The impact of information-aware routing on road traffic

Theophile Cabannes

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2023-52
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2023/EECS-2023-52.html

May 1, 2023

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
Control, Intelligent Systems, and Robotics

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 dissertation of Theophile Charles Prosper Cabannes, titled The impact of information-aware routing on road traffic
From case studies to game-theoretical analysis and simulations, is approved:

Chair _____ Date _____

_____ Date _____

_____ Date _____

University of California, Berkeley

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
Control, Intelligent Systems, and Robotics

University of California, Berkeley

Professor Alexandre M. Bayen, Chair

During the 2010s decade, the increase of connectivity in the world has led to the development of navigation applications that help vehicles to travel within the transportation network using real-time traffic information. *Information-aware routing* has 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. As a consequence, with these new routing behaviors, some traffic plans are now outdated.

From a game theoretical point of view, by providing vehicles with the fastest path to reach their destination, information-aware routing suggestions have directed the state of traffic toward a *Nash equilibrium*. In game theory, 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) have increased the overall transportation efficiency, some local roads now 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. To our knowledge, 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 that are 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 due to the lack of available floating-car data (trajectory data). Qualitatively, many cities and neighborhoods have 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, Fremont, CA described 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 are changing the design of their road network. Others are using 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 that 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 applications 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 had been 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. I have not heard about U.C. Berkeley by then, and I was not able to say if San Francisco and California were on the West coast or on the East coast of the U.S.. However, doing an internship with Alex Bayen seemed my best options 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 during 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. Beside influencing the beginning of my professional life, Alex has had a tremendous impact on my personal life. First, Alex enabled me feeling home 10,000 km away from Paris: we speak the same language, went to the same school, had the same beginning of 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. Especially, I am delighted that you let me coordinating the work of 16 GSIs for the EECS127/227AT class. The feeling I had when giving a 45 minutes lecture about PCA that involves a (successful) live experiment in front of 400 students will always remain in my memory. Thank you, Alex!

Over 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 learn 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 actually 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 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 selfless will continue to strike me for many years. I hope that I could follow your example, and keeping my peace of mind when I will be hit hard by the workload. Thank you, Eric, for all your support and your leading example!

Along the people I met during the course of 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 the 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 the fate, again, that lead you to suggest me to join full-time Google researches the day when my flight back to the U.S. was cancelled, because the 3rd COVID-19 wave was rising all over the world?

During the 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. Then, I am grateful to Laurent El-Ghaoui that taught the EECS227B class, and has always helped me during the Ph.D.. Shankar Sastry has blown my mind in 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 with 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 facts that in the discussion I had had with Claire and Shankar, they have never said something that seemed irrelevant to me. I was blessed to meet Scott Shenker during 30 minutes to talk about the 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 been 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 that give 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.

Beside outstanding professors, I met remarkable students and researchers. First, my labmates were a second family to me. I have been involved as the "Chief Hapiness Officier" of the lab in order 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 definitely 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 much when Marine was visiting. I am grateful to the fun people group: Lucas Fischer, Vincent Figheria, Marco Sangiovanni-Vincentelli, Fangyu Wu, Jiayi Li, Marsalis Gibson, Zhe Fu, Askhan Yousefpour, Nathan Litche, Shuxia Tang, Joy Carpio, and Yiling You, and to the other labmates: Jessica Lazarus, Aboudy Kriechen, Alexander Keimer, Ayush Jain, Harry Dong, Junghwan Lee, Jon Davis, Eugene Vinitsky, Cathy Wu, 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 larger re-used for the 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, 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 would 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 some time challenging. I hope that I was a great manager for this project.

Finally, at the U.C. Berkeley side, 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 to 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, and Ruben Lozano-Aguilera. It is amazing to see how much impact your team can have with the new eco-routing feature in Google Maps. Then, I am appreciated to the mean-field game (MFG) team between DeepMind and Google Brain that welcomed me in France. Thank you for your help and your support, Romuald Elie and Olivier Pietquin. Thank you for the theoretical discussions,

Mathieu Lauriere and Paul Muller. Thank you for the code reviews, Marc Lanctot, Sergan Girgin, Julien Perolat, and Raphael Marinier. Beside the MFG team, thank you to the policy team that have also welcomed me in 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 that gave me freedom in my personal life, while giving me strong incentives in my professional life. I would like to thank my mother that gave 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 with my twin brother (that does very similar research as me) about mathematics. Thank you to my aunts, to my uncles, and my cousins. Thank you to Meg, that introduced me to Linda, that kindly hosted me during 5 months at Berkeley in a hippy-like fashion. Thank you to my parents-in-law, and to 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, 6^th, 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

To do. This thesis aims to understand ...

1.2 Contributions

To do. Its contribution is too

1.3 Necessary background: definitions and notations

To do. In this section, we present some elementary notions from mathematics, game theory, traffic engineering, and probabilities, 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 definitions

- Numbers $\mathbb{N}$ and $\mathbb{R}$
- Sets A^B
- Measures $\mathcal{P}(X)$

- Norms $\|x\|_p$
- Optimization min, argmin

Elementary game theory definitions

- Player / agent
- State
- Trajectory
- Action / Strategy
- Policy
- Best response
- Deviation incentive
- Nash equilibrium
- ϵ-Nash equilibrium.

Fundamental traffic engineering definitions

- Vehicle
- Network
- Link
- Path
- Flow (link flow and path flow)
- Travel time (link and path)
- Demand
  - Trips
  - TAZs
  - Matrix

Elementary probability theory definitions

- Random variables
- Stochastic processes
- Expectations
- Probabilities

Make sure that all notations in macro.tex are defined here.

Part I

Traffic control in the age of information technology

Chapter 2

Traffic engineering 101

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. Naturally, the transportation system has a tremendous impact on the daily lives and well-being of both the users and nonusers of the system at any given time. Transportation system **users** benefit from access to the transportation system in so much as they are able to 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 [151]. 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 [50]. 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 **supply** (the road network) and **demand** (goods or travelers). Planners and engineers working for cities, public institutions, or consulting firms and transportation operators can adapt the supply to the transportation de-

mand as theorized and recommended by Adam Smith in 1776 in *The Wealth of Nations* [154]. 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* [107]. In 1930, the adaptation of the demand to the available transportation supply was reported by the City of St. Louis in [49]. 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 [167]. 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* [111] and *Capital, A Critique of Political Economy* in 1867 and 1885 [110].

Influenced by Alfred Marshall – who defined mathematically the concept of supply and demand in 1890 in *Principle of Economics* [109] – Arthur Pigou introduced the notion of transportation supply and transportation demand in 1920 in *The Economics of Welfare* [137].

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 the geographic area – 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 [5, 175]: arterial roads – including Interstate, other freeways and expressways, other principal arterials, and minor arterials –, collectors – including major collectors and minor collectors – and local roads. For each type of road, two notions of capacity are defined:

	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 type and their expected annual average daily traffic (AADT) and percentage of vehicle miles traveled (VMT) on each type of roads in the U.S. [175]. Expected AADT and VMT percentage depends on whether the road is in a rural or urban network. Rural network means that less than 75% of the network population is located in urban centers.

1. The **expected annual average daily traffic (AADT)** [5, 6] 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.

2. The **critical density** [120] is the density of vehicle 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 [153] or derived based on the road geometry and other features [120].

Paradoxically, due to latent demand, creating new infrastructures might not decrease congestion in the network [49]. As an illustration, in Paris, the average daily time in transportation is about 1h30min per person [80]. 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. Through this approach, the city of Paris does not only manage the supply, but implicitly manage the transportation demand. Cite

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 [167]. When one person travels, they will sequentially make 5 distinct choices that are expressed by the trip chain (fig. 2.1) [178]:

- Destination-choice; first, the traveler chooses their destination.
- Time-of-day choice; then, the traveler chooses when they want to arrive or to depart.
- Mode-choice; then, the traveler chooses in which mode (e.g., public transportation, private car, carpool, bike, walk, etc.) they want to travel.
- Route-choice; after choosing 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 [178, 180]:

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

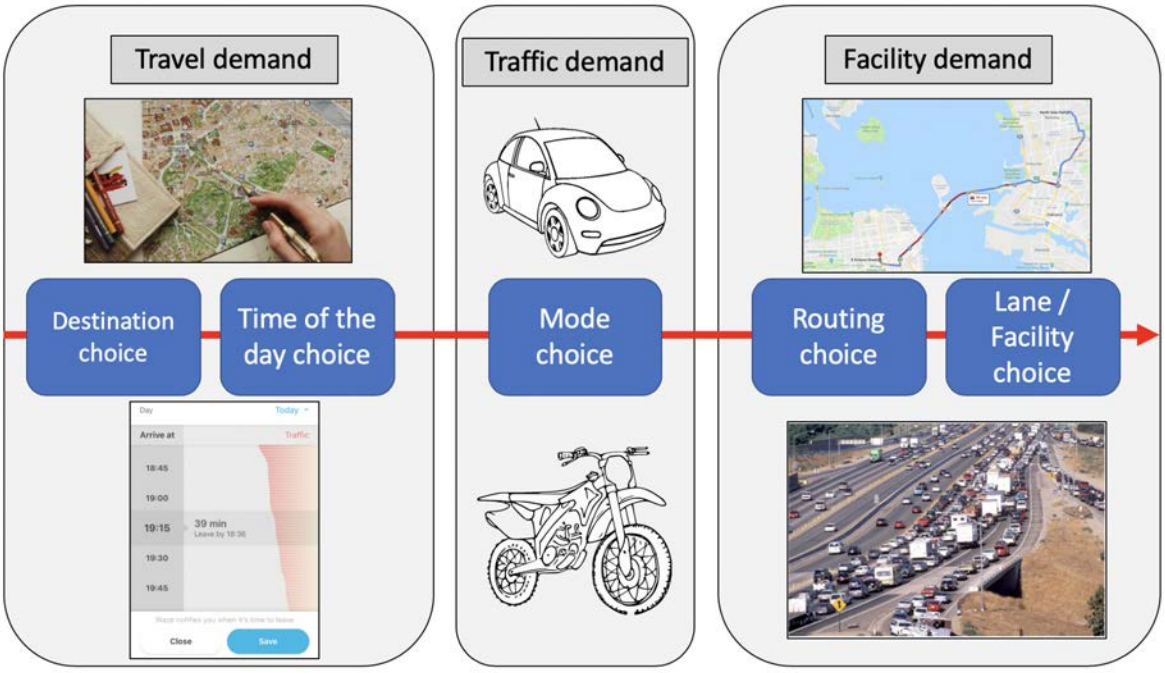


Figure 2.1: The trip chain [178, Figure 4]. First, travelers choose their destination and time of departure. This creates the travel demand of the network. Then, they choose their mode. The ones traveling by car establish to the road traffic demand. Then, they choose their route and their lane, contributing to the facility demand of the route or the lane.

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 [99, 180] and measures of city efficiency. In short-term planning, and in 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 [99, 180, 27] and include:

cite.

- 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 [64], 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, O3, Carbon, NO2, SO2, CO, 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 meet with traffic engineers of six different cities in the U.S. and in France – the measures of city efficiency are less 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 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 they 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:

citation for each metrics

- Road maintenance cost and infrastructure damage.
- Public transit subsidy cost.
- Operational cost.
- Operational revenue.
- Investment.
- Institution debt.

Economic metrics might include:

- City productivity.

- Job accessibility [101].
- Business accessibility [101].
- Travel cost.
- Fuel consumption and fuel cost.
- 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).
- Transportation equity (including underserved community accessibility).

Environmental metrics can include total greenhouse gas emissions due to the transportation system (either instantaneous emissions, either a life-cycle emissions assignment). 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 efficiency. Long-term planning is referred to as transportation planning, while short-term planning and management is 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 station, 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.

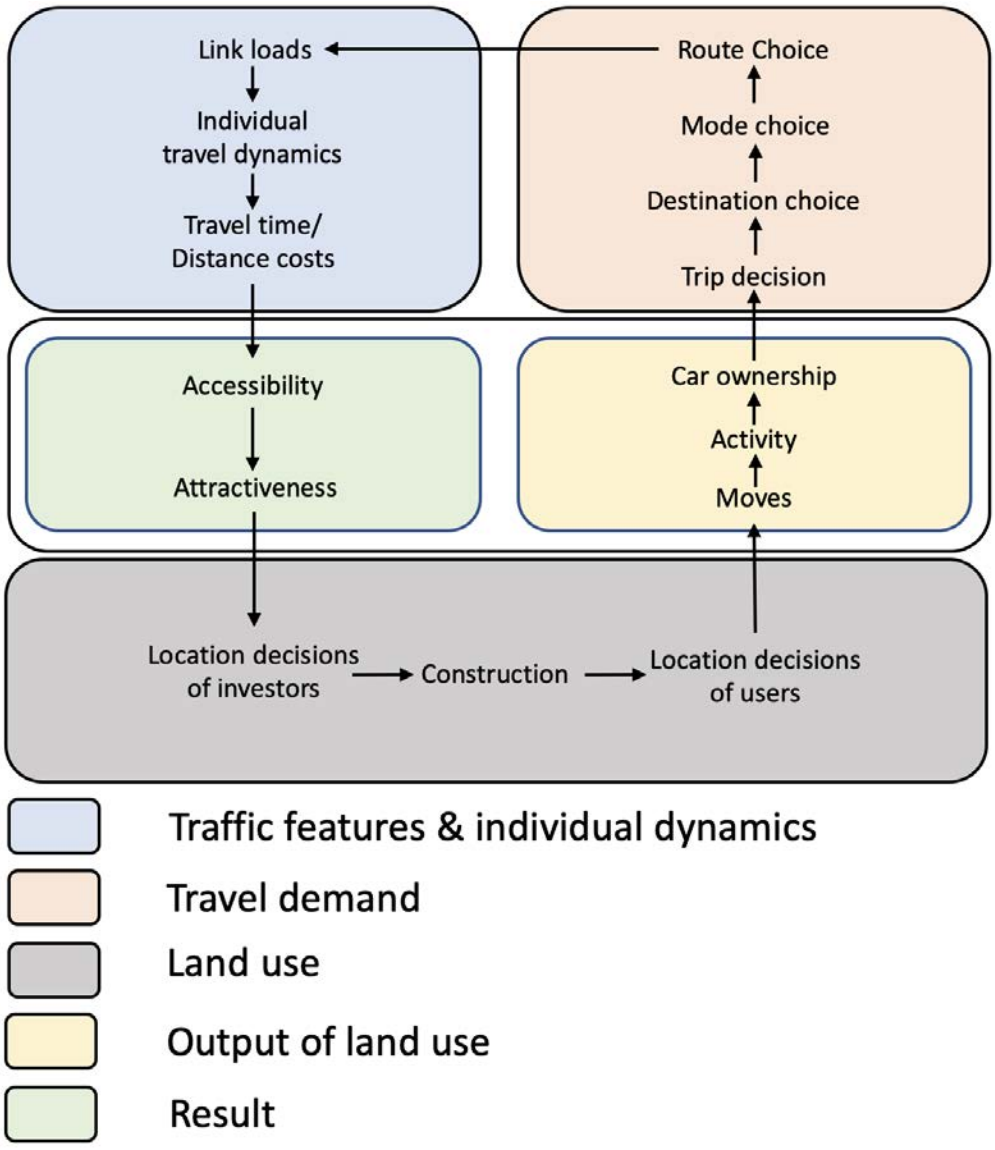


Figure 2.2: The Land Use Cycle [192]: land use determines the need for transport, and transport, in return, further determines spatial development.

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 spatial interactions or trips in the transport system to overcome the distance between the locations of activities. Through the decisions need to be made for trips including whether to make a trip, destination choice, mode choice and route choice, 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 the land accessibility and attractiveness. Depending on the accessibility and attractiveness of each area, investors make decision on choosing appropriate locations for different human activities. Then, constructions – including building transportation infrastructure supplies – take place in the area. After constructions, users could make decisions on where to live, work, entertain or get education. The distribution of land uses, such as residential, industrial or commercial, over the urban area would determine the locations of human activities such as living, working, entertaining and education. Such variation in locations further influence people's moves and mode of transportation, such as car ownership.

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 [183].

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 [173]. 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 can be divided into four branches (see fig. 2.3):

1. Traffic monitoring
2. Transportation performance measures
3. Traffic management strategies
4. Implementation and trade-off

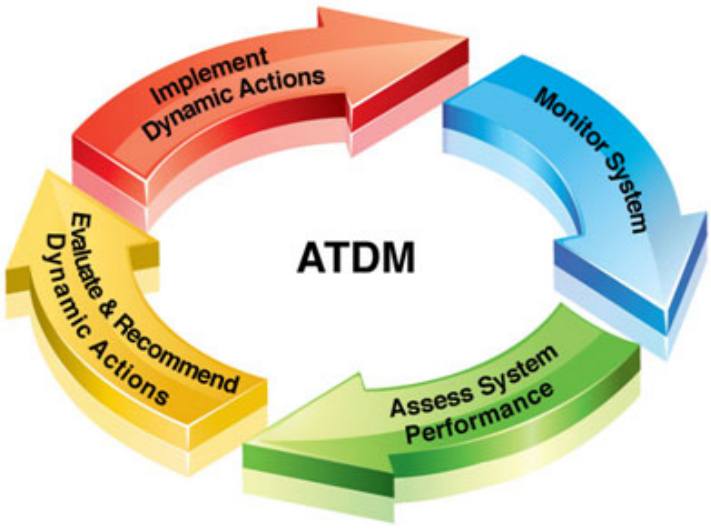


Figure 2.3: Active traffic and demand management (ATDM) chart from the Federal Highway Administration [180]. Traffic management can be divided into four combined branches that enable the optimization of the traffic system [178]. First, the traffic state is estimated through traffic monitoring. Then, 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. Finally, 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 [89] (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:

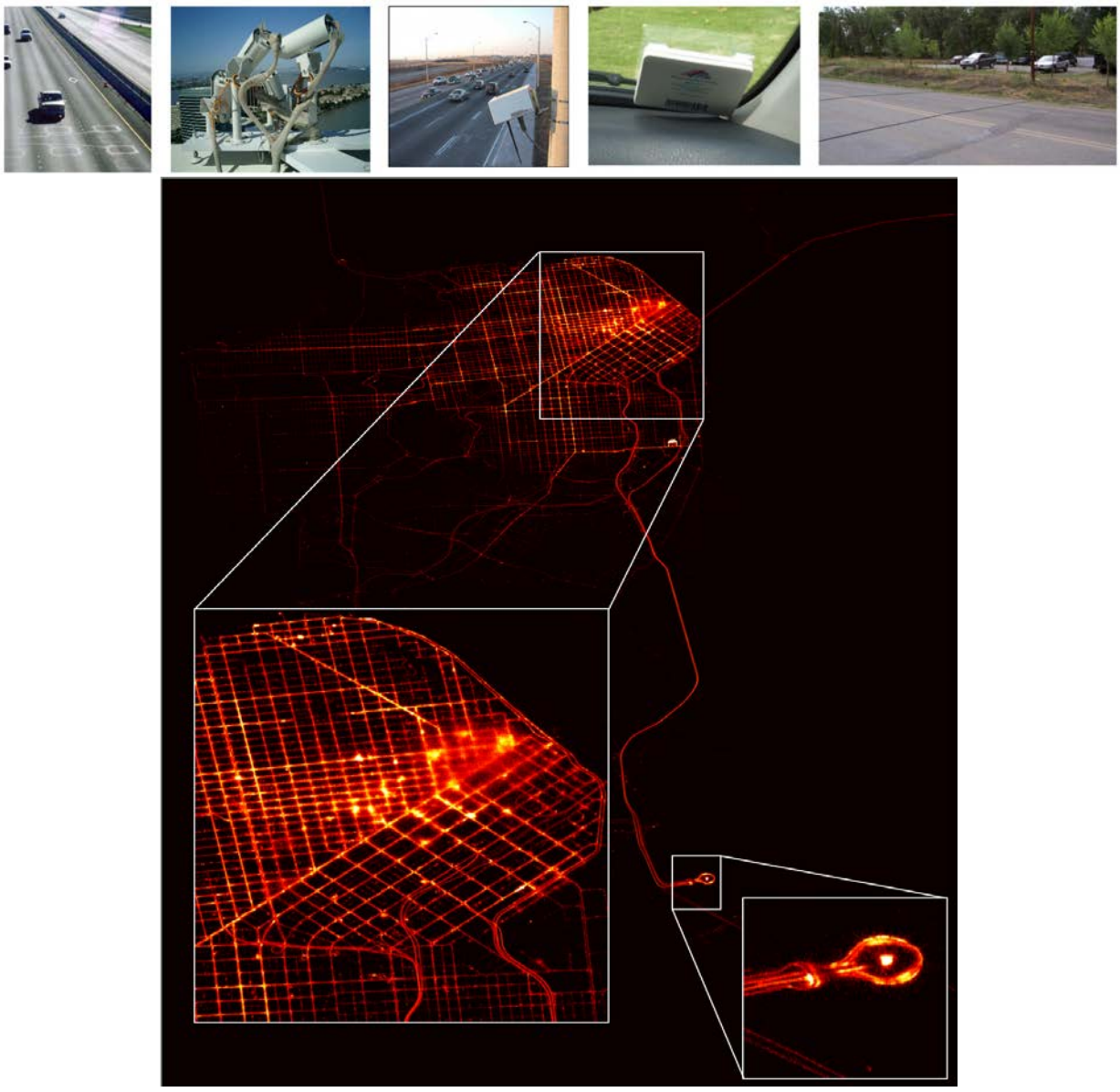


Figure 2.4: Sensing infrastructures and devices. **On the top:** Classical sensing infrastructures [89] used in the pre-smartphone era (and still used as the backbone of traffic control by cities without access to mobile data in sufficient quantities. **(A)**: Loop detectors;**(B):** CCTV cameras;**(C):** radar;**(D)**: tolling RFID transponders used for traffic monitoring;**(E)**: traffic counting tubes. **On 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 [132].

- 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, or access eligibility (like high occupancy vehicle (HOV) and high-occupancy toll (HOT) lanes), speed limits. To promote safe and livable streets, it might include complete street design (like in Los Angeles, CA [47]). More dynamic infrastructure changes might include dynamic lane assignment, dynamic lane restriction, temporal shoulder use, and managed lanes.

- Traffic control strategies [180]; 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 [180]; 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 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.

The transportation planning tools can be divided between building or updating infrastructures, and giving incentives to travelers [162]. 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 [162]. 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.

Predict 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) has on cut-through traffic [170], the traffic engineers can attempt to extrapolate from what the city of Leonia did for Ford Lee Road [84] or what the city of Pleasanton did for Dublin Canyon Road [82]. 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 [170], keeping only those policies that have the most desirable impact on the public. Third, a city can develop a digital twin of its road traffic network with which to try out and learn policies in a virtual setting [112]. 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.

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 impact on the trip chain [178]. An effort of listing all the possible tools can be found in [162]. HOT = high-occupancy toll lane, HOV = high-occupancy vehicle lane

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

Proofread this section.

This dissertation aims to give clues on understanding the impact of information-aware routing (more specifically fastest-path routing) on transportation system, in order to enable studying potential use cases of traffic management through navigational apps.

Over the 2010s decade, the emergence of navigational apps has enables new traffic management strategies, that includes:

- Fastest-path routing [193].
- Eco-routing [11, 72]
- Mass evacuation routing guidance [122, 138]

Eventually all vehicles in the road network will be connected, traffic managers will be able to considerably improve the efficiency of the transportation system through intelligent and adaptive routing suggestions.

The evolution of the traffic information systems until 2020

Until World War II, traffic information related to routing had mostly been map information: the only tool ever used by motorists had been paper maps. In the 1940s the Highway Advisory Radio was among the earliest services to provide traffic information to its users, accessed from radio receivers at home, work or public places. In the 1950s Variable Message Signs appeared, also in the New York area, conveying some local information about traffic conditions (see fig. 2.6). The late 1960s saw the birth of route guidance systems, in 1969, in the Washington, DC area, in which mostly static routing information was provided to users, soon followed by Japan and Germany (see fig. 2.7). It was not until the later age of consumer electronics that traffic information started to be provided to users through a variety of devices, starting by radios onboard cars, later followed by connected devices. However, the nature and quantity of information shared with motorists remained modest for decades, mostly because of the lack of data. Indeed, for decades, only few highways were instrumented by sensors, the geographic extent of traffic information coverage focused on large metropolitan areas, and live traffic information was hard to obtain for networks broadcasting it. In fact, until almost the 2010s, numerous channels gathered their live information by helicopter or CCTV cameras. Over time and with the improvement of communication, this information progressively became live information on both radio and TV, a practice abandoned today, except in special circumstances such as holidays (Thanksgiving in particular), or natural disasters and evacuation. In some areas, journalists are still famous for being multi-decades standing of motorists on the radio in the morning, for example the legendary Joe McConnel in the Bay Area (on NPR/KQED).

Over the 2010 decade, thanks to the internet and crowdsourced GPS data, the information provided to drivers has seen a boom. First, the connectivity between people has 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 have been counted [156]. In the U.S. 95% of the population uses the Internet monthly [155, slide 33]. Second, this connectivity has led to the production of an extensive quantity of traffic data, often referred to as big data. In fall 2018, it has been estimated that every smartphone uses on average 7.0 GB of data every month [155, slide 178]. This number is 10 times higher than in 2012 [155, 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. Eventually, the decade has seen the emergence of new technologies like demand-responsive public transportation, ride-hailing services [51], and GPS-enable navigational apps (fig. 2.7), that enables traffic managers and operators to dynamically impact mode-choice and route-choice. In April 2018, Statista estimated that Google Maps receive 154.4 million monthly users in the U.S. [158].

cite + date

cite + date

Add references and statis-

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.

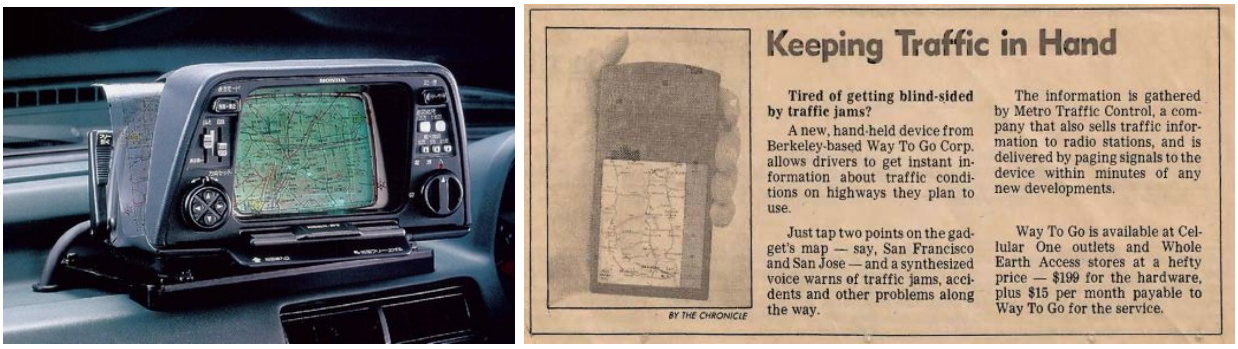


Figure 2.7: **Left:** Onboard navigation unit prototype built by Honda in 1981 for route guidance. **Right:** Newspaper commercial for *Way to Go* aftermarket device circa 1990, one of the early prototypes of navigational apps built by UC Berkeley.

