It is the y distance between A and B (equal to 1 in this scenario). The difference in the cost of p_3 between the scenario with regulation, and the scenario with optimal regulation ($\gamma = 1$ and $\boldsymbol{\pi} = 0$) is the cost of imperfect differentiation. As an example, the cost of imperfect differentiation for p_2 can be seen as the distance between B and the red curve (equals to 15 in this scenario).

- Change the **timing of the signals** [81], as done by Fremont (see fig. 2.5).

Access restriction can be done using:

- **Diverters**; Berkeley, CA uses diverters to forbid the usage of some roads. This had been done to redirect traffic on main roads in a very grid-like traffic network with small residential roads.
- **Turn restrictions**; In Sherman Oaks' in Los Angeles, CA "no-turn" road signs have been set up to redirect cut-through traffic [23].
- **Dynamic turn restrictions**; In Fremont, CA, variable message signs [253] that indicate dynamic turn restriction have been used (see fig. 2.5).
- **Controlled access**; In Leonia, NJ, the police enforced the road closure of 60 streets to all drivers aside from residents and people employed in the borough during the morning and afternoon peak hours in 2018 [145]. The city of Leonia, NJ was soon joined by Weehawken, NJ, a nearby city [125].
- **Road closure** [234]; In Baxter street, Los Angeles, CA went even further and closed one direction of the road [54] to prevent crashes from drivers following directions from navigation apps that are inept to handle the steep incline drive.

Other tools have been implemented. For example, the city of Fremont used variable message signs to incentive drivers to not take the cut-through roads (fig. 2.5).

The facility demand can also be reduced by reducing travel and traffic demand. Some tools have been shown in table 2.2. However, decreasing the travel or the traffic demand is not something that can be done by cities alone (e.g., Fremont, CA is in the San Francisco Bay Area or Leonia, NJ is in the New York City metropolitan area). And it requires long-term thinking about urban planning, which is not directly related to traffic management.

Pleasanton case study. Before navigational apps, the city of Pleasanton, CA was already fitting against cut-through traffic on their local roads [141]. In 2000, during morning peak hours, vehicles coming from the cities east of Pleasanton (like Livermore, Tracy, or Stockton), and going to the Bay Area (especially going to the Silicon Valley) were facing congestion to reach the Southbound I-680 from the Westbound I-580 (and vice versa in the afternoon). To avoid congestion, some vehicles cut through Pleasanton local roads, creating a lot of traffic in residential areas.

Following the traffic management steps outlined in section 2.2 (fig. 2.3), the city of Pleasanton first evaluated the state of traffic using speed surveys in 2002 [33]. The city of Pleasanton then decided to implement several traffic-calming measures, including ramp metering, decreasing speed limits, and closing some roads.

Later, to answer each resident complaint, the city launched its traffic-calming program: the City of Pleasanton Neighborhood Traffic Calming Program [63]. As of 2022, the program

filters the complaints of the residents regarding speeding or high volumes and sends them to the Traffic Unit Supervisor. After determining the importance of the traffic concerns, the traffic unit supervisor ensures effective and proportionate solutions. First, speed trailers and speed limit signs are set up to decrease speeding. Then, speed radars might be introduced. If the problem persists, the city may implement speed lumps, street closures, or turn restrictions. For example, residents complained about cut-through traffic on Laurel Creek drive [123]. Traffic studies in 2016 showed that the cut-through drivers amount to 84% of the traffic [123]. The city decided to implement the first set of traffic-calming measures, which lead to a 60% decrease in the cut-through traffic. To have a larger impact, the city eventually decided to close the Southbound Laurel Creek to all turns from Dublin Canyon.

Out-of-the-box mitigation techniques

While the city of Pleasanton addressed the cut-through traffic challenges using techniques from a traffic management textbook, other cities and residents have used out-of-the-box techniques.

Fake street signs were posted to confuse drivers. For instance, in the Sherman Oaks neighborhood in Los Angeles, CA, "no-turn" signs were installed by residents [23]. Signs that aim to raise the awareness of commuters about cut-through traffic have been observed by [196, 191].

Some residents went further and tried to spoof the apps by injecting fake data into the app, asking residents to walk slowly with the app turned on to provide slow-moving GPS points to the app, or reporting false accidents. This technique had been previously used by Technion students to highlight weaknesses of Waze against cyberattacks [172]. The web contained several discussion forums on how to "spoof" apps (or "keep apps out of neighborhoods"), see for example [6].

Part II

Simulating routing behaviors

Chapter 5

Calibration and validation of Aimsun traffic microsimulation

In order to pick the best mitigation techniques against cut-through traffic, traffic planners and operators should predict or evaluate the impact of possible of the management strategies they can implement (see fig. 2.3). Because transportation systems are complex, this might be challenging. As stated in section 2.2, cities can perform case studies (by looking at chapter 4 for example), they can use trial and errors, or they can use models and simulations to predict the impact of traffic management techniques.

In this chapter, we desire to understand the impact of information-aware routing and traffic-calming measures (e.g., traffic signal timing changes) on congestion and business attractivity in the Mission San Jose district in Fremont, CA. The different type of simulations a city can perform are presented in section 5.1. Most of the challenges of running a traffic simulation are not apparent until one creates their own. Therefore, this chapter first gives an overview of traffic microsimulations and describes the process to create, calibrate and validate them (section 5.2). To illustrate the process, a realistic, open-source simulation with a 2000-link network where traffic conditions dictate the routing behaviors of around 75,000 vehicles every day – namely, the Mission San Jose district in Fremont – is provided to researchers that seek to try out and test ideas in a similar environment [135]. Section 5.3 specifies the input data needed, and section 5.4 enumerates the steps to follow to run an Aimsun simulation [227]. Finally, section 5.5 explains the model calibration process and section 5.6 describes the output analysis that can be done after running the simulation.

The work presented in this chapter was done in collaboration with a team of UC Berkeley researchers.

5.1 Existing traffic simulator

Every year, each person living in a city in the United States loses in average 84 hours and 33 gallons (124.92 liters) of fuel in traffic congestion [150]. To address congestion, cities can

understand the impact of traffic congestion mitigation techniques using digital twins (i.e., simulations) of their traffic (see section 2.2).

As stated in section 2.2, transportation planning is inherently related and interconnected to urban planning. Specific interaction within the land use cycle (fig. 2.2) can be cast using different type of models, depending on the problem a planner might want to solve.

Land use models

Land use models [249] estimate how accessibility and attractiveness impact the land use. As an illustration, using traffic data and demographics as input, UrbanSim [245] models household, employment, real estate, and potential evolution in job location, urban density, and building constructions.

Travel demand models and simulations

Using land use data coupled with demographics data, travel demand models estimate the need for transportation [169, 201]. More specific, they estimate and forecast the travel and traffic demand (definition 2.1.3 and definition 2.1.4). These models are relevant to estimate mode-shifting opportunities or latent demand, for example. There are three main approaches of transport demand modeling.

Trip-based models [169] are the historical way to generate facility demand from sociodemographic data. In particular, since the 1950s, the classical four-step model [161] has been used extensively by planners. In this model, a trip is an origin, and a destination. Trips are aggregated across space in origin-destination demand matrix (**OD matrix**) (definition 1.2.10) using transportation analysis zones (**TAZs**) (definition 1.2.9. The model first generates a distribution of outbound and inbound trips across transportation analysis zones (TAZs) (*trips generation*). Second, the inbound trips and outbound trips across all TAZs are matched into origin-destination (OD) matrices (*trip distribution*). Third, for each *od* pairs, the modal split is estimated (*modal split*). Finally, the traffic demand is assigned to routes using the static traffic assignment (definition 3.1.12) (*route assignment*). The four-step model is a static model, that estimates total travel during a day without modeling any dynamics in the network.

Activity-based models [32] extends the trip-based models by associating to each trip a purpose (work, shopping, or leisure). This feature enables modeling heterogeneous departure times. It also enables taking into account the demand elasticity with respect to the network traffic conditions. MATSim [21], TransCAD [151], Visum, or CEMDAP [185] are few software products that enable solving activity-based models.

Finally, **agent-based models** extend the activity-based model by modeling each agent, instead of each trip [257]. Therefore, agent-based models can reproduce carpooling behaviors, and model that most commuters come back to their home after working during the day. MATSim [21], BEAM [19] (that was built on MATSim), or ActivitySim [95] are few software

products that enable solving agent-based models. Agent-based models can also be solved by researchers using their own algorithm, like [160].

The travel demand models do not represent the traffic dynamics on individual links.

Traffic models and simulators

Assigning the OD demand (definition 1.2.10) to dynamic routes is done by **dynamic traffic assignment** models [59, 231]. Once the route of each vehicle is known, computing the dynamic road section traffic loads in the network is called **dynamic network loading** [59].

In this dissertation, we refer to the traffic demand (definition 2.1.4) as the demand that travels using the road network, without a route being specified (see fig. 2.1). However, in the transportation engineering academic literature, the traffic demand might refer to a different notion: in the four-step model [161], it refers to what we called in fig. 2.1 the link facility demand (definition 2.1.5). Our choice is motivated by the fact that we are interested in understanding the impact of information-aware routing on traffic. With the rise of information-aware routing behaviors, the author believes that the **traffic demand should not include the individual route choice** anymore. Consequently, traffic models might refer to models that perform both the dynamic traffic assignment and the dynamic network loading, but it might also refer to models that only perform dynamic network loading.

Three types of traffic flow modeling exist [57, 59].

Macroscopic models aggregate vehicles into traffic flow (definition 1.2.5), making the internal assumption that traffic behaves like a fluid [59]. These models can be static (like the four-step model [161], or the static traffic assignment (definition 3.1.12)) or dynamic. Aimsun [227], Cube, POLARIS [16], Visum, TransCAD [151] are few macrosimulation software products.

Microscopic models represent each vehicle in the network [59]. Microsimulation reproduces individual driving behaviors and fine-grain traffic resolution. The interaction of each vehicle with its environment is modeled using car-following models, lane-changing models, and route-choice models [59]. Aimsun [227], Cube, TransModeler, POLARIS [16], SUMO [27], Vissim [86], MISIMLab, Synchro, BEAM [19], Paramics [50], ActivitySim [95] are few microsimulation software products.

Mesoscopic models fit in between and represent a compromise between macroscopic and microscopic modelling [59]. They model every vehicle, but only the interaction between a vehicle and the traffic flow is modeled. The interaction between a vehicle and the traffic flow can be modeled with fundamental diagrams of traffic flow, exit-functions, or queuing models [59]. Some mesoscopic models are event-based models: they model interaction between vehicles and the traffic flow as events that might not require to model time continuously. The mean-field routing game model presented in chapter 6 is an event-based mesoscopic model that uses fundamental diagrams of traffic flow. Aimsun [227], Cube, TransModeler, POLARIS [16], SUMO [27], Vissim [86], MISIMLab, Synchro, BEAM [19], Paramics [50], ActivitySim [95] are few mesosimulation software products.

In the team's knowledge, within the traffic simulations options, only microsimulations can help understand the impact of a change in traffic signal timing plans on road traffic at scale. Within the available traffic microsimulator, Aimsun is used in this chapter because it includes macroscopic, mesoscopic and microscopic simulation options, and it can be used with an academic license.

Motivation

This chapter aims to give city planners and traffic engineers the necessary tools and methodology to create, calibrate, and validate a large-scale road traffic microsimulation (i.e., where routing behaviors impact the state of traffic). An accurate traffic simulation model will facilitate transportation engineering in multiple aspects, from traffic congestion improvement [206] to applications for autonomous driving [57]. Existing literature provides a high-level overview and comparison of traffic simulation and its development [183, 9, 57, 206], the benefits and applications of traffic simulation [57], and an abstract framework for constructing a simulation [143, 236]. Existing literature also includes a closer examination of simulation models with various techniques for calibration and validation of simulations under different scenarios [154, 30, 115, 178, 220, 41]. This chapter has a similar pattern of creating a microsimulation compared to [236], but it also introduces a machine learning based method for model calibration. The team found that none of the above-mentioned literature shares any generalizable or transferable blueprint for the end-to-end process of creation, calibration, and validation of a microsimulation. Because the processes of developing traffic microsimulations are very similar across different cities, this chapter aims to provide a detailed handbook and publicly available code source for creating a microsimulation.

Note that traffic microsimulation only makes sense when fine-grain traffic data needs to be modeled and when case studies or A/B experiments are unavailable or unrealistic (see section 2.2). Traffic microsimulations cannot be used for demand analysis (such as for assessing the impact of ride-hailing companies with the respect to the number of trips [97]) or mode shift analysis [32, 169, 257]. In addition, microsimulations are not relevant when data is missing to calibrate the simulation.

5.2 Simulation overview and its creation process

Traffic simulations [57] provide traffic information visualizations and related figures, which include vehicle hours traveled (VHT), vehicle miles traveled (VMT), mean delay per vehicle, gas emission, accessibility index, etc.. It facilitates a comprehensive analysis of the design and efficiency of the transportation system in question (section 5.6).

In this chapter, we refer to traffic simulations as the simulation of road traffic (vehicle flows in the network over time), given a traffic demand (people's origin, destination, and departure time grouped into timed origin-destination matrices) and a road network (including road sections, lanes, intersections, road signs, and traffic signal timing plans) as inputs.

Simulations can be aggregated macroscopically, mesoscopically, and microscopically [57] (see section 5.1). In this work, Aimsun Next 22 [227] is used to perform microsimulations, where the focus is on the individual elements in a transportation system.

In a microsimulation, individual vehicles are generated and assigned to a route. Their movements are simulated across the road sections [57]. Before being generated, each vehicle is defined by an origin, destination, departure time, and optionally, a vehicle type (definition 1.2.4). The vehicle input data are aggregated across space and time into timed origin-destination (OD) matrices for each vehicle type (definition 1.2.10). Space aggregation uses transportation analysis zones (TAZs) (definition 1.2.9), while time aggregation uses time buckets (definition 1.2.2). Lanes of contiguous road sections are connected through unsignalized intersections (with yield or stop signs) or signalized intersections (with given traffic signal timing plans and a master control plan). Assigning each vehicle to a route is sometimes referred to as route assignment, while the simulation of vehicle movement through the network is commonly referred to as dynamic network loading [57]. Simulated link flows (definition 1.2.5) and network traversal times can be compared to ground data with which to calibrate and validate a proposed model. For the Fremont San Jose Mission district microsimulation, input data are described in (section 5.3).

Generally, when it comes to modeling transportation systems, there exists a notable tradeoff between the number of model variables and the risk of overfitting, a result of the large quantity of data needed to calibrate complex transportation models [41]. With this in mind, real data – set aside for model calibration – should be split into training and testing data to decrease the overfitting risk [255].

Before calibrating a microsimulation (section 5.5), one needs first to fix any existing network and demand issues (connectivity issues, wrong number of lanes, wrong traffic signal plans, small mistakes in the master traffic control plan, obvious error in the demand data). The first phase of calibration is done without simulation by matching simulated and ground total counts of vehicles entering or exiting the network. This is then followed up by the second phase of calibration, done through macrosimulation. Once the OD demand is calibrated, the driving behaviors (routing, car-following, lane-changing models), and microsimulation parameters (like simulation time step) can be calibrated using optimization algorithms that work with expensive function evaluations (this work uses a genetic algorithm that is highly parallelizable).

Once calibrated, the microsimulation can be validated using eyeball estimation or concrete metrics alike. Eyeballing mainly consists of understanding where and when the congestion occurs in the input network and checking for consistency with any prior knowledge about the network's congestion. Metrics of effectiveness (MOE) can then be used for a more rigorous second validation. For example, the mean delay per vehicle over time in the network indicates when the peak hour happens in the network and is a strong indicator of the global quality of the simulation. Finally, more specific data like detector flows and network traversal times can be used to validate the simulation against ground data.

Once the simulation is created, calibrated, and validated, it can be used for traffic analysis (section 5.6). For example, the causes of congestion can be derived from the simulation, and

it can be used to test congestion mitigation policies.

5.3 Microsimulator input data

The required inputs for a microsimulation are a network (definition 1.2.3) and a dataset of timed origin-destination demand (definition 1.2.10). Traffic data may also be used to calibrate and validate the simulation. The data used by the team to simulate the traffic in the Mission San Jose district around Interstate 680 and Mission Boulevard (State Route 238/262) in Fremont, CA is openly available and the process of calibrating the input data and importing it in Aimsun is reproducible [135].

Network

The road network is made up of constituent road sections connected through signalized or unsignalized intersections. To create this network, the team downloaded the OpenStreetMap (OSM) [109] network model using the bounding box defined by the following coordinates: north: 37.5524, east: -121.9089, south: 37.4907, and west: -121.9544 and first cleaned it in ArcGIS [38] (see fig. 5.1). After importing the network into Aimsun [227], Google satellite, Google Maps, and Google StreetView images were used to perform manual adjustments to ensure the accuracy of connectivity, yield and stop sign locations, and lane counts (see fig. 5.2). Speed limits were calibrated using the data provided by the City of Fremont and road capacities were adjusted using the data from the Behavior, Energy, Autonomy, and Mobility (BEAM) model which is an open-source agent-based regional transportation model [19]. Then, traffic signal plans (including the ramp meters and the master control plan) from the city and CalTrans were added using the Aimsun graphical user interface (GUI). Finally, traffic-calming measures (primarily turn-restrictions) were created in the simulator. In summary, the modeled network has 5,626 links, including 111 freeway sections, 373 primary road sections, 2,916 residential road sections, and 2,013 nodes (intersections), 313 of which have stop signs and 37 of which have traffic lights (26 operated by the city and 11 operated by CalTrans). The overall process to create and fix the network (with traffic signal plans) took our team about 600 person-hours to complete.

Origin-Destination Demand

The origins, destinations, and departure times for every vehicle are aggregated into timed origin-destination demand matrices (definition 1.2.10). Origins (or destinations) are clustered into transportation analysis zones (TAZ) (definition 1.2.9), which are bijective to the set of centroids connected to internal or external entry/exit nodes in the network. The 16 square-kilometers network area is divided into 84 internal centroids and 11 external centroids (see fig. 5.3). Departure times are aggregated into 15-minutes time intervals. Between 2pm and 8pm, 75,000 vehicles are modeled (including 45,000 commuters and 30,000 residents).

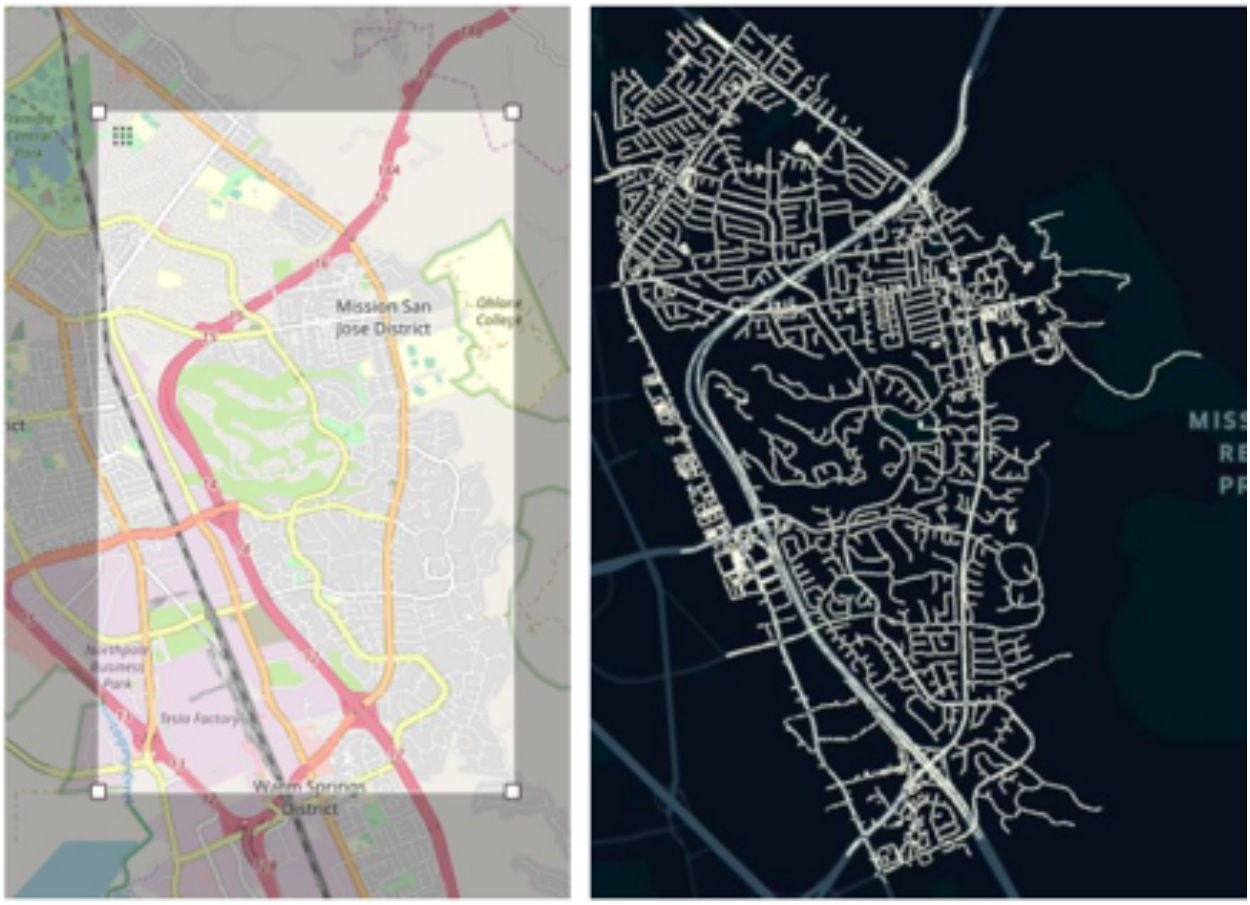


Figure 5.1: On the left: OSM network with the bounding box. On the right: corresponding Aimsun network after cleaning.

In this work, the demand data was derived from the SF-CHAMP demand model [126] from the San Francisco County Transportation Authority and from a StreetLight study [192] performed for the City of Fremont. Unfortunately, there is no reproducible process to create accurate demand data as of now, which is where most of the major challenges of realistic traffic simulation remain. However, demand data accuracy can still be slightly improved through calibration against ground data (section 5.5).

The overall process to create and calibrate the origin-destination demand took our team around 600 person-hours to complete.

