Dynamics and games for information-aware routing in traffic networks

Tommaso Toso

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : EEATS – Électronique, Électrotechnique, Automatique, Traitement du Signal
Spécialité : Automatique – Productique
Unité de recherche : Grenoble Images Parole Signal Automatique

Dynamiques et jeux pour le routage informé dans les réseaux des transport

Dynamics and games for information-aware routing in traffic networks

Présentée par:

Tommaso TOSO

Direction de thèse :

Alain KIBANGOU MAITRE DE COONFERENCES, Université Grenoble Alpes	Directeur de thèse
Paolo FRASCA CHARGE DE RECHERCHE, CNRS - Délégation Alpes	Co-directeur de thèse

Rapporteurs :

Jason R. MARDEN
FULL PROFESSOR, University of California, Santa Barbara

Simona SACONE
FULL PROFESSOR, Università degli Studi di Genova

Thèse soutenue publiquement le **04 octobre 2024**, devant le jury composé de :

Marie-Laure ESPINOUSE PROFESSEURE DES UNIVERSITES, Université Grenoble Alpes	Présidente
Alain KIBANGOU MAITRE DE COONFERENCES, Université Grenoble Alpes	Directeur de thèse
Paolo FRASCA CHARGE DE RECHERCHE, CNRS - Délégation Alpes	Co-directeur de thèse
Jason R. MARDEN FULL PROFESSOR, University of California, Santa Barbara	Rapporteur
Simona SACONE FULL PROFESSOR, Università degli Studi di Genova	Rapporteure
Paola GOATIN DIRECTRICE DE RECHERCHE, Centre INRIA de l'Université Côte d'Azur	Examinatrice
Ashish K. CHERUKURI ASSISTANT PROFESSOR, Rijksuniversiteit Groningen	Examinateur
Carlos CANUDAS-DE-WIT DIRECTEUR DE RECHERCHE, CNRS - Délégation Alpes	Examinateur

A mia nonna Danzica

Acknowledgment

These acknowledgments must begin by expressing my immense gratitude to my two advisors, Alain Y. Kibangou and Paolo Frasca. Your guidance and mentorship have been essential to my growth over the past three years, and I am confident that everything I have learned from you will continue to shape me into a better researcher every day.

I also extend my heartfelt thanks to Francesca Parise, who hosted me during my visit to Cornell University. Our collaboration was a profoundly enriching experience that undoubtedly contributed to my personal and academic development.

My gratitude extends to the members of my Comité de suivi individuel, Federica Garin and Nicolas Gast, for their insightful advice, which helped me focus my efforts more effectively. A particular thanks to Federica, for her constant support as a senior member of my team.

I am deeply thankful to the reviewers of this dissertation, Jason Marden and Simona Sacone. Jason, especially, for the illuminating discussions we've shared. I am equally grateful to those who accepted the role of examiners for my defense: Marie-Laure Espinouse, Paola Goatin, Carlos Canudas-de-Wit, and Ashish K. Cherukuri. Their observations have opened up compelling avenues for future research.

To my colleagues and peers, thank you for the time we shared and the inspiration you provided. Special recognition goes to Mladen and Martin, whose experience was a constant source of inspiration and encouragement.

My deepest affection goes to the friends who have accompanied me on this journey: Giovanni, Ivana, Maria, Gian Marco, Charline, Martina, Jacopo, Pietro, and Valeria. Sharing this adventure with you has made me a better person—and perhaps even a better researcher.

And last, but by no means least, my endless gratitude, though it will never suffice, goes to my family: my mother Antonietta, my father Luca, my brother Matteo, and my uncle Giuseppe. Your love and support will always be my anchor in challenging times.

Tommaso Toso,
Grenoble, December 2024

Résumé

Les réseaux de transport jouent un rôle clé dans nos économies modernes en facilitant la circulation des personnes et des marchandises. Cependant, malgré leur importance, ces réseaux sont confrontés à des problèmes de congestion, qui affectent l'économie, la qualité de vie, l'environnement et l'équité sociale. La demande croissante et la complexité des réseaux rendent la congestion toujours plus problématique.

Pour répondre à ce défi, la mobilité intelligente a émergé comme une solution efficace ces dernières années. Les innovations dans les technologies de l'information et la diffusion des appareils GPS et smartphones ont favorisé l'apparition de services de mobilité qui transforment les systèmes de transport urbains. Des services comme les applications de navigation et les VTC sont devenus omniprésents, modifiant profondément les réseaux de transport.

Le routage informé, qui exploite les données de trafic en temps réel pour optimiser les itinéraires, est un élément central de cette révolution. Ce concept est crucial pour les services de mobilité, de la navigation aux plateformes de VTC. L'impact du routage informé sur l'efficacité des réseaux de trafic est de plus en plus étudié, en se demandant si ces usagers informés, en cherchant à optimiser leur parcours, génèrent des gains globaux ou de nouvelles inefficacités.

Parmi les premiers modèles de réseaux de trafic les plus utilisés figurent les jeux de routage, des modèles statiques dont l'objectif est de caractériser les flux d'équilibre de trafic d'un réseau de transport, en se basant sur des hypothèses concernant le comportement des voyageurs, notamment celle selon laquelle tous les utilisateurs cherchent à minimiser leur propre temps de parcours (routage égoïste). Bien que ces modèles aient considérablement enrichi notre compréhension des flux et de l'efficacité du trafic, ils présentent des limites importantes lorsqu'ils sont appliqués aux réseaux modernes. Cette thèse étend le cadre traditionnel des jeux de routage égoïste afin de surmonter ces limites et de mieux capturer les effets du routage informé.

Ci-dessous, un résumé est proposé pour chacun des chapitres de la thèse, incluant les méthodologies utilisées, les résultats obtenus, ainsi que les perspectives futures concernant le travail réalisé dans cette thèse.

Chapitre 3: Routage égoïste sur des réseaux avec contraintes d'offre et démande

Les jeux de routage non atomiques classiques [1, 2] ont prouvé leur utilité pour comprendre divers aspects des réseaux de trafic. Néanmoins, leur formulation de

base présente certaines limites lorsqu'ils sont appliqués aux réseaux de trafic. Tout d'abord, les liens des réseaux n'ont pas de contraintes de capacité limitant la quantité de flux qu'ils peuvent accueillir. Cela ne permet pas de saisir correctement les phénomènes de congestion typiques des réseaux de trafic. De plus, les coûts des liens sont généralement modélisés comme des fonctions croissantes du flux. Étant donné que les coûts des liens représentent généralement les temps de trajet des liens dans la plupart des applications aux réseaux routiers, cela n'est pas cohérent avec la modélisation du trafic, selon laquelle la relation entre le flux de trafic et le temps de trajet est non monotone.

Dans ce chapitre, nous proposons un nouveau type de jeu de routage non atomique en exploitant des concepts du modèle CTM de Daganzo [3, 4] pour définir une structure de réseau. L'innovation clé réside dans l'incorporation de variables de densité aux côtés des flux de trafic traditionnels. En considérant à la fois les flux (nombre de véhicules par unité de temps) et les densités (nombre de véhicules par unité de longueur) sur chaque lien, nous pouvons définir un mécanisme d'offre et de demande qui impose des contraintes de capacité. Ce mécanisme limite le flux qui peut traverser un lien en fonction de sa densité actuelle, permettant ainsi d'identifier les sections congestionnées. De plus, les temps de trajet sur chaque lien deviennent directement dépendants de sa densité, conformément à la modélisation du trafic [5]. Cette approche combinée fournit une représentation plus précise des phénomènes réels du trafic.

Ce modèle permet d'identifier une conséquence critique du routage égoïste qui va au-delà du problème bien connu de l'efficacité réduite du trafic en raison de l'augmentation du temps total de trajet : *le transfert partiel des flux d'équilibre*, aussi connus comme *équilibres de Wardrop (WE)*. Plus précisément, dans certains cas, le WE résultant de l'interaction des usagers égoïstes consiste en un modèle de congestion qui permet seulement à une partie du flux exogène auquel le réseau est soumis d'entrer dans le réseau et de le traverser.

Les principales contributions de ce chapitre sont les suivantes :

1. Nous proposons un nouveau type de jeu de routage égoïste mieux adapté à la modélisation des réseaux routiers basé sur le CTM. Dans ce modèle, les liens sont traités comme des cellules avec des contraintes de capacité qui dépendent de la densité au sein de la cellule, et les temps de trajet des liens sont des fonctions croissantes de la densité plutôt que du flux. Nous nous concentrons sur les *réseaux parallèles*, où le réseau est constitué de N routes parallèles, chacune composée de plusieurs liens.

2. Nous caractérisons les WE et l'optimum social (SO) de ce jeu et prouvons leur unicité essentielle.

3. Enfin, nous introduisons le concept de transfert partiel des WE et montrons que, sous certaines conditions, le WE unique du jeu peut être partiellement transférant, même lorsque la demande exogène du réseau est inférieure à la capacité de min-cut. Cela fournit de nouvelles preuves de l'inefficacité du routage égoïste.

Chapitre 4: Impact des systèmes de navigation : recommandations en temps réel

Dans ce chapitre, nous allons au-delà d'une description statique du trafic et proposons une modélisation dynamique, dans le but d'offrir une représentation plus réaliste des réseaux de transport. Nous modélisons l'impact de l'utilisation des applications de navigation sur l'efficacité des réseaux de trafic. Pour ce faire, nous définissons un modèle dynamique de flux sur un réseau qui décrit la dynamique du trafic d'un réseau parallèle soumis à un flux de trafic exogène. Nous supposons qu'une partie du trafic est constituée d'usagers qui suivent les recommandations d'une application de navigation, qui les dirige vers l'itinéraire ayant le temps de trajet le plus court, tandis que l'autre partie choisit ses itinéraires en fonction de croyances préalables. Contrairement aux travaux précédents et en continuité avec le chapitre précédent, le modèle est défini sur un réseau soumis à des contraintes d'offre et de demande sur ses liens. Ainsi, similaire au chapitre 3, notre analyse de l'efficacité se concentrera principalement sur la question du transfert partiel de demande introduit précédemment, en mettant l'accent sur le rôle que jouent les applications de navigation dans l'émergence de ce phénomène.

Le modèle présenté dans ce chapitre peut être considéré, dans une certaine mesure, comme une extension dynamique du modèle présenté au chapitre 3, bien qu'il soit restreint à un réseau avec seulement deux itinéraires, chacun composé d'un seul lien. La demande de trafic est divisée en deux classes : l'une suit une stratégie de routage fixe, tandis que l'autre utilise une application de navigation pour minimiser le temps de trajet. Nous démontrons la stabilité asymptotique globale pour une large famille de dynamiques des préférences des utilisateurs. Ensuite, nous étudions les propriétés de l'équilibre unique en supposant que les préférences des utilisateurs suivent le modèle de choix logit. Cette analyse est effectuée dans deux régimes limites. Dans le régime de forte conformité aux recommandations de l'application, nous montrons que l'équilibre approche l'équilibre de Wardrop de l'instance correspondante du jeu de routage non atomique défini dans le chapitre 3. Dans le régime de faible conformité, nous dérivons une approximation linéaire des dynamiques des préférences des utilisateurs. Cette étude en deux volets montre que les applications de navigation peuvent dégrader l'efficacité du réseau, en augmentant le temps de trajet moyen (en accord avec les travaux précédents [6, 7]) et en menant à un transfert partiel de demande. La variable clé dans notre analyse à l'état stationnaire est le taux de pénétration, c'est-à-dire la part des utilisateurs informés par l'application dans la demande totale. Notre analyse montre qu'un taux de pénétration élevé est susceptible de dégrader l'efficacité du réseau lorsque la conformité est élevée. Ce constat est également confirmé par l'analyse numérique du modèle dans une étude de cas réaliste basée sur la ville de Grenoble, France.

Chapitre 5: Impact des systèmes de navigation : recommandations de routage retardées

Les recommandations de routage fournies par les applications de navigation à leurs utilisateurs sont dérivées des données de trafic collectées par l'application. Il ex-

iste un délai inévitable entre le moment où les données de trafic sont collectées et le moment où les recommandations basées sur ces données sont fournies aux utilisateurs. Ce retard est dû au temps nécessaire pour les opérations de collecte, de communication et de traitement des données.

Dans ce chapitre, nous cherchons à évaluer quel impact les retards affectant les recommandations de routage peuvent avoir sur l'efficacité du trafic. Pour ce faire, nous utilisons une version légèrement modifiée du modèle développé dans le chapitre 4, en tenant compte de ce délai.

En supposant que les deux itinéraires du réseau ont la même longueur et la même vitesse, nous réduisons le problème à une équation différentielle ordinaire scalaire en considérant la différence de temps de trajet entre les deux itinéraires. En menant une analyse de stabilité de cette dynamique scalaire, nous pouvons caractériser de manière exhaustive l'impact du retard informationnel sur la stabilité et l'efficacité du système de trafic. Cette analyse implique trois paramètres clés : le flux exogène à travers le réseau, le taux de pénétration des utilisateurs informés et leur conformité.

Nos résultats indiquent que lorsque ces trois paramètres sont suffisamment faibles, le retard dans les recommandations de routage n'altère pas le comportement asymptotique du système de trafic. Cependant, lorsque le produit de ces paramètres dépasse un certain seuil, qui peut être exprimé en termes d'autres paramètres du système, des retards suffisamment importants entraînent une déstabilisation du système. Cela provoque un état de trafic oscillant et peut entraîner un transfert partiel de demande périodique.

Chapitre 6: Routage coordonné

Ce chapitre explore l'impact de la présence d'une flotte centralisée sur l'efficacité du trafic. Chaque flotte, sous le contrôle d'un opérateur centralisé, optimise ses opérations grâce au routage coordonné des véhicules. Nous modélisons le problème comme une instance des jeux de routage à comportement mixte [8], impliquant deux classes d'utilisateurs distinctes. La première classe comprend des utilisateurs individuels, égoïstes, cherchant à minimiser leurs propres temps de parcours. La deuxième classe est composée de véhicules qui coordonnent leurs décisions de routage afin de minimiser le temps de parcours moyen de la flotte.

Cette étude vise à améliorer notre compréhension de la manière dont les services de mobilité récemment nés, notamment les VTC, influencent l'efficacité du trafic. Les entreprises de ce secteur exploitent les données de trafic en temps réel pour développer des plans de déploiement de flotte stratégiques qui rationalisent les opérations, garantissent un service de qualité et maximisent la rentabilité, tout en utilisant des stratégies de routage coordonné pour optimiser les performances de la flotte. Il est important de noter que cette étude se concentre exclusivement sur les stratégies de routage et ne fournit pas une analyse exhaustive des impacts plus larges de ces services sur l'efficacité du trafic, tels que la prise en compte des kilomètres de véhicules vides ou des effets potentiels sur l'utilisation des transports publics.

Les principales contributions de ce chapitre sont les suivantes. Après avoir observé que, sous des hypothèses légères, ce jeu à deux classes est équivalent à un jeu

convexe à deux joueurs, nous exploitons une reformulation bien connue en termes de solution à une inéquation variationnelle (IV) pour étudier le problème (voir [9, 10]). Plus précisément, nous identifions des conditions suffisantes pour que l'opérateur de l'IV soit fortement monotone. D'une part, la monotonie forte garantit l'unicité de l'équilibre. D'autre part, elle permet de fournir des aperçus significatifs sur l'impact de la part de la flotte coordonnée sur l'efficacité globale du trafic dans les réseaux à deux terminaux.

Nous utilisons le Prix de l'Anarchie (PoA) comme métrique de l'efficacité du trafic. Nous démontrons que l'équilibre unique et le PoA présentent une continuité de Lipschitz par rapport à la part des véhicules coordonnés. De plus, nous établissons des conditions garantissant un seuil minimum pour cette part. En dessous de ce seuil, la présence de véhicules coordonnés n'a aucun impact sur l'efficacité du trafic.

Enfin, pour les réseaux parallèles, nous montrons que le PoA, les flux des utilisateurs individuels et le temps de trajet le plus court à l'équilibre diminuent tous lorsque la part de véhicules coordonnés augmente. Cela suggère que de plus grandes flottes coordonnées conduisent à une meilleure efficacité.

Conclusion

Cette thèse explore le routage informé et son impact sur l'efficacité des réseaux de trafic. En étendant les modèles et théories existants, elle permet de proposer une analyse plus complète sur la manière dont les technologies modernes et les services basés sur le routage informé, tels que les applications de navigation et les services de VTC, influencent la dynamique du trafic.

Plusieurs pistes de recherche prometteuses s'ouvrent pour étendre et affiner les travaux présentés dans cette thèse. Une extension clé est la généralisation des modèles à des topologies de réseaux plus complexes. Les modèles développés se sont concentrés sur des réseaux simplifiés, mais les réseaux réels, avec une diversité de paires origine-destination et de routes croisées, nécessitent des modèles plus généraux pour mieux caractériser les phénomènes observés et étendre leur applicabilité.

La généralisation des jeux de routage non atomiques avec contraintes d'offre et de demande nécessitera de formaliser ces modèles comme des problèmes d'optimisation, en tenant compte des spécificités de chaque itinéraire, notamment les limites de capacité des liens et intersections imposées par le méchanisme d'offre et demande.

Les modèles dynamiques avec contraintes d'offre et de demande posent des défis supplémentaires. L'intégration d'informations en temps réel rend la dynamique des flux plus complexe, nécessitant de nouvelles approches pour analyser la stabilité des systèmes.

Une fois ces modèles dynamiques développés, une voie importante de recherche consistera à comprendre comment l'introduction de contraintes de capacité modifie notre compréhension des problèmes de routage. L'incorporation de contraintes réalistes pourra améliorer la gestion du trafic, en particulier dans les réseaux à usagers égoïstes, et nécessitera une évaluation des mécanismes existants comme la tarification et les incentives.

La validation des modèles à travers des simulations microscopiques est essentielle pour garantir la précision des résultats théoriques. Ces simulations, réalisées avec des plateformes de micro-simulation, permettront de tester les modèles sur des réseaux et dynamiques de trafic réalistes.

Enfin, l'étude du routage coordonné dans des systèmes de transport basés sur des flottes offre un grand potentiel. Le routage coordonné pourrait être crucial pour optimiser l'efficacité des réseaux, notamment dans les systèmes de VTC ou de livraison. Ces recherches pourraient aboutir à de nouvelles approches de contrôle du trafic.

Contents

List of Symbols

$\mathcal{G}$	Network
$\mathcal{N}$	Set of nodes
$\mathcal{L}$	Set of links
$\mathcal{K}$	Set of OD pair
$\mathcal{P}$	Set of routes
x	Traffic density
f	Link traffic flow
v	Free-flow speed
w	Congestion wave speed
z	Route traffic flow
t	Continuous time
s	Supply function
d	Demand function
Φ	Exogenous flow
τ	Travel time
R	Routing ratios
α	Penetration rate
$\frac{1}{\eta}$	User compliance
γ	Coordinated fleet share

List of Acronyms

CTM	Cell Transmission Model
GAS	Global Asymptotic Stability
MTNM	Macroscopic traffic network model
NRG	Non-atomic routing games
OD	Origin-destination
ODE	Ordinary differential equation
PDE	Partial differential equation
PoA	Price of Anarchy
SO	Social optimum
VI	Variational inequality
WE	Wardrop equilibrium

Part I

Introduction & Preliminaries

Chapter 1

Introduction

1.1 The rise of information-aware routing

Navigating through traffic networks has become increasingly challenging in recent years. Urban areas are often plagued with heavy congestion, and the complexity of modern traffic networks adds another layer of difficulty for commuters. As cities grow and populations increase, the demand for efficient movement intensifies.

In response to this challenge, a plethora of services aimed at simplifying transportation have emerged. Navigation apps like Google Maps and Waze provide real-time guidance, while ride-hailing services such as Uber and Lyft offer convenient alternatives to traditional taxis. Car-pooling and car-sharing platforms like BlaBlaCar and Zipcar encourage shared usage of vehicles, reducing the number of cars on the road. Collectively, these services aim to enhance the mobility experience of urban commuters by offering flexible, on-demand solutions, that are often more efficient and cost-effective than traditional methods.

A common thread among many of these modern transportation solutions is their reliance on *information-aware routing*. Much of these solutions, in fact, rely on real-time traffic data to maximize the quality of the service offered and the efficiency of their operations. This is particularly the case for navigation apps and ride-hailing services, which critically rely on information-aware routing.

1.1.1 Overview of practices

Navigation apps

Navigation apps are software designed to help travellers find the most efficient routes to their destinations. Using GPS technology, digital maps, and real-time traffic data, these apps provide optimized routing recommendations to minimize travel time by avoiding slowdowns and bottlenecks. Navigation apps have achieved near-ubiquitous adoption. This is evident in the staggering user base of popular navigation apps, with Google Maps surpassing 23 million downloads in the US for 2023 alone, and Gaode Map boasting a colossal 730 million monthly active users in China as of 2022 [11, 12].

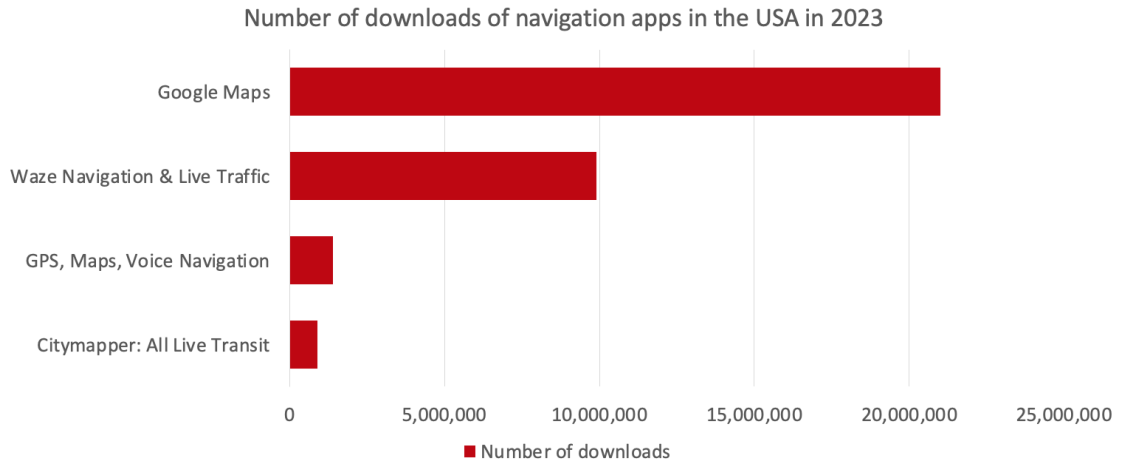

Figure 1.1: *Source:* AppMagic, Statista.

Ride-hailing services

Ride-hailing services, such as Uber and Lyft, operate through a digital platform that connects passengers with drivers via a mobile application. When a passenger requests a ride, the app matches them with a nearby driver who accepts the request. The app then provides the driver with the passenger's location and destination, and the driver uses this information to navigate to the pick-up point, transport the passenger to their destination, and complete the ride.

Information-aware routing is crucial for enhancing the efficiency and effectiveness of ride-hailing services. These services exploit real-time traffic data to optimize their fleet deployment, in order to ensure efficient passenger transportation and minimizing driver downtime between pick-up points. This approach not only improves service reliability but also maximizes profitability by streamlining operations.

Similarly to navigation apps, ride-hailing services are witnessing parallel growth, driven by increasing urbanization and disposable income. Forecasts predict their user base will reach 1.97 billion by 2028 (Statista [13]).

1.1.2 Unexpected consequences

Initially, the proliferation of services based on information-aware routing was met with great optimism. Information-aware routing was then considered a tool that would certainly improve traffic flow and reduce congestion phenomena. However, these expectations have largely been unmet. Despite the widespread adoption of these services, significant improvements in traffic congestion have not been materialized. Instead, unexpected and undesirable phenomena associated with navigation apps and ride-hailing services have emerged.

Notably, navigation apps have been accused of contributing to the rise of cut-through traffic in residential areas across the United States and Europe. Many municipalities have reported a marked increase in traffic volumes on their local roads as navigation apps direct users through quieter suburban streets to avoid congested highways [14, 6, 15, 16]. This phenomenon has prompted action from authorities. In the US, municipalities have tried various solutions, such as limiting

road use to residents [17] or implementing tolls [18], while France has issued a national decree regulating navigation apps [19].

Ride-hailing services have also been criticized for being responsible of increasing urban congestion [20, 21, 22, 23]. While factors like no passenger travels and competition with public transportation are often cited, the role of information-aware routing in fleet deployment deserves scrutiny. Just as navigation apps have shown, these routing strategies, designed to optimize individual rides, may inadvertently contribute to traffic flow inefficiencies.

The proliferation of navigation apps and mobility services such ride-hailing services and the unexpected effects stemming from their use underline the necessity of a more comprehensive assessment of the implications of information-aware routing. Such an analysis should aim to comprehend the full range of consequences associated with it, highlighting potential shortcomings and drawbacks.

The aim of this thesis is to contribute to the analysis of the impact of information-aware routing on traffic efficiency in traffic networks, with a specific focus on navigation apps and ride-hailing services. We consider information-aware routing in two primary contexts. First, we model the impact of users leveraging traffic information provided by navigation apps to minimize their travel time. Second, we assess the effects of coordinated routing within ride-hailing service fleets, controlled by a central operator aiming to optimize overall fleet metrics, on overall traffic efficiency.

1.2 Previous work

The impact of information-aware routing on traffic efficiency has received significant attention in recent years. This section reviews relevant studies to provide context for our analysis.

The problem has been tackled primarily through the lens of *macroscopic traffic network models* (MTNM). MTNMs aim to provide a mathematical representation of a traffic network capable of describing the behavior of traffic flows. At their core, MTNMs represent a road network by means of a directed graph. Graph links represent the network roads, and nodes represent the junctions between them. Links are assigned a number of attributes, such as length and capacity, and a travel time function, quantifying the amount of time it takes to traverse the link given a certain level of traffic volume on that link. Each ordered pair of nodes in the network is assumed to have a non-negative demand of traffic of users, who aim to cross the network in order to move between the origin node and the destination node of that pair. MTNMs focus on the bigger picture: the number of users is assumed to be very large and each user controls a negligible fraction of traffic, so traffic is modeled in an aggregate, continuous manner by link flows and densities.

In MTNMs, the behavior of traffic flows is modeled through the imposition of mathematical conditions that reflect specific behavioral assumptions. The derivation of such mathematical conditions is done through the adoption of a game-theoretic approach, that allows for elegantly capture the reactive behavior to congestion of users informed about traffic conditions over the network. Typically, these conditions translate into users favouring routes associated with shorter travel times.

1.2.1 Navigation apps

The impact of navigation apps' usage has been studied through two main families of models: heterogeneous non-atomic routing games and dynamical network flows.

Non-atomic heterogeneous routing games

Non-atomic routing games (NRG) represent one of the most widely used MTNMs. NRGs are static models, so their goal is to identify steady-state traffic patterns emerging from individual user behavior. In their older and arguably most studied variant, NRGs assumes that all users behave according to the selfish routing model: all of them has complete information about traffic conditions and their goal is to minimize their own traversal time. This core assumption aligns perfectly with the behavior of users of navigation apps, who rely on the routing recommendations allowing them to minimize their travel time. In this model, equilibrium configurations are those traffic patterns such that no user can improve their travel time by unilaterally changing routes. These traffic patterns are known in the literature as *Wardrop equilibria* [24]. The study of the efficiency of Wardrop equilibria has demonstrated that selfish routing is generally inefficient. Users act in an uncoordinated manner and disregard the impact of their actions on other users, leading to sub-optimal traffic configurations in terms of total travel time on the network [1]. Various metrics have been developed to evaluate the inefficiency of Wardrop equilibria, with the Price of Anarchy being one of the most widely used [1].

Originally, the selfish routing model assumed complete homogeneity in user behavior. However, it was later generalized to account for heterogeneity in the traffic information available to users [2]. This generalization introduced multiple classes of users, each characterized by a specific set of travel time functions. This heterogeneity is particularly valuable for investigating the impact of navigation apps. Firstly, it allows for addressing cases where not all users rely on routing recommendations or on the same navigation apps, resulting in varying levels of knowledge. Secondly, it allows for the assessment of selfish routing impacts at different levels of navigation app penetration among users. For this reason, all relevant work that focused on assessing the impact of information-aware routing included heterogeneity in their models.

The impact of navigation apps was first investigated within the framework of heterogeneous non-atomic routing games in [25]. In this work, users are divided into two classes: app-informed users, who follow navigation app recommendations, and non-routed users who do not. Network roads are also divided into high-capacity and low-capacity roads. The two classes have different travel time functions. Routed users' travel time functions represent actual link travel times without differentiating by road capacity, whereas non-routed users' functions lead them to favor high-capacity roads most of the time, even when low-capacity routes had lower travel times. Two important facts emerge from the numerical experiments performed on the model in [25]. First, the introduction of app-informed users on the network leads to an improvement in traffic efficiency, alleviating traffic congestion on high-capacity roads. On the other hand, reduced pressure on high-capacity roads comes at the cost of significantly increasing congestion on secondary roads. The first

fact suggests that although selfish routing behaviors are sub-optimal in terms of traffic efficiency, information-aware selfish routing can help to relieve congestion when routing is performed under biased or incomplete information. The second fact, instead, provides a clear explanation for the emergence of the phenomenon of cut-through traffic[1].

The idea that providing users with more information always leads to an efficiency improvement was contradicted a few years later. In [7], the concept of Informational Braess' Paradox (IBP) was introduced. To investigate the impact of additional route information provided through navigation apps, a non-atomic routing game was developed where users have varying information sets about available routes and can only use routes within their information set. The focus is on whether increasing the information set of one of the classs can be detrimental by increasing total travel times. It is shown that in networks outside a rather narrow class of networks (series of linearly independent networks), there exists a configuration of link travel time functions where IBP will occur. Interestingly, it is also shown that the maximum inefficiency occurs when all users' information sets correspond to the entire network.

In line with this work, Cabannes considered a two-class non-atomic routing game with app-informed users and non-informed users [6]. Utilizing the concept of average marginal regret, it is proved that increasing the penetration level of navigation apps steers traffic patterns towards Wardrop equilibria, potentially degrading traffic efficiency through increasing total travel time and causing cut-through traffic on secondary roads.

This body of work underscores the complexity and potential unintended consequences of information-aware selfish routing, highlighting the need for careful consideration in their use.

Finally, another model offering interesting insights on the use of navigation apps is presented in [26] through the framework of *Bayesian routing games*. In that work, the authors examine a heterogeneous NRG, where each class of users is associated with a distinct navigation app, and the network is influenced by an uncertain state affecting traffic conditions. Each navigation app sends a noisy signal about the state to its users, who then make route choices based on their own beliefs and the information received from the app. In this scenario, users select their routes to minimize their expected travel time. The authors fully characterize the equilibrium structure and analyze how population sizes influence the difference in expected travel times between different populations, essentially the advantage that one population has over another. They identify population-specific size thresholds for each pair of populations, showing that one population benefits more than another if and only if its size is below the corresponding threshold. The key takeaway is that users within a class significantly benefit from the information provided by their navigation app only when a relatively small number of users are utilizing that same information.

[1] Cut-through traffic means traffic that passes through a given residential neighborhood that has neither an origination nor destination point in that neighborhood.

Dynamical network flows

While NRGs have proven to be a highly useful tool for evaluating the impact of real-time routing recommendations, they present a crucial limitation. The equilibria of an NRG are assumed to be the traffic pattern naturally resulting from user interaction. However, since NRGs do not provide any dynamic description of traffic, it is unclear whether the network will converge to any of these equilibria from a given initial condition. In an effort to provide a more comprehensive analysis of the implications of navigation apps' usage, researchers have explored the use of dynamic models to investigate the problem.

To study the stability of Wardrop equilibria in NRGs, evolutionary dynamics [27] have been proposed. Evolutionary dynamics consist of an ODE system that describes the evolution of user preferences over time, where the portion of users choosing a specific route decreases as its cost (e.g. travel time) increases. Two evolutionary dynamics that have received attention for traffic applications are the *replicator dynamics* and the *logit dynamics*. Briefly, according to the replicator dynamics, users tend to update their route choice when the route travel time exceeds the average travel time. In contrast, the logit dynamics is a perturbed best-response dynamic: users always aim for the shortest travel time route, but the costs they perceive are affected by noise, which can lead them to choose a sub-optimal strategy. The stability of Wardrop equilibria under general evolutionary dynamics has already been established for potential games, as homogeneous NRGs are [27, Chapter 7]. For heterogeneous games, no such general result is available. To the best of the authors' knowledge, until now, stability results in heterogeneous settings have only been proposed for the logit dynamics [28]. This work proved that a subclass of the Wardrop equilibria of the NRG is asymptotically stable. Despite their attempt to dynamically characterize the problem, evolutionary dynamics do not really model actual traffic dynamics, as they assume that traffic flows propagate instantaneously across the network.

Another approach in the literature relies on *dynamical network flow* models [29]. Dynamical network flows are inspired by compartmental models [30, 31]. The intuition behind these models is to think of the links of a traffic network as a set of interconnected cells exchanging mass with each other. The flows from one cell to another are determined by a *supply and demand* mechanism, inspired by Daganzo's cell transmission model (CTM) [3, 4]. Unlike evolutionary dynamics, these models not only capture the dynamics of user preferences, but also of the traffic dynamics over each network link. Recent works [32, 33] have provided results concerning the stability of Wardrop equilibria under the replicator and logit dynamics for dynamical network flows.

In [32], user preferences are modeled through the logit dynamics. The authors show that the traffic system admits a unique globally asymptotically stable fixed point, which converges to the Wardrop equilibrium, as the noise vanishes. In [33], the evolution of route preferences is modeled through the well-known *replicator dynamics*. Sufficient conditions for convergence to a Wardrop equilibrium are provided, and numerical simulations also show that when such conditions are not met, user strategic behavior can lead to instability, causing the traffic state to oscillate.

In both of these works, information is homogeneous across users, meaning all users have full knowledge of the traffic state in real-time.

The problem of assessing the impact of selfish routing induced by real-time information and recommendations has also been explored through macroscopic PDE models. These models couple a mass conservation law with a Hamilton-Jacobi equation to model both the traffic dynamics and the strategic choices of users. In [34, 35, 36], it is shown that these types of models can capture the inefficiencies previously highlighted in static frameworks (sub-optimality with respect to total travel time, Braess' Paradox). In [35, 36, 37], the case of heterogeneous class of users characterized by different levels of information is considered. The conclusions drawn are analogous to those stemming from their static counterparts [6, 25, 7]: real-time information provision can help to relieve network congestion, but increasing levels of penetration of information-aware routing lead to Wardrop equilibria and typically result in higher pressure on secondary roads.

1.2.2 Ride-hailing services

The assessment of the impact of routing policies adopted by ride-hailing services has been addressed through *mixed behavior non-atomic routing games* and *coordinated routing*. Similarly to heterogeneous NRGs, mixed behavior NRGs also involve dividing traffic flow into separate users classes, only this time classes are characterized by different behavioral assumptions, rather than different sets of travel time functions. Mixed behavior NRGs were considered for the first time in [8], which introduced the concept of coordinated routing within classes. The mixed behavior NRG developed in that work is characterized by a class of selfish users and several coordinated classes, consisting of users that act coordinately in order to minimize the fleet total travel time, instead of seeking to minimize their own travel time. Sufficient conditions for equilibrium existence and uniqueness are established. More general conditions have been provided later in [38].

However, the above works do not investigate the impact that coordination among users in a same class has on the overall traffic efficiency. The first work to consider this problem is [39], where the authors consider a three-class problem with a class of selfish users, a coordinated fleet (aiming at reducing fleet average travel time) and a system optimum fleet (aiming at reducing the system's average travel time). The provided numerical experiments show that sufficiently large coordinated and system-optimal fleets can lead to system optimality, even in the presence of individual users, thus improving the overall traffic efficiency.

More recently, the impact on overall efficiency has been analyzed within the framework of two-class problems. Studies such as [40, 41] consider a two-class problem, where one class consists of selfish users and the other comprises coordinated vehicles. In [40], the authors develop an algorithm to compute the traffic equilibrium resulting from the interaction of selfish users and the coordinated fleet. Through numerical experiments, they highlight that traffic efficiency improves as the fraction of coordinated vehicles on the total increases.