Classical route-choice models

Route-choice and direction signs

In order to increase the transportation system efficiency through economies of scale [154], the road are designed such that the traffic is aggregated to high-category roads where the speed and the capacity are designed to be high [5]. 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 road signs, that were designed to control the route-choice. Cite

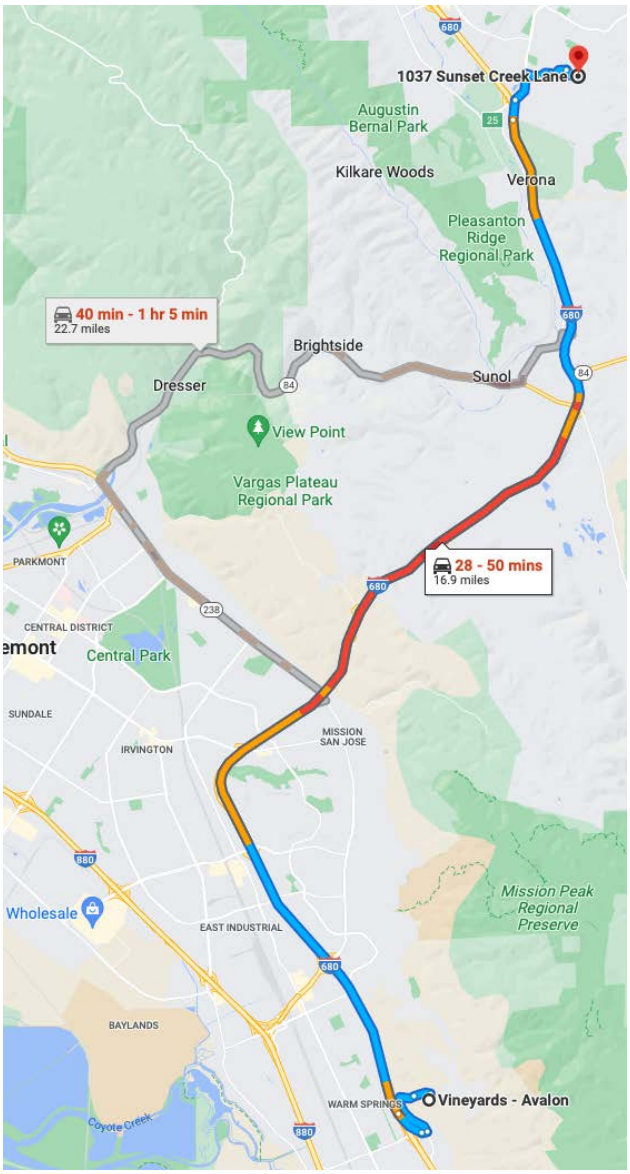


Figure 2.8: July 14th, 2022 4:30pm, traffic guidance on Google Maps within the Fremont, CA area. The request is done on July 13th and uses Google Maps traffic prediction system. The route suggested uses a local road, then a minor arterial road, then a collector road and then an interstate before using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

Wardrop equilibrium

Due to the double definition of the capacity (daily urban planning capacity and hourly road traffic capacity section 2.1), two classical models of route-choice can be found in the early transportation literature (since 1950s):

1. The all-or-nothing route assignment model [133], where vehicles follow the fastest path between their origin and destination under free-flow conditions. This model represents that vehicles follow the road signs, and that the road signs provide the fastest path between any origin to any destination if there is no congestion on the road.

2. The static traffic assignment (STA) model [133], where vehicles choose their route such that they minimize their experienced travel time (potentially accounting for congestion created by other vehicles).

In the second model, the vehicle are assumed to adapt their route based on congestion. This assumption has been made back in 1920 by Arthur Pigou [137]. Answering to Pigou's notion of social cost, Frank Knight first described the notion of traffic equilibrium (equilibrium resulting from no incentive of vehicles to change their route choice from one day to another one) in 1924 in *Some fallacies in the interpretation of social cost* [90]. In 1950, John Nash developed conditions to prove the existence of non-cooperative game equilibrium, now called Nash equilibrium [119]. In 1952, the concept of Nash equilibrium was applied to the road transportation network by John Wardrop [190]. 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 the Wardrop equilibrium model, daily urban planning road capacity are not taking into account. Therefore, some trips might include a decrease and then an increase in the categories of the roads used by the trip (see fig. 2.8).

Dissertation outline

Using a combination of the two classical routing models introduced in the previous subsection, chapter 3 models the impact of an increase in app usage by dividing vehicles between app users (following the STA) and non-app users (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 of app usage. The resulting change in the state of traffic leads to some negative externalities on local roads where the induced facility demand exceed the urban planning capacity, as show by chapter 4.

In both classical routing models, the vehicles do not dynamically change their route: either they follow the direction signs or they follow the same route as the previous day (such that the traffic is in an equilibrium state). To model dynamic rerouting models, traffic

microsimulation can be used. However, the creation, calibration, and validation of a traffic microsimulation is difficult. Therefore, chapter 5 provides a blueprint for such task. Finally, chapter 6 introduces a game-theory dynamic-routing model to decrease the computation runtime of 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 behavior might be managed to improve transportation efficiency given a traffic demand, assumed to be fixed. As a consequence, this thesis does not take into account mode shift (Pigou–Knight–Downs paradox [9]), departure time shift (rescheduling behavior [60]), or willingness to travel shift ("reduced demand" [178]) due to change in traffic implied by routing behavior changes.

Chapter 3

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

Proofread. Remove we and article. Move some notations to the introduction. Unify notations. Search for [â-zA-Z]d[â-z].

Navigational apps allegedly provide vehicles with the fastest path to reach their destinations [193]. 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 [190]. This translates to an equalization of the travel times of every path used between any specific origin and specific destination fig. 3.1).

To model route-choice, transportation planners use the static traffic assignment (STA) [133, 113]. 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 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 app usage 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 [37, 35, 36].

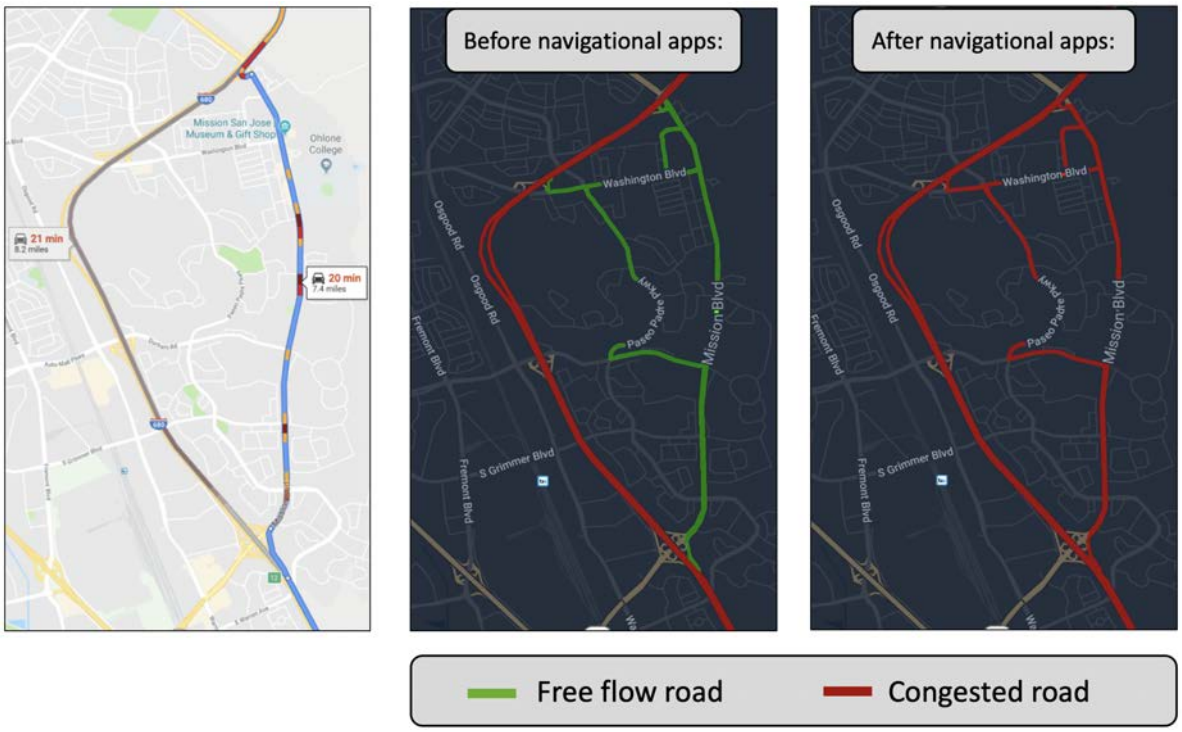


Figure 3.1: Illustrations of the impact of navigational apps on the Fremont, CA network. **On the left:** travel time equalization of the highway commute and of the short-cut commute (screenshot of Google Maps proposed directions to go North at 4pm on Monday July, 8^{th} 2019). **In the middle:** before 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 navigational app usage (or information-aware routing) on the road traffic. First, the section describes the classical static traffic assignment (STA) model that is also referred to as the static non-atomic routing game. Then, the section introduces the restricted path choice model that can be 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 [133] is traditionally defined on a network $\mathcal{G}$. For each path p of the network, the path flow h_p of the path p is the flow of vehicles using this path p. Assuming that the path flow allocation is static, the link flow on an link l – the flow of vehicles using the link l of $\mathcal{G}$ – is the sum of the path flows of all paths using the link l. Then, for a given flow demand vector $\boldsymbol{d}$ – which assigns for each origin o and destination d in the network, a flow demand d_{od} – 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. We assume that the travel time t_p of each path p is the sum of the travel times t_l 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, etc.). A flow allocation is called a *user equilibrium* if all paths used between an origin o and a destination d have the same travel time for every o, d pair (Wardrop's first condition [190]). We specify this in the following:

Static Traffic Assignment framework [133]

This subsection first formalizes the notion of network and paths using graphs.

Cite graph theory textbook

Definition 3.1.1 (Network, paths and demand)**.** *Given a finite strongly connected directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of vertices, and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ the set of links (links), we call, for each origin $o \in \mathcal{V}$ and destination $d \in \mathcal{V}$, $\mathcal{P}_{od}$ the set of feasible paths without cycles from o to d.*

Traffic micro-models model each individual vehicle and their dynamics. 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 3.1.1 (Micro to macro: mean-field theory and non-atomic 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. Aggregating a high-dimension of individual stochastic dynamics into deterministic macroscopic dynamics is known as mean-field theory in physics. Connecting both the game theory and the statistical physics community, non-atomic dynamic games with stochastic noise are referred to as mean-field games. By extension, in this dissertation, we coined any non-atomic dynamic game with the name mean-field game. In chapter 6, we will define a dynamic routing game that we will solve with a mean-field game approximation in order to improve the realism of the routing game models.*

Traffic macro-model model vehicles using traffic flow. More specifically, vehicles are aggregated into path flows and link flows.

Definition 3.1.2 (Path flows). *For each path $p \in \mathcal{P} = \bigcup_{o,d \in \mathcal{V}} \mathcal{P}_{od}$, we define:*

- *The flow h_p using the path p (path flows). We note the path flow vector $\mathbf{h} = (h_p)_{p \in \mathcal{P}}$.*
- *The indicator $\delta_p = (\delta_p(l))_{l \in \mathcal{E}}$, where $\delta_p(l) = \begin{cases} 1 \text{ if } l \in p \\ 0 \text{ else} \end{cases}$*
  This vector is called indicator of links included in the path p. We denote the incidence matrix $\Delta = (\delta_p)_{p \in \mathcal{P}}$

Definition 3.1.3 (Link flows). *For each link $l \in \mathcal{E}$, we define f_l as the flow using link l (link flow). We note the link flow vector $\mathbf{f} = (f_l)_{l \in \mathcal{E}}$.*

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.

Remark 3.1.2 (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 3.1.2).*

Remark 3.1.3 (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.*

Definition 3.1.4 (Static traffic demand). *We define the (static) traffic demand as $\boldsymbol{d} \in \mathbb{R}_+^{|\mathcal{V}| \times |\mathcal{V}|}$. For any origin $o \in \mathcal{V}$ and any destination $d \in \mathcal{V}$, d_{od} denotes the number of vehicles traveling for the origin to the destination.*

Definition 3.1.5 (Feasible assignment). *Given a demand $\boldsymbol{d} \in \mathbb{R}_+^{|\mathcal{V}| \times |\mathcal{V}|}$, 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.*

Figure 3.2 illustrates this framework on a benchmark network.

Traffic engineers have shown that, in average over an hour, travel time can be related to flow on highway road sections [133, 120, 153]. 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.

Definition 3.1.6 (Travel time function). *For each link $l \in \mathcal{E}$ we define the link (link) travel time function which gives the link (link) travel time given the link (link) flow: $t_l : f_l \to t_l(f_l)$. We assume t_l to be continuous. We then define the travel time vector $\mathbf{t}(\mathbf{f}) = (t_l(f_l))_{l \in \mathcal{E}}$.*

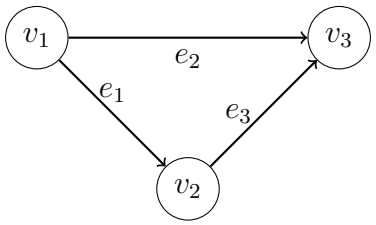


Figure 3.2: Network to illustrate the framework. We have $\mathcal{V} = \{v_1, v_2, v_3\}$. $\mathcal{E} = \{e_1, e_2, e_3\}$ where $e_1 = (v_1, v_2)$, $e_2 = (v_1, v_3)$ and $e_3 = (v_2, v_3)$. $\mathcal{P}_{v_1,v_3} = \{(e_1, e_3), (e_2)\}$. If we name $p = (e_1, e_3) \in \mathcal{P}_{v_1,v_3}$, then $\delta_p = (1, 0, 1)$. $\mathcal{P} = \{(e_1), (e_2), (e_3), (e_1, e_3)\}$. If we have 100 travelers with $(o, d) = (v_1, v_3)$, then we would have $\mathcal{F}_{\boldsymbol{d}} = \{(x, 100 - x, x),\ x \in [0, 100]\}$.

For each path $p \in \mathcal{P}$, we define the travel time function of the path p as: $t_p : \mathbf{f} \to t_p(\mathbf{f}) = \sum_{l \in \mathcal{E}} t_l(f_l) \cdot \delta_p(l)$.

We sometimes refer to the travel time of a path as the cost of the path.

Remark 3.1.4 (Travel time function)**.** *During the remainder of this dissertation, travel time functions are assumed to be increasing as functions of the corresponding 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, gas emission [87]), and it will not change the following results.*

As stated in section 2.3, transportation planners assume that vehicles choose their path such that they minimize their travel time [190]. By interpreting each vehicle in the network as a player, each route-choice of a vehicle as a player's action, and the travel time of a route as the cost of the action, then the static traffic assignment can be understood as a game: the non-atomic routing game.

Definition 3.1.7 (Non-atomic routing game (chapter 18 of [125]))**.** *The set $(\mathcal{G}, \boldsymbol{d}, \mathbf{t})$ defines the non-atomic routing game [125, 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). However, the cost a player experienced might depend on the other players' actions. Therefore, it is hard for a player to know which action to play without knowing the actions 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. It is modeled that the vehicle will take the route that is predicted to be the fastest one. The mapping between the state of the game (or the knowledge of the state of the game) and the action of a player is called the player's strategy. After the facts, a vehicle might realize that the path it took was not the fastest one to reach its destination. In this case, the vehicle will have counterfactual regrets: it should have chosen another path.

Remark 3.1.5 (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 [116, 149]. Some notions defined in a classic non-cooperative game with finite number of players [17] (like deviation incentives or regret, ϵ-Nash equilibrium, or average-ϵ-Nash equilibrium) can be extended to non-atomic games using measure-theoretical tools to deal with the continuum of players.*

Definition 3.1.8 (Counterfactual regret)**.** *We define the counterfactual regret of a vehicle as the difference between their travel times and the optimal travel times between their origins to their destinations. Because in the non-atomic routing game, the number of vehicles is uncountable, we define the regret for the type of vehicles on a path p (called then regret of path p) as:*

Cite.

$$r_p(\mathbf{f}) = t_p(\mathbf{f}) - \min_{\tilde{p} \in \mathcal{P}_{od}} t_{\tilde{p}}(\mathbf{f})$$

Remark 3.1.6 (Counterfactual regret, marginal regret, deviation incentives)**.** *In the context of repeated games, counterfactual regret is sometimes referred to as marginal regret [41]. To distinguish the concept of cumulative regret and marginal regret, the counterfactual regret (or marginal regret) is sometimes also referred to as deviation incentive: if the game would happen again, then the vehicle that did not take the fastest path to reach its destination will have some incentive to deviate from its previous strategy. In the remaining of the dissertation, we will use the term* ***deviation incentive****.*

Cite.

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

If no player has any incentive to change its action, then the game reaches an equilibrium: for every instance of the game, each player will play the same action as in the previous game instance. This equilibrium state is referred to as a Nash equilibrium. When dealing with vehicles that choose their route in other to minimize their travel time, the notion of Nash equilibrium is sometimes called Wardrop equilibrium [190] or user equilibrium [133].

Definition 3.1.9 (User equilibrium [133])**.** *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:*

$$\forall o, d \in \mathcal{V},\ \forall p \in \mathcal{P}_{od},\ h_p \cdot \left(t_p(\mathbf{f}) - \min_{\tilde{p} \in \mathcal{P}_{od}} t_{\tilde{p}}(\mathbf{f}) \right) = 0 \tag{3.1}$$

Theorem 3.1.1 (Routing game Nash equilibrium is the STA user equilibrium [125])**.** *The user equilibrium is the Nash equilibrium of the non-atomic routing game $(\mathcal{G}, \boldsymbol{d}, \mathbf{t})$ [144, 116, 149, 125].*

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-user equilibrium is such that no one has a deviation incentive higher than ϵ.

Definition 3.1.10 (ϵ-user equilibrium)**.** *For the traffic demand $\boldsymbol{d}$, 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:*

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

Remark 3.1.7 (ϵ-user equilibrium)**.** *A strategy profile of a non-cooperative game is an ϵ-Nash equilibrium if no one can increase their payoff by ϵ by changing their strategy unilaterally [17]. In an ϵ-Nash equilibrium, no one has a deviation incentive above ϵ. This notion is an extension from the discrete game framework to non-atomic routing game framework using measure theoretical tools.*

As an ϵ-user equilibrium considers the worst case for one path, we introduce the notion of average ϵ-user equilibrium that is better suited for non-atomic games (i.e., continuum of players):

Definition 3.1.11 ([26] Average ϵ-user equilibrium)**.** *Let $\epsilon \in \mathbb{R}_{>0}$. For the traffic demand $\boldsymbol{d}$, a flow allocation $\mathbf{f} = \Delta\mathbf{h} \in \mathcal{F}_{\boldsymbol{d}}$ is an average ϵ-user equilibrium if and only if:*

$$\frac{1}{\|\boldsymbol{d}\|_1} \sum_{o,d \in \mathcal{V}} \sum_{p \in \mathcal{P}_{od}} h_p \cdot \Big(t_p(\mathbf{f}) - \min_{\tilde{p} \in \mathcal{P}_{od}} t_{\tilde{p}}(\mathbf{f}) \Big) < \epsilon \tag{3.3}$$

Remark 3.1.8 (Average ϵ-user equilibrium)**.** *In an average ϵ-user equilibrium, an average "player" (infinitesimal fraction of traffic) can expect to save ϵ by changing unilaterally their strategy (path). As highlight in [26], this notion is similar to the definition of (ϵ, δ)-Nash equilibria in [63].*

Counterintuitively, the user equilibrium route assignment might not minimize the average cost in the network [145]. The route assignment such that the average cost in the network is minimized is called the **social optimum** (or system-optimum) [190].

Definition 3.1.12 (Social optimum)**.** *For the traffic demand $\boldsymbol{d}$, a flow allocation $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$ is a social optimum if and only if:*

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

The fact that a user equilibrium is not necessarily a social optimum is known as the price of anarchy [145].

Variational inequality and minimization problem formulation

For a non-atomic routing game, Nash equilibria are known to be easy to compute as they can be expressed as the solution to a convex optimization problem, using a convex potential function (Rosenthal function) [144].

Definition 3.1.13 (Variational inequality [133]). *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 3.1.14 (Minimization problem [144]). *We define the following optimization problem with the classical Rosenthal potential:*

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

Property 3.1.1 (Interpretation and equivalences [144]). *If for all links* $l \in \mathcal{E}$*, 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*t_l(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 3.1.14), 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 separate vehicles into two populations: those who use navigational applications to route themselves (app users) and those who do not (non-app users). In the restricted path choice model, non-app users are following the road 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 3.1.1). 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 3.1.2). 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 3.1.14 and [133]. In this case, every vehicle possesses perfect information, the user equilibrium is reached.

Proofread everything below in this chapter. Potentially use [37, Property 4.2].

Definition 3.1.15 (User equilibrium of the restricted path choice model)**.** *The Wardrop's first condition for the restricted path choice model can be express as:*

$$h_p^a \cdot (t_p(\Delta\mathbf{h}) - \pi_{od}^a) = 0 \qquad \forall o, d \in \mathcal{V},\ \forall p \in \mathcal{P}_{od} \tag{3.7}$$

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

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

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

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

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

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

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

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 ([30]). Following remark 3.1.2, definition 3.1.2 and definition 3.1.5, we denote it $\mathbf{f}_\alpha^\star = \Delta(\mathbf{h}^{a,\star} + \mathbf{h}^{na,\star})$.

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 (or average marginal regret, or average counterfactual regret, or exploitability, or relative gap to user equilibrium) best meets the target.

Cite.

Cite.

Evaluating the distance between a traffic state and a user equilibrium

An observed traffic state – defined as a feasible flow allocation $\mathbf{f}$ (definition 3.1.5) – could be neither a user equilibrium nor a social optimum. 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 later to experimentally assess the impact of routing apps on traffic. It can also be used to understand under which conditions the use of apps will have an impact on a given traffic assignment (given that app usage decreases deviation incentives of vehicles). To achieve this goal, we define a function $\bar{\mathcal{R}}$ 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 $\mathcal{G}$, a traffic demand $\boldsymbol{d}$, and a travel time vector function $\mathbf{t}$, $\bar{\mathcal{R}}$ should satisfy the following properties:

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

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

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

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

v. $\bar{\mathcal{R}}$ 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 3.1.9 (Distance)**.** *As we want $\bar{\mathcal{R}}$ to be only a function of the observed state of traffic $\mathbf{f}$, it cannot be defined as a mathematical distance. We presents later why defining $\bar{\mathcal{R}}(\mathbf{f}) = \min_{\mathbf{f}^{ue} \in S_{ue}} d(\mathbf{f}, \mathbf{f}^{ue})$ – where S_{ue} is the set of user equilibrium flow allocation and d is a metric like $d(x, y) = \|x - y\|_2$ – would not satisfied properties we want for $\bar{\mathcal{R}}$.*

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

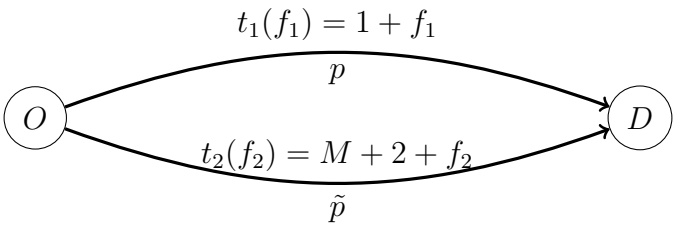


Figure 3.3: Benchmark network example to illustrate the average deviation incentive

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 $\bar{\mathcal{R}}$ than the other candidates of functions.

The case considered. Let us consider the network in Figure 3.3. The network consists of two nodes (O and D) and two paths (p and $\tilde{p}$). The cost of each path depends on the flow on this path (f_1 and f_2). The functions are given in Figure 3.3: $t_1(f_1) = 1 + f_1$ and $t_2(f_2) = M + 2 + f_2$ with $M \geq 0$ given. We consider a demand of $d_{OD} = 1$ between O and D. The user equilibrium flow can be directly determined from the Wardrop first principle: $f_1^{\text{ue}} = 1$ and $f_2^{\text{ue}} = 0$. We note $\mathbf{f}^{\text{ue}}$ the vector $\mathbf{f}^{\text{ue}} = (f_1^{\text{ue}}, f_2^{\text{ue}})$. Now, imagine that we observe a traffic flow $\hat{\mathbf{f}}$ where $\hat{f}_1 = 1 - \alpha$ and $f_2^{\text{ue}} = \alpha$ 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 $\hat{\mathbf{f}}$ is from a Nash equilibrium (here $\mathbf{f}^{\text{ue}}$).

Inadequacy of a flow-based function. A first intuition approach might consist of comparing the link flows between the observed state $\hat{\mathbf{f}}$ and the Nash equilibrium flow $\mathbf{f}^{\text{ue}}$: $\bar{\mathcal{R}}(\hat{\mathbf{f}}) = \|\mathbf{f}^{\text{ue}} - \hat{\mathbf{f}}\|_2$. In the case considered, we have that $\bar{\mathcal{R}}(\hat{\mathbf{f}}) = \sqrt{2}\alpha$. The function $\bar{\mathcal{R}}(\hat{\mathbf{f}})$ satisfies properties (i), (ii) and (iv) for the case we consider. But the function $\bar{\mathcal{R}}(\hat{\mathbf{f}})$ does not satisfy the property (iii). Indeed, $\bar{\mathcal{R}}(\hat{\mathbf{f}})$ 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 $\bar{\mathcal{R}}(\hat{\mathbf{f}}) = \|\mathbf{t}(\mathbf{f}^{\text{ue}}) - \mathbf{t}(\hat{\mathbf{f}})\|_2$. In the case considered, the travel time function is just a translation of the flow vector, so that we still have $\bar{\mathcal{R}}(\hat{\mathbf{f}}) = \sqrt{2}\alpha$. 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 $\mathbf{f} \cdot \mathbf{t}(\mathbf{f})$: $\bar{\mathcal{R}}(\hat{\mathbf{f}}) = |\mathbf{f}^{\text{ue}} \cdot \mathbf{t}(\mathbf{f}^{\text{ue}}) - \hat{\mathbf{f}} \cdot \mathbf{t}(\hat{\mathbf{f}})|$ (like in [141]). But, as seen in definition 3.1.12, $\mathbf{f} \cdot \mathbf{t}(\mathbf{f})$ 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, $\bar{\mathcal{R}}(\hat{\mathbf{f}}) = \alpha|M - 1 + 2\alpha|$. If $|M - 1 + 2\alpha| = 0$, $\hat{\mathbf{f}}$ is not a Nash equilibrium but $\bar{\mathcal{R}}(\hat{\mathbf{f}}) = 0$. So $\bar{\mathcal{R}}(\hat{\mathbf{f}})$ does not satisfy the property (ii).