Traffic data

To calibrate against ground data, one can utilize ground flow, speed, or/and travel time data, each of which can be directly imported within Aimsun as a Real Data Set. In this study, flow

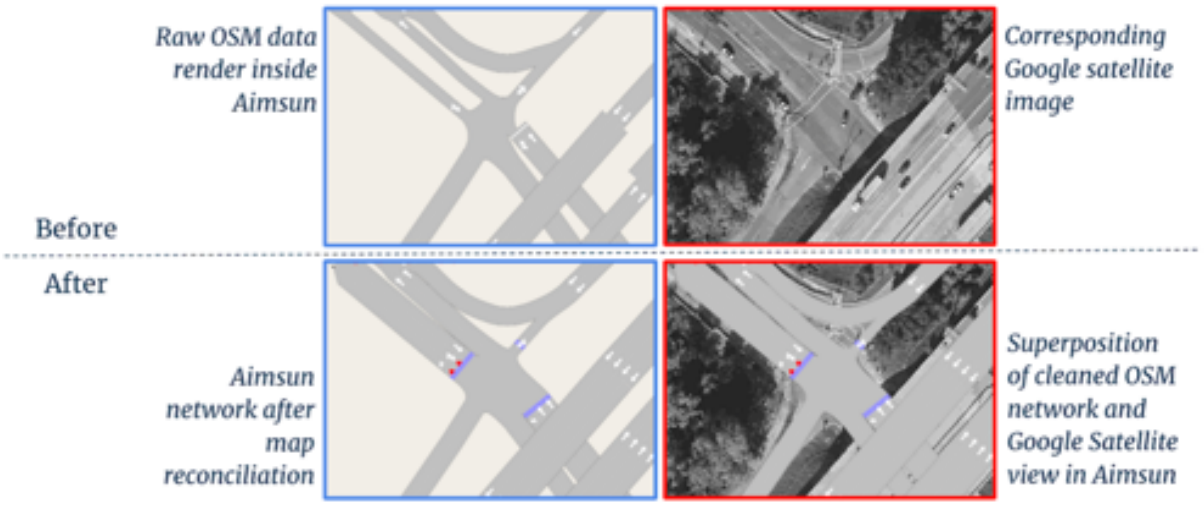


Figure 5.2: An intersection of the Aimsun network before and after manually editing the OSM network using Google Satellite images.

data is generated from 56 city flow detectors and 27 CalTrans Performance Measurement Systems (PeMS) detectors [243]. Speed and travel time data can be acquired using the Google Maps API [102]. In this study, travel time data was gathered from driving in the area. The overall process to create traffic data took our team 400 person-hours to complete.

5.4 Simulation

Once the input data is imported into the Aimsun simulator, simulations can be run, generating simulated traffic data as output.

Running a simulation

To run a traffic simulation in Aimsun, one first needs to create/import the network. Second, the OD demand data should be imported, and a traffic demand (a list of timed OD demand matrices with scaling factors) should be generated. Optionally, traffic data can also be imported. Once all imports are complete, a traffic simulation can be created, with many mutable parameters (see full list in section 5.5). From here, the simulation can be run.

Creating and running a simulation can be done using the Aimsun's GUI, but, for the sake of reproducibility, the team opted to write Python scripts for each step of the process. The open-source repository is self-contained, and any readers with an Aimsun 22.0.1 license should be able to reproduce all the steps explained below and run the same simulation performed by the team [135].

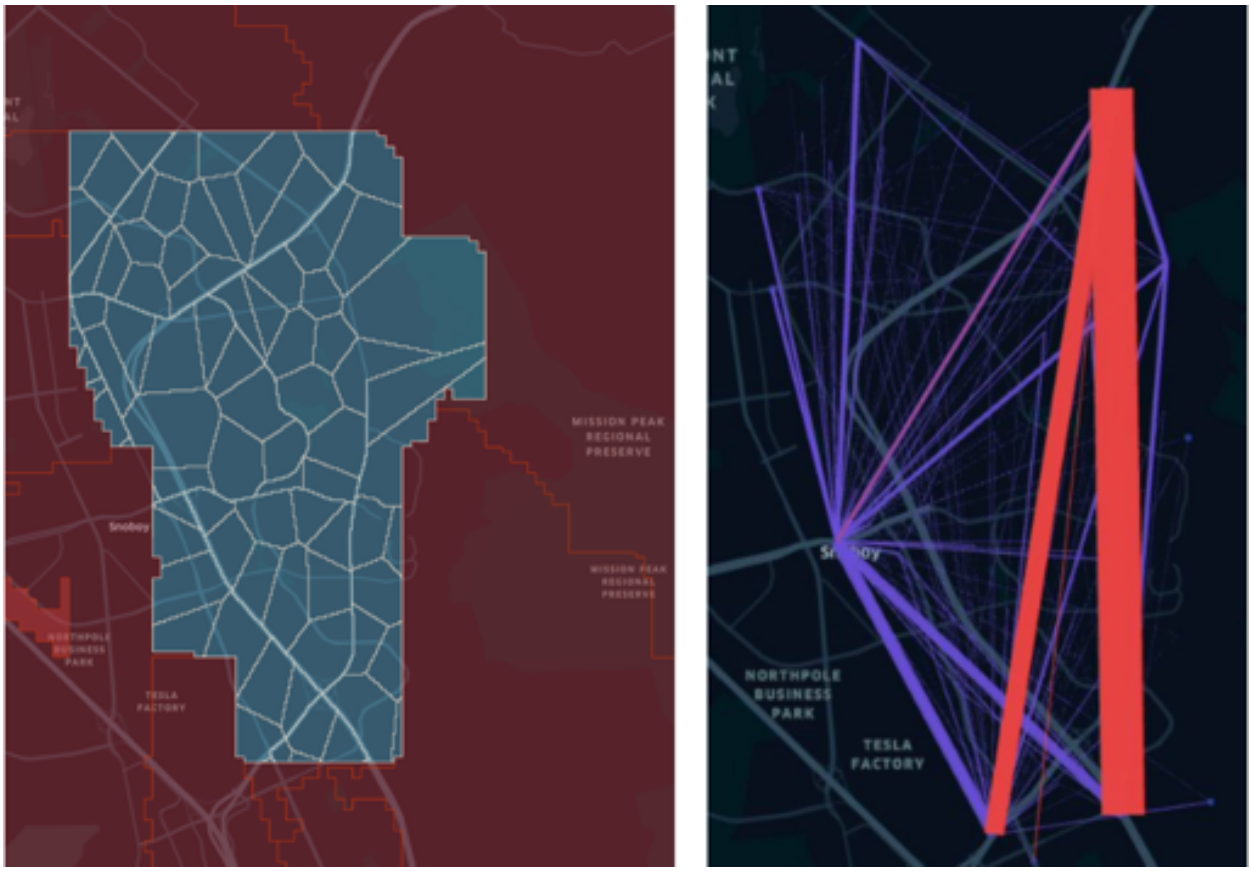


Figure 5.3: On the left: Transportation analysis zones (TAZ). On the right: The OD demand plotted with desire lines. A commuter (aggregated into red lines on the right plot) is a vehicle with an origin and a destination, which are both external centroids (red TAZ on the left plot). A resident (aggregated into blue lines on the right plot) is a vehicle departing or/and arriving from or/and to an internal centroid (blue TAZ on the left plot).

First, the network can be imported directly from OpenStreetMap (OSM) in Aimsun. For this work, the team did some processing of the OSM network data in ArcGIS before importing the OSM network from an external file in Aimsun. Speed limits and vehicle capacities for each of the road sections can then be updated. Then, the ramp meters, the traffic control plans, the master control plan, and the traffic management strategies are imported. Among these, correctly configuring the traffic control plans took most of the time because of the many parameters that needed to be changed for each traffic light to consider various settings and actuation. The demand data is also imported into Aimsun, which is created by importing the centroid data with the centroid connectors. The centroid data contains the OD demand data, which is converted to traffic demand. The ground flow data, the final bit of input, is then imported as a real data set inside Aimsun.

A Jupyter Notebook was designed to define any configurable model characteristics outside Aimsun (including time step size, routing model, driving behavior, or output database location) needed to generate the simulation. Finally, the simulation can be run. Running the 6-hour-long simulation takes between 30 minutes and 4 hours, depending on the data to output and the input data size.

Simulation Output Data Description

Microsimulation models can generate detailed data for every vehicle corresponding to its car following, lane-changing, and gap acceptance behavior. These characteristics can be observed through simulation playback, using Aimsun's GUI to visualize each vehicle's motion through the network. These results can also be aggregated to compare the macroscopic simulation results vs. real data sets with respect to factors such as flow, speed, and travel time.

In Aimsun, while the outputs can be accessed using the GUI, it is also possible to save them as SQLite tables. The output tables [228] contained in the output database are defined in the microsimulation configuration. Each table contains different types of statistics, and it is important to identify which tables are necessary to generate before running the simulation to prevent data cluttering. For example, the MISYS (microsimulation system) table contains system-level statistics about the entire network, such as VMT, VHT, total gas emissions, or average delay across all road sections. The SQLite output databases are used as primary data sources for simulation outputs throughout calibration (section 5.5), validation, and analysis (section 5.6).

5.5 Calibration

The challenges of creating a realistic microsimulation of city traffic lie in its calibration. To calibrate a microsimulation, one needs first to fix all the network issues. Then, OD demand can be calibrated without simulation by matching the total counts of vehicles entering or exiting the network. Macrosimulation can then be used to conclude the demand calibration, by matching all detector counts in the network. Once the OD demand is calibrated, and

after having chosen a route choice model, the driving behaviors are calibrated using a genetic algorithm. The OD demand calibration and driving behavior calibration procedures can be reproduced from the provided open-source code [135].

Network and demand apparent issue fixes

A first test to ensure that the input network and demand data are not flawed can be done by running a simulation with 50% of the demand and checking that no congestion occurs in the network. Congestion can be detected with the Aimsun GUI by playing the simulation. It can also be detected with total delay in the network across time, or with the mean travel time per vehicle-mile. A second check – with a full-demand simulation – can be done by looking at gaps between the ground flow and the simulated flow that are below 50% or above 200%.

The team used GitHub [98] to report, follow and solve any apparent network or demand issues. On average, simulation tests found 4 to 5 issues. Each issue was assigned to a team member, and around 2 issues were solved per team member per week. 43 issues of this variety were reported in total. The overall apparent issues fixes process took around 500 person-hours. The issues included:

- Updating incorrect lane connections at intersections.
- Changing erroneous road section geometries.
- Changing wrong numbers of lanes on road sections.
- Removing parking lots, where cut-throughs were performed in the simulation to avoid congestion at traffic lights.
- Updating improperly imported traffic signal plans.
- Updating master traffic control plans to solve missing synchronization between traffic signal plans. The team found one ground truth congestion issue that could be solved by synchronizing traffic lights operated by the state with the traffic lights operated by the city on Auto-Mall Pkwy and the I-680.

The team realized during this process that the input demand data used was biased towards the northwest, which is likely a byproduct of the demand being sourced from the SFCTA CHAMP demand model, which aims to replicate traffic in San Francisco (located to the northwest of Fremont). An accurate initial demand data is key for a realistic simulation.

OD demand calibration without simulation

A first calibration of the OD demand described in (section 5.3) can be done without simulation. To do so, the total demand is scaled up or down such that the ground flow data at

every external entry or exit point matches the demand data that enter or exit the network at said point (each of which is represented by an external centroid derived from external TAZs). This approach is very similar to the one used in [105]. The objective function minimizes is shown in eq. (5.1):

$$\min_{\alpha\in\mathbb{R}} \sum_{\theta\in\Theta}\sum_{d\in\mathcal{C}_{ext}} \left(f_{mapd(d),\theta} - \alpha \sum_{o\in\mathcal{C}} \mathrm{d}(o,d,\theta) \right)^2 + \sum_{\theta\in\Theta}\sum_{o\in\mathcal{C}_{ext}} \left(f_{mapo(o),\theta} - \alpha \sum_{d\in\mathcal{C}} \mathrm{d}(o,d,\theta) \right)^2 \tag{5.1}$$

Where, using notations and definitions from section 1.2; $\mathcal{C}$ is the set of centroids, o and d are origin and destination centroids, Θ is the set of time bucket, and θ is one time bucket, $f_{l,\theta}$ is the ground flow data for the time bucket θ on the detector l, and $\mathrm{d}(o,d,\theta)$ is the number of vehicle that exit the origin o to reach the destination d during the time bucket θ.

And, where; $\mathcal{C}_{ext}$ is the subset of external centroids, $mapd(d)$ is the detector associated with the destination centroid d, and $mapo(o)$ is the detector associated with the origin centroid o.

This approach can be used to derive how the demand changes over time in cases when the current OD demand matrices and the flow data over the years are available.

OD demand calibration with macrosimulation

Once the OD demand is calibrated against entry or exit flows, it can be calibrated against all detector flows in the network by assigning the OD demand to routes and counting the number of vehicles going over each detector. To assign the OD demand to routes without simulating each individual dynamics, the static traffic assignment can be used [181]. OD-demand calibration aims to better align the simulated and ground detector flows. This OD adjustment is done by solving the constrained generalized least-squares described in eq. (5.2) as adopted from [29]:

$$\min_{\hat{d}\in\mathbb{R}_+^{\mathcal{C}\times\mathcal{C}\times\Theta}} \sum_{\theta\in\Theta}\sum_{l\in\mathcal{D}} (f_{l,\theta} - \hat{f}_{l,\theta}(\hat{d}))^2 + \gamma \sum_{\theta\in\Theta}\sum_{o,d\in\mathcal{C}} (\mathrm{d}(o,d,\theta) - \hat{d}(o,d,\theta))^2 \tag{5.2}$$

Where, using notations and definitions from section 1.2; $\hat{d}$ is the calibrated timed OD demand, $\mathcal{D} \subset \mathcal{L}$ is the set of detectors, $\hat{f}_{l,\theta}(\hat{d})$ is the simulated flow on detector l during the time bucket θ given the demand $\hat{d}$, $\boldsymbol{d}$ is the prior demand before the calibration with macrosimulation and after the calibration without simulation, and γ is a scaling factor to avoid overfitting the flow data. In this subsection, the simulation to create $\hat{f}_{l,\theta}(\hat{d})$ is the static traffic assignment (definition 3.1.7).

To reduce the risk of overfitting, a regularization term that penalizes large modifications of the prior demand can be added to provide a balancing effect [29], formulated as the Frobenius norm [28] of the difference between the calibrated and the original OD demand matrices. This approach is not exclusive to the Frobenius norm – other norms such as the

nuclear norm [28] could be considered for regularization. In [162], the l_1 norm is used to compare the OD matrix elements. The comparison of results between the macrosimulation results after OD adjustment using the training and testing sets with a small regularization term is shown in fig. 5.4, where overfitting can be observed. In this work, to avoid overfitting, the regularization term was scaled by a large factor $\gamma = 10$ such that the OD demand matrix after the macrosimulation calibration was very similar to the OD demand matrix after the calibration without simulation.

Finally, the validation of the calibrated matrix was done with flow regression plots. Flow regression plots compare simulated values with real-world values by scatter-plotting them as y and x-axes, respectively. A linear regression is then fitted onto the data points to draw the line of best fit [83]. The slope and intercept of the regression can then be compared to the ideal $y = x$ line to determine whether the simulation model tends to over/underestimate the plotted metrics and whether there is bias in the model. Performance metrics of the linear regression [83], such as the coefficient of determination (R^2), root-mean-square error (RMSE, or nRMSE when normalized), and the mean absolute percentage error (MAPE) can be computed to determine the accuracy of the simulated flow.

Choice of the routing model

Once the network is bug-free and the OD demand is calibrated, the microsimulation can be run. By design, the microsimulation has many modifiable parameters to calibrate individual driver behaviors. Some of the most important microsimulation parameters are about routing behaviors.

A routing model assigns travelers to a series of links to get from one centroid (origin) to another (destination). There exist two types of routing models [59]:

1. The **one-shot assignment model** assigns routes and runs the simulation once. When assigning vehicles to route, only past and current information are used, and no assumptions are made about the future. The route is given following a stochastic route choice (SRC) model [59].

2. The **iterative assignment model** assigns routes and runs the simulation iteratively until the travel cost experienced by each vehicle at the end of their trip cannot be minimized by unilaterally changing the route of the vehicle. This equilibrium state is referred to as the dynamic user equilibrium [59] (sometimes called Wardrop equilibrium or Nash equilibrium, see definition 3.1.7).

Because running many simulations iteratively takes a lot of time, the team opted for a stochastic route choice (SRC) model. Several SRC models are available in Aimsun (fixed-route under free-flow conditions, fixed-route under warm-up period traffic conditions, binomial model, proportional model, logit model, and C-logit model). Considering the tradeoff between accuracy and simulation run time, the team chose to use the C-logit model [53] after experimenting with the different models.

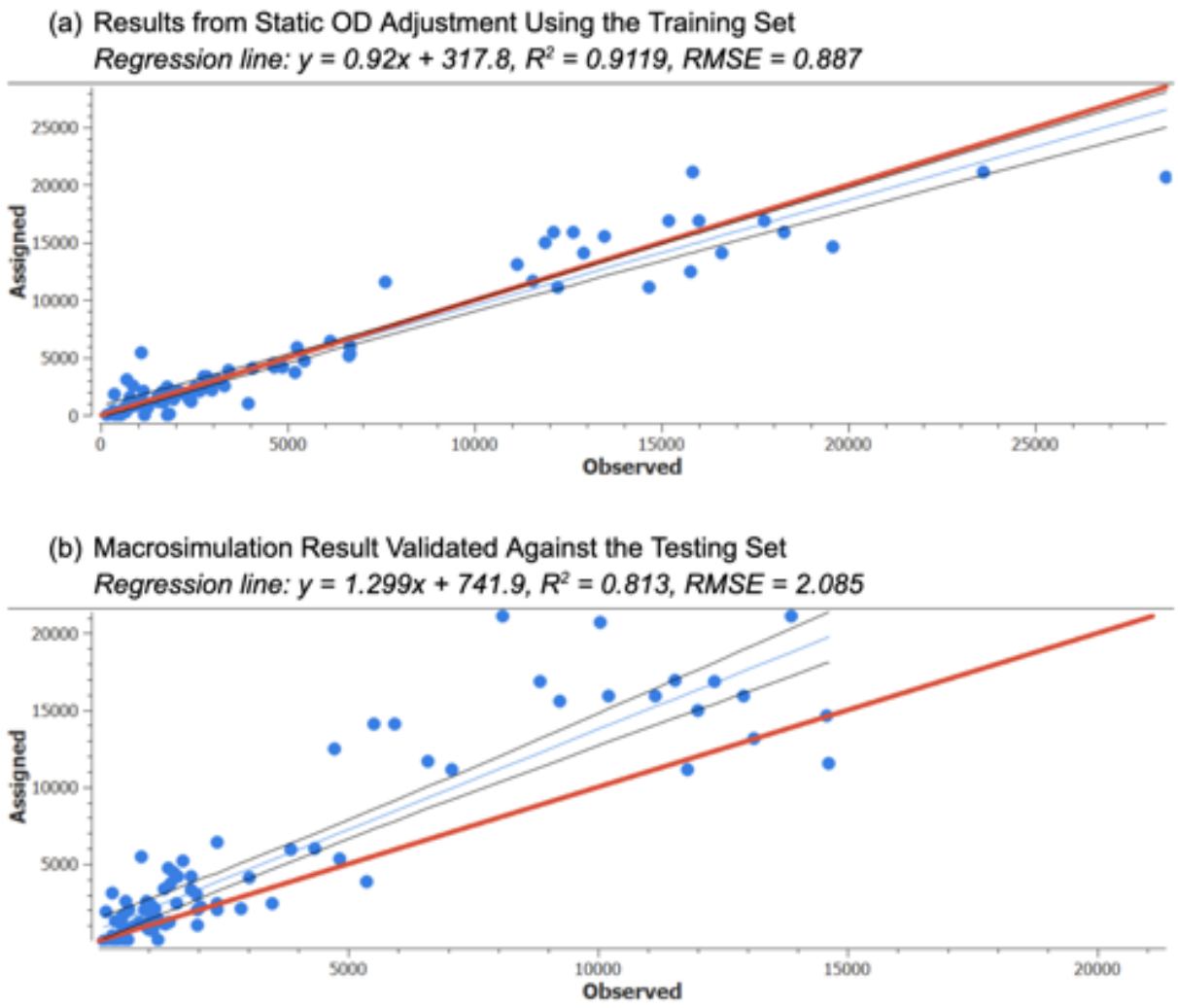


Figure 5.4: Assigned/simulated traffic vs. observed/actual traffic flow regression plots after Aimsun's default OD demand calibration with macrosimulation. On top: Training results. At the bottom: Testing results. Very good training results (slope of 0.96 and R^2 of 0.9, both close to 1) accompanying poor-quality testing results (slope of 1.11 and R^2 of 0.81, further away to 1) show that the calibration has over-fitted the training demand data.

The C-logit route choice parameters include [53]:

- The number of alternate routes considered for each vehicle route choice.
- The cost update frequency (definition 3.2.1)
- The percentage of vehicles allowed to reroute en route.
- The averaging parameters for past and current (instantaneous) travel cost parameters.
- The route cost function parameters (utility, scaling cross-factor, overlap parameter).

The calibration of the C-logit route choice parameters was done as part of the microsimulation calibration.

Driving behavior calibration

Once the routing choice model is chosen, the microsimulation-specific parameters (like simulation time step length) and the driving behavior can be calibrated. The driving behavior parameters include the routing behavior parameters (like cost update frequency), the car following parameters (like reaction time), and the lane changing parameters (like aggressiveness). The full list of parameters that can be calibrated can be found in [135].

The routing calibration aims to find the configuration of parameters that minimizes the difference between simulated data and ground truth data, without overfitting. A first manual calibration can be done based on intuition with simulation recording (for example, reaction time can be adjusted if the output to input flow ratio at some intersection seems low). Then, bounds can be set for each parameter based on physical intuitions (reaction time is between 0.2 to 3 seconds) and a systematic calibration can be performed.

In the systematic calibration, an objective function, described in eq. (5.3), is set to be minimized (similarly to OD demand calibration in eq. (5.2)):

$$\min_{\phi} \sum_{\theta \in \Theta} \sum_{l \in \mathcal{D}} (f_{l,\theta} - \tilde{f}_{l,\theta}(\phi))^2 \tag{5.3}$$

Where ϕ is a set of parameters to calibrate, $\tilde{f}_{l,\theta}(\phi)$ is the simulated flow on detector l for the time bucket θ given the calibrated demand, and the set of parameters ϕ. If we have any prior on the set of parameters ϕ, a regularization term can be added to avoid overfitting.