In [41], the authors study the impact of coordinated fleets on traffic efficiency. First, they provide an example on a network with multiple origin-destination pairs

and show that coordinated fleets can have detrimental effects on efficiency. Then, they investigate the minimum fleet size necessary to achieve system optimality and the maximum fleet size for which the user equilibrium persists, developing mathematical programs to compute them. They also provide analytical results about the threshold effect associated with the coordinated fleet size on efficiency, but only for parallel networks.

This review focused only on how information-aware routing strategies affect traffic efficiency. However, a complete picture of ride-hailing services' impact requires considering additional factors. These factors include, for example, no passenger trips and a shift away from public transportation. While these aspects are beyond the scope of this thesis, the provided references [23, 42, 43, 44, 45] offer starting points for further exploration.

1.3 Thesis contributions

This thesis investigates the impact of information-aware routing on the efficiency of traffic network.

1.3.1 Navigation apps & selfish routing

Despite the valuable insights provided by the studies mentioned in Section 1.2, many existing models have design limitations that may prevent them from fully capturing the complexities of traffic phenomena. For example, all cited works do not incorporate capacity constraints at all [6, 25, 7], or only include partial constraints [32, 33]. Instead, traffic networks have inherent limitations in the number of vehicles they can handle: the traffic volume on a road is limited upwards and affects the link's ability to accommodate additional traffic. Neglecting these constraints raises questions about the completeness of the investigations carried out with these models.

In Part II, we address this gap by investigating information-aware routing associated with navigation apps using traffic network models that more accurately describe road networks, incorporating the aforementioned capacity constraints. We next summarize out three main contributions on this topic.

Selfish routing on networks with supply and demand constraints

In their standard formulation, non-atomic selfish routing models do not consider any capacity constraints on the network links. In this thesis, we extend upon previous work [1] the analysis of these models, applying them to networks that better represent road networks. We propose a new model that addresses the limitations of standard routing games by incorporating several key features, inspired by the well-established Cell Transmission Model (CTM) by Daganzo [3, 4]. By including traffic density in the state description, the model surpasses the constraints of flow-based models. Network links are characterized by a supply and demand mechanism that regulates capacity constraints based on density, reflecting the fact that congestion affects available capacity. Additionally, link travel times are now functions

of density, rather than flow [5]. Focusing on parallel networks and assuming users' homogeneity, we comprehensively characterize Wardrop Equilibria (WE) and social optima (SO). Significantly, the model uncovers a potential drawback of selfish routing behavior: the presence of *partially transferring Wardrop equilibria*. With partially transferring, we mean that the WE consists in a congestion pattern that allows only a part of the exogenous flow which the network is subject to to enter the network and cross it. We provide sufficient conditions for this phenomenon to occur.

Assessing the impact of navigation apps' usage in networks with supply and demand constraints

Most studies on traffic dynamics with information-aware routing assume overly simplistic capacity constraints. In [32, 33], for example, although the outgoing flow is limited by the link's capacity, the incoming flow can exceed this threshold. In this thesis, we study this problem through dynamical network flow that incorporates the same supply and demand mechanism of the above non-atomic routing game. This model can be thought of as its dynamic counterpart, offering a more comprehensive representation of traffic flow over time. Furthermore, in this case we account for heterogeneity among users. Specifically, we assume that the traffic demand is divided into two classes, where one class splits on the two routes according to a fixed routing strategy, whereas the second class relies of the routing recommendations of a navigation app and strategically select the route to minimize travel time. The analysis focuses on a network with two alternative routes. We begin by proving that the system exhibits global asymptotic stability for a broad family of user preference dynamics. Next, we investigate the characteristics of the unique equilibrium under the assumption that user preferences adhere to the logit choice model. This examination is conducted in two different compliance regimes to the app's recommendations. First, in the scenario of high compliance regime, we demonstrate that the equilibrium approximates the Wardrop equilibrium of a non-atomic routing game analogous to the one described in Chapter 3. In the low compliance regime, we derive a linear approximation for the dynamics of user preferences. This dual approach reveals that navigation apps can lead to traffic states affected by partial transfer of demand. A critical factor in our steady-state analysis is the level of penetration of app-informed users within the total demand. Our analysis indicates that the higher the level of penetration, the more likely partial demand transfer will emerge (particularly when user compliance is high). Additionally, corroborating previous findings [6, 25, 7], we show that an excessive penetration rate can also negatively affect efficiency, by increasing total travel time.

The impact of navigation apps' usage under informational delay

Although navigation apps increasingly rely on real-time traffic data, there is inevitably a time lag in information provision due to the need for data collection, communication, and processing before it can be used. Therefore, it is crucial to evaluate the impact that delays in routing recommendations can have on traffic

efficiency. This issue has not been addressed in previous studies, making our investigation the first to consider informational delays. We consider the aforementioned dynamical network flow model with user preferences modeled through the logit dynamics, and we modify it to account for delay affecting information communication, i.e., app-informed users base their choice on a past traffic state. Through the stability analysis of this model, we show that informational delays can negatively impact traffic efficiency, fully characterizing this phenomenon. Specifically, we show that for when the exogenous flow, the noise level and the fraction of app-informed users are low, the system admits a globally asymptotically stable equilibrium, regardless of the delay. However, when these parameters are sufficiently high, there exists a critical delay threshold beyond which the delay destabilizes the traffic dynamics and the trajectories becomes oscillatory. Finally, through numerical experiments, we show that in some cases the oscillations in the traffic state are so pronounced to lead to a periodic partial transferring of demand.

1.3.2 Ride-hailing services & coordinated routing

Although the problem of assessing the impact of ride-hailing services has already been considered in the literature and studied through the lens of coordinated routing, existing works have mostly been limited to quantitative approaches [39, 40, 44, 41]. In this thesis, we propose an analytical study of the problem.

The impact of coordinated routing on traffic efficiency

We formulate the problem as a mixed behavior non-atomic routing game with two user classes, where the first class consists of selfish users, whereas the second one consists in a coordinated fleet of vehicles that strategically select their route to minimize the average travel time of the whole fleet. We show that this game is equivalent to a two-user game, where one user is associated with the individual users and the other with the coordinated fleet. We study this game by using a well-known reformulation in terms of solution to a variational inequality (VI) (see [9, 10]). Specifically, we establish conditions ensuring that the operator of the VI associated to our game is strongly monotone. On the one hand, strong monotonicity ensures equilibrium uniqueness. On the other hand, through this property we are able to provide meaningful insights about the relationship between traffic efficiency and the share of the coordinated fleet in two-terminal networks. Using the Price of Anarchy (PoA) as a measure of traffic efficiency [1], we prove that the unique equilibrium and the PoA are Lipschitz continuous functions of the fleet share. Additionally, we derive sufficient conditions for the existence of a minimum share below which the presence of a coordinated fleet has no effect on traffic efficiency. Finally, for parallel networks, we show that the PoA, the flow of individual users, and the shortest travel time at equilibrium are monotonically non-increasing functions of the fleet share, suggesting improved efficiency for larger fleet share. Notice that in this analysis, we do not take into consideration the capacity constraints used for the analysis of the impact of navigation apps.

1.4 Thesis outline

This section provides the reader with an overview of the thesis. The thesis consists of four parts:

- Introduction & Preliminaries,
- Selfish routing & navigation apps,
- Coordinated routing & optimized fleets,
- Conclusion.

This chapter is part of the first section of the thesis. Within it, we detail the content and contributions of each chapter, highlighting the publications which they are based on.

Chapter 2: Preliminaries

In this chapter, the reader is provided with an overview on the main approaches adopted in this thesis to address the problem of traffic modeling. Section 2.1 briefly introduces multi-digraphs. Section 2.2 is dedicated to static models, specifically non-atomic routing games. Section 2.3 briefly describes Daganzo's CTM. Finally, Section 2.4, instead, delves into dynamic traffic modeling, presenting dynamical network flows.

Part II: Selfish routing & navigation apps

Chapter 3: Selfish routing on networks with supply and demand constraints

The second part of the thesis focuses on the modeling of selfish routing. In this chapter, the non-atomic routing game on networks with supply and demand constraints is presented and analyzed. Section 3.1 opens the chapter, summarizing the main contributions and briefly discussing related works. In Section 3.2, we detail the network structure. We describe the supply and demand mechanism for each link, define feasible traffic assignments, and introduce density-dependent link travel times. In Section 3.3, we define a non-atomic selfish routing game on the network. Here, we also fully characterize the Wardrop equilibria and social optima of this game and provide necessary and sufficient conditions for the emergence of partially transferring Wardrop equilibria. Section 3.4 presents an example showing that partial demand transfer persists also in non-parallel networks. Section 3.5 terminates the section with some concluding remarks.

- T. Toso, A. Y. Kibangou, and P. Frasca, "Selfish routing on transportation networks with supply and demand constraints," under review in *IEEE Transactions on Intelligent Transportation Systems*, 2024.

Chapter 4: Impact of navigation systems: real-time routing recommendations

This chapter offers insights about the impact exerted by the presence of app-informed users on traffic efficiency. Section 4.1 introduces the main contributions and related works. In Section 4.2, the dynamical network flow model is presented. In Section 4.3, it is proved that the system exhibits global asymptotic stability for a wide range of user behavior models. In Section 4.4, the properties of the unique equilibrium of the system are analyzed, assuming that user preferences follow the logit choice model. We characterize the phenomenon of partial transfer demand and the impact on total travel time with respect to the fraction of informed users. This is done both in the high user compliance and low noise regimes. Section 4.5 proposes simulations against which the theoretical findings are validated. Section 4.6 presents further experiments demonstrating that our findings apply beyond the specific setup analyzed in this chapter. Section 4.7 concludes the chapter.

The chapter is based on the following publication:

- T. Toso, A. Y. Kibangou, and P. Frasca, "Modeling the impact of route recommendations in road traffic," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 4179–4185, 2023. 22nd IFAC World Congress.
- T. Toso, A. Y. Kibangou, and P. Frasca, "Potential detrimental effects of real-time routing recommendations in traffic networks," under review in Transportation Research: Part C, 2024.

Chapter 5: Impact of navigation systems: delayed routing recommendations

In this chapter, we propose a dynamical network flow model to analyze the impact of delayed traffic information on traffic efficiency. First, Section 5.1 opens the chapter. In Section 5.2, we introduce the model. This model is analogous to the one presented in Chapter 4, but now also accounts for informational delay. In Section 5.3, we show that, under a route homogeneity assumption, the system can be reduced to a scalar dyanmics, simplifying the analysis. In Section 5.4, we conduct the stability analysis of the model and provide a sufficient condition for delay-independent stability, as well as a sufficient condition for destabilization caused by delay. In Section 5.5, we present some numerical experiments confirming our theoretical findings. Section 5.6 concludes the chapter.

The chapter is based on the following publication:

- T. Toso, A. Y. Kibangou, and P. Frasca, "Impact on traffic of delayed information in navigation systems," *IEEE Control Systems Letters*, vol. 7, pp. 1500–1505, 2023.

Part III: Coordinated routing & optimized fleets

Chapter 6: The impact of coordinated routing on traffic efficiency

The third part of the thesis delves into coordinated routing. In this chapter, we study the impact that the presence of a coordinated fleet has on traffic efficiency.

Section 6.1 opens the chapter. In Section 6.2, the model and the main concepts are defined. In Section 6.3, we give strong monotonicity, existence and uniqueness conditions. Section 6.4 discusses the effect of a coordinated fleet on traffic efficiency as a function of the fleet size, leveraging strong monotonicity. Section 6.5 proposes numerical experiments corroborating the findings reported in the previous sections. Section 6.6 contains concluding remarks.

The chapter is based on the following publication:

- T. Toso, F. Parise, P. Frasca and A. Y. Kibangou, "On the impact of coordinated fleets size on traffic efficiency," accepted for presentation in *63rd IEEE Conference on Decision and Control (CDC)*, Milan, Italy, 2024.

Conclusion

The thesis concludes with a summary of the work presented in the previous chapters and an outline of future research directions.

Chapter 2

Preliminaries

This chapter provides a foundation in mathematical models that will be instrumental in understanding the analysis presented in this thesis. In Section 2.1 the mathematical representation of a traffic network is provided. Section 2.2 consists in an overview on non-atomic routing games. Section 2.3 describes Daganzo's Cell Transmission Model. Finally, Section 2.4 introduces dynamical network flows.

2.1 Traffic networks as multi-digraphs

The models adopted in this thesis rely on a mathematical representation of traffic networks based on multi-digraphs.

Definition 2.1. *A* multi-digraph (directed multigraph) *is a pair* $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, *where:*

- $\mathcal{N}$ *is the set of nodes with cardinality* $|\mathcal{N}|$*;*
- $\mathcal{L}$ *is the set of links with cardinality* $|\mathcal{L}|$*. Each link* l *is an ordered pair of nodes, i.e.,* $l = (u, v)$, $u, v \in \mathcal{N}$*.* u *is called the* tail*, whereas* v *the* head*.*

Let a, b be the *tail* and *head functions*, associating each link with its head and its tail, respectively. Being *ordered* pairs of nodes, links have a direction, i.e, each link connects its tail to its head. Links are not univocally determined by their tail and head, as multiple links can share the same tail-head pair. Nodes which are not directly connected by a link can be connected through a route. Given a node u, let u^+ indicate the set of $l \in \mathcal{L}$ such that $b(l) = u$ and let u^- indicate the set of $l \in \mathcal{L}$ such that $a(l) = u$.

Definition 2.2. *A* route p *between two nodes* u, v *is a sequence of non-repeating links* $(l_1, \dots, l_n)$, $l_i \neq l_j$, $\forall i \neq j$, *such that* $b(l_1) = u$, $a(l_n) = v$ *and* $a(l_i) = b(l_{i+1})$, $i = 1, \dots, n-1$.

Let $\mathcal{P}_{(u,v)}$ be the set of all routes connecting u to v. The multi-digraph is also equipped with a set of *origin-destination pairs* (*OD pairs*) $\mathcal{K} = \{(u, v) : u, v \in \mathcal{N}\} \subseteq \mathcal{N} \times \mathcal{N}$. Given an OD pair k, $\mathcal{P}_k$ is the set of all routes connecting its origin to its destination. It is rather straightforward how to interpret a multi-digraph as

a traffic network. The links of the multi-digraph stand for the roads of the network and nodes for junctions between them. Routes between a pair of nodes are all the possible routes connecting the starting node to the end node. OD pairs account for those pairs of nodes subject to an exogenous traffic demands aiming to go from the origin to the destination by crossing the network. From now on, we will always refer to a multi-digraph as *network*. In this thesis, special attention will be devoted to the following classes of networks.

Definition 2.3. *A* two-terminal network *is a network with a single OD pair* (o, d).

Definition 2.4. *A network* $\mathcal{G}$ *is a* parallel network *if its routes are all parallel, i.e., each of its links belongs to one route only.*

2.2 Non-atomic routing games

Non-atomic routing games provide a powerful framework for analyzing traffic networks, specifically for identifying traffic patterns that emerge from the interactions of strategic network users. Users adapt their routing strategy (which route to take) based on their knowledge of the current traffic state in the network. NRGs are macroscopic traffic models. As the identifier *non-atomic* suggests, the number of users is assumed to be very large, so that each user controls only a negligible fraction of traffic, and traffic is modeled in an aggregate continuous manner by link flows. Also, user interaction is *anonymous*, i.e., these games only account for the fractions of users using each route, rather than who is choosing what.

2.2.1 Homogenous NRGs

Homogeneous NRGs, also known as selfish routing games [1], represent the foundation and the simplest instance of NRGs. In this setting, users are modeled as selfish agents seeking to minimize their own travel time. Consider a network $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, with set of OD pairs $\mathcal{K}$. Each OD pair k is subject to fixed non-negative traffic demand Φ_k. Again, each traffic demand consists of a very large number of users, each of them corresponding to an infinitesimal part of the total demand. Let also $\mathcal{P} = \cup_{i=1}^{|\mathcal{K}|} \mathcal{P}_k$ be the set of all routes on $\mathcal{G}$. The way traffic is distributed over the network is described by the *route flow vectors* $z \in \mathbb{R}_{\geq 0}^{|\mathcal{P}|}$. Given $p \in \mathcal{P}_k$, z_p allocated on route p. The set of feasible flows is

$$\mathcal{Z} := \left\{ z \in \mathbb{R}_{\geq 0}^{|\mathcal{P}|} \, : \, \sum_{p \in \mathcal{P}_k} z_p = \Phi_k \right\}.$$

After defining the *link-route incidence matrix* $A \in \{0,1\}^{|\mathcal{L}| \times |\mathcal{P}}$ as

$$A_{lp} = \begin{cases} 1 & \text{if } l \in p \\ 0 & \text{otherwise} \end{cases},$$

it is possible to define the *link flow vectors* $f \in \mathbb{R}_{\geq 0}^{|\mathcal{L}|}$ as $f = Az$. These vectors quantify the link flow on each link. Each link is also characterized by a *travel time*

$\tau_l : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, which is assumed to be continuous, non-negative and increasing functions of the link flow f_l. Typically, link travel times are *separable*, i.e., the travel time on link l depends on the flow on link l only. From the definition of link travel time, it easily follows that of *route travel time* $p \in \mathcal{P}$, which is nothing but the sum of all travel times τ_l such that $l \in p$. The interest of homogeneous NRGs is to determine equilibrium patterns emerging from the interaction of selfish users.

Definition 2.5 (Wardrop equilibrium (WE))**.** *A route flow $z^W \in \mathcal{Z}$ is called an* WE *for the heterogeneous NRG if*

$$\forall k \in \mathcal{K}, \quad \forall p, q \in \mathcal{P}_k, \qquad z_p^W > 0 \;\Rightarrow\; \tau_p(f^W) \leq \tau_q(f^W).$$

In words, a WE is route flow vector such that users choose only minimum travel time routes and none of them has an incentive to modify its choice. The WEs admit an alternative characterization as the solutions to the following optimization problem:

$$\begin{aligned} \min_{z} \quad & \sum_{l \in \mathcal{L}} \int_0^{f_l} \tau_l(r)\, dr \\ \text{s.t.} \quad & \sum_{p \in \mathcal{P}_k} z_p = \Phi_k, \;\; \forall k \\ & f_l = \sum_{p\,:\, l \in p} A_{lp} z_p \\ & z_p \geq 0, \;\; \forall p \in \mathcal{P}. \end{aligned} \tag{2.1}$$

The objective function in (2.1) is a Rosenthal-like potential and it is called the *Beckmann transformation* [46]. To see that the solutions to Problem (2.1) satisfy to the definition of WE, it suffices to derive the Kuhn-Tucker conditions associated to the problem [47]. Because of the assumptions on the link travel times, (2.1) is a convex separable optimization problem, it is ensured that a homogeneous NRG always admits a WE, and if the link travel time functions are strictly increasing, then there exists a unique link equilibrium flow f^W.

One of the main concerns with WEs is assessing their efficiency with respect to a measure that a hypothetical network manager is interested in optimizing, which also depends on the network's congestion. This measure typically corresponds to the *total travel time* realized at equilibrium:

$$S(z) = \sum_{p \in \mathcal{P}} z_p \cdot \tau_p(z) = \sum_{l \in \mathcal{L}} f_l \cdot \tau_l(f_l).$$

A route flow minimizing $S(z)$ is called *system optimum* (SO). The SO can be retrieved as the solution to the following optimization problem:

$$\begin{aligned} \min_{z} \quad & \sum_{l \in \mathcal{L}} f_l \cdot \tau_l(f_l) \\ \text{s.t.} \quad & \sum_{p \in \mathcal{P}_k} z_p = \Phi_k, \;\; \forall k \\ & f_l = \sum_{p\,:\, l \in p} A_{lp} z_p \\ & z_p \geq 0, \;\; \forall p \in \mathcal{P}. \end{aligned} \tag{2.2}$$

The efficiency of a WE is then assessed by calculating its *Price of anarchy* (PoA) [1, 48], which corresponds to the ratio between the total travel time attained by the WE and the minimum attained by the SO:

$$\mathrm{PoA}(z^W) = \frac{S(z^W)}{\min_{y \in \mathcal{Z}} S(y)}.$$

While homogeneous NRGs have proven useful in studying road traffic, they have certain limitations that subsequent works have attempted to address. Among these, the homogeneity among network users has garnered the most attention. The next section will focus on heterogeneous games. The literature also includes refinements of the model that remove other limiting assumptions, such as the separability of travel time functions. However, this goes beyond the scope of the current discussion, and we refer interested readers to the following references [49, 50].

2.2.2 Heterogeneous NRGs

The heterogeneous routing framework was initially introduced in [2]. In this case, traffic consists of multiple classes of users, each of them associated with a different set of travel time functions. Users are still assumed to behave selfishly, according to the travel time functions associated with the user class it belongs to. Heterogeneity allows to model more complex scenarios, where users have diverse or limited knowledge of the network or of the traffic conditions, or where users have different preferences beyond travel time minimization, e.g., monetary tolls. In the following, we define the setting proposed in [2].

Consider the same network $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, with set of OD pairs $\mathcal{K}$, defined for homogeneous NRGs, only this time each OD pair k is subject to fixed non-negative traffic demands from I different classes of users. Let $\Phi_k^i \geq 0$ be the traffic demand of each class. The way traffic is distributed over the network is now described by class-specific *route flow vectors* $z^i \in \mathbb{R}_{\geq 0}^{|\mathcal{P}|}$. Given $p \in \mathcal{P}_k$, z_p^i allocated on route p associated with class i. For every class, the set of feasible flows is

$$\mathcal{Z}^i := \left\{ z^i \in \mathbb{R}_{\geq 0}^{|\mathcal{P}|} \,:\, \sum_{p \in \mathcal{P}_k} z_p^i = \Phi_k^i \right\}.$$

Let $f^i \in \mathbb{R}_{\geq 0}^{|\mathcal{L}|}$, $f^i = Az^i$ be the class-specific *link flow vectors*. This vectors gives information on what is the link flow which each link of network is subject to for each class. Let z and f be the concatenation of class route flows and link flows, respectively, and let $Z = \sum_{i \in \mathcal{I}} z^i$, $F = \sum_{i \in \mathcal{I}} f^i$. Finally, each class is characterized by a set of *link travel times* $\tau_l^i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, which, again, are assumed to be continuous, non-negative, increasing and separable functions of link flow F_l. Notice that congestion is not class-specific, meaning that any class affect the link travel time in the same way as the others. The *route travel time* $p \in \mathcal{P}$ for class i, τ_p^i, is the sum of all the travel time functions τ_l^i such that $l \in p$.

For heterogeneous NRGs, WEs are defined as follows:

Definition 2.6 (Wardrop equilibrium)**.** *A route flow* $z^W \in \mathcal{Z}$ *is called a* WE *for the heterogeneous NRG if*

$$\forall k \in \mathcal{K},\ \forall i \in I,\ \forall p, q \in \mathcal{P}_k, \qquad (z^W)^i_p > 0 \ \Rightarrow\ \tau^i_p(F^W) \leq \tau^i_q(F^W).$$

In words, for all classes, all used routes share the same travel time, which is the minimum travel time for that OD pair. Heterogeneous NRG can be seen as I-person games, where each user corresponds to a specific class and aims to minimize the following cost function:

$$U^i(f) = \sum_{l \in \mathcal{L}} \int_0^{f^i_l} \tau^i_l(r + f^{-i}_l)\, dr, \quad f^{-i} := \sum_{j \neq i} f^j_l, \qquad i \in I.$$

Then, the I-person game correspond to the following interconnection of optimization problems:

$$\begin{aligned} \min_{z^i} \quad & \sum_{l \in \mathcal{L}} \int_0^{f^i_l} \tau^i_l(r + f^{-i})\, dr \\ \text{s.t.} \quad & \sum_{p \in \mathcal{P}_k} z^i_p = \Phi^i_k,\ \forall k \\ & f^i_l = \sum_{p\,:\,l \in p} A_{lp} z^i_p \\ & z^i_p \geq 0,\ \forall p \in \mathcal{P}. \end{aligned} \quad , \quad i = 1, \ldots, I. \tag{2.3}$$

Again, to see that the solutions to Problem (2.3) satisfy to the definition of WE, it suffices to derive the Kuhn-Tucker conditions associated to the problem. Although existence is guaranteed also for heterogeneous NRGs, uniqueness does not, even when user travel times are strictly increasing, in general [28, 51].

2.2.3 Mixed behavior NRGs

The heterogeneous setting inspired the definition of a more general problem, where some of classes consists of users that, instead of adopting a selfish behavior, act *coordinately* to achieve a common goal, e.g., minimizing the class total travel time. Coordination within classes of vehicles was initially introduced in [8]. Let $C \subseteq I$ indicate the set of coordinated classes. Resorting once again the a I-person game formulation, the cost function of users in C takes the form

$$U^i(f) = \sum_{l \in \mathcal{L}} f^i_l \cdot \tau^i_l(f^i_l + f^{-i}_l).$$

The expression *mixed behavior* aims to highlight the presence of both selfish and coordinated behavioral instances. The equilibrium flows of a mixed behavior NRG can be defined as follows.

Definition 2.7 (Mixed equilibrium (ME))**.** *A route flow $z^* \in \mathcal{Z}$ is called an* ME *for the mixed behavior NRG if*

$$\begin{aligned} z^{i*} = \min_{z^i} \quad & \sum_{l\in\mathcal{L}} \int_0^{f_l^i} \tau_l^i(r + f^{-i})\, dr \\ s.t. \quad & \sum_{p\in\mathcal{P}_k} z_p^i = \Phi_k^i, \ \forall k \\ & f_l^i = \sum_{p\,:\,l\in p} A_{lp} z_p^i \\ & z_p^i \geq 0, \ \forall p \in \mathcal{P}. \end{aligned} \quad , \quad i \in I \setminus C; \tag{2.4}$$

$$\begin{aligned} z^{i*} = \min_{z^i} \quad & \sum_{l\in\mathcal{L}} f_l^i \cdot \tau_l^i(f_l^i + f_l^{-i}) \\ s.t. \quad & \sum_{p\in\mathcal{P}_k} z_p^i = \Phi_k^i, \ \forall k \\ & f_l^i = \sum_{p\,:\,l\in p} A_{lp} z_p^i \\ & z_p^i \geq 0, \ \forall p \in \mathcal{P}. \end{aligned} \quad , \quad i \in C. \tag{2.5}$$

If existence is guaranteed after simply imposing conditions to guarantee that the optimization problem associated with each user is convex, uniqueness requires more tight conditions on the travel time functions, in general. Existence and uniqueness conditions for mixed behavior NRGs where first studied in [8], and then further explored in later works [38, 52].

2.2.4 Bayesian routing games

In *Bayesian routing games*, the network is characterized by a state described by a random variable s, aiming to reflect conditions of the network, which takes values on a probability space $(\mathcal{S}, \mathcal{A}, \mathbb{P})$. Each edge $l \in \mathcal{L}$ has a state-dependent travel time function $\tau_l^s(\cdot)$, which is a positive, increasing, and differentiable function of the link flow. For simplicity, we consider the case of two-terminal networks.

The network serves an exogenous flow Φ of non-atomic users. All travelers receive information from a navigation app, which provides a signal $\sigma : \mathcal{S} \to \Delta(\mathcal{P})$ that, for each network state, provides the portion of exogenous flow recommended to take each route [53]. As the noisy signal and the prior distribution of the network state are known to users, users are then able to form beliefs about the network state s based on the received signal σ_p, according to Bayes' formula:

$$d\mathbb{P}_p(s) = \frac{\sigma_p(s) d\mathbb{P}(s)}{\int_{\mathcal{S}} \sigma_p(r) d\mathbb{P}(r)}.$$

Let $\mathbb{E}_p(\cdot) = \int_{\mathcal{S}} \cdot\, d\mathbb{P}_p(r)$ indicate the expected value after receiving recommendation p. Users are provided with a recommendation each, but they follow the recommendation only if they think it is a best response according to their belief. Let

$c \in \mathbb{R}_{\geq 0}^{|\mathcal{P}|} \times \mathbb{R}_{\geq 0}^{|\mathcal{P}|}$ be a matrix such that each entry c_{pq} represents the fraction of users that received recommendation p but take route q. Then, in this case the route and link flow vectors take the following form:

$$z^{\sigma,c}(s) = c'\sigma(s), \qquad f^{\sigma,c}(s) = Az^{\sigma,c}(s).$$

Definition 2.8 (Bayesian Wardrop Equilibrium (BWE))**.** *Given a signal σ, a route flow vector $z^{\sigma,c}(s) = c'\sigma(s)$ is said to be a* BWE *if*

$$\sigma_p(s)c_{pq} > 0 \;\Rightarrow\; \mathbb{E}_p(\tau_q^s(f^{\sigma,c}(s))) \leq \mathbb{E}_p(\tau_k^s(f^{\sigma,c}(s))) \quad \forall p, q, k \in \mathcal{P}.$$

Bayesian routing games model user decision-making under uncertain network conditions. This approach is particularly useful when considering the accuracy of information provided to app users, addressing scenarios with incomplete information. However, for this specific thesis, the focus was on other aspects, and incomplete information was not the main area of investigation.

2.3 Cell Transmission Model

Daganzo's Cell transmission Model was introduced in [3, 4]. For a given stretch of road divided into N cells of length L_n, the model describes the evolution of the traffic density in each cell as the difference between the cell's inflow and outflow:

$$\dot{x}_n(t) = f_n^{\text{in}}(t) - f_n^{\text{out}}(t).$$

By connecting in series such that $f_n^{\text{out}}(t) = f_{n+1}^{\text{in}}(t)$, one recovers the following equation for the evolution of the traffic density on cell n:

$$\dot{x}_n(t) = \frac{1}{L_n}\left(f_{n-1}(t) - f_n(t)\right),$$

where now $f_n(t)$ is the flow from cell n to cell $n+1$. The flow $f_n(t)$ is determined as the minimum between the demand of cell n and the supply of cell $n+1$:

$$s_l(x_l) = \min\{\overline{f}_l, w_l(\overline{x}_l - x_l)\}, \tag{2.6}$$

$$d_l(x_l) = \min\{v_l x_l, \overline{f}_l\}, \tag{2.7}$$

where $\overline{f}_l$ stands for the capacity of the cell, i.e., the maximum flow that can flow through the cell. Supply and demand functions represent the maximum possible inflow and outflow for cell n when the density equals x_n, respectively.

2.4 Dynamical network flows

Dynamical network flows are inspired by compartmental models [30, 31]. The idea is to think of the links of a traffic network as a set of interconnected cells exchanging mass with each other. The flows from one cell to another are determined by a *supply and demand* mechanism, inspired by the CTM.

Consider a network topology similar to the one defined in Section 2.1, but this time with only one OD pair, for simplicity. Let the origin node be equipped with an on-ramp o, through which the travel demand $\Phi(t)$ has access to the network $\mathcal{G}$. Let also $\mathcal{L}^{\text{out}}$ be the set of links such that $b(l)$ is the destination node of the network. In addition, each link of the network is now characterized by a capacity $\overline{f}_l$. Let $x_l(t)$ represent the traffic density of link l at time t. By letting $x \in \mathbb{R}_{\geq 0}^{|\mathcal{L}|+|\mathcal{O}|}$ represent the state of the traffic system, then its rate of change is determined by the following ODE:

$$\dot{x}_l(t) = \frac{1}{L_l}\left(f_l^{\text{in}}(x(t)) - f_l^{\text{out}}(x(t))\right), \quad l \in \mathcal{L}.$$

Each link l is associated with a demand function $d_l(x_l)$ that represents the ideal outflow from the link. This function is Lipschitz-continuous, non-decreasing, and concave, with $d_l(0) = 0$. This implies that the function is smooth, the outflow does not decrease as the density increases, and it exhibits diminishing returns, meaning the rate of increase of the outflow slows down as x_l increases. Additionally, the function has an upper bound for the outflow, denoted by $\overline{f}_l$:

$$\sup_{x_l \geq 0} d_l(x_l) = \overline{f}_l.$$

This means that regardless of how large x_l becomes, the outflow cannot exceed $\overline{f}_l$. For the on-ramps $o \in \mathcal{O}$, $d_o(t) = \Phi(t)$. Each link l is also associated with a supply function $s_l(x_l)$, standing for the maximum flow that can enter the link when its density equals x_l. Supply functions are designed to capture spill-back effects and back-propagation of congestion. These functions are chosen to be Lipschitz-continuous, non-increasing, concave, upper-bounded the link capacity $\overline{f}_l$ and such that $s_l(\overline{x}_l) = 0$. Typically, supply and demand are chosen so as to satisfy to $s_l(x_l^c) = d_l(x_l^c) = \overline{f}_l$.

The splitting onto downstream nodes at each diverge non-destination node occurs according to routing ratios $R_{jl}(x(t)) \in \mathbb{R}_{\geq 0}$. These can be either fixed quantities or time-varying feedback control variables (either static or dynamic), evolving according to a specific routing policy. Routing ratios provide significant flexibility in designing the routing policy according to specific modeling preferences and needs.

Now, the expression of the inflow and outflow terms of each link is

$$f_l^{\text{in}} = \begin{cases} \Phi(t), & l = o \\ \sum_{j \in a(l)^+} f_{jl}(x), & l \in \mathcal{L} \end{cases},$$

$$f_l^{\text{out}} = \begin{cases} d_l(x_l), & l \in \mathcal{L}^{\text{out}} \\ \sum_{j \in a(l)^-} f_{lj}(x), & l \in \mathcal{L} \end{cases}.$$

The function $f_{jl}(x)$ represents the flow from link j directed toward link l, and takes the following form:

$$f_{jl}(x) = \gamma_l(x) R_{jl}(x) d_j(x_j).$$

The terms $\gamma_l(x)$ account for supply constraints, and they are designed so as to guarantee that

$$\sum_{j \in a(l)^+} \gamma_l(x_l) R_{jl}(x) d_j(x_j) \leq s_l(x_l), \quad \forall t \geq 0, \quad \forall l \in \mathcal{L},$$

by both implementing both FIFO or non-FIFO rules (see [29] for more details). For example, if one were to choose a non-FIFO rule, wanting to capture the fact that the supply constraint of a link do not affect the flow directed toward other links, then $\gamma_l(x)$ can be written as

$$\gamma_l(x) = \min\left\{1, \frac{s_l(x_l)}{\sum_{j\in a(l)^+} R_{jl}(x) d_j(x_j)}\right\},$$

so that

$$f_{jl}(x) = \min\left\{R_{jl}(x)d_j(x_j), \frac{R_{jl}(x)d_j(x_j)}{\sum_{j\in a(l)^+} R_{jl}(x)d_j(x_j)} s_l(x_l)\right\}.$$

From the brief description above, it is evident that dynamical network flows offer several key advantages. Firstly, the traffic system is modeled using a ODE system, which is even autonomous when the traffic throughput vector is constant. Plenty of tools are available to perform qualitative analysis of its asymptotic behavior. Secondly, the incorporation of supply and demand constraints allows for the imposition of capacity limitations, enabling to capture of significant traffic phenomena such as spill-backs, thereby enhancing the realism of traffic dynamics. Lastly, routing ratios provide a direct and highly flexible way to model the routing dynamics at non-destination nodes.

2.4.1 Alternative dynamical network flows

The literature also provides models that offer a more faithful description of traffic dynamics by defining a law that governs traffic propagation on a link. This can be achieved through models based on ODE systems [54, 55] as well as PDE systems [56, 57]. Briefly, in the case of ODE systems, this is done by establishing a temporal link between incoming and outgoing flows of a link:

$$\begin{cases} \dot{x}_l(t) = f_l^{\text{in}}(t) - f_l^{\text{out}}(t), \\ \tau(t) = \tau(x(t), t), \\ f_l^{\text{out}}(t + \tau(t)) = \frac{f_l^{\text{in}}(t)}{1+\dot{\tau}(t)}. \end{cases}$$

PDE models provide a much more detailed description of the traffic dynamics, allowing to describe the variation in the traffic density on a link not only over time, but also at each position s of the road. This is done by resorting to the well-know Lighthill-Whitman-Richards (LWR) model [58, 59]:

$$\partial_t x(t,s) + \partial_s \varphi(x(t,s)) = 0.$$

The function φ is the called the *flux* function and it is taken as the product of the traffic density and the traffic speed, i.e., $\varphi = x \cdot v(x)$. The traffic speed is a function of the traffic density and it is chosen so that the relationship between the traffic density and the flux function is consistent with the *fundamental diagram of traffic* [5]. Notice that The CTM can be seen as a space discretization of the LWR [60].A even more accurate modeling of the traffic dynamics can be done by using the Aw-Rascle-Zhang (ARZ) model [61], which exploits second-order PDEs.