Inadequacy of a price of anarchy-like function. Considering an idea similar to the price of anarchy [145], one can define $\bar{\mathcal{R}}(\hat{\mathbf{f}}) = \frac{\hat{\mathbf{f}} \cdot \mathbf{t}(\hat{\mathbf{f}})}{\mathbf{f}^{\text{ue}} \cdot \mathbf{t}(\mathbf{f}^{\text{ue}})}$ where $\mathbf{f}^{\text{ue}}$ is a user equilibrium. This is similar to the hybrid based function: $\frac{\hat{\mathbf{f}} \cdot \mathbf{t}(\hat{\mathbf{f}})}{\mathbf{f}^{\text{ue}} \cdot \mathbf{t}(\mathbf{f}^{\text{ue}})} = 1 + \frac{1}{\mathbf{f}^{\text{ue}} \cdot \mathbf{t}(\mathbf{f}^{\text{ue}})} \cdot (\hat{\mathbf{f}} \cdot \mathbf{t}(\hat{\mathbf{f}}) - \mathbf{f}^{\text{ue}} \cdot \mathbf{t}(\mathbf{f}^{\text{ue}}))$. 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 $\bar{\mathcal{R}}$ using the game theoretical framework. Using deviation incentives, we do not

need to know every user equilibrium flow allocation $\mathbf{f}^{\text{ue}}$ to find out whether a state of traffic is a user equilibrium. A user equilibrium 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 3.1.9): i.e., no one has deviation incentive. One idea is to define $\bar{\mathcal{R}}(\hat{\mathbf{f}})$ as the worst deviation incentive of vehicles in the observed traffic assignment. This is equivalent to define $\bar{\mathcal{R}}(\hat{\mathbf{f}})$ as the smallest ϵ such that $\hat{\mathbf{f}}$ is a ϵ-Nash equilibrium:

$$\bar{\mathcal{R}}(\hat{\mathbf{f}}) = \max_{o,d} \max_{\substack{p \in \mathcal{P}_{od} \\ h_p > 0}} \left(t_p(\hat{\mathbf{f}}) - \min_{\tilde{p} \in \mathcal{P}_{od}} \mathbf{t}_{\tilde{p}}(\hat{\mathbf{f}}) \right)$$

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 $|\mathcal{E}|$.

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

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

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

The average "deviation incentive" of vehicles function. Based on the definition of the average-ϵ-Nash equilibrium, we define $\bar{\mathcal{R}}(\hat{\mathbf{f}})$ as the smallest ϵ such that $\hat{\mathbf{f}}$ is at average-ϵ-Nash equilibrium:

$$\bar{\mathcal{R}}(\hat{\mathbf{f}}) = \frac{1}{\|\boldsymbol{d}\|_1} \cdot \sum_{(o,d) \in \mathcal{V}^2} \sum_{p \in \mathcal{P}_{od}} h_p \cdot (t_p(\hat{\mathbf{f}}) - \pi_{od}(\hat{\mathbf{f}}))$$

where $\pi_{od}(\hat{\mathbf{f}}) = \min_{\tilde{p} \in \mathcal{P}_{od}} t_p(\hat{\mathbf{f}})$. We will use this function as $\bar{\mathcal{R}}$ 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 the average deviation incentive, which quantifies how much time an average vehicle can expect to save by changing their path to the optimal one. Then, some properties of the average deviation incentive are presented. In particular, the average deviation incentive satisfies the properties stated before.

Definition 3.1.16 (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* $\mathbf{y}(\mathbf{f})$ *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* $\mathbf{f}$*.*

Remark 3.1.10 (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 define the average deviation incentive 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.

Definition 3.1.17 (Average deviation incentive)**.** *We define the average deviation incentive of the flow pattern* $\mathbf{f} \in \mathcal{F}_{\boldsymbol{d}}$ *as follows:*

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

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 3.1.16.*

Remark 3.1.11 (Measuring the average deviation incentive)**.** *Because the average deviation incentive is a function of only the link flow, 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* $\bar{\mathcal{R}}$ *satisfies the property (iv).*

After defining the average deviation incentive, we introduce its properties.

Definition 3.1.18 (Shortest travel time)**.** *Any optimal path* $p^*_{od}(\mathbf{f})$ *(see definition 3.1.16) is a shortest path (with respect to cost) between o and d given the cost on each link* $\mathbf{t}(\mathbf{f})$*:* $\mathbf{t}(\mathbf{f})^\top \delta_{p^*_{od}(f_l)} = \min_{p \in \mathcal{P}_{od}} \mathbf{t}(\mathbf{f})$*. For every* $o, d \in \mathcal{V}$*, we define:*

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

Remark 3.1.12 (Interpretation of the average deviation incentive)**.** *Since* $t_p(\mathbf{f}) - \pi_{od}(\mathbf{f})$ *represents the time a vehicle on path* $p \in \mathcal{P}_{od}$ *could save by choosing the best path for their trip,* $\bar{\mathcal{R}}$ *can be interpreted as the average time a vehicle could save by changing unilaterally their path.*

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

$$\mathbf{t}(\mathbf{f})^\top \mathbf{f} = \sum_{(o,d)\in\mathcal{V}^2} \sum_{p\in\mathcal{P}_{od}} h_p \cdot t_p(\mathbf{f})$$

$$\mathbf{t}(\mathbf{f})^\top \mathbf{y}(\mathbf{f}) = \sum_{(o,d)\in\mathcal{V}^2} \mathbf{t}(\mathbf{f})^\top \delta_{p^*_{od}(f_l)} \cdot d_{od} = \sum_{(o,d)\in\mathcal{V}^2} \pi_{od}(\mathbf{f}) \cdot d_{od}$$

$$\bar{\mathcal{R}}(\mathbf{f}) = \frac{1}{\|\boldsymbol{d}\|_1} \cdot \sum_{(o,d)\in\mathcal{V}^2} \left(\left(\sum_{p\in\mathcal{P}_{od}} h_p \cdot t_p(\mathbf{f}) \right) - d_{od} \cdot \pi_{od}(\mathbf{f}) \right)$$

$$\sum_{p\in\mathcal{P}_{od}} h_p = d_{od} \implies \quad \bar{\mathcal{R}}(\mathbf{f}) = \frac{1}{\|\boldsymbol{d}\|_1} \cdot \sum_{(o,d)\in\mathcal{V}^2} \sum_{p\in\mathcal{P}_{od}} h_p \cdot (t_p(\mathbf{f}) - \pi_{od}(\mathbf{f}))$$

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

Property 3.1.2 (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:

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

Equation (3.16) implies that $\forall \mathbf{f} \in \mathcal{F}_{\boldsymbol{d}},\ \bar{\mathcal{R}}(\mathbf{f}) \geq 0$*:* $\bar{\mathcal{R}} : \mathcal{F}_{\boldsymbol{d}} \mapsto \mathbb{R}_+$*, 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 3.1.13) – provides that $\bar{\mathcal{R}}(\mathbf{f}) = 0 \iff \mathbf{f}$ *is a user equilibrium (property (ii)).*

Remark 3.1.13. *The variational inequality tells us that none of the players can have a better outcome by choosing a different path in isolation. The average deviation incentive identifies which travel time any player could expect to save by rerouting.*

Property 3.1.3 (The average deviation incentive as a measure of vehicle efficiency)**.** *Given* $\epsilon \in \mathbb{R}_{>0}$*, from remark 3.1.12, it is straightforward that* $\forall \mathbf{f} \in \mathcal{F}_{\boldsymbol{d}},\ \bar{\mathcal{R}}(\mathbf{f}) \leq \epsilon \iff \mathbf{f}$ *is an average-*ϵ*-Nash equilibrium (definition 3.1.11). 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 best route between their origin and destination as their path. Then $\bar{\mathcal{R}}$ *can be interpreted as a measure of the efficiency of the vehicles.*

The closer $\bar{\mathcal{R}}$ *is to 0, the less inclined players are to change their paths. If* $\bar{\mathcal{R}} = 0$*, we are at a user equilibrium.*

Property 3.1.4 (Continuity)**.** *The average deviation incentive* $\bar{\mathcal{R}}(\mathbf{f})$ *is continuous with respect to* $\mathbf{f}$*. Property (v) is satisfied.*

Proof. We have $\bar{\mathcal{R}}(\mathbf{f}) = \frac{1}{\|\boldsymbol{d}\|_1}\mathbf{t}(\mathbf{f})^\top(\mathbf{f}-\mathbf{y}(\mathbf{f})) = \frac{1}{\|\boldsymbol{d}\|_1}\left(\mathbf{t}(\mathbf{f})^\top\mathbf{f} - \min_{\tilde{\mathbf{f}}\in\mathcal{F}_{\boldsymbol{d}}}\mathbf{t}(\mathbf{f})^\top\tilde{\mathbf{f}}\right)$. Because $\mathbf{t}(\mathbf{f})$ is continuous with respect to $\mathbf{f}$ (definition 3.1.6), 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 3.1.3 is a polytope). The optimal objective value of an LP is continuous with respect to perturbation on the objective function [30]. □

Remark 3.1.14 (Computational time)**.** *An optimal path p^*_{od} can be found in $\mathcal{O}(|\mathcal{E}|\cdot\log(|\mathcal{V}|))$ with Dijkstra's algorithm. $\bar{\mathcal{R}}$ can be found in $\mathcal{O}(|\mathcal{E}|\cdot|\mathcal{V}|\log(|\mathcal{V}|))$ with a sequential application of $|\mathcal{V}|$ Dijkstra's algorithms. Therefore, the average deviation incentive satisfies property (iv).*

The average deviation incentive $\bar{\mathcal{R}}$ can be computed with "local data" only, i.e. link cost, link flow and demand. Definition 3.1.17 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 3.1.15 (Frank Wolfe's Algorithm [67, 26])**.** *The user equilibrium can be seen as a routing game [125, Chapter 18]. It has been shown that a no-regret learning algorithm in selfish routing converges to a user equilibrium of the system [26, 92, 125, Chapter 4].*

For solving the minimization problem (3.6)*, we can use Frank Wolfe's algorithm [67], a projected gradient descent algorithm. This algorithm minimizes $\bar{\mathcal{R}}$ over each iteration of the algorithm. It is equivalent to a no-regret learning algorithm [26]:*

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

Data: $(\mathcal{V},\mathcal{E})$, $\epsilon\in\mathbb{R}_{>0}$, $\boldsymbol{d}\in\mathbb{R}_+^{|\mathcal{V}|\times|\mathcal{V}|}$, $\mathbf{t}\in\mathbb{R}_+^{|\mathcal{E}|}\to\mathbb{R}_+^{|\mathcal{E}|}$
Set $k=1$;
Take any $\mathbf{f}^k\in\mathcal{F}_{\boldsymbol{d}}$;
while $\bar{\mathcal{R}}(\mathbf{f}^k)>\epsilon$ **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: $\mathbf{f}^k$

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

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, app users do not have "regrets" while non-app users "regret" to not know $\mathcal{P}_{od}$. We can express the average deviation incentive associated with $\mathbf{f}_\alpha^\star$ (as in remark 3.1.12):

$$\bar{\mathcal{R}}\left(\mathbf{f}_\alpha^\star\right) = \sum_{o,d\in\mathcal{V}} \sum_{p\in\mathcal{P}_{od}} h_p \cdot \left(t_p(\mathbf{f}_\alpha^\star) - \min_{\tilde{p}\in\mathcal{P}_{od}} t_{\tilde{p}}(\mathbf{f}_\alpha^\star)\right)$$

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

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

Then eq. (3.13) gives:

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

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 $\bar{\mathcal{R}}\left(\mathbf{f}_\alpha^\star\right)$ with respect to α.

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

Put the full proof here (from TRB paper).

Proof. This follows from the continuity of the average deviation incentive (see property 3.1.4) and of the continuity of $\mathbf{f}_\alpha^\star$ with respect to α. The continuity of $\mathbf{f}_\alpha^\star$ with respect to α is due to the convexity of the restricted path choice model (definition 3.1.15, see [76, Theorem 1] for a more detailed proof). □

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

$$\bar{\mathcal{R}}\left(\mathbf{f}_{\alpha_2}^\star\right) \leq \bar{\mathcal{R}}\left(\mathbf{f}_{\alpha_1}^\star\right) \textit{ and } \lim_{\alpha_2\to 1} \bar{\mathcal{R}}\left(\mathbf{f}_{\alpha_2}^\star\right) = 0$$

Proof. First, given that $\bar{\mathcal{R}}\left(\mathbf{f}^{\star}_{\alpha}\right)$ is continuous with α (theorem 3.2.1), $\bar{\mathcal{R}}\left(\mathbf{f}^{\star}_{\alpha=1}\right) = 0$ (eq. (3.18)) gives that $\lim_{\alpha \to 1} \bar{\mathcal{R}}\left(\mathbf{f}^{\star}_{\alpha}\right) = 0$.

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 [56, 118]). By using eq. (3.18), we have:

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

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 [56, 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 3.1.15) 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 [56]. Then, the Dafermos sensitivity analysis of the travel cost with respect to the demand [56, 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:

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

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

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 converges to 0 as the app usage increases. Then, on the full Los Angeles, CA network, simulations show the same phenomenon with the restricted path choice model. Because the restricted path choice model is a static model, we also simulation traffic with the Aimsun microsimulator, that simulates each vehicle independently. Calibrated simulations of the I-210 traffic in Los Angeles, CA confirm the trends when representing information-aware routing with dynamic rerouting models. A blueprint to create, calibration and validate an Aimsun traffic microsimulator is shared in chapter 5.

The Bureau of Public Roads (BPR) function [120]. For each link we define a link capacity $c_l \in \mathbb{R}_{>0}$ which is roughly the total flow a link can have before being congested [120]. Then we define the link travel function as $t_l(f_l) = \frac{d_l}{v_l}\left(1 + 0.15\left(\frac{f_l}{c_l}\right)^4\right)$ with $d_e \in \mathbb{R}_{>0}$ and $v_l \in \mathbb{R}_{>0}$ the length and the free-flow speed of the link. $\frac{d_l}{v_l}$ is the free-flow travel time of the link l. The function t_l is referred to as the BPR function in transportation literature [133, 120].

Open source code to solve static traffic assignment used for the work. Transportation networks with road capacity, free-flow travel times and travel demand are available in open source form on the website [75]. A static traffic assignment solver can be found on the website [94]. Open Street Map, and OSMNX [28] can be used to model the considered road network.

App usage on a benchmark network

The first computation (shown in figure 3.4) studies the impact of the number of app users on the traffic state. The highway capacity is 6,000 veh./h. The arterial road capacity is fixed at 2,000 veh./h. The *od* demand is set to 20,000 veh./h (more than the capacity to observe congestion). 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 3.1.15), non-app users stay on the highway regardless of the traffic conditions.

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

Figure 3.4: Benchmark network (above) and path travel times (on the left) and the average deviation incentive (on the right) as a function of the percentage of app usage. On a benchmark network, the traffic converges to a user equilibrium state when app usage increase.

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.

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 [55]. 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 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 LA network

The second simulation is performed on the LA network (figure 3.5). Simulations use the cognitive cost static model [169] 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* pairs. The network is built from Open

Street Map. Traffic demand is set consistently at rush hour levels to find the effects of app usage when networks are congested.

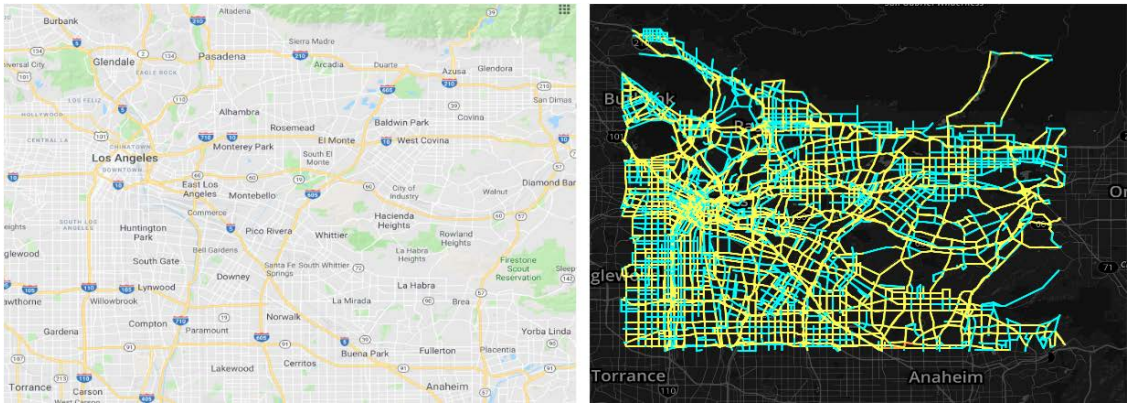


Figure 3.5: LA network considered for experiment. On the left: a map of the LA basin, on the right: the graph we use to model the LA basin.

We see that the average deviation incentive decreases monotonically with the increase of navigational app usage (figure 3.6). The fact that the decrease is monotonic is important here. This shows that, whatever the percentage of app users is, the traffic will be closer to a user equilibrium when app usage increases.

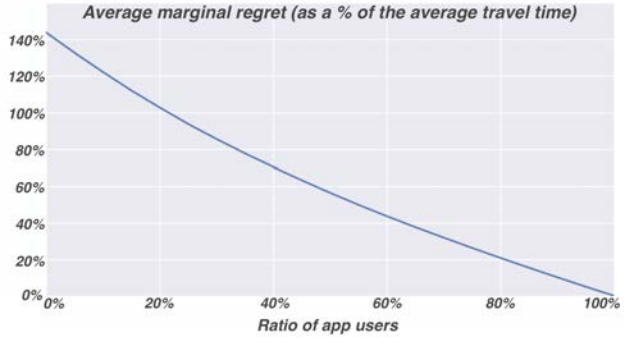


Figure 3.6: Average deviation incentive as a function of the percentage of app usage on the LA network. The average deviation incentive decreases monotonically to 0 when app usage increases.

Remark 3.2.1. *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.*

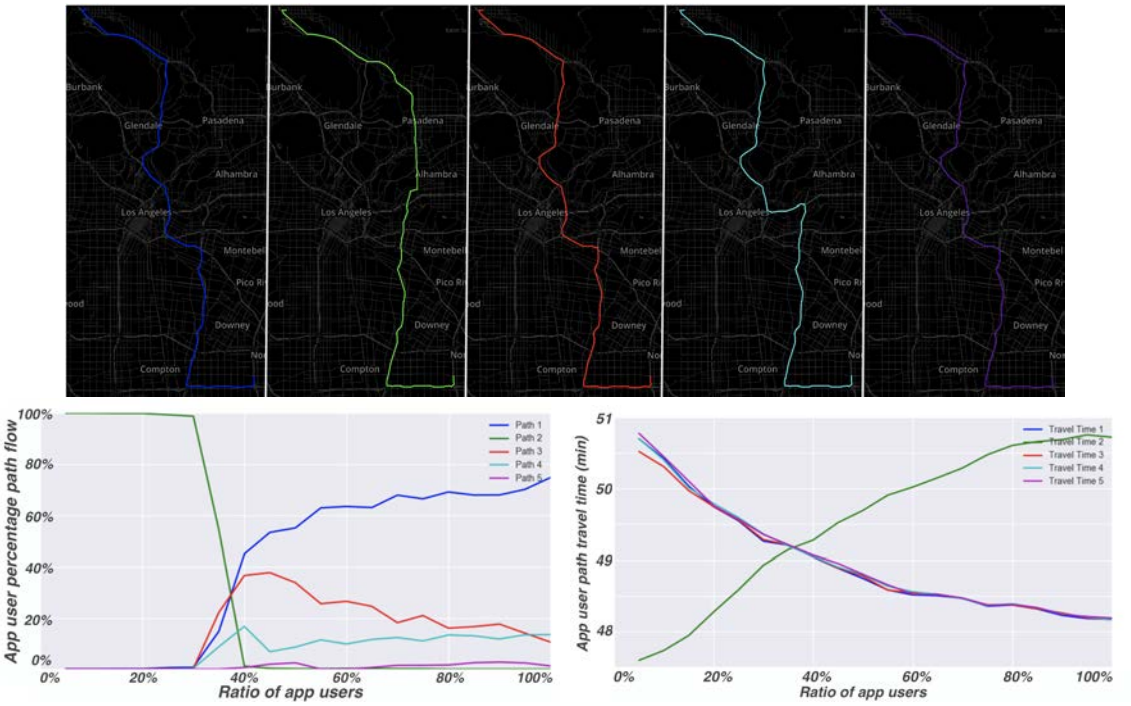


Figure 3.7: Impact of the increase of app usage on path choice and path travel time for a specific (*od*) pair with the increase of app usage. Above: the 5 main paths are used by app users. The blue path is the main path used by non app users. Below on the right: the travel time of the 5 paths as a function of the app usage. Below on the left: the percentage of flow of app users on the 5 paths as a function of the 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.

We show the evolution of path flow for a particular *od*. This (*od*) has been chosen to be one of the (*od*s) with the highest demand. This particular (*od*) starts in slightly southeast of Compton and ends just north of Burbank. Figure 3.7 shows the top five paths taken for this *od*. 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 3.2.2 (Travel time evolution)**.** *It is important to see that here the travel time of these paths depends on other ods. Even after 35% app usage, when no rerouting occurs, the path travel time still varies, Mainly because vehicles from other ods 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 have implemented the concepts embedded in these models into microsimulations (which integrate app usage at the individual vehicle level). This was completed using TSS' Aimsun on the I-210 corridor segment between Pasadena and Monrovia (Figure 3.8). *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 [131]. Data from the California 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 behavior 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, 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 high cost cycle (e.g. 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 the path, which will lead to the congestion of the path. However, the cost of the path is not updated (since the cost cycle is high) so app users will continue to route onto the path, further worsening the congestion. To prevent such effects, we use a one-minute cost 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,

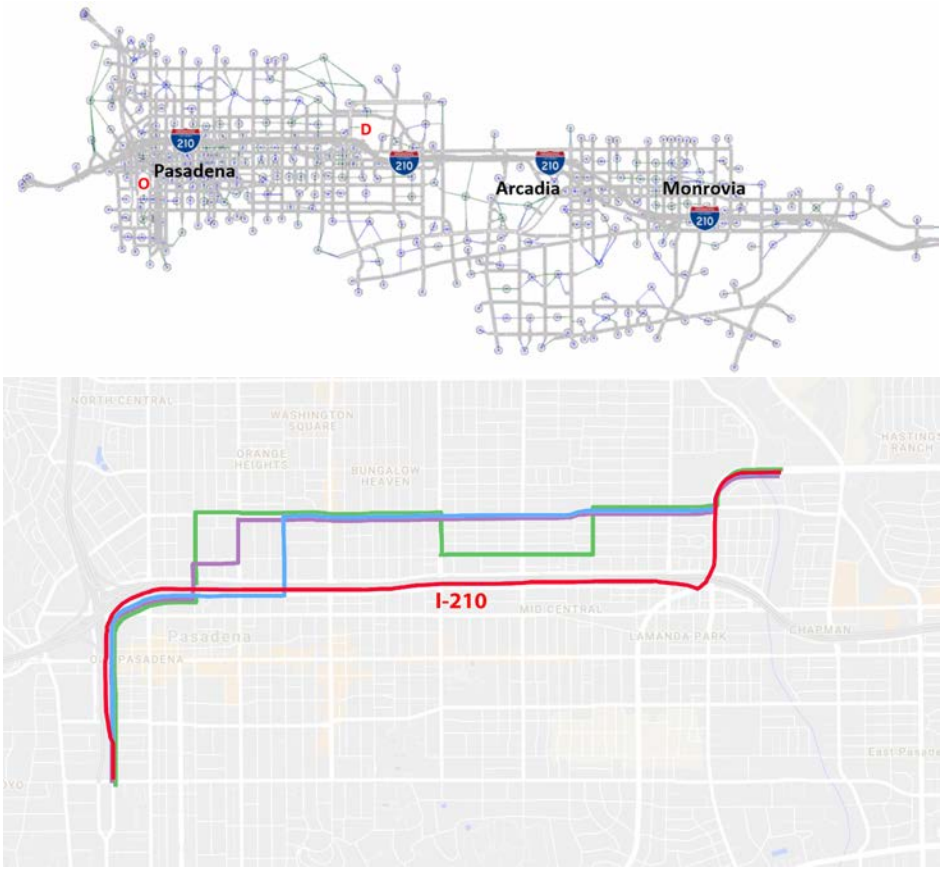


Figure 3.8: (a) The selected *od* pair for the Aimsun experiments. (b) A selection of alternate paths that app users took instead of the I-210 and the I-210 route (taken by non-app users) shown in red.

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 (6:00 AM - 7:00 AM) is obtained from PeMS and city data collected for the Connected Corridors project [185]. During these peak hours, 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 as shown in Figure 3.8. 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.

We perturb the number of app users between this OD pair from 10% to 90% with results shown in fig. 3.9. As described previously, non-app users follow paths determined by computing a static user equilibrium in the network given the same demand. Similar to the results in the benchmark network, we observe that the path flows and travel times of the main path

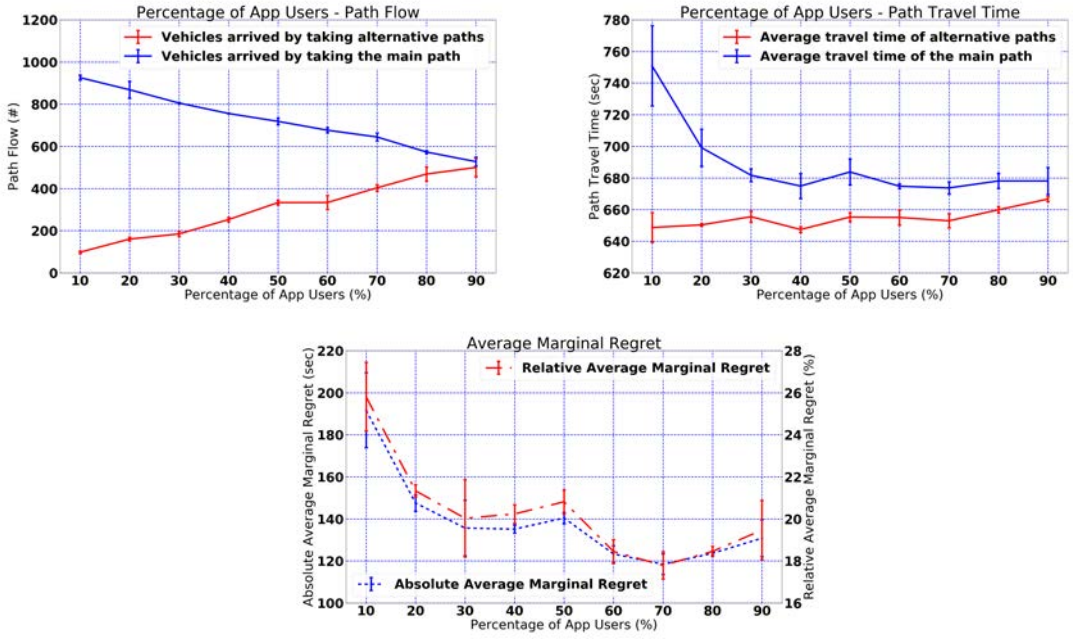


Figure 3.9: I-210 network without accident. **Top left:** Path flow on the main path (freeway shown in red in fig. 3.8) and all alternative paths (shown in blue, purple, and green in fig. 3.8). **Top right:** Path travel time convergence between the main freeway path and all alternative paths. **Bottom:** 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.