An optimization algorithm can be run to find the optimal microsimulation-specific parameters and driving behavior parameters. Optimization algorithms include brute force algorithms like grid search, random search [190], classical optimization algorithms that can be found in the SciPy.Optimize toolbox [244], neural-network [140], or genetic algorithm [163]. Because evaluating the objective function given the input parameters is costly, since it requires running a microsimulation, the team decided to use a genetic algorithm. Genetic algorithms are particularly efficient for opaque box functions with a high stochastic effect. In addition, they are easily parallelizable and can handle multi-criteria optimization [163].

Genetic algorithm for microsimulation calibration

The team followed the approach presented by [154] to fine-tune the driving and microsimulation parameters.

Because of the significant number of model parameters (35 for micro-simulation using the C-logit stochastic route choice model) and the runtime for each simulation on the team's computers (30-40 minutes), the search space (i.e., number of model parameters to calibrate) has been decreased through a sensitivity analysis. Only the 10 parameters that have the most impact on the measures of effectiveness were selected by comparing the relative metric evolution over small increments of each parameter using the Latin hypercube sampling (LHS) algorithm [119]. The calibration of the 10 selected parameters was done using the Non-dominated Sorting Genetic Algorithm-III [75] from the Distributed Evolutionary Algorithms in Python library [89] using the objective function described in eq. (5.3). The initial value of the parameters was set to the Aimsun default ones if not updated based on intuition after few initial simulations.

5.6 Post-Processing Analysis

Once the simulator is calibrated, one needs to validate the accuracy of the simulation to ensure the credibility of its results. Once the accuracy of the simulation is satisfactory, traffic analysis can be conducted to observe how certain metrics change in different scenarios.

Validation

The first step in validating the simulation is to use the Measures of Effectiveness (MOE) [36, 143] (see section 2.2), which serves as an indicator for general correctness of system-wide results. Some examples of metrics that can be used for MOE are average delay time, total distance traveled by vehicles in the entire network, and average number of vehicles in the virtual queue at each time step.

After determining that the MOE are accurate, the next step is to validate system-wide and location-specific metrics for the simulation. System-wide metrics denote data encapsulating property or properties across the entire simulation network, such as flow at all detectors. Location-specific metrics, on the other hand, deal with data specific to a subset of the network, such as a corridor.

Validation of system-wide metrics

For validation of system-wide metrics, regression plots comparing simulated versus real-world data are commonly used across previous literature [206, 12, 56]. In this work, detector flow, OD travel times, OD route distances, and system-wide metrics were used to validate the simulation results (fig. 5.5).

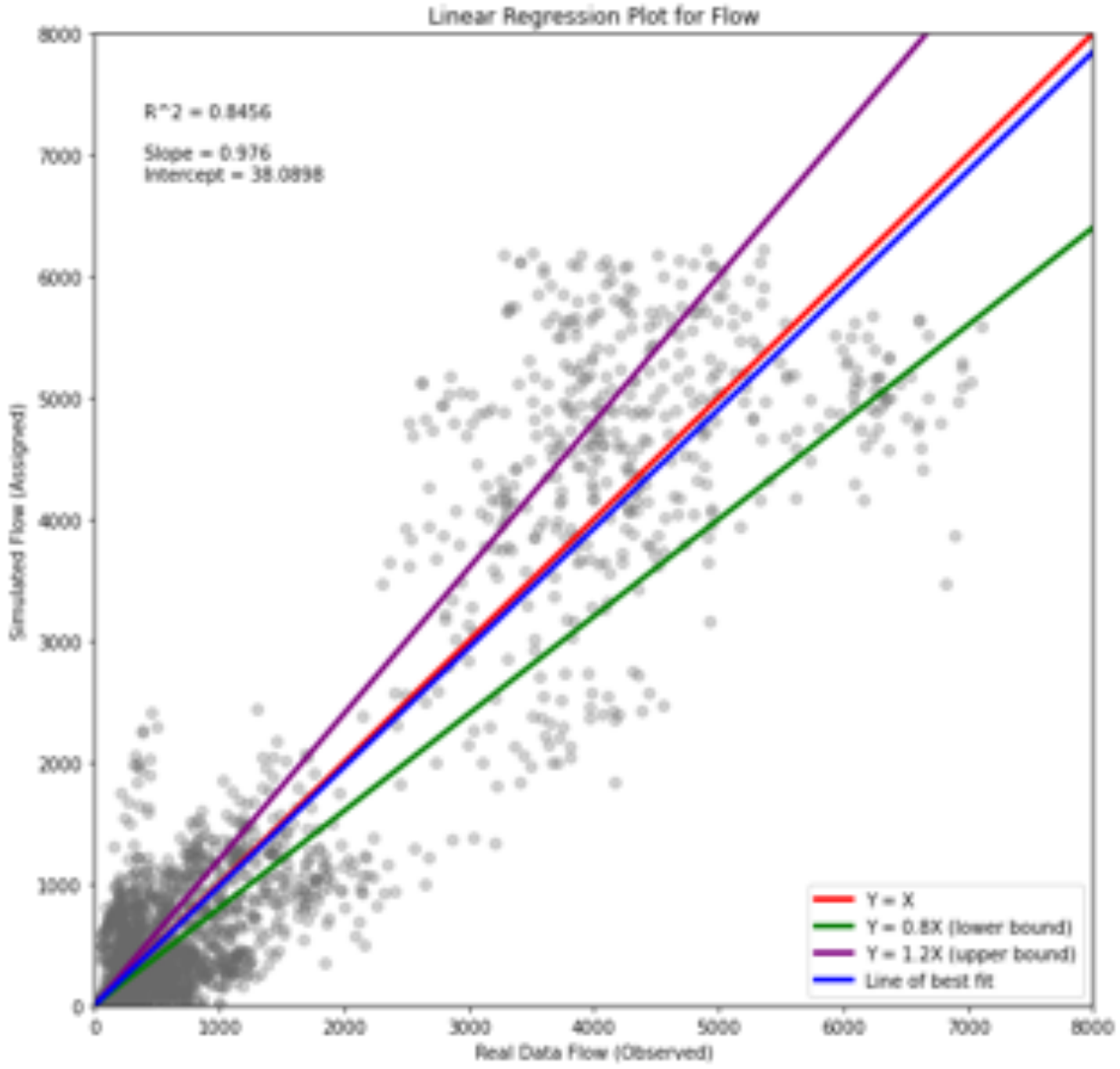


Figure 5.5: Linear regression plot of simulated vs. ground vehicle counts during 15-minutes time buckets across 83 detectors in the networ.

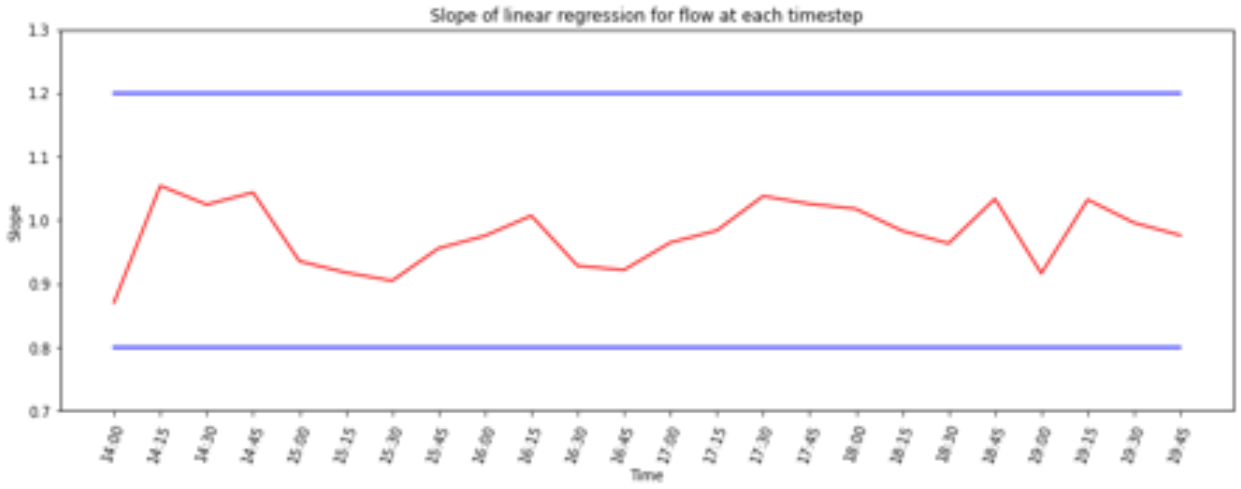


Figure 5.6: Slope of the simulated vs. ground 15-minutes vehicle count linear regression for as a function of the time bucket (red). Setting arbitrary lower and upper bounds of 0.8 and 1.2 (blue), respectively, shows that the simulation does not over- or under-estimate the real flow.

To conduct a more detailed analysis of regression plots, a time series of regression statistics for each regression plot can be plotted, which gives insight into specific points in time when the simulation needs more calibration (fig. 5.6).

In addition, it might be useful to identify the times for each detector in which the simulated flow was over 1.2 or below 0.8 times the ground flow to know which detectors need more calibration.

Distributions of actual and simulated values can also be used for eyeball validation of the simulation (fig. 5.7). Though distributions show that the general trend that the simulated data is correctly approached, they provide less insight into the individuality of each data point than regression plots.

Validation of specific location metrics

After validation of system-wide metrics, one can validate location-specific metrics in areas of high importance to ensure the accuracy of the simulation. To do so, we need to identify the scope of the location critical to our study. In this example, the team set the I-680 corridor as the scope of specific location metric validation. Metrics one can validate include flow (veh/h), speed (km/h), delay time (h or h/veh), and density of vehicles (veh/km) observed at each flow detector within the corridor (fig. 5.8).

The team divided up validation of location specific metrics into two processes. The first process is to visualize a time-series of each metric observed in each detector. The trend and

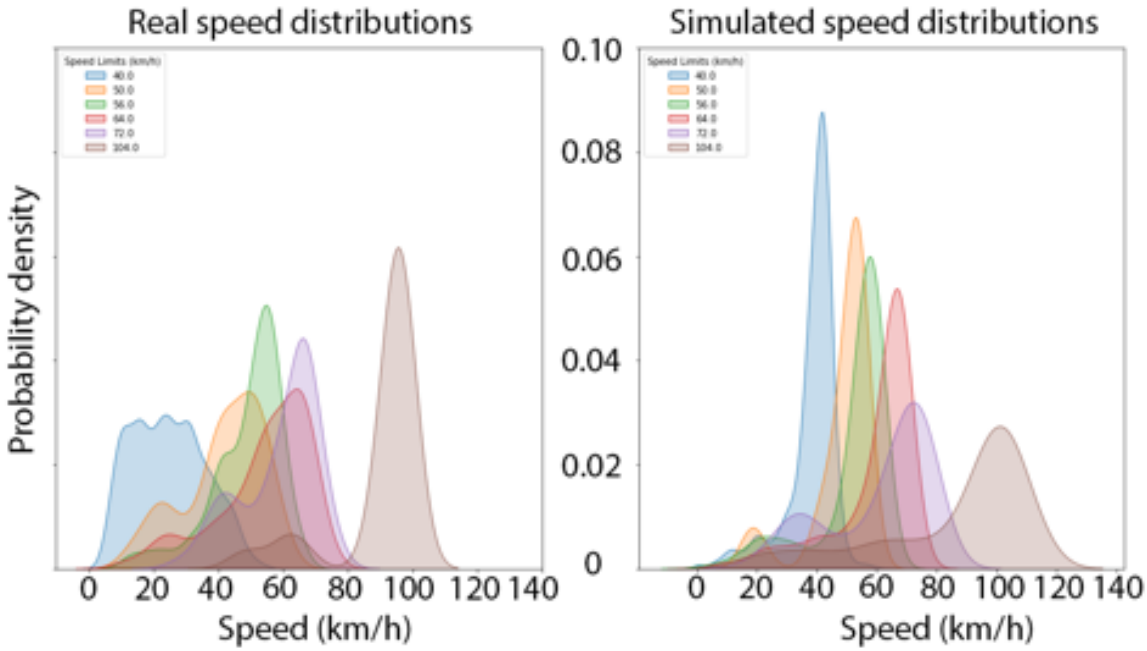


Figure 5.7: On the left: Kernel density estimation (KDE) plot of real speeds at each road section. On the right: Corresponding simulated speeds. Each distribution is grouped by the speed limit on the road where the speed was observed. The simulation is able to capture a majority of trends seen in real average speed distributions. The simulation better performs on highway data than local road data.

values observed at each time step help determine the accuracy of each metric at each critical geographic point at the granular level.

Then, a time-space diagram (fig. 5.9) is used to validate the macroscopic spatio-temporal relations across detectors in one corridor. Time-space graphs describe the relationship between the location of vehicles in a traffic stream and the time as the vehicles progress along the highway [96]. Note that on or off-ramp detectors are not included when creating the flow time-space diagram of the highway.

Analysis

After successful validation, the microsimulation can be used for traffic analysis study. The study can help understand the current traffic in a city or predict the impact of a "what-if" scenario. Note that microsimulation can only be used to study effects that do not require changing the input demand data. Changing the demand data will result in comparisons of two different traffic simulation models.

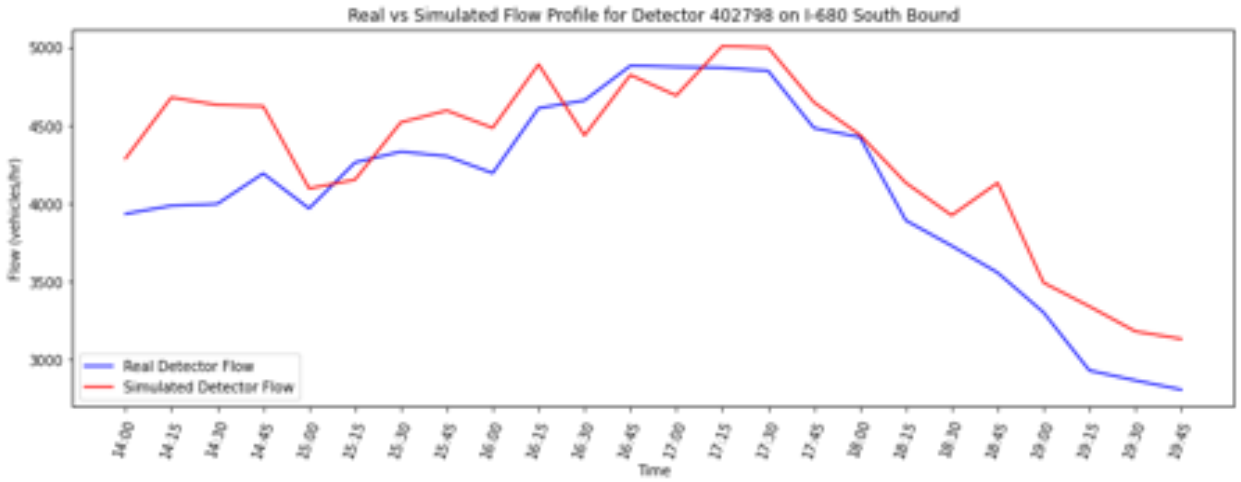


Figure 5.8: Simulated and ground flow profile for a detector on the I-680 South corridor. The trends observed align with each other.

Some examples of scenario changes include implementing new traffic-calming strategies or changes in routing behaviors or driving behaviors. Motivating examples of scenarios that can be analyzed include:

- Changing traffic signal timing plans [225].
- Changing speed limit or adding speed bumps [225].
- Adding turn and/or access restrictions [225].
- Understand the impact of increase in usage of navigational-apps on traffic (see section 3.2).
- Understand the impact of eco-routing adoption [13, 104] by changing the cost function that a portion of the drivers minimizes when choosing their route.
- Changing the type of some vehicles to study the impact of mixed-autonomy in traffic [254].

The all the measures of effectiveness metrics that have been enumerated in section 2.2 can be computed with the microsimulation outputs.

5.7 Conclusion

Through the development of a traffic microsimulation of the San Jose Mission district in Fremont, CA, the team designed and shared a reproducible process to create, calibrate and

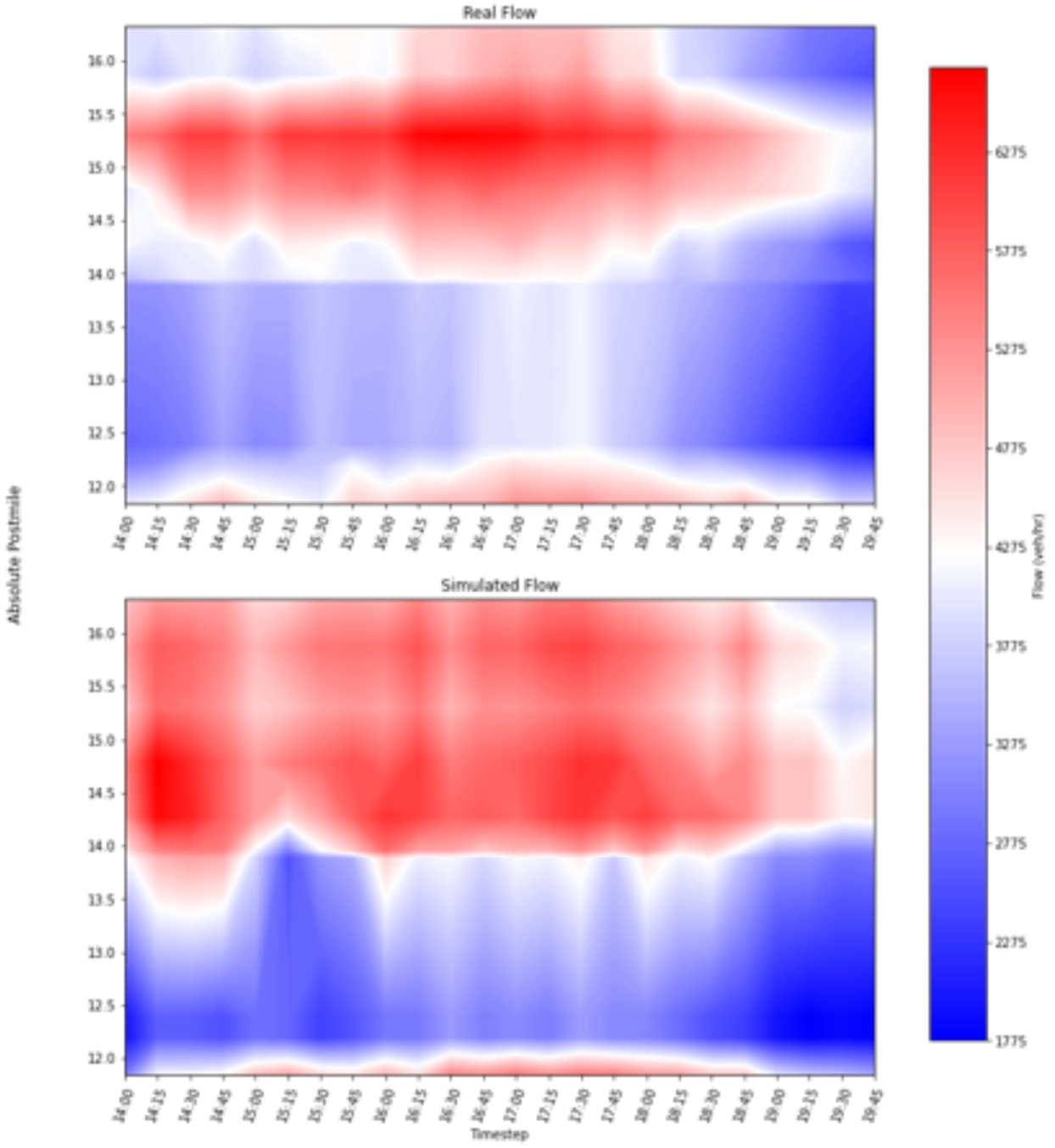


Figure 5.9: Time-space diagram of real versus simulated flow in I-680 South. Patterns of congestion are similar across real and simulated plots.

validate a large-scale microsimulation. The development of a large-scale traffic microsimulation is a tedious process that took the team around 2,500 person-hours, and it is relevant only if case studies or A/B experiments cannot be performed and if enough data is available to accurately reproduce the demand data. A realistic traffic microsimulation can be used to understand current traffic and estimate policies that might impact routing or driving behaviors without changing the traffic demand (number of trips, departure times, and origins and destinations). If the simulation quality is very high compared to existing literature (slope of the flow regression line of 0.976 and corresponding R^2 of 0.8456), it is not good enough yet to be used off-the-shelves by the city of Fremont traffic engineers (flow nRMSE of 47%).

If microsimulations can reproduce individual dynamics in the road network, one of the barriers to develop microsimulations is the simulation running time. To tackle this hindrance, chapter 6 develops a dynamic routing game, where each individual vehicle route-choice and dynamics are modeled. Mean-field game theory is used to efficiently solve the routing game with numerous vehicles.

Chapter 6

Computable dynamic routing game: the mean-field routing game

With the emergence of navigational applications and consequent information-aware routing behaviors, the traffic patterns evolved (chapter 2, chapter 3, chapter 4). Being able to model traffic and especially route choice adequately can enable traffic control to leverage the new routing behaviors in order to improve the network efficiency.

However, solving realistic route choice problems requires the consideration of large scale multi-agent systems, where the number of vehicles making strategic decisions is not tractable within classical algorithms [73]. This chapter focuses on finding a scalable approach to model the routing behaviors of each vehicle in a dynamic road traffic environment as a large multi-agent dynamic system. First, section 6.1 presents a review of dynamic routing models. Second, section 6.2 introduces a novel routing model that can represent dynamically individual behaviors using dynamic game theory: the **N-player dynamic routing game**. The N-player dynamic routing game is a dynamic mesoscopic traffic model with explicit congestion dynamics (i.e., congestion impacts the evolution of the traffic state). In addition, section 6.3 introduces the **mean-field routing game** to efficiently simulate the novel routing model. Finally, section 6.4 proves experimentally that the mean-field routing game provides a scalable approach to model the routing behaviors of numerous agents (14,000) in road networks.

The chapter is derived from [46], and it is still ongoing work.