While these models offer advantages over dynamic network flows (Section 2.4), they also present challenges that have led us to refrain from using them. First, it is difficult for these models to complete a qualitative analysis of their dynamic behavior. Second, in both cases, routing is derived from defining an optimal control problem implying the minimization over a certain time horizon of a cost functional or through the imposition of time-dependent Wardrop conditions in the form of variational inequalities. Although this solution is well suited for modeling certain scenarios, it certainly represents a less flexible solution than routing ratios. Finally, in the case of models using ODEs, it is not clear how to include supply and demand mechanisms. Some models include a point-queue model at non-destination nodes, which involve increasing a buffer whenever the outgoing flow exceeds the link capacity. However, this approach still has limitations, as the accumulation of vehicles in the buffer does not alter traffic dynamics on the link.

Part II

Selfish routing & navigation apps

Chapter 3

Selfish routing on networks with supply and demand constraints

3.1 Introduction

Standard non-atomic routing games [1, 2] have proven useful in understanding various aspects of traffic networks. Nevertheless, their basic formulation presents some shortcomings when addressing traffic networks. Firstly, networks links have no capacity constraints limiting the amount of flow that they can accommodate. This does not allow for properly capturing congestion phenomena typical of traffic networks. Additionally, link costs are generally modeled as increasing functions of flow. As link costs typically represent link travel times in most applications to road networks, this is not consistent with traffic modeling, according to which the relationship between traffic flow and travel time is non-monotonic.

In this chapter, we propose a novel type of non-atomic routing game by leveraging concepts from Daganzo's CTM [3, 4] to define a network structure. The key innovation lies in incorporating density variables alongside traditional traffic flows. By considering both flows (number of vehicles per unit time) and densities (number of vehicles per unit length) on each link, we can define a supply and demand mechanism that enforces capacity constraints. This mechanism limits the flow that can traverse a link based on its current density, essentially allowing us to identify congested sections. Furthermore, travel times on each link become directly dependent on its density, consistently with traffic modeling [5]. This combined approach provides a more accurate representation of real-world traffic phenomena.

This model allows for identifying a critical consequence of selfish routing that goes beyond the well-known problem of reduced traffic efficiency due to increased total travel time: *partially transferring WE*. Specifically, in some cases, the WE resulting from the interaction of selfish users consists in a congestion pattern that allows only a part of the exogenous flow which the network is subject to enter the network and cross it.

3.1.1 Summary of results

First, we propose a novel type of selfish routing game that is more suitable for modeling road networks based on the CTM. In this model, links are treated as cells with capacity constraints that depend on the density within the cell, and link travel times are increasing functions of density rather than flow. We focus on *parallel networks*, where the network consists of N parallel routes, each composed of multiple links. Second, we characterize the WE and the SO of this game and prove their essential uniqueness. Finally, we introduce the concept of partially transferring WE and demonstrate that under certain conditions, the unique WE of the game can be partially transferring, even when the exogenous demand on the network is less than the min-cut capacity. This provides further evidence of the inefficiency of selfish routing.

3.1.2 Related work

To the best of the authors' knowledge, non-cooperative routing games accounting for capacity constraints and congested traffic regimes have only been proposed in [62], so far. In this work, the authors analyze a Stackelberg routing game on a parallel network, where a central authority can control a fraction of the total traffic demand to improve the total cost on the network, thus improving efficiency. From the point of view of the network design, their model differs from ours in that it does not entail any supply and demand mechanism. Although their model anticipates the possibility that a WE might not be able to fully satisfy the exogenous traffic flow, the absence of a supply and demand mechanism does not allow for the characterization of partially transferring equilibria. Their analysis is therefore limited to the case where the exogenous flow is guaranteed to admit fully transferring WEs.

Capacitated networks cannot accommodate any exogenous flow. It is a well-known fact that for networks with capacitated links, assuming that the link capacity does not depend on the flow traversing the link, the maximum exogenous flow that can be successfully transferred through the network equals its min-cut capacity [63, 64]. Specifically, there exists a flow allocation that transfers this maximum flow. The problem of identifying routing policies that prompt fully transferring flow allocations has received significant attention in the last years, mostly through dynamical models [65, 66, 67, 68, 69, 70]. In particular, in [68] the authors study the behavior of a dynamical network flow model governed by a distributed local routing policies allowed to depend on the network state. These policies are characterized by the routing decisions at each non-destination node being made independently based only on the state of incoming and outgoing links, without considering the state of other nodes in the network. Nonetheless, capacity constraints are only applied at the exits of the links, allowing any amount of flow to enter a link. It is shown that if the exogenous flow which the network is subject to does not exceed the min-cut capacity, then the class of monotone distributed routing policies ensures that the system globally asymptotically converges to a state where the flow is fully transferred. In [69, 71], the authors propose a dynamical network flow model encompassing the CTM with fixed routing policies (not necessarily fully transferring) at non-destination nodes. The main results concern convergence to equilibria. In [69], the

authors develop a ramp metering control strategy for maximizing the transferred flow.

3.1.3 Chapter organization

Section 3.2 delves into the details of the proposed network structure. Here, we describe the mechanism for supply and demand on each link, define what constitutes valid traffic assignments, and introduce factors influencing travel time based on link density. Section 3.3 establishes a non-atomic selfish routing game on the network. We comprehensively analyze the Wardrop equilibria and social optima of this game, identifying necessary and sufficient conditions for the occurrence of partially transferring Wardrop equilibria. In Section 3.4, we provide an example showing that the problem of partial demand transfer occurs also in more complex network topologies. Finally, Section 3.5 concludes the chapter with some closing remarks

3.2 Network modeling

We consider a parallel network consisting of a single OD pair and N distinct non-intersecting routes connecting them. Each route p is composed of n_p links. The network is subject to a constant positive exogenous flow of vehicles Φ that distributes among its routes. In the following, we describe the functioning of each network link in relation to the traffic density within it.

3.2.1 Characterization of the network links

Given a link $l \in \mathcal{L}$, let x_l and f_l indicate its density, corresponding to the number of vehicles per unit of length, and its flow, corresponding to the number of vehicles per unit of time. Let $\overline{x}_l$ (veh/km), $\overline{f}_l$ (veh/h), v_l (km/h) and L_l (km) be positive finite constants representing the *jam density* (maximum attainable density), the *capacity* (maximum attainable flow), the *free-flow speed* and the *length* of the link. Now, associate with each link supply and demand functions $s_l(x_l)$, $d_l(x_l)$, depending on the link density. Supply and demand functions are inspired by Daganzo's fundamental diagram [5] and take the following form:

$$s_l(x_l) = \min\{\overline{f}_l, w_l(\overline{x}_l - x_l)\}, \tag{3.1}$$

$$d_l(x_l) = \min\{v_l x_l, \overline{f}_l\}. \tag{3.2}$$

Both functions are continuous and piece-wise linear. The supply function is non-increasing with density, reflecting the fact that as more vehicles occupy the link, the fewer additional vehicles the link can accommodate. In contrast, the demand function is non-decreasing, meaning that as more vehicles are on the link, the higher the number of vehicles aiming to leave it. If we define the *critical density* of a link as $x_l^c := \overline{f}_l / v_l$, then $w_l = \overline{f}_l / (\overline{x}_l - x_l^c)$, so as to guarantee

$$s_l(x_l^c) = d_l(x_l^c) = \overline{f}_l.$$

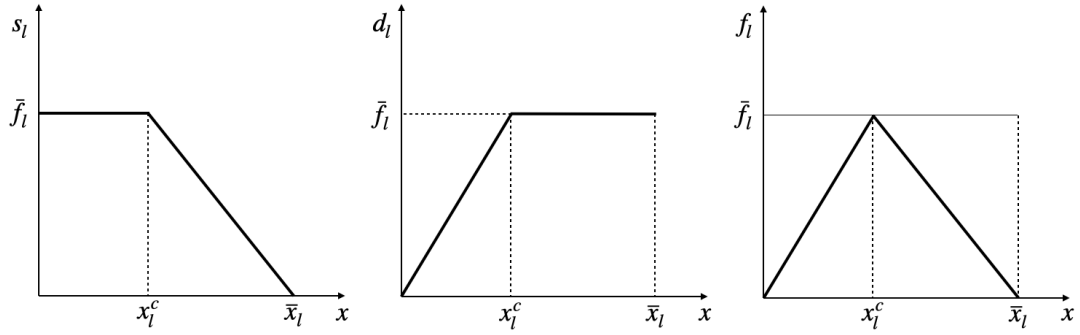

Figure 3.1: From left to right: supply function as in (3.1), demand function as in (3.2), Daganzo's fundamental diagram of traffic.

When $x_l \leq x_l^c$, we say that the link is in *free-flow* or *free-flow regime*, whereas if $x_l > x_l^c$, the link is said to be *congested* or in *congested regime*.

In the following section, we characterize the feasible density and flow vectors for a network whose links exhibit such a supply and demand mechanism.

3.2.2 Traffic assignments

Definition 3.1. *A vector* $R = (R_1, \dots, R_N) \in \mathbb{R}^N_{\geq 0}$ *such that* $\sum_{i=1}^N R_i = 1$ *is called a* routing vector. *Each element of* R *is called* routing ratio.

The supply and demand functions determine the exchange flow at the interface between contiguous cells. Let $x \in \mathbb{R}^{n_1+\cdots+n_N}_{\geq 0}$, $f \in \mathbb{R}^{p+n_1+\cdots+n_N}_{\geq 0}$ be the *density* and *flow vectors*, respectively. Consider a route p and two of its consecutive links, l_p and $(l+1)_p$. Then, the inflow from link l_p to link $(l+1)_p$ is

$$f_{l_p}(x) = \min\{d_{l_p}(x_{l_p}), s_{(l+1)_p}(x_{(l+1)_p})\}. \tag{3.3}$$

The inflow of the first link of a route is

$$f_{0_p}(x) = \min\{\Phi R_p, s_{0_p}(x_{0_p})\}, \tag{3.4}$$

Finally, since the final link of each route is not connected to any other link,

$$f_{n_p}(x) = d_{n_p}(x_{n_p}). \tag{3.5}$$

With abuse of notation, we will indicate the density and flow vectors associated with route p as $x_p \in \mathbb{R}^{n_p}_{\geq 0}$, $f_p \in \mathbb{R}^{n_p+1}_{\geq 0}$.

Given a routing vector, we are interested in identifying all the density vectors associated with it.

Definition 3.2. *Given a routing vector* R, *a* consistent density vector x^R *is a density vector satisfying to*

$$f_{0_p}(x^R) = f_{1_p}(x^R) = \cdots = f_{n_p}(x^R), \quad p = 1, \dots, N. \tag{3.6}$$

Let $C(R)$ *be the set of all consistent density vectors associated with* R.

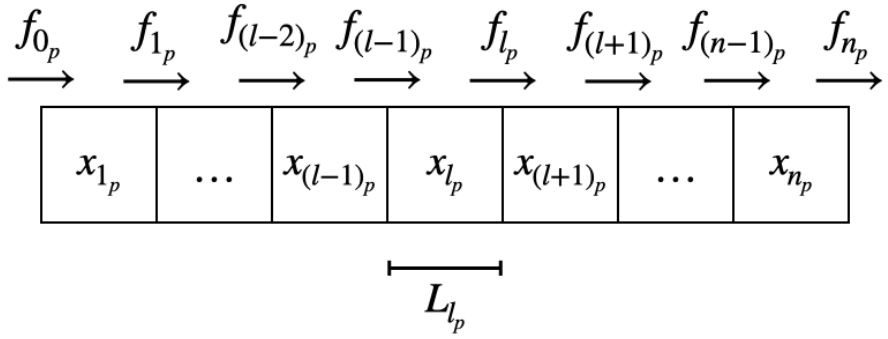

Figure 3.2: Structure of route p.

Consistent density vectors are density vectors such that the inflow and the outflow of each link correspond and this flow is equal for all the links of the route.

Definition 3.3. *A* traffic assignment *is a pair (R, x^R), where R is routing vector and x^R is a consistent density vector of it.*

Depending on the routing vector, the associated consistent density vectors might be characterized by some congested links or not.

Definition 3.4. *The* capacity of route p, $\overline{z}_p$, *is the capacity of the route's lowest capacity link:*

$$\overline{z}_p := \min_{l \in p} \overline{f}_l. \tag{3.7}$$

Given an exogenous flow Φ and a routing vector R, consider the following sets:

$$\begin{aligned} P_F &= \{p \in \{1, \dots, N\} |\, \Phi R_p < \overline{z}_p\}, \\ P_C &= \{p \in \{1, \dots, N\} |\, \Phi R_p = \overline{z}_p\}, \\ P_S &= \{p \in \{1, \dots, N\} |\, \Phi R_p > \overline{z}_p\}. \end{aligned} \tag{3.8}$$

The set P_F consists of the routes assigned a fraction of exogenous flow smaller than their capacity. The set P_C consists of the routes assigned a fraction of exogenous flow equal to their capacity. Finally, The set P_S consists of the routes assigned a fraction of exogenous flow exceeding their capacity. Then, let us discuss about the shape of the consistent density vectors for a given routing vector.

To ease the discussion, we assume that each route has a unique link of minimum capacity.

Assumption 3.1. *Route p has a unique minimum capacity link $b_p \in \{1_p, \dots, n_p\}$, $p = 1, \dots, N$.*

This assumption, while simplifying the model, will not undermine the relevance of the findings presented in the following sections. Instead, it allows us to focus on specific aspects of the problem and draw conclusions that are still applicable to a wide range of scenarios.

Proposition 3.1. *Let Assumption 3.1 hold. Then, $\forall p \in P_F$, there exists a unique consistent density vector for route p which is as follows:*

$$\begin{aligned} x^R_{l_p} &= \frac{\Phi R_p}{v_{l_p}}, && l_p = 1_p, \dots, n_p, \\ f_{l_p}(x^R) &= \Phi R_p, && l_p = 1_p, \dots, n_p; \end{aligned} \tag{3.9}$$

$\forall p \in P_S$, there exists a unique consistent density vector for route p which is as follows:

$$\begin{aligned} x^R_{l_p} &= \overline{x}_{l_p} - \frac{\overline{z}_p}{w_{l_p}}, && l_p = 1_p, \dots, (b-1)_p \\ x^R_{l_p} &= \frac{\overline{z}_p}{v_{l_p}}, && l_p = b_p, \dots, n_p, \\ f_{l_p}(x^R) &= \overline{z}_p, && l_p = 1_p, \dots, n_p; \end{aligned} \tag{3.10}$$

Finally, $\forall p \in P_C$, a consistent density vector on p is any vector such that, given $k_p \in \{1, \dots, (b-1)_p\}$:

$$\begin{aligned} x^R_{l_p} &= \frac{\overline{z}_p}{v_{l_p}}, && l_p = 1_p, \dots, (k-1)_p; \\ x^R_{k_p} &\in \left[x^c_{k_p}, \overline{x}_{l_q} - \frac{\overline{z}_p}{w_{l_p}} \right], \\ x^R_{l_p} &= \overline{x}_{l_p} - \frac{\overline{z}_p}{w_{l_p}}, && l_p = (k+1)_p, \dots, (b-1)_p, \\ x^R_{l_p} &= \frac{\overline{z}_p}{v_{l_p}}, && l_p = b_p, \dots, n_p, \\ f_{l_p}(x^R) &= \overline{z}_p, && l_p = 1_p, \dots, n_p. \end{aligned} \tag{3.11}$$

Remark 3.1. *It follows from (3.1) and (3.2) that, for any link l_p,*

$$x^R_{l_p} = \frac{\Phi R_p}{v_{l_p}} < x^c_{l_p}, \qquad x^R_{l_p} = \overline{x}_{l_p} - \frac{\overline{z}_p}{w_{l_p}} > x^c_{l_p}.$$

This means that if $p \in P_F$, then all links of route p are in free-flow regime. Contrarily, if $p \in P_S$, then the first $b-1$ links of route p are in congested regime. Finally, if $p \in P_C$, depending on the value of k_p, the route might present links with congested regime or not, extending backward from the least capacity link to the origin.

Proof. The proof is split into three parts, each for one of the sets P_F, P_C and P_S.

1. Consider a route $p \in P_F$. Suppose that $f_{0_p} = \Phi R_p$. Then, from (3.6), $f_{l_p} = \Phi R_p$, $l_p = 1_p, \dots, n_p$. This implies that

   $$d_{l_p}(x_{l_p}) = \Phi R_p \quad \text{or} \quad s_{(l+1)_p}(x_{(l+1)_p}) = \Phi R_p.$$

   Suppose that $f_{l_p} = d_{l_p}$, $l_p = 1_p, \dots, n_p$. Then, it is straightforward that (3.9) is the only possible density vector with this form satisfying to (3.6). Suppose now that there exists $k_p \in p$ such that $f_{k_p}(x_p) = s_{(k+1)_p}(x_{(k+1)_p})$. Since

   $$f_{k_p}(x_p) = \Phi R_p < \overline{f}_{(k+1)_p},$$

it must be that

$$x_{(k+1)_p} > x^c_{(k+1)_p},$$

so that

$$s_{(k+1)_p}(x_{(k+1)_p}) < d_{(k+1)_p}(x_{(k+1)_p}).$$

From (3.6), the last inequality implies that $f_{(k+1)_p}(x_p) = s_{(k+2)_p}(x_{(k+2)_p})$. The same argument can be applied inductively to the subsequent route links, up to the final link of the route. Nevertheless, since $f_{(n-1)_p}(x_p) = s_{n_p}(x_{n_p})$, then $x_{n_p} > x^c_{n_p}$, which in turn implies that outflow of link n_p is equal to $\overline{f}_{n_p}$. As this violates (3.6), we proved that there exists no consistent density vector where some links are in congested regime. Hence, the consistent density vector is unique and as in (3.9). Using the same argument, it follows that any density vector such that $f_{0_p} < \Phi R_p$ is not a consistent density vector.

2. Consider a route $p \in P_S$. Clearly, since $\Phi R_p > \overline{z}_p$, only part of ΦR_p can be accommodated. Suppose that $f_{0_p} = \overline{z}_p$. This imposes that all links preceding b_p, which have higher capacity, must be in congested regime so as to guarantee that the flow transferred from a link to the following is $\overline{z}_p$:

$$x_{l_p} = \overline{x}_{l_p} - \frac{\overline{z}_p}{w_{l_p}}, \quad f_{(l-1)_p}(x) = s_{l_p}(x_{l_p}) = \overline{z}_p,$$

$l_p = 1_p, \ldots, (b-1)_p$. Then, the density on b_p must be equal to $\overline{z}_p / v_{l_p}$. As for the links from $(b+1)_p$ to n_q, one can apply the same argument as in 1. to the sub-route they form. Again, similarly to 1, density vectors such that $f_{0_p} < \overline{z}_p$ are not consistent density vectors, as they do not fulfil to (3.6). Finally, density vectors such that $f_{0_p} > \overline{z}_p$ cannot be consistent, as this implies $f_{0_p} > \overline{f}_{b_p}$, which contradicts (3.6).

3. Consider a route $p \in P_C$. Since $\Phi R_p = \overline{z}_p$, all the inflow can be accommodated. It is easy to verify that all density vectors of the form as in (3.11) satisfy (3.6) and entirely accommodate ΦR_p. All consistent density vectors cannot take any different form. For the same argument in 1 and 2, links from $(b+1)_q$ to n_q cannot be in a congested regime. As for the links preceding b_q, from (3.3), any congested link must be followed by a congested link that limits the incoming flow from its predecessor to be equal to $\overline{z}_p$. Finally, also in this case density vectors such that $f_{0_p} < \overline{z}_p$ are not consistent density vectors, as they do not fulfil to (3.6).

□

Proposition 3.1 prescribes that when a routing vector R violates the capacity constraints of some routes, i.e., $P_S \neq \emptyset$, then the unique consistent density vector associated with it is characterized by congested links. As implied by equation (3.11), these traffic assignments transfer only a fraction of the exogenous flow directed to that route. We call such traffic assignments *partially transferring*. On the other hand, traffic assignments such that P_S are called *fully transferring*. Given an exogenous flow Φ exceeding the min-cut capacity of the network, which in our case

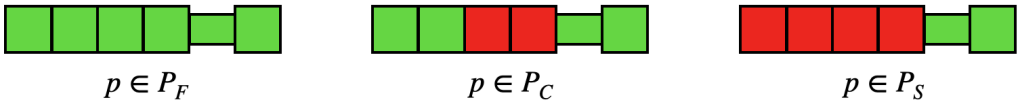

Figure 3.3: Congestion patterns for routes in P_F, P_C, P_S. Green links are in free-flow regime, red links are congested.

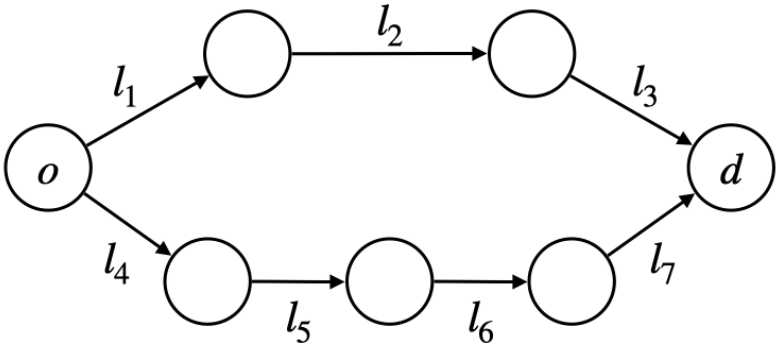

Figure 3.4: A two-route network. Route 1 consists of three links, whereas Route 2 of four.

simply corresponds to the sum of all route capacities, clearly all of its traffic assignments are partially transferring. Therefore, we turn our attention to exogenous flows that do not exceed the min-cut capacity of $\mathcal{G}$.

Assumption 3.2. *The exogenous flow Φ does not exceed the min-cut capacity of $\mathcal{G}$:*

$$\Phi \leq \sum_{p=1}^{N} \overline{z}_p.$$

Although any Φ satisfying to Assumption 3.2 admits a fully transferring traffic assignment, such exogenous flows still admit partially transferring traffic assignments, in general.

Example 3.1. *Consider the network in Figure 3.4, and assume it is characterized as follows:*

$$\begin{aligned} &\overline{f} = (1500, 1500, 1000, 1500, 1500, 1500, 1500), \\ &\overline{z} = (1000, 1500), \\ &\overline{x} = (187.5, 187.5, 125, 187.5, 187.5, 187.5, 187.5), \\ &v_l = 40, \ \forall l \in \mathcal{L}. \end{aligned} \tag{3.12}$$

Suppose that the network is subject to a constant exogenous flow $\Phi = 1500$, which satisfies Assumption 3.2. Consider the three following routing vectors:

$$R^{(1)} = (1/3, 2/3), \quad R^{(2)} = (3/4, 1/4) \quad R^{(3)} = (2/3, 1/3).$$

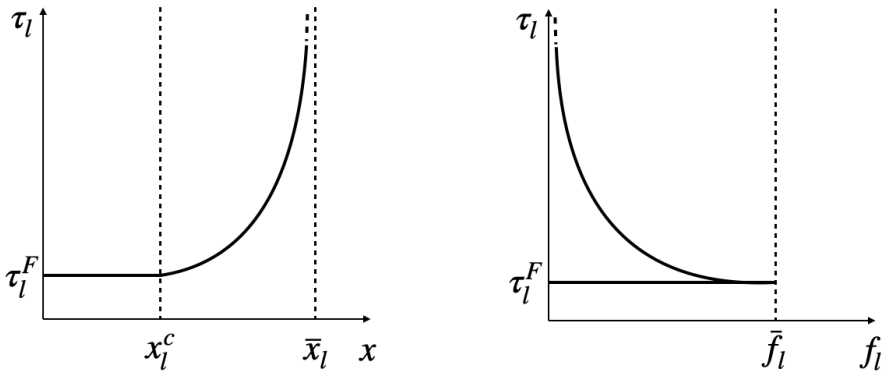

Figure 3.5: Link travel time as a function of the link density (left) and the relationship between link travel time and link flow (right).

When Φ is assigned according to $R^{(1)}$, then both routes belong to P_F, and the unique consistent density vector is

$$x^{R^{(1)}} = (12.5, 12.5, 12.5, 25, 25, 25, 25).$$

Hence, the traffic assignment $(R^{(1)}, x^{R^{(1)}})$ is unique and fully transferring.

When Φ is assigned according to $R^{(2)}$, then route 1 belongs to P_S, whereas route 2 to P_F. The unique consistent density vector associated with this routing vector is

$$x^{R^{(2)}} = (87.5, 87.5, 25, 9.375, 9.375, 9.375, 9.375).$$

The assignment $(R^{(2)}, x^{R^{(2)}})$ is clearly partially transferring, and the amount of flow that does not get transferred equals 375 veh/h.

Finally, when Φ is assigned according to $R^{(3)}$, then route 1 belongs to P_C, whereas route 2 to P_F. In this case, there exist multiple consistent density vectors, which take one of the two following form:

$$x^{R^{(3)}} = (25, x_{l_2} \in [25, 87.5], 25, 12.5, 12.5, 12.5, 12.5),$$
$$x^{R^{(3)}} = (x_{l_1} \in [25, 87.5], 87.5, 25, 12.5, 12.5, 12.5, 12.5).$$

In this case, all possible traffic assignments $(R^{(3)}, x^{R^{(3)}})$ are fully transferring. □

3.2.3 Link travel times

Consistently with the fundamental diagram of traffic, we model travel times as in [62]:

$$\begin{aligned} \tau_l : [0, \overline{x}_l] &\to \mathbb{R}_{>0} \cup +\infty \\ x_l &\mapsto \tau_l(x_l) = L_l \frac{x_l}{f_l(x_l)}, \quad l \in \mathcal{L}, \end{aligned} \tag{3.13}$$

where L_l is the length of link l. From the shape of the travel time functions, for a link l_p, when $f_{l_p}(x_{l_p}) = d_{l_p}(x_{l_p})$, then $\tau_{l_p}(x_{l_p}) = L_{l_p}/v_{l_p}$. Thus, when the link is in free-flow, its travel time is constant and equal to the free-flow travel time $\tau_{l_p}^F := L_{l_p}/v_{l_p}$. On the contrary, when $f_{l_p}(x_{l_p}) = s_{l_p}(x_{l_p})$, $\tau_{l_p} \geq \tau_{l_p}^F$ and it is increasing in x_{l_p}.

The travel time of each route $\tau_p(x)$ is simply defined as the sum of the link travel times of all route links:

$$\tau_p(x_p) = \sum_{l \in p} \tau_l(x_l), \quad p = 1, \dots, N. \tag{3.14}$$

For a given routing vector, the travel time of a route p, depending on which among the sets P_F, P_S and P_C it belongs to, will be as follows:

- if $p \in P_F$, then the route attains the lowest possible value of travel time, the *free-flow route travel time*:
$$\tau_p^F := \sum_{l \in p} \tau_l^F;$$
- if $p \in P_S$, then the route attains the following value of travel time:
$$\tau_p^S := \sum_{l=1_p}^{(b-1)_p} \tau_l \left(\overline{x}_l - \frac{\overline{z}_p}{w_l} \right) + \sum_{l=b_p}^{n_p} \tau_l^F;$$
- if $p \in P_C$, then the route can attain any value of travel time between τ_p^F and τ_p^S, precisely
$$\begin{aligned} \tau_p(x_p) = & \sum_{l=1_p}^{(k-1)_p} \tau_l^F + \tau_{k_q}(x_{k_q}^R) + \\ & + \sum_{l=(k+1)_p}^{(b-1)_p} \tau_l \left(\overline{x}_l - \frac{\overline{z}_p}{w_l} \right) + \sum_{l=b_p}^{n_p} \tau_l^F. \end{aligned}$$

Before moving to the next section, it proves convenient to define the following quantities. Given a route $p \in \{1, \dots, N\}$, with abuse of notation, let $\tau_p^{-1} : [\tau_p^F, \tau_p^S] \to \mathbb{R}^{n_p}_{\geq 0}$ be the function that, given $\tau \in [\tau_p^F, \tau_p^S]$, returns a unique consistent density vector x_p^τ of the form (3.11) such that

$$\tau_p(x_p^\tau) = \tau.$$

3.3 Non-atomic routing game (NRG)

Let us indicate the NRG as $(\mathcal{G}, \Phi)$. Each vehicle chooses its route to minimize its travel time according to the link travel time functions τ_l. We assume that the routes are ordered by increasing free-flow travel time, and, to ease the discussion, the travel times τ_p^S are assumed to be distinct.

Assumption 3.3. *The free-flow travel times τ_p^F and the travel times τ_p^S are all distinct, and routes are ordered by increasing free-flow travel times:*

$$\begin{aligned} & \tau_1^F < \tau_2^F < \dots < \tau_N^F, \\ & \tau_p^S \neq \tau_q^S, \ \forall p, q \in \{1, \dots, N\}. \end{aligned} \tag{3.15}$$

3.3.1 Wardrop equilibrium

Now, we formalize the notion of Wardrop equilibrium in our setting.

Definition 3.5 (Wardrop equilibrium)**.** *A* Wardrop equilibrium (WE) *of the NRG* $(\mathcal{G}, \Phi)$ *is a traffic assignment* (R^W, x^W) *such that*

$$R_p^W > 0 \;\Rightarrow\; \tau_p(x^W) \leq \tau_q(x^W), \quad \forall q = 1, \dots, N. \tag{3.16}$$

The following result states characterizes the WE of the NRG $(\mathcal{G}, \Phi)$, and establishes whether they are fully or partially transferring. In order to state it, let us define

$$k := \min\left\{ p \in \{1, \dots, N\} \,\middle|\, \Phi - \sum_{j=1}^{k} \overline{z}_j \leq 0 \right\},$$

$$U := \left\{ p \in \{1, \dots, k\} \,\middle|\, \tau_p^S \leq \tau_j^F, \text{ for some } j \in \{p+1, \dots, k\} \right\}.$$

The index k represents how many of the most efficient routes, i.e., routes with smallest travel time, it takes to fully accommodate the exogenous flow Φ, while the set U consists of those routes such that their free-flow travel time exceeds τ_p^S, for some p among the first k most efficient routes.

Theorem 3.1 (Characterization of the WEs)**.** *Consider the NRG* $(\mathcal{G}, \Phi)$ *and suppose that Assumption 3.1, 3.2 and 5.2 hold. Then, if* $U = \emptyset$*, the game admits a unique fully transferring WE* (R^W, x^W)*, such that*

$$\begin{aligned} \Phi R_p^W &= \begin{cases} \overline{z}_p, & p = 1, \dots, k-1 \\ \Phi - \sum_{p=1}^{z-1} \overline{z}_p, & p = k \\ 0, & p = k+1, \dots, N \end{cases}, \\ x_p^W &= \tau_p^{-1}(\tau_k^F), \quad p = 1, \dots k-1, \\ x_k^W &= \left(\frac{\Phi - \sum_{p=1}^{k-1} \overline{z}_p}{v_{1_k}}, \dots, \frac{\Phi - \sum_{p=1}^{k-1} \overline{z}_p}{v_{n_k}} \right), \end{aligned} \tag{3.17}$$

and all used routes share the same travel time τ_k^F*.*

If $U \neq \emptyset$*, let* $u := \min U$ *and let*

$$j := \min\{p = u+1, \dots, k \,|\, \tau_p^F \geq \tau_u^S\}.$$

Then,

- *if* $\tau_j^F > \tau_u^S$*, then the game admits a unique partially transferring WE* (R^W, x^W)*, such that*

$$\begin{aligned} \Phi R_p^W &= \begin{cases} \overline{z}_p, & p \in \{1, \dots, j-1\} \setminus u \\ \Phi - \sum_{p=1, p\neq u}^{j-1} \overline{z}_p, & p = u \\ 0, & p = j, \dots, N \end{cases}, \\ x_p^W &= \tau_p^{-1}(\tau_u^S), \quad p = 1, \dots, j-1, \end{aligned} \tag{3.18}$$

  and all used routes share the same travel time τ_u^S*;*

- *if* $\tau_j^F = \tau_u^S$, *then* (R^W, x^W) *reads*

$$\begin{aligned}
\Phi R_p^W &= \begin{cases} \overline{z}_p, & p \in \{1,\dots,j-1\} \setminus u \\ \Phi R_u^W, & p = u \\ \Phi(1 - \sum_{i=1}^{j-1} R_j^W), & p = j \\ 0, & p = j+1,\dots,N \end{cases}, \\
\Phi R_u^W &\in \left[\Phi - \sum_{p=1, p\neq u}^{j} \overline{z}_p, \Phi - \sum_{p=1, p\neq u}^{j-1} \overline{z}_p \right], \\
x_p^W &= \tau_p^{-1}(\tau_u^S), \quad p = 1,\dots,j-1, \\
x_j^W &= \left(\frac{\Phi(1 - \sum_{i=1}^{j-1} R_j^W)}{v_{1_j}}, \dots, \frac{\Phi(1 - \sum_{i=1}^{j-1} R_j^W)}{v_{n_j}} \right).
\end{aligned} \tag{3.19}$$

  If $j < k$, *then* (R^W, x^W) *is partially transferring, and all used routes hare the same travel time* $\tau_j^F = \tau_u^S$. *If* $j = k$, *then* (R^W, x^W) *is fully transferring if and only if* $\Phi R_u^W = \overline{z}_u$, *and all used routes hare the same travel time* $\tau_k^F = \tau_u^S$.

Before providing the proof of the theorem, we provide the reader with some intermediate results.

Lemma 3.1. *Suppose* (R^W, x^W) *is a WE of* $(\mathcal{G}, \Phi)$. *If* $R_p^W > 0$, *then* $\Phi R_q^W \geq \overline{z}_q$, $\forall q < p$.

Proof. By contradiction, assume that $\Phi R_q^W < \overline{z}_q$ for some $q < p$. Then, for any consistent density vector x^W of R^W, $\tau_q(x^W) = \tau_q^F < \tau_p^F \leq \tau_q(x^W)$, which contradicts (3.16). □

Lemma 3.2. *Suppose* (R^W, x^W) *is a WE of* $(\mathcal{G}, \Phi)$. *Then,* $\operatorname{supp}(R^W) \subseteq \{1,\dots,k\}$.

Proof. By contradiction, suppose that $\max \operatorname{supp}(R^W) > k$. By Lemma 3.1, it should be that $\Phi R_q^W \geq \overline{z}_q$, $\forall q < \max \operatorname{supp}(R^W)$, which contradicts the definition of k. □

We are now ready to provide the proof of Theorem 3.1.

Proof. Lemmata 3.1 and 3.2 imply that any WE of the game has support of the form $\{1,\dots,p\}$, $p \leq k$. We split the proof into three parts: the first part is dedicated to characterize the WE of $(\mathcal{G}, \Phi)$ when $U = \emptyset$, the second one addresses the case $U \neq \emptyset$ and $\tau_j^F > \tau_u^S$, and the third one the case $U \neq \emptyset$ and $\tau_j^F = \tau_u^S$.

1. $U = \emptyset$: in this case, there cannot be any routes such that $\Phi R_q^W > \overline{z}_q$, as this would imply that
$$\tau_q(x_q^W) = \tau_q^S > \tau_k^F, \quad \forall q < k,$$
contradicting the Wardrop condition (3.16). This also implies that $\operatorname{supp}(R^W) = \{1\dots,k\}$, as if $\operatorname{supp}(R^W) = \{1\dots,p\}$, with $p < k$, them by definition of k, there should exist $q \in \{1\dots,p\}$ such that $\Phi R_q^W > \overline{z}_q$. By combining these facts with Lemma 3.1, it becomes straightforward that the only possible traffic assignment (R^W, x^W) satisfying to the Wardrop condition (3.16) is that in (3.17). Clearly, the traffic assignment in (3.17) is fully transferring.