(freeway) and alternative paths (shown in orange and green in fig. 3.8) converge as the number of app users between the specific OD pair increases. The the average deviation incentive shown in fig. 3.9 also decreases as the percentage of app users increases. However, there are some slight increases in the average deviation incentive, specifically at 50%, 80%, and 90% app users. This phenomenon may be caused by the cost (i.e., travel time) of a path, which is calculated based on the current network state instead of the future or predicted state. As a result, large numbers of app users will reroute themselves onto alternative routes with low costs before the costs of these alternatives are updated, leading to a travel time longer than expected. Hence, the graph 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.

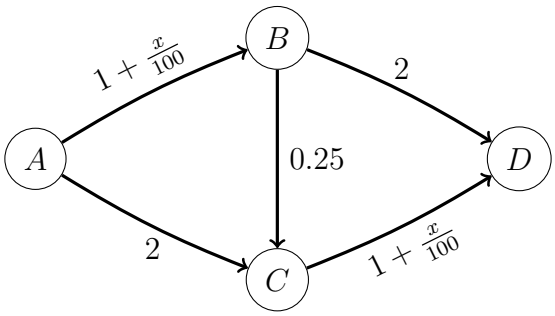


Figure 3.10: Braess network considered. Cost on every link are given as a function of the link flow.

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 this increases or decreases the 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 network and demand considered.

On the Braess network [31], we show that an increase of app usage might decrease the individual travel time of each vehicle (app users and non-app users) in the network, or might also increase the individual travel time of each vehicle. The fact that the travel time of a vehicle that switches from non using an app to using an app is known as the Braess paradox [31]. The Braess paradox is a replica of the prisoner's dilemma in routing games. Cite.

When apps 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 a network 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 cost functions are defined in figure 3.10. This is equivalent to the experiment of [53].

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 the evolution of the travel time of app users and non app users as a function of the percentage of app users (figure 3.11).

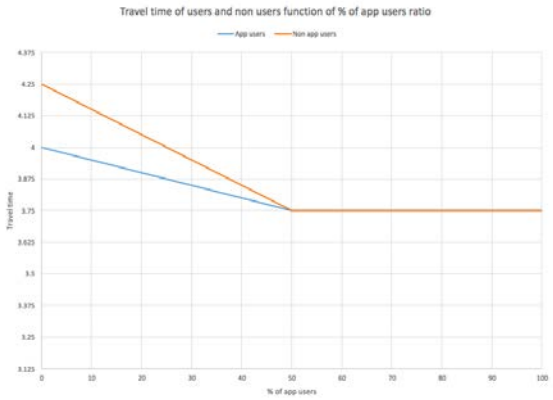


Figure 3.11: Impact of the increase of app usage on the Braess network: everybody gets a better travel time when the app usage increase.

When app usage is worse for everybody

In his book, 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 increase the network performance).

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* [103], where the cows have been replaced by motorists, the grass has been replaced by road capacity, and the action of eating the grass is now embodied by the ubiquitous use of information-aware routing. 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 socially optimal traffic assignment; this is the price of anarchy [119, 133, 145]. We claim that routing apps can reproduce the Braess paradox [31]! 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, let us consider the same setup as the previous subsection (fig. 3.10), but with a different knowledge of the network 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 path BC. Therefore, they are routed on the shortest path between ABD and ACD.

We observe the evolution of the travel time of app users and non app users as a function of the percentage of app users (figure 3.12).

In this particular case, even if the usage of apps decreases the average deviation incentive, 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 usually is not socially optimal.

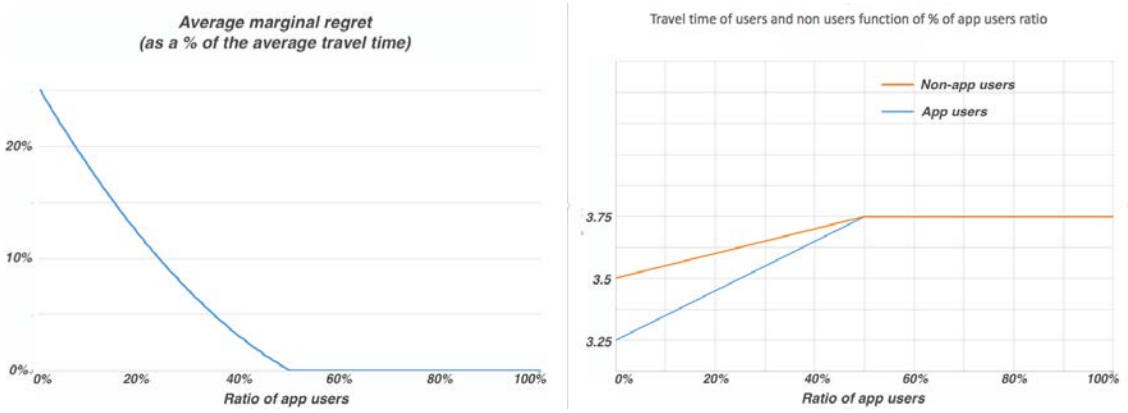


Figure 3.12: Impact of the increase of app usage on the Braess network: everybody gets a worse travel time when the app usage increases. On the left: the average deviation incentive as a function of the app usage. The average deviation incentive of the vehicles decreases monotonically when app usage 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.

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 capacity and not the daily urban planning capacity. Ultimately, navigation apps influence the distribution of vehicles across space and time within the transportation system, without taking into account urban planning road capacity, 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 among others. Chapter 4 studies the negative externalities of cut-through traffic 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: Out 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*

Information-aware routing has changed the state of traffic (see chapter 3), mainly by spreading congestion away from congested areas. In 1920, Arthur Pigou [137] showed that information-aware routing (or selfish-routing) can lead to a suboptimal state of traffic due to the price of anarchy[145]. Beside a possible decrease in the transportation operational performance in the Pigou paradox thought-experiment, information-aware routing occurs a cut-through traffic on some local roads. Section 4.1 showed that both travel time equalization (Wardrop equilibrium condition, see definition 3.1.9) and 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 capacity by the facility demand. As predicted by Thomas Malthus [107], this

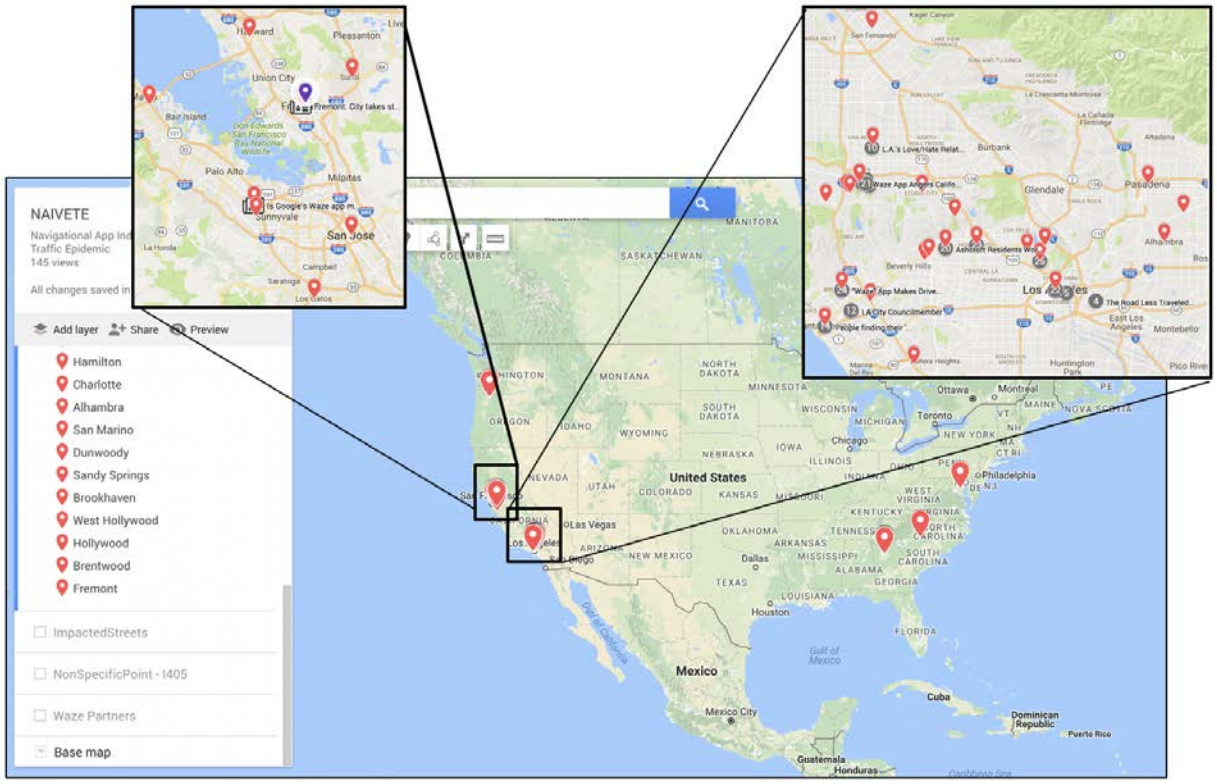


Figure 4.1: Sample of the occurrences of the negative externalities of the information-aware routing in the popular media in the US and specifically Northern and Southern California.

lead to negative impact on the system efficiency. Indeed, congestion incurs fixed, non-market and indirect costs for the society [102], which can lead to anger from residents negatively impacted by traffic, as shown in section 4.2. On the ground, in neighborhoods across the U.S. and abroad, residents have alerted the public to their experiences. Countless reports in the media (see fig. 4.1) exhibit 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 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 tackle by impacting the route-choice of the vehicles. Unfortunately, few traffic-management tools can impact route-choice; some of which are presented in section 4.3.

In this chapter, section 4.1 uses ground truth data to show the existence of cut-through traffic due to an increase in navigational app-usage. Some materialization of the negative

externalities of cut-through traffic due to information-aware routing are shown in section 4.2. Finally, section 4.3 tries to showcase some mitigation techniques that have been used or can be used by transportation planners, managers, and operators.

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

Due to information-aware routing, vehicles – that will usually stay on the main road for their trips – might now 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, app-users will take such short-cuts, changing traffic patterns and changing of the facility demand. Eventually, it will create traffic on local roads in residential areas. If the new facility demand implied by cut-through traffic in local roads exceeds the daily urban planing 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 – will incur.

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

In this section, a case study – borrowed from [37] – 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, five routes – routinely suggested by Google Maps – 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 in March 10th, 2014, the travel time on the freeway increased from 3:30pm to 4pm until it is not beneficial to use the freeway anymore. At this point (4pm), then the travel times of all paths increased until 4:20pm, and they remain almost equal between 4pm to 5pm. This travel time equalization between the paths between 4pm to 5pm translate that the state of traffic is in a Nash equilibrium. This is expected if some vehicles in the network follow navigational-app information-aware routing suggestions.

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

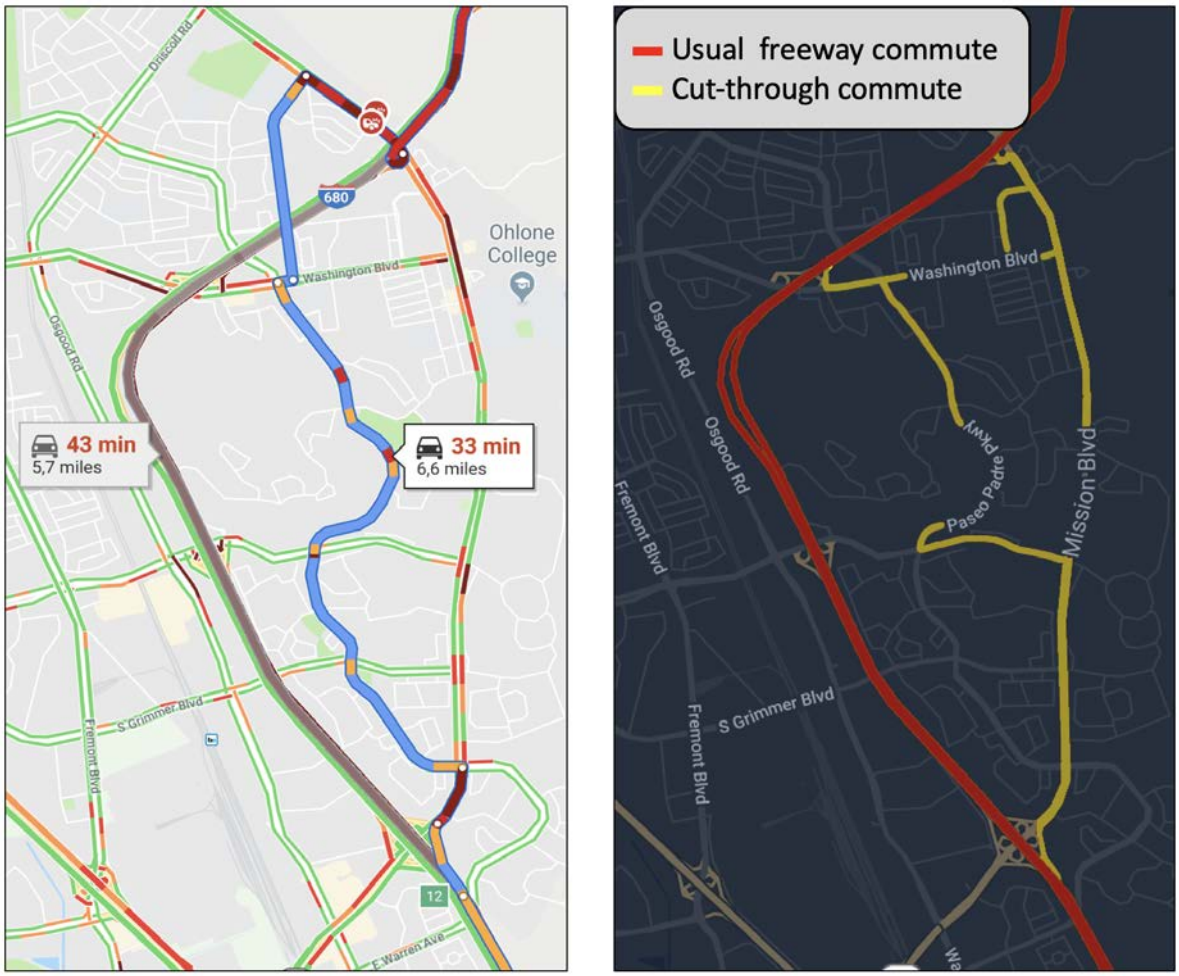
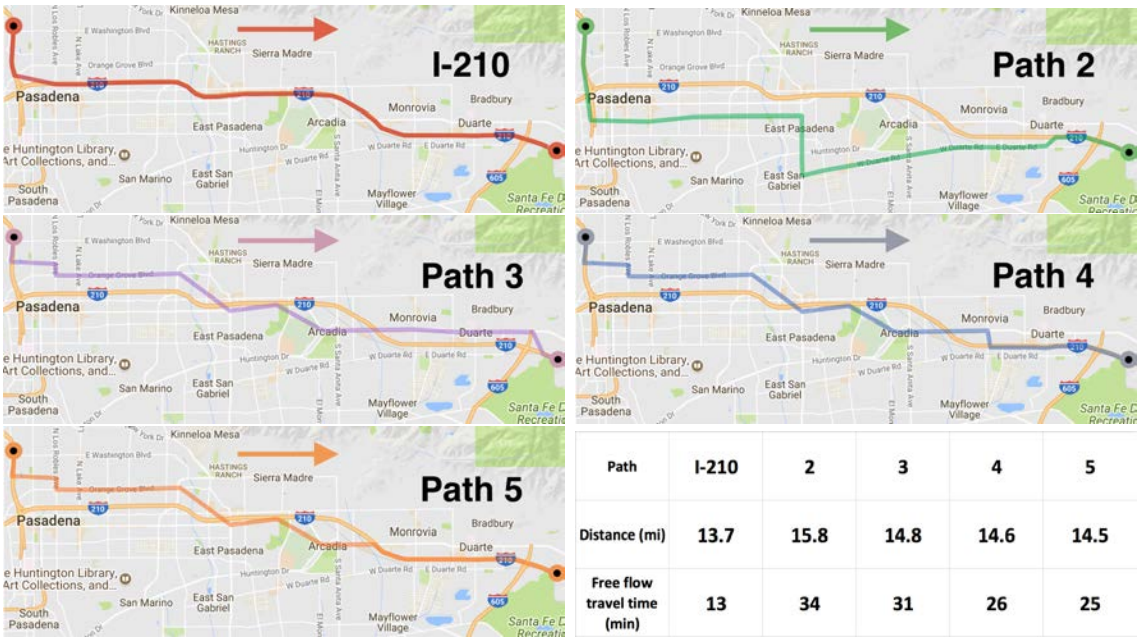


Figure 4.2: Cut-through traffic in Fremont, CA. **On the left:** Screenshot of directions proposed by Google Maps in July 2019 to go from the South of the map to the North East of the map. Distance between the west and the east side of the map is around 2 miles. **On the right:** Fremont, CA map with the localization of usual freeway commute and cut-though commute.



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 considered for this work, with four alternative arterial paths. 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.

This equalization of the travel times translates in the spread of freeways' congestion to parallel side-roads, as shown in chapter 3 on fig. 3.1. If the spread of congestion from highways to side-roads can be seen as an increase of the network efficiency (at least by the public relations departments of the companies developing navigational apps [164], [191]), it has been shown that this phenomenon might generate high flow on low-volume capacity roads [169, Figure 6].

Human-based knowledge (newspaper articles, resident complaints), tracking of the navigational apps route suggestions (fig. 4.2), and consulting case studies [129] has qualitatively exhibited that app-usage leads to an increase of cut-through traffic. To quantify how much the increase of traffic flow is due to navigational app usage and how much is due to the increase of the traffic demand, traffic data are 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 [173]:

- **Cross-sectional data** are traffic counts. At a specific location, a detector counts the

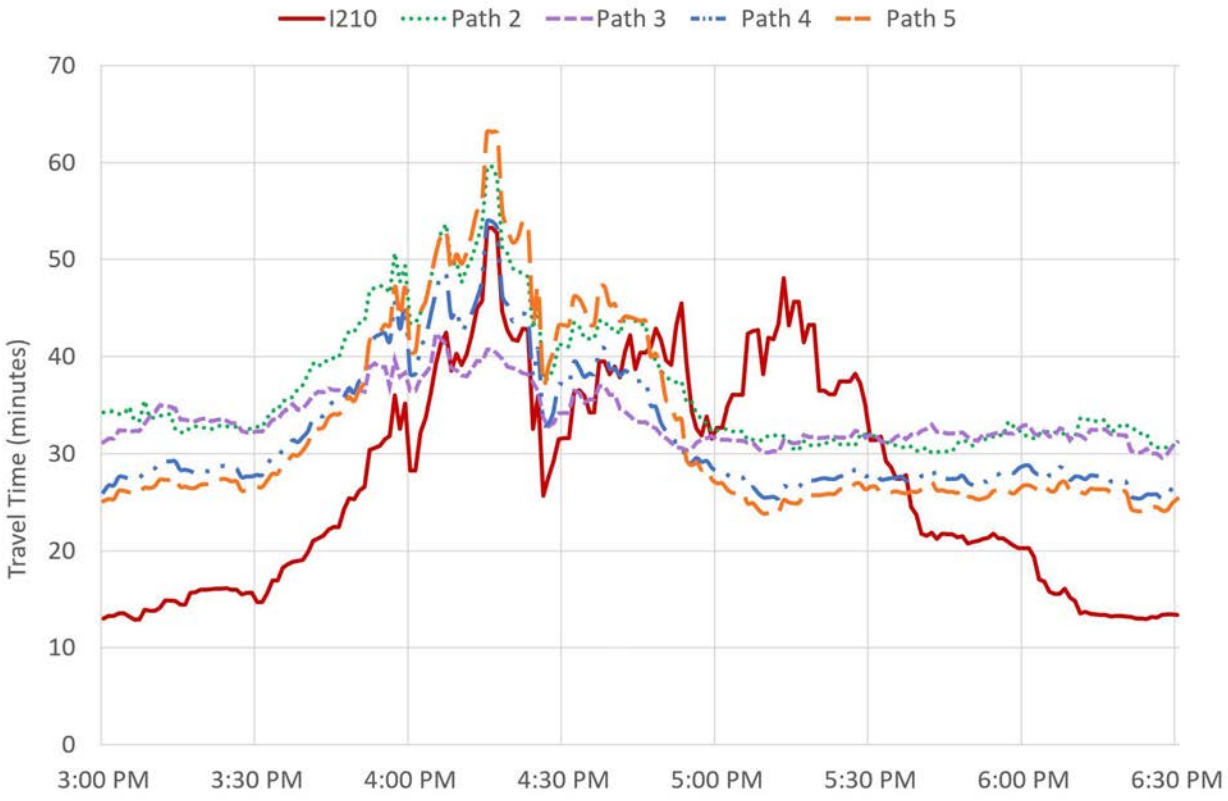


Figure 4.4: Evolution of travel times on the paths parallel to the I-210 considered in fig. 4.3 (March 10, 2014). At the beginning and end of the peak hour, the travel time on the freeway is close to the free flow travel time, and the freeway is faster than the arterial road routes. However, when the freeway travel time increases, the arterial detours become beneficial alternative routes. This figure shows that in high congestion, drivers can reroute themselves to arterial roads in order to reduce their travel time, leading to travel time equalization among possible parallel routes.

number of vehicles going through a given road section (see fig. 4.5). It provides traffic engineers the traffic flow of the road section depending on the hours of the day.

- **Trajectory and floating-car data** are the trajectories of every or a subset of vehicles that are inside the road network (see fig. 4.5). It requires vehicle identification and provides to traffic engineers how the traffic flow moves inside the network.

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

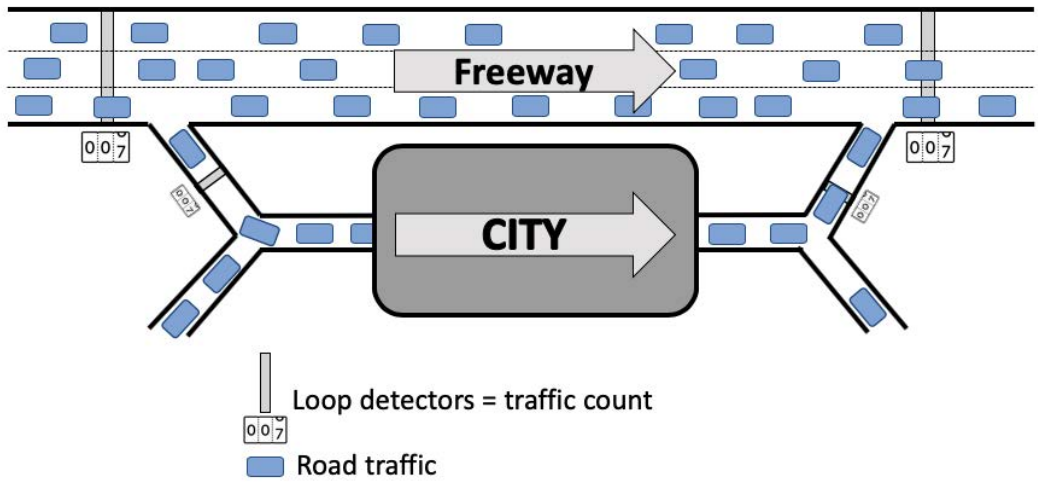
Cross-sectional data can be seen as traffic counts: they give the traffic flow at a specific location at any time [173]. From a fluid mechanics point of view, cross-sectional data are called Eulerian data. They might also be called static data.

Cross-sectional data can be obtained from several types of sensors including [89]:

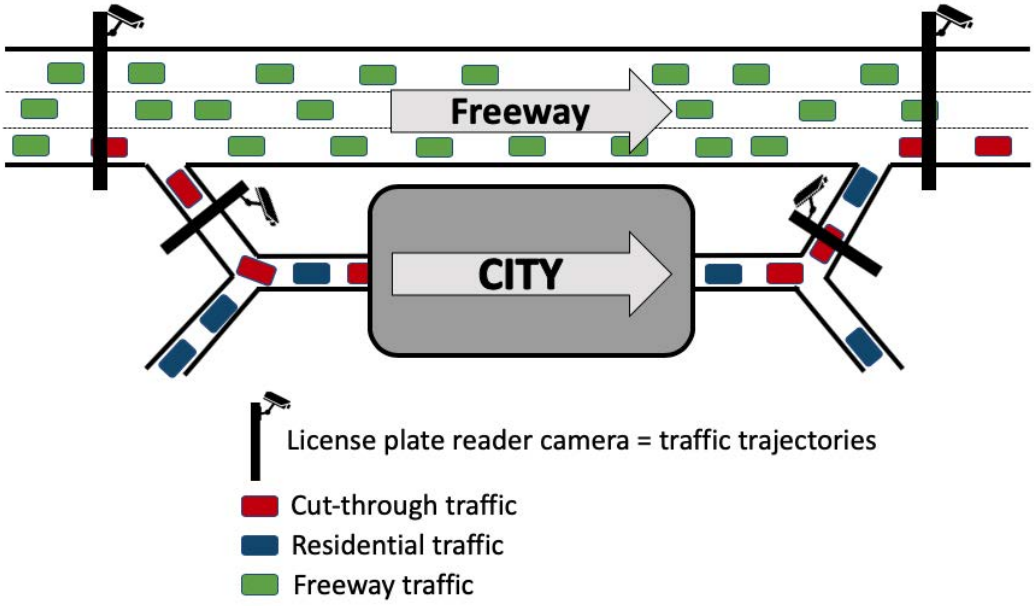
- Inductive Loops, show in fig. 2.4
- Magnetometer
- Magnetic
- Microwave radars, shown in fig. 2.4
- Active infrared or laser radars
- Passive infrared
- Ultrasonic
- Acoustic
- Video Image Processors, show in fig. 2.4

Combined inside a system, cross-sectional data can indicate the traffic flow on almost ever road section of the network (called link flow allocation). For example, current traffic flow on Californian highways is monitored by PeMS [186]. PeMS data, which consisted of flow data from inductive loop sensors, are publicly available.

PeMS data along the I-210 roadway (path 1 of fig. 4.3) and its associated ramps was analyzed. The median (chosen to reduce the effect of outliers) of the total evening peak flow was found over the weekdays of the month of March. Specifically, flows exiting the freeway were analyzed at 4 different off ramps and the yearly trend was examined. The selected off ramps include exits commonly suggested by applications like Google Maps (Michillinda Ave and Baldwin Ave exits), and the exits directly before and after them. All off ramps along this route were examined, and the selected off ramps showed significant change, as they were most often suggested by the applications.



(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. They count the number of vehicles going through specific points. Aggregated in a system, these data can estimate the number of vehicles on any road section. However, they cannot explicitly show the existence of cut-through traffic as they do not know the routes of the vehicles. (b) Trajectory data can be obtained through license plate reader cameras. They identify every vehicle and derive its route. Combined in a system, they can estimate the number of vehicles on any route of the network. They can explicitly show the existence of cut-through traffic (i.e., the vehicles in red in the figure).