6.1 Background on dynamic routing models

Dynamic traffic assignment

Dynamic traffic simulations are used to model the evolution of the locations of the vehicles in the road network across time. Within dynamic models, simulations are divided between macro, meso and microsimulations (section 5.1).

Dynamic routing models are divided between one-shot assignments and iterative assignments (see choice of the routing model subsection in section 5.5).

In static routing games (definition 3.1.4), the congestion on each link does not evolve over time. These games cannot replicate dynamic phenomena like departure time choice or congestion spill back. Therefore, dynamic extensions of the static routing game were introduced. Using the potential formulation of the static traffic assignment (definition 3.1.12), a first dynamic traffic assignment model was defined as the solution of a dynamical variational inequality in 1993 [93]. The variational inequality approach was later extended in 2013 to consider both route choices and departure time choices [111]. However, in such dynamical models, the game theoretical aspect is not explicit. In parallel, a first dynamic routing game was defined in 1993 [250] using differential games [92]. The resolution of the dynamic routing game via multi-agent reinforcement learning has been studied in 2020 [207]. Nevertheless, as of 2022, to the knowledge of the author, current multi-agent learning algorithms do not scale in terms of population size [73]. This inherently calls for new methods to study route choice problems with very numerous vehicles. We next explain how a mean-field game perspective can help to make an important step in this direction.

Mean-field game

Mean-field games (MFGs) were introduced in [139, 118] to model differential game dynamics between infinitely many players with symmetric interactions. Similar to mean-field theory from statistical physics[22], the key idea is to use a macroscopic approximation of a large population with anonymous and symmetric players. When the population is infinite, each player has no influence on the population distribution. So, in a MFG, one does not need to study the pairwise interactions between all players, but simply the interaction between a representative infinitesimal player and the distribution of players' states. Because players are assumed to be identical, it suffices to determine the strategy of a representative player in response to the full population behavior. From a mathematical point of view, the solution of a MFG can be characterized by a coupled system of a forward equation for the population evolution and a backward dynamic programming one for the player's value function. This system is easier to solve than the Nash equilibrium of a finite-player dynamic game with numerous players (solved through a large system of coupled Bellman equations). Intuitively, this is true when representing each player is more costly than representing a distribution over all possible states, which is the case for instance when the number of players is larger than the number of possible states.

Applications of MFGs include crowd modeling [136, 17, 2], energy management [199, 142, 5], epidemiology [79, 84, 18] or financial markets [137, 51, 85]. In traffic theory, similar approaches model the evolution of vehicles in the network using traffic flow (remark 1.2.1). Modeling microscopic vehicles on a link as a macroscopic traffic flow has been interpreted as a mean-field game in 2015 [58], developing the application of MFGs to road traffic management. Extending on this interpretation, [116] connected the Lighthill-Whitham-Richards (LWR) model [144] on a single road to a MFG model with myopic players in 2019. In this mean-

field game [116], vehicles are assumed to choose their speed to minimize a cost function that penalizes traveling near other vehicles. In 2021, the authors of [116] extended the vehicles' controls to route choice in [117]. In parallel, mean-field routing games have been developed by [198] and [219] building on the construction of MFGs on graphs done in 2015 by [108].

Related work

Several works study mean-field routing games. First, continuous time models have been studied. The existence and uniqueness of the Nash equilibrium of a MFG with congestion on a graph has been shown in [108], where the state space is the set of nodes. Models in which the state space is given by the edges have been analyzed in [3], which proved existence and uniqueness for a forward-backward system of equations with suitable conditions at the vertices of the network. In [25], the authors analyzed a MFG model for traffic flow on networks by using an extended state space that includes the distribution of players on the network. They also connected the dynamic Wardrop equilibria condition to the Nash equilibria of the MFG. In existing discrete time MFG models for routing, the players move one edge per time step and pay a cost that increases with the proportion of players on the same edge. In [198], the authors analyzed a MFG model and studied the impact of adding or removing edges on the equilibrium traffic flow. Their work provides a discrete time resolution of a mean-field routing game on an 11 link network with 6 time steps. In [219], the authors proposed a MFG model that reduces to a linearly solvable Markov Decision Process and showed connections with Fictitious Play [43] in some cases. In [117], the route choice is coupled with the speed choice to model how autonomous vehicles might choose both their speed and their route to optimize traffic.

The fact that existing models take congestion into account only through the cost functions leads to an ambiguity about the definition of travel time: the graph traversal time and the player cost can differ. Such issues make these models hardly applicable for traffic engineering. Also, in this chapter, the main motivation for using a MFG-based routing methods is to obtain an efficient equilibrium policy in the finite-player routing game, which has not been performed in previous works.

Section 6.2 introduces a N-player dynamic routing game with explicit congestion dynamics. Section 6.3 introduces the corresponding mean-field routing game. Finally, section 6.4 solves N-player dynamic routing games with the corresponding mean-field routing game.

6.2 Dynamic N-player game

This section introduces the dynamic routing game. The dynamic routing game models the evolution of N vehicles on a road network. The vehicles are described by their current link location, the time they will spend on the link before exiting it (i.e., their waiting time), and their destination. The action of a vehicle is the successor link they want to reach when exiting a given link. When arriving at a link, the waiting time of the player is assigned based

on the number of players on the link at this time. As time goes by, the waiting time of a vehicle decreases until it becomes negative, then the vehicle moves to a successor link and the waiting time gets reassigned. The total cost for the vehicle is its travel time. In the corresponding mean-field game (MFG), the vehicles of the N-player game are replaced by a representative vehicle and the probability distribution of the vehicles states.

Traffic flow environment

In this chapter, we reuse the notations introduced in section 1.2. For the sake of convenience, we assume that the time horizon T (definition 1.2.1) is large enough so that any vehicle will have time to travel through the network.

The number of players in the game $N \in \mathbb{N}$ (notation 1.2.3) is not necessarily the total number of vehicles N_0 in the real-life scenario (definition 1.2.4). Each player of the model corresponds to a proportion of the real number of vehicles, which allows defining a player as an infinitesimal portion of flow that does not impact network travel time in the corresponding mean-field game (MFG) (see section 6.3). In the case where $N = N_0$, a player is a vehicle. A player $i \in \mathcal{N}$ starts at an origin link $L_0^i \in \mathcal{L}$ with a departure time $W_0^i \in \mathcal{T}$, and has a destination link $D_0^i \in \mathcal{L}$. This is the initial state of the player: the player wants to depart at time W_0^i from L_0^i and tries to reach D_0^i. We assume that the players' initial state $(L_0^i, W_0^i, D_0^i) \in (\mathcal{L} \times \mathcal{T} \times \mathcal{L})^{\mathcal{N}}$ are distributed according to a finite-support distribution $m_0 \in \mathcal{M}(\mathcal{L} \times \mathcal{T} \times \mathcal{L})$. Both the origin and the destination are modeled as links, so that the location of the vehicle is always described as a link. In experimental setups, an origin link is added before each origin node and a destination link is added after each destination node. Being on the origin link means having not departed yet, and being on the destination link means having finished the trip.

At any time t, the state of a player i is not only the link $L_t^i \in \mathcal{L}$ where they stand, but also their waiting time $W_t^i \in \mathcal{T}$ before exiting this link together with their destination $D_t^i \in \mathcal{L}$. L_t^i and W_t^i are random variables due to the randomness in the action choices. Even though the destination is constant through time ($D_t^i = D_0^i$ for all t), including this information in the state allows keeping track of the objective in the player's policy. So the state space for each vehicle is $\mathcal{X} = \mathcal{L} \times \mathcal{T} \times \mathcal{L}$, where the first component is for the current location and the last one is for the destination (recall that the destination is represented by a link in our model). Then, the space of vehicle trajectories is $\boldsymbol{\mathcal{X}} = \mathcal{X}^{\mathcal{T}}$. The state trajectories are in the space of triples (location, waiting time, destination), which provide more information than the physical trajectories just in terms of locations. At the population level, the states of all the agents is a vector $\underline{X} = (X^i)_{i \in \mathcal{N}}$. The state space for the whole population is $\underline{\mathcal{X}} = \mathcal{X}^{\mathcal{N}}$, and the corresponding space of trajectories is $\underline{\boldsymbol{\mathcal{X}}} = \mathcal{X}^{\mathcal{T} \times \mathcal{N}}$. We respectively call game state and game trajectory, the state and trajectory of the population.

Routing policy

When at a link $l \in \mathcal{L}$, a player can try to move to another link among the successors of l, and the transition is realized provided the waiting time is non-positive. The players are allowed to randomize their actions. We thus call strategy function and denote by $\boldsymbol{\pi}$ a function from $\mathcal{X} \times \mathcal{T}$ to $\mathcal{M}(\mathcal{L})$ such that for any $x = (l, w, d)$, $\boldsymbol{\pi}_t(x)$ has support in the successors of l. Each player has the same set of strategies, therefore we denote by $\boldsymbol{\Pi}$ the set of strategy functions (notation 1.2.4). Therefore, the set of strategy profile is $\underline{\boldsymbol{\Pi}} = \boldsymbol{\Pi}^{\mathcal{N}}$ A strategy might also be called a (feedback or closed-loop) policy. The notation $\boldsymbol{\pi}_t(\tilde{l}|l, w, d)$ represents the probability at time t with which the agent would like to go from l to $\tilde{l}$ given the fact that their waiting time is w and their destination is d.

State dynamics

Since the players' initial states and actions are randomized, their trajectories are stochastic. Given a policy profile $\underline{\boldsymbol{\pi}} \in \underline{\boldsymbol{\Pi}}$, $\underline{X}_t = (\underline{L}_t, \underline{W}_t, \underline{D}_t) \in \underline{\mathcal{X}}$ denotes the random variable corresponding to the links, waiting times and destinations for all the players at the time $t \in \mathcal{T}$. The stochastic process of the population state is denoted $\underline{\boldsymbol{X}} = (\underline{X}_t)_{t\in\mathcal{T}} \in \underline{\boldsymbol{\mathcal{X}}}$. The players' interactions are through the travel time functions $(t_l)_{l\in\mathcal{L}}$ (definition 1.2.7). When arriving at a link, a player will get assigned to a travel time that is given by t_l as a function of the number of vehicles on the link. So, the interaction between a vehicle and the rest of the vehicles is only modeled through the number of vehicles on the same link. It is thus convenient to introduce the empirical vehicles' locations distribution $\nu_{\underline{l}} \in \mathcal{M}(\mathcal{L})$ corresponding to a location profile $\underline{l} = (l^i)_{i\in\mathcal{N}} \in \mathcal{L}$. For every $\tilde{l} \in \mathcal{L}$, $\nu_{\underline{l}}(\tilde{l}) = \frac{1}{N}\#\{i \,|\, l^i = \tilde{l}\} \in [0, 1]$ is the proportion of players on the link $\tilde{l} \in \mathcal{L}$, given $\underline{l} \in \mathcal{L}^{\mathcal{N}}$. This is all the information one needs from the game state to compute the interactions between players at link $\tilde{l}$. Note that $\nu_{\underline{l}}(\tilde{l})$ is invariant by permutation of the components of the vector $\underline{l}$.

Let us fix a policy profile $\underline{\boldsymbol{\pi}} \in \underline{\boldsymbol{\Pi}}$. We denote by $\underline{\boldsymbol{U}}$ the $\mathcal{L}^{\mathcal{T}\times\mathcal{X}\times\mathcal{N}}$-valued random variable assigned to the probability distribution given by the policy profile: for each $(t, x, i) \in \mathcal{T} \times \mathcal{X} \times \mathcal{N}$, $U^i_t(x)$ is an $\mathcal{L}$-valued random variable with probability distribution $\boldsymbol{\pi}^i_t(x)$.

The evolution of the state of the game $\underline{X}_t = (\underline{L}_t, \underline{W}_t, \underline{D}_t)$ is given by the following

dynamics. At initial time, (L_0^i, W_0^i, D_0^i), $i \in \mathcal{N}$ are given, and then the dynamics is:

$$\begin{aligned}
t_{k+1} &= t_k + \min\{W_{t_k}^i,\ i \in \mathcal{N}\} \\
L_{t_{k+1}}^i &= \begin{cases} U_{t_{k+1}}^i(X_{t_k}^i) & \text{if } i \in I_{t_{k+1}} \\ L_{t_k}^i & \text{otherwise;} \end{cases} \\
W_{t_{k+1}}^i &= \begin{cases} t_{L_{t_{k+1}}^i}\left(\nu_{\underline{L}_{t_{k+1}}}(L_{t_{k+1}}^i)\right) & \text{if } i \in I_{t_{k+1}} \\ W_{t_k}^i - (t_{k+1} - t_k) & \text{otherwise;} \end{cases} \\
L_t^i &= L_{t_k}^i \qquad \forall k, \forall t \in [t_k, t_{k+1}[, \forall i \in \mathcal{N} \\
W_t^i &= W_{t_k}^i - (t - t_k) \qquad \forall k, \forall t \in [t_k, t_{k+1}[, \forall i \in \mathcal{N} \\
D_t^i &= D_0^i, \qquad t \in \mathcal{T},
\end{aligned}$$

where $I_{t_{k+1}} := \{i \in \mathcal{N},\ W_{t_k}^i + t_k - t_{k+1} = 0\}$ and using $(t_k)_{k\in\mathbb{N}} \in \mathcal{T}^{\mathbb{N}}$ the sequence of times when one of the vehicles changes link with $t_0 = 0$ and, for any $k \in \mathbb{N}$, $t_k = T$ if all the players have arrived at their destination. The destination is constant through time and is not affected by the policy's randomness. Note that $\boldsymbol{U}^i = (U_t^i)_{t\in\mathcal{T}}$ is defined for all t but used only when the player moves from one link to the next one, i.e., when the waiting time has vanished. This enables reducing any pure (i.e., deterministic) policy as a path choice.

Cost function

Given a policy profile $\underline{\boldsymbol{\pi}} \in \underline{\boldsymbol{\Pi}}$, the cost for the player $i \in \mathcal{N}$ (notation 1.2.5) is its average arrival time, which can be defined as:

$$J^i(\underline{\boldsymbol{\pi}}) = \mathbb{E}_{\underline{\boldsymbol{\pi}}}\left[\min\{t \in \mathcal{T},\ L_t^i = D^i\}\right] = \mathbb{E}_{\underline{\boldsymbol{\pi}}}\left[\int_{t\in\mathcal{T}} r(X_t^i)dt\right]$$

where the instantaneous cost is defined as: for every $x = (l, w, d)$, $r(x) = \delta_{l\neq d}$ (notation 1.2.1). Note that the running cost is independent of the rest of the population state, contrary to other models for routing or crowd motion in which the interactions are not in the dynamics but in the cost function, like [108, 25, 198, 219, 117].

Furthermore, the population is homogeneous (all players have the same dynamics evolution and same running cost). The player $i \in \mathcal{N}$ interacts with the other players only through $\boldsymbol{\nu}$ and for this reason, the cost function J^i does not depend directly on the index i but only on $\boldsymbol{\pi}^i$: as a function, $J^i = J^j$ for all $j \in \mathcal{N}$. The policy profile $\underline{\boldsymbol{\pi}}^{-i}$ for the rest of the population is used only to compute $\boldsymbol{\nu} = (\nu_t)_{t\in\mathcal{T}}$. Although $\underline{\boldsymbol{\pi}}^{-i}$ is necessary to compute $\boldsymbol{\nu}$, it is not sufficient because $\boldsymbol{\nu}$ is also influenced by the policy $\boldsymbol{\pi}^i$ chosen by the player under consideration. However, the influence of each player decays as N increases. This will be the basis for the mean-field approach presented in section 6.3, where $\boldsymbol{\nu}$ can be computed with $\underline{\boldsymbol{\pi}}^{-i}$ only.

Nash equilibrium

In this chapter, we are interested in computed a Nash equilibrium of the game (definition 1.2.12), assuming that all the players are individually optimizing their own. Therefore, we call a Nash equilibrium a solution of the game.

Theorem 6.2.1 (Existence of N-player Nash equilibrium, Kakutani-Fan-Glisckberg theorem [99])**.** *Assuming the continuity of the cost function with respect to the policy profiles, there exists a Nash equilibrium in the N-player routing game.*

Proof of Theorem 6.2.1. The proof of the existence of a Nash equilibrium relies on the Kakutani-Glicksberg-Fan theorem [99] which needs the continuity of the cost function with respect to the policy profile.

Because $\mathcal{N}$ and $\mathcal{L}$ are finite and original waiting times are given, the time between two actions in the N-player game is necessary of the form

$$\sum_{l,k} \alpha_{l,k} t_l \left(\frac{k}{N}\right) + \sum_{i \in \mathcal{N}} \beta_i W_0^i \,,$$

where $(\alpha_l)_{l \in \mathcal{L}} \in \{-1,0,1\}^{\mathcal{L}}$, $k \in \{0,\dots,N\}$ and $\beta_i \in \{-1,0,1\}^N$. This generates at most $3^{|\mathcal{L}|(N+1)+N}$ possibilities. Therefore, there is a minimum time between two actions. As the time horizon T is fixed, this implies that there is a maximum number of times M where an action can be taken. Therefore, a policy only needs to define an action on all the possible tuples of times when an action should be taken. The number K of possible tuples is smaller than $\binom{3^{|\mathcal{L}|(N+1)+N}}{M}$. Without loss of generality, we restrict ourselves to the set of policies which is a subset of $\mathcal{M}(\mathcal{L})^{\mathcal{L}\times K}$, making the set of pure policies a subset of $\mathcal{L}^{\mathcal{L}\times K}$ finite, and the set of mixed-policies (its convex hull) a compact subset of the Euclidean vector space $\mathbb{R}^{\mathcal{L}\times\mathcal{L}\times K}$. We will later show that any Nash equilibrium in the set of restricted mixed-policies is a Nash equilibrium in the set of non-restricted mixed-policies.

We denote by $\mathcal{K} \subseteq \mathbb{N}$ the set of pure policy indices. We denote $\mathcal{P} = \{p_k, k \in \mathcal{K}\}$ the set of path (definition 1.2.3) (in this case the set of pure policies). The set of mixed-policies is a subset of the simplex over the set of pure policy $\mathcal{M}(\mathcal{P})$. Therefore, for all $\boldsymbol{\pi} \in \mathcal{M}(\mathcal{P})$, there exists $\alpha \in \mathcal{M}(\mathcal{K})$ identified with a vector of length $|\mathcal{K}|$ such that $\boldsymbol{\pi} = \alpha \cdot (p_k)_{k\in\mathcal{K}}$. We denote by $\underline{\widetilde{\boldsymbol{\Pi}}} \subseteq \mathcal{M}(\mathcal{P})$ the set of restricted mixed-policies and by $\underline{\widetilde{\boldsymbol{\Pi}}} = \widetilde{\boldsymbol{\Pi}}^N$ the set of restricted mixed-policy profiles.

Lemma 2 (Cost is a convex combination of pure strategy cost)**.** *For any $\underline{\boldsymbol{\pi}} \in S$, and for any $i \in \mathcal{N}$ with corresponding $\alpha \in \mathcal{M}(\mathcal{K})$ such that $\boldsymbol{\pi}^i = \alpha \cdot (p_k)_{k\in\mathcal{K}}$*

$$J^i(\boldsymbol{\pi}^i, \underline{\boldsymbol{\pi}}^{-i}) = \sum_{k\in\mathcal{K}} \alpha_k J^i(p_k, \underline{\boldsymbol{\pi}}^{-i}).$$

Proof. The equality is a direct consequence of the linearity of the expected value with respect to the probability distribution. □

Next, let the (set-valued) best response map $\phi : \widetilde{\underline{\mathbf{\Pi}}} \to 2^{\widetilde{\underline{\mathbf{\Pi}}}}$ be defined by for all: for $\underline{\boldsymbol{\pi}} \in \widetilde{\underline{\mathbf{\Pi}}}$,

$$\phi(\underline{\boldsymbol{\pi}}) = \bigtimes_{i \in \mathcal{N}} \operatorname*{argmin}_{\boldsymbol{\pi}' \in \widetilde{\mathbf{\Pi}}} J^i(\boldsymbol{\pi}', \underline{\boldsymbol{\pi}}^{-i}).$$

Lemma 3 (Best response map has a closed graph)**.** *ϕ has a closed graph.*

Proof. For any sequence $(\underline{\boldsymbol{\pi}}_q)_{q \in \mathbb{N}} \in \widetilde{\underline{\mathbf{\Pi}}}^{\mathbb{N}}$ converging to $\underline{\boldsymbol{\pi}}_\infty$ and any sequence $(x_q)_{q \in \mathbb{N}} \in \widetilde{\underline{\mathbf{\Pi}}}^{\mathbb{N}}$ converging to x_∞ such that $x_q \in \phi(\underline{\boldsymbol{\pi}}_q)$ for all $q \in \mathbb{N}$, we have that, for any $i \in \mathcal{N}$:

$$J^i(x_q^i, \underline{\boldsymbol{\pi}}_q^{-i}) \leq \alpha \cdot \left(J^i(p_k, \underline{\boldsymbol{\pi}}_q^{-i})\right)_{k \in \mathcal{K}} \qquad \forall \alpha \in \mathcal{M}(\mathcal{K}),$$

which holds in the limit provided that J^i is continuous w.r.t. the policy profile (our regularity assumption). Therefore, $x_\infty \in \phi(\underline{\boldsymbol{\pi}}_\infty)$ and ϕ has a closed graph. □

Proposition 1 (Best response map has a fixed-point)**.** *The map ϕ is a Kakutani-Glicksberg-Fan map, therefore it admits a fixed-point [99], which is a Nash equilibrium of the N-player game.*

Proof.

The set $\widetilde{\underline{\mathbf{\Pi}}}$ is:

- non-empty (we assume the graph is non-empty, so at least one path exists)
- compact (as a probability distribution over finite set)
- convex (as a probability distribution over finite set)
- subset of a Hausdorff locally convex topological vector space (as a probability distribution over finite set)

The function $\phi(\underline{\boldsymbol{\pi}})$ is non-empty and convex for all $\underline{\boldsymbol{\pi}} \in \underline{\mathbf{\Pi}}$ (because the minimization is a linear problem). The function ϕ has a closed graph (lemma 3). Hence, by Kakutani-Glicksberg-Fan theorem [99], ϕ has a fixed point.