2. $U \neq \emptyset$, $\tau_j^F > \tau_u^S$: we start by observing that $\text{supp}(R^W) \subseteq \{1 \dots, j-1\}$, as all routes $q \in \{j, \dots, k\}$ have free-flow travel time greater than τ_u^S. Observe also that there cannot be any routes such that $\Phi R_q^W > \overline{z}_q,\ q \in \{1 \dots, j-1\} \setminus u$. In fact, as $u = \min U$ and all maximum route travel time are distinct, it holds that
$$\tau_u^S = \min_p \tau_p^S.$$
Thus, $\Phi R_q^W > \overline{z}_q,\ q \in \{1 \dots, j-1\} \setminus u$, would imply
$$\tau_q(x_q^W) = \tau_q^S > \tau_u^S,$$
violating the Wardrop condition (3.16). These facts, combined with Lemma 3.1, imply that the only possible traffic assignment (R^W, x^W) satisfying to the Wardrop condition (3.16) is that in (3.18).

3. $U \neq \emptyset$, $\tau_z^F = \tau_u^S$: analogously to the previous case, $\text{supp}(R^W) \subseteq \{1 \dots, j\}$, as all routes $q \in \{j+1, \dots, k\}$ have free-flow travel time greater than τ_u^S, and there cannot be any routes such that $\Phi R_q^W > \overline{z}_q,\ q \in \{1 \dots, j\} \setminus u$, as it would result in contradicting the Wardrop condition (3.16). By combining these facts with Lemma 3.1, it follows that all traffic assignments (R^W, x^W) that take the form in (3.19) satisfy to the Wardrop condition (3.16). As all such routing vectors satisfy to $\Phi R_u^W \geq \overline{z}_u$ and $\Phi R_j^W \leq \overline{z}_j$, so they attain the maximum travel time on route u and the free-flow travel time on route j. Among these traffic assignments, it is straightforward to see that the only one which is fully transferring is the one associated with the case $j = k$ and such that $\Phi R_u^W = \overline{z}_u$.

□

Theorem 3.1 highlights a potential drawback of selfish routing: *partially transferring Wardrop equilibria*. Even when the network is subject to an exogenous flow smaller than its min-cut capacity, users' selfish behavior can lead to traffic assignments that only partially transfer the exogenous demand. In a sense, we might think of this as selfish routing reducing the effective capacity of the network, as vehicles would never use routes that are sub-optimal in terms of travel time. Because all users aim for the shortest travel time routes and share the same queue before entering the network, the exogenous flow may be accommodated only partly, leading to congestion at the origin. In the following, we characterize the exact amount of exogenous flow loss due to partial demand transfer.

Corollary 3.1. *Consider a partially transferring* WE (R^W, x^W). *Let us indicate* Ψ *the amount of non-transferred exogenous flow. Then:*

- *if* (R^W, x^W) *takes the form in* (3.18)*, then*
$$\Psi = \Phi - \sum_{p=1}^{j-1} \overline{z}_p; \tag{3.20}$$

- *if* (R^W, x^W) *takes the form in* (3.19)*, then:*

– *if* $j < k$, *then*

$$\Psi \in \left[\Phi - \sum_{p=1}^{j} \overline{z}_p, \Phi - \sum_{p=1}^{j-1} \overline{z}_p\right]; \tag{3.21}$$

– *if* $j = k$, *then*

$$\Psi \in \left[0, \Phi - \sum_{p=1}^{k-1} \overline{z}_p\right]; \tag{3.22}$$

Example 3.2. *Consider the network in Figure 3.4 with capacities, jam densities and speeds as in (3.4). Suppose also that link lengths are as follows:*

$$L = (1, 1, 0.5, 2, 2, 2, 2)\,.$$

Suppose that $\Phi = 1000$*, so that Assumption 3.1 is satisfied. In this case,* $k = 1$ *and* $U = \emptyset$*, so the unique* WE (R^W, x^W) *is*

$$R^W = (1, 0)\,, \quad x^W = (25, 25, 25, 0, 0, 0, 0)\,.$$

Now, assume that $\Phi = 1500$*, which still satisfies to Assumption 3.1. In this other case,* $k = 2$*, but* $U = \{1\}$*, since*

$$\tau_1^S = 11.25 \text{ min} < 12 \text{ min} = \tau_2^F\,.$$

As a result, the unique WE *of the game is the following partially transferring traffic assignment:*

$$R^W = (1, 0), \quad x^W = (87.5, 87.5, 25, 0, 0, 0, 0)\,.$$

The amount of non-transferred flow Ψ *amounts to* 500*.*

Wardrop equilibria are said to be *essentially unique* when they all share the same minimum travel time [1]. Theorem 3.1 implies that the game $(\mathcal{G}, \Phi)$ exhibits essential uniqueness. Specifically, when $U = \emptyset$, the WE is unique. When $U \neq \emptyset$, if $\tau_j^F > \tau_u^S$, the WE is unique; however, if $\tau_j^F = \tau_u^S$, the WE is not unique, but all WEs have the same travel time.

Remark 3.2. *Assumptions 3.1 and 5.2 were made to simplify the analysis of the Wardrop Equilibria (WEs) of* $(\mathcal{G}, \Phi)$*. Assumption 3.1 certainly limits the generality of the model. Without Assumption 3.1, routes can be characterized by multiple minimal capacity links. In this more general case, routes would have multiple bottlenecks, and the categories of consistent density vectors for routes in sets* P_C *and* P_S *would become richer, encompassing a wider variety of congestion patterns. On the other hand, Assumption 5.2 imposes minimal limitations on the set of parameters. We underscore that these two assumptions allow for capturing the problem of partial demand transfer and are not the cause of it. As we will show in one of the next sections with an example, this issue also presents in networks where these two assumptions do not hold.*

Remark 3.3 (Comparison with [62]). *As mentioned in Section 3.1.2, a non-atomic selfish routing game relying on a description of the traffic state based both on density and flow, accounting for capacity constraints and congested traffic regimes has already been proposed in [62], but that model does not include a supply and demand mechanism. This leads to two important differences. First, our model exhibits essential uniqueness, whereas that model does not. Second, in some cases, that model does not admit a WE for certain values of exogenous flow, even when the latter is less than the min-cut capacity of the network. In contrast, our model admits a WE for any possible exogenous flow.*

3.3.2 Social optimum

In general, *social optima* are assignments minimizing some system cost. Here, we provide a definition of social optimum that accounts for both the minimization of the total travel time over the network and the full transfer of the exogenous flow Φ.

Definition 3.6. *Given an exogenous flow Φ satisfying to Assumptions 3.1 and 3.2, a* social optimum *(SO) of the game $(\mathcal{G}, \Phi)$ is a traffic assignment (R^O, x^O) such that*

$$\begin{aligned}(R^O, x^O) = \arg\min_{z,x} & \sum_{p=1}^{N} \Phi R_p \tau_p(x) = \sum_{p=1}^{N} \Phi R_p \left(\sum_{l_p=1_p}^{n_p} \tau_{l_p}(x_{l_p}) \right) \\ s.t. \quad & x \in C(R), \\ & \Phi R_p \leq \overline{z}_p, \\ & \sum_{p=1}^{N} R_p = 1.\end{aligned} \tag{3.23}$$

We prove that, in our setting, there exists a unique SO.

Proposition 3.2. *Suppose that Assumption 3.1 is satisfied. Then, the NRG $(\mathcal{G}, \Phi)$ admits a unique SO (R^O, x^O), whose expression is as follows:*

$$\begin{aligned}\Phi R^O &= \left(\overline{z}_1, \dots, \overline{z}_{(k-1)}, \Phi - \sum_{p=1}^{k-1} \overline{z}_p, 0 \dots, 0 \right), \\ x_p^O &= \left(\frac{\overline{z}_p}{v_{1_p}}, \dots, \frac{\overline{z}_p}{v_{n_p}} \right), \quad p = 1, \dots, k-1, \\ x_k^O &= \left(\frac{\Phi - \sum_{p=1}^{k-1} \overline{z}_p}{v_{1_k}}, \dots, \frac{\Phi - \sum_{p=1}^{k-1} \overline{z}_p}{v_{n_k}} \right).\end{aligned} \tag{3.24}$$

Proof. Suppose that $(R^{O'}, x_p^{O'})$ is a social optimum of the NRG and suppose that $R^{O'} > 0$. It is straightforward that every link of route p is in free-flow. In fact, since $\Phi R^{O'} \leq \overline{z}_p$, $p \in F \cup S$, which means $R^{O'}$ admits a consistent flow vector such that

all links are in free-flow. Hence, $x_p^{O'}$ cannot present saturated links, as otherwise would not be minimizing the cost in (3.23). Then, (3.23) reduces to

$$\begin{aligned}(R^O, x^O) = \arg\min_{z,x} & \sum_{p=1}^{N} \Phi R_p \tau_p^F \\ \text{s.t.} \quad & x_p = \left(\frac{\Phi R_p}{v_{1_p}}, \dots, \frac{\Phi R_p}{v_{n_p}}\right), \qquad p = 1, \dots, N, \\ & \Phi R_p \le \overline{z}_p. \\ & \sum_{p=1}^{N} R_p = 1.\end{aligned}$$

It follows immediately that the unique SO is the one using the first k routes as in (3.24). □

One of the measure most commonly used to quantify the inefficiency of WEs in routing games is the Price of Anarchy (PoA) [72]. The PoA of a WE corresponds to the total travel time realized by the WE and the minimum total travel time achievable, the one realized by the SO:

$$\text{PoA}(R^W, x^W) = \frac{\sum_{p=1}^{N} \Phi R_p^W \tau_p(x^W)}{\sum_{p=1}^{N} \Phi R_p^O \tau_p(x^O)}.$$

In our model the PoA turns out not to be the most appropriate measure of inefficiency. In fact, for partially transferring WEs, the PoA loses its significance, as the WE is transferring a flow less than that transferred by the SO. In this case, a WE might even realize a total travel time smaller than the SO, but this comes from the fact that the WE is transferring less flow. On the other hand, when (R^W, x^W) is fully transferring, the PoA is well-defined and takes the following form:

$$\text{PoA}(R^W, x^W) = \frac{\Phi \cdot \tau_k^F}{\sum_{p=1}^{k-1} \overline{z}_p \cdot \tau_p^F + \left(\Phi - \sum_{p=1}^{k-1} \overline{z}_p\right) \tau_k^F} \ge 1. \tag{3.25}$$

Another interesting fact to remark is that if the WE of the NRG is fully transferring, then the WE and the SO share the same routing vector, i.e., $R^W = R^O$ (see (3.17) and (3.24)). As this might sound contradictory, let us discuss it more in detail.

Example 3.3. *Consider the network in Figure 3.4 with capacities, jam densities and speeds as in (3.4) and link lengths*

$$L = (1.5, 1.5, 1.5, 2, 2, 2, 2)\,.$$

as in (3.2). Assume that $\Phi = 1500$, so that Assumption 3.1 is satisfied. The WE *in this case is unique and corresponds to*

$$R^W = \left(\frac{2}{3}, \frac{1}{3}\right), \quad x^W = (25, 83.3, 25, 12.5, 12.5, 12.5, 12.5)\,.$$

Indeed, such traffic assignment implies that the two route travel times satisfy two

$$\tau_1(x^W) = \tau_2(x^W) = 12 \text{ min}. \tag{3.26}$$

On the other hand, the SO *corresponds to*

$$R^O = \left(\frac{2}{3}, \frac{1}{3}\right), \quad x^O = (25, 25, 25, 12.5, 12.5, 12.5, 12.5)\,.$$

In this case,

$$\text{PoA}(R^W, x^W) = \frac{24}{17}$$

The SO fully transfer the whole exogenous demand, while also minimizing the total travel time, keeping all used routes in free-flow regime. We can provide the following explanation to this phenomenon. under Wardrop Equilibrium (WE), each user selfishly chooses their route to minimize their own travel time. This selfish behavior leads to a density vector x^W as given in (3.3). Consequently, the flow at the origin is split between the two roads in a way that results in the routing vector R^W. Conversely, in the SO scenario, the objective is to minimize the overall travel time for all users. A central planner determines the optimal routing vector R^O, which results in a specific density vector x^O. The density vector x^O ensures that all traffic routes used are in the free-flow regime, meaning they are not congested.

Therefore, even though the two routing vectors coincide, R^W can be seen as the routing vector imposed by the Wardrop condition to ensure that the used routes have the same travel time, while R^O is the routing vector that induces an optimal utilization of the network.

3.4 Beyond parallel networks

In the previous sections, we analyzed selfish routing on parallel networks. This section aims to provide an example showing that selfish routing can cause the same type of issues, such as partial demand transfer, in more complex network topologies beyond parallel networks. Consider the network in Figure 3.6 and suppose that the network geometry is the following:

$$\begin{aligned}
&\overline{f} = (1500, 1500, 800, 1500, 1500),\\
&\overline{x} = (187.5, 187.5, 100, 187.5, 187.5),\\
&v_l = 40, \ \forall l \in \mathcal{L},\\
&L = (8, 16, 4, 16, 8)
\end{aligned}$$

Suppose that the network is subject to an exogenous flow $\Phi = 1600$. A WE for this network is given by

$$R^W = \left(0, \frac{9}{16}, \frac{7}{16}\right), \qquad x^W = (107.5, 51.41, 20, 0, 37.5)$$

From the expression of x^W, one can see that both link 1 and 2 are in congested regime. The travel times of the used routes, Route 2 and Route 3, is 1 h 23 min,

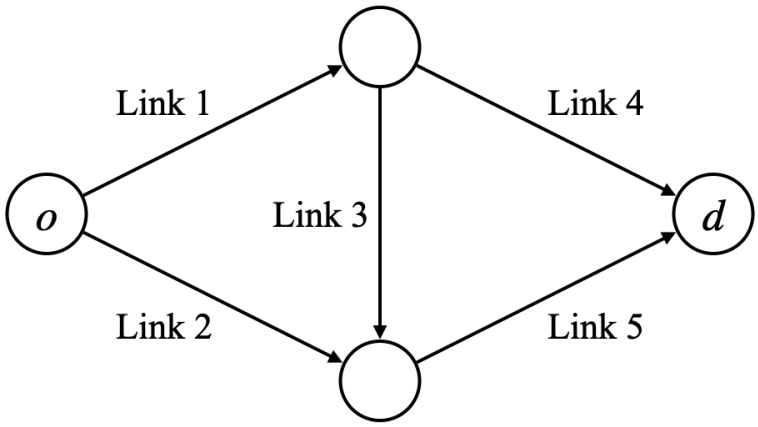

Figure 3.6: Wheatstone's network.

approximately. The travel time of the unused route, Route 1, is 1 h 29 min, instead. Then, one can notice that (R^W, x^W) is a partially transferring. Indeed, from (3.1), the supply of link one is exactly 800 veh/h. As the fraction of exogenous flow aiming to use Route 2 must pass through link 1, it is clear that the exogenous flow cannot be fully accommodated. Also in this case, users' selfish behavior leads to an inefficient traffic pattern that causes partial demand transfer.

This example demonstrates that partial demand transfer is a fundamental issue of selfish routing. Moreover, its occurrence is not limited to parallel networks but can also arise in more complex network topologies.

3.5 Concluding remarks

The main contribution of this chapter lies in the analysis of the selfish routing model in a network subject to supply and demand constraints on its links, inspired by Daganzo's cell transmission model. This approach effectively characterizes the congestion phenomena typical of traffic networks. Our analysis highlights that the issues associated with selfish routing extend beyond a mere reduction in traffic efficiency in terms of total travel time. We have demonstrated that selfish routing can lead to suboptimal utilization of the road network's capacity. Even when the network is subject to an exogenous flow less than its min-cut capacity, which can theoretically be fully transferred across the network, the traffic distribution caused by the selfish behavior of users results in only a portion of the traffic being transferred, leaving part of the exogenous flow unserved at the network's origin.

This study opens several avenues for further research. The first potential extension involves applying the model to more complex network topologies beyond the family of parallel networks. This would significantly enhance the model's applicability to real-world scenarios. The main challenge in generalizing to arbitrary networks lies in computing the Wardrop equilibria. In the current setting, we found that these equilibria can be computed easily, and an algorithm for their computation is straightforward. This is not the case for more complex networks. Therefore, a primary future objective will be to determine if the calculation of Wardrop equi-

libria can be framed as an optimization problem similar to how Wardrop equilibria are calculated in the classical routing games formulation (see Section 2.2.1).

A second important extension involves analyzing scenarios where the management of exogenous flow at the network's origin differs from what we have considered. In our model, the flow is assumed to access the network through a single entry point, with all routes sharing the same queue before entering the network. While this situation may correspond to certain real-world cases, there are scenarios that fall outside this framework and would be better modeled if each road had its own queue, i.e., the entry to one route does not depend on the others. In such cases, the travel time for each route should account for the waiting time in the queue to enter that route. In this case, the problem of partial demand transfer would probably be mitigated, as excessively long waiting times for one route would prompt users to consider alternative routes.

Future work should also aim to encompass heterogeneity, so as to account for users with different levels of information or preferences, and mixed behaviors, to capture the presence of user classes that act coordinately. Heterogeneity will be addressed in Chapters 4 and 5.

The final extension of the model involves its dynamization. Throughout this thesis, we have repeatedly emphasized the importance of studying the stability of Wardrop equilibria. The current model is entirely static, making it unclear whether traffic dynamics actually converge to these traffic assignments. To address this, we need to design dynamic network flows based on CTM principles, similar to the approaches used in previous works by [69, 71]. However, it is crucial to incorporate routing policies that reflect the selfish behavior of users. A few steps toward this direction have already been made and they are presented in Chapters 4 and 5.

Chapter 4

Impact of navigation systems: real-time routing recommendations

4.1 Introduction

In this chapter, we model the impact of using navigation apps on the efficiency of traffic networks. To do this, we will define a dynamic flow model on a network that describes the traffic dynamics of a parallel network subject to exogenous traffic flow. We assume that part of the traffic consists of users who follow the recommendations of a navigation app, which directs them to the route with the shortest travel time, while the remaining part selects routes based on prior beliefs. Unlike previous works and in continuity with the previous chapter, the model is defined on a network subject to supply and demand constraints on its links. Thus, similar to Chapter 3, our analysis of efficiency will focus primarily on the issue of partial demand transfer introduced earlier, emphasizing the role that navigation apps play in the emergence of this phenomenon.

4.1.1 Summary of results

The model can be considered, to some extents, as the extension to the dynamic framework of the model presented in Chapter 3, albeit restricted to a network with only two routes, each consisting of a single link. Traffic demand is divided into two classes: one follows a fixed routing strategy, while the other uses a navigation app to minimize travel time. We demonstrate global asymptotic stability for a broad family of user preference dynamics. Then, we study the properties of the unique equilibrium assuming that user preferences follow the logit choice model. This analysis is performed in two limit regimes. In the regime of high compliance to app's recommendations, we show that the equilibrium approximates the Wardrop equilibrium of the corresponding instance of the non-atomic routing game defined in Chapter 3. In the low compliance regime, we derive a linear approximation of the user preferences' dynamics. This twofold study shows that navigation apps can degrade the network efficiency, by increasing the average travel time (in line with previous works [6, 7]) and by leading to partial demand transfer. The key variable in our steady-state analysis is the penetration rate, that is, the share of app-informed

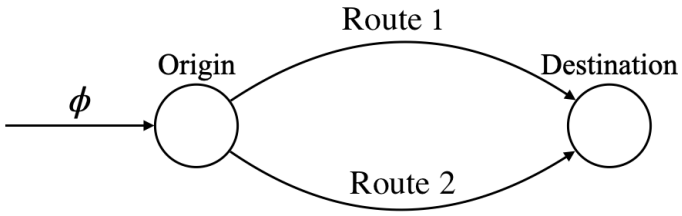

Figure 4.1: The origin-destination pair $\mathcal{G}$.

users in the total demand. Our analysis shows that a high penetration rate is likely to degrade the network efficiency when compliance is high. This picture is also confirmed by the numerical analysis of the model in a realistic case study based on the city of Grenoble, France.

4.1.2 Related work

The stability of dynamical network flows has been extensively studied in the literature [29, 65, 66, 73, 68, 70, 33], often relying on monotonicity [74] and contractivity [75] properties to demonstrate asymptotic stability. However, these studies typically overlook supply and demand mechanisms that regulate capacity constraints on network links. Unlike traffic networks, these models assume unconstrained link inflows, which can be arbitrarily high but are constrained only by link capacities for outflows. This setup ensures that traffic dynamically adjusts to minimize travel times, facilitating complete transferability of exogenous flows up to the network's min-cut capacity. In contrast, our model incorporates a realistic supply and demand mechanism on each link. This addition fundamentally alters the traffic dynamics and disrupts the monotonicity and contractivity properties. Supply and demand models are also present in [69, 71]. However, in those works, user preferences are kept fixed and do not evolve according to the state of the network.

4.1.3 Chapter organization

Section 4.2 introduces the model for dynamical network flow. In Section 4.3, we demonstrate that the system exhibits global asymptotic stability under a broad range of user behavior models. Section 4.4 focuses on the properties of the unique equilibrium when user preferences follow the logit choice model. Here, we comprehensively analyze the phenomenon of partial transfer demand and analyze the impact on total travel time, considering both high and low noise scenarios. Section 4.5 contains numerical experiments corroborating the theoretical findings. Section 4.6 proposes additional experiments showing that our findings extend beyond the setup analyzed in this chapter. Finally, Section 4.7 concludes the chapter.

4.2 Model description

Consider an origin-destination pair $\mathcal{G}$ connected by two parallel routes (see Figure 4.1), each of them consisting in a single link. Let $\overline{x}_l$, $\overline{f}_l$, v_l and L_l be the jam

density, the capacity, the free-flow speed and the length of link l. We assume that a constant exogenous flow $\Phi > 0$ (veh/h) enters the network from the origin to reach the destination. The traffic demand Φ splits on the two routes according to the *routing ratios* R_1, R_2, corresponding to the fraction of exogenous flow directed toward Route 1 and Route 2, respectively. The traffic dynamics over the network is captured by the following ODE system, consisting in a conservation law describing how the density $x_l \in [0, \overline{x}_l]$ of Route l evolves according over time:

$$\dot{x}_l = \frac{1}{L_l}\left(\min\{\Phi R_l(\tau(x)), s_l(x_l)\} - d_l(x_l)\right), \quad l = 1, 2, \tag{4.1}$$

where $x = (x_1, x_2)^T$ in $\Omega := [0, \overline{x}_1] \times [0, \overline{x}_2]$. Equation (4.1) states that the instant variation of traffic density on a route equals the difference between its inflow and its outflow. The inflow term corresponds to the minimum between the fraction of exogenous flow directed toward the route and the route supply. On the other hand, the route outflow corresponds to the route supply. The definitions of supply and demand functions is analogous to those provided in Chapter 3:

$$s_l(x_l) = \min\{\overline{f}_l, w_l(\overline{x}_l - x_l)\}, \tag{4.2}$$

$$d_l(x_l) = \min\{v_l x_l, \overline{f}_l\}, \tag{4.3}$$

where $w_l = \overline{f}_l/(\overline{x}_l - x_l^c)$, with $x_l^c := \overline{f}_l/v_l$.

Partial demand transfer and congestion

As in Chapter 3, the supply and demand mechanism naturally allows to define states that exhibit *partial demand transfer* and *congestion*. When $\Phi R_l(x) \leq s_l(x_l)$, then the demand can enter the route freely; the demand is satisfied (S). On the contrary, if $\Phi R_l(x) > s_l(x_l)$, then the first term in (4.1) gets saturated. In this case, the exogenous flow is partially transferred. When the route supply equals its capacity, i.e., $x_l \leq x_l^c$, then the route is in free-flow (F), otherwise, when $x_l > x_l^c$, it is congested (C). Therefore, each route is characterized by four possible *route modes* (see Table 4.1), allowing to rewrite (4.1)-(4.3) as follows:

$$\dot{x}_l = \begin{cases} \frac{\Phi R_l(\tau(x)) - v_l x_l}{L_l}, & \text{if } x_l \leq x_l^c,\ R_l(\tau(x)) \leq \frac{s_l(x)}{\Phi}, \quad \text{SF} \\ \frac{\overline{f}_l - v_l x_l}{L_l}, & \text{if } x_l \leq x_l^c,\ R_l(\tau(x)) > \frac{S_l(x)}{\Phi}, \quad \text{UF} \\ \frac{\Phi R_l(\tau(x)) - \overline{f}_l}{L_l}, & \text{if } x_l > x_l^c,\ R_l(\tau(x)) \leq \frac{S_l(x)}{\Phi}, \quad \text{SC} \\ \frac{\overline{f}_l(x_l^c - x_l)}{L_l(\overline{x}_l - x_l^c)}, & \text{if } x_l > x_l^c,\ R_l(\tau(x)) > \frac{s_l(x)}{\Phi}, \quad \text{UC}, \end{cases} \qquad l = 1, 2. \tag{4.4}$$

In order to avoid discussing uninteresting cases, we shall assume that demand does not exceed the network capacity, that is, it is possible to divide the demand so as to satisfy it completely.

Assumption 4.1 (Satisfiable traffic demand)**.** *Traffic demand is such that*

$$\Phi < \overline{f}_1 + \overline{f}_2. \tag{4.5}$$

Table 4.1: Notation for the four route modes in (4.4).

	Route demand	Traffic regime
SF	satisfied	free-flow
UF	unsatisfied	free-flow
SC	satisfied	congested
UC	unsatisfied	congested

4.2.1 Routing ratios and travel times

Each route is characterised by a *travel time* that is assumed to be a strictly increasing C^1 function $\tau_l : [0, \overline{x}_l] \to \mathbb{R}_{>0}$ of density x_l, $l = 1, 2$. Routing ratios R_l quantify the ratio of demand directed toward each route and they are modeled in order to account for the presence of users that rely on the routing recommendations of a navigation system to choose which route to take (*informed users*). Assuming that a fraction of users $\alpha \in (0, 1]$, which we refer to as the *penetration rate*, is influenced by the routing recommendations, whereas the remaining fraction $1-\alpha$ splits according to fixed routing ratios $r^0 = (r_1^0, r_2^0)$, $r_1^0 + r_2^0 = 1$, we write routing ratios as follows

$$R_l(\tau(x)) := (1 - \alpha)r_l^0 + \alpha r_l(\tau(x)), \quad l = 1, 2. \tag{4.6}$$

The second term in (4.6) captures the behavior of the informed users with respect to travel times, where $r_l(\tau) : \mathbb{R}^2_{>0} \to [0, 1]$ is a globally Lipschitz C^1 function, $l = 1, 2$. These functions are assumed to satisfy $r_1(\tau) + r_2(\tau) = 1$, $\forall \tau \in \mathbb{R}^2_{>0}$, so that $0 \leq R_l(\tau) \leq 1$, $\forall \tau \in \mathbb{R}^2_{>0}$, $i = 1, 2$, and that $R_1(\tau) + R_2(\tau) = 1$.

We now focus on a special class of routing ratios, *monotone routing ratios*, originally introduced in [65].

Definition 4.1 (Monotone ratios)**.** *Routing ratios* (4.6) *are said to be monotone if*

$$\frac{\partial R_l}{\partial \tau_j} \geq 0, \quad l \neq j, \ l = 1, 2. \tag{4.7}$$

If the inequality in (4.7) *is strict, then they are strictly monotone.*

Monotone routing ratios ensure that the higher the travel time on a route, the fewer informed users are directed toward it. This dependence captures the fact that app-informed users seek to minimize their travel time. Since travel times are strictly increasing in the traffic densities, for monotone routing ratios we have

$$\frac{\partial R_l(\tau(x))}{\partial x_j} \geq 0, \quad \frac{\partial R_l(\tau(x))}{\partial x_l} \leq 0, \quad l \neq j, \ l = 1, 2. \tag{4.8}$$

In what follows, we assume:

Assumption 4.2 (Strict monotonicity)**.** *Routing ratios* (4.6) *are strictly monotone.*

A well-established model of routing ratios, which falls under this assumption, is the logit routing ratios:

$$R_l(\tau) = (1-\alpha)r_l^0 + \frac{\alpha}{1 + \frac{r_j^0}{r_l^0}\exp\left(\frac{1}{\eta}(\tau_l - \tau_j)\right)}, \tag{4.9}$$

$l \neq j,\ l = 1, 2$. The state-dependent term takes the form of the *logit choice model* [27, 76, 77], where $\eta > 0$ is the so-called *noise* parameter. When η approaches zero, the logit model approximates a best response dynamics. The noise parameter can be interpreted as measuring how reliable or how *accurate* is the travel-time information provided by the navigation app. In this work, we mainly interpret the parameter $1/\eta > 0$ as measuring the *users' compliance*. When $1/\eta \to 0$, i.e., users' compliance is very low, users do not really exploit the information and the demand splitting stays close to r^0. On the contrary, when $1/\eta \to +\infty$, all users tend to take the route with shortest travel time.

4.3 Equilibria and stability analysis

Existence and uniqueness of the solutions of (4.4) are ensured by the fact that the system is Lipschitz continuous. We remark that (4.4) is well-posed with respect to $\Omega = [0, \overline{x}_1] \times [0, \overline{x}_2]$, i.e., Ω is positively invariant. Indeed, the vector field is always pointing inward on the boundaries of Ω. Hence, if $x_0 \in \Omega$ and $x(t, x_0)$ is the solution issuing from x_0, then $x(t, x_0) \in \Omega,\ \forall t \geq 0$.

Observe that (4.4) is a state-dependent switched system, where each *system mode* is a combination of route modes M_1-M_2, where $\mathrm{M}_1, \mathrm{M}_2 \in \{\mathrm{SF}, \mathrm{UF}, \mathrm{SC}, \mathrm{UC}\}$ indicate the modes of Route 1 and Route 2, respectively. We will refer to the sub-system associated with system mode M_1-M_2 with the notation $\Sigma^{\mathrm{M}_1\text{-}\mathrm{M}_2}$ and we will indicate as $\Omega^{\mathrm{M}_1\text{-}\mathrm{M}_2}$ the open region of the state space where sub-system $\Sigma^{\mathrm{M}_1\text{-}\mathrm{M}_2}$ is active. Notice that $\Omega^{\mathrm{UF\text{-}UF}}$ will always be empty because of Assumption 4.1.

We are now going to present some preliminary results that allow us to simplify the stability analysis of (4.4). Before presenting them, let us define the two following regions:

$$P := \{x \in \Omega |\ 0 \leq x_1 \leq x_1^c,\ 0 \leq x_2 \leq x_2^c\}, \quad Q := \Omega \setminus P. \tag{4.10}$$

Notice that $P = \overline{\Omega}^{\mathrm{SF\text{-}SF}} \cup \overline{\Omega}^{\mathrm{UF\text{-}SF}} \cup \overline{\Omega}^{\mathrm{SF\text{-}UF}}$, and that all regions $\Omega^{\mathrm{M}_1\text{-}\mathrm{M}_2}$ such that $\mathrm{M}_1 \in \{\mathrm{SC}, \mathrm{UC}\}$ or $\mathrm{M}_2 \in \{\mathrm{SC}, \mathrm{UC}\}$ are contained in Q.

Lemma 4.1 (Properties of region P)**.** *Given Assumption 4.1, region P is positively invariant and globally attractive.*

Proof. For positive invariance, let $x(t)$ be a solution. If $x(t)$ enters P, both routes will be either in mode SF or UF. From (4.4), we see that $x_l = x_l^c \Rightarrow \dot{x}_l \leq 0,\ l = 1, 2$, ensuring that trajectories cannot escape P.

For global attractivity, consider now $x \in Q$. Then, from the definition of Q, at least one of the two routes is in mode SC or UC. If Route i is in mode SC or UC, we can write

$$\dot{x}_l \leq -\frac{\overline{f}_l}{L_l(\overline{x}_l - x_l^c)}(x_l - x_l^c).$$

This inequality implies the convergence to P. □

Remark 4.1 (Traffic interpretation of the properties of region P)**.** *From a traffic perspective, the positive invariance of P implies that, in our model, congestion cannot arise from a free-flow condition. Also, global attractiveness implies that congestion always vanishes over time. This property is due to the implicit assumption of infinite capacity at the destination node and unconstrained route outflows.*

We shall prove that system (4.4) admits a globally asymptotically stable equilibrium, contained in region P. The proof relies on the fact that (4.4) admits a unique equilibrium and is monotone.

Definition 4.2 (Monotone system)**.** *A system $\dot{y} = h(y)$ with $h : \mathbb{R}^d \to \mathbb{R}^d$ is said to be monotone if $y_0 \leq \tilde{y}_0$ implies that $\varphi_t(y_0) \leq \varphi_t(\tilde{y}_0)$, $\forall t \geq 0$, where $\varphi_t(y^\circ)$ is the solution to $\dot{y} = h(y)$ with initial condition $y(0) = y^\circ$.*

Proposition 4.1 (Monotonicity)**.** *Given Assumptions 4.1 and 4.2, System (4.4) is monotone.*

Proof. See Appendix A.1. □

Monotonicity imparts a high degree of structure to the system, making it easier to establish its stability properties. Before proceeding, we make the following assumption.

Assumption 4.3 (Demand upper-bound)**.** *The following conditions hold:*

$$\Phi < v_l \overline{x}_l, \quad l = 1, 2. \tag{4.11}$$

Although representing a formal constraint, condition (4.11) will not prove to be restrictive at all in practical cases.

Theorem 4.1 (Global Asymptotic Stability)**.** *Given Assumptions 4.1, 4.2, and 4.3, system (4.4) admits a globally asymptotically stable equilibrium $x^* \in P$.*

Proof. See Appendix A.1 □

Theorem 6.2 characterizes the asymptotic behavior of (4.4), guaranteeing convergence of all solutions to a unique equilibrium $x^* \in P$ in free-flow regime. Therefore, we can study its properties to draw conclusions about the impact of the app's recommendations on traffic and network efficiency at steady-state.

Remark 4.2 (Beyond piece-wise linear supply and demand functions)**.** *Theorem 4.1 assumes that the supply and demand functions are piece-wise linear as per Equations (4.2)-(4.3). However, this result can be extended to more general supply and demand functions. Specifically, it is possible to show (with minor changes to the proof) that the result still holds when the supply and demand functions for $l = 1, 2$ take the form*

$$s_l(x_l) = \min\{\overline{f}_l, \tilde{s}_l(x_l)\}, \quad d_l(x_l) = \min\{\tilde{d}_l(x_l), \overline{f}_l\}$$

where $\tilde{s}_l : [0, \overline{x}_l] \to \mathbb{R}_{>0}$ is a strictly decreasing C^1 function such that $\tilde{s}_l(x_l^c) = \overline{f}_l$, $\tilde{s}_l(\overline{x}_l) = 0$, and $\tilde{d}_l : [0, \overline{x}_l] \to \mathbb{R}_{>0}$ is a strictly increasing and $\tilde{d}_l(0) = 0$, $\tilde{d}_l(x_l^c) = \overline{f}_l$. In this case, Assumption 4.3 should be replaced by $\Phi < \tilde{d}_l^{-1}(\overline{x}_l)$.

4.4 Equilibrium efficiency for logit routing

This section is dedicated to the analysis of the traffic implications of the model, focusing on the study of its unique equilibrium: the latter being globally asymptotically stable, its properties fully describe the steady-state of the system. Studying these properties in full generality is made hard by the lack of an analytic characterization of the equilibrium. For this reason, in the following, we shall study system (4.4) with logit routing ratios (4.9) and affine travel times, i.e.,

$$\tau_l(x_l) = \frac{a_l}{\overline{x}_l} x_l + \frac{L_l}{v_l}, \quad l = 1, 2. \tag{4.12}$$

The choice of affine travel time functions is largely used in the traffic literature, especially when considering a free-flow regime [44, 45, 54]. Although explicit expressions of x^* are not available, a detailed analysis becomes possible in the *high compliance* and *low compliance* regimes. Under high user compliance, i.e., when $\eta \to 0$, we will prove that the equilibrium converges to the Wardrop equilibrium of an underlying routing game almost identical to those introduced in Chapter 3. The properties of this Wardrop equilibrium can be extended by continuity to the equilibrium[1]. For low compliance, we will show that a linearization of (4.4) equipped with (4.9) provides a suitable approximation.