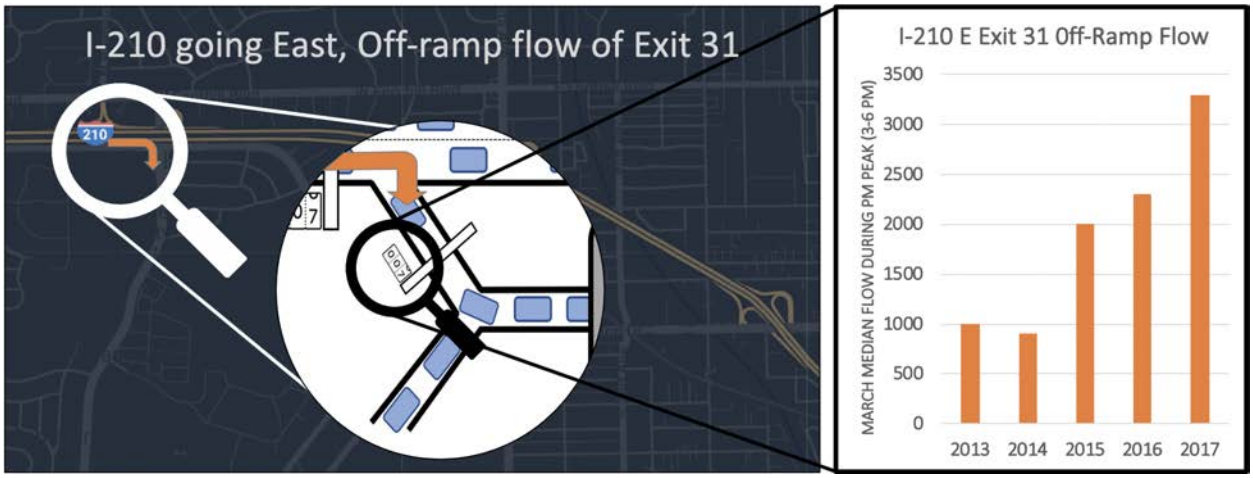


Figure 4.6: **Left:** localization of the off-ramp traffic detector of the exit 31 of the I-210 East in Arcadia, CA. **Right**: Evolution of the median off-ramp flow from I-210 during March weekday peak hours between 2013 and 2017. The 3-fold increase in off-ramp flow that occurs between 2014 and 2017 coupled with the decrease in speed on parallel paths provides evidence in favor of app-induced arterial rerouting patterns.

A significant increase in flow using these off ramps is observed between 2013 and 2017, as shown in fig. 4.6. Specifically, at the exits for Michillinda Avenue and Baldwin Avenue, we see a 1.5- and 3-fold increase respectively over 4 years. While some of this increase can be explained by an increase in demand to Arcadia, it is unlikely that this can be explained solely by demand growth. Additionally, since these exits are often suggested by navigation-apps which have increased in popularity during the same time period, it is likely that the increased flow can be partially explained by app usage. Nevertheless, cross-sectional data do not reveal the cut-through caused by the drivers using exit 31.

Remark 4.1.1. *Remark that 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 hour of a single day (in fig. 4.4, or examining off-ramp flow increase on congested highway across the year, 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 in each year. As expected, in a one-year time span, the travel times along the alternative paths increased by roughly 20%, i.e., around five minutes (fig. 4.7).

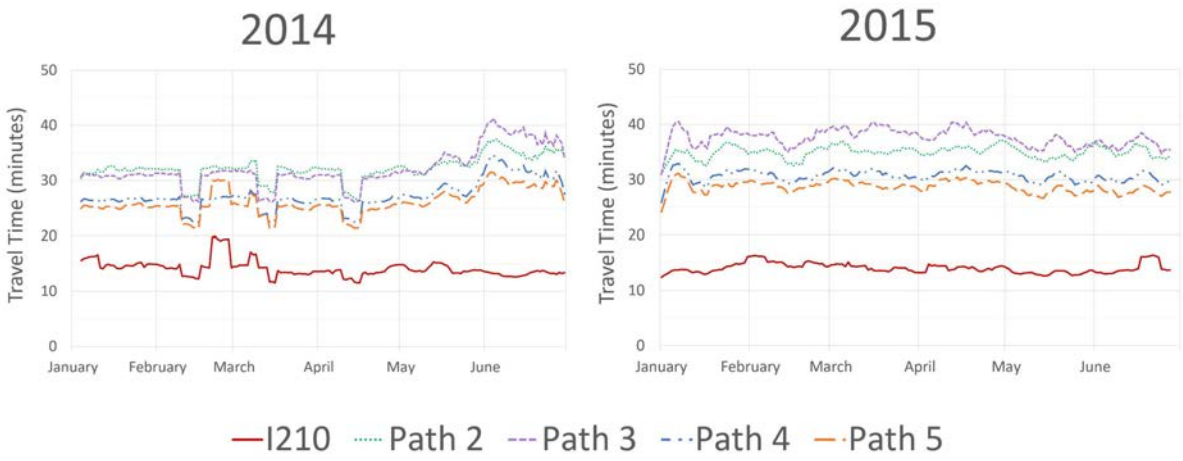


Figure 4.7: Evolution of the average travel time computed with INRIX data on the five paths during peak hours (4:30 to 5:30 PM) considered in 2014 and 2015, for each week from January to June. While the travel time on the I-210 remains roughly constant over two years, alternative paths suffer a 20% increase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

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

The travel time on the I-210 path oscillates around 15 minutes over the two years, remaining roughly constant over time. This occurrence can be explained by *latent demand* 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 (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 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 are the trajectory of every vehicle that is inside the road network. Floating-car data are a subset of the trajectory data: it is the trajectory of some vehicles inside the road network. From a fluid mechanics point of view, trajectory data are called Lagrangian data. They might also be called dynamic data.

Current trajectory and floating-car data require identifying the vehicles inside the network. They can be obtained from [173]:

- License plate readers
- Camera vehicles tracking with drones or fixed cameras (like Next Generation Simulation data [184])
- GPS devices that give GPS position in real-time of vehicles
- GPS traces that give past GPS position of vehicles

High penetration floating-car data can explicitly show the existence of the cut-through traffic caused by the navigational app users. These data will show the cut-through used by the navigational app users, they will also reveal the quantity of people using the different cut-through routes.

Currently, the authors are not aware of any traffic control center or any city traffic engineers 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 the GPS tracking of the 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 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 the navigational application to share their traffic data[12]. However, they are obvious privacy issues that might disable navigational applications to share any user data to any government entities. Also, a lot of navigational apps are trying to store fewer users' location data on their server to increase users' privacy.

cite

cite

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 the traffic congestion is now more spread onto local roads near congested highways than before the emergence of smartphones. It becomes a challenge when some of these roads receive a higher traffic flow than the one they are designed to sustain (i.e., when the facility demand

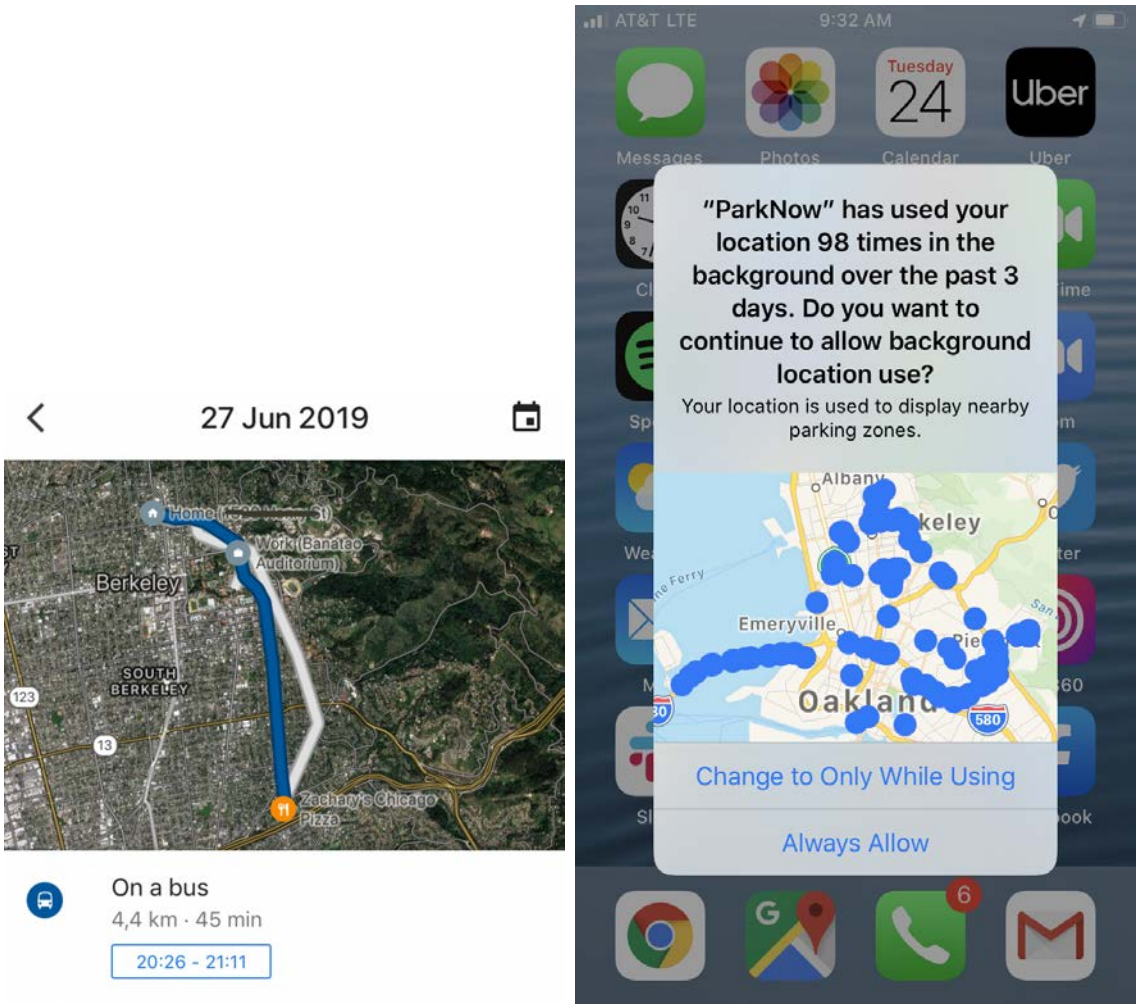


Figure 4.8: Google Maps localization and transportation mode GPS tracking. Screenshot of June 2019.

on some local roads exceed their corresponding urban planning capacity). It eventually turns into a nightmare for the residents living near these streets. Inherently, this is due to the fact that navigational-app 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 do 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 crossing, location of elementary schools, or difficult intersection-crossing.

The geometry of the roads are designed as a function of the expected flow that they would receive [5] based on their category (fig. 2.8). As shown in fig. 2.8, local roads are designed to receive less than 400 vehicles per day in average in rural areas and 700 in urban areas [6]. In fact, a countless number of complaints has 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 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 [64], noise, gas emissions, traffic accidents (fig. 4.9), decrease accessibility in affected neighborhoods, bad directions [91], infrastructure damage, etc.

Cut-through traffic changes the spatial distribution of noise level and emission in the neighborhood [126, 62]. The health impact would only be approached from a long-term perspective because transport's detrimental impact on health is a long-term accumulation. From individual's long-term perspective, congestion increases psychological stress [189, 126] and decreases people's leisure time, which consequently decrease people's physical activities. Many studies and media reports have shown that road traffic system has a strong association with the decreased mental wellbeing 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 [126].

Infrastructure damage. As the cumulative load on the road is a function of the number and weight of vehicles, infrastructure damage is an increasing function of the number of vehicles using the road [123, 88, 166]. Recognizing that current road users usually adopt various transportation modes including driving, bicycling and walking, policy-makers adjust the component of transportation funding such that everyone bears the cost of transportation [171]. Therefore, there are three major funding resources (cost of driving): (1). general taxes, such as incomes and sales tax ; (2). gas taxes; (3) revenue from tolls and fares [171, 25]. Most of these revenues go to the state, therefore, the state mainly bear the cost of trans-

portation. Especially, the state is in charge of maintaining highways; the roads designed to be used by commuters. The state is not in charge of maintaining local roads. Because small cities cannot afford damages cause by heavy usage of their local roads, if the commuters use some local roads, then the state should subsidy 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 [38]. In Leonia, NJ, one pedestrian died struck by a bus in 2014 in the Ford Lee Road, a road with a lot of cut-through traffic [65].

On 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 of them 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 [139]. For example, Google Maps recently launched eco-routing to account for fuel consumption [72, 11].*

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

Cut-through traffic and resultant 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, LA, Fremont, CA, Bordeaux, France, Lyon, France, Meudon, France, have been vocal about these issues [105].

In early 2018 residents of Echo Park, Los Angeles County, CA began reporting excessive through 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 brake failures. Many vehicles even spun out onto neighbors' gardens [161] (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 [40]. In Los Angeles, residents of the Sherman Oaks neighborhood have also reported high traffic due to cut-through [16].

In Leonia, NJ (9,200 residents), the traffic flow on Fort Lee Road has increased from 4,000 to 14,000 vehicles per day between 2014 and 2017 according to the Mayor Judah Zeigler [127].

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

If the traffic is uniformly distributed over 11 hours, it represents a traffic flow that **increased 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 [175] (see fig. 2.8). Consequently, safety concerns have been raised by the residents [84]. Indeed, a pedestrian died in 2014 on the Ford Lee Road after being struck by a bus [65].

In Fremont, CA, faced with increased congestion due to altered commuter behaviors, Mission San Jose neighborhood residents encounter heavy traffic in not just driving to and from work, but also in 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

Figure 4.10: Photograph of a congested Fremont, CA neighborhood near I-680 taken from a drone camera, from [46].

has always been severe [3]. 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 [128]. 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 the Silicon Valley, which might have counter balanced the improvement due to the city's policies [170]. With the continued increase in congestion, local residents have started blaming the problem on city policies for allowing the development of new housing units and office parks nearby.

In Lyon, France, 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 South-West suburbs that do not take the congested N118 highway.

Section 4.1 shows that the cut-through traffic only happens during peak congestion hours. Therefore, 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 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 of the

travel demand or of the traffic demand 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 [181]. So, as a consequence of having more people, they are more travelers. Secondly, people are more clustered in cities. The **urbanization** (percentage of people living in urban areas) is growing everywhere in the world: in 2020 55% of the world's population lives in urban areas, in 2050 the number is expected to be 68% [182]. As a consequence, the increase of travelers is focused on metropolitan areas, creating more travel demand for these specific areas. An indirect consequence of the urbanization – that increases even more the travel demand – is the development of urban sprawl, leading to an increase of the average commute distance. For example, in France, the median commute distance has increased from 7 km in 1975 [168, Table 2] to 8 km in 1982 [168, Table 2], 13 km 1999 [54, Table 6], 15 km 2013 [54, Table 6].

Fremont, CA is a striking example where the increase of travel demand has largely exacerbated cut-through traffic (see fig. 4.11). Fremont, CA is located in the San Francisco Bay Area at the East of the Silicon Valley. Between 2014 and 2018, 152,000 new jobs have been created in the Silicon Valley. However, only 28,000 housing unit have been created in the Valley. On 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). The traffic on Mission boulevard is now so heavy that even small businesses have complained about the decrease of business accessibility. cite

Increase of the traffic demand

As a result of the increase of 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 the travel demand because of mode shift implied by **transportation network companies** (TNCs) and the increase of **on-demand delivery**.

Traffic demand has been considerably increased due to the sudden emergence of transportation network companies (TNCs) like Uber and Lyft [34]. Uber was created in March 2009. Ten years after, in 2019, Uber is estimated to have 110 million worldwide monthly users [157]. Statista states that 996.7 millions people use TNCs in 2019 [159]. 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, 59% of trips made by TNC are actually adding vehicles to the road network (see fig. 4.12 or [69, Figure 11]). Accordingly, **the worldwide increase of traffic demand due to TNCs can be approximately estimate to be 588 millions annual riders** (56% of 996.7 millions TNCs users). In

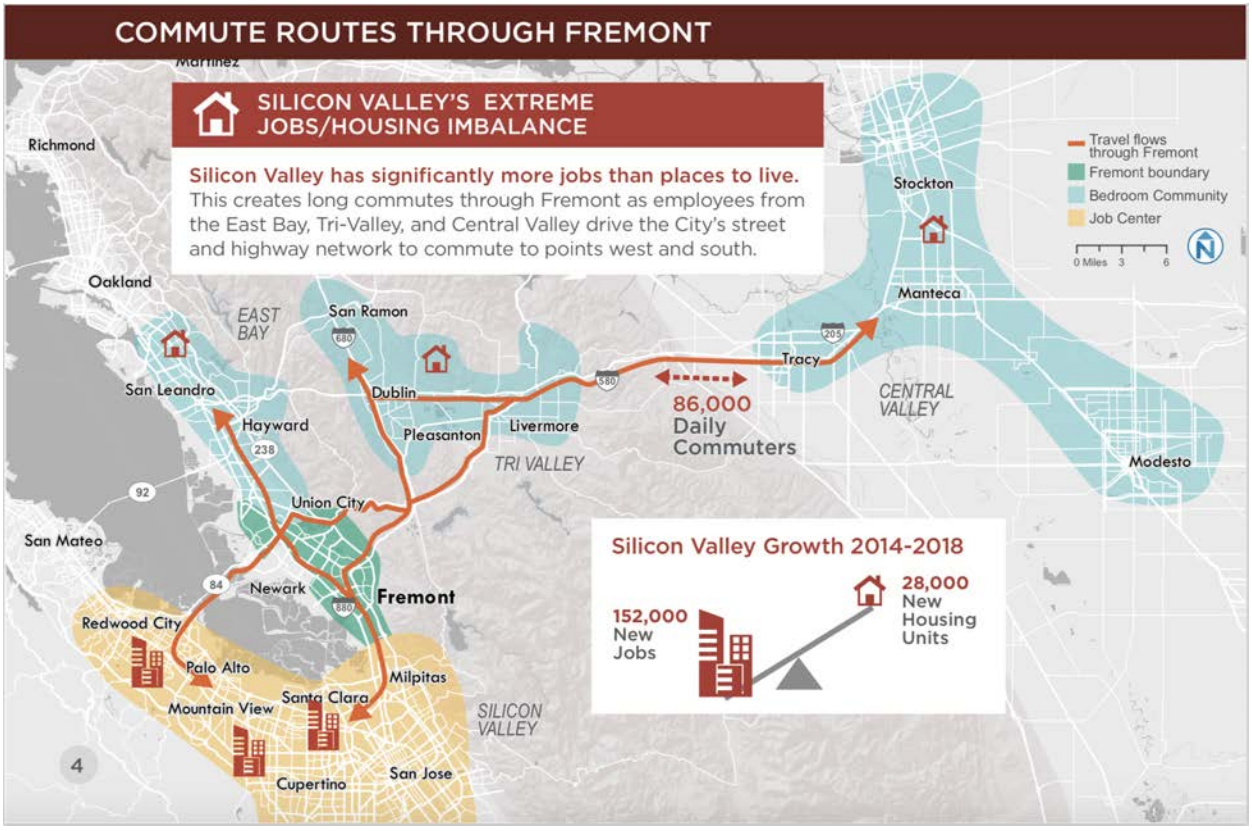


Figure 4.11: Schematic illustration of traffic created in Fremont, CA due to the demographic increase in the San Francisco Bay Area. Screenshot from March 2019 Fremont Mobility Action Plan [170].

San Francisco, CA, Lyft, and Uber are responsible for 13% of the combined Vehicle Miles Traveled inside the city [15].

Concurrently to the increase of the traffic demand due to TNCs, another increase of the traffic demand is due to the development of **last mile delivery**. According to the economic research department of the Federal Reserve Bank of St.Louis [8], e-commerce retail sales has heavily grown during the last decades. Representing 1% of the total sales in 2000, it now (2019) represents 10% of the total sales. Consequently, last mile delivery has densely increased the number of trucks and small delivery trucks on the roads. Brookings's analysis of Federal Highway Administration data [33] show that trucks – which used to drive 40% of the time on urban roadways (and 60% on highways) in 1966 – drive 60% of the time on urban roadways in 2014. This leads to an **8-fold increase of urban vehicle miles traveled of all trucks between 1966 and 2014**.

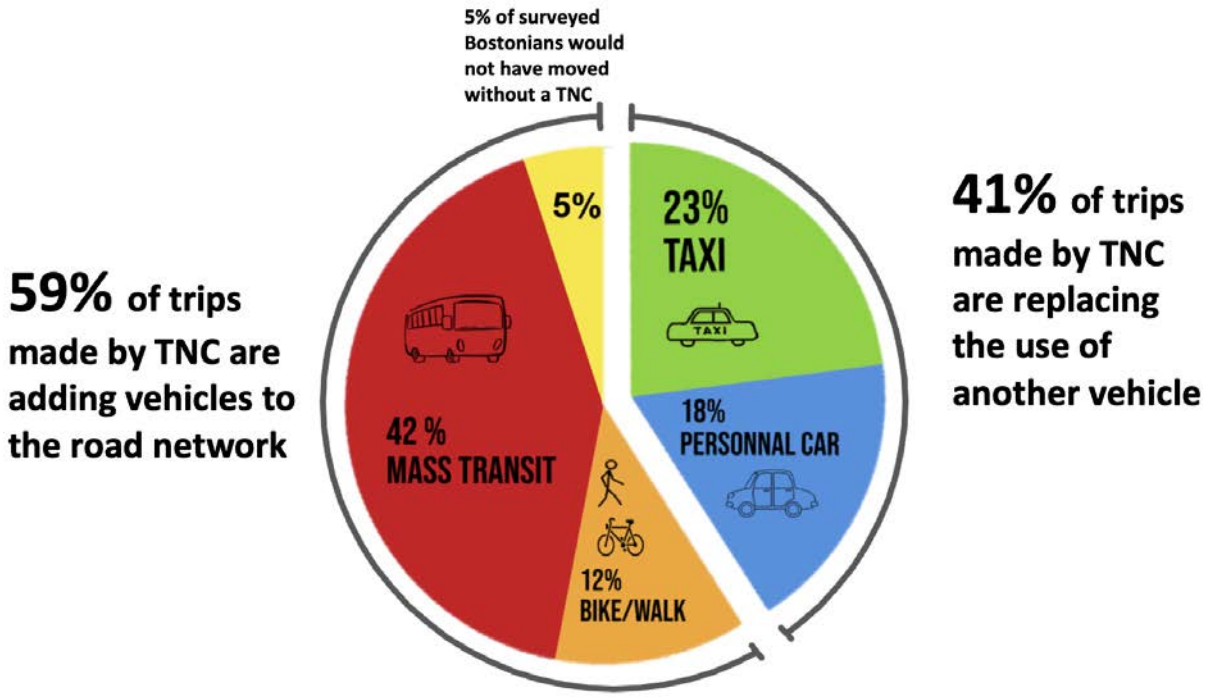


Figure 4.12: Mode of transportation replaced by transportation network companies (TNCs) like Uber and Lyft in Boston in February 2018 [69, Figure 11].

4.3 Cut-through traffic mitigation techniques

In order to mitigate the negative externalities of cut-through traffic, cities can use several approaches, including the ones listed in table 2.2. Generally, externalities is a concept from economics where free-market lead to some indirect cost to the society. Interestingly, the concept has first been introduced by Pigou in [137], 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 the previous sections of this chapter, 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 [45]:

- **Cap-and-trade**: assigning property rights for the supply usage.

- **Pigovian tax**: taxing the market that generate 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 [74].

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 economy theory, this approach was theorized by Ronald Coase in [52]; property rights are assigned to use of supply.

This approach has been used in the U.S. by Leonia, NJ. On January 22th, 2018, under the pressure of inhabitants of the city, the city closed 60 streets to thru-traffic during rush hours [100, 127]. Any commuter going through the 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 complained accusing Leonia of "infringing on her access to public roads and violating the public's right to freedom of travel" [140]. The road is opened back in August 2018 after the invalidation of the restriction ordinances by state Superior Court [93].

In France, some roads can be marked as residential-only. Commuters do not have the right to use residential-only road. 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 has decided to directly regulate the navigational apps [13]. In Tel Aviv, Israel, some residents have gathered to create a group that has filed suit against this navigational app providers [197].

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 restriction on non-residents create issues for local businesses because it was decreasing the city accessibility and, as a consequence, its attractivity [163].

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 (or market failures) of a market is to internalize them. More specifically, a regulator can tax the cost of the externalities to the market participants. Such taxes are called Pigovian taxes [150], as they were derived from

Arthur Pigou's work [137]. For example, a sugar tax that discourages unhealthy diets and offsets the economic cost of obesity is a Pigovian tax.

In the case of information-aware routing, an obvious Pigovian tax will be to charge commuters for using a specific route. As an illustration, between Monterey, CA, and Carmel-by-the-Sea, CA, the 17-miles drive is tolled in order to keep quiet the residential area near around it [176].

Lombard Street pricing. Another example is the 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 [148]. During a busy day, the scenic section of Lombard Street (between Hyde St and Leavenworth St) is 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 complain about the mess led by the visitors' spillover: indeed, according to a survey, the noise, the littering and trespassing incidents are making that place less bearable than a common residential area [148]. In order to counter those issues, many propositions have 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 for the purpose of dissuade visitors to use that road. A long-term solution was to implement a cut-through pricing to regulate the numbers of comings and goings. In June 2014, the "crooked street" have been completely closed for vehicles (except for residents) during the 7 most crowded hours of the day (12:00 PM to 6:00 PM). The result was that there were no more congestion above the problematical section of the street, and the pedestrians were more numerous [148]. Regulating the number of cars and modeling 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 will have to make a reservation and pay a fee ($5 or $10 depending on the day). The bill has been accepted by both the Californian assembly and senate with a large majority [148].

Check that the reference is the correct one.

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 [71]. Combined with license plate readers, the ERP system in Singapore is able to understand the routes of cars in the network. Therefore, this system enables to influence any vehicle's route choice through specific financial incentives. However, the author believes that due to potential privacy threads, it is unlikely that such systems will exist in the U.S. or in 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 [18]. The traffic management strategies presenting 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 [58]: it impacts every driver the same way, even if cut-through travelers have a higher negative externality on local roads than residential drivers.

Add specific equations about the Fremont game here.

Instead of increasing the cost of the market participants (the vehicles) that creates externalities, a regulator can subsidy market participants that do not create externalities. This can be done by improving the congestion across the network, to reduce the incentives for cut-through traffic. It can also be done by subsidizing more 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. The author had mainly found that decreasing the speed on cut-through roads (Pigovian taxes) and adding access restriction (cap-and-trade) have been implemented to mitigate cut-through traffic.

In order to **decrease the speed of cut-through roads**, cities can:

- Set up **stop signs** like in Fremont, CA (see fig. 2.5) [81].
- Set up **speed bumps/humps/lumps**, like in Passadena, CA [108] or partially in Fremont, CA [18].
- Reduce the **speed limits** [38]. This can be achieved by setting up low-speed alerts or low-speed zone around schools or senior apartments [160].
- Change the **timing of the signal** [59] like in Fremont too (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 in order to redirect traffic on main roads in a very grid-like traffic network by making small residential roads are to use.
- **Turn restrictions**; In Sherman Oaks' in Los Angeles, CA "no-turn" road signs have been set-up to redirect cut-through traffic [16].
- **Dynamic turn restrictions**; In Fremont, CA, variable message signs [194] 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 [100]. The city of Leonia, NJ was soon joined by Weehawken, NJ, a nearby city [85].