□

The proof of theorem 6.2.1 is concluded by noting that a fixed point $\underline{\boldsymbol{\pi}} \in \widetilde{\underline{\mathbf{\Pi}}}$ of ϕ is a Nash equilibrium in $\underline{\mathbf{\Pi}}$. Indeed, by definition of ϕ:

$$\forall i \in \mathcal{N},\ \forall \boldsymbol{\pi}' \in \widetilde{\mathbf{\Pi}},\ J^i(\boldsymbol{\pi}^i, \underline{\boldsymbol{\pi}}^{-i}) \leq J^i(\boldsymbol{\pi}', \underline{\boldsymbol{\pi}}^{-i}).$$

From here, by lemma 2 and by construction of $\widetilde{\mathbf{\Pi}}$:

$$\forall i \in \mathcal{N},\ \forall \boldsymbol{\pi}' \in \mathbf{\Pi},\ J^i(\boldsymbol{\pi}^i, \underline{\boldsymbol{\pi}}^{-i}) \leq J^i(\boldsymbol{\pi}', \underline{\boldsymbol{\pi}}^{-i}).$$

The last line states that the Nash equilibrium in the set $\widetilde{\underline{\mathbf{\Pi}}}$ is a Nash equilibrium in the set $\underline{\mathbf{\Pi}}$ and finishes the proof of existence of the Nash equilibrium.

□

6.3 Mean-field routing game

Solving the N-player game in section 6.2 is infeasible when N is very large. We thus turn to a MFG version of the above routing game, which can be used to provide approximate Nash equilibria and whose quality improves as $N \to +\infty$. This is based on considering the interactions between a typical player and a distribution representing the rest of the population. This is possible thanks to the anonymity and the symmetry in the interactions, which allows us to focus on symmetric Nash equilibria. Intuitively, the law of large numbers allows considering the state distribution instead of numerous random variables induced by it.

Set-up

The state of a typical player at time t is a random variable denoted by $X_t = (L_t, W_t, D_t)$ which takes values in $\mathcal{X} = \mathcal{L} \times \mathcal{T} \times \mathcal{L}$. At time 0, the population's state distribution is m_0 and is known to the players.

The space of policies is still $\mathbf{\Pi}$. For a policy $\boldsymbol{\pi}$, we denote by $\boldsymbol{\pi}_t(\tilde{l}|l, w, d)$ the probability with which a typical player using the policy $\boldsymbol{\pi}$ would like to go from l to $\tilde{l}$ given that their waiting time is w and their destination is d. The routing random variable is denoted by U.

Assume that an infinitesimal agent uses policy $\boldsymbol{\pi}$ while the rest of the population uses $\boldsymbol{\pi}'$. Let $\boldsymbol{\nu} = (\nu_t)_{t\in\mathcal{T}} \in \mathcal{M}(\mathcal{L})^{\mathcal{T}}$ be the flow of distributions on $\mathcal{L}$ induced by the population that uses $\boldsymbol{\pi}'$. The evolution of a typical player's state is given by the following dynamics. Let $t_0 = 0$ and let (L_0, W_0, D_0) be a given initial state. Then, the dynamics follow:

$$\begin{aligned}
t_{k+1} &= W_{t_k} + t_k \\
L_{t_{k+1}} &= U_{t_k}(X_{t_k}) \\
W_{t_{k+1}} &= t_{L_{t_{k+1}}}\left(\nu_{t_{k+1}}(L_{t_{k+1}})\right) \\
L_t &= L_{t_k} \qquad \forall k, \forall t \in [t_k, t_{k+1}[\\
W_t &= W_{t_k} - (t_k - t) \qquad \forall k, \forall t \in [t_k, t_{k+1}[\\
D_t &= D_0, \qquad t \in \mathcal{T}.
\end{aligned}$$

Here $(t_k)_{k\in\mathbb{N}} \in \mathcal{T}^{\mathbb{N}}$ denotes the sequence of times when the representative player changes link (we take $t_k = T$ when there are no more changes), and $\nu_t(l) \in [0, 1]$ is the proportion of the mean-field population on link l at time t.

The cost of the typical player using policy $\boldsymbol{\pi}$ when the population uses policy $\boldsymbol{\pi}'$ is defined as:

$$J(\boldsymbol{\pi}, \boldsymbol{\pi}') = \mathbb{E}_{\boldsymbol{\pi},\boldsymbol{\pi}'}\left[\int_{t\in\mathcal{T}} r(X_t)dt\right]$$

where the state of the representative player $\boldsymbol{X} = (X_t)_{t\in\mathcal{T}}$ has the above dynamics with policy $\boldsymbol{\pi}$, and the instantaneous cost function r is the same function as in the finite player game (see section 6.2). Analogously to the N-player game, the policy $\boldsymbol{\pi}'$ is used only to deduce

$\boldsymbol{\nu} = (\nu_t)_{t\in\mathcal{T}}$ that appears in the evolution of $\boldsymbol{W}$. So the cost function J could alternatively be written as a function of $(\boldsymbol{\pi}, \boldsymbol{\nu})$ instead of $(\boldsymbol{\pi}, \boldsymbol{\pi}')$. In contrast with the finite player regime, we highlight that here $\boldsymbol{\pi}'$ completely determines $\boldsymbol{\nu}$ because the player under consideration is infinitesimal and hence their policy $\boldsymbol{\pi}$ does not affect the flow $\boldsymbol{\nu}$ of distributions of locations of the population.

Nash equilibrium

The counterpart of the N-player Nash equilibrium in the mean-field regime can now be introduced.

Definition 6.3.1 (Mean-field Nash equilibrium (Definition 3.1. of [197]))**.** *A mean-field Nash equilibrium (MFNE) is a policy $\boldsymbol{\pi}^\star \in \boldsymbol{\Pi}$ such that: $J(\boldsymbol{\pi}^\star, \boldsymbol{\pi}^\star) \leq J(\boldsymbol{\pi}', \boldsymbol{\pi}^\star)$ for all $\boldsymbol{\pi}'$, or equivalently:*

$$\boldsymbol{\pi}^\star \in \underset{\boldsymbol{\pi}\in\boldsymbol{\Pi}}{\operatorname{argmin}}\, J(\boldsymbol{\pi}, \boldsymbol{\pi}^\star).$$

Theorem 6.3.1 (Existence of mean-field Nash equilibrium, Kakutani-Fan-Glisckberg theorem [99])**.** *Assuming the continuity of the cost function with respect to the policy profiles, and assuming that the support of the initial distribution of the waiting time is a finite set, there exists a mean-field Nash equilibrium.*

Proof of theorem 6.3.1. This proof is similar to the proof of theorem 6.2.1 as long as the cost function is continuous with respect to the policy profile, which is assumed here. In the mean-field game, given a departure time the set of pure policies can be restricted to the set of path $\mathcal{P}^{\text{supp}(W_0)}$ given the departure time W_0, with K the number of possible path (with less than a given number of cycles, which is possible due to minimum travel time on each link and finite time horizon) times the number of possible departure time. It is assumed that the support of the departure time random variable $\text{supp}(W_0)$ is finite by design. Therefore, a pure policy, should be understood as choosing a path $p \in \mathcal{P}$ given a departure time W_0. Encoding the policy in $\boldsymbol{\Pi}$ that correspond to any choice of a path given a departure time might require some notation work, that we will not do here. The reader should understand that the proof relies on finding a Nash equilibrium where the set of policies is the set of probability distributions on the set of paths given the departure time, and then showing that this Nash equilibrium policy can be translated into a policy in $\boldsymbol{\Pi}$, and then showing that in the set of $\boldsymbol{\Pi}$ the policy is still a Nash equilibrium. To establish the existence of a Nash equilibrium in the set of probability distribution on the set of paths given a departure time $S = \mathcal{M}\left(\mathcal{P}^{\text{supp}(W_0)}\right)$, the same arguments that the ones in the proof of theorem 6.2.1 can be used with the map $\phi : S \to 2^S$ defined by:

$$\phi(\boldsymbol{\pi}) = \bigtimes_{i\in\mathcal{N}} \underset{\boldsymbol{\pi}'\in S}{\operatorname{argmin}}\, J(\boldsymbol{\pi}', \boldsymbol{\pi})$$

which is a Kakutani-Fan-Glisckberg map. Then, one can convert the path choice given a departure time to a list of actions in time, therefore convert it to a policy $\boldsymbol{\pi}^\star$ in $\boldsymbol{\Pi}$. Then,

one can show by contradiction that they cannot exist a policy $\boldsymbol{\pi}$ in $\boldsymbol{\Pi}$ that gives a strictly better outcome than the policy $\boldsymbol{\pi}^\star$ in the cost function $J(\boldsymbol{\pi}, \boldsymbol{\pi}^\star)$. Therefore, $\boldsymbol{\pi}^\star$ is also a Nash equilibrium in the set $\boldsymbol{\Pi}$. □

The continuity of the cost function with respect to the policies plays a crucial role.

Remark 6.3.1 (Counter-example of the existence of a Nash equilibrium without continuity)**.** *Here, we present a counter-example of the existence of a Nash equilibrium of the mean-field game when congestion functions are not continuous.*

Consider a network with two links, say l, $\tilde{l}$, connecting the origin and destination links. Let the congestion function be fined as:

$$t_l(\mu) = \begin{cases} 1 & \text{if } \mu < 0.5 \\ 2 & \text{otherwise} \end{cases}$$

$$t_{\tilde{l}}(\mu) = \begin{cases} 1 & \text{if } \mu \leq 0.5 \\ 2 & \text{otherwise} \end{cases}$$

Then the mean-field game admits no Nash equilibrium. Indeed, if $\nu_t(l) < 0.5$, then $t_l = 1$ and $t_{\tilde{l}} = 2$, and if $\nu_t(l) \geq 0.5$, then $t_l = 2$ and $t_{\tilde{l}} = 1$. In this case there is no flow allocation such that travel times are equal on the paths that are used, therefore there is no Nash equilibrium. This is due to the discontinuity of the cost of pure actions with respect to the state distribution.

One of the advantages of considering a mean-field setting, is that any MFNE is a dynamic Wardrop equilibrium.

Theorem 6.3.2 (Dynamic Wardrop equilibrium [247])**.** *For any mean-field Nash equilibrium, all induced trajectories of players with the same initial state (origin, waiting time, destination), have the same travel time (i.e., the same total cost).*

Proof of theorem 6.3.2. Lemma 2 is still valid in the mean-field game setup, with the set of pure policies corresponding to a choice of paths. This implies that if the equilibrium policy $\underline{\boldsymbol{\pi}}^\star$ is a mix-policy; i.e., there exist $\mathcal{K}^\star \subset \mathcal{K}$ and $\alpha \in \mathcal{M}(\mathcal{K})$ such that:

$$\begin{aligned} &\underline{\boldsymbol{\pi}}^\star = \alpha \cdot (p_k)_{k \in \mathcal{K}} \\ &\alpha_k > 0 \qquad \forall k \in \mathcal{K}^\star \\ &\alpha_k = 0 \qquad \forall k \in \mathcal{K} \backslash \mathcal{K}^\star. \end{aligned}$$

Then $J(\underline{\boldsymbol{\pi}}^\star, \underline{\boldsymbol{\pi}}^\star) = J(p_k, \underline{\boldsymbol{\pi}}^\star)$ for all $k \in \mathcal{K}^\star$. In the mean-field game, this translates in the equality of the travel time on all path p_k with $k \in \mathcal{K}^\star$. □

Any mean-field Nash equilibrium policy $\boldsymbol{\pi}^\star$ can be used by the players in an N-player game. Intuitively, the larger N is, the closer the population is to the mean-field regime. In

fact, it can be shown under suitable conditions that $\underline{\boldsymbol{\pi}}^\star = (\boldsymbol{\pi}^\star, \dots, \boldsymbol{\pi}^\star) \in \underline{\boldsymbol{\Pi}}$ is an approximate Nash equilibrium whose quality improves with N in the sense that:

$$\bar{\rho}^N(\underline{\boldsymbol{\pi}}^\star) \to 0, \qquad \text{as } N \to +\infty.$$

Where $\bar{\rho}^N(\underline{\boldsymbol{\pi}}^\star)$ is the average deviation incentive (notation 1.2.6) of the optimal mean-field policy $\underline{\boldsymbol{\pi}}^\star$ in the N-player routing game. So if all the agents use the mean-field Nash equilibrium policy, then any single player's incentive to deviate decreases when the population becomes larger. For example, [197] prove in their setting that: if $\boldsymbol{\pi}^\star$ is a MFNE, then for every $\epsilon > 0$, there exists $N_0 \in \mathbb{N}$ such that for every $N \geq N_0$, the N-player policy profile $(\boldsymbol{\pi}^\star, \dots, \boldsymbol{\pi}^\star) \in \underline{\boldsymbol{\Pi}}$ satisfies: $\bar{\rho}^N(\underline{\boldsymbol{\pi}}) \leq \epsilon$.

Remark 6.3.2 (Hamilton-Jacobi-Bellman equation)**.** *One can show that the backward Hamilton-Jacobi-Bellman equation that needs to be solved for the representative player when the player state distribution is known is equivalent to finding the shortest path for the representative player. When several shortest paths exist, any shortest path will minimize the Hamiltonian of the system. The Hamiltonian is a linear combination of the travel time of the different possible paths weighted by the representative player action choice distribution. However, the fix-point condition indicates that the player state distribution should be split between the shortest paths such that their travel times are still equal.*

Next, to illustrate this property in our model, an explicit computation is carried out in the simple Pigou network and then is empirically verified on both Pigou (fig. 1.1) and Braess (fig. 1.2) networks.

Mean-field equilibrium policy in the N-player Pigou game

For the sake of illustration, we present a toy example for which the solution can be computed analytically.

The network (fig. 1.1) has 2 nodes and 2 parallel links, $l, \tilde{l}$, relating these 2 nodes. The cost function is: $t_l(x) = 0.5T$, $t_{\tilde{l}}(x) = xT$ for all $x \in [0, 1]$, where x is the proportion of traffic on each link (this is similar to set the demand $\mathrm{d}_{od} = 1$. The departure time (initial waiting time) is the same for all the agents. The mean-field Nash equilibrium can be computed and yields an equilibrium distribution with proportions $\nu_t(l) = \nu_t(\tilde{l}) = 0.5$. On the other hand, the Nash equilibrium for the N player game is such that $\nu_t(l)$ – the current proportion of players on $\tilde{l}$ – is included in $[\frac{1}{2} - \frac{1}{N}, \frac{1}{2}]$. The average deviation incentive of the mean-field equilibrium policy $\underline{\boldsymbol{\pi}}^\star$ in the N-player game is

$$\bar{\rho}^N(\underline{\boldsymbol{\pi}}^\star) = \frac{T}{N2^N} \sum_{m=1}^{N} \binom{N-1}{m} \max\left\{\frac{N}{2} - m - 1, m + 1 - \frac{N}{2}\right\},$$

which goes to 0 when $N \to \infty$.

Proof. Proof of average deviation incentive of the Pigou mean field equilibrium policy in the corresponding N player game.

For the point of view of one player, if the other players are following the mean field equilibrium policy, then the probability that m other players are on the link $\tilde{l}$ is $\binom{N-1}{m}0.5^{N-1}$. In this case, the travel time on link $\tilde{l}$ is $2\frac{mT}{N}$ if the player does not use it, or $2\frac{(m+1)T}{N}$ if the player uses it. If the player follows the mean field equilibrium policy, with 50% it will use the link l and have a deviation incentive of $0.5T\max\{0, 1-2\frac{m+1}{N}\}$, and with 50% it will use the link $\tilde{l}$ and have a deviation incentive of $0.5T\max\{0, 2\frac{m+1}{N}-1\}$. Therefore, the player deviation incentive is:

$$\begin{aligned}\bar{\rho}^N(\underline{\boldsymbol{\pi}}^\star) &= 0.5T\sum_{m=1}^{N-1}\binom{N-1}{m}0.5^{N-1}\Big(.5\max\{0, 2\frac{m+1}{N}-1\} \\ &\qquad + 0.5\max\{0, 1-2\frac{m+1}{N}\}\Big) \\ &= \frac{T}{N2^N}\sum_{m=1}^{N}\binom{N-1}{m}\max\{\frac{N}{2}-m-1, m+1-\frac{N}{2}\}\end{aligned}$$

□

6.4 Experiments

Experiments show that (1) computing the mean-field equilibrium is easier than computing the N-player Nash equilibrium using state-of-the-art algorithms (sampled counterfactual regret minimization [258]) and (2) it gives an excellent approximation of the N-player equilibrium when N is large (above 30 in the case of the Pigou [184] and the Braess [40] network). The experiments also show that (3) online mirror descent algorithm [182] enables computing the mean-field equilibrium on the Sioux Falls network, a classic use case in road traffic network games [1], with 14,000 vehicles (across two origin-destination pairs) and realistic travel time functions.

Context

All the experiments are conducted within the OpenSpiel framework [138], an open source library that contains a collection of environments and algorithms to apply reinforcement learning and other optimization algorithms in games. The code is publicly available on GitHub [76].

Goal of the experiments. The experiments aim to show that the mean-field equilibrium policy is faster to compute than the N-player policy and approximates well an equilibrium policy in the corresponding N-player game. With this perspective, the mean-field approach solves the curse of dimensionality regarding the number of players in N-player games. Intuitively, the MFG approach is relevant when the number of possible states for any

player is lower than the number of players. In that case, computing the population's distribution probabilities over the possible states is faster than simulating each player trajectory. The approximation is correct when representing the probability distribution over the state space is equivalent to representing the sum of each individual player random variable state, which is the case with large number of players thanks to the central limit theorem [31]. In the MFG, heterogeneity between the players is encoded in the state, to use the same policy for each player without a loss of generality. As an example, in our model, the destination of a player is represented in its state.

Metrics. The quality of the approximation of the Nash equilibrium policy completed by the candidate policy is measured using the average deviation incentive (notation 1.2.6).

Implementation. The N-player game is encoded as a simultaneous, perfect information, general sum game. The corresponding MFG is encoded as a mean-field, perfect information, general sum game. OpenSpiel provides many algorithms to find Nash equilibria of simultaneous games or MFGs. These algorithms include model-free algorithms such as Neural Fictitious Self-Play [112] and model-based algorithms such as Counterfactual Regret Minimization (and some variants) [258] which we use to solve the N-player game. The experiments solve the MFG using the online mirror descent algorithm [182]. The experiments performed in OpenSpiel use a fixed time discretization.

Networks. As classical network games consider demand between nodes, we add artificial origin and destination links before and after each node in the network (Pigou [184], Braess [40] and Sioux Falls). This enables defining vehicle location only using links, and defining state of not having begun a trip and having finished it.

The *Pigou network* [184] has two links l, $\tilde{l}$ and two nodes (an origin and a destination one) which come from the conversion of the origin and the destination nodes (fig. 1.1). A time discretization of 0.01, with a time horizon of 2 is used. The cost functions are $t_l(x) = 2$, $t_{\tilde{l}}(x) = 1 + 2x$ and all the demand leaves the origin link at time 0 and head towards the destination link.

The *Braess network* game is the dynamic extension of the game described in [40] (fig. 1.2). The network has 5 links AB, AC, BC, BD and CD, one origin node A converted to an origin link OA and a destination node D converted to a destination link DE. The cost functions are $t_{AB}(x) = 1 + x$, $t_{AC}(x) = 2$, $t_{AB}(x) = 0.25$, $t_{BD}(x) = 2$, $t_{CD}(x) = 1 + x$. All the demand leaves the origin link at time 0 and head towards the destination link. We use a time step of 0.05 and a time horizon of 5.

In the *augmented Braess network* game, a destination link CF is added to the Braess network (fig. 1.2) and 50 more vehicles leave the origin to DE at time 0, 0.5 and 1, while 50 others leaves the origin to CF at times 0 and 1, totaling 250 vehicles with 2 different destinations and 3 different departure times.

The *Sioux Falls network* game is used by the traffic community for proof of concepts on a network with around 100 links. The network (76 links without the origin and destination links), the link congestion functions, and an origin-destination traffic demand are open source [1] (the network is shown with the results in fig. 6.5). As the classical routing game [173, Chapter 18] is a static game, the demand is only a list of tuple origin, destination

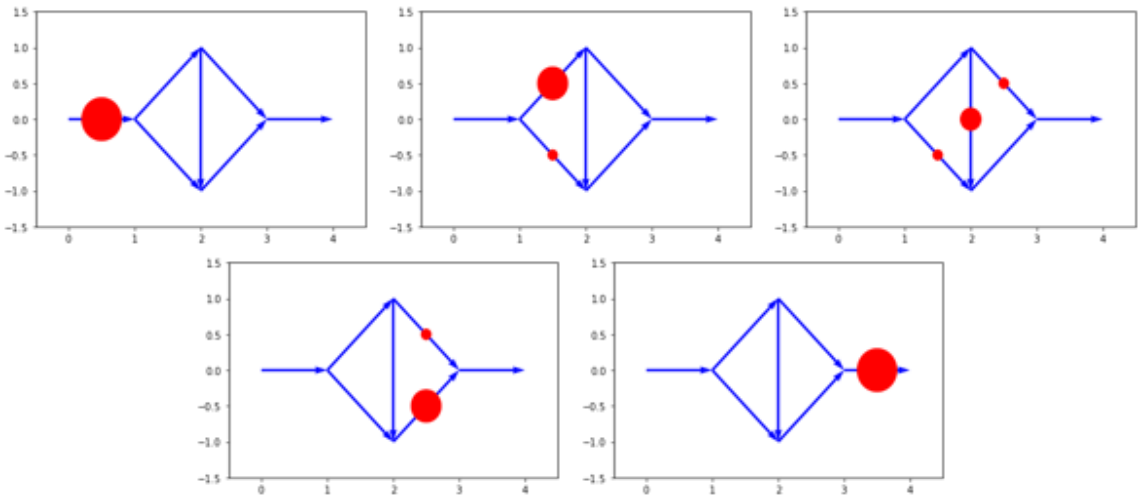


Figure 6.1: Braess network dynamics in the mean-field Nash equilibrium. From left to right, top to bottom: the locations of the cars at time 0.0, from time 0.25 to 1.75, at time 2.0, from time 2.25 to 3.75, at time 4.0. The travel time on each path is equal to 3.75. The travel time equalization characterizes the mean-field Nash equilibrium (theorem 6.3.2). This figure is reproduced from [46, Figure 1].

and counts, and does not provide any departure time. We use the network data (including the congestion functions) and generate a demand specific to the game. We model 7,000 vehicles departing at time 0 from node 1 to node 19, and 7,000 vehicles departing at time 0 from node 19 to node 1. We use a time step of 0.5 and a time horizon of 50.