We will evaluate the efficiency of the equilibrium by establishing whether the equilibrium features partial demand transfer or not. In the case where demand is fully transferred, we will also evaluate the Price of Anarchy at equilibrium:

$$\mathrm{PoA}(x^*) = \frac{\sum_{l=1,2} \Phi R_l(x^*) \tau_l(x_l^*)}{\sum_{l=1,2} \Phi R_l^O \tau_l(x_l^O)}, \tag{4.13}$$

where (R_l^O, x^O) is such that $R_l^O = \frac{v_l x_l^O}{\Phi}$, $\Phi R_l \leq \overline{f}_l$, $l = 1, 2$, and minimizes the total travel time. As we will see, this second type of analysis will lead us to conclusions similar to those in [6, 25, 7]. In order to disregard trivial cases in which unsatisfied demand arises independently of routing recommendations, we make the following assumption.

Assumption 4.4 (Fixed routing ratios)**.** *The following condition holds:*

$$r_l^0 < \frac{\overline{f}_l}{(1-\alpha)\Phi}, \quad l = 1, 2, \quad \forall \alpha. \tag{4.14}$$

4.4.1 High users' compliance

As anticipated, when user compliance is very high, the properties of the equilibrium of (4.4) are inferred from those of the Wardrop equilibrium of an underlying routing game, whose structure is almost identical to that of the non-atomic routing games introduced in Chapter 3.

[1] The fact that the equilibria of a logit-based dynamics converge to the Wardrop equilibria of an associated game has already been exploited in the literature [27], including for similar traffic models that did not account for route's capacity saturation [28].

Underlying routing game

Consider the instance of NRGs introduced in Section 3.3 for the network with two parallel single-link routes which (4.4) is defined on. Suppose this network is subject to an exogenous flow $\Phi > 0$. Differently from Section 3.3, suppose that Φ consists of both informed users and non-informed users. Informed users represent a fraction $\alpha \in [0, 1]$, whereas the remaining part is represented by non-informed users. As for system (4.4), non-informed users split on the two routes according to prior beliefs $r^0 = (r_1^0, r_2^0)$, while informed users choose their route to minimize their travel time, but this time according to the travel time functions (4.12).

We start by observing the routing vectors of this game take the form

$$R_1 \in [(1-\alpha)r_1^0, (1-\alpha)r_1^0 + \alpha], \quad R_2 = 1 - R_1, \tag{4.15}$$

due to the fact that the splitting of non-informed users is fixed. Also, condition 3.6, that allows for identifying consistent density vectors, in this case reduces to

$$\min\{\Phi R_l, s_l(x_l)\} = d_l(x_l), \qquad l = 1, 2,$$

implying that the consistent density vector x^R for a given routing vector R takes the form

$$x_l^R = \begin{cases} \frac{\Phi R_l}{v_l}, & \Phi R_l < \overline{f}_l \\ x_l^c, & \Phi R_l \geq \overline{f}_l \end{cases}, \qquad l = 1, 2. \tag{4.16}$$

Hence, as (4.4), this NRG does not present traffic assignments with congested links. Definition (3.16) needs to be slightly modified in order to account for the fact that now only a fraction of the exogenous flow behaves strategically with respect to travel times.

Definition 4.3 (Wardrop equilibrium (WE))**.** *A traffic assignment* $(R^W(\alpha), x^W(\alpha))$ *of the NRG defined on the network with two parallel single-link routes and affine travel times* (4.12) *is a* Wardrop equilibrium (WE) *if and only if*

$$R_l^W(\alpha) > (1-\alpha)r_l^0 \;\Rightarrow\; \tau_l(x_l^W(\alpha)) \leq \tau_k(x_k^W(\alpha)), \quad k \neq l,\ l = 1, 2. \tag{4.17}$$

In this case, the WE is expressed as a function of α, to emphasize the influence of the penetration rate on its shape. We next characterize the Wardrop equilibrium WE(α) of the NRG. Let us define the following quantities to ease the notation:

$$c_l := \frac{a_l}{v_l \overline{x}_l}, \quad b_l := \frac{L_l}{v_l}, \quad l = 1, 2.$$

For convenience, similarly to what we have done in Chapter 3, we make the following assumption on the route travel times.

Assumption 4.5 (Route labeling)**.** *Route* 1 *has the shortest travel time route for* $\alpha = 0$*, i.e.,* $b_1 + c_1\Phi r_1^0 \leq b_2 + c_2\Phi r_2^0$.

For convenience in stating our results, we also define the following quantities:

$$\overline{\Phi} := \overline{f}_1 \left(1 + \frac{c_1}{c_2}\right) - \frac{b_2 - b_1}{c_2}, \tag{4.18}$$

$$\alpha^{\mathrm{M}} := \frac{1}{\Phi r_2^0} \frac{c_2 \Phi r_2^0 - c_1 \Phi r_1^0 + b_2 - b_1}{c_1 + c_2}, \tag{4.19}$$

$$\alpha^{\mathrm{U}} := \frac{\overline{f}_1 - \Phi r_1^0}{\Phi r_2^0}, \tag{4.20}$$

$$\alpha^{\mathrm{UM}} := 1 - \frac{1}{\Phi r_2^0} \left(\frac{c_1}{c_2} \overline{f}_1 - \frac{b_2 - b_1}{c_2} \right). \tag{4.21}$$

Lemma 4.2 (Unsatisfied demand, Wardrop equilibrium)**.** *Suppose Assumptions 4.1, 4.3, 4.4 and 4.5 holds. Then, the underlying routing game admits a unique Wardrop equilibrium* $(R^W(\alpha), x^W(\alpha))$ *and the following characterization holds:*

- *If* $\Phi \leq \overline{\Phi}$*, then no route is affected by unsatisfied demand. Moreover,*
  - *if* $\alpha \leq \alpha^{\mathrm{M}}$*, then*

$$\begin{aligned} R_1^W(\alpha) &= \alpha + (1-\alpha) r_1^0, \\ R_2^W(\alpha) &= (1-\alpha) r_2^0, \\ x^W(\alpha) &= \left(\frac{\Phi R_1^W(\alpha)}{v_1}, \frac{\Phi R_2^W(\alpha)}{v_2} \right); \end{aligned} \tag{4.22}$$

  - *if* $\alpha > \alpha^{\mathrm{M}}$*, then*

$$\begin{aligned} R_l^W(\alpha) &= \frac{c_k \Phi + b_k - b_l}{c_l + c_k}, \quad k \neq l, \ l = 1, 2, \\ x^W(\alpha) &= \left(\frac{\Phi R_1^W(\alpha)}{v_1}, \frac{\Phi R_2^W(\alpha)}{v_2} \right). \end{aligned} \tag{4.23}$$

- *If* $\Phi > \overline{\Phi}$*, then Route* 1 *will be affected by unsatisfied demand for* $\alpha > \alpha^{\mathrm{U}}$*. Moreover:*
  - *if* $\alpha \leq \alpha^{\mathrm{U}}$*, then* $(R^W(\alpha), x^W(\alpha))$ *is as in (4.22);*
  - *if* $\alpha \leq \alpha^{\mathrm{U}}$*, then* $R^W(\alpha)$ *is as in (4.22) and*

$$x^W(\alpha) = \left(x_1^c, \frac{(1-\alpha)\Phi r_2^0}{v_2} \right); \tag{4.24}$$

  - *if* $\alpha > \alpha^{\mathrm{UM}}$*, then*

$$\begin{aligned} R_1^W(\alpha) &= \Phi - \frac{c_1}{c_2} \overline{f}_1 + \frac{b_2 - b_1}{c_2}, \\ R_2^W(\alpha) &= \frac{c_1}{c_2} \overline{f}_1 - \frac{b_2 - b_1}{c_2}, \\ x^W(\alpha) &= \left(x_1^c, \frac{(1-\alpha)\Phi r_2^0}{v_2} \right); \end{aligned} \tag{4.25}$$

Proof. From the definition of Wardrop equilibrium, informed users distribute only on shortest travel time routes. From Assumption 4.5, it is clear that for α small enough, $\alpha\Phi$ will distribute entirely on Route 1. As α increases, the demand routed toward Route 1 increases as well, until either travel times equalize or Route 1 gets saturated. α^M in expression (4.19) is the maximum value of α such that, by allocating all informed users on Route 1, Route 1 still is the shortest travel time route, and can be calculated by solving $\tau_1((1-\alpha)\Phi r_1^0 + \alpha\Phi) = \tau_2((1-\alpha)\Phi r_2^0)$. α^U in expression (4.20), instead, is the maximum value of α such that, by allocating all informed users on Route 1, the capacity on Route 1 is not exceeded and can be calculated by solving $(1-\alpha)\Phi r_l^0 + \alpha\Phi = \overline{f}_l$. One can verify that $\Phi \leq \overline{\Phi} \iff \alpha^{\mathrm{M}} \leq \alpha^{\mathrm{U}}$.

This leads to consider two cases, depending on the value of Φ. If $\Phi \leq \overline{\Phi}$, then by increasing α, travel times equalize before Route 1 gets congested, since $\alpha^{\mathrm{M}} \geq \alpha^{\mathrm{U}}$. Once travel times are equal, $R^W(\alpha)$ takes the form in (4.23), which no longer depends on α. Hence, increasing α will affect no more the shape of $R^W(\alpha)$, as the additional app-informed users will distribute on the two routes so as to keep travel times even. Expression (4.23) can be retrieved by imposing the two routes' travel times to be equal. On the contrary, if $\Phi > \overline{\Phi}$, then Route 1 gets congested before travel times equalize. Nevertheless, analogously to the previous case, the informed demand keeps selecting Route 1, which is still the shortest travel time route, until travel times equalize, which now happens at α^{UM}. Again, one can verify that $\Phi \geq \overline{\Phi} \iff \alpha^{\mathrm{U}} \leq \alpha^{\mathrm{UM}}$. After travel times even out, $R^W(\alpha)$ takes the form in (4.25) and further increase do not affect the demand distribution anymore. Expression (4.25) can be retrieved by imposing the two routes' travel times to be equal, accounting for the fact that Route 1 is saturated.

The uniqueness of $R^W(\alpha)$ follows from the fact that the above cases are exhaustive and mutually exclusive. To conclude, the expressions of $x^W(\alpha)$ easily follow from (4.16). □

Convergence to the Wardrop equilibrium

We now prove how x^* converges to $x^W(\alpha)$.

Proposition 4.2 (Equilibrium approximation)**.** *Let Assumptions 4.1, 4.3, 4.4 hold. The unique equilibrium x^* of (4.4) equipped with logit routing ratios (4.9) converges to $x^W(\alpha)$, as $\eta \to 0$.*

Proof. From the proof of Theorem 4.1, the unique equilibrium of (4.4) corresponds to the fixed point of the map

$$G_l(x_l, \eta) = \frac{\min\{\Phi R_l(\tau(x_l)), \overline{f}_l\}}{v_l}, \quad l = 1, 2,$$

with $R_l(\tau(x))$ as in (4.9). Consider a sequence $(\eta_n)_{n\in\mathbb{N}} | \eta_n \to 0$. Let $x^{*(n)}$ the unique equilibrium associated with the corresponding instance of (4.4). Since $x^{*(n)} \in P$, $\forall n$, and P is compact, the sequence $\{x^{*(n)}\}_{n\in\mathbb{N}}$ is bounded. From compactness, every sequence admits a converging sub-sequence. Let $E \subseteq P$ be the set of accumulation points of all converging sequences of equilibria of (4.4). Pick

$e \in E$ and the corresponding sequence $\{x^{*(n_k)}\}_{k\in\mathbb{N}}$. Assume that one of the two routes is sub-optimal at e, i.e., $\exists l \,|\, \tilde{\tau}_l(e_l) > \tilde{\tau}_{l'}(e_{l'})$, for some l. Then

$$e_l = \lim_{\eta\to 0} x^{*(n_k)} = \lim_{\eta\to 0} G_l(x^{*(n_k)}, \eta) = \frac{(1-\alpha)\Phi r_l^0}{v_l}. \tag{4.26}$$

Hence, e corresponds to $x^W(\alpha)$, since none of the informed demand $\alpha\Phi$ is allocated on the sub-optimal route l. Since $(R^W(\alpha), x^W(\alpha))$ is the unique WE of the underlying routing game, it follows that all sequences converge to $e = x^W(\alpha)$. Again, compactness ensures that all the sub-sequences of $\{x^{*(n_k)}\}_{n\in\mathbb{N}}$ admit a sub-sequence converging to $x^W(\alpha)$. This is equivalent to say that $\{x^{*(n_k)}\}_{n\in\mathbb{N}}$ converges to $x^W(\alpha)$, as well. □

Remark 4.3 (Relationship to non-atomic routing games)**.** *The relationship between our model and* NRGs *actually represents an important element that justifies their validity. On the one hand, it proves that our dynamical model is coherent with the Wardrop's framework. On the other hand, it clarifies the stability properties of the* WE *of this* NRG*, showing that it is globally asymptotically stable.*

Since x^* converges to $x^W(\alpha)$, by continuity we can extend the properties of the latter to the former.

Corollary 4.1 (Unsatisfied demand for high compliance)**.** *Let Assumptions 4.1, 4.3, 4.4 and 4.5 hold. If $\Phi > \overline{\Phi}$ and $\alpha > \alpha^{\mathrm{U}}$, then, for small enough η, the equilibrium x^* of (4.4) equipped with logit routing ratios (4.9) presents unsatisfied demand on Route* 1*.*

Lemma 4.2 and Corollary 4.1 showcase the relevance of parameter α to the problem at hand. Nevertheless, these results also highlight the importance of Φ. Indeed, a higher traffic demand makes the system more sensitive to the penetration rate, lowering the threshold α^{U} of penetration rate beyond which unsatisfied demand emerges.

Price of Anarchy

We now investigate what impact the penetration rate has on the Price of Anarchy, under Assumptions 4.5 and 4.4. We will first perform the analysis on $(R^W(\alpha), x^W(\alpha))$ and then extend it by continuity to x^* when (4.4) is equipped with logit routing ratios and affine travel times, in the limit of high compliance. Since we consider PoA meaningful only when there is full demand transfer, we assume $\Phi \leq \overline{\Phi}$. With abuse of notation, we will write

$$\mathrm{PoA}(\alpha) = \frac{\sum_{l=1,2} \Phi R_l^W(\alpha)\tau_l(x_l^W(\alpha))}{\sum_{l=1,2} \Phi R_l^O \tau_l(x_l^O)}.$$

Proposition 4.3 (Price of anarchy in URG(α))**.** *Suppose that Assumptions 4.4 and 4.5 holds and that $\Phi \leq \overline{\Phi}$. Then,* PoA($\alpha$) *is strictly convex in α in $[0, \alpha^{\mathrm{M}}]$ and constant for $\alpha > \alpha^{\mathrm{M}}$. Moreover, let*

$$\alpha^{opt} := \frac{1}{\Phi r_2^0} \frac{2c_2\Phi r_2^0 - 2c_1\Phi r_1^0 + b_2 - b_1}{2(c_1+c_2)}. \tag{4.27}$$

Then,

- *if* $\alpha^{opt} < 0$, *then* $\mathrm{PoA}(\alpha)$ *is increasing in* $[0, \alpha^{\mathrm{M}}]$ *and constant in* $[\alpha^{\mathrm{M}}, 1]$;
- *if* $\alpha^{opt} \in [0, \alpha^{\mathrm{M}}]$, *then* $\mathrm{PoA}(\alpha)$ *attains its minimum at* α^{opt};
- *if* $\alpha^{opt} > \alpha^{\mathrm{M}}$, *then* $\mathrm{PoA}(\alpha)$ *is decreasing in* $[0, \alpha^{\mathrm{M}}]$ *and constant in* $[\alpha^{\mathrm{M}}, 1]$.

Proof. Strict convexity in $[0, \alpha^{\mathrm{M}}]$ can be checked through a second order condition. The fact that $\mathrm{PoA}(\alpha)$ is constant for $\alpha > \alpha^{\mathrm{M}}$ follows from $R^W(\alpha)$ is constant for $\alpha > \alpha^{\mathrm{M}}$. The expression of α^{opt} can be retrieved by solving $\partial_\alpha \mathrm{PoA}(\alpha) = 0$. □

One can verify that $\alpha^{\mathrm{opt}} < \alpha^{\mathrm{M}}$ if and only if $b_1 < b_2$, which means that Route 1 is faster than Route 2 when both are empty. Therefore, in this case an excessive penetration rate, notably $\alpha > \alpha^{\mathrm{opt}}$, leads to an increased number of users favoring the shortest travel time route. Consequently, this elevates the average travel time for users, thereby reducing the efficiency of the network.

Considering Proposition 2 and that the PoA is a continuous function of the density, the above characterization of the average travel time at Wardrop equilibrium as a function of the penetration rate constitutes a good approximation of the average travel time at the equilibrium x^* of (4.4) equipped with logit routing ratios (4.9), for sufficiently small η. This fact is confirmed by the simulations proposed in Section 4.5.

4.4.2 Low users' compliance

So far we have analyzed the equilibrium of (4.4) equipped with logit routing ratios (4.9) under the assumption of high users' compliance. The opposite case of low compliance can be addressed by studying a suitable linearization of the system. Its analysis yields results that are qualitatively consistent with those for high compliance.

Assume that $|\tau_1 - \tau_2|/\eta \to 0$, i.e., the argument of the exponential in (4.9) approaches zero. Then, (4.9) admits the following first order approximation:

$$R_l(\tau) = r_l^0 + \frac{\alpha r_l^0 r_k^0}{\eta} (\tau_k - \tau_l), \quad l \neq k, \ l = 1, 2. \tag{4.28}$$

For Eq. (4.28) to be a valid approximation of (4.9), i.e., to guarantee that (4.28) satisfies to $0 \leq R_l(\tau) \leq 1, \ l = 1, 2$, it is necessary that

$$\frac{1}{\eta} \leq \frac{1}{\alpha \Delta \max_l r_l^0} \tag{4.29}$$

where $\Delta := \max_{x \in P} |\tau_1(x_1) - \tau_2(x_2)|$.

Since (4.28) falls within the category of monotone routing ratios, Theorem 4.1 holds and (4.4) admits a globally asymptotically stable equilibrium, which can now also be calculated explicitly. So, we can proceed to analyse demand satisfaction and network performance.

Proposition 4.4 (Partial demand transfer for low compliance)**.** *Let us assume that Assumptions 4.1, 4.3, 4.4 and condition (4.29) hold. If* $\Phi > \overline{\Phi}$ *and*

$$\alpha > \tilde{\alpha}^U := \frac{\eta}{r_1^0(c_2\Phi + b_2 - b_1 - \overline{f}_1(c_1 + c_2))}\alpha^{\mathrm{U}}, \tag{4.30}$$

then the equilibrium x^* *of (4.4) equipped with routing ratios as in (4.28) presents unsatisfied demand on Route* 1*.*

Proof. In this case,

$$x_l^* = \frac{\eta\Phi r_l^0 + \alpha\Phi r_l^0 r_k^0(c_k\Phi + b_k - b_l)}{v_l(\eta + \alpha\Phi r_l^0 r_k^0(c_l + c_k))}, \quad l = 1, 2.$$

Unsatisfied demand emerges on Route 1 when $v_1 x_1^* > \overline{f}_1$. By plugging into the latter condition the above expression of x_1^*, after rearranging some terms one finds the following equivalent condition:

$$\alpha\Phi r_1^0 r_2^0(c_2\Phi + b_2 - b_1 - \overline{f}_1(c_1 + c_2)) > \eta(\overline{f}_1 - \Phi r_1^0).$$

This condition is met if and only if $\Phi > \overline{\Phi}$ and (4.30) holds. □

Proposition 4.4, akin to Lemma 4.2 and Corollary 4.1, demonstrates that a higher traffic demand increases the system's sensitivity to the penetration rate. In fact, $\tilde{\alpha}^{\mathrm{U}}$ is decreasing in Φ.

We omit the proof of next result, since it is analogous to that of Proposition 4.3.

Proposition 4.5 (Price of Anarchy for low compliance)**.** *Defining*

$$\tilde{\alpha}^{opt} := \frac{2\eta}{r_1^0(b_2 - b_1)}\alpha^{opt}, \tag{4.31}$$

if assumptions 4.1, 4.3, 4.4, 4.5 and condition (4.29) hold then, the average travel time of equilibrium $\overline{x}$ *of (4.4) with routing ratios (4.28),* $\mathrm{PoA}(\alpha)$*, is convex in* α*. Moreover*

- *if* $\tilde{\alpha}^{opt} < 0$*, then* $\mathrm{PoA}(\alpha)$ *is increasing in* α*;*
- *if* $\tilde{\alpha}^{opt} \in [0, 1]$*, then* $\mathrm{PoA}(\alpha)$ *attains its minimum at* $\tilde{\alpha}^{opt}$*;*
- *if* $\tilde{\alpha}^{opt} > 1$*, then* $\mathrm{PoA}(\alpha)$ *is decreasing in* $[0, 1]$*.*

One can notice the similarity in the conditions provided in Propositions 4.3 and 4.5. Another interesting aspect is that $\tilde{\alpha}^{\mathrm{opt}}$ is proportional to η, which means that for lower levels of compliance, it takes a higher penetration rate to attain the optimum.

Figure 4.2: Two possible routes connecting the Crolles basin to the Minatec area in Grenoble, France.

4.5 Case study: two crossings of Grenoble (France)

In this section, we propose numerical simulations that confirm and complement the theoretical results previously presented. We consider the following set of parameters:

$$\begin{aligned}
&\overline{f}_1 = 1700 \text{ veh/h}, \ \overline{f}_2 = 3500 \text{ veh/h} \\
&v_1 = 50 \text{ km/h}, \ v_2 = 70 \text{ km/h} \\
&\overline{x}_1 = 170 \text{ veh/km}, \ \overline{x}_2 = 250 \text{ veh/km}, \\
&L_1 = 7.5 \text{ km}, \ L_2 = 21 \text{ km}, \\
&a_1 = 1 \text{ h}, \ a_2 = 0.5 \text{ h}, \ r^0 = (0.25, 0.75).
\end{aligned}$$

These parameters have been chosen to represent realistic travel conditions in Grenoble (France), based on data from the GTL traffic platforms [78, 79]. Route 1 corresponds to an urban itinerary crossing the city center, whereas Route 2 emulates the Grenoble South Ring (see Figure 4.2). The choice of r^0 is motivated by the observation that non-informed users typically prefer main routes with higher capacity.

Figures 4.3 and 4.4 show simulations for two realistic values of demand ($\Phi = 2000$ and $\Phi = 4000$) and for three values of compliance ($1/\eta = 10$, $1/\eta = 100$ and $1/\eta = 500$), which cover both high compliance and low compliance cases. As expected, the approximation given by $(R^W(\alpha)), x^W(\alpha))$ better suits the logit model for high values of compliance, whereas the linearized model better captures low compliance scenarios. As for the qualitative behavior of the equilibrium point x^* as a function of α, the experiments are in line with the results provided in Section 4.4. To see this, let us comment upon Figures 4.3 and 4.4 more in detail. Notice that $\overline{\Phi}_1 = 3450$ veh/h, and that Assumption 4.5 is satisfied for both values of demand.

The case of $\Phi = 2000$ is illustrated in Figure 4.3: in this case, $\alpha^{\text{opt}} \approx 0.29$, $\alpha^{\text{M}} \approx 0.61$. Consistently with the theoretical results, unsatisfied demand does not arise

on any route, for any of the parameter configurations. For high compliance, the PoA reaches its minimum for low values of penetration rates, but grows rapidly for higher values. For low compliance, instead, the PoA reaches its minimum for higher values of α, but stays relatively low afterwards. This is consistent with the general observation that high compliance and high penetration rates have compounding effects on the network efficiency.

The case of $\Phi = 4000$ is shown in Figure 4.4: in this case, $\Phi > \overline{\Phi}$ and unsatisfied demand emerges on Route 1 for $1/\eta = 100$ and $1/\eta = 500$, in a neighborhood of $\alpha^{\mathrm{U}} \approx 0.23$. Instead, the curve associated with $1/\eta = 10$ is not affected by unsatisfied demand, in that the demand directed toward to Route 1 does not exceed its capacity. This is consistent with the fact that $\tilde{\alpha}^U > 1$. Again, the average travel time stays low for high penetration rate in the case of low compliance.

A relevant fact emerging from these simulations is that greater users' compliance can result in a decrease in the efficiency of the equilibrium, especially for high penetration rates. This finding is further supported by the plots in Figure 4.5. These plots depict the variation of two key metrics, α^U and $\tilde{\alpha}^U$ for $1/\eta = 10$, with respect to increasing exogenous flow Φ.

In the high compliance regime, the critical penetration rate – the minimum proportion of compliant users needed to trigger partial demand transfer – plummets rapidly as the exogenous flow increases. Conversely, the low compliance regime exhibits a more resilient behavior. The critical penetration rate only starts to decrease for very high exogenous flow values. Even then, the network can tolerate significantly higher penetration rates of compliant users compared to the high compliance scenario.

Since both high penetration rate and high compliance imply more sharing of information, these results align with the evidence in the literature about the information paradoxes in traffic networks [7, 26].

4.6 Beyond parallel single-link networks

In the previous sections, we analyzed the impact of app-informed users using a simplified network model consisting of two parallel routes, each with a single link. This framework allowed for analytical tractability and initial exploration of key concepts. However, real-world traffic networks exhibit far greater complexity. To assess the generalizability of our findings and their applicability to broader scenarios, we conducted numerical simulations using a more complex network structure.

Consider the network in Figure 4.6 and suppose that the network structure is determined by the following parameters:

$$\overline{f} = (1500 \text{ veh/h}, 1500 \text{ veh/h}, 800 \text{ veh/h}, 1500 \text{ veh/h}, 1500 \text{ veh/h}),$$
$$v_l = 40 \text{ km/h}, \quad \forall l \in \mathcal{L},$$
$$\overline{x} = (187.5 \text{ veh/km}, 187.5 \text{ veh/km}, 100 \text{ veh/km}, 187.5 \text{ veh/km}, 187.5 \text{ veh/km}),$$
$$L = (8 \text{ km}, 16 \text{ km}, 4 \text{ km}, 16 \text{ km}, 8 \text{ km}).$$

The network parameters were chosen to create a scenario where Routes 1 and 3 are high-capacity routes with similar travel times. There is also a connection between these routes that saves time but has a lower capacity. The network faces

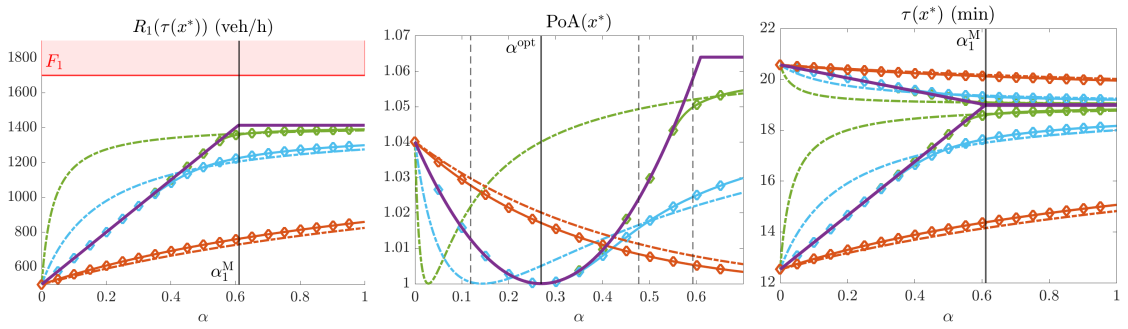

Figure 4.3: Simulations for $\Phi = 2000$. The three plots respectively show the demand directed toward Route 1 (left), average travel time T (middle) and route travel times (right) at equilibrium, as functions of penetration rate α. The diamond-marked lines are the logit routing ratios, while the dashed lines correspond to the linearized model. We draw the curves for $1/\eta = 10$ in orange, for $1/\eta = 100$ in light-blue, and for $1/\eta = 500$ in green. The limit Wardrop equilibrium $(R^W(\alpha), x^W(\alpha))$ is drawn as solid violet lines. In the left-most plot, the area highlighted in red identifies the cases in which the demand toward Route 1 is unsatisfied. In the right-most plot, the increasing travel time refers to Route 1, the decreasing one to Route 2.

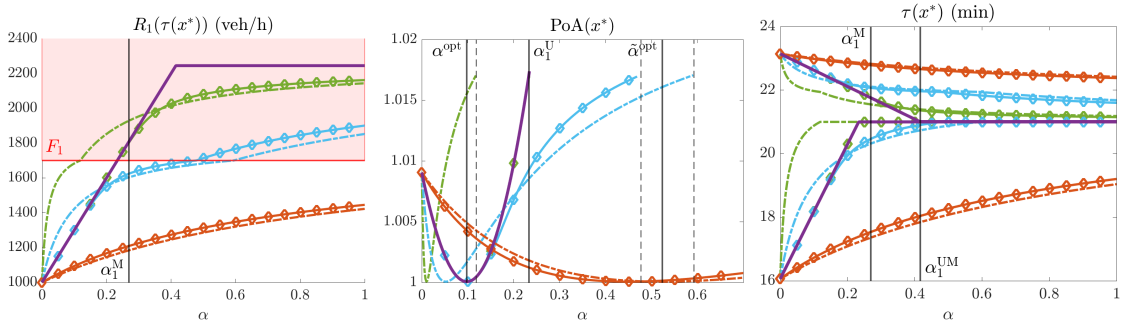

Figure 4.4: Simulations for $\Phi = 4000$. The three plots respectively show the demand directed toward Route 1 (left), average travel time T (middle) and route travel times (right) at equilibrium, as functions of the penetration rate α. The diamond-marked lines are the logit routing ratios, while the dashed lines correspond to the linearized model. We draw the curves for $1/\eta = 10$ in orange, for $1/\eta = 100$ in light-blue, and for $1/\eta = 500$ in green. The limit Wardrop equilibrium $(R^W(\alpha), x^W(\alpha))$ is drawn as solid violet lines. In the left-most plot, the area highlighted in red identifies the cases in which the demand toward Route 1 is unsatisfied. In the middle plot, the lines are truncated at the value of α at which unsatisfied demand emerges. In the right-most plot, the increasing travel time refers to Route 1, the decreasing one to Route 2.

an exogenous demand $\Phi = 2500$ vehicles per hour, with a fraction $\alpha \in [0, 1]$ of this demand consisting of users following navigation app recommendations. Let A represent the informed users and N the non-informed users. Non-informed users do not change routes once chosen and split among the three routes according to fixed ratios $R^N = (R^N_{p_1}, R^N_{p_2}, R^N_{p_3})$. Informed users can update their route choices

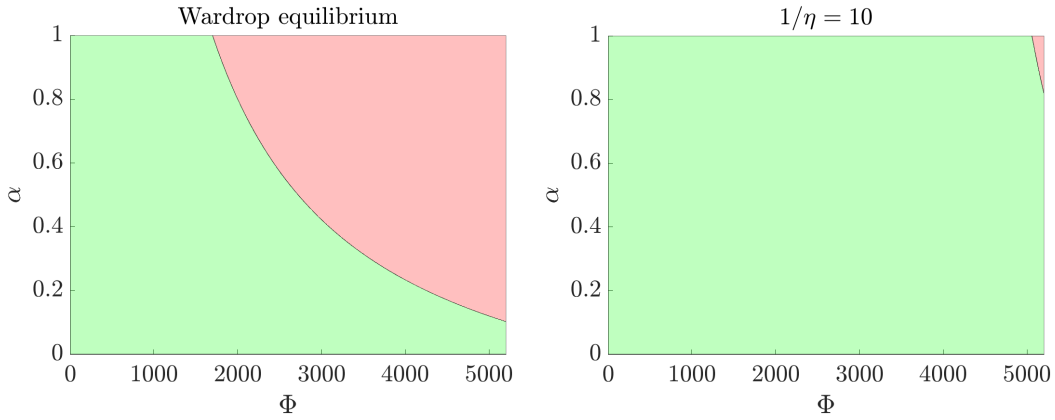

Figure 4.5: The partial demand transfer phenomenon in high and low compliance regimes is illustrated in the following manner. The green zones represent the regions of the (Φ, α) plane where the network equilibrium is unaffected by partial demand transfer. Conversely, the red zones indicate the regions where the network equilibrium is affected by partial demand transfer. The plot on the left depicts the Wardrop equilibrium $(R^W(\alpha), x^W(\alpha))$, which approximates the high compliance regime. The plot on the right illustrates the linearized system for $1/\eta = 10$. When user compliance is high, the network shows significantly more sensitivity to penetration rates as the exogenous flow increases.

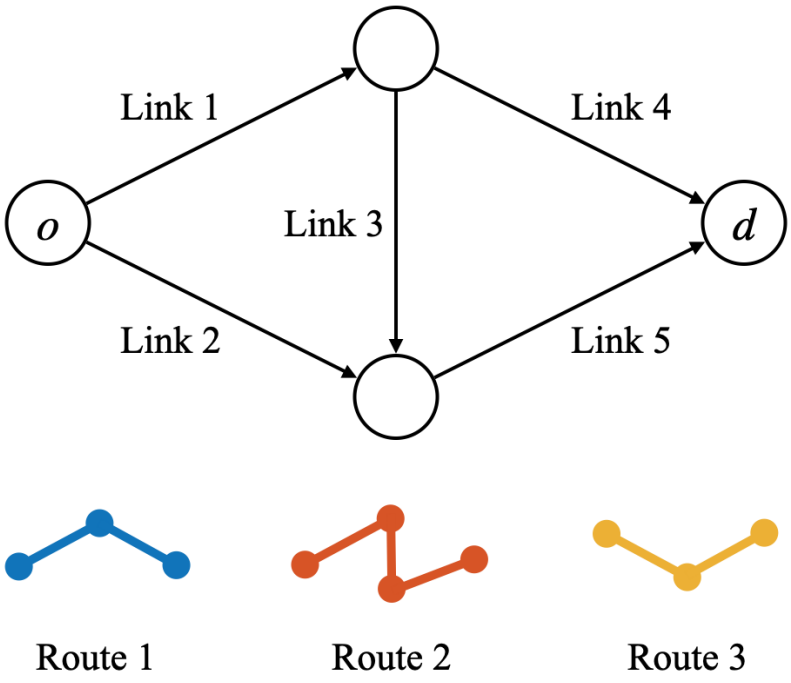

Figure 4.6: Wheatstone's network.

at each non-destination node. Let $\Phi^A = \alpha\Phi$, $\Phi^N = (1-\alpha)\Phi$, and let x_l^A indicate the density of informed users on link l and x_l^N that of non-informed users, then the evolution of the system may be described through the following set of ODEs (see Section 2.4):

$$\dot{x}^i(t) = f^{\text{in},i}(x^i(t)) - f^{\text{out},i}(x^i(t)), \tag{4.32}$$

$$f_l^{\text{in},i} = \begin{cases} \Phi^i(t), & l = o \\ \sum_{j\in a(l)^+} f_{jl}^i(x), & l \in \mathcal{L} \end{cases}, \tag{4.33}$$

$$f_l^{\text{out},i} = \begin{cases} d_l(x_l^i), & l \in \mathcal{L}^{\text{out}} \\ \sum_{j\in a(l)^-} f_{lj}^i(x), & l \in \mathcal{L} \end{cases}, \tag{4.34}$$

$$x_l = x_l^A + x_l^B, \tag{4.35}$$

$$f_{jl}^i(x) = \gamma_l(x) R_{jl}^i(x) d_j(x_j^i). \tag{4.36}$$

$$\gamma_l(x) = \min\left\{1, \frac{s_l(x_l)}{\sum_{i=A,N}\sum_{j\in a(l)^+} R_{jl}^i(x) d_j(x_j^i)}\right\}, \tag{4.37}$$

$i = A, N$. Each equation describes the evolution of one user class on one network link. The density variation on a link for each class depends only on the inflow and outflow of that class. However, congestion on links is determined by the aggregate state and is not class-specific, as indicated by the terms $\gamma_l(x)$ being the same for both classes on each link $l \in \mathcal{L}$. Notably,

- link 1 receive the flows of the two user classes using Route 1 and 2;
- link 2 receive the flows of the two user classes using Route 3;
- link 4 receives the outflow of link 1 that decided for Route 1;
- link 3 receives the outflow of link 1 that decided for Route 2;
- link 5 receives the outflows of link 2 and 3.