- **Road closure** [177]; In Baxter street, Los Angeles, CA went even further and closed one direction of the road [40] to prevent crashes from drivers following directions from navigation apps that are inept to handle the steep incline drive.

Other tools have been implemented, 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 the travel and the traffic demand. Some tools have been shown in table 2.2. However, decreasing travel or traffic demand is not something that can be done by cities alone (c.f. Fremont, CA is in the San Francisco Bay Area or Leonia, NJ is in the New York City metropolitan area). And this requires long term thinking about urban planning, which is not directly related to traffic engineering.

Pleasanton case study. Before navigational apps, the city of Pleasanton, CA was already fitting against cut-through traffic on their local roads[98]. 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. Cite.

Following the traffic management steps outlined in section 2.2, the city of Pleasanton first evaluated the state of traffic using speed surveys in 2002 [24]. The city of Pleasanton then decided to implement several traffic-calming measures, including ramp metering, decrease in the speed, and some road closure. Cite.

Later, in order to answer to each resident complains, the city launched their traffic-calming program: the City of Pleasanton Neighborhood Traffic Calming Program [48]. 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 in the Laurel Creek drive [83]. Traffic studies in 2016 showed that the cut-through drivers amount to 84% of the traffic. The city decided to implement a first set of traffic-calming measures, that lead to a 60% decrease of 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.

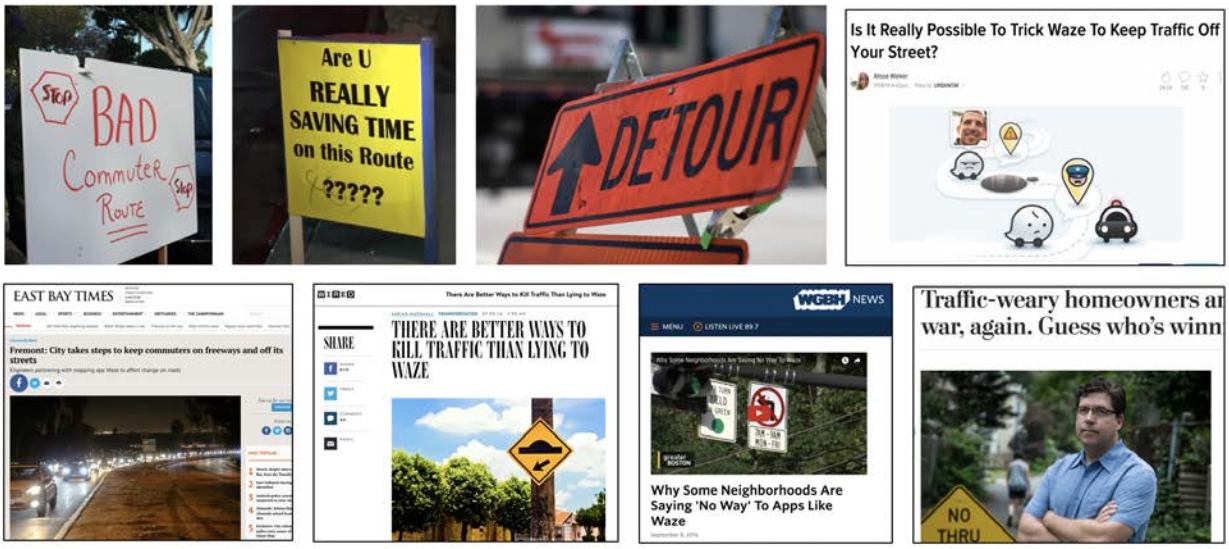


Figure 4.13: Example of mitigation strategies used by private citizens, as covered in the media. (**(A) and (B):** "shaming" signs posted by private citizens along cut-through routes; **(C):** "fake" (not legitimate) detour sign, posted on purpose to confuse through-traffic; **(D):** occurrences in the popular media of "manuals" at the disposal of citizens to *spoof* the app to prevent through traffic, for example by reporting fake accidents or recording slow moving GPS points to simulate traffic jams; **(E):** specific case of the city of Fremont, CA, in which traffic light signal timing plans were reprogrammed to minimize exit traffic from freeway to freeway through the city; **(F):** traffic slowing infrastructure built by cities to slow down through traffic (bumps, stops) in the hope of generating longer travel time (hence resulting in the app avoiding the corresponding areas); **(G):** turn restrictions put in place to limit the number of options for cut through traffic; **(H):** access restriction (prohibition of through traffic to limit the amount of motorists through the city.

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 have been posted to confuse motorists. For instance, in Sherman Oaks' neighborhood in Los Angeles, CA, "no-turn" signs were installed by residents [16].

Signs that aim to raise the awareness of commuters about cut-through traffic have been observed by [146, 143], as shown in fig. 4.13.

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 w.r.t. cyberattacks [124]. In fact, the web already contains several discussion forums on how to "spoof" apps (or "keep apps out of [your] neighborhoods"), see for example [4].

Transition.

Part II

Simulating routing behaviors

Chapter 5

Calibration and validation of Aimsun traffic microsimulation

Proofread once the previous chapter are finished.

5.1 Introduction

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 [104]. To address congestion, cities can understand the impact of traffic congestion mitigation techniques using digital twins/simulations of their traffic. 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.

Existing traffic simulator.

Work on the flow below.

As stated in section 2.2, ...

Urban road transportation network often involves large amount of active participants, wide physical area, independent agent who individually has unpredictable behavior patterns and the difference between social optimum and user optimum, etc. [135]. These listed features make the problems of traffic system extremely challenging to solve analytically [135, 7]. Therefore, simulation is an effective tool to apply to field of transportation system to help with testing, evaluating and demonstrating proposed strategies before actually implementing in physical world[135].

Building a model implies stating about the beliefs about the real world that the builders consider to be relevant to the objective of the problem [7]. Simulation models help with

predicting future steps in physical systems, under assumptions that are derived from known data.

Several types of digital twins/simulations can be constructed: land use models [192] help define zoning, activity-based models [23] grasp commutes or accessibility and attractiveness of a district, as shown in fig. 2.2, microsimulation reproduces individual driving behaviors and fine-grain traffic resolution.

Small def for each type of models.

Citation for each simulators.

- Land use model; survey [192], UrbanSim [188]
- Travel demand model: (static and dynamic)
  - Trip-based model: four step model [113], book [121]
  - Activity-based model: handbook [23], MATSim, TransCAD, Visum, CEMDAP
  - Agent-based: MATSim, extended by BEAM, ActivitySim [112] [14]
- Traffic simulators (and a lot of models)
  - Macrosimulator (static and dynamic): Aimsun, Cube, POLARIS, Visum, TransCAD.
  - Mesosimulator: Aimsun, Cube, TransModeler, POLARIS, SUMO
  - Microsimulator: Aimsun, Cube, TransModeler, POLARIS, SUMO, Vissim, MISIMLab, Synchro, BEAM, Paramics, ActivitySim.

To the author's knowledge, within the digital twin options, only microsimulations can help understand the impact of a change in traffic signal timing plans on road traffic at scale.

Work on the flow above.

Why do we choose Aimsun?

Add white paper as motivations.

Motivation

This article 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 [152] to applications for autonomous driving [43]. Existing literature provides a high-level overview and comparison of traffic simulation and its development [135, 7, 43, 152], the benefits and applications of traffic simulation [43], and an abstract framework for constructing

a simulation [99, 179]. Existing literature also includes a closer examination of simulation models with various techniques for calibration and validation of simulations under different scenarios [106, 21, 78, 130, 165, 32]. This chapter has a similar pattern of creating a microsimulation compared to [179], but it also introduces a machine learning based method for model calibration. The authors 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. Nonetheless, the authors are aware that some transportation consulting firms have internal methodologies that are not public. 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 in general.

The authors 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. 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) or mode shift analysis [23, 121, 198]. In addition, microsimulations are not relevant when data is missing to calibrate the simulation.

Outline

The author believes that 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 an example 2000-link network where traffic conditions dictate the dynamic routing of an estimated 75,000 vehicles/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[95]. Section 5.3 specifies the input data needed, and section 5.4 enumerates the steps to follow to run an Aimsun simulation. 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.

5.2 Simulation overview and its creation process

Traffic simulations [43] 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. to provide for a comprehensive analysis of the design and efficiency of the transportation system in question (section 5.6).

In this article, the authors refer to traffic simulations as the simulation of road traffic (vehicle flows in the network over time) with key inputs of traffic demand (people's origin, destination, and departure time grouped by timed origin-destination matrices) and a road

network (including road sections, lanes, intersections, road signs, and traffic signal timing plans).

Simulations can be aggregated macroscopically, mesoscopically, and microscopically [43]. Macrosimulation focuses on the aggregation of traffic flow and demand, while mesosimulation breaks traffic flow into smaller groups and examines the behavior of the whole in those groups. In this work, Aimsun Next 22 [172] 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, which is then simulated across the lanes of the input network's road sections [43]. Before being generated, each vehicle is defined by an origin, destination, departure time, and optionally, a vehicle type. The vehicle input data are aggregated across space and time into timed origin-destination (OD) matrices for each vehicle type. Space aggregation uses transportation analysis zones (TAZs), while time aggregation uses time buckets. 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 [43]. Simulated link flows and network traversal times can be compared to ground data with which to calibrate and validate a proposed model. For the example 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 [32]. With this in mind, real data set aside for model calibration should be further split into training and testing data to decrease the overfitting risk [196].

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 here 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 similarly be used to validate the simulation against ground data.

Once the simulation is created, calibrated, and validated, it can be used for analysis (section 5.6). For example, causes of congestion can be derived and policies to mitigate these can be tested.

5.3 Microsimulator input data

Chapter 3 and 6 of the Waze book.

The required inputs for a microsimulation are a network and a dataset of timed origin-destination demand. Traffic data may also be used to calibrate and validate the simulation. The data used by the authors 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[95].

Network

The road network is made up of constituent road sections connected through signalized or unsignalized intersections. To create this network, the authors downloaded the OpenStreetMap (OSM) [75] 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 [29] (see fig. 5.1). After importing the network into Aimsun [172], Google satellite, Maps, and 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 are calibrated using the data provided by the City of Fremont and road capacities are adjusted using the data from the Behavior, Energy, Autonomy, and Mobility (BEAM) model which is an open-source agent-based regional transportation model [14]. 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 (TODD) matrices. Origins (or destinations) are clustered into transportation analysis zones (TAZ), which are bijective to the set of centroids connected to

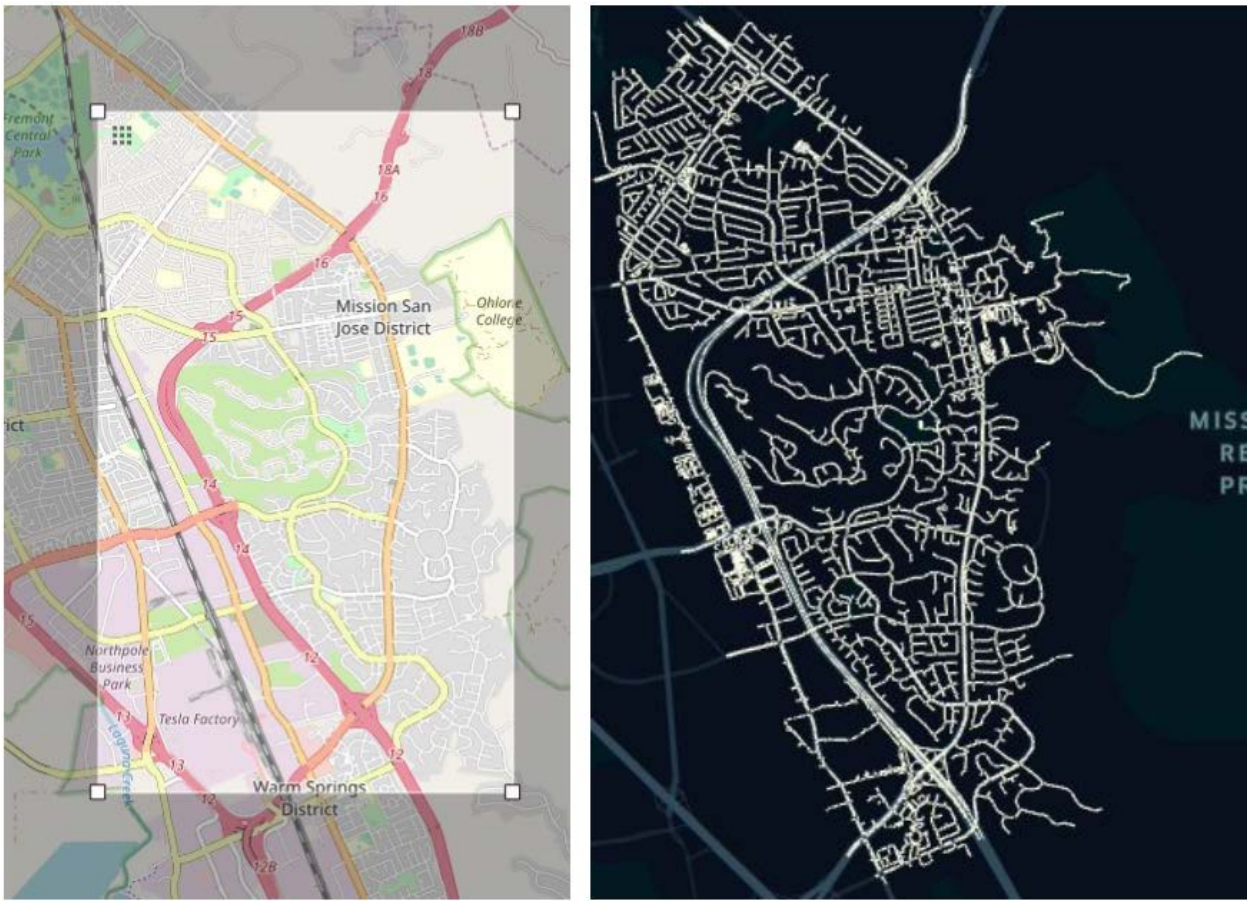


Figure 5.1: OSM network with the bounding box on the left, corresponding Aimsun network after cleaning on the right.

internal or external entry/exit points in the network. The connections between the centroids and these points (nodes) in the network are called centroid connectors [66]. 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). In this work, the demand data was derived from the SF-CHAMP demand model [86] from the San Francisco County Transportation Authority and from a StreetLight study 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.

Figure 5.2: Before and after manual editing of the OSM network in Aimsun with comparison to the Google Satellite image.

Update fig. 5.3

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 data is generated from 56 city flow detectors and 22 CalTrans Performance Measurement Systems (PeMS) detectors [186]. Speed and travel time data can be acquired using the Google Maps API[1]. 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

[1] `https://developers.google.com/maps/documentation/directions`

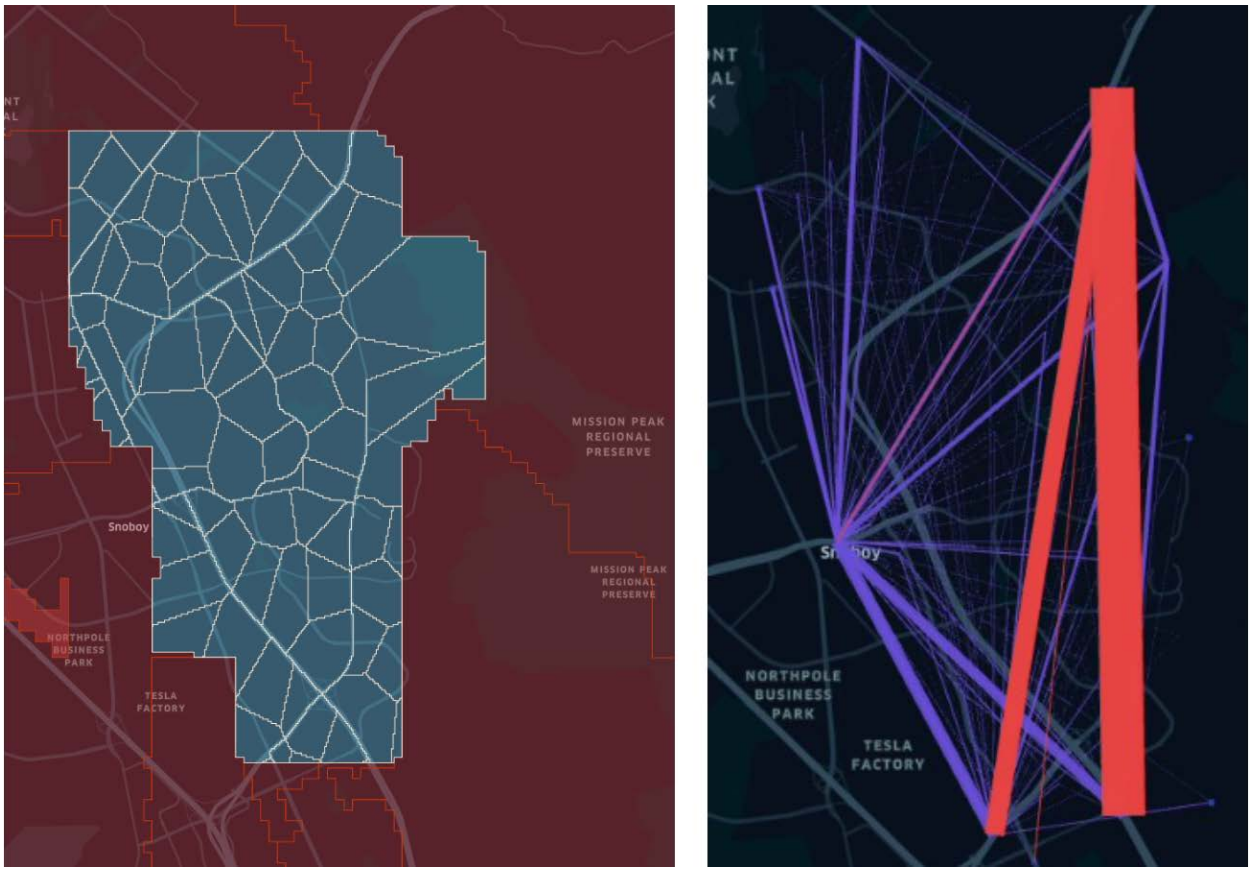


Figure 5.3: Transportation analysis zones (TAZ) (on the left) and demand plotted with desire lines (on the right). A commuter (aggregated into red lines on the right plot) is a vehicle with an origin and a destination, which are both external centroids (red TAZ on the left plot). A resident (aggregated 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).

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 authors 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 authors.

First, the network can be imported directly from OpenStreetMap (OSM) in Aimsun. For this work, the authors 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 capacity for each of the road sections can then be updated, followed by importing the ramp meters, traffic control plans, the master control plan, and traffic management strategies. 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 real flow data, the final bit of input, is then imported as a real data set inside Aimsun.

Then, a Jupyter Notebook is 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, with total runtime depending on the list of 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[2] 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

[2]Described in `https://docs.aimsun.com/next/22.0.1/UsersManual/OutputDatabaseDefinition.html`

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 followed by calibration with macrosimulation by matching all detector counts in the network. Once the OD demand is calibrated in this manner and after having chosen a route choice model, the driving behaviors are then calibrated using a genetic algorithm. The OD demand calibration and driving behavior calibration procedures can be reproduced from the provided open-source code[95].

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, or mean travel time per vehicle-mile. A second check – running the full demand – can be done by looking at gaps between the ground flow and the simulated flow that are below 50% or above 200%.

The authors used GitHub issues[3] to report, follow and solve any apparent network or demand issues. On average, such checks 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 authors found that there was one ground truth congestion issue that

[3] `https://github.com/features/issues`

could be solved by synchronizing traffic lights operated by the state with the traffic lights operated by the city on Auto-Mall Pkwy and I-680.

The authors 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 (which is 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 [73]. The objective function of the calibration, shown in eq. (5.1), was adopted from [2] and [136] to ensure that the demand was minimally modified in hopes of avoiding overfitting the flow data.

Update the equation. Add macros. $\min_\alpha F(v, \hat{v}(\alpha g))$ with $\hat{v}(g)$ the assigned flow corresponding to the demand g. α is the scaling factor that we are trying to learn.

$$\begin{aligned} \min_{g,v} F(g,v) &= \gamma_1 F_1(g,\hat{g}) + \gamma_2 F_2(v,\hat{v}) \quad v,g \geq 0 \\ \text{s.t.} \quad & v = \text{assign}(g) \end{aligned} \tag{5.1}$$

where F_1 (respectively F_2) represent the generalized distance measures between the estimated g (resp. v) and the reference $\hat{g}$ (resp. $\hat{v}$) values of OD matrices (resp. links flows); γ_1 and γ_2 are the parameters that reflect the uncertainty in the information contained in the prior OD matrix $\hat{g}$ and in the observed link flows $\hat{v}$; assign(g) is the process of assigning g, to the transportation network.

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 [133]. 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 [20].

Update the equation.

$$\begin{aligned} &\min_{g,v} F(g,v) = \tfrac{1}{2}\left[(g-\hat{g})^T X^{-1}(g-\hat{g})\right] + \tfrac{1}{2}\left[(v-P(g)g)^T Y^{-1}(v-\hat{P}(g)g)\right] \\ &\text{s.t. } g_{ij} \geq c_{ij} \end{aligned} \tag{5.2}$$

where X and Y are the variance-covariance matrices from the prior estimates and from the link flow measurements respectively, and $P(g)$ is the matrix of link choice proportions. Other notations are similar to eq. (5.1). The constraint ensures that the OD movement g_{ij}, must be greater than or equal to some lower bound c_{ij}.

This step can be done inside Aimsun directly with the static OD demand adjustment scenarios. Because of the high number of variables that can be calibrated (namely, each element of each OD demand matrix) and the relatively low volume of ground data (flow for each time step for each detector), overfitting is a major risk and must be accounted for. Therefore, the flow data is divided into training data and testing data. Then, the objective function is minimized against the training data and tested against testing data. If the demand is changed such that the training data is perfectly fitted, but the testing data is badly reproduced, then the calibration has overfitted the training data. 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 [20], formulated as the Frobenius norm [19] 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 [19] could be considered for regularization. In [114] 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 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 to the data points to draw the line of best fit [61]. 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 [61], 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.

Update fig. 5.4

Choice of the routing model

Routing calibration overleaf

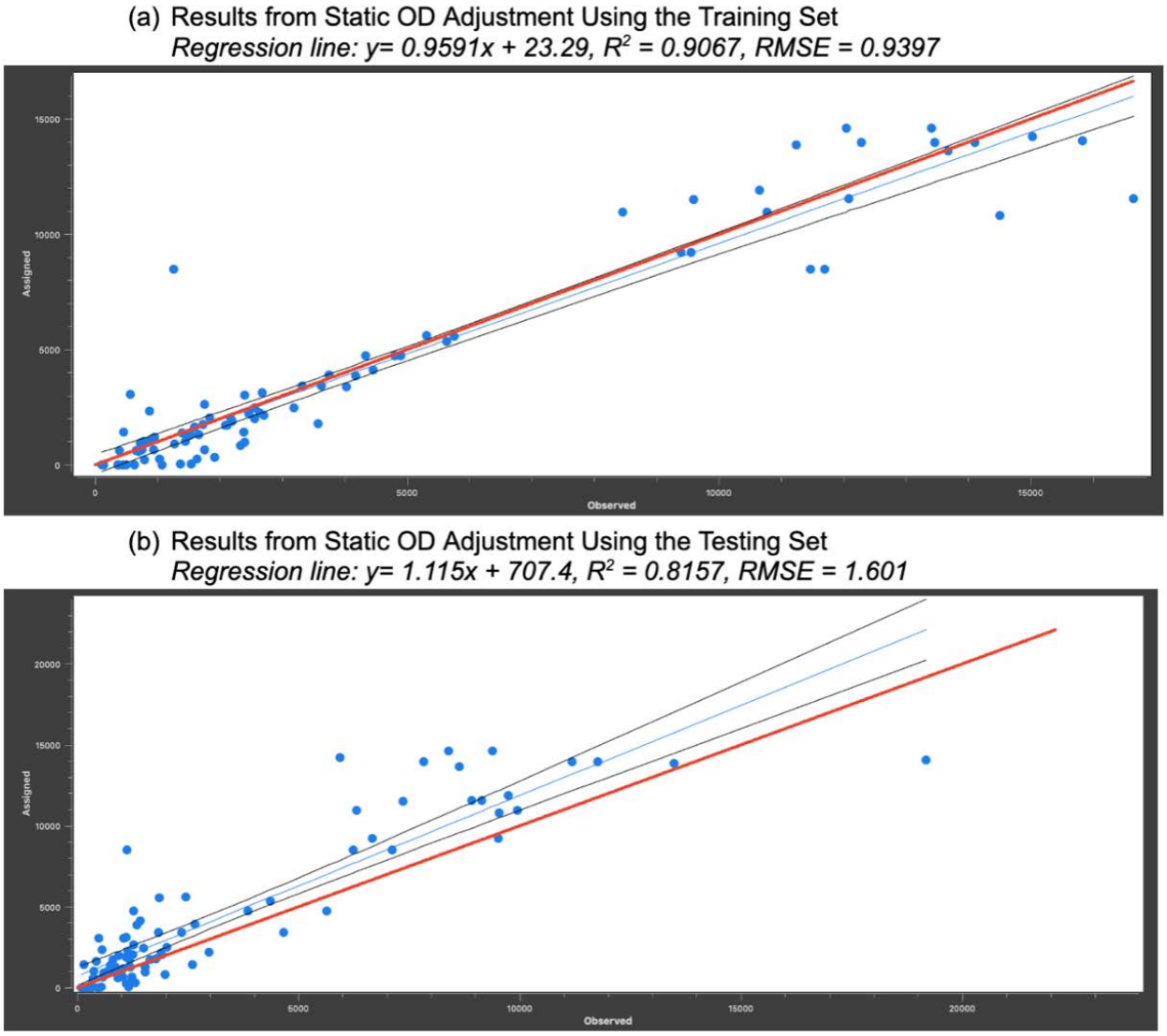


Figure 5.4: Assigned/Simulated Traffic vs. Observed/Actual Traffic Flow Bi-plot after Aimsun's default OD-demand calibration with macrosimulation. Training results are reported on the top figure, while testing results are on the bottom figure. Very good training results (slope of 0.96 and R^2 of 0.9, 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.

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 [44]:

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.

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 [44] (sometimes called Wardrop equilibrium or Nash equilibrium).

Because running many simulations iteratively takes a lot of time, the authors 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 trade-off between accuracy and simulation run time, the authors chose to use the C-logit model [39] after experimenting with the different models.

The C-logit route choice parameters include:

- Number of alternate routes considered at each routing time step.
- Size of each rerouting time interval.
- Percentage of vehicles allowed to reroute en route.
- Averaging parameters for past and current (instantaneous) travel cost parameters.
- 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

Alice's report

Once the routing choice model has been 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 rerouting time interval), 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 [95].

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 is set to be minimized (similarly to OD demand calibration in eq. (5.2)). Then, 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 [142], classical optimization algorithms that can be found in the SciPy.Optimize toolbox [187], neural-network [97], or genetic algorithm [115]. Because evaluating the objective function given the input parameters is costly, since it requires running a microsimulation, the authors 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 [115].