Mean-field game solves the curse of dimensionality in the number of players

In this section, the mean-field equilibrium policy is computed for both the Braess and the Pigou network games. In addition to being considerably faster to compute compared to the N-player Nash equilibrium, the mean-field equilibrium provides an excellent approximation when N is above 30.

The evolution of the Braess mean-field Nash equilibrium policy is given in fig. 6.1. The travel time on the three possible paths are equals, which encodes the Nash equilibrium condition of the MFG provided that the travel time on each link is a multiple of the time step, accordingly to theorem 6.3.2.

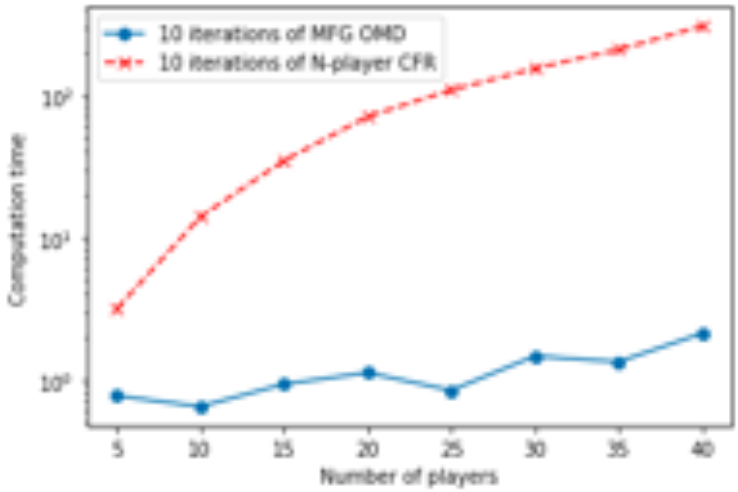


Figure 6.2: Computation time of 10 iterations of Online Mirror Descent in the MFG and of 10 iterations of sampled Counterfactual regret minimization as a function of the number of players N. As a first approximation, the running time of the MFG algorithm does not depend on the number of players. This figure is reproduced from [46, Figure 2].

While solving N-player game is intractable for numerous players, this can be done for the mean-field game.

We compare the running time of the algorithms for solving the N-player game and the mean-field player game, depending on the number of players it models. The counterfactual regret minimization with external sampling (ext CFR) is used in the N-player game, as it is the fastest algorithms to solve the dynamic routing N-player game within the OpenSpiel library of algorithms (comparison done within the OpenSpiel framework are not reported here). Online mirror descent (OMD) is used in the MFG. The running time of 10 iterations of ext CFR and OMD as a function of the number of vehicles modeled are reported in fig. 6.2. On one hand, as the mean-field Nash equilibrium does not depend on the number of vehicles the MFG models, the computation time of 10 iterations of OMD is independent of the number of vehicles modeled. On the other hand, the computational cost of 10 iterations of ext CFR increases exponentially with the number of players, making the computation of a Nash equilibrium with many players intractable using the OpenSpiel library algorithms.

The mean-field equilibrium policy is a good approximation of the N-player equilibrium policy whenever N is large enough.

In the Pigou network game, the mean-field equilibrium policy is almost a Nash equilibrium in the N-player game as soon as N is larger than 20 players (fig. 6.3). This was shown

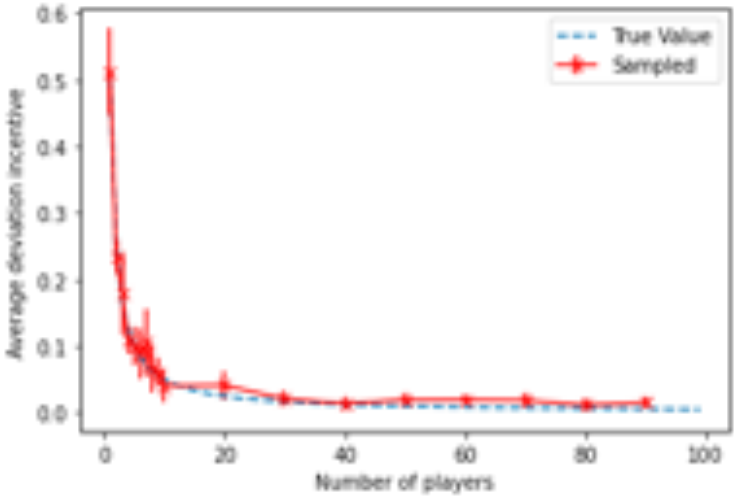


Figure 6.3: Average deviation incentive of the Nash equilibrium mean-field policy in the N-player game as a function of N in the case of the Pigou game. The sampled values are the values computed in OpenSpiel by testing all the possible pure best responses, and sampling game trajectories to get the expected returns. This figure is reproduced from [46, Figure 3].

theoretically in section 6.3, and is confirmed using approximate average deviation incentive of the mean-field equilibrium policy in the N-player game.

In the Braess network game, the mean-field equilibrium policy is almost a Nash equilibrium in the N-player game as soon as N is larger than 30 players.

Experiments show the ability to learn the mean-field equilibrium policy on the 76 links Sioux Falls network, with 14,000 vehicles going to two different destinations. Using online mirror descent, we see that the average deviation incentive decreases to 1.55 (for a travel time of 27) over 100 iterations, see fig. 6.4. We use a fixed learning rate of 1 in the 30 first iterations of the algorithm, 0.1 in the 31 to the 60 first iterations and a fixed learning rate of 0.01 in the 40 remaining iterations to produce fig. 6.4.

The resulting mean-field policy is not exactly the Nash equilibrium policy of the MFG as its average deviation incentive is 1.55 (for a travel time of 27.5). The game evolution displayed in fig. 6.5 shows that some vehicles going from node 19 to node 1 have a longer travel time than others: on time step 26.5 (section 6.4) some vehicles have arrived to node 1 (top left) and some have not.

Average deviation of the learned mean-field policy cannot be computed numerically in the 14,000 player game, due to the large number of players.

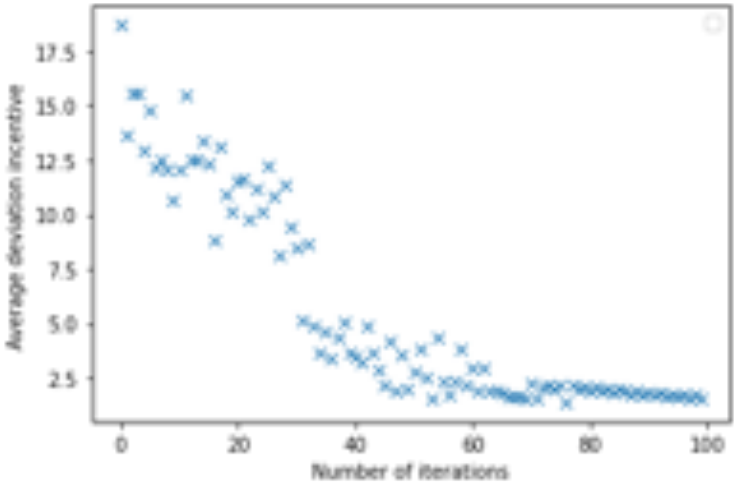


Figure 6.4: Online mirror descent average deviation incentive in the Sioux Falls MFG as a function of the number of iterations of the descent algorithm. After 100 iterations of the algorithm, the average deviation incentive is 1.55. This figure is reproduced from [46, Figure 4].

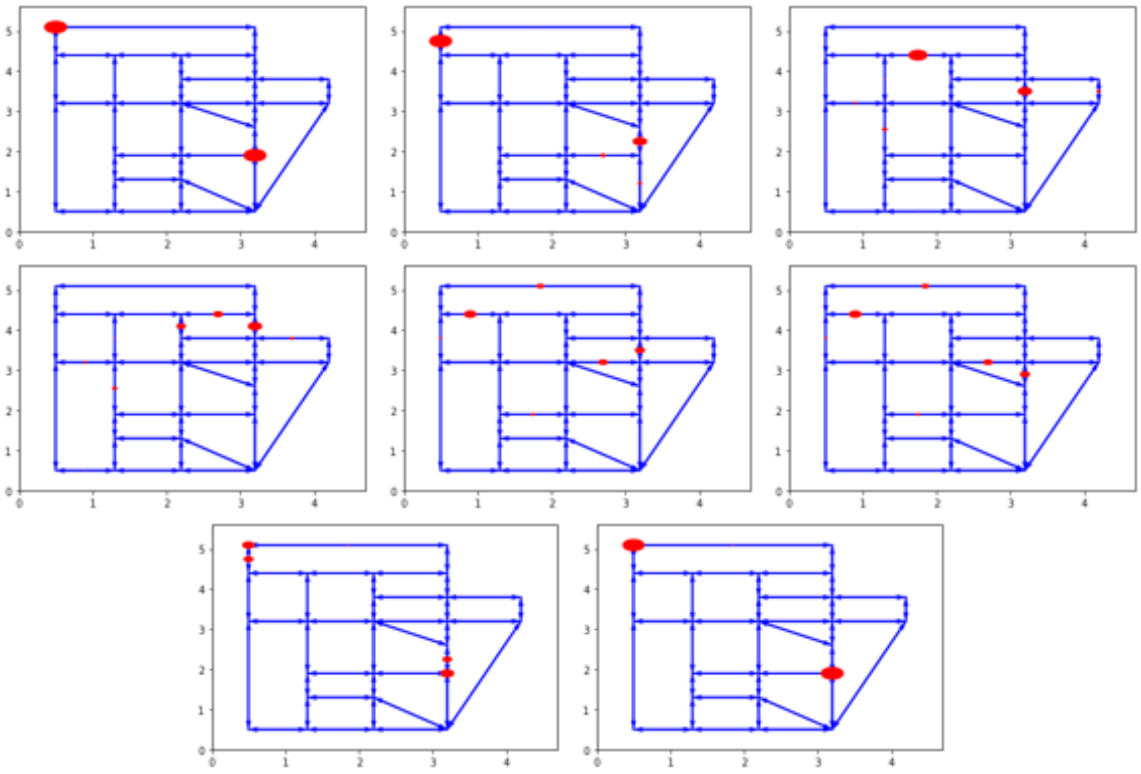


Figure 6.5: Sioux Falls network dynamics in the mean-field Nash equilibrium. From left to right, top to bottom: the locations of the cars at time 0.0, 2.5, 10.0, 12.5, 21.0, 22.0, 26.5, 27.5. Some vehicles arrived at their destination before some other vehicles that left the origin at the same time: the Nash equilibrium has not been reached. On average, players can expect saving 1.55 time by being the only one to be rerouted on a better path. This figure is reproduced from [46, Figure 5].

Chapter 7

Conclusion

During more than 5 years of collaboration with the mobile sensing lab, I was able to show that **information-aware routing steers traffic to a Nash equilibrium**. For that purpose, I developed the *restricted path-choice model* picturing that the vehicles that do not have access to information are following known paths, and that the vehicles that have access to information are minimizing their experienced travel time. I showed that the *average deviation incentive*, the *average marginal regret*, the *average counterfactual regret*, the *exploitability*, and the *relative gap to the user equilibrium* are the same mathematical object used in different contexts. I demonstrated that the *average deviation incentive decrease monotonically to* 0 *with an increase of information-aware routing behaviors in the restricted path-choice model*. I further showed on microsimulation that information-aware routing also decreases the average deviation incentive of the vehicles. In addition, the microsimulations exhibited that navigational-apps should take into account their own impact on traffic in their traffic predictions when they are used by more than 30% of the vehicles. On small thought experiments, I illustrated that information-aware routing behaviors might not increase the transportation efficiency.

I connected the traffic engineering literature with case studies about the impact of cut-through traffic induced by information-aware routing behaviors. I displayed that the travel time of some parallel paths to the I-210 in Los Angeles, CA became balanced with the highway travel time during peak hours, giving clues about the existence and the intensity of information-aware routing behaviors. I gave evidences about cut-through traffic induced by these information-aware routing behaviors in the Los Angeles, CA network. Case studies indicate that in Fremont, CA, during the 2019 evening peak hours, up to 10% of the traffic supposed to take the I-680 will take local roads instead. In Pleasanton, CA, cut-through traffic accounted for up to 84% of the traffic on some local roads in 2016.

In order to help cities to improve possible mitigation techniques against cut-through traffic, I have provided a *blueprint for developing, calibrating, and validating traffic microsimulations*. I used this sketch to create a traffic microsimulation of a Fremont neighborhood. For the sake of reproducibility, the full process has been made open source, along with the Fremont traffic microsimulation. Finally, I have developed the *mean-field routing game*, to

enable simulating promptly how individual dynamic routing behaviors impact the overall state of traffic. The mean-field routing game enables simulating numerous vehicles together, as illustrated on the Sioux Falls network with 76 links and 100 time steps with 14,000 vehicles.

The way forward

The Ph.D. journey helped me envision what the future of transportation research should be focusing on. First, in my opinion, the research community should continue bridging economics, urban planning, traffic engineering, and optimal control of dynamical systems when working on transportation related problems. Second, the transportation planning community should establish and communicate clear criteria to know when case studies, A/B experiments or simple models should be used instead of simulations. As part of this recommendation, the community should also establish which type of simulations (land-use, travel demand, traffic model, micro/macro/meso, static/dynamic) should be used based on the use cases that cities have.

Within the transportation research community working on transportation simulation, the effort to develop large-scale simulation models should be continued to improve traffic at the metropolitan scale. Especially, with the rise of information-aware routing behaviors, I believe that the traffic demand (i.e., the output of the travel demand models) should not include the individual route choice anymore. I expect that mean-field routing games will be a useful tool to portray route choice efficiently at a metropolitan scale. Moreover, standardized validation toolboxes for traffic microsimulation should be developed. These toolboxes should leverage the work done in machine learning about model validation. In particular, all traffic models and all traffic simulations should be validated against ground truth with data separated between training and testing data. Building on this dissertation, the traffic simulation community can develop a calibration toolbox for traffic microsimulations. I believe that new tools can be brought to the transportation data analysis (like Koopman operator, principal component analysis, or sensitivity analysis). I also think that the work done to decrease the level of under-determination in traffic simulator calibration can leverage work done by the machine learning community (for example, using robust principal component analysis on the OD demand matrices).

Finally, I think that the transportation research community should try as much as possible to stay connected to the needs of transportation planners and engineers by working on ground truth use cases. Especially, research projects should be built in collaboration with cities and traffic engineers.

Bibliography

[1] Mustafa Abdulaal and Larry J LeBlanc. "Continuous equilibrium network design models". In: *Transportation Research Part B: Methodological* 13.1 (1979), pp. 19–32.

[2] Yves Achdou and Jean-Michel Lasry. "Mean field games for modeling crowd motion". In: *Contributions to partial differential equations and applications*. Springer, 2019.

[3] Yves Achdou et al. "Finite horizon mean field games on networks". In: *Calculus of Variations and Partial Differential Equations* 59.5 (2020), pp. 1–34.

[4] Alameda County Transportation Commission. *Interstate 680 Sunol Express Lanes (Phase 1 and Phase 2)*. Aug. 2018.

[5] Clémence Alasseur, Imen Ben Taher, and Anis Matoussi. "An extended mean field game for storage in smart grids". In: *Journal of Optimization Theory and Applications* 184.2 (2020).

[6] Alissa Walker for Gizmodo. *Is It Really Possible To Trick Waze To Keep Traffic Off Your Street?* 2014. URL: https://gizmodo.com/is-it-really-possible-to-trick-waze-to-keep-traffic-off-1660273215 (visited on 07/15/2022).

[7] American Association of State Highways and Transportation Officials (AASHTO). *A Policy on Geometric Design of Highways and Streets*. 2013, p. 934. ISBN: 9781560515081.

[8] American Association of State Highways and Transportation Officials (AASHTO). *Guidelines for geometric design of very low-volume local roads (ADT$¡=$400)*. Vol. L. 202. 2001, p. 72. ISBN: 1560511664.

[9] C Antoniou et al. *Fundamentals of Traffic Simulation*. Ed. by Jaume Barceló. Vol. 145. International Series in Operations Research & Management Science. New York, NY: Springer New York, 2010. ISBN: 978-1-4419-6141-9. DOI: 10.1007/978-1-4419-6142-6. URL: http://iet.jrc.ec.europa.eu/%20http://link.springer.com/10.1007/978-1-4419-6142-6.

[10] Archival Econmic data, St.Louis FED. *E-Commerce Retail Sales as a Percent of Total Sales*. 2019. URL: https://alfred.stlouisfed.org/series?seid=ECOMPCTSA (visited on 08/06/2019).

[11] Richard Arnott and Kenneth Small. "The economics of traffic congestion". In: *American scientist* 82.5 (1994), pp. 446–455.

[12] Neha Arora et al. "An Efficient Simulation-Based Travel Demand Calibration Algorithm for Large-Scale Metropolitan Traffic Models". In: *arXiv* (2021). DOI: https://doi.org/10.48550/arXiv.2109.11392. URL: https://arxiv.org/abs/2109.11392.

[13] Neha Arora et al. "Quantifying the sustainability impact of Google Maps: A case study of Salt Lake City". In: *arXiv preprint arXiv:2111.03426* (2021).

[14] Assemblée nationale, France (French Parliament). *LOI n° 2021-1104 du 22 août 2021 portant lutte contre le dérèglement climatique et renforcement de la résilience face à ses effets, Article 109*. https://www.legifrance.gouv.fr/eli/loi/2021/8/22/2021-1104/jo/article_109. 2021.

[15] Assemblée nationale, France (French Parliament). *LOI n° 2021-1104 du 22 août 2021 portant lutte contre le dérèglement climatique et renforcement de la résilience face à ses effets, Article 122*. https://www.legifrance.gouv.fr/eli/loi/2021/8/22/2021-1104/jo/article_122. 2021.

[16] Joshua Auld et al. "POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations". In: *Transportation Research Part C: Emerging Technologies* 64 (2016), pp. 101–116.

[17] Alexander Aurell and Boualem Djehiche. "Mean-field type modeling of nonlocal crowd aversion in pedestrian crowd dynamics". In: *SIAM J. Control Optim.* 56.1 (2018), pp. 434–455. ISSN: 0363-0129. DOI: 10.1137/17M1119196. URL: https://doi.org/10.1137/17M1119196.

[18] Alexander Aurell et al. "Optimal incentives to mitigate epidemics: a Stackelberg mean field game approach". In: *arXiv preprint arXiv:2011.03105* (2020).

[19] Sangjae Bae et al. *Behavior, Energy, Autonomy, Mobility Modeling Framework*. Tech. rep. 7 Summits IT AG LTD, Zurich, Zurich; Skylite Networks, Fremont, CA; Lawrence ..., 2019.

[20] Melissa Balding et al. *Estimated TNC Share of VMT in Six US Metropolitan Regions*. Fehr & Peers, 2019. URL: https://1y4yclbm79aqghpm1xoezrdw-wpengine.netdna-ssl.com/wp-content/uploads/2019/08/TNC-VMT-Findings.pdf.

[21] Michael Balmer et al. "MATSim-T: Architecture and simulation times". In: *Multi-agent systems for traffic and transportation engineering*. IGI Global, 2009, pp. 57–78.

[22] Albert-László Barabási, Réka Albert, and Hawoong Jeong. "Mean-field theory for scale-free random networks". In: *Physica A: Statistical Mechanics and its Applications* 272.1-2 (1999), pp. 173–187.

[23] Dana Bartholomew. "Sherman Oaks residents blame Waze navigation app for clogging streets." In: *Los Angeles Daily News* (Jan. 2017).

[24] Tamer Başar and Geert Jan Olsder. *Dynamic noncooperative game theory*. Vol. 23. Siam, 1999.

[25] Dario Bauso, Xuan Zhang, and Antonis Papachristodoulou. "Density flow in dynamical networks via mean-field games". In: *IEEE Transactions on Automatic Control* 62.3 (2016), pp. 1342–1355.

[26] Bea Karnes. *Speed Lumps Being Installed to Slow Traffic*. 2016. URL: https://patch.com/california/fremont/speed-lumps-being-installed-slow-traffic (visited on 09/28/2019).

[27] Michael Behrisch et al. "SUMO–simulation of urban mobility: an overview". In: *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind. 2011.

[28] Genrich Belitskii et al. *Matrix norms and their applications*. Vol. 36. Birkhäuser, 2013.

[29] Michael GH Bell. "The estimation of origin-destination matrices by constrained generalised least squares". In: *Transportation Research Part B: Methodological* 25.1 (1991), pp. 13–22.

[30] E. Bert, A. Torday, and A. Dumont. "Calibration of Urban Network Microsimulation Models". In: *Proc., 5th Swiss Transport Research Conf., Ascona, Switzerland* January (2005).

[31] Dimitri P Bertsekas and John N Tsitsiklis. *Introduction to probability*. Vol. 1. Athena Scientific Belmont, MA, 2002.

[32] Chandra R. Bhat and Frank S. Koppelman. "Activity-Based Modeling of Travel Demand". In: *Handbook of Transportation Science*. Boston: Kluwer Academic Publishers, 2006, pp. 39–65. DOI: 10.1007/0-306-48058-1_3. URL: http://link.springer.com/10.1007/0-306-48058-1_3.

[33] Jeb Bing. *City cameras now watch your car, but not you*. 2002. URL: https://www.pleasantonweekly.com/morgue/2002/2002_05_10.cameras10.html (visited on 07/15/2022).

[34] Laura Bliss. *U.S. Transportation Funding Is Not Created Equal*. 2017. URL: https://www.citylab.com/transportation/2017/07/us-transportation-funding-is-not-created-equal/534327/ (visited on 08/18/2019).

[35] Avrim Blum, Eyal Even-Dar, and Katrina Ligett. "Routing without regret: on convergence to nash equilibria of regret-minimizing algorithms in routing games". In: *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*. Vol. 6. PODC '06 1. New York, NY, USA: ACM, 2006, pp. 45–52. DOI: 10.4086/toc.2010.v006a008. URL: http://www.theoryofcomputing.org/articles/v006a008.

[36] Transportation Economics Committee Transportation Research Board. *Transportation Benefit-Cost Analysis*. URL: http://bca.transportationeconomics.org/.

[37] Geoff Boeing. "OSMnx: A Python package to work with graph-theoretic OpenStreetMap street networks". In: *Journal of Open Source Software* 2.12 (2017).