Non-informed users split on the three routes according to the following routing ratios:

$$R_{p_1}^N = R_{p_3}^N = 0.475, \quad R_{p_2}^N = 0.05,$$

which result in the following routing ratios at the non-destination nodes:

$$R_{o1}^N = R_{p_1}^N + R_{p_2}^N = 0.525, \quad R_{o3}^N = R_{p_3}^N = 0.475, \quad R_{14}^N \approx 0.905, \quad R_{13}^N \approx 0.095,$$

Informed users' routing preferences follow the logit choice model:

$$R_{o1}^A(x) = \frac{R_{p_1}^N \exp(-\tau_{p_1}(x)/\eta) + R_{p_2}^N \exp(-\tau_{p_2}(x)/\eta)}{\sum_{j=1}^3 R_{p_j}^N \exp(-\tau_{p_j}(x)/\eta)},$$

$$R_{o3}^A(x) = \frac{R_{p_3}^N \exp(-\tau_{p_3}(x)/\eta)}{\sum_{j=1}^3 R_{p_j}^N \exp(-\tau_{p_j}(x)/\eta)},$$

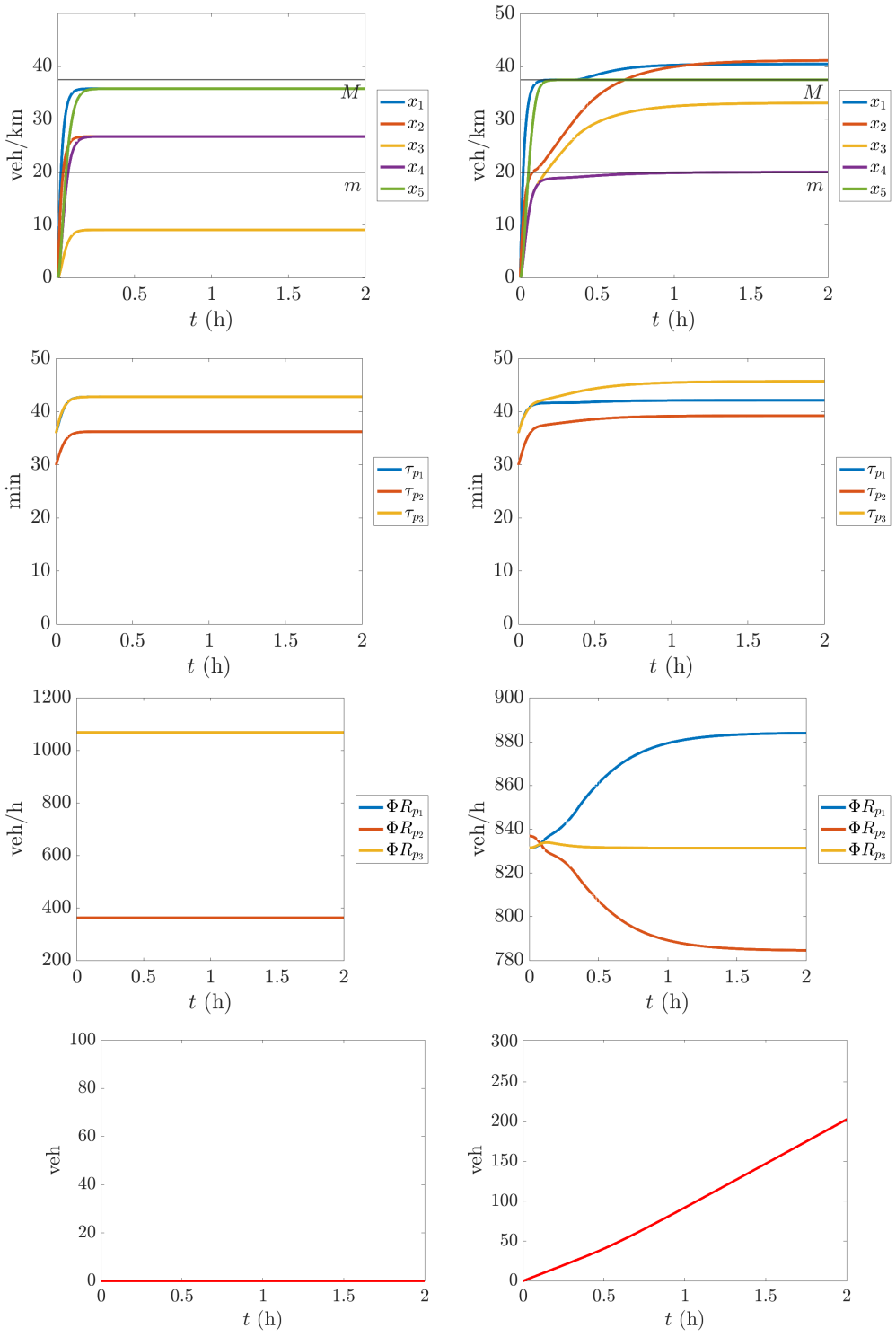

Figure 4.7: Simulations of dynamics (4.32)-(4.37) on the Wheatstone graph (Figure 4.6). The left column corresponds to $\alpha = 0.1$, the right column to $\alpha = 0.3$. The top row shows the aggregate link density plots. The second row displays the route travel times. The third row presents the fraction of exogenous demand directed towards each route. The bottom row illustrates the cumulative number of unserved vehicles stuck at the network entrance.

$$R^A_{13}(x) = \frac{R^N_{p_2} \exp(-(\tau_{l_3}(x) + \tau_{l_5}(x))/\eta)}{R^N_{p_1} \exp(-\tau_{l_4}(x)/\eta) + R^N_{p_2} \exp(-(\tau_{l_3}(x) + \tau_{l_5}(x))/\eta)},$$
$$R^A_{14}(x) = 1 - R^A_{13}(x).$$

This time, link travel are modeled as [80]:

$$\tau_l(x_l) = \frac{L_l}{v_l \left(1 - \frac{x_l}{\overline{x}}\right)},$$

and the travel time for a route is the sum of the travel times for its links.

Figure 4.7 displays the results of a numerical experiment conducted on the Wheatstone network (Figure 4.6), with $\alpha = 0.1$ and $\alpha = 0.3$, $1/\eta = 100$, and initial conditions $x^i_l(0) = (0, 0, 0, 0, 0)$ for $i = A, N$. The left column shows the results for $\alpha = 0.1$, while the right column shows the results for $\alpha = 0.3$.

In both experiments, the traffic state eventually converges to a steady state. Comparing the two sets of plots, we observe how the increase in the fraction of informed users negatively impacts traffic efficiency, leading to congestion and partial demand transfer. For $\alpha = 0.1$, most users do not follow the routing recommendations, favoring higher capacity routes even if they are sub-optimal in travel time. Nonetheless, the network remains in a free-flow regime, fully accommodating the exogenous demand. However, increasing α to 0.3 significantly worsens the conditions. Most informed users take Route 2, causing significant congestion on links 1, 2, and 3. Link 5 cannot handle the increased flow from Routes 2 and 3, resulting in spill-backs and eventual congestion up to the network origin. Consequently, the exogenous flow is not fully transferred, and congestion builds at the origin, as shown in the bottom-right plot.

This experiment demonstrates that the negative impact of increased proportions of informed users on traffic efficiency, as well as partial demand transfer due to navigation apps, holds true even in more complex network configurations.

4.7 Concluding remarks

The contribution of this chapter is twofold. First, the proposed model represents an initial attempt to dynamize the model presented in Chapter 3. Although this is done on a very simple network, the results suggest that the Wardrop equilibria of the game developed in that chapter are stable and reachable by the traffic system when the traffic consists of selfish users interacting in an uncoordinated manner.

The second contribution is successfully linking the study of the impact of navigation apps to the theoretical framework developed in Chapter 3, specifically the issue of partial demand transfer. It is shown here how a too high penetration rate of informed users can lead to unsatisfied demand, as informed users, all directed towards the route with the shortest travel time, create a bottleneck at the origin of the network. Finally, we also emphasize that our study highlights results aligned with those proposed in previous works ([6, 25, 7]) regarding the effect of the penetration rate on network efficiency in terms of total travel time. We show that as the penetration rate of informed users increases, the total travel time in the network can also increase.

As with the model proposed in Chapter 3, the main directions for improving the model are twofold. The first is the generalization to more complex topologies to increase the applicability of the model. The second is to generalize the traffic dynamics at the origin node of the network to describe more realistic cases, not just those where entry to the network is through a single road, with each user forced to wait in the same queue regardless of the road they want to select.

The model can be easily extended to networks with multiple multi-link routes, such as those considered in Chapter 3. However, beyond the case studied in this chapter (Figure 4.1), the properties of equilibrium uniqueness and monotonicity are not satisfied. As a result, analyzing stability for these and more complex networks is challenging. Furthermore, as noted at the beginning of the chapter, the supply and demand constraints mean that many techniques used in previous works, which are based on the contractivity property [65, 66, 68, 73], are not applicable in our case. Our system does not possess these properties. Therefore, efforts will be directed towards identifying new techniques for stability analysis.

Chapter 5

Impact of navigation systems: delayed routing recommendations

5.1 Introduction

The routing recommendations provided by navigation apps to their users are derived from traffic data collected by the app. There is an inevitable time lag between the moment the traffic data is collected and the moment the recommendations based on this data are provided to users. This delay is due to the time required for the operations of data collection, communication, and processing. In this chapter, we aim to evaluate what impact delays affecting routing recommendations can exert on traffic efficiency. To do this, we use a slightly modified version of the model developed in Chapter 4, accounting for this delay.

5.1.1 Summary of the results

Assuming the two routes in the network have the same length and speed, we reduce the problem to a scalar ODE by considering the difference in travel times between the two routes. By conducting a stability analysis of this scalar dynamic, we can comprehensively characterize the impact of informational delay on the stability and efficiency of the traffic system. This analysis involves three key parameters: the exogenous flow through the network, the penetration rate of informed users, and their compliance.

Our findings indicate that when these three parameters are sufficiently small, the delay in routing recommendations does not compromise the asymptotic behavior of the traffic system. However, when the product of these parameters exceeds a certain threshold, which can be expressed in terms of other system parameters, sufficiently large delays lead to system destabilization. This results in an oscillating traffic state and can cause periodic partial transfer of demand.

5.1.2 Related works

In the literature, almost all relevant works within the macroscopic dynamic routing framework do not consider information delay affecting routing [32, 33, 35, 36, 68].

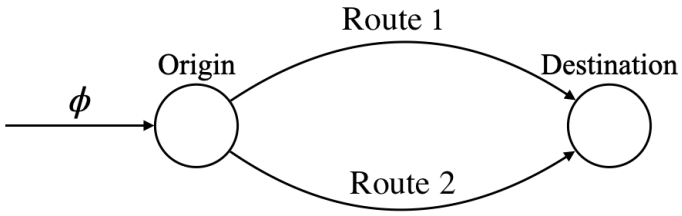

Figure 5.1: Graph representation of the origin-destination pair considered.

It is worth to mention that in [81] the authors propose a framework that allows for dynamic routing of users based on delayed information, but they do not elaborate on the effects caused by the latter. In [82, 83], simulations using a microscopic traffic model demonstrated that information based on floating car data, which is intrinsically affected by delay, can lead to oscillating trajectories of the system. However, these works do not address the problem of partial demand transfer.

5.1.3 Chapter organization

In Section 5.2, we introduce the model, which is similar to the one described in Chapter 4, but now includes informational delay. In Section 5.3, under the assumption that routes are homogeneous (same length and free-flow speed), we show that the system introduced in the previous section can be reduced to a scalar dynamics. In Section 5.4, we analyze the stability of the model by studying the asymptotic behavior of the scalar dynamics retrieved in the previous section, offering a sufficient condition for delay-independent stability and another for destabilization due to delay. Section 5.5 presents numerical experiments that both support our theoretical results and suggest that they extend beyond the set up analyzed in this chapter. Finally, Section 5.6 concludes the chapter.

5.2 Model definition

The model proposed in this chapter extends the one analyzed in Chapter 4 so as to account for the presence of delay affecting the routing recommendations provided by the navigation system. Once again, we consider a network with two parallel single-link routes, subject to an exogenous flow Φ, whose traffic dynamics on the two routes is described by the following conservation law:

$$\dot{x}_l(t) = \frac{1}{L_l}\left(\min\{\Phi R_l(\tau(x(t-\theta))), s_l(x_l(t))\} - d_l(x_l(t))\right), \qquad l = 1, 2. \tag{5.1}$$

Equation (5.1) almost coincides with Equation (4.1), except for the fact that in this case the routing ratios depend on the route travel times associated with a delayed state of the system $x(t-\theta)$, capturing the fact that routing recommendations provided to informed users are affected by a delay $\theta \geq 0$. This delay is due to the time required for collecting, communicating, and processing the traffic data.

Supply and demand function take the same form as in (4.2), (4.3) and establish analogous capacity constraints on the two routes. Also in this case, the system can

be rewritten in the following form:

$$\dot{x}_l(t) = \begin{cases} \frac{\Phi R_l(\tau(x(t-\theta)))-v_l x_l(t)}{L_l}, & \text{if } x_l(t) \leq x_l^c,\ R_l(\tau(x(t-\theta))) \leq \frac{s_l(x(t))}{\Phi}, \\ \frac{\overline{f}_l-v_l x_l(t)}{L_l}, & \text{if } x_l(t) \leq x_l^c,\ R_l(\tau(x(t-\theta))) > \frac{s_l(x(t))}{\Phi}, \\ \frac{\Phi R_l(\tau(x(t-\theta)))-\overline{f}_l}{L_l}, & \text{if } x_l(t) > x_l^c,\ R_l(\tau(x(t-\theta))) \leq \frac{s_l(x(t))}{\Phi}, \\ \frac{\overline{f}_l(x_l^c-x_l(t))}{L_l(\overline{x}_l-x_l^c)}, & \text{if } x_l(t) > x_l^c,\ R_l(\tau(x(t-\theta))) > \frac{s_l(x(t))}{\Phi}, \end{cases} \quad l = 1, 2. \tag{5.2}$$

Routing ratios are also analogous to those in (4.6), but they now account for the delay affecting recommendations:

$$R_l(\tau(x(t-\theta))) = (1-\alpha)r_l^0 + r_l(\tau(x(t-\theta))). \tag{5.3}$$

We still assume that only a part of the exogenous flow Φ consists of informed users that rely on the recommendations of a navigation system. The remaining part of users, the non-informed users, splits according to fixed preferences $r^0 = (r_1^0, r_2^0)$. Again, we assume that the demand can be fully accommodated, i.e., $\Phi < \overline{f}_1 + \overline{f}_2$. We are interested in analyzing the stability of the system (5.1) to understand how the presence of the delay θ affects traffic efficiency.

We start by observing that the right-hand side of (5.1) is Lipschitz continuous. Hence, there exists a unique solution to (5.1) that continuously depends on the initial data for every initial condition $x(\omega) = \chi(\omega)$, $\omega \in [-\theta, 0]$, where $\chi(\omega) \in C([-\theta, 0], \Omega)$ and $\Omega := [0, \overline{x}_1] \times [0, \overline{x}_2]$ [84][Section 1.3.1].

Now, consider the region $P = [0, x_1^c] \times [0, x_2^c]$ introduced in Section 4.3.

Proposition 5.1. *P is a positively invariant and attractive region for (5.1).*

Proof. The proof is analogous to that of Lemma 4.1. □

Proposition 5.1 allows us for restricting the stability analysis of (5.1) to region P only, as we did for the traffic system influenced by real-time recommendations (4.1). Within region P, (5.2) reduces to the following differential system of equations, stating that each route is characterized by two possible *modes*:

$$\dot{x}_l(t) = \begin{cases} \frac{1}{L_l}\left(\Phi R_l(\tau(x(t-\theta))) - v_l x_l(t)\right), & \text{if } R_l(\tau(x(t-\theta))) \leq \frac{\overline{f}_l}{\Phi}, \\ \frac{1}{L_l}\left(\overline{f}_l - v_l x_l(t)\right), & \text{if } R_l(\tau(x(t-\theta))) > \frac{\overline{f}_l}{\Phi}, \end{cases} \quad l = 1, 2. \tag{5.4}$$

The first equation represents traffic dynamics when the exogenous flow assigned to the route doesn't exceed its capacity. In contrast, the second equation captures dynamics when the assigned flow surpasses the route's capacity.

5.2.1 Routing ratios and travel time

In this chapter, the stability analysis of the system will be conducted assuming from the outset that routing ratios follow the logit choice model and that the route travel times are linear function of the route densities:

$$\begin{aligned} R_1(\delta(t-\theta)) &:= (1-\alpha)r_1^0 + \alpha \frac{1}{1 + \frac{r_2^0}{r_1^0}\exp\left(-\frac{1}{\eta}\delta(t-\theta)\right)}, \\ R_2(\delta(t-\theta)) &= 1 - R_1(\delta(t-\theta)), \end{aligned} \tag{5.5}$$

where

$$\delta(t) := \tau_2(x_2(t)) - \tau_1(x_1(t)),$$

and

$$\tau_l(x_l(t)) = a_l \frac{x_l(t)}{\overline{x}_l} + \frac{L_l}{v_l}, \quad a_l > 0, \quad l = 1, 2. \tag{5.6}$$

As in Chapter 4, we interpret $1/\eta > 0$ as the *users' compliance*, which quantifies the tendency of informed users to actually follow the routing recommendations from the navigation system. Again, when $1/\eta \to 0$, i.e., users' compliance is very low, users do not really exploit information and the demand splitting stays close to r^0. On the contrary, when $1/\eta \to +\infty$, all users tend to take the shortest travel time route recommended by the navigation system.

5.3 The case of homogeneous routes

Section 5.4 is dedicated to the stability analysis of the model introduced in Section 5.2. In this section, we will introduce an assumption that allows for, at the cost of losing some generality, to reduce (5.4) to a scalar dynamics.

Assumption 5.1 (Homogeneous free-flow speeds and route lengths)**.** *The free-flow speed and the route length are the same for both routes:*

$$v_l = v, \quad L_l = L, \quad l = 1, 2. \tag{5.7}$$

From now on, we will work on this special case. Assumption 5.1 refers to a scenario in which the two routes are of similar length and subject to the same speed limit but can have different capacities, e.g., two different itinerary in a urban road network.

Now, let us multiply $\dot{x}_l(t)$ by $a_l/\overline{x}_l$, $l = 1, 2$ in (4.1), and then subtract the first equation from the the second one. Thanks to Assumption 5.1, we get the following scalar delay-differential equation in $d(t)$:

$$\dot{\delta}(t) = -\frac{v}{L}\delta(t) + g(\delta(t-\theta)), \tag{5.8}$$

where

$$g(\delta(t-\theta)) := \frac{1}{L}\left(\frac{a_2}{\overline{x}_2}\min\left(\overline{f}_2, \Phi(1 - R_1(\delta(t-\theta)))\right) - \frac{a_1}{\overline{x}_1}\min\left(\overline{f}_1, \Phi R_1(\delta(t-\theta))\right)\right) \tag{5.9}$$

is a globally Lipschitz continuous function with Lipschitz constant

$$K = \frac{\alpha\Phi}{4\eta L}\left(\frac{a_2}{\overline{x}_2} + \frac{a_1}{\overline{x}_1}\right). \tag{5.10}$$

From Proposition 5.1, the state space of (5.8) is given by $[-a_1 x_1^c/\overline{x}_1, a_2 x_2^c/\overline{x}_2]$.

Observe that because of the one-to-one correspondence between $R_1(\delta(t))$ and $\delta(t)$, any conclusion about the stability of the trajectories of $\delta(t)$ will be valid for the stability of the trajectories of $R_1(\delta(t))$, as well.

5.4 Stability analysis

In this section, we analyze the stability of (5.8). First, we demonstrate that (5.8) has a single equilibrium point. Next, we use a Lyapunov approach to provide sufficient conditions for this equilibrium point to be globally asymptotically stable. Finally, we identify sufficient conditions under which the system becomes unstable when the delay θ is sufficiently large. Special emphasis is placed on the relationship between these conditions and the parameters Φ, α, and $1/\eta$.

Proposition 5.2 (Uniqueness of the equilibrium point)**.** *The dynamics* (5.8) *admits a unique equilibrium point.*

Proof. Define $G(\delta) := Lg(\delta)/v$. We observe that δ^* is an equilibrium of (5.8) if and only if is a fixed point of $G(\delta)$. Since $G(\delta)$ is a continuous and strictly decreasing function in δ, it admits a unique fixed point. Therefore, (5.8) has a unique equilibrium point δ^*. □

5.4.1 Delay-independent global asymptotic stability

In the following, we provide a sufficient condition for the delay-independent global asymptotic stability (GAS) of (5.8).

Theorem 5.1 (Delay-independent GAS)**.** *The unique equilibrium point δ^* of the dynamics* (5.8) *is globally asymptotically stable for all $\theta \geq 0$ if $K < v/L$.*

Proof. For convenience, consider the dynamics obtained by shifting (5.8) so that δ^* is at the origin of the system:

$$\dot{u}(t) = -\frac{v}{L}u(t) - h(u(t-\theta)), \tag{5.11}$$

where $u(t) = \delta(t) - \delta^*$, $h(u(t)) := g(u(t) + \delta^*) - x(\delta^*)$. Clearly, the asymptotic properties of (5.8) coincide with those of (5.11). Define now the following Lyapunov functional:

$$V(t) := \frac{1}{2}u^2(t) + \frac{v}{2L}\int_{t-\theta}^{t} u^2(s)ds. \tag{5.12}$$

First of all, notice that

$$V(u(t)) \geq \frac{1}{2}|u(t)|^2, \quad V(u(t)) \leq \frac{1}{2}\left(1 + \frac{v}{L}\theta\right)\max_{s\in[t-\theta,t]}|u(s)|^2.$$

Moreover, by taking its derivative, we find that

$$\begin{aligned}\dot{V}(t) &= u(t)\dot{u}(t) + \frac{v}{2L}\left(u^2(t) - u^2(t-\theta)\right) \leq \\ &\leq -\frac{v}{2L}u^2(t) - \frac{v}{2L}u^2(t-\theta) + K|u(t)||u(t-\theta)| = \\ &= -\begin{pmatrix}|u(t)| & |u(t-\theta)|\end{pmatrix}\begin{pmatrix}\frac{v}{2L} & -\frac{K}{2} \\ -\frac{K}{2} & \frac{v}{2L}\end{pmatrix}\begin{pmatrix}|u(t)| \\ |u(t-\theta)|\end{pmatrix}.\end{aligned}$$

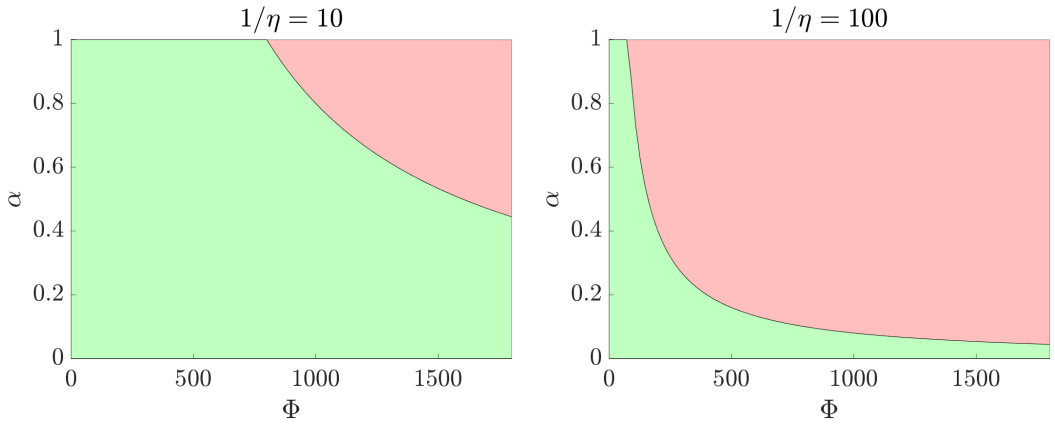

Figure 5.2: Stability regions of the dynamics (5.8) on the (Φ, α)-plane are depicted. The plots on the left refer to the case where $1/\eta = 10$, while the plots on the right correspond to the case where $1/\eta = 100$. The green areas indicate regions of the plane where the dynamics are guaranteed to be delay-independent globally asymptotically stable. In contrast, the red areas indicate regions where stability is not guaranteed. In this example, the values assigned to the other system parameters were taken from Section 5.5. The plots support the observation that increasing any of the following factors – exogenous flow, penetration rate, or user compliance – can undermine the stability of the system.

If the matrix defining the quadratic form above is positive definite, then there exists $\gamma > 0$ such that $\dot{V}(t) < -\gamma|u(t)|^2$ and global asymptotic stability of (5.8) comes from [84][Theorem 3.1]. The matrix is positive definite if and only if $v/L > K$. □

The inequality $K < v/L$ can be rewritten equivalently as follows:

$$\frac{\alpha\Phi}{\eta} < \frac{4v\overline{x}_1\overline{x}_2}{a_2\overline{x}_1 + a_1\overline{x}_2}. \tag{5.13}$$

Then, some considerations can be made about Theorem 5.1. Condition (5.13) clearly highlights how the stability of the system and its sensitivity to delays crucially depend on the traffic demand on the network Φ, the penetration rate α, and the user compliance $1/\eta$. Specifically, a sufficiently high number of users following the recommendations makes the system sensitive to delays, potentially compromising its stability.

5.4.2 Instability and oscillations for large demand and delay

When condition (5.13) is not met, the system may become unstable if the delay is sufficiently large. In the following, we perform a local stability analysis around the unique equilibrium point δ^* of the system to investigate its stability properties when condition (5.13) does not hold.

We will focus on a relevant subset of the parameter set, in which the following three facts hold.

1. There is no unsatisfied demand when the penetration rate is zero, which is equivalent to
$$\Phi r_l^0 < \overline{f}_l, \quad l = 1, 2. \tag{5.14}$$

2. There is no unsatisfied demand at equilibrium, which is equivalent to
$$1 - \frac{\overline{f}_2}{\Phi} < R_1(\delta^*) < \frac{\overline{f}_1}{\Phi}. \tag{5.15}$$

3. The user demand Φ and the penetration rate α are large enough to allow unsatisfied demand to emerge on one of the two routes. This requirement will be satisfied by the following condition:
$$\Phi > \overline{f}_l, \ \alpha > \underline{\alpha}_l := \frac{\overline{f}_l - \Phi r_l^0}{\Phi(1 - r_l^0)}, \quad l = 1, 2. \tag{5.16}$$

Assumption 5.2. *Conditions (5.14), (5.15) and (5.16) are satisfied.*

Theorem 5.2 (Local stability)**.** *Suppose Assumption 5.2 holds. Then, the following assertions hold true for dynamics (5.8):*

1. *if $|g'(\delta^*)| < \frac{v}{L}$, then δ^* is asymptotically stable for all $\theta \geq 0$.*
2. *if $g'(\delta^*) < -\frac{v}{L}$, then δ^* is asymptotically stable for $\theta < \theta^*$ and unstable for $\theta > \theta^*$, where*
$$\theta^* := \frac{1}{\sqrt{(g'(\delta^*))^2 - \frac{v^2}{L^2}}} \arccos\left(\frac{v}{L g'(\delta^*)}\right), \tag{5.17}$$
undergoing a Hopf bifurcation at $\delta = \delta^*$ when $\theta = \theta^*$.

Proof. Assumption 5.2 ensures that no terms of (5.9) are saturated at equilibrium. Then, (5.8) can be rewritten as follows:
$$\dot{\delta}(t) = -\frac{v}{L}\delta(t) + \frac{\Phi}{L}\left(\frac{a_2}{\overline{x}_2} - \left(\frac{a_1}{\overline{x}_1} + \frac{a_2}{\overline{x}_2}\right) R_1(\delta(t-\theta))\right). \tag{5.18}$$

Since the inequalities in (5.15) are strict, we are able to find a neighborhood $I \subset \mathbb{R}$ of δ^* where (5.8) takes the form (5.18), with a differentiable right-hand side. Therefore, we are able to perform a local stability analysis within I, by linearizing (5.18) in I and studying the behavior of its eigenvalues. The linearization of (5.18) in I takes the form
$$\dot{\delta}(t) = -\frac{v}{L}\delta(t) + \frac{\Phi}{L}\left(\frac{a_1}{\overline{x}_1} + \frac{a_2}{\overline{x}_2}\right) R_1'(\delta^*)(\delta(t-\theta) - \delta^*),$$
and it scharacteristic equation is given by
$$\lambda = -\frac{v}{L} + \frac{\Phi}{L}\left(\frac{a_1}{\overline{x}_1} + \frac{a_2}{\overline{x}_2}\right) R_1'(\delta^*) e^{-\lambda\theta}.$$
The statement follows after applying [85][Theorem 2.3] to (5.8). □

This result provides a necessary and sufficient condition for instability. However, it involves several conditions that cannot easily be tested, as the equilibrium δ^* is not known in closed form. In order to obtain testable conditions, we begin by deriving a sufficient condition to replace (5.15), always assuming to be in a sufficiently small neighborhood of δ^* where (5.8) reads as (5.18), as in the proof of Theorem 5.2. To this purpose, one can readily verify that condition (5.15) is equivalent to

$$\delta_{FC} < \delta^* < \delta_{CF}, \tag{5.19}$$

where

$$\delta_{CF} := \eta \log\left(\frac{r_2^0}{r_1^0}\frac{\gamma_1}{\alpha\Phi - \gamma_1}\right), \quad \delta_{FC} := \eta \log\left(\frac{r_2^0}{r_1^0}\frac{\alpha\Phi - \gamma_2}{\gamma_2}\right),$$

with $\gamma_l := \overline{f}_l/\Phi - (1-\alpha)r_l^0$, $l = 1, 2$. Next, we derive the following sufficient condition.

Lemma 5.1. *Given $\Phi > 0$ satisfying (5.14) and (5.16). Then, (5.19) is satisfied for all $\alpha > \max\{\underline{\alpha}_1, \underline{\alpha}_2\}$ and $1/\eta > 0$, if the following condition hold:*

$$r_1^0 < \frac{a_2\overline{x}_1}{a_1\overline{x}_2 + a_2\overline{x}_1} < \frac{\overline{f}_1}{\Phi} \quad \text{or} \quad r_2^0 < \frac{a_1\overline{x}_2}{a_1\overline{x}_2 + a_2\overline{x}_1} < \frac{\overline{f}_2}{\Phi}. \tag{5.20}$$

Proof. See Appendix A.2 □

Based on this analysis, we can now state the following assumption, which can be tested on the system parameters.

Assumption 5.3. *Conditions (5.14), (5.16) and (5.20) are satisfied.*

Finally, we combine Theorem 5.2 with the following lower and upper bounds on the absolute value of $g'(\delta)$.

Lemma 5.2. *Suppose that Assumption 5.3 holds. Then, the following inequalities hold:*

$$Q < |g'(\delta^*)| < K, \quad Q := \min(|g'(\delta_{CF})|, |g'(\delta_{FC})|) \tag{5.21}$$

where

$$\begin{aligned} |g'(\delta_{CF})| &= \frac{\Phi}{\eta L}\left(\frac{a_1}{\overline{x}_1} + \frac{a_2}{\overline{x}_2}\right)\gamma_1\left(1 - \frac{\gamma_1}{\alpha}\right), \\ |g'(\delta_{FC})| &= \frac{\Phi}{\eta L}\left(\frac{a_1}{\overline{x}_1} + \frac{a_2}{\overline{x}_2}\right)\gamma_2\left(1 - \frac{\gamma_2}{\alpha}\right). \end{aligned} \tag{5.22}$$

Proof. See Appendix A.3. □

We thus get the following result only involving testable conditions.

Corollary 5.1. *Under Assumption 5.3, the following assertions hold:*

1. *if $K < \frac{v}{L}$, then δ^* is asymptotically stable for any $\theta \geq 0$.*

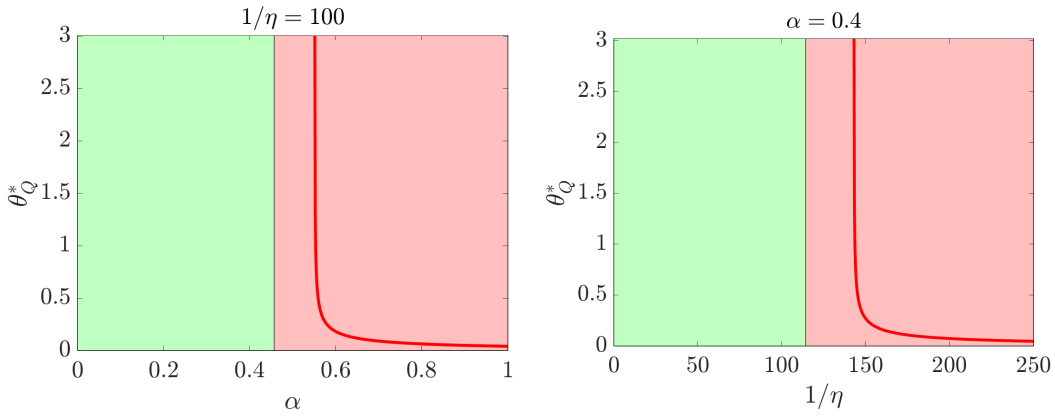

Figure 5.3: Qualitative behavior of θ_Q^* with respect to α and $1/\eta$. The left plot shows the behavior of θ_Q^* with respect to α for a fixed $1/\eta = 100$, while the right plot shows the behavior of θ_Q^* with respect to $1/\eta$ for a fixed $\alpha = 0.4$. The green zones indicate intervals of α and $1/\eta$ where delay-independent globally asymptotic stability is guaranteed, whereas the red zones indicate intervals where stability is not guaranteed. Note that θ_Q^* cannot be computed over the entire red intervals. For these values of α and $1/\eta$, we are unable to characterize the stability properties of the dynamics (5.8).

2. *if* $\frac{v}{L} < Q$, *then the second assertion of Theorem 5.2 holds and the critical delay value satisfies*

$$\theta^* < \theta_Q^* := \frac{1}{\sqrt{Q^2 - \frac{v^2}{L^2}}} \arccos\left(-\frac{v}{LQ}\right). \tag{5.23}$$

Proof. The first assertion follows trivially from the first assertion of Theorem 5.2 and the second inequality in (5.21). Similarly, the second assertion follows directly from the second assertion of Theorem 5.2, the first inequality in (5.21) and the fact that θ^* is an increasing function of $g'(\delta)$ when $g'(\delta) < -v/L$. □

The first assertion of Corollary 5.1 is in fact a special case of Theorem 5.1. The second assertion, instead, provides a sufficient condition for the instability of d^* and an upper bound for the critical delay θ^*, which are explicitly written as functions of the system parameters. We can easily deduce from (5.22) that Q is increasing in α and $1/\eta$ and therefore θ_Q^* is decreasing α and $1/\eta$. This fact provides us with some indications about the qualitative behavior of θ^* with respect to the above mentioned parameters, suggesting that increases in the penetration rate and in users' compliance reduce the delay threshold after which the system equilibrium is sure to loose its stability (see Figure 5.3). One can also verify that θ^* is decreasing in Φ, even though the relevance of this observation is tempered by the fact that too large demand can lead outside the set of assumptions under consideration.

Overall, the results provided in this section are consistent with those presented in Section 5.4.1: increases in Φ, α and $1/\eta$ negatively affect the system stability.