Genetic algorithm for microsimulation calibration

The authors followed the approach presented by [106] 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 authors' 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 have been selected by comparing the relative metric evolution over small increments of each parameter using the Latin hypercube sampling (LHS) algorithm [79]. The calibration of the 10 selected parameters was done using the Non-dominated Sorting Genetic Algorithm-III [57] from the DEAP Python library[4]. The initial value of the parameters was set to the Aimsun default ones if not updated based on intuition after few initial simulations.

Add objective function equations here.

Update or remove fig. 5.5

The overall process is presented in fig. 5.5.

[4] `https://deap.readthedocs.io/en/master/`

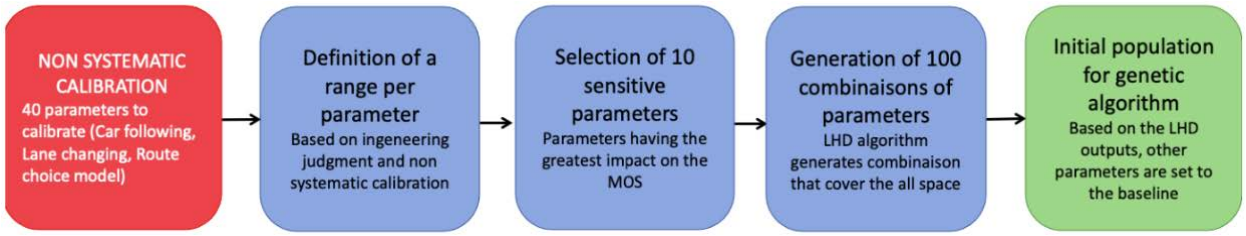


Figure 5.5: The systematic calibration process with the NDSGA-III algorithm.

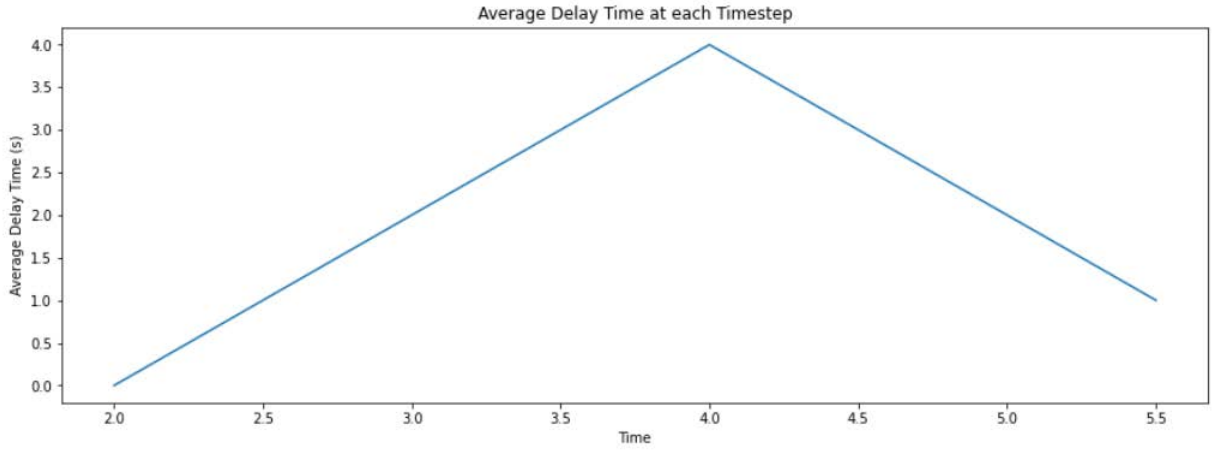


Figure 5.6: Time-series of average delay time across the entire network in Fremont, CA. We are able to identify that peak congestion occurs at 4:00PM.

5.6 Post-Processing Analysis

After the simulation is created, 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 Measure of Effectiveness (MOE) [27, 99], 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 timestep.

Update fig. 5.6.

After determining that MOE correctness is validated, 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, biplots comparing simulated versus real-world data are commonly used across previous literature [152, 10, 42]. In this work, detector flow, OD travel times, OD route distances, and system-wide metrics were used to validate the simulation results (Figure 5.7).

Update figure 5.7.

To conduct a more detailed analysis of biplots, one can create a separate biplot of system-wide data points for each time frame. Then, a time series of regression statistics for each biplot can be plotted, which gives insight into specific points in time where the simulation needs more calibration.

Update figure 5.8.

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 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. 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 biplots.

Update figure 5.9.

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 authors 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 detector within the corridor.

The authors 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 like Figure X. The trend and values observed at each timestep help determine the accuracy of each metric at each critical geographic point at the granular level.

Update figure 5.10.

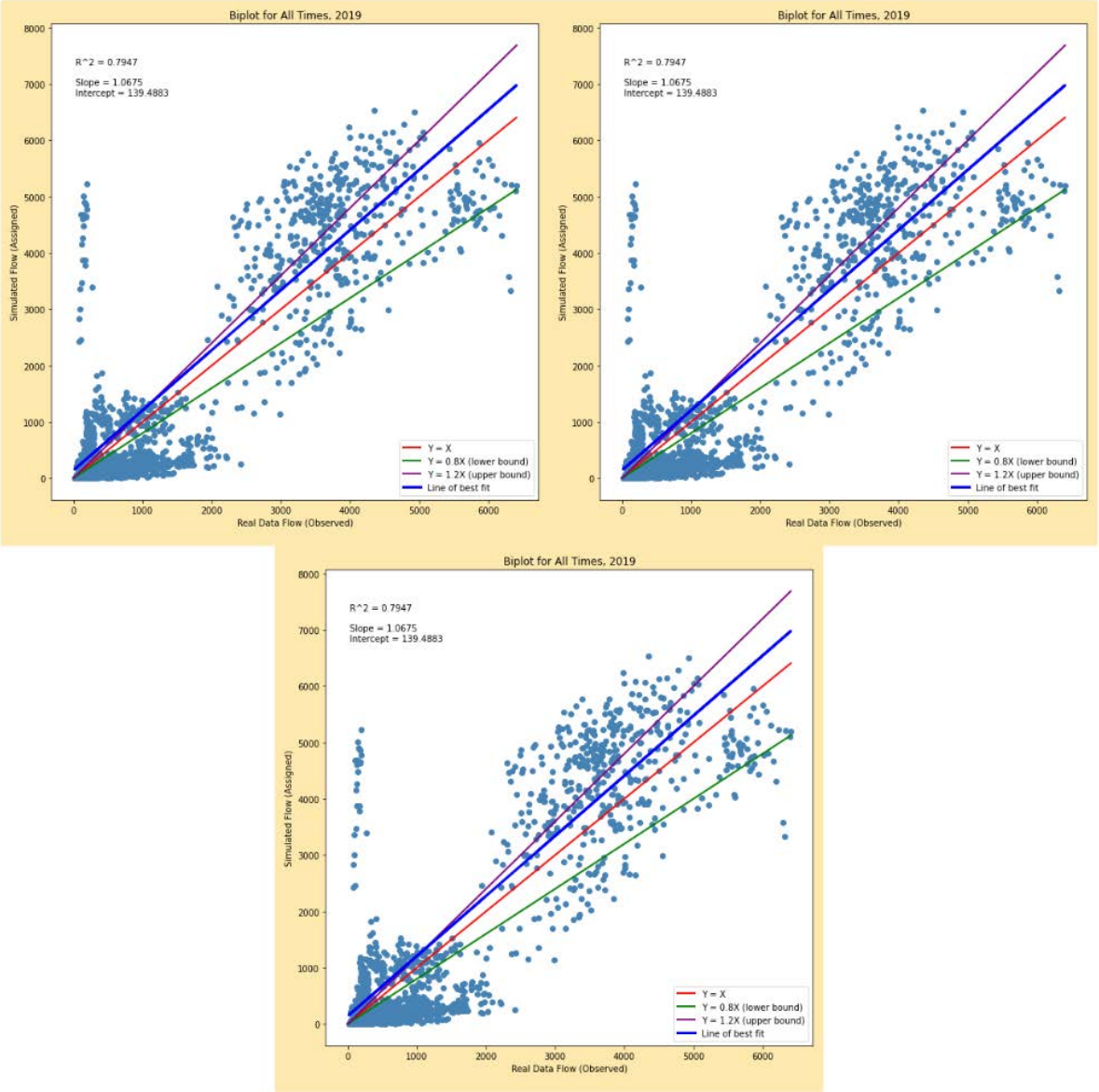


Figure 5.7: Biplot of flow (top left), OD travel times (top right), and OD route distances (center) in the entire network alongside its regression summary.

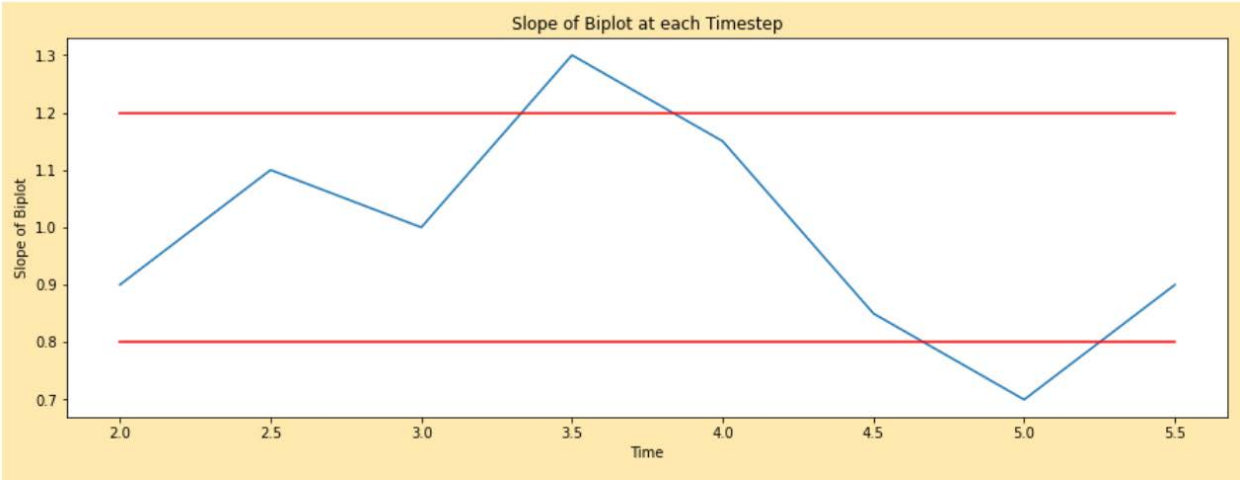


Figure 5.8: Slope of the linear regression on biplots at each timestep. Setting arbitrary lower and upper bounds of 0.8 and 1.2, respectively, shows that the simulation at 3:30PM tends to overestimate the flow while underestimating the flow at 5:00PM.

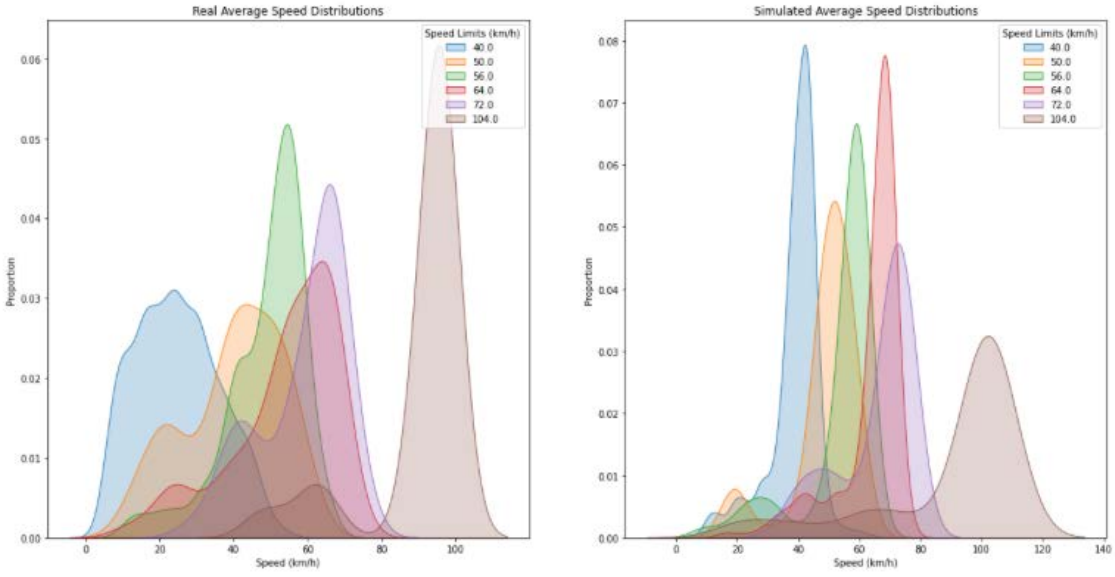


Figure 5.9: Kernel density estimation (KDE) plot of real versus simulated speeds at each road section. Each distribution is grouped by the speed limit on the road where the speed was observed.

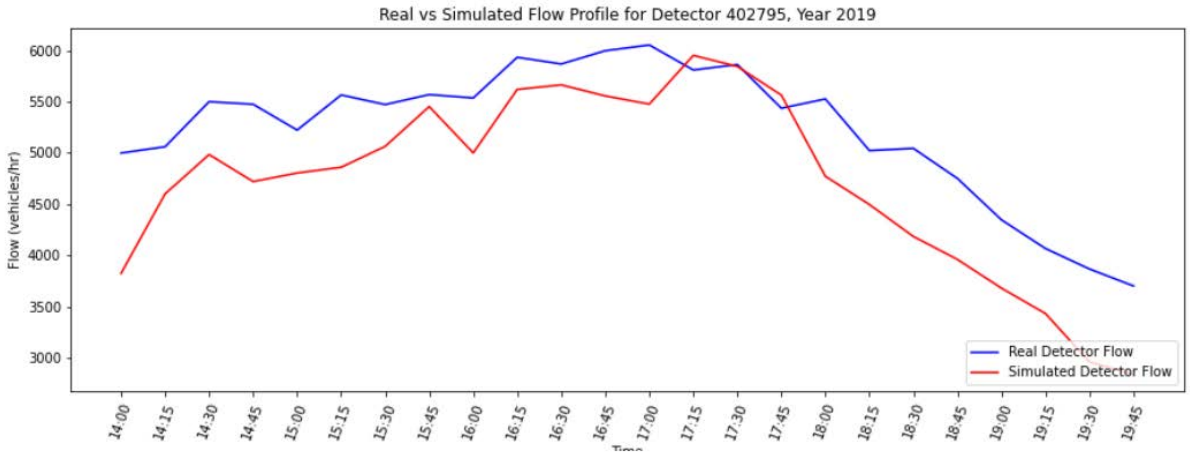


Figure 5.10: Flow profile for a detector on the I-680 South corridor. The trend of observed flow in real and simulated detectors align with each other.

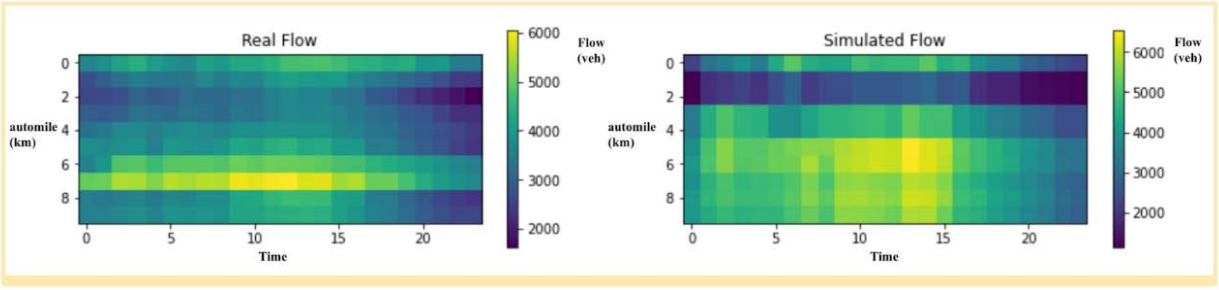


Figure 5.11: Space-time diagram of real versus simulated flow in I-680 South. Peak congestion occurs near automiles Y-Y at times X-X while least congestion occurs near automiles Y-Y at times X-X & X-X for both real and simulated flow.

Then, a time-space diagram 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 [68]. Note that on or off-ramp detectors are not included when creating a time-space diagram of detectors on a highway.

Update figure 5.11.

Analysis

Chapter 10 of the Waze book.

Here cross ref and reuse part of the work done in section 1.1 (tools).

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.

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 are as follows:

- Changing traffic signal timing plans [170].
- Changing speed limit or adding speed bumps [170].
- Adding turn and/or access restrictions [170].
- Understand the impact of increase in usage of Navigational apps on traffic (see section 3.2)
- Changing cost function that a portion of the drivers minimizes in their routing choice to understand the impact of eco-routing adoption [11, 72].
- Changing the type of some vehicles to study the impact of mixed-autonomy in traffic [195].

Table 5.1 enumerates some of the metrics that can be used to analyze the impact of such scenarios, alongside with the granularity of measurement.

The table needs formatting. A possible way to solve this is to include the table as a figure within the table environment or use https://www.tablesgenerator.com/ to create the table and generate Latex code.

5.7 Conclusion

Through the development of a traffic microsimulation of the San Jose Mission district in Fremont, CA, the authors designed and shared a reproducible process to create, calibrate and validate a large-scale microsimulation. The development of a large-scale traffic microsimulation is a tedious process that took the authors around 2,500 person-hours and it is relevant only if case study or A/B experiments cannot be performed and if enough data is available to accurately reproduce demand data. A realistic traffic microsimulation can be used to

Type	Output Statistic (Units)	Output Table
Transportation Effectiveness	Experienced Relative Gap	DUERGAP
Economic	Battery Consumption	MISYS, MICENT_O, MICENT_D, MISECT
Environmental	VOC Emissions	MISYIEM, MIODPAIIEM, MISECTIEM
Social	Safety Accessibility	

Table 5.1: Examples of metrics that can be used to analyze scenario changes. Relative Gap refers to the comparison between the current assignment solution to the ideal shortest-route time for all O-D pairs and all departure intervals [174].

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 (flow nRMSE of X), it is not good enough yet to be used off-the-shelves by the city of Fremont traffic engineers.

Move this to dissertation conclusion.

The authors envision that future research directions about traffic microsimulation should include:

- Considering physical constraints on OD demand calibration regularization terms (like physics-informed neural network [CITE], or nuclear matrix norms [CITE]).
- Building on this work to develop a standardized validation toolbox for traffic microsimulation.
- Building on this work to develop a standardized calibration toolbox for microsimulation.
- Establish clear use cases for when each type of traffic simulation are relevant and for when simulation are not relevant and case study, A/B experiment or simple models should be used instead of simulation.
- Continue developing large-scale simulation models and techniques to continue improving traffic at the metropolitan scale.

Chapter 6

Computable dynamic routing game: the mean-field routing game

Add some introduction. State still work in progress.

6.1 Dynamic N-player game

This section introduces the dynamic routing game and the corresponding MFG. We show that the MFG approach allows recovering the dynamic routing game Nash equilibrium with a very large number of vehicles. Before moving to the mathematical details, the model we propose can be summarized in the following way. This 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, and their destination. The action of a vehicle is the successor link they want to reach when exiting a given link. Actions are encoded as integers from 0 to K. Pure actions for a player on link l, with a negative waiting time, are the successors link of l. 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 MFG, the vehicles of the N-player game are replaced by a representative vehicle and the probability distribution of the vehicles states.

Network and game set-up

Time is represented as an interval $\mathcal{T} = [0, T]$ of $\mathbb{R}$. The road network is described by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ respectively denote the sets of vertices and links of the road network. When exiting a link $l \in \mathcal{E}$, a vehicle chooses one of the possible successor links. In case the link has no successor, the vehicle stays on the link until the end of time. When joining a link, a vehicle get assigned a travel time on this link, that depends on the

volume of traffic on the link. More specifically, congestion induces a travel time spent on the link $l \in \mathcal{E}$ which is a function $t_l \in \mathbb{R}_{>0}^{[0,1]}$ of the proportion of vehicles on the link l. We assume that t_l is continuous. The congestion functions $(t_l)_{l\in\mathcal{E}}$ encode the heterogeneity of the roads' sensitivity to traffic volume within the network. The following typical congestion function is based on an example provided by the 1964 traffic assignment manual of the U.S. Bureau of public road functions, see [133, table 1.1]: $t_l : \mu \mapsto t_0(1 + \alpha(\mu/\mu_{l,c})^\beta)$, where α and β are positive constants, t_0 is the free flow travel time (i.e., the travel time when the link is empty), and $\mu_{l,c}$ is the relative capacity of the link l (which, in our context, is to be understood as a capacity in terms of proportion of players).

Let us stress that the congestion functions are functions of the *proportion* of vehicles within the link. However, in practice, congestion effects scale with the *number* of vehicles (as well as other factors such as the width and the length of the road, which are considered fixed for a given network). One should thus interpret t_l as being tuned for a given number of vehicles, say N_0. Concretely, in the previous example, if n vehicles out of N_0 are on link ℓ, we have $t_l(n/N_0) = t_0(1 + \alpha(n/(N_0\mu_{l,c}))^\beta)$. For $N_0' \neq N_0$, we have $t_l(n/N_0') = t_0(1 + \alpha(n/(N_0'\mu_{l,c}))^\beta)$. So the travel time can be viewed as a function of the number of vehicles n on the link l, provided the relative capacity is scaled by the total number of vehicles.

Traffic flow environment

For the sake of convenience, we assume that the time horizon T is large enough so that any driver will have time to travel through the network.

Let N be a positive integer. The set of players is $\mathcal{N} = \{1, 2, \dots, N\}$. The number of players in the game is not necessarily the total number of vehicles N_0 in the real-life scenario. 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 MFG. In the case where $N = N_0$, a player is a vehicle. A player $i \in \mathcal{N}$ starts at an origin link $L_0^i \in \mathcal{E}$ with a departure time $W_0^i \in \mathcal{T}$, and has a destination link $D_0^i \in \mathcal{E}$. This is the initial state of the player. Intuitively, the player wants to start moving at time W_0^i from L_0^i and tries to reach D_0^i. We assume that the players' initial state are distributed according to a finite-support distribution m_0. 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 their trip.

Then, at any time step t, the state of a player i is not only the link L_t^i where they stand, but also their waiting time W_t^i before exiting this link together with their destination D_t^i. 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 driver is $\mathcal{X} = \mathcal{E} \times \mathcal{T} \times \mathcal{E}$, 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 link 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 0. The players are allowed to randomize their actions. We thus call strategy function and denote by π a function from $\mathcal{X}$ to $\mathcal{P}(\mathcal{E})$ such that for any $x = (l, w, d)$, $\pi(x)$ has support in the successors of l. We denote by Π the set of such strategy functions. A (feedback or closed-loop) policy $\boldsymbol{\pi}$ is a function that associates to each time a strategy function, so it is an element of the set $\boldsymbol{\Pi} = \Pi^{\mathcal{T}}$ of policies. 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. A policy profile $\underline{\boldsymbol{\pi}}$ is a vector of policy functions with one policy for each player, i.e., it is an element of $\underline{\boldsymbol{\Pi}} = \boldsymbol{\Pi}^{\mathcal{N}}$. Studying this class of policies can be justified by the fact that it allows each player to take a decision based only on their own state, which is realistic if the players do not know the situation of the rest of the population. More information could be added in the inputs of the policy (e.g., the proportion of agents on the current link), but this is beyond the scope of this work.

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 $\underline{\boldsymbol{X}} = (\underline{X}_t)_{t\in\mathcal{T}} \in \underline{\boldsymbol{\mathcal{X}}}$. As indicated above, the players' interactions are through the travel time functions $(t_l)_{l\in\mathcal{E}}$, taking into account congestion levels. So the interaction between a driver and the rest of the vehicles is only through the proportion of vehicles on the same link. It is thus convenient to introduce the empirical distribution $\nu^N_{\underline{l}} \in \mathcal{P}(\mathcal{E})$ corresponding to a location profile $\underline{l} = (l^i)_{i\in\mathcal{N}} \in \mathcal{E}^N$: for every $l \in \mathcal{E}$, $\nu^N_{\underline{l}}(l') = \frac{1}{N}\#\{i \,|\, l^i = l'\} \in [0, 1]$, which is the proportion of players on the link l', given $\underline{l}$. This is all the information one needs from the game state to compute the interactions between players at link l'. Note that $\nu^N_{\underline{l}}(l')$ 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 $\mathcal{E}^{\mathcal{T}\times\mathcal{X}\times\mathcal{N}}$-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 $\mathcal{E}$-valued random variable with distribution $\pi^i_t(\cdot|x)$.

The evolution of the state of the game $\underline{\boldsymbol{X}}_t = (\underline{L}_t, \underline{W}_t, \underline{D}_t)$ 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:

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

where $I_{t_{k+1}} := \{i \in \mathcal{N},\ W_{t_k}^i + t_k - t_{k+1} = 0\}$ and using $(t_k)_{k\in\mathbb{N}}$ the sequence of times when one of the vehicles changes link with $t_0 = 0$ and $t + k = T$ 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 player i is the average arrival time which can be defined as:

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

where $\boldsymbol{\pi}^{-i} = (\boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^{i-1}, \boldsymbol{\pi}^{i+1}, \ldots, \boldsymbol{\pi}^N)$, and the instantaneous cost is defined as: for every $x = (l, w, d)$, $r(x) = \nexists_{l\neq d}$. 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.

Furthermore, the population is homogeneous (all players have the same dynamics evolution and same running cost), and the player i interacts with the other players only through $\boldsymbol{\nu}^N$ and for this reason, the cost function J_i^N does not depend directly on the index i but only on $\boldsymbol{\pi}^i$: as a function, $J_i^N = J_{i'}^N$ for all i'. The policy profile $\boldsymbol{\pi}^{-i}$ for the rest of the population is used only to compute $\boldsymbol{\nu}^N = (\nu_t^N)_{t\in\mathcal{T}}$. Although $\boldsymbol{\pi}^{-i}$ is necessary, it is not sufficient because $\boldsymbol{\nu}^N$ 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, which will be the basis for the mean-field approach presented in section 6.2.

Nash equilibrium

Considering that all the players are individually optimizing their own cost leads to the following notion of solution for the game. We refer to e.g. [117] for more details.

Definition 6.1.1 (Nash equilibrium)**.** *A Nash equilibrium is a policy profile* $\underline{\boldsymbol{\pi}}^\star = (\boldsymbol{\pi}^{i\star})_{i\in\mathcal{N}} \in \underline{\boldsymbol{\Pi}}$ *such that:*

$$\forall i \in \mathcal{N},\ \forall \boldsymbol{\pi} \in \boldsymbol{\Pi},\ J_i^N(\boldsymbol{\pi}^{i\star}, \boldsymbol{\pi}^{-i\star}) \leq J_i^N(\boldsymbol{\pi}, \boldsymbol{\pi}^{-i\star}).$$

The following result says that, in our model, such equilibria exist.