[38] Bob Booth, Andy Mitchell, et al. *Getting started with ArcGIS*. 2001.

[39] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004. ISBN: 0521833787.

[40] Dietrich Braess. "Über ein Paradoxon aus der Verkehrsplanung". In: *Unternehmensforschung* 12.1 (1968), pp. 258–268.

[41] Elmar Brockfeld, Reinhart D Kühne, and Peter Wagner. "Calibration and Validation of Microscopic Traffic Flow Models". In: *Transportation Research Record* 1876.1 (2004), pp. 62–70. DOI: 10.3141/1876-07. URL: https://doi.org/10.3141/1876-07.

[42] Brookings. *New e-commerce entry Jet means rock-bottom prices ... and more city trucks*. 2015. URL: https://www.brookings.edu/blog/the-avenue/2015/07/28/new-e-commerce-entry-jet-means-rock-bottom-prices-and-more-city-trucks/ (visited on 08/06/2019).

[43] George W Brown. "Iterative solution of games by fictitious play". In: *Activity analysis of production and allocation* 13.1 (1951), pp. 374–376.

[44] Bruce Schaller. *The New Automobility: Lyft, Uber and the future of American cities*. 2018. URL: http://www.schallerconsult.com/rideservices/automobility.pdf.

[45] Theophile Cabannes et al. "Measuring Regret in Routing: Assessing the Impact of Increased App Usage". In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* November (2018), pp. 2589–2594. DOI: 10.1109/ITSC.2018.8569758.

[46] Theophile Cabannes et al. "Solving N-Player Dynamic Routing Games with Congestion: A Mean-Field Approach". In: *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. 2022, pp. 1557–1559.

[47] Théophile Cabannes et al. "Regrets in Routing Networks". In: *ACM Transactions on Spatial Algorithms and Systems* 5.2 (July 2019), pp. 1–19. ISSN: 23740353. DOI: 10.1145/3325916. URL: http://dl.acm.org/citation.cfm?doid=3350424.3325916.

[48] Théophile Cabannes et al. "The impact of GPS-enabled shortest path routing on mobility: a game theoretic approach". In: *Transportation Research Board 97th Annual Meeting* (2018).

[49] California State Legislature. *AB-1605 City and County of San Francisco: Crooked Street Reservation and Pricing Program.* https://leginfo.legislature.ca.gov/faces/billVotesClient.xhtml?bill_id=201920200AB1605. 2019. (Visited on 08/06/2022).

[50] Gordon DB Cameron and Gordon ID Duncan. "PARAMICS—Parallel microscopic simulation of road traffic". In: *The Journal of Supercomputing* 10.1 (1996), pp. 25–53.

[51] Pierre Cardaliaguet and Charles-Albert Lehalle. "Mean field game of controls and an application to trade crowding". In: *Mathematics and Financial Economics* 12.3 (2018).

[52] Tom Carmody et al. "2012 - forbes etal. - Methods and Practices for Setting Speed Limits-An Informational Report.pdf". In: ().

[53] Ennio Cascetta et al. "A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks". In: *Transportation and Traffic Theory. Proceedings of The 13th International Symposium On Transportation And Traffic Theory, Lyon, France, 24-26 July 1996*. 1996.

[54] CBS Los Angeles. *Steep Street Plagued By Navigation App Traffic Will Go One-Way*. 2018. URL: https://losangeles.cbslocal.com/2018/05/21/navigation-app-street-goes-one-way/ (visited on 08/15/2019).

[55] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

[56] Cy Chan et al. "Quasi-Dynamic Traffic Assignment using High Performance Computing". In: *Open Access Publications from the University of California* (2021). URL: https://escholarship.org/uc/item/0f09r1x7.

[57] Qianwen Chao et al. "A Survey on Visual Traffic Simulation: Models, Evaluations, and Applications in Autonomous Driving". In: *Computer Graphics Forum* 39.1 (2020), pp. 287–308. DOI: https://doi.org/10.1111/cgf.13803. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13803.

[58] Geoffroy Chevalier, Jerome Le Ny, and Roland Malhamé. "A micro-macro traffic model based on mean-field games". In: *2015 American Control Conference (ACC)*. IEEE. 2015, pp. 1983–1988.

[59] Yi-Chang Chiu et al. "Dynamic traffic assignment: A primer (Transportation Research Circular E-C153)". In: (2011).

[60] Vidar Christiansen and Stephen Smith. "Externality-Correcting Taxes and Regulation". In: *The Scandinavian journal of economics* 114.2 (2012), pp. 358–383.

[61] City of Fremont. *Taming Traffic in Fremont*. Tech. rep. 2017, pp. 2–3. URL: https://fremont.gov/DocumentCenter/View/36420/FINAL-Traffic-Newsletter-11-1-17?bidId=.

[62] City Of Los Angeles. *Complete Streets*. Tech. rep. 2014, p. 252.

[63] City of Pleasanton. *Traffic Calming Program*. Tech. rep. 2012. URL: https://www.cityofpleasantonca.gov/civicax/filebank/blobdload.aspx?BlobID=23868.

[64] City of St. Louis. "Report of the Transportation Survey Commission of the City of St. Louis". In: (1930).

[65] Civilengineeringbible.com. *What is Transportation Engineering?* URL: https://civilengineeringbible.com/article.php?i=113 (visited on 09/28/2018).

[66] R. Clewlow and G. Mishra. "Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States (NO. UCD-ITS-RR-17-07). University of Clifornia, Davis, Institute of Transportation Studies, Davis, CA." In: October (2017).

[67] R. H. Coase. "The Problem of Social Cost". In: *The Journal of Law & Economics* (1960). URL: www.jstor.org/stable/724810.

[68] Code général des collectivités territoriales. *Article L2213-4*. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000039783412/. 1996.

[69] Serdar Çolak, Antonio Lima, and Marta C González. "Understanding congested travel in urban areas". In: *Nature communications* 7 (2016), p. 10793.

[70] Maud Coud\'ene and David Levy. "De plus en plus de personnes travaillent en dehors de leur commune de r\'esidence". In: *Insee Premi\'ere* 1605.Juin (2016), pp. 1–4.

[71] Adolf D.May. *The Highway Congestion Problem and the Role of In-Vehicle Information Systems*. 1989.

[72] Stella Dafermos and Anna Nagurney. "Sensitivity analysis for the asymmetric network equilibrium problem". In: *Mathematical programming* 28.2 (1984), pp. 174–184.

[73] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. "The complexity of computing a Nash equilibrium". In: *SIAM Journal on Computing* 39.1 (2009), pp. 195–259.

[74] Quentin David, Matteo Del Fabbro, Paul Vertier, et al. "Etude sur la "gratuité" des transports en commun à Paris". In: *Paris, France* (2018).

[75] Kalyanmoy Deb and Himanshu Jain. "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints". In: *IEEE transactions on evolutionary computation* 18.4 (2013), pp. 577–601.

[76] DeepMind. *Open Spiel*. https://github.com/deepmind/open_spiel. Accessed: 2022-08-07.

[77] Peter A Diamond. "Consumption externalities and imperfect corrective pricing". In: *The Bell Journal of Economics and Management Science* (1973), pp. 526–538.

[78] Edsger W Dijkstra. "A note on two problems in connexion with graphs". In: *Edsger Wybe Dijkstra: His Life, Work, and Legacy*. 2022, pp. 287–290.

[79] Josu Doncel, Nicolas Gast, and Bruno Gaujal. "A mean field game analysis of sir dynamics with vaccination". In: *Probability in the Engineering and Informational Sciences* (2020), pp. 1–18.

[80] Warren F Dorsey. *Highway advisory radio operational site survey and broadcast equipment guide*. Tech. rep. 1979.

[81] Mariagrazia Dotoli, Maria Pia Fanti, and Carlo Meloni. "A signal timing plan formulation for urban traffic control". In: *Control Engineering Practice* 14.11 (2006), pp. 1297–1311. ISSN: 09670661. DOI: 10.1016/j.conengprac.2005.06.013.

[82] Anthony Downs. "The law of peak-hour expressway congestion". In: *Traffic Quarterly* 16.3 (1962).

[83] Norman R Draper and Harry Smith. *Applied regression analysis*. Vol. 326. John Wiley & Sons, 1998.

[84] Romuald Elie, Emma Hubert, and Gabriel Turinici. "Contact rate epidemic control of COVID-19: an equilibrium view". In: *Mathematical Modelling of Natural Phenomena* (2020).

[85] Romuald Elie, Tomoyuki Ichiba, and Mathieu Laurière. "Large banking systems with default and recovery: A mean field game model". In: *arXiv preprint arXiv:2001.10206* (2020).

[86] Martin Fellendorf and Peter Vortisch. "Microscopic traffic flow simulator VISSIM". In: *Fundamentals of traffic simulation*. Springer, 2010, pp. 63–93.

[87] Jay M. Finkelman et al. "Noise and driver performance." In: *Journal of Applied Psychology* 62.6 (1977), pp. 713–718. ISSN: 1939-1854. DOI: 10.1037/0021-9010.62.6.713. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0021-9010.62.6.713.

[88] Simon Fischer, Harald Räcke, and Berthold Vöcking. "Fast convergence to Wardrop equilibria by adaptive sampling methods". In: *SIAM Journal on Computing* 39.8 (2010), pp. 3700–3735.

[89] Félix-Antoine Fortin et al. "DEAP: Evolutionary Algorithms Made Easy". In: *Journal of Machine Learning Research* 13 (July 2012), pp. 2171–2175.

[90] Mogens Fosgerau. "The valuation of travel-time variability". In: *Internatioanl Transport Forum* (2017), pp. 39–56. DOI: 10.1787/9789282108093-3-en.

[91] Fredrick Kunkle for the Washington Post. *Everybody stop: N.J. finds simple way to cut crashes involving pedestrians*. 2017. URL: https://www.washingtonpost.com/news/tripping/wp/2017/07/18/everybody-stop-n-j-finds-simple-way-to-cut-crashes-involving-pedestrians/ (visited on 07/13/2022).

[92] Avner Friedman. *Differential games*. Courier Corporation, 2013.

[93] Terry Friesz et al. "A variational inequality formulation of the dynamic network user equilibrium problem". In: *Operations Research* 41.1 (1993), pp. 179–191.

[94] Masao Fukushima. "A modified Frank-Wolfe algorithm for solving the traffic assignment problem". In: *Transportation Research Part B: Methodological* 18.2 (1984), pp. 169–177.

[95] Emanuele Galli et al. "ActivitySim: large-scale agent-based activity generation for infrastructure simulation". In: *Proceedings of the 2009 spring simulation multiconference*. 2009, pp. 1–9.

[96] NJ Garber and LA Hoel. "Fundamental Principles of Traffic Flow". In: *Traffic and highway engineering*. 4th Ed. Cengage Learning, 2009, pp. 213–214.

[97] Steven R Gehrke, Alison Felix, and Timothy Reardon. "Fare choices: A survey of ride-hailing passengers in metro Boston". In: *Metropolitan Area Planning Council* (Feb. 2018), p. 19. URL: http://www.mapc.org/wp-content/uploads/2018/02/Fare-Choices-MAPC.pdf.

[98] GitHub. *Project planning for developers*. https://github.com/features/issues. Accessed: 2022-08-07.

[99] Irving L Glicksberg. "A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points". In: *Proceedings of the American Mathematical Society* 3.1 (1952), pp. 170–174.

[100] Chris Godsil and Gordon Royle. *Algebraic Graph Theory*. Vol. 207. Springer New York, 2001. ISBN: 978-0-387-95220-8. DOI: 10.1007/978-1-4613-0163-9. URL: http://link.springer.com/10.1007/978-1-4613-0163-9.

[101] Mark Goh. "Congestion management and electronic road pricing in Singapore". In: *Journal of transport geography* 10.1 (2002), pp. 29–38.

[102] Google. *Direction API*. https://developers.google.com/maps/documentation/directions. Accessed: 2022-08-07.

[103] Google. *Google privacy*. 2022. URL: https://policies.google.com/privacy?hl=en-US (visited on 08/07/2022).

[104] Google Maps. *Google Maps Eco-Friendly Routing: Accelerating the journey to sustainability for one billion people*. 2021. URL: https://www.gstatic.com/gumdrop/sustainability/google-maps-eco-friendly-routing.pdf (visited on 07/09/2022).

[105] Jason B Gordon, Haris N Koutsopoulos, and Nigel HM Wilson. "Estimation of population origin–interchange–destination flows on multimodal transit networks". In: *Transportation Research Part C: Emerging Technologies* 90 (2018), pp. 350–365.

[106] Governor Gavin Newsom. *AB-1605 Veto letter*. https://www.gov.ca.gov/wp-content/uploads/2019/10/AB-1605-Veto-Message.pdf. 2019. (Visited on 08/06/2022).

[107] Bruce C. Greenwald and Joseph E. Stiglitz. "Externalities in Economies with Imperfect Information and Incomplete Markets". In: *The Quarterly Journal of Economics* (1986). URL: www.jstor.org/stable/1891114.

[108] Olivier Guéant. "Existence and uniqueness result for mean field games with congestion effect on graphs". In: *Applied Mathematics & Optimization* 72.2 (2015), pp. 291–303.

[109] Mordechai Haklay and Patrick Weber. "Openstreetmap: User-generated street maps". In: *IEEE Pervasive computing* 7.4 (2008), pp. 12–18.

[110] Michael A. Hall. "Properties of the Equilibrium State in Transportation Networks". In: *Transportation Science* 12.3 (1978), pp. 208–216. ISSN: 00411655, 15265447. URL: http://www.jstor.org/stable/25767914.

[111] Ke Han, Terry L Friesz, and Tao Yao. "Existence of simultaneous route and departure choice dynamic user equilibrium". In: *Transportation Research Part B: Methodological* 53 (2013), pp. 17–30.

[112] Johannes Heinrich and David Silver. "Deep reinforcement learning from self-play in imperfect-information games". In: *arXiv preprint arXiv:1603.01121* (2016).

[113] Ronald Paul Hill and Myron Laible. "Changeable message signs: a technology whose time has come". In: *Journal of Public Policy & Marketing* 16.1 (1997), pp. 173–176.

[114] James R Hines. "Three sides of Harberger triangles". In: *Journal of Economic Perspectives* 13.2 (1999), pp. 167–188.

[115] Yaron Hollander and Ronghui Liu. "The principles of calibrating traffic microsimulation models". In: *Transportation* 35.3 (2008), pp. 347–362. ISSN: 00494488. DOI: 10.1007/s11116-007-9156-2.

[116] Kuang Huang et al. "A game-theoretic framework for autonomous vehicles velocity control: Bridging microscopic differential games and macroscopic mean field games". In: *Discrete & Continuous Dynamical Systems - B* 22.11 (2017). ISSN: 1553-524X. DOI: 10.3934/dcdsb.2020131. URL: http://dx.doi.org/10.3934/dcdsb.2020131.

[117] Kuang Huang et al. "Dynamic driving and routing games for autonomous vehicles on networks: A mean field game approach". In: *Transportation Research Part C: Emerging Technologies* 128 (2021), p. 103189.

[118] Minyi Huang, Roland P. Malhamé, and Peter E. Caines. "Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle". In: *Communications in Information and Systems* 6.3 (2006). ISSN: 1526-7555. URL: http://projecteuclid.org/euclid.cis/1183728987.

[119] DE Huntington and CS Lyrintzis. "Improvements to and limitations of Latin hypercube sampling". In: *Probabilistic engineering mechanics* 13.4 (1998), pp. 245–253.

[120] DRIEAT Île-de-France. "Enquête globale de transport". In: *Enquêtes transport et déplacements* (2019).

[121] Jennifer Olney. *Bay Area transportation experts say apps may make traffic worse.* 2018. URL: https://abc7news.com/technology/bay-area-transportation-experts-say-apps-may-make-traffic-worse/4688216/ (visited on 09/28/2019).

[122] Jeremy Walsh for Pleasanton Weekly. *City OKs partial street closure to curtail cut-through traffic*. 2018. URL: https://pleasantonweekly.com/news/2018/02/20/city-oks-partial-street-closure-to-curtail-cut-through-traffic (visited on 07/08/2022).

[123] Jeremy Walsh for Pleasanton Weekly. *City OKs partial street closure to curtail cut-through traffic*. 2018. URL: https://pleasantonweekly.com/news/2018/02/20/city-oks-partial-street-closure-to-curtail-cut-through-traffic (visited on 07/15/2022).

[124] John Cichowski for North Jersey. *How Leonia found smooth sailing with its last-ditch traffic fix to stop bridge commuters*. 2018. URL: https://www.northjersey.com/story/news/columnists/john-cichowski/2018/01/27/leonias-last-ditch-traffic-fix-stop-george-washington-bridge-commuters/1068170001/ (visited on 08/15/2019).

[125] Joshua Jongsma. *Reports: Weehawken joins Leonia in closing local roads during commute*. 2018. URL: https://www.northjersey.com/story/news/bergen/leonia/2018/02/13/reports-weehawken-joins-leonia-closing-local-roads-during-commute/335521002/ (visited on 09/28/2019).

[126] Alireza Khani et al. *Integration of the FAST-TrIPs person-based dynamic transit assignment model, the SF-CHAMP regional, activity-based travel demand model, and san francisco's citywide dynamic traffic assignment model*. Tech. rep. 2013.

[127] A. Khiyami, A. Keimer, and A. Bayen. "Structural Analysis of Specific Environmental Traffic Assignment Problems". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Nov. 2018, pp. 2327–2332. DOI: 10.1109/ITSC.2018.8569488.

[128] Aman Kishore et al. "Prevention of Highway Infrastructure Damage Through Commercial Vehicle Weight Enforcement Annual Indian Roads Congress (Irc) Session". In: 306 (2000).

[129] Lawrence A Klein, Milton K Mills, David R P Gibson, et al. *Traffic detector handbook: Volume I*. Tech. rep. Turner-Fairbank Highway Research Center, 2006.

[130] Frank H Knight. "Some fallacies in the interpretation of social cost". In: *The Quarterly Journal of Economics* 38 (4 1924), pp. 582–606.

[131] Rachel Kraus. *Out of traffic, into a ditch: Why Waze on snowy mountain roads could be a bad idea*. URL: https://finance.yahoo.com/news/traffic-ditch-why-waze-snowy-192643966.html.

[132] Walid Krichene, Benjamin Drighès, and Alexandre Bayen. "On the convergence of no-regret learning in selfish routing". In: *Proceedings of the 31st International Conference on Machine Learning* 32.2 (2014), pp. 163–171. URL: http://proceedings.mlr.press/v32/krichene14.html.

[133] Kristie Cattafi for North Jersey. *Leonia, trying to block commuters from its local roads, applauds latest court decision*. 2020. URL: https://eu.northjersey.com/story/news/bergen/leonia/2020/03/07/leonia-nj-george-washington-bridge-commuter-ban-gets-new-life/4975666002/ (visited on 07/15/2022).

[134] Mobile Sensing lab. *Open source static traffic assignment solver*. https://github.com/megacell. Accessed: 2018-09-25.

[135] Mobile Sensing lab. *Traffic microsimulation toolbox*. https://github.com/Fremont-project/traffic-microsimulation. Accessed: 2022-07-15.

[136] Aimé Lachapelle and Marie-Therese Wolfram. "On a mean field game approach modeling congestion and aversion in pedestrian crowds". In: *Transportation research part B: methodological* 45.10 (2011), pp. 1572–1589.

[137] Aimé Lachapelle et al. "Efficiency of the price formation process in presence of high frequency participants: a mean field game analysis". In: *Mathematics and Financial Economics* 10.3 (2016).

[138] Marc Lanctot et al. "OpenSpiel: A Framework for Reinforcement Learning in Games". In: *CoRR* abs/1908.09453 (2019). arXiv: 1908.09453 [cs.LG]. URL: http://arxiv.org/abs/1908.09453.

[139] Jean-Michel Lasry and Pierre-Louis Lions. "Mean field games". In: *Japanese Journal of Mathematics* 2.1 (2007). ISSN: 0289-2316. DOI: 10.1007/s11537-007-0657-8. URL: http://dx.doi.org/10.1007/s11537-007-0657-8.

[140] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), pp. 436–444.

[141] Leslie Fulbright for SFGate. *I-580 home to 3 of Bay Area's worst traffic bottlenecks / East Bay gridlock fueled by growth in San Joaquin Valley*. 2005. URL: https://www.sfgate.com/bayarea/article/I-580-home-to-3-of-Bay-Area-s-worst-traffic-2706604.php (visited on 07/15/2022).

[142] Feng Li, Roland P Malhamé, and Jerome Le Ny. "Mean field game based control of dispersed energy storage devices with constrained inputs". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 2016.

[143] Edward Lieberman and Ajay K. Rathi. "Traffic simulation". In: *Federal Highway Administration* (1992). ISSN: 0097-8418. DOI: 10.1145/274790.273160.

[144] Michael James Lighthill and Gerald Beresford Whitham. "On kinematic waves II. A theory of traffic flow on long crowded roads". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 229.1178 (1955), pp. 317–345.

[145] Lisa W. Foderaro for NYT. *Navigation Apps Are Turning Quiet Neighborhoods Into Traffic Nightmares*. URL: https://www.nytimes.com/2017/12/24/nyregion/traffic-apps-gps-neighborhoods.html (visited on 12/24/2017).

[146] Todd Litman. "Evaluating Accessibility for Transport Planning: Measuring People's Ability to Reach Desired Goods and Activities. Victoria Transport Policy Institute.—Canada". In: *Victoria Transportation Policy Institute* January 2008 (2012).

[147] Todd Alexander Litman. *Transportation Cost and Benefit Analysis: Techniques, Estimates and Implications*. Victoria Transport Policy Institute, 2009.

[148] William Forster Lloyd. *Two lectures on the checks to population. England: Oxford University*. Oxford University, 1833.

[149] Edward Lockhart et al. "Computing approximate equilibria in sequential adversarial games by exploitability descent". In: *arXiv preprint arXiv:1903.05614* (2019).

[150] Tim Lomax, David Schrank, Bill Eisele, et al. "2021 urban mobility report". In: (2021).

[151] Hong Q Lu and Priya Nimbole. "Intro to TransCAD GIS". In: *Model Research and Development Unit Transportation Planning Branch* (2002).

[152] Jane Macfarlane. "When apps rule the road: The proliferation of navigation apps is causing traffic chaos. It's time to restore order". In: *IEEE Spectrum* 56.10 (2019), pp. 22–27.