Remark 5.1. *Assumptions 5.2 and 5.3 have two interesting features. First, they select a set of parameters that is large enough to include realistic traffic scenarios, as will be demonstrated in the next section. Second, they focus on a very interesting situation, in which despite not having unsatisfied demand conditions both at equilibrium and in the absence of informed users, the destabilising effect of delays might lead to emergence of partial demand transfer, as we are going to show in the next section.*

5.5 Numerical examples

5.5.1 Homogeneous routes

Consider a network as the one in Figure 5.1 characterized by the following parameters:

$$\begin{gathered}\overline{f}_1 = 1200 \text{ veh/h}, \ \overline{f}_2 = 600 \text{ veh/h}, \\ x_1^c = 24 \text{ veh/km}, \ x_2^c = 12 \text{ veh/km}, \\ \overline{x}_1 = 120 \text{ veh/km}, \ \overline{x}_2 = 60 \text{ veh/km}, \\ a_1 = a_2 = 0.1 \text{ h (6 min)}, \ r_1^0 = 0.66, \ r_2^0 = 0.34.\end{gathered}$$

Suppose that the network is subject to a constant user demand of $\Phi = 1750$ veh/km, the length of the two routes is $L = 1.5$ km and the average free-flow speed is 50 km/h. These parameters have been chosen to represent a generic two-lane urban route and a generic one-lane urban route. Consider now two distinct values of θ, 1 minute and 6 minutes (which are realistic in a real world scenario [86, 87]), two distinct values of α, 0.4 and 0.7, and two values of $1/\eta$, 100 and 200. The numerical simulations in Figure 5.4 show the system behavior for the delays taken into consideration in three different cases:

- $\alpha = 0.4$, $1/\eta = 100$: in this case, $K \approx 29.17$, less than $v/L = 33.33$. Hence, Proposition 5.1 holds and, as one can see from the plots in the first column of Figure 5.4, the increase of delay does not alter the stability of the equilibrium point of the system.

- $\alpha = 0.7$, $1/\eta = 100$: in this case, $K \approx 51.04$ and Proposition 5.1 is no longer applicable. Moreover, $Q \approx 43.14$ and exceeds $v/L = 33.33$. Hence, by Corollary 5.1, we know that sufficiently high delays destabilise the system. Indeed, since $\theta_Q^* \approx 5$ min and 22 s, for $\theta = 6$ min the equilibrium point of the system is unstable and the trajectory of the system is oscillating.

- $\alpha = 0.4$, $1/\eta = 200$: similarly to the previous case, the decrease of noise destabilises the equilibrium point of the system when the system is affected by a delay of 6 minutes. Consistently with Corollary 5.1, in this case $K \approx 58.33$ and $Q \approx 46.60$, which are both greater than $v/L = 33.33$, and $\theta_Q^* \approx 4$ min and 22 s.

Observe how in the two cases in which the equilibrium point gets destabilised, the oscillations characterising the system trajectories cause periodic partial demand transfer.

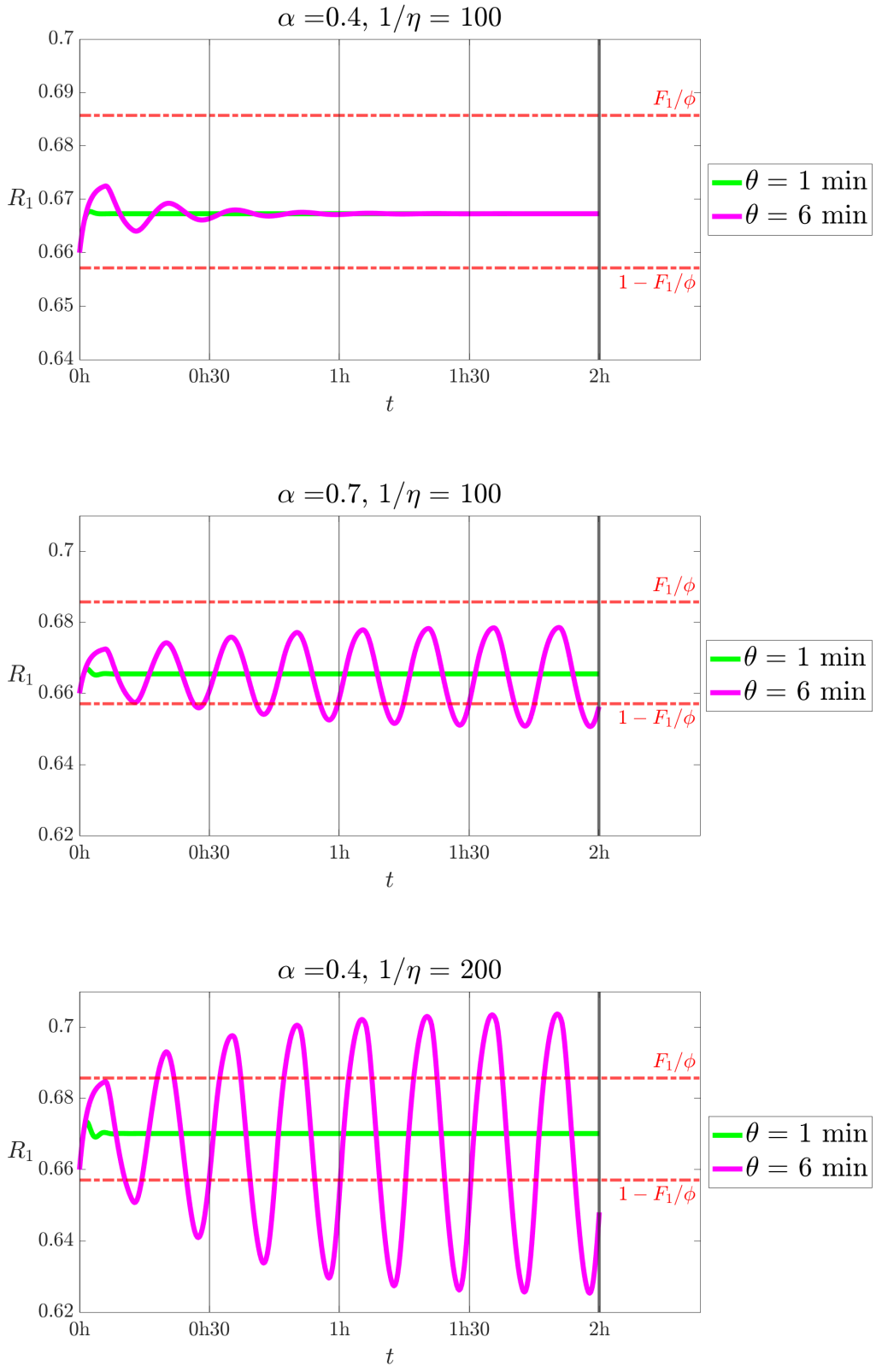

Figure 5.4: Each column of plots is associated with different values of the pair of parameters α, $1/\eta$, each line with different delay θ. The red dashed lines delimit the states for which unsatisfied demand is absent (in-between) or present (outside).

5.5.2 Beyond homogeneous routes

Although Assumption 5.1 allows us to reduce the system (5.4) to the scalar dynamics (5.8), greatly simplifying the stability analysis, this assumption limits the generality of the model. In this section, we propose numerical experiments to show that the qualitative behavior of the system with respect to the delay θ remains the same, even when Assumption 5.1 is not satisfied.

Consider a network as the one in Figure 5.1 characterized by the following parameters:

$$\begin{gathered}\overline{f}_1 = 1200 \text{ veh/h}, \ \overline{f}_2 = 600 \text{ veh/h}, \\ x_1^c = 24 \text{ veh/km}, \ x_2^c = 12 \text{ veh/km}, \\ \overline{x}_1 = 120 \text{ veh/km}, \ \overline{x}_2 = 60 \text{ veh/km}, \\ L_1 = 2 \text{ km}, \ L_2 = 1.5 \text{ km}, \ v_1 = 60 \text{ km/h}, \ v_2 = 50 \text{ km/h}, \\ a_1 = a_2 = 0.1 \text{ h (6 min)}, \ r_1^0 = 0.66, \ r_2^0 = 0.34.\end{gathered}$$

Clearly, this set of parameters does not satisfy Assumption 5.1, and therefore, one cannot reduce (5.4) to (5.8). Nevertheless, the plots in Figure 5.5 show that the behavior of the dynamics (5.4) is analogous to that of dynamics (5.8) in Section 5.5. When $\alpha = 0.4$, $1/\eta = 100$, increasing the delay affecting routing recommendations does not compromise the stability of the traffic dynamics. However, by either increasing α to $\alpha = 0.7$ or $1/\eta$ to $1/\eta = 200$, the equilibrium of the dynamics gets destabilised when the delay θ is increased from 1 min to 6 min.

This demonstrates that the destabilizing effect of delay on the traffic dynamics, along with the relationships between delay, penetration rate, and user compliance, persist even in more general network topologies that violate Assumption 5.1.

5.6 Concluding remarks

This chapter characterizes the impact of users who base their routing choices on recommendations affected by delays on the efficiency of a traffic network. The stability analysis of the model proposed in this chapter has shown that the impact of delay depends crucially on the penetration rate of informed users, their compliance, and the exogenous flow to which the network is subjected. For sufficiently high values of these parameters, the delay in routing recommendations can negatively affect the stability of the traffic state, causing destabilization. Such instability can adversely affect traffic flow efficiency, resulting in periodic partial demand transfer.

The results presented here closely align with those in Chapter 4. In both cases, it is evident that the penetration rate of informed users, their compliance, and the exogenous flow are directly related to network efficiency. Excessively high values for these parameters are likely to cause problems, specifically traffic state oscillations and partial demand transfer.

As with the model studied in Chapter 4, the primary direction for expanding this work involves an analysis of the proposed model for more complex network families. Again, the main challenge lies in conducting the stability study for this model, which is further complicated by the presence of delay.

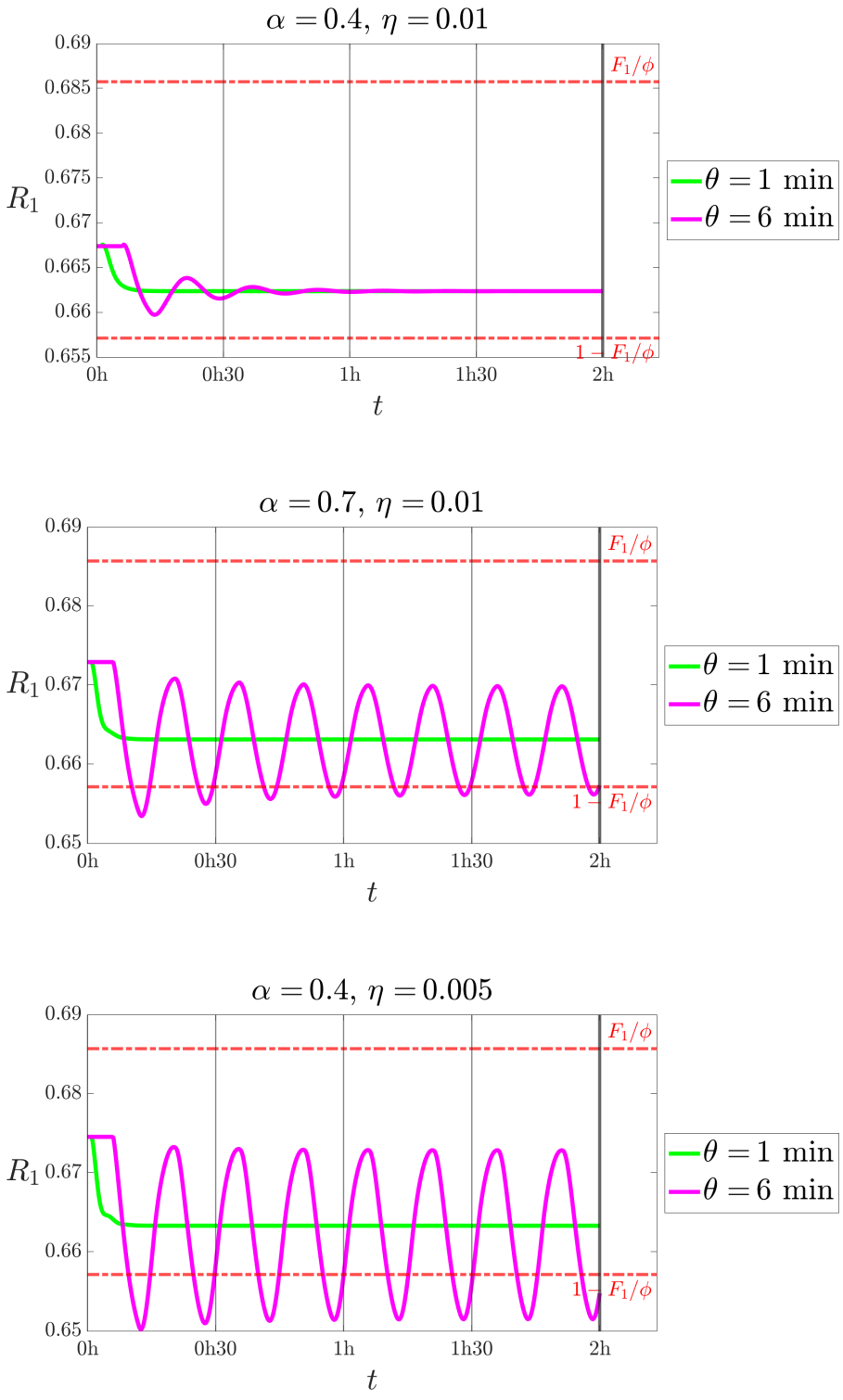

Figure 5.5: Each column of plots is associated with different values of the pair of parameters α, $1/\eta$, each line with different delay θ. The red dashed lines delimit the states for which unsatisfied demand is absent (in-between) or present (outside).

Part III

Coordinated routing & optimized fleets

Chapter 6

Coordinated routing

6.1 Introduction

This chapter delves into evaluating the impact of the presence of centralized fleet on traffic efficiency. Each fleet, under the control of a centralized operator, optimizes its operations through coordinated vehicle routing. We model the problem as an instance of the *mixed behavior NRGs* introduced in Section 2.2.3, involving two distinct user classes. The first class comprises individual, self-interested users aiming to minimize their personal travel times. The second class consists of vehicles that coordinate their routing decisions to minimize the fleet's average travel time.

This study aims to enhance our understanding of how recently born in mobility services, particularly ride-hailing, influence traffic efficiency. Companies in this sector leverage real-time traffic data to develop strategic fleet deployment plans that streamline operations, ensuring quality service and maximizing profitability, and employ coordinated routing strategies to optimize fleet performance. It is important to note that this study focuses exclusively on routing strategies and does not provide a comprehensive analysis of the broader impacts of these services on traffic efficiency, such as considerations of empty vehicle miles or potential effects on public transportation usage (see [23, 42, 43, 44, 45]).

The content presented in this chapter arises from a productive research visiting period at Cornell University. During my time there, from September to November 2023, I had the opportunity to collaborate with Professor Francesca Parise.

6.1.1 Summary of result

After observing that, under mild assumptions, this two-class game is equivalent to a two-person convex game, we exploit a well-known reformulation in terms of solution to a Variational Inequality (VI) to study the problem (see [9, 10]). Specifically, we identify sufficient conditions for the operator of the VI to be strongly monotone. On the one hand, strong monotonicity ensures equilibrium uniqueness. On the other hand, it allows for providing meaningful insights about the impact of the share of the coordinated fleet on the overall traffic efficiency in two-terminal networks. We employ the Price of Anarchy (PoA) as a metric for traffic efficiency. We demonstrate that both the unique equilibrium and the PoA exhibit Lipschitz continuity with

respect to the share of coordinated vehicles. Additionally, we establish conditions that ensure a minimum threshold for this share. Below this threshold, the presence of coordinated vehicles has no impact on traffic efficiency. Finally, for parallel networks, we show that the PoA, individual user flows, and the shortest travel time at equilibrium are all decreasing in the share of coordinated vehicles increases. This suggests that larger coordinated fleets lead to improved efficiency.

6.1.2 Related works

To the best of the authors' knowledge, information-aware coordinated routing has primarily been studied through mixed behavior NRGs. The concept of coordinating users within the same class was first introduced in [8]. The NRG in this study features a class of selfish users and a finite number of optimized fleets, establishing sufficient conditions for equilibrium existence and uniqueness. More genral conditions were later provided in [38].

Prior to [39], existing studies did not explore the overall impact on traffic efficiency. In [39], the authors address this gap by examining a three-class problem: selfish users, a coordinated fleet aiming to reduce fleet average travel time, and a system-optimal fleet aiming to minimize system-wide travel time. Numerical experiments from this work demonstrate that sufficiently large coordinated and system-optimal fleets can achieve system optimality, thereby enhancing overall traffic efficiency.

Recent attention has focused on two-class problems. Specifically, [88, 89, 90] investigate scenarios with selfish users and a system-optimal fleet. [88] and [89] derive methods to calculate the minimum proportion of system-optimal users needed to achieve system optimality, while [90] examines the trade-off between improvement magnitude and deployment costs for network managers.

In contrast, in [40, 41, 52] authors address two-class problems with selfish users and a coordinated fleet. In [40], an algorithm to compute traffic equilibrium resulting from interactions between selfish users and a coordinated fleet is developed, illustrating improved traffic efficiency with higher fractions of coordinated vehicles.

In [41], the authors initially provide an example in a network with multiple origin-destination pairs, demonstrating that coordinated fleets may adversely affect efficiency. They also develop mathematical programs to compute the minimum fleet size required for achieving system optimality and the maximum fleet size allowing user equilibrium to persist. Additionally, they offer analytical insights into the threshold effect of coordinated fleet size on efficiency, but for parallel networks only.

In [52], sufficient conditions for equilibrium existence and uniqueness are derived. Their work also proposes two algorithms for the computation of the equilibrium and a control scheme to converge to the equilibrium in a dynamical framework.

6.1.3 Chapter organization

Section 6.2 outlines the model and introduces the key concepts. Section 6.3 presents the conditions for strong monotonicity, existence, and uniqueness. In Section 6.4,

we analyze how the size of a coordinated fleet impacts traffic efficiency. Section 6.5 offers numerical experiments that support the findings from the previous sections. Finally, Section 6.6 provides concluding remarks.

6.2 Model definition

Consider a network $\mathcal{G} = (\mathcal{N}, \mathcal{L})$. Let $\mathcal{K} := \{k | k = (u, w),\ u, w \in \mathcal{N}\}$ be the set of OD pairs. Let $\mathcal{P} = \cup_{i=1}^{|\mathcal{K}|} \mathcal{P}_k$ be the set of all routes on $\mathcal{G}$. Suppose that $\mathcal{G}$ supports two classes of demand, namely class S and class C. Class S consists of selfish individual users, whereas class C consists of a coordinated fleet of vehicles.

Let Φ^i be the total demand of class i, $i = S, C$. Each OD pair k is subject to fractions $\Phi_k^i > 0$, $i = S, C$, of the total demand over $\mathcal{P}_k$, i.e, $\sum_{k \in \mathcal{K}} \Phi_k^i = \Phi^i$. Let $\Phi := \Phi^S + \Phi^C$ be the total demand. For each class $i \in \{S, C\}$, define the *route flow vector of class* i $z^i \in \mathbb{R}^P_{\geq 0}$ representing the traffic assignment of traffic demand Φ^i over the network routes. The set of feasible route flows of class i is

$$\mathcal{Z}^i := \left\{ z^i \in \mathbb{R}^P_{\geq 0} \ : \ \sum_{p \in \mathcal{P}_k} z_p^i = \Phi_k^i,\ \forall k \in \mathcal{K} \right\}$$

and let $\mathcal{Z} = \mathcal{Z}^S \times \mathcal{Z}^C$. Each flow vector z^i is associated with the *link flow vector of class* i, $f^i := Az^i$, $i = S, C$ representing the flow on each link of the network for class i, where A is the link-route incidence matrix:

$$A_{lp} = \begin{cases} 1 & \text{if } l \in p \\ 0 & \text{otherwise} \end{cases}.$$

Then, the set of feasible link flows of class i is

$$\mathcal{F}^i := \{ f^i \in \mathbb{R}^L_{\geq 0} \ : \ f^i = Az^i, \text{ for some } z^i \in \mathcal{Z}^i \}$$

and let $\mathcal{F} = \mathcal{F}^S \times \mathcal{F}^C$. Let the *route flow vector* $z := (z^S, z^C)$ and the *link flow vector* $f := (f^S, f^C)$ be the concatenations of the route and link flow vectors of the two classes and let $Z := z^S + z^C$, $F := f^S + f^C$ be the *aggregate route flow* and *aggregate link flow* vectors, respectively. The assignment of the two classes of vehicles is determined by the link travel time functions $\tau_l, l \in \mathcal{L}$, which are assumed to be the same for the two classes of vehicles.

Assumption 6.1. *For every $l \in \mathcal{L}$, the link travel time $\tau_l : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a non-negative, strictly increasing and $C^2([0, +\infty))$ function of the aggregate load F_l with $\tau_l'(0) > 0$.*

Assumption 6.1 imposes that link travel times are separable and that congestion is not-class specific. Therefore, the travel time of link l depends on the aggregate flow of link l only.

We are interested in characterizing the equilibrium link flow vectors of the traffic assignment problem emerging from the interaction of the vehicle classes S and C. To do this, we reformulate the problem as a *two-player game*, by associating each

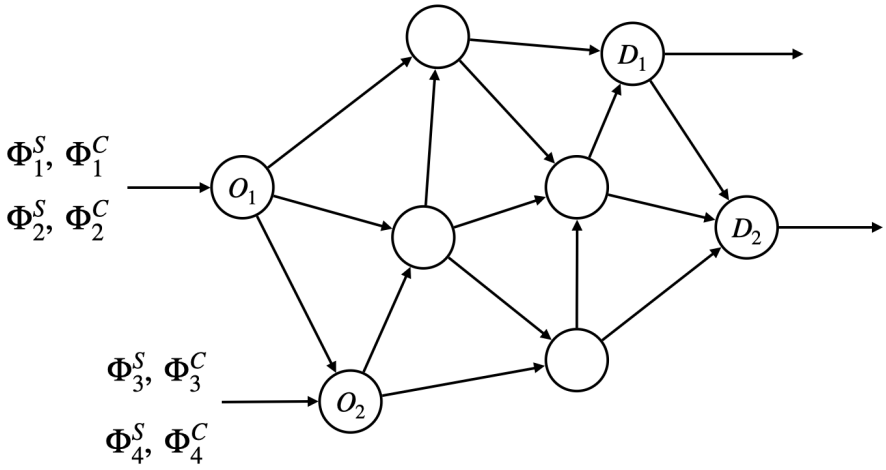

Figure 6.1: A multi-origin multi-destination network. This network has two origin nodes and two destination nodes, which amounts to four OD pairs: $k_1 = \{O_1, D_1\}$, $k_2 = \{O_1, D_2\}$, $k_3 = \{O_2, D_1\}$ and $k_4 = \{O_2, D_2\}$. Each OD pair k is assigned traffic demands Φ_k^S, Φ_k^C, from the two user classes S and C.

class to a strategic player (see Chapter 2). The strategy of each player corresponds to the link flow vector f^i with strategy set $\mathcal{F}^i$, $i = S, C$, respectively. The cost functions that player S and player C have to minimize in order to attain the goals of the traffic assignment problem are the following:

$$U^S(f) := \sum_{l \in \mathcal{L}} \int_0^{f_l^S} \tau_l(r + f_l^C) dr, \tag{6.1}$$

$$U^C(f) := \sum_{l \in \mathcal{L}} f_l^C \cdot \tau_l(F_l). \tag{6.2}$$

Definition 6.1 (Equilibria)**.** *An* equilibrium link flow *of the two-class congestion game is a link flow vector* $f^* = (f^{S*}, f^{C*})$ *such that*

$$\begin{aligned} f^{S*} &:= \arg\min_{f^S \in \mathcal{F}^S} U^S(f^S, f^{C*}), \\ f^{C*} &:= \arg\min_{f^C \in \mathcal{F}^S} U^C(f^{S*}, f^C). \end{aligned} \tag{6.3}$$

All the feasible route flows $z^* = (z^{S*}, z^{C*})$ *such that* $f^{i*} = Az^{i*}$, $i = S, C$ *are called* equilibrium route flows.

Observe that,

$$f_l^C \cdot \tau_l(F_l) = \int_0^{f_l^C} \left(\tau_l(f_l^S + r) + r \cdot \tau_l'(f_l^S + r)\right) dr.$$

Hence (6.2) can be rewritten as

$$U^C(f) = \sum_{l \in \mathcal{L}} \int_0^{f_l^C} \left(\tau_l(f_l^S + r) + r \cdot \tau_l'(f_l^S + r)\right) dr. \tag{6.4}$$

The functions inside the integral in (6.4), that is,

$$\mu_l(f_l) := \tau_l(F_l) + f_l^C \cdot \tau_l'(F_l). \tag{6.5}$$

are known as *marginal travel time functions* [1, Chapter 18].

We prove that under appropriate assumptions on the marginal travel times $\mu_l(f_l)$, the game in (6.3) is convex.

Lemma 6.1. *$U^S(f)$ is convex in f^S for any f^C. Moreover, if*

$$\frac{\partial \mu_l(f_l)}{\partial f_l^C} > 0, \quad \forall f_l^S,\ f_l^C \geq 0,\ \forall l \in \mathcal{L}, \tag{6.6}$$

then $U^C(f)$ is convex in f^C for any f^S.

Proof. First, since τ_l is twice continuously differentiable, $\forall l \in \mathcal{L}$, the same is true for U^s. The Hessian matrix of U^S with respect to f^S is

$$\nabla^2 U^S(f) = \operatorname{diag}\left(\tau_l'(F_l)\right) \succ 0$$

Hence, U^S is convex in f^S, for any f^C. As for U^C, condition (6.6) ensures that its Hessian matrix with respect to f^C is positive definite:

$$\nabla^2 U^C(f) = \operatorname{diag}\left(\frac{\partial \mu_l(f_l)}{\partial f_l^C}\right) \succ 0$$

Hence, U^C is convex in f^C, for any f^S. □

Remark 6.1. *The convexity of the cost functions (6.1) and (6.2) implies that any equilibrium route flow z^* must satisfy the following Wardrop conditions* [47, Chapter 3]*:*

$$z_p^{S*} > 0 \Rightarrow \sum_{l \in \mathcal{L}} A_{lp} \tau_l(F_l^*) \leq \sum_{l \in \mathcal{L}} A_{lr} \tau_l(F_l^*),\ \forall r \in \mathcal{P}, \tag{6.7}$$

$$z_p^{C*} > 0 \Rightarrow \sum_{l \in \mathcal{L}} A_{lp} \mu_l(f_l^*) \leq \sum_{l \in \mathcal{L}} A_{lr} \mu_l(f_l^*),\ \forall r \in \mathcal{P}. \tag{6.8}$$

In words, at equilibrium, each vehicle in class S uses a route among those of shortest travel time, whereas each vehicle in class C uses a route among those of shortest marginal travel time. Conditions (6.7) and (6.8) will be of key importance when proving the results in Section 6.4.

6.3 Variational inequality formulation

Under condition (6.6), the two-class routing game is convex and is equivalent to the following variational inequality [9, Proposition 1.4.2]:

$$(\phi - f^*)^\top H(f^*) \geq 0, \quad \forall \phi \in \mathcal{F}, \tag{6.9}$$

where

$$H(f) = \left((\tau_l(F_l))_{l \in \mathcal{L}}, (\mu_l(f_l))_{l \in \mathcal{L}}\right), \tag{6.10}$$

that is, equilibria of the two-class routing game correspond to solutions of (6.9).

The main result of this section consists in providing sufficient conditions for the operator H of such VI to be strongly monotone on $\Omega := [0, \Phi]^{2|\mathcal{L}|} \supset \mathcal{F}$, that is, for guaranteeing that

$$\exists c > 0 \, : \, (H(x) - H(y))^\top (x - y) \geq c||x - y||^2, \ \forall x, y \in \Omega. \tag{6.11}$$

The strong monotonicity of H not only ensures the uniqueness of the solution of (6.9) [9, Theorem 2.3.3], but also allows us to assess the impact of the fleet size onto traffic efficiency, as we shall demonstrate in the next section.

Proposition 6.1. *The operator H in (6.10) is strongly monotone on Ω if (6.6) holds and*

$$\tau_l'(F_l) > \frac{1}{4} \frac{\partial \mu_l(f_l)}{\partial f_l^C}, \quad \forall f_l^S, \, f_l^C \geq 0, \ \forall l \in \mathcal{L}. \tag{6.12}$$

Proof. From [9, Proposition 2.3.2], the operator H is strongly monotone on an open set $\mathcal{U}$ if and only if its jacobian matrix is uniformly positive definite on $\mathcal{U}$, i.e.,

$$\exists \zeta > 0 : \ \phi^\top J_H(f) \phi \geq \zeta ||\phi||^2, \ \forall \phi \in \mathbb{R}_+^{2 \cdot |\mathcal{L}|}, \ \forall f \in \mathcal{U}.$$

The condition above is equivalent to

$$\exists \zeta > 0 : \ J_H(f) - \zeta I \succeq 0, \ \forall f \in \mathcal{U}. \tag{6.13}$$

The proof proceeds in two steps: i) using the fact above we show that (6.12) implies that H is strongly monotone on $\text{int}(\Omega)$, ii) we show that strong monotonicity extends to Ω by continuity.

i) We study the positive semi-definiteness of $J_H - \zeta I$ by examining its symmetric part $J_H^{\text{sym}}(f) - \zeta I$, where $J_H^{\text{sym}}(f)$ is the symmetric part of J_H. Define

$$\Sigma_1(f) := \text{diag}(\tau_l'(F_l)) - \zeta I, \quad \Sigma_2(f) := \text{diag}\left(\frac{1}{2} \frac{\partial \mu_l(f_l)}{\partial f_l^C} \right),$$

$$\Sigma_3(f) := \text{diag}\left(\frac{\partial \mu_l(f_l)}{\partial f_l^C} \right) - \zeta I,$$

then

$$J_H^{\text{sym}}(f) - \zeta I = \begin{pmatrix} \Sigma_1(f) & \Sigma_2(f) \\ \Sigma_2(f) & \Sigma_3(f) \end{pmatrix},$$

where we used the identity $\tau_l'(F_l) + \frac{\partial \mu_l(f_l)}{\partial f_l^S} = \frac{\partial \mu_l(f_l)}{\partial f_l^C}$. If $\Sigma_1(f)$ is positive definite, then $J_H^{\text{sym}}(f) - \zeta I$ is positive semi-definite if and only if its Schur complement $\Sigma(f) := \Sigma_3(f) - (\Sigma_1(f))^{-1} \Sigma_2^2(f)$ is. $\Sigma_1(f)$ and $\Sigma(f)$ are positive definite and positive semi-definite in $\text{int}(\Omega)$, respectively, if the following conditions hold for all $l \in \mathcal{L}$:

$$\tau_l'(F_l) - \zeta > 0, \quad \forall f_l^S, \, f_l^C \in (0, \Phi), \tag{6.14}$$

$$\frac{\partial \mu_l(f_l)}{\partial f_l^C} - \zeta - \frac{\left(\frac{\partial \mu_l(f_l)}{\partial f_l^C} \right)^2}{4(\tau_l'(F_l) - \zeta)} \geq 0, \quad \forall f_l^S, \, f_l^C \in (0, \Phi). \tag{6.15}$$

By Assumption 6.1,

$$\exists \zeta_1 > 0 \,:\, \zeta_1 < \min_{l\in\mathcal{L}} \min_{f_l^S,\, f_l^C \in [0,\Phi]} \tau_l'(F_l).$$

Hence (6.14) holds for any $\zeta \leq \zeta_1$. Now, let

$$K_l(\zeta, f_l) = \frac{\partial \mu_l(f_l)}{\partial f_l^C} - \zeta - \frac{\left(\frac{\partial \mu_l(f_l)}{\partial f_l^C}\right)^2}{4(\tau_l'(F_l) - \zeta)},$$

$$K(\zeta) := \min_{l\in\mathcal{L}} \min_{f_l^S,\, f_l^C \in [0,\Phi]} K_l(\zeta, f_l).$$

We aim at proving that $K(\zeta) > 0$ for ζ small enough, as that would imply (6.15). To this end, observe that given (6.6), (6.12) is equivalent to

$$\frac{\partial \mu_l(f_l)}{\partial f_l^C} - \frac{\left(\frac{\partial \mu_l(f_l)}{\partial f_l^C}\right)^2}{4\tau_l'(F_l)} > 0, \quad \forall f_l^S,\, f_l^C \geq 0,\ \forall l \in \mathcal{L}.$$

Since the left-hand side of the above condition is continuous in f_l and the condition holds strictly for every $l \in \mathcal{L}$ and any $f_l^S,\, f_l^C \in [0, \Phi]$, then $K(0) > 0$. We next prove that $K(\zeta)$ is continuous in $\zeta \in I = [0, \zeta_1)$ by showing that

$$K_l(\zeta) := \min_{f_l^S,\, f_l^C \in [0,\Phi]} K_l(\zeta, f_l)$$

is continuous, for every l. By continuity in both arguments of $K_l(\zeta, f_l)$, for every $\epsilon > 0$,

$$\exists \delta > 0 \,:\, |\zeta - \varrho| + ||f_l - g_l|| < \delta \Rightarrow |K_l(\zeta, f_l) - K_l(\varrho, g_l)| < \epsilon. \tag{6.16}$$

Take $\zeta_3,\, \zeta_4 \in I$ such that $|\zeta_3 - \zeta_4| < \delta$ and define the minimizers $f_l^{\zeta_i} \in [0, \Phi]^2 \,:\, K_l(\zeta_i, f_l^{\zeta_i}) = K_l(\zeta_i)$. Then, by (6.16) with $f_l = g_l = f_l^{\zeta_i}$ we obtain

$$0 \leq |K_l(\zeta_i, f_l^{\zeta_i}) - K_l(\zeta_j, f_l^{\zeta_i})| < \epsilon, \quad i = 3, 4,\ i \neq j.$$

Hence,

$$K_l(\zeta_i, f_l^{\zeta_i}) > K_l(\zeta_j, f_l^{\zeta_i}) - \epsilon \geq K_l(\zeta_j) - \epsilon,\ i = 3, 4,\ i \neq j.$$

The above implies

$$K_l(\zeta_i) > K_l(\zeta_j) - \epsilon, \quad i = 3, 4,\ i \neq j.$$

Combining the two conditions above we get

$$|K_l(\zeta_3) - K_l(\zeta_4)| < \epsilon.$$

Hence, $K_l(\zeta)$ is continuous, $\forall l \in \mathcal{L}$, thus $K(\zeta)$ is continuous, as it is point-wise minimum of continuous functions.

The continuity of $K(\zeta)$ together with $K(0) > 0$, implies that there exists ζ_2 such that (6.15) is satisfied for all $l \in \mathcal{L}$, for any $\zeta \in [0, \zeta_2)$. The existence of ζ_1 and ζ_2 ensure the existence of $\zeta > 0$ such that (6.14) and (6.15) hold for all $l \in \mathcal{L}$.

Therefore, there exists a ζ small enough such that (6.13) holds on $\text{int}(\Omega)$, thus H is strongly monotone in $\text{int}(\Omega)$:

$$\exists c > 0 \,:\, (H(\tilde{x}) - H(\tilde{y}))^\top (\tilde{x} - \tilde{y}) \geq c||\tilde{x} - \tilde{y}||^2, \;\forall \tilde{x}, \tilde{y} \in \text{int}(\Omega). \tag{6.17}$$

ii) Now, observe that $\text{cl}(\text{int}(\Omega)) = \Omega$. Then, consider any $x, y \in \Omega$ and let $\{x^{(n)}\}$, $\{y^{(n)}\} \subset \text{int}(\Omega)$ be two sequences converging to x and y, respectively. Then,

$$(H(x^{(n)}) - H(y^{(n)}))^\top (x^{(n)} - y^{(n)}) \geq c||x^{(n)} - y^{(n)}||^2, \;\forall n.$$

By taking the limit and using the continuity of H,

$$(H(x) - H(y))^\top (x - y) \geq c||x - y||^2.$$

This means that strong monotonicity of H extends to Ω. □

The strong monotonicity of H ensures the uniqueness of the solution of (6.9), that is, of the equilibrium link flow f^*. In [52], weaker conditions similar to (6.12) were derived to ensure the uniqueness of the equilibrium link flow. Our slightly stronger conditions are needed to guarantee that H is strongly monotone and that thus the following assumption holds.