Theorem 6.1.1 (Existence of N-player Nash equilibrium, Kakutani-Fan-Glisckberg theorem [70])**.** *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. The proportion of players on each link is always a multiple of $1/N$. Since the number of links and the time horizon are finite, there is a finite set of times at which a vehicle can switch link. The set of policy profiles can thus be restricted to a finite set. Therefore, the game can be restated as a game with finite state and action spaces. Assuming the continuity of the cost function with respect to the policy, it has a Nash equilibrium. Further development of the proof are provided in the appendix. □

Put the full proof here.

Besides the above definition, another way to express that a policy profile $\underline{\boldsymbol{\pi}}$ is a Nash equilibrium is to say that the deviation incentive is 0 for every player, where the deviation incentive for player i is:

$$D_i^N(\boldsymbol{\pi}^i, \boldsymbol{\pi}^{-i}) = J_i^N(\boldsymbol{\pi}^i, \boldsymbol{\pi}^{-i}) - \operatorname*{argmin}_{\boldsymbol{\pi}'\in\boldsymbol{\Pi}} J_i^N(\boldsymbol{\pi}', \boldsymbol{\pi}^{-i}).$$

This also serves as a basis to assess the convergence of algorithms towards a Nash equilibrium using the average deviation incentive:

$$\overline{D}^N(\underline{\boldsymbol{\pi}}) = \frac{1}{N}\sum_{i=1}^{N} D_i^N(\boldsymbol{\pi}^i, \boldsymbol{\pi}^{-i}). \tag{6.1}$$

6.2 Mean field routing game

Set-up

To do.

Nash equilibrium

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

Definition 6.2.1 (Mean field Nash equilibrium (Definition 3.1. of [147]))**.** *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:*

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

Another way to express that $\boldsymbol{\pi} \in \boldsymbol{\Pi}$ is a MFNE is to say that the average deviation incentive $\overline{D}$ vanishes, where:

$$\overline{D}(\boldsymbol{\pi}) = J(\boldsymbol{\pi}, \boldsymbol{\pi}) - \underset{\boldsymbol{\pi}' \in \boldsymbol{\Pi}}{\operatorname{argmin}}\, J(\boldsymbol{\pi}', \boldsymbol{\pi}).$$

Theorem 6.2.1 (Existence of mean-field Nash equilibrium, Kakutani-Fan-Glisckberg theorem [70])**.** *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. The set of pure policies for the representative player can be restricted to the choice of a path given a departure time. This set is finite as long as the support of the initial distribution of waiting time is. Therefore, the argmin map (also called Best response map) is a Kakutani-Fan-Glisckberg map providing the continuity of the cost with respect to the policy of the representative player and of the mean-field. The proof is further developed in appendix. □

Put the full proof here.

The continuity of the cost function with respect to the policies plays a crucial role. A counter-example of the existence of a Nash equilibrium with a discontinuous cost is shown in the appendix.

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

Theorem 6.2.2 (Dynamic Wardrop equilibrium [190])**.** *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. In case a trajectory used by the representative player has a higher travel time than another one, then the player has an incentive to deviate, and the game is not a Nash equilibrium. The proof is further developed in the appendix. □

Put the full proof here.

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 $\underline{\boldsymbol{\pi}}^\star = (\boldsymbol{\pi}^\star, \dots, \boldsymbol{\pi}^\star) \in \boldsymbol{\Pi}^N$ is an approximate Nash equilibrium whose quality improves with N in the sense that:

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

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, [147] prove in their setting that: if $\boldsymbol{\pi}^\star$ is an 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 $(\boldsymbol{\pi}^\star, \dots, \boldsymbol{\pi}^\star) \in \boldsymbol{\Pi}^N$ satisfies: $\overline{D}^N(\underline{\boldsymbol{\pi}}) \leq \epsilon$.

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 and Braess 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 graph has 2 nodes and 2 parallel links, say $l, \tilde{l}$, relating these 2 nodes. The cost function is: $t_l(x) = 0.5T$, $t_{\tilde{l}}(x) = xT$ for all $x \in [0.1]$. 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 $[\frac{1}{2} - \frac{1}{N}, \frac{1}{2}]$. As detailed in the appendix, we can check that the average deviation incentive of the mean-field equilibrium policy in the N-player game is

$$\frac{T}{N2^N} \sum_{m=1}^{N} \binom{N-1}{m} \max\left\{\frac{N}{2} - m - 1, m + 1 - \frac{N}{2}\right\},$$

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

6.3 Experiments

Experiments show ...

This section shows experimentally 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 [199]) and (2) it gives an excellent approximation of the N-player equilibrium when N is large (above 30 in the case of the Pigou [137] and the Braess [31] network). The experiments also show that (3) online mirror descent algorithm [134] enables computing the mean-field equilibrium on the Sioux Falls network, a classic

use case in road traffic network games, with 14,000 vehicles (across two origin-destination pairs) and realistic congestion function (from the open source dataset [1]).

Context

All the experiments are conducted within the OpenSpiel framework [96], 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[1].

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, showing that 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 similar 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 [22]. In the MFG, heterogeneity between the players is encoded in the state, to use the same policy for each player without a loss a generality. As an example, in our model, the destination of the player is represented in the state.

Metrics.

X-REF chapter 2.

The quality of the approximation of the Nash equilibrium policy completed by the candidate policy is measured using the average deviation incentive defined in (6.1) (also known as the average marginal regret, or the relative gap to the dynamic user equilibrium [44] in traffic engineering).

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 [77] and model-based algorithms such as Counterfactual Regret Minimization (and some variants) [199] which we use to solve the N-player game. The experiments solve the MFG using the online mirror descent algorithm [134]. The experiments performed in OpenSpiel use a fixed time discretisation.

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 [137], Braess [31] and Sioux Falls). This enables defining vehicle location only using links, and defining state of not having begun a trip and having finished it.

[1] `https://github.com/deepmind/open_spiel`

The *Pigou network* [137] 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. 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 [31]. 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 network 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]. As the classical routing game [125, Chapter 18] is a static game, the demand is only a list of tuple origin, destination 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 on 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 6.2.2.

While solving N-player game is intractable for large number of 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

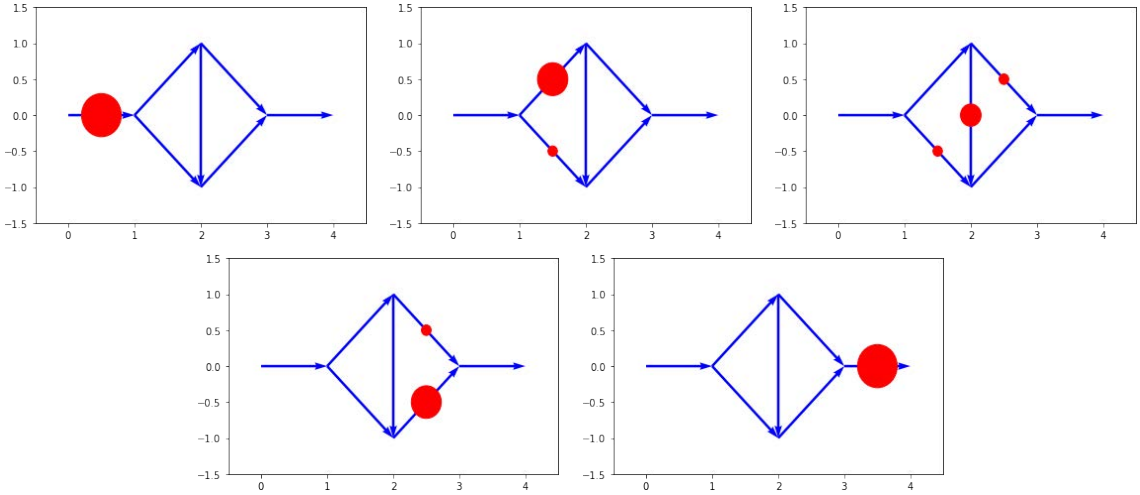


Figure 6.1: The dynamic of the Braess network in the mean-field Nash equilibrium; the locations of the cars at time 0.0 (Figure 6.3), from time 0.25 to 1.75 (Figure 6.3), at time 2.0 (Figure 6.3), from time 2.25 to 3.75 (Figure 6.3), at time 4.0 (Figure 6.3). The travel time on each path are equal to 3.75, travel time equalization defines the mean-field Nash equilibrium.

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. Comparison between the running time of 10 iterations of ext CFR and OMD are done as a function of the number of vehicles modeled in fig. 6.2. 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 with the algorithms of the OpenSpiel library.

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, see fig. 6.3. This was shown theoretically in section 6.2, 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.

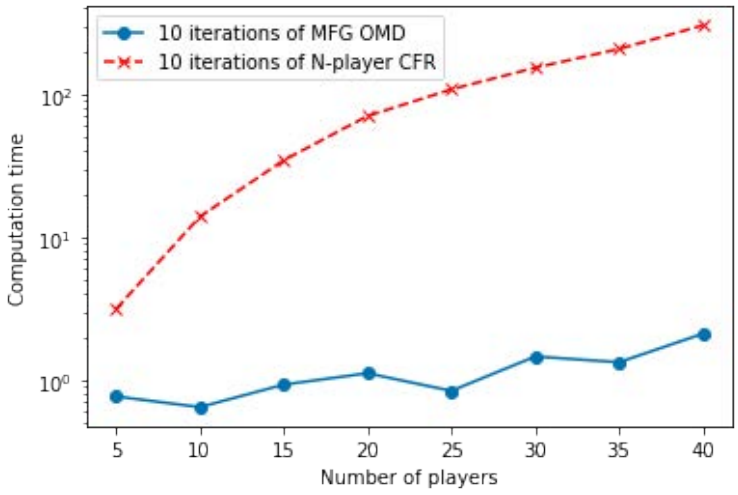


Figure 6.2: Computation time of 10 iterations of Online Mirror Descent in the MFG and of 10 iterations of sampled Counterfactual regret minimization as a function of the number of players N.

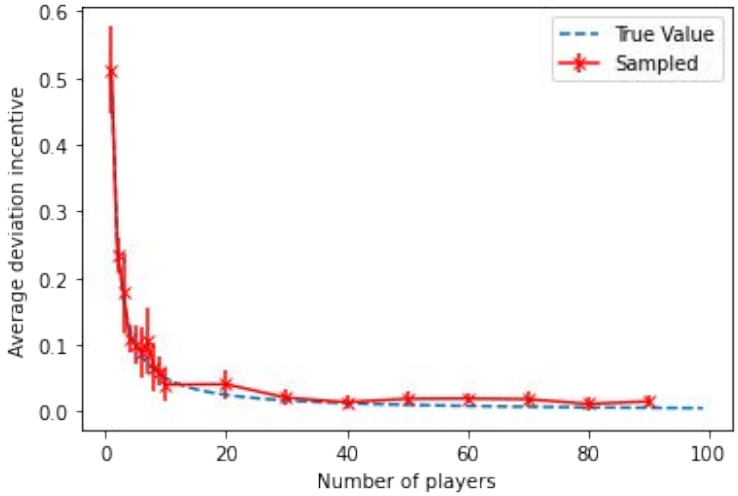


Figure 6.3: Average deviation incentive of the Nash equilibrium mean-field policy in the N-player game as a function of N in the case of the Pigou game. The sampled value is the value computed in OpenSpiel by testing all the possible pure best responses, and sampling game trajectories to get the expected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

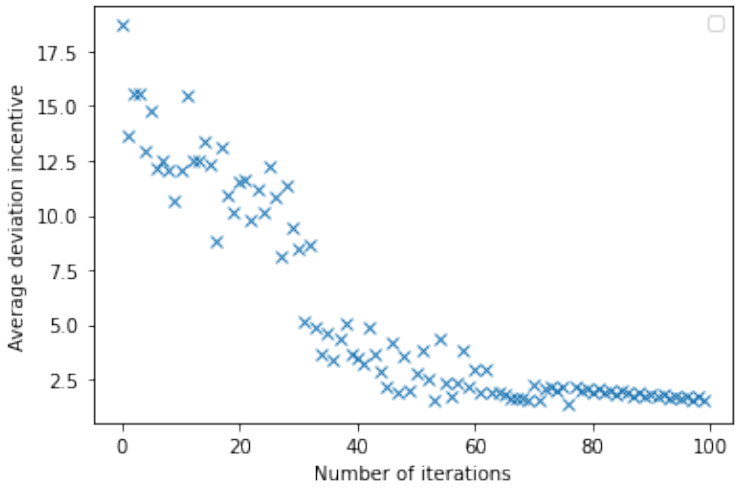


Figure 6.4: Online mirror descent average deviation incentive in the Sioux Falls MFG as a function of the number of iterations of the descent algorithm.

The experiment shows the ability to learn the mean-field equilibrium policy on this 76 links 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 6.3) 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.

If more time, add next steps.

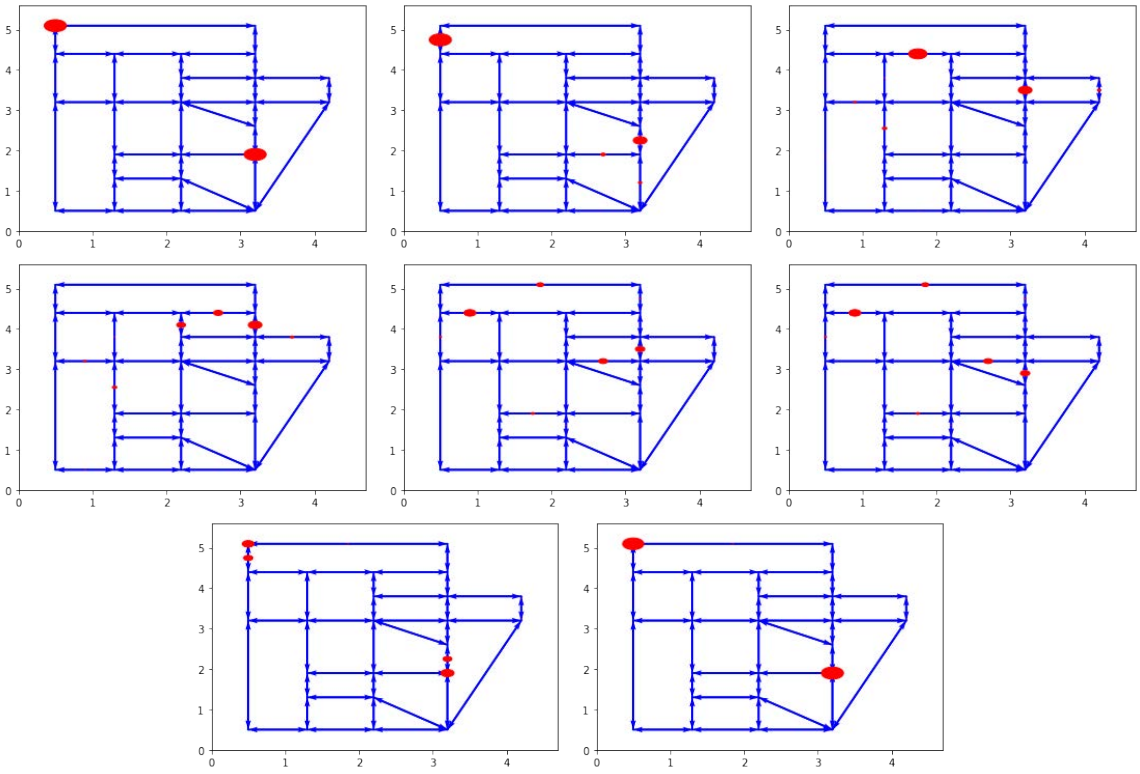


Figure 6.5: Dynamics of the Sioux Falls network in the mean-field Nash equilibrium. Road network; location of the cars at time 0.0 (6.3), 2.5 (6.3), 10.0 (6.3), 12.5 (6.3), 21.0 (6.3), 22.0 (6.3), 26.5 (6.3), 27.5 (6.3). Some vehicles arrived at their destination after some that left the origin at the 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.

Conclusion

During more than 5 years of collaboration with the mobile sensing lab, I was able to:

-

This journey helped me envision that the future of transportation research should be focusing on:

-

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] T. Abrahamsson. *Estimation of Origin-Destination Matrices Using Traffic Counts - A Literature Survey*. IIASA Interim Report. IIASA, Laxenburg, Austria, May 1998. URL: http://pure.iiasa.ac.at/id/eprint/5627/.

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

[4] 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).

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

[6] 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.

[7] 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.

[8] 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).

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

[10] 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.

[11] 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).

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

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

[18] 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).

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

[20] 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.

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

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

[23] 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.

[24] 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).

[25] Luara 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).

[26] 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.

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

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

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

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

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

[32] 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.

[33] 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).

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

[35] 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.

[36] 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.

[37] 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).

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

[39] 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.

[40] 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).

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

[42] 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.

[43] 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.

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

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

[46] 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=.

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

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

[49] City of St. Louis. "Report of the Transportation Survey Commission of the City of St. Louis". In: (1930).

[50] Civilengineeringbible.com. *What is Transportation Engineering?* URL: https://civilengineeringbible.com/article.php?i=113 (visited on 09/28/2018).

[51] 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).

[52] R. H. Coase. "The Problem of Social Cost". In: *The Journal of Law & Economics* (1960). URL: www.jstor.org/stable/724810.

[53] Serdar *c*Colak, Antonio Lima, and Marta C González. "Understanding congested travel in urban areas". In: *Nature communications* 7 (2016), p. 10793.

[54] 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.

[55] Adolf D.May. *The Highway Congestion Problem and the Role of In-Vehicle Information Systems*. 1989.

[56] Stella Dafermos and Anna Nagurney. "Sensitivity analysis for the asymmetric network equilibrium problem". In: *Mathematical programming* 28.2 (1984), pp. 174–184.

[57] 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.

[58] Peter A Diamond. "Consumption externalities and imperfect corrective pricing". In: *The Bell Journal of Economics and Management Science* (1973), pp. 526–538.

[59] 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.

[60] Anthony Downs. "The law of peak-hour expressway congestion". In: *Traffic Quarterly* 16.3 (1962).

[61] Norman R Draper and Harry Smith. *Applied regression analysis*. Vol. 326. John Wiley & Sons, 1998.

[62] 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.

[63] 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.

[64] Mogens Fosgerau. "The valuation of travel-time variability". In: *Internatioanl Transport Forum* (2017), pp. 39–56. DOI: 10.1787/9789282108093-3-en.

[65] 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).

[66] Markus Friedrich and Manuel Galster. "Methods for generating connectors in transport planning models". In: *Transportation Research Record* 2132 (2009), pp. 133–142. ISSN: 03611981. DOI: 10.3141/2132-15.

[67] 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.

[68] NJ Garber and LA Hoel. "Fundamental Principles of Traffic Flow". In: *Traffic and highway engineering*. 4th Ed. Cengage Learning, 2009, pp. 213–214.

[69] 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.

[70] 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.

[71] Mark Goh. "Congestion management and electronic road pricing in Singapore". In: *Journal of transport geography* 10.1 (2002), pp. 29–38.

[72] Google Maps. *Google Maps Eco-Friendly Routing: How it works*. 2021. URL: https://www.gstatic.com/gumdrop/sustainability/google-maps-eco-friendly-routing.pdf (visited on 07/09/2022).

[73] 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.

[74] 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.

[75] Mordechai Haklay and Patrick Weber. "Openstreetmap: User-generated street maps". In: *IEEE Pervasive computing* 7.4 (2008), pp. 12–18.

[76] 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.

[77] Johannes Heinrich and David Silver. "Deep reinforcement learning from self-play in imperfect-information games". In: *arXiv preprint arXiv:1603.01121* (2016).

[78] 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.

[79] DE Huntington and CS Lyrintzis. "Improvements to and limitations of Latin hypercube sampling". In: *Probabilistic engineering mechanics* 13.4 (1998), pp. 245–253.

[80] DRIEAT Île-de-France. "Enquête globale de transport". In: *Enquêtes transport et déplacements* (2019).

[81] 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).

[82] 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).

[83] 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).

[84] 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).

[85] 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).

[86] 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.

[87] 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.

[88] Aman Kishore et al. "Prevention of Highway Infrastructure Damage Through Commercial Vehicle Weight Enforcement Annual Indian Roads Congress (Irc) Session". In: 306 (2000).

[89] 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.

[90] Frank H Knight. "Some fallacies in the interpretation of social cost". In: *The Quarterly Journal of Economics* 38 (4 1924), pp. 582–606.

[91] 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.

[92] 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.

[93] 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).

[94] Mobile Sensing lab. *Open source static traffic assignment solver*. https://github.com/megacell. Accessed: 2018-09-25.

[95] Mobile Sensing lab. *Traffic microsimulation toolbox*. https://github.com/Fremont-project/traffic-microsimulation. Accessed: 2022-07-15.

[96] 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.

[97] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), pp. 436–444.

[98] 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).

[99] Edward Lieberman and Ajay K. Rathi. "Traffic simulation". In: *Federal Highway Administration* (1992). ISSN: 0097-8418. DOI: 10.1145/274790.273160.

[100] 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).

[101] 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).

[102] Todd Alexander Litman. *Transportation Cost and Benefit Analysis: Techniques, Estimates and Implications*. Victoria Transport Policy Institute, 2009.

[103] William Forster Lloyd. *Two lectures on the checks to population. England: Oxford University*. Oxford University, 1833.

[104] Tim Lomax, David Schrank, Bill Eisele, et al. "2021 urban mobility report". In: (2021).

[105] 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.

[106] 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.

[107] 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.

[108] 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/.

[109] Alfred Marshall. *Principles of Economics*. Great Mind Series, 1890.

[110] Karl Marx. *Capital. A Critique of Political Economy*. Verlag von Otto Meisner, 1867.

[111] Karl Marx and Friedrich Engels. *The Communist Manifesto*. London: Communistischer Arbeiterbildungsverein, 1848.

[112] 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.

[113] Michael G McNally. "The four-step model". In: *Handbook of transport modelling*. Emerald Group Publishing Limited, 2007.

[114] 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.

[115] Seyedali Mirjalili. "Genetic algorithm". In: *Evolutionary algorithms and neural networks*. Springer, 2019, pp. 43–55.

[116] Dov Monderer and Lloyd S Shapley. "Potential games". In: *Games and economic behavior* 14.1 (1996), pp. 124–143.

[117] Roger B Myerson. *Game theory*. Harvard university press, 2013.

[118] 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.

[119] John Nash. "Equilibrium points in n-person games". In: *Proceedings of the National Academy of Sciences* 36 (1 1950), pp. 48–49.

[120] National Academies of Sciences, Engineering, and Medicine and others. "Highway Capacity Manual". In: (2016).

[121] National Academies of Sciences, Engineering, and Medicine and others. "Travel demand forecasting: Parameters and techniques". In: (2012).

[122] 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.

[123] 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.

[124] 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).

[125] N. Nisam et al. *Algorithmic game theory*. Cambridge University Press, 2007.

[126] 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.

[127] 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).

[128] Jennifer Olney. *Bay Area transportation experts say apps may make traffic worse*. Nov. 2018.

[129] 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/.

[130] 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.

[131] 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).

[132] 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.

[133] Michael. Patriksson. *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015. ISBN: 0486787907.

[134] Julien Perolat et al. "Scaling up Mean Field Games with Online Mirror Descent". In: *arXiv preprint arXiv:2103.00623* (2021).

[135] 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.

[136] Anders Peterson. *Origin-destination matrix estimation from traffic counts*. 2003.

[137] Arthur Cecil Pigou. *The Economics of Welfare*. 1920. DOI: 10.4337/9781788118569.00060.

[138] 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.

[139] 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.

[140] 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).

[141] 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).

[142] LA Rastrigin. "Random search as a method for optimization and adaptation". In: *Stochastic Optimization*. Springer, 1986, pp. 534–544.

[143] Gary Richards. *Is Google's Waze app making traffic worse?* Jan. 2016.

[144] 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.

[145] 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.

[146] 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).

[147] Naci Saldi, Tamer Basar, 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.

[148] 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.

[149] 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.

[150] 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.

[151] David Schrank, Bill Eisele, and Tim Lomax. "2019 Urban Mobility Report". In: *Texas A&M Transportation Institute* August (2019).

[152] 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.

[153] Florian Siebel and Wolfram Mauser. "On the fundamental diagram of traffic flow". In: *SIAM Journal on Applied Mathematics* 66.4 (2006), pp. 1150–1162.

[154] Adam Smith. *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations*. W. Strahan and T. Cadell, 1776.

[155] We are social. *Digital in 2019*. 2019. URL: https://wearesocial.com/uk/digital-2019 (visited on 08/11/2019).

[156] Statista. *Global digital population as of July 2019*. 2019. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ (visited on 08/11/2019).

[157] 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).

[158] 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).

[159] Statista. *Ride Hailing; worldwide*. 2019. URL: https://www.statista.com/outlook/368/100/ride-hailing/worldwide (visited on 08/06/2019).

[160] 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).

[161] 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).

[162] 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/.

[163] 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).

[164] 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).

[165] 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.

[166] 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.

[167] 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.

[168] 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.

[169] 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.

[170] 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=.

[171] 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).

[172] Transport Simulation Software. *Aimsun Next 22*.

[173] 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.

[174] 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.

[175] U.S. Department of Transportation Federal Highway Administration. *Highway Functional Classification: Concepts Criteria, and Procedure*. 2013. URL: https://trid.trb.org/view/1266295.

[176] 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).

[177] 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).

[178] 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.

[179] 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).

[180] U.S. Department of Transportation Federal Highway Administration, Office of Operations. *Overview of Active Transportation and Demand Management*. 2018. URL: https://ops.fhwa.dot.gov/atdm/about/overview.htm.

[181] 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).

[182] 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).

[183] US Department of Transportation. *Manual on Uniform Traffic Control Devices; for Streets and Highways*. US Department of Transportation, Federal Highway Administration., 2009.

[184] US Department Of Transportation Federal Highway Administration. *Next Generation SIMulation (NGSIM)*. 2005. URL: www.ngsim-community.org..

[185] Pravin Varaiya. *Freeway Performance Measurement System (PeMS), PeMS 6*. Tech. rep. California Center for Innovative Transportation, University of California, Berkeley, Feb. 2006.

[186] Pravin Varaiya. "Freeway Performance Measurement System (PeMS), Version 4". In: UCB-ITS-PR.April (2004).

[187] 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.

[188] 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.

[189] 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.

[190] 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.

[191] Waze Connected Citizen Program. *Case Studies*. 2019. URL: https://www.waze.com/en/ccp/casestudies (visited on 08/14/2019).

[192] 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.

[193] Wiki Waze. *Routing server*. 2019. URL: `https://wiki.waze.com/wiki/Routing_server` (visited on 06/13/2019).

[194] 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`.

[195] Cathy Wu et al. "Flow: Architecture and benchmarking for reinforcement learning in traffic control". In: *arXiv preprint arXiv:1710.05465* 10 (2017).

[196] 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.

[197] Naomi Zeveloff. "Israelis Sue Waze Navigation App for Creating Neighborhood Traffic Jam." In: *Forward* (Dec. 2016).

[198] Hong Zheng et al. "A primer for agent-based simulation and modeling in transportation applications". In: (2013).

[199] Martin Zinkevich et al. "Regret minimization in games with incomplete information". In: *Advances in neural information processing systems* 20 (2007), pp. 1729–1736.

Appendix

- Traffic equilibrium sensitivity
- White paper on sustainability
- Socially-aware routing algorithm
- Koopman operator
- PCA on traffic flows
- EECS127 class
- Waze book