[153] Jenny Mageean and John D Nelson. "The evaluation of demand responsive transport services in Europe". In: *Journal of Transport Geography* 11.4 (2003), pp. 255–270.

[154] Palak Maheshwary et al. "A methodology for calibration of traffic micro-simulator for urban heterogeneous traffic operations". In: *Journal of Traffic and Transportation Engineering (English Edition)* 7.4 (2020), pp. 507–519. ISSN: 20957564. DOI: 10.1016/j.jtte.2018.06.007. URL: https://doi.org/10.1016/j.jtte.2018.06.007.

[155] Thomas Robert Malthus. *An essay on the principle of population: Or, a view of its past and present effects on human happiness, with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions*. London, Reeves and Turner, 1798.

[156] Aarian Marshall. *There are better ways to kill traffic than lying to Waze*. July 2016. URL: https://www.wired.com/2016/07/better-ways-kill-traffic-lying-waze/.

[157] Alfred Marshall. *Principles of Economics*. Great Mind Series, 1890.

[158] Karl Marx. *Capital. A Critique of Political Economy*. Verlag von Otto Meisner, 1867.

[159] Karl Marx and Friedrich Engels. *The Communist Manifesto*. London: Communistischer Arbeiterbildungsverein, 1848.

[160] Millard McElwee, Bingyu Zhao, and Kenichi Soga. "Real-time Analysis of City Scale Transportation Networks in New Orleans Metropolitan Area using an Agent Based Model Approach". In: *MATEC Web of Conferences* 271.8 (Apr. 2019). Ed. by H. Sadek, p. 06007. ISSN: 2261-236X. DOI: 10.1051/matecconf/201927106007. URL: http://leader.pubs.asha.org/article.aspx?doi=10.1044/leader.PPL.21082016.20%20https://www.matec-conferences.org/10.1051/matecconf/201927106007.

[161] Michael G McNally. "The four-step model". In: *Handbook of transport modelling*. Emerald Group Publishing Limited, 2007.

[162] Aditya Krishna Menon et al. "Fine-grained OD estimation with automated zoning and sparsity regularisation". In: *Transportation Research Part B: Methodological* 80 (2015), pp. 150–172.

[163] Seyedali Mirjalili. "Genetic algorithm". In: *Evolutionary algorithms and neural networks*. Springer, 2019, pp. 43–55.

[164] Dov Monderer and Lloyd S Shapley. "Potential games". In: *Games and economic behavior* 14.1 (1996), pp. 124–143.

[165] Roger B Myerson. *Game theory*. Harvard university press, 2013.

[166] A. Nagurney. *Network Economics: A Variational Inequality Approach*. Advances in Computational Economics. Springer US, 1998. ISBN: 9780792383505. URL: https://books.google.com/books?id=dJwL96FVPE8C.

[167] John Nash. "Equilibrium points in n-person games". In: *Proceedings of the National Academy of Sciences* 36 (1 1950), pp. 48–49.

[168] National Academies of Sciences, Engineering, and Medicine and others. "Highway Capacity Manual". In: (2016).

[169] National Academies of Sciences, Engineering, and Medicine and others. "Travel demand forecasting: Parameters and techniques". In: (2012).

[170] Pascal Neis, Peter Singler, and Alexander Zipf. "Collaborative mapping and Emergency Routing for Disaster Logistics - Case studies from the Haiti earthquake and the UN portal for Afrika The Case of Africa : A Geoportal for the UN Joint Logistics Cluster". In: *GI_Forum* December 2014 (2010), p. 6.

[171] B Y David M Newbery. "Road Damage Externalities and Road User Charges Author (s): David M . Newbery Published by : The Econometric Society Stable URL : http://www.jstor.org/stable/1911073 . ROAD DAMAGE EXTERNALITIES AND ROAD USER CHARGES ' that road users create externalitie". In: *Econometrica* 56.2 (1988), pp. 295–316.

[172] Nicholas Tufnell for Wired. *Students hack Waze, send in army of traffic bots*. 2014. URL: https://www.wired.co.uk/article/waze-hacked-fake-traffic-jam (visited on 07/15/2022).

[173] N. Nisam et al. *Algorithmic game theory*. Cambridge University Press, 2007.

[174] Raymond W. Novaco and Oscar I. Gonzalez. "Commuting and well-being". In: *Technology and Psychological Well-being*. Ed. by Yair Amichai-Hamburger. August. Cambridge: Cambridge University Press, 2009, pp. 174–205. ISBN: 9780511635373. DOI: 10.1017/CBO9780511635373.008. URL: https://www.cambridge.org/core/product/identifier/CBO9780511635373A016/type/book_part.

[175] Samantha Raphelson for NPR. *New Jersey Town Restricts Streets From Commuters To Stop Waze Traffic Nightmare*. 2018. URL: https://www.npr.org/2018/05/08/609437180/new-jersey-town-restricts-streets-from-commuters-to-stop-waze-traffic-nightmare (visited on 08/15/2019).

[176] Jennifer Olney. *Bay Area transportation experts say apps may make traffic worse*. Nov. 2018.

[177] Kaleb Osagie. *The Waze Effect: 4 Steps For Cities To Fight Back*. 2018. URL: https://www.streetlightdata.com/waze-effect-4-steps-for-cities-to-fight-back/.

[178] Irena Ištoka Otković, Aleksandra Deluka-Tibljaš, and Sanja Šurdonja. "Validation of the calibration methodology of the micro-simulation traffic model". In: *Transportation Research Procedia* 45.2019 (2020), pp. 684–691. ISSN: 23521465. DOI: 10.1016/j.trpro.2020.02.110. URL: https://doi.org/10.1016/j.trpro.2020.02.110.

[179] Institute Transportation Studies Partners for Advanced Transportation Technology. *Connected Corridors*. 2018. URL: https://connected-corridors.berkeley.edu/resources/document-library (visited on 05/01/2018).

[180] Anthony D. Patire et al. "How much GPS data do we need?" In: *Transportation Research Part C: Emerging Technologies* 58 (2015), pp. 325–342. ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2015.02.011. URL: http://www.sciencedirect.com/science/article/pii/S0968090X15000662.

[181] Michael. Patriksson. *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015. ISBN: 0486787907.

[182] Julien Perolat et al. "Scaling up Mean Field Games with Online Mirror Descent". In: *arXiv preprint arXiv:2103.00623* (2021).

[183] Matti Persula. "Simulation of Traffic Systems - An Overview". In: *Journal of Geographic Information and Decision Analysis* 3.1 (1999), pp. 1–8. ISSN: 1435-5949.

[184] Arthur Cecil Pigou. *The Economics of Welfare*. 1920. DOI: 10.4337/9781788118569.00060.

[185] Abdul Pinjari et al. "Cemdap: Modeling and microsimulation frameworks, software development, and verification". In: *Proceedings of the transportation research board 87th annual meeting, Washington DC*. 2008.

[186] Thiago Henrique Poiani et al. "Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the Nepal earthquake 2015". In: *Proceedings of the Annual Hawaii International Conference on System Sciences* 2016-March (2016), pp. 188–197. ISSN: 15301605. DOI: 10.1109/HICSS.2016.31.

[187] Jacopi Prisco. *Why UPS trucks (almost) never turn left*. 2017. URL: https://www.cnn.com/2017/02/16/world/ups-trucks-no-left-turns/index.html.

[188] Rachel Dovey. *Lawsuit Filed Against N.J. Town That Banned Waze Drivers*. 2018. URL: https://nextcity.org/daily/entry/lawsuit-filed-against-nj-town-that-banned-waze-drivers (visited on 09/28/2019).

[189] Gabriel de O Ramos, Ana LC Bazzan, and Bruno C da Silva. "Analysing the impact of travel information for minimising the regret of route choice". In: *Transportation Research Part C: Emerging Technologies* 88 (2018).

[190] LA Rastrigin. "Random search as a method for optimization and adaptation". In: *Stochastic Optimization*. Springer, 1986, pp. 534–544.

[191] Gary Richards. *Is Google's Waze app making traffic worse?* Jan. 2016.

[192] Robert Rees and Ronald Ramos, Fehr & Peers. *SR-262 (Mission Boulevard) Cross Connector Project – Local Street Analysis*. 2019. URL: https://www.alamedactc.org/wp-content/uploads/2020/12/SR_262-Local_Existing_Memorandum_Final_Dec2019.pdf (visited on 08/06/2022).

[193] Dan A Rosen, Frank J Mammano, and Rinaldo Favout. "An electronic route-guidance system for highway vehicles". In: *IEEE Transactions on Vehicular Technology* 19.1 (1970), pp. 143–152.

[194] Robert W. Rosenthal. "A class of games possessing pure-strategy Nash equilibria". In: *International Journal of Game Theory* 2.1 (Dec. 1973), pp. 65–67. ISSN: 1432-1270. DOI: 10.1007/BF01737559. URL: https://doi.org/10.1007/BF01737559.

[195] Tim Roughgarden and Eva Tardos. "How bad is selfish routing?" In: *Journal of the ACM (JACM)* 49.2 (Mar. 2002), pp. 236–259. ISSN: 00045411. DOI: 10.1145/506147.506153. URL: http://doi.acm.org/10.1145/506147.506153%20http://portal.acm.org/citation.cfm?doid=506147.506153.

[196] Ryan Bradley for The New Yorker. *Waze and the traffic panopticon*. 2015. URL: https://www.newyorker.com/business/currency/waze-and-the-traffic-panopticon (visited on 07/15/2022).

[197] Naci Saldi, Tamer Başar, and Maxim Raginsky. "Markov–Nash Equilibria in Mean-Field Games with Discounted Cost". In: *SIAM Journal on Control and Optimization* 56.6 (2018), pp. 4256–4287.

[198] Rabih Salhab, Jerome Le Ny, and Roland P Malhamé. "A mean field route choice game model". In: *2018 IEEE Conference on Decision and Control (CDC)*. IEEE. 2018, pp. 1005–1010.

[199] Sumudu Samarakoon et al. "Energy-efficient resource management in ultra dense small cell networks: A mean-field approach". In: *IEEE Global Communications Conference (GLOBECOM)*. 2015.

[200] San Francisco County Transportation Authority. *Lombard Study : Managing Access to the " Crooked Street "*. City of San Francisco, 2017. URL: https://www.sfcta.org/sites/default/files/2019-03/Lombard_existing_conditions_report_092816.pdf.

[201] San Francisco County Transportation Authority and Cambridge Systematics Inc. "San Francisco Travel Demand Forecasting Model Development: Executive Summary". In: 29 (2002).

[202] William H. Sandholm. "Potential Games with Continuous Player Sets". In: *Journal of Economic Theory* 97.1 (2001), pp. 81–108. ISSN: 00406376. DOI: 10.1136/thx.2005.042960.

[203] Agnar Sandmo. *Pigouvian Taxes*. London: Palgrave Macmillan UK, 2016, pp. 1–4. ISBN: 978-1-349-95121-5. DOI: 10.1057/978-1-349-95121-5_2678-1. URL: https://doi.org/10.1057/978-1-349-95121-5_2678-1.

[204] David Schmeidler. "Equilibrium points of nonatomic games". In: *Journal of statistical Physics* 7.4 (1973), pp. 295–300.

[205] David Schrank, Bill Eisele, and Tim Lomax. "2019 Urban Mobility Report". In: *Texas A&M Transportation Institute* August (2019).

[206] Sajjad Shafiei, Ziyuan Gu, and Meead Saberi. "Calibration and validation of a simulation-based dynamic traffic assignment model for a large-scale congested network". In: *Simulation Modelling Practice and Theory* 86 (Aug. 2018), pp. 169–186. ISSN: 1569190X. DOI: 10.1016/j.simpat.2018.04.006.

[207] Zhenyu Shou and Xuan Di. "Multi-Agent Reinforcement Learning for Dynamic Routing Games: A Unified Paradigm". In: *CoRR* abs/2011.10915 (2020). arXiv: 2011.10915. URL: https://arxiv.org/abs/2011.10915.

[208] Florian Siebel and Wolfram Mauser. "On the fundamental diagram of traffic flow". In: *SIAM Journal on Applied Mathematics* 66.4 (2006), pp. 1150–1162.

[209] Adam Smith. *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations*. W. Strahan and T. Cadell, 1776.

[210] We are social. *Digital in 2019*. 2019. URL: https://wearesocial.com/uk/digital-2019 (visited on 08/11/2019).

[211] Statista. *Global digital population as of July 2019*. 2019. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ (visited on 08/11/2019).

[212] Statista. *Monthly number of Uber's active users worldwide from 2016 to 2019 (in millions)*. 2019. URL: https://www.statista.com/statistics/833743/us-users-ride-sharing-services/ (visited on 08/06/2019).

[213] Statista. *Most popular mapping apps in the United States as of April 2018, by monthly users*. 2019. URL: https://www.statista.com/statistics/865413/most-popular-us-mapping-apps-ranked-by-audience/ (visited on 08/11/2019).

[214] Steve Hendrix for The Washington Post. *Traffic-weary homeowners and Waze are at war, again. Guess who's winning?* 2016. URL: https://www.washingtonpost.com/local/traffic-weary-homeowners-and-waze-are-at-war-again-guess-whos-winning/2016/06/05/c466df46-299d-11e6-b989-4e5479715b54_story.html (visited on 07/15/2022).

[215] Steve Lopez for the Los Angeles Times. *On one of L.A.'s steepest streets, an app-driven frenzy of spinouts, confusion, and crashes*. 2018. URL: https://www.latimes.com/local/california/la-me-lopez-echo-park-traffic-20180404-story.html (visited on 07/13/2022).

[216] Knowledgebase on Sustainable Urban Land use and Transport. *A Policy Guidebook*. Intelligent Energy Europe Programme of the European Union, 2016. URL: http://www.konsult.leeds.ac.uk/pg/.

[217] Svetlana Shkolnikova and Joshua Jongsma for North Jersey. *Business owners protest as new traffic signage planned in Leonia*. 2018. URL: https://www.northjersey.com/story/news/bergen/leonia/2018/02/14/new-signage-aims-make-leonia-more-accessible-business-protest-planned/339241002/ (visited on 08/15/2019).

[218] Tala Salem. *Why some cities have had enough of Waze*. 2018. URL: https://www.usnews.com/news/national-news/articles/2018-05-07/why-some-cities-have-had-enough-of-waze (visited on 05/07/2018).

[219] Takashi Tanaka et al. "Linearly solvable mean-field traffic routing games". In: *IEEE Transactions on Automatic Control* 66.2 (2020), pp. 880–887.

[220] Mostafa H. Tawfeek et al. "Calibration and validation of micro-simulation models using measurable variables". In: *12th International Transportation Specialty Conference 2018, Held as Part of the Canadian Society for Civil Engineering Annual Conference 2018* 6 (2019), pp. 12–22.

[221] Brian Taylor et al. "The importance of commercial vehicle weight enformcement in safety and road asset management". In: *Traffic Technology International* 2000 Annual Review (2000), pp. 234–237. URL: http://engrwww.usask.ca/entropy/tc/publications/pdf/irdtraffictechwhyweighv2finalpostedpdf.pdf.

[222] Dušan Teodorović et al. "Transportation Demand Analysis". In: *Transportation Engineering* (Jan. 2017), pp. 495–568. DOI: 10.1016/B978-0-12-803818-5.00008-1. URL: https://www.sciencedirect.com/science/article/pii/B9780128038185000081?via%3Dihub.

[223] Christophe Terrier. "Les d\'eplacements domicile-travail en France : \'evolution de 1975 à 1982". In: *Espace, populations, soci\'et\'es* 4.2 (1986), pp. 333–342. ISSN: 0755-7809. DOI: 10.3406/espos.1986.1145. URL: https://www.persee.fr/doc/espos_0755-7809_1986_num_4_2_1145.

[224] J. Thai, N. Laurent-Brouty, and A. M. Bayen. "Negative externalities of GPS-enabled routing applications: A game theoretical approach". In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. Nov. 2016, pp. 595–601. DOI: 10.1109/ITSC.2016.7795614.

[225] The Fremont Mobility Task Force. *Fremont Mobility Action Plan*. City of Fremont, 2019. URL: https://www.fremont.gov/DocumentCenter/View/40583/FREMONT-Mobility-Action-Plan-Final-3-1-19?bidId=.

[226] Tony Dutzik and Gideon Weissman from Frontier. *Who Pays for Roads*. 2015. URL: https://frontiergroup.org/reports/fg/who-pays-roads (visited on 08/18/2019).

[227] Transport Simulation Software. *Aimsun Next 22*.

[228] Transport Simulation Software. *Aimsun Next 22.0.1 Users Manual: OutputDatabaseDefinition*. https://docs.aimsun.com/next/22.0.1/UsersManual/OutputDatabaseDefinition.html. Accessed: 2022-08-07.

[229] Martin Treiber and Arne Kesting. *Traffic flow dynamics: Data, models and simulation*. Springer Berlin Heidelberg, 2013, pp. 1–503. ISBN: 9783642324604. DOI: 10.1007/978-3-642-32460-4. URL: http://link.springer.com/10.1007/978-3-642-32460-4.

[230] Albert W Tucker. "A two-person dilemma". In: *Prisoner's Dilemma* (1950).

[231] U.S. Department of Transportation Federal Highway Administration. *Guidebook on the Utilization of Dynamic Traffic Assignment in Modeling*. May 2020. URL: https://ops.fhwa.dot.gov/publications/fhwahop13015/sec2.htm.

[232] U.S. Department of Transportation Federal Highway Administration. *Highway Functional Classification: Concepts Criteria, and Procedure*. 2013. URL: https://trid.trb.org/view/1266295.

[233] U.S. Department of Transportation Federal Highway Administration. *Non-Interstate System Toll Roads in the United States*. 2015. URL: https://www.fhwa.dot.gov/policyinformation/tollpage/t1part4.cfm (visited on 07/15/2022).

[234] U.S. Department of Transportation Federal Highway Administration. *Road Closure and Lane Closure*. 2019. URL: https://ops.fhwa.dot.gov/wz/construction/full_rd_closures.htm (visited on 06/25/2019).

[235] U.S. Department of Transportation Federal Highway Administration. *The active transportation and demand management program (ATDM): Lessons learned*. 2013. URL: http://ops.fhwa.dot.gov/publications/fhwahop13018/fhwahop13018.pdf.

[236] U.S. Department of Transportation Federal Highway Administration. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software 2019 Update to the 2004 Version*. Apr. 2019. URL: https://ops.fhwa.dot.gov/publications/fhwahop18036/index.htm (visited on 07/14/2022).

[237] United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects: The 2017 Revision, Online Demographic Profiles*. 2017. URL: https://population.un.org/wpp/Graphs/DemographicProfiles/ (visited on 06/01/2019).

[238] United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision*. 2018. URL: https://population.un.org/wpp/Graphs/DemographicProfiles/ (visited on 06/01/2019).

[239] United States. Bureau of Public Roads. *Traffic assignment manual for application with a large, high speed computer*. Vol. 2. US Department of Commerce, Bureau of Public Roads, Office of Planning, Urban Planning Division, 1964.

[240] US Department of Transportation. *Manual on Uniform Traffic Control Devices; for Streets and Highways*. US Department of Transportation, Federal Highway Administration., 2009.

[241] US Department Of Transportation Federal Highway Administration. *Next Generation SIMulation (NGSIM)*. 2005. URL: www.ngsim-community.org..

[242] Pravin Varaiya. *Freeway Performance Measurement System (PeMS), PeMS 6*. Tech. rep. California Center for Innovative Transportation, University of California, Berkeley, Feb. 2006.

[243] Pravin Varaiya. "Freeway Performance Measurement System (PeMS), Version 4". In: UCB-ITS-PR.April (2004).

[244] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[245] Paul Waddell. "UrbanSim: Modeling urban development for land use, transportation, and environmental planning". In: *Journal of the American planning association* 68.3 (2002), pp. 297–314.

[246] Abraham Wandersman and Maury Nation. "Urban neighborhoods and mental health: Psychological contributions to understanding toxicity, resilience, and interventions." In: *American Psychologist* 53.6 (1998), pp. 647–656. ISSN: 1935-990X. DOI: 10.1037/0003-066X.53.6.647. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0003-066X.53.6.647.

[247] John Glen Wardrop. "Some Theoretical Aspects of Road Traffic Research." In: *Proceedings of the Institution of Civil Engineers* 1 (3 May 1952), pp. 325–362. ISSN: 1753-7789. DOI: 10.1680/ipeds.1952.11259. URL: http://www.icevirtuallibrary.com/doi/10.1680/ipeds.1952.11259.

[248] Waze Connected Citizen Program. *Case Studies*. 2019. URL: https://www.waze.com/en/ccp/casestudies (visited on 08/14/2019).

[249] Michael Wegener. "Overview of Land Use Transport Models". In: Aug. 2004, pp. 127–146. DOI: 10.1108/9781615832538-009. URL: http://www.emeraldinsight.com/doi/10.1108/9781615832538-009.

[250] Byung-Wook Wie. "A differential game model of Nash equilibrium on a congested traffic network". In: *Networks* 23.6 (1993), pp. 557–565.

[251] Wiki Waze. *Routing server*. 2019. URL: https://wiki.waze.com/wiki/Routing_server (visited on 06/13/2019).

[252] Wikipedia. *Glossary of mathematical symbols*. 2022. URL: https://en.wikipedia.org/wiki/Glossary_of_mathematical_symbols (visited on 07/13/2022).

[253] Wisconsin Department of Transportation. "Variable-message sign". In: *Intelligent Transportation Systems (ITS), Design Manual*. 2000. Chap. 6. URL: http://www4.uwm.edu/cuts/itsdm/chap6.pdf.

[254] Cathy Wu et al. "Flow: Architecture and benchmarking for reinforcement learning in traffic control". In: *arXiv preprint arXiv:1710.05465* 10 (2017).

[255] Yun Xu and Royston Goodacre. "On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning". In: *Journal of analysis and testing* 2.3 (2018), pp. 249–262.

[256] Naomi Zeveloff. "Israelis Sue Waze Navigation App for Creating Neighborhood Traffic Jam." In: *Forward* (Dec. 2016).

[257] Hong Zheng et al. "A primer for agent-based simulation and modeling in transportation applications". In: (2013).

[258] Martin Zinkevich et al. "Regret minimization in games with incomplete information". In: *Advances in neural information processing systems* 20 (2007), pp. 1729–1736.