Assumption 6.2. *Suppose that the operator H in (6.10) is Lipschitz and strongly monotone in $\Omega = [0, \Phi]^{2|\mathcal{L}|}$.*

Again, we remark that sufficient conditions for strong monotonicity to hold are given in Proposition 6.1, whereas Lipschitz continuity follows from the smoothness of travel time and marginal travel time functions (defined on a compact set).

Remark 6.2. *A class of travel time functions that satisfy conditions (6.6) and (6.12), thereby ensuring strong monotonicity of (6.10), consists of polynomial functions of degree at most 3 with non-negative coefficients and strictly positive derivatives on $[0, +\infty)$, see [52] for similar examples. This demonstrates that assuming strong monotonicity is not too restrictive, as this property holds for a relevant class of travel time functions.*

6.4 Price of Anarchy

For this two-class game, the PoA is still defined as the ratio between the total travel time attained at the (unique under Assumption 6.2) equilibrium f^* and the minimum total travel time:

$$\text{PoA}(f^*) := \frac{\sum_{l \in \mathcal{L}} F_l^* \cdot \tau_l(F_l^*)}{\sum_{l \in \mathcal{L}} F_l^\omega \cdot \tau_l(F_l^\omega)} \geq 1. \tag{6.18}$$

We aim to study how the size of the coordinated fleet affects the PoA of the overall system. From now on, we focus our attention on *two-terminal networks*.

Assumption 6.3. *The network has a single OD pair. Let $\Phi^S = (1 - \gamma)\Phi$ and $\Phi^C = \gamma\Phi$ represent the demand of class S and C entering the network from its unique origin, where γ is the* share of class C*, which we refer to as the fleet share.*

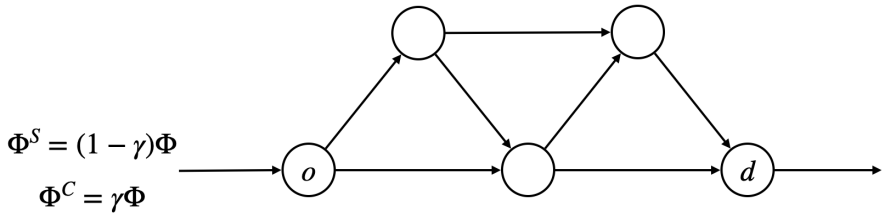

Figure 6.2: A two-terminal network.

We provide three main results in this section. First, we prove that the equilibrium link flow and the PoA are Lipschitz continuous functions of the fleet share γ. Second, we derive a sufficient condition for the existence of a minimum fleet share below which the coordinated fleet has no impact on the PoA. Finally, we show that the PoA of the equilibrium link flow is a non-increasing function of γ for the case of parallel networks. To make explicit their dependence on γ, we will indicate the feasible set by $\mathcal{F}(\gamma)$ the equilibrium link flow as $f^*(\gamma)$ and we will indicate as $\text{PoA}(\gamma)$, $\tau_l(\gamma)$, $\mu_l(\gamma)$ the PoA and the associated travel time and marginal travel time functions at equilibrium.

6.4.1 Lipschitz continuity

Proposition 6.2. *Let Assumptions 6.2 and 6.3 hold. The equilibrium link flow $f^*(\gamma)$ is Lipschitz continuous in γ, i.e., there exists $k > 0$ such that*

$$\forall \gamma_1, \gamma_2, \quad ||f^*(\gamma_2) - f^*(\gamma_1)|| \leq k|\gamma_2 - \gamma_1|. \tag{6.19}$$

Proof. Without loss of generality, assume that $\gamma_1 < \gamma_2$. Define $v := (-v^S, v^C)$, where

$$v^S := \frac{\gamma_2 - \gamma_1}{1 - \gamma_1} f^{S*}(\gamma_1), \quad v^C := \frac{\gamma_2 - \gamma_1}{\gamma_2} f^{C*}(\gamma_2),$$

are scaled versions of $f^{S*}(\gamma_1)$ and $f^{C*}(\gamma_2)$, respectively, both associated with a total demand equal to $(\gamma_2 - \gamma_1)\Phi$ and such that

$$0 \leq v^S \leq f^{S*}(\gamma_1), \quad 0 \leq v^C \leq f^{C*}(\gamma_2).$$

Since v^S and v^C are both associated with a total demand equal to $(\gamma_2 - \gamma_1)\Phi$, it must hold that

$$0 \leq v_l^i \leq (\gamma_2 - \gamma_1)\Phi, \quad \forall l \in \mathcal{L},\ i = S, C,$$

which implies that $||v||^2 \leq 2|\mathcal{L}|\Phi^2(\gamma_2 - \gamma_1)^2$. Hence $||v|| \leq k'|\gamma_2 - \gamma_1|$, with $k' = \sqrt{2|\mathcal{L}|}\Phi$. Now, define

$$f^{(1)} = f^*(\gamma_2) - v \in \mathcal{F}(\gamma_1), \quad f^{(2)} = f^*(\gamma_1) + v \in \mathcal{F}(\gamma_2).$$

By (6.9), one can write

$$(f^{(i)} - f^*(\gamma_i))^\top H(f^*(\gamma_i)) \geq 0, \quad i = 1, 2.$$

By summing these two inequalities and using the definition of $f^{(i)}$, one gets

$$\begin{aligned}&(H(f^*(\gamma_2)) - H(f^*(\gamma_1)))^\top v \geq \\ &(H(f^*(\gamma_2)) - H(f^*(\gamma_1)))^\top (f^*(\gamma_2) - f^*(\gamma_1)) \geq \\ &c||f^*(\gamma_2) - f^*(\gamma_1)||^2,\end{aligned} \tag{6.20}$$

where the last line follows from strong monotonicity of H over Ω (notice that $\mathcal{F}(\gamma_i) \subset \Omega,\ i = 1, 2$). From Cauchy-Schwartz inequality and (6.20)

$$\begin{aligned}&c||f^*(\gamma_2) - f^*(\gamma_1)||^2 \leq ||H(f^*(\gamma_1)) - H(f^*(\gamma_2))||\,||v|| \leq \\ &\leq Q \cdot ||f^*(\gamma_1) - f^*(\gamma_2)|| \cdot k'|\gamma_2 - \gamma_1|,\end{aligned}$$

where Q is the Lipschitz constant of H. The proof is complete by picking $k = Qk'/c$. □

Since the PoA is Lipschitz continuous in the equilibrium link flow and link flows are defined on a bounded set, the PoA is also a Lipschitz continuous function of γ.

6.4.2 Critical fleet share

A first question that one may ask is if introducing a coordinated fleet always helps in reducing the PoA. In this section, we show that this is not the case. Specifically, we derive a sufficient condition under which there is a positive minimum critical fleet size needed to induce changes in the PoA.

Theorem 6.1 (Critical fleet size)**.** *Let $\mathcal{P}^i(z(\gamma))$ indicate the set of routes used by class i at the equilibrium route flow $z(\gamma)$, $i = C, S$, respectively. Let Assumptions 6.2 and 6.3 hold. Suppose that $\exists \tilde{\gamma} \in (0, 1)$ such that $f^*(\tilde{\gamma})$ admits an equilibrium route flow $z^*(\tilde{\gamma})$ such that $\mathcal{P}^C(z^*(\tilde{\gamma})) \subseteq \mathcal{P}^S(z^*(\tilde{\gamma}))$. Then,*

$$f^*(\gamma) = \left(f^{S^*}(\tilde{\gamma}) + \frac{\tilde{\gamma} - \gamma}{\tilde{\gamma}} f^{C^*}(\tilde{\gamma}), \frac{\gamma}{\tilde{\gamma}} f^{C^*}(\tilde{\gamma})\right), \tag{6.21}$$

and $\mathrm{PoA}(\gamma) = \mathrm{PoA}(0),\ \forall \gamma \in [0, \tilde{\gamma}]$.

Proof. Consider the equilibrium link flow $f^*(\tilde{\gamma})$ and the associated equilibrium route flow $z^*(\tilde{\gamma})$. Clearly,

$$\begin{aligned}z_p^{S^*}(\tilde{\gamma}) > 0 \Rightarrow \sum_{l \in \mathcal{L}} A_{lp}\tau_l(\tilde{\gamma}) \leq \sum_{l \in \mathcal{L}} A_{lq}\tau_l(\tilde{\gamma}),\ \forall q \in \mathcal{P}, \\ z_p^{C^*}(\tilde{\gamma}) > 0 \Rightarrow \sum_{l \in \mathcal{L}} A_{lp}\mu_l(\tilde{\gamma}) \leq \sum_{l \in \mathcal{L}} A_{lq}\mu_l(\tilde{\gamma}),\ \forall q \in \mathcal{P}.\end{aligned} \tag{6.22}$$

Consider the following feasible route flow (obtained by moving part of the flow from C to S):

$$z^*(\gamma) = \left(z^{S^*}(\tilde{\gamma}) + \frac{\tilde{\gamma} - \gamma}{\tilde{\gamma}} z^{C^*}(\tilde{\gamma}), \frac{\gamma}{\tilde{\gamma}} z^{C^*}(\tilde{\gamma})\right). \tag{6.23}$$

We show that $z^*(\gamma)$ is an equilibrium route flow when the fleet share is γ, i.e.,

$$\begin{aligned} z_p^{S*}(\gamma) > 0 \Rightarrow \sum_{l\in\mathcal{L}} A_{lp}\tau_l(\gamma) \leq \sum_{l\in\mathcal{L}} A_{lq}\tau_l(\gamma),\ \forall q \in \mathcal{P}, \\ z_p^{C*}(\gamma) > 0 \Rightarrow \sum_{l\in\mathcal{L}} A_{lp}\mu_l(\gamma) \leq \sum_{l\in\mathcal{L}} A_{lq}\mu_l(\gamma),\ \forall q \in \mathcal{P}. \end{aligned} \tag{6.24}$$

We prove each of the above conditions. The first easily follows after noticing that i) $z^*(\gamma)$ and $z^*(\tilde{\gamma})$ induce the same aggregate link flow, i.e., $F_l^*(\gamma) = F_l^*(\tilde{\gamma})$, so none of the route travel times has changed, and ii) the set of routes used by vehicles in class S is the same, i.e., $\mathcal{P}^S(z^*(\tilde{\gamma})) = \mathcal{P}^S(z^*(\gamma))$ (since $\mathcal{P}^C(z^*(\tilde{\gamma})) \subseteq \mathcal{P}^S(z^*(\tilde{\gamma}))$). Hence, the first inequality in (6.22) ensures that all vehicles in class S still use shortest travel time routes. As for the second condition, similarly, one has to prove that vehicles in class C still use shortest marginal travel time routes. Because of the expression of (6.23), one can observe that:

- $\mathcal{P}^C(z^*(\gamma)) = \mathcal{P}^C(z^*(\tilde{\gamma})) \subseteq \mathcal{P}^S(z^*(\tilde{\gamma}))$;
- for every $p \in \mathcal{P}$, since the aggregate link flows have not changed, the marginal travel time is
$$\mu_p(\gamma) = \sum_{l\in\mathcal{L}} A_{lp}\left(\tau_l(\tilde{\gamma}) + \frac{\gamma}{\tilde{\gamma}} f_l^{C*}(\tilde{\gamma})\tau_l'(\tilde{\gamma})\right),$$

By multiplying the first inequality in (6.22) by $1-\gamma/\tilde{\gamma}$, the second one by $\gamma/\tilde{\gamma}$, then summing them, one obtains

$$\begin{aligned} \mu_p(\gamma) &= \sum_{l\in\mathcal{L}} A_{lp}\left(\tau_l(\tilde{\gamma}) + \frac{\gamma}{\tilde{\gamma}} f_l^{C*}(\tilde{\gamma})\tau_l'(\tilde{\gamma})\right) \leq \\ &\leq \sum_{l\in\mathcal{L}} A_{lq}\left(\tau_l(\tilde{\gamma}) + \frac{\gamma}{\tilde{\gamma}} f_l^{C*}(\tilde{\gamma})\tau_l'(\tilde{\gamma})\right) = \mu_q(\gamma), \end{aligned} \tag{6.25}$$

$\forall p \in \mathcal{P}^C(z^*(\gamma)),\ \forall q \in \mathcal{P}$. Hence, every $\mathcal{P}^C(z^*(\gamma))$ is still a shortest marginal travel time route. Therefore, $z^*(\gamma)$ is a equilibrium route flow when the fleet share is equal to γ. The equilibrium link flow associated with $z^*(\gamma)$ is

$$f^*(\gamma) = \left(f^{S*}(\tilde{\gamma}) + \frac{\tilde{\gamma}-\gamma}{\tilde{\gamma}} f^{C*}(\tilde{\gamma}), \frac{\gamma}{\tilde{\gamma}} f^{C*}(\tilde{\gamma})\right),$$

which must correspond to the unique equilibrium of the problem.

To conclude, notice that for all $\gamma \in [0, \tilde{\gamma}]$ all links have the same aggregate link flow. Hence $\mathrm{PoA}(\gamma) = \mathrm{PoA}(0)$ for all $\gamma \in [0, \tilde{\gamma}]$. □

6.4.3 PoA monotonicity for Parallel Networks

In this section, we show that the PoA is non-increasing in the fleet share γ under the following assumptions.

Assumption 6.4. *$\mathcal{G}$ consists of an OD pair connected by finitely many links directed from the origin node to the destination node. Again, let γ be the fleet share.*

Assumption 6.5. *The travel time function τ_l is convex, $\forall l \in \mathcal{L}$.*

Note that Assumption 6.4 pertains to a subclass of parallel networks (see Section 2.1), where each route consists of a single link. For simplicity, the following results are formulated based on this specific subclass. However, these results can straightforwardly be extended to encompass the entire class of parallel networks. Briefly, for any parallel network, one can trace back the problem to a network where each route corresponds to a single link, with the link travel time corresponding to the route travel time.

The assumption of parallel networks simplifies the analysis as, in that case, the notion of link and route coincides. Let $\mathcal{L}^i(\alpha)$ indicate the set of links used by class i at the equilibrium $f^*(\alpha)$, $i = S, C$. The convexity of the travel time functions instead ensures that $\tau_l'(F_l)$ is non-decreasing in F_l. Note that in particular this implies the following monotonicity property

$$\bar{F}_l > F_l \text{ and } \bar{f}_l^C > f_l^C \quad \Rightarrow \quad \mu_l(\bar{f}_l) > \mu_l(f_l). \tag{6.26}$$

Let $\overline{\tau}(\gamma)$ and $\overline{\mu}(\gamma)$ indicate the *minimum travel time* and the *minimum marginal travel time* realised at equilibrium when the fleet share is γ, respectively. Observe that, since links and routes coincide, the equilibrium condition implies

$$l \in \mathcal{L}^S(\gamma) \Rightarrow \tau_l(F_l^*(\gamma)) = \overline{\tau}(\gamma),$$

$$l \in \mathcal{L}^C(\gamma) \Rightarrow \mu_l(f_l^*(\gamma)) = \overline{\mu}(\gamma).$$

Proposition 6.3. *Let Assumptions 6.2, 6.4 and 6.5 hold. Suppose there exists $\gamma_1, \gamma_2 \in (0,1)$ such that $\gamma_1 < \gamma_2$ and $\mathcal{L}^S(\gamma_1) = \mathcal{L}^S(\gamma_2)$ and $\mathcal{L}^C(\gamma_1) = \mathcal{L}^C(\gamma_2)$. Then,*

1. $\overline{\tau}(\gamma_1) \geq \overline{\tau}(\gamma_2)$;
2. $\overline{\mu}(\gamma_1) \leq \overline{\mu}(\gamma_2)$;
3. $f_l^{S*}(\gamma_1) \geq f_l^{S*}(\gamma_2), \ \forall l \in \mathcal{L}$;
4. $f_l^{C*}(\gamma_1) \leq f_l^{C*}(\gamma_2), \ \forall l \in \mathcal{L}$.

Proof. Since $\mathcal{L}^i(\gamma_1) = \mathcal{L}^i(\gamma_2)$, $i = S, C$, let us indicate both as $\mathcal{L}^i$, $i = S, C$ for convenience. Along with them, consider also the set $\mathcal{L}^{C\setminus S} := \mathcal{L}^C \setminus (\mathcal{L}^S \cap \mathcal{L}^C)$, corresponding to the set of links used by class C only. Notice that also this set remains constant in passing from γ_1 to γ_2. Also, since it is used by vehicles of class C only,

$$f_l^{C*}(\gamma_i) = F_l^*(\gamma_i), \ \forall l \in \mathcal{L}^{C\setminus S}, \ i = 1, 2. \tag{6.27}$$

We distinguish two cases: If $\mathcal{L}^{C\setminus S} = \emptyset$ the conclusion follows from Theorem 6.1. We next discuss the case in which $\mathcal{L}^{C\setminus S} \neq \emptyset$.
1) By contradiction, suppose that $\overline{\tau}(\gamma_1) < \overline{\tau}(\gamma_2)$. This implies that the aggregate link flow increased on all links in $\mathcal{L}^S$, i.e., $F_l^*(\gamma_1) < F_l^*(\gamma_2)$, $\forall l \in \mathcal{L}^S$. Now, since the demand of class S decreased, there must exist a link $j \in \mathcal{L}^S$ such that the link flow of class S on it decreased, i.e., $f_j^{S*}(\gamma_1) > f_j^{S*}(\gamma_2)$. The latter fact, combined with the increase of the aggregate link flows on all link in $\mathcal{L}^S$, implies that the link

flow of class C on link j increased, i.e., $f_j^{C*}(\gamma_1) < f_j^{C*}(\gamma_2)$. By (6.26), the increase of both the aggregate link flow and the link flow of class C on link j implies that its marginal travel time increased. Hence, $\overline{\mu}(\gamma_1) < \overline{\mu}(\gamma_2)$.

On the other hand, the fact that the aggregate link flow increased on all links in $\mathcal{L}^S$ implies that the aggregate demand directed toward the set $\mathcal{L}^S$ increased, which is equivalent to say that the aggregate demand toward the set $\mathcal{L}^{C\setminus S}$ decreased. Then, there must be at least one link $e \in \mathcal{L}^{C\setminus S}$ whose aggregate link flow decreased, i.e., $F_e^*(\gamma_1) > F_e^*(\gamma_2)$. From (6.27), this is equivalent to $f_e^{C*}(\gamma_1) > f_e^{C*}(\gamma_2)$, which implies that $\overline{\mu}(\gamma_1) > \overline{\mu}(\gamma_2)$, contradicting what proved above. Therefore, $\overline{\tau}(\gamma_1) \geq \overline{\tau}(\gamma_2)$.

2) From 1), $\overline{\tau}(\gamma_1) \geq \overline{\tau}(\gamma_2)$, which implies that the aggregate link flow on none of the links in $\mathcal{L}^S$ can increase. This implies that the aggregate demand toward $\mathcal{L}^S$ cannot increase, which is equivalent to say that the aggregate demand toward $\mathcal{L}^{C\setminus S}$ cannot decrease. From (6.27), this means that the demand associated with class C directed toward $\mathcal{L}^{C\setminus S}$ did not decreased. Hence, there exists $e \in \mathcal{L}^{C\setminus S}$ such that $f_e^{C*}(\gamma_1) \leq f_e^{C*}(\gamma_2)$. Hence, $\overline{\mu}(\gamma_1) \leq \overline{\mu}(\gamma_2)$.

3) By contradiction, suppose that $\exists l \in \mathcal{L}^S \mid f_l^{S*}(\gamma_1) < f_l^{S*}(\gamma_2)$. Since on all links in $\mathcal{L}^S$ the aggregate link flow did not increase ($F_l(\gamma_1) \geq F_l(\gamma_2)$), the above implies that $f_l^{C*}(\gamma_1) > f_l^{C*}(\gamma_2)$. This implies $\overline{\mu}(\gamma_1) > \overline{\mu}(\gamma_2)$, contradicting point 2).
4) Suppose that there $\exists l \in \mathcal{L}^C \mid f_l^{C*}(\gamma_1) > f_l^{C*}(\gamma_2)$. By point 3) we also know that $f_l^{S*}(\gamma_1) \geq f_l^{S*}(\gamma_2)$. Hence $F_l(\gamma_1) > F_l(\gamma_2)$. By (6.26), this implies $\overline{\mu}(\gamma_1) > \overline{\mu}(\gamma_2)$, which contradicts 2). □

Remark 6.3. *The result above and its proof implicitly assumes that* $(\mathcal{L}^S \cap \mathcal{L}^C) \neq \emptyset$. *To see that this is always true, assume by contradiction that* $(\mathcal{L}^S \cap \mathcal{L}^C) = \emptyset$. *Then, it follows*

$$\forall l \in \mathcal{L}^S(\gamma), \quad \mu_l(f_l^*(\gamma)) = \tau_l(F_l^*(\gamma)) = \overline{\tau}(\gamma),$$
$$\forall e \in \mathcal{L}^C(\gamma), \quad \mu_e(f_e^*(\gamma)) > \tau_e(F_e^*(\gamma)) \geq \overline{\tau}(\gamma),$$

which is impossible as vehicles in class C at equilibrium must use links of minimal marginal travel time.

Proposition 6.4. *Let Assumptions 6.2, 6.4 and 6.5 hold. Suppose there exists* $\gamma_1, \gamma_2 \in (0,1)$ *such that* $\gamma_1 < \gamma_2$ *and* $\mathcal{L}^S(\gamma_1) = \mathcal{L}^S(\gamma_2)$ *and* $\mathcal{L}^C(\gamma_1) = \mathcal{L}^C(\gamma_2)$. *Then,* $\mathrm{PoA}(\gamma_1) \geq \mathrm{PoA}(\gamma_2)$.

Proof. First of all, notice that it suffices to consider only the numerator of the PoA (6.18), as its denominator is constant. Let us indicate the numerator of (6.18) as $T(f^*(\gamma))$. Now, $T(f^*(\gamma))$ can be written as follows:

$$T(f^*(\gamma)) = \sum_{l\in\mathcal{L}} f_l^{S*}(\gamma)\cdot\tau_l(\gamma) + \sum_{l\in\mathcal{L}} f_l^{C*}(\gamma)\cdot\tau_l(\gamma) =:$$
$$:= T^S(f^*(\gamma)) + T^C(f^*(\gamma)).$$

From 1) of Proposition 6.3,

$$\begin{aligned} T^S(f^*(\gamma_2)) &= \sum_{l\in\mathcal{L}} f_l^{S^*}(\gamma_2)\cdot\tau_l(F_l^*(\gamma_2)) = \\ &= \overline{\tau}(\gamma_2)\sum_{l\in\mathcal{L}} f_l^{S^*}(\gamma_2) \leq \\ &\leq \overline{\tau}(\gamma_1)\sum_{l\in\mathcal{L}} f_l^{S^*}(\gamma_2) = \\ &= \sum_{l\in\mathcal{L}} f_l^{S^*}(\gamma_2)\cdot\tau_l(F_l^*(\gamma_1)). \end{aligned} \tag{6.28}$$

Moreover, because of 3) of Proposition 6.3, one can observe that

$$f^{C^*}(\gamma_1) + (f^{S^*}(\gamma_1) - f^{S^*}(\gamma_2)) \in \mathcal{F}^C(\gamma_2).$$

Therefore, if one defines $\varphi := f^{S^*}(\gamma_1) - f^{S^*}(\gamma_2) \geq 0$:

$$\begin{aligned} T^C(f^*(\gamma_2)) &= \sum_{l\in\mathcal{L}} f_l^{C^*}(\gamma_2)\cdot\tau_l(F_l^*(\gamma_2)) \leq \\ &\leq \sum_{l\in\mathcal{L}} (f_l^{C^*}(\gamma_1) + \varphi)\cdot\tau_l(F_l^*(\gamma_1)), \end{aligned} \tag{6.29}$$

where the inequality follows from the fact $f_l^{C^*}(\gamma_2)$ minimizes $\sum_{l\in\mathcal{L}} f_l^C\cdot\tau_l(f_l^{S^*}(\gamma_2) + f_l^C)$. The proof is concluded after noticing that summing the inequalities (6.28) and (6.29) one gets

$$\begin{aligned} T(f^*(\gamma_2)) &\leq \sum_{l\in\mathcal{L}} (f_l^{S^*}(\gamma_1) + f_l^{C^*}(\gamma_1))\cdot\tau_l(F_l^*(\gamma_1)) = \\ &= \sum_{l\in\mathcal{L}} F_l^*(\gamma_1)\cdot\tau_l(F_l^*(\gamma_1)) = T(f^*(\gamma_1)). \end{aligned}$$

□

Theorem 6.2 (PoA monotonicity)**.** *Let Assumptions 6.2, 6.4 and 6.5 hold.* PoA(γ) *is non-increasing in the fleet share* γ*.*

Proof. Proposition 6.2 establishes that the equilibrium link flow $f^*(\gamma)$ is a Lipschitz continuous function of γ and Proposition 6.4 ensures that on any interval over which the support of the two vehicles classes is constant, the flow of links used by class S can only decrease and that of class C can only increase. Hence it must be that for any $\gamma_1 < \gamma_2$, $\mathcal{L}^S(\gamma_2) \subseteq \mathcal{L}^S(\gamma_1)$ and $\mathcal{L}^C(\gamma_1) \subseteq \mathcal{L}^C(\gamma_2)$. Since there are a finite number of links, there are a finite number of points in which the support of either class S or C changes. Since: i) the PoA is Lipschitz continuous, ii) it is non-increasing with γ for any interval in which the support doesn't change and iii) the support changes in a finite number of points, one can conclude that the PoA is non-increasing with γ everywhere.

□

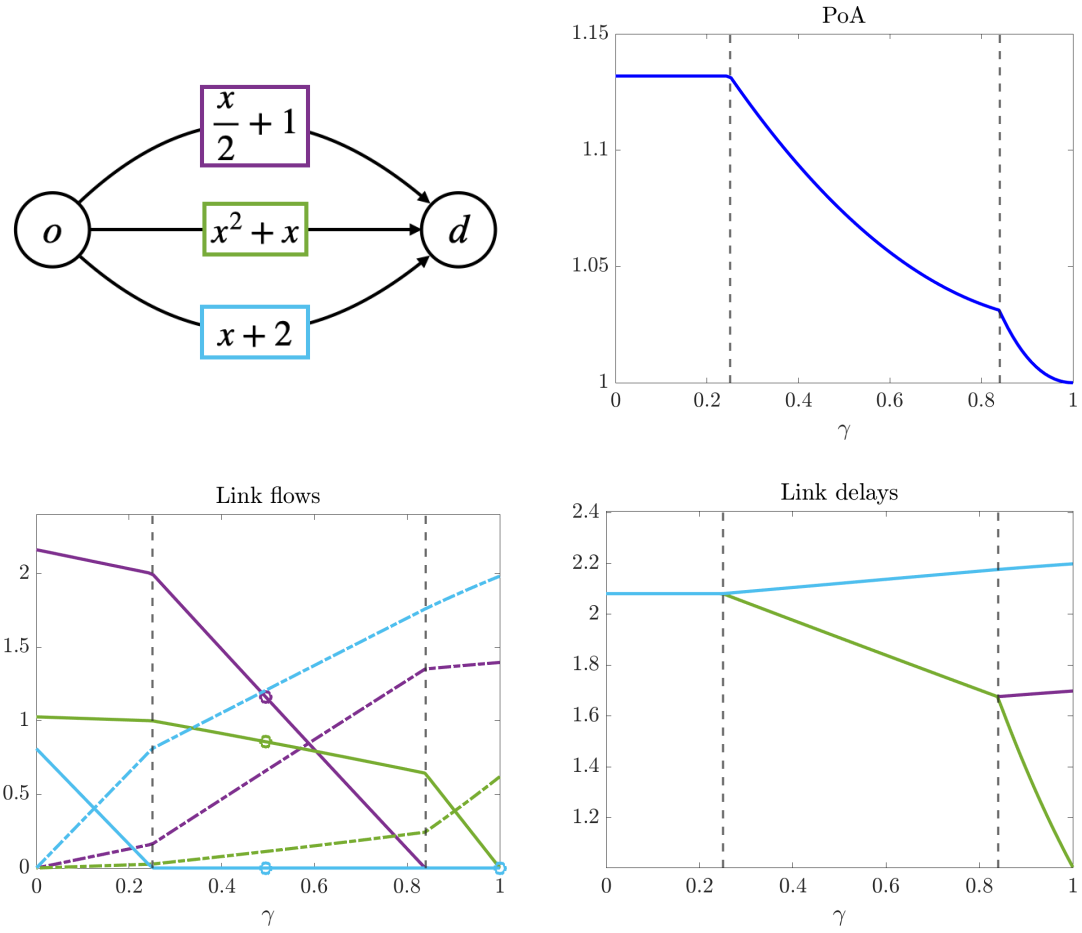

Figure 6.3: Example of parallel network. Link labels stand for the link travel times. We set $\Phi = 4$. In the bottom row, violet lines refer to the top link, green lines to the middle link and light-blue lines to the bottom link. Solid lines with circles refer to the flows of S, while dashed lines to those of class C. The vertical gray dashed lines identify changes occurring in the support of the vehicle classes as γ varies.

6.5 Numerical experiments

Below, we present two examples. The first example aims to illustrate the theoretical results presented in the preceding section. The second example, on the other hand, aims to suggest which results can be expected to persist in more general contexts and which may not.

Example 6.1. *Consider the example in Figure 6.3. The plots showcase the evolution of the* $\mathrm{PoA}(\gamma)$*, the equilibrium link flows* $f_l^{i*}(\gamma)$*,* $l = 1, 2, 3$*,* $i = S, C$*, and the link travel times* $\tau_l(\gamma)$*,* $l = 1, 2, 3$*, as functions of* γ*, for* γ *varying in* $[0, 1]$*. According to Proposition 6.3 and Theorem 6.2, the three plots demonstrate that Price of Anarchy, the flows associated with class* S *and the minimum travel time at equilibrium are non-increasing in the fleet share* γ*, while the flows associated with class* C *are non-decreasing in* γ*. Notice also that, as long as* $\gamma \leq \tilde{\gamma} \approx 0.25$*, the support of* C *is included in that of* S *and* $\mathrm{PoA}(\gamma) = \mathrm{PoA}(0)$ *for any* $\gamma \leq \tilde{\gamma}$*, consistently with*

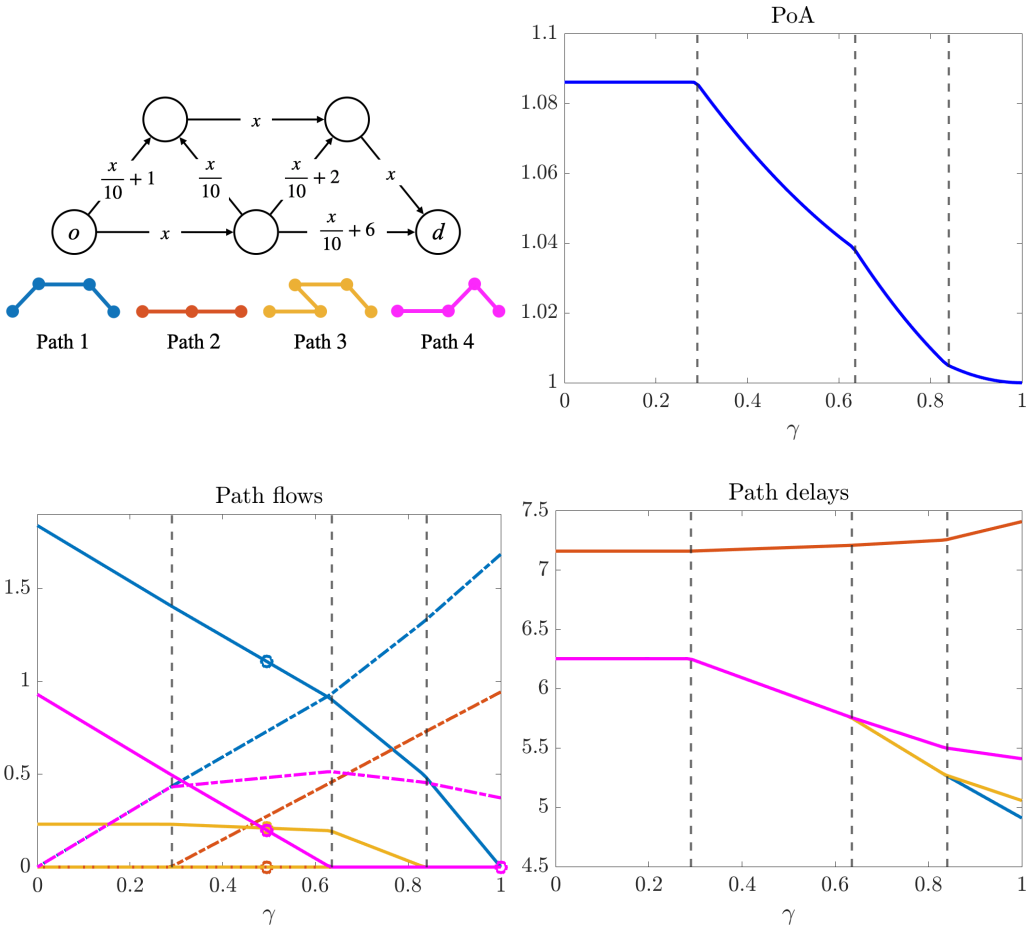

Figure 6.4: Example of network consisting of seven links and four routes. Link labels stand for the link travel times. We set $\Phi = 3$. In the bottom row, blue lines refer to Route 1, orange lines refer to Route 2, yellow lines refer to Route 3 and magenta lines to Route 4. Solid circled lines refer to the route flows associated with class S, while dashed lines to those associated with class C. Finally, the vertical grey dashed lines identify the changes occurring in the support of the two vehicle classes as α varies.

Theorem 6.1. Hence this is an example in which a minimum fleet size ($\tilde{\gamma}$) is needed for affecting the PoA.

Example 6.2. *Consider the example in Figure 6.4. The plots depict the behavior of the* PoA(γ)*, the unique equilibrium route flows* $z_p^{i*}(\gamma)$*,* $p = 1, 2, 3, 4$*,* $i = S, C$ *and the route travel times* $\tau_p(\gamma)$*,* $p = 1, \ldots, 4$*, as functions of* γ*, for* γ *varying in* $[0, 1]$*. Although not guaranteed in general, in this case uniqueness of the equilibrium route flow* $z^*(\gamma)$ *follows from the uniqueness of the equilibrium link flow* $f^*(\gamma)$*,* $\forall \gamma \in [0, 1]$*. This is due to the fact that each route of the network in Figure 6.4 possesses a link not shared with any other route. This means that the link flow of a class on that link determines the flow of the class on the corresponding route. Hence, since the equilibrium link flow is unique, so is the equilibrium route flow. Now, as in the parallel network case, the* PoA*, the equilibrium route flows associated with class S and the minimum route travel time are non-increasing with respect to* γ*. Different*

from parallel networks, we note that in this simulation the route flows associated with class C are instead not necessarily monotone (see the flow of Route 4). Whether monotonicity of the PoA *can be proven in this more general case remains an open problem.*

6.6 Concluding remarks

This chapter contributes to the understanding of the implications of the presence of a coordinated fleet of vehicles aiming to minimize the total travel time of that class on the efficiency of the entire traffic network. The results found, valid for two-terminal networks, suggest that the presence of such a fleet, when small relative to the total traffic volume, does not alter the way traffic is distributed across the network. However, for parallel networks, when the fraction of coordinated vehicles becomes significant, their coordinated behavior benefits the efficiency of the entire network.

Future work aims to extending the analysis to more general networks. Initially, the efforts should target arbitrary two-terminal networks to verify if the observed benefits hold beyond parallel networks. Further exploration should then encompass general multi-origin, multi-destination networks to comprehensively understand the impact of coordinated routing. The extension of the analysis to multi-origin multi-destination networks appears of crucial importance, considering the existence of counterexamples in the literature based on this type of networks, where it is shown that coordinated routing might have negative consequences. This fact emphasizes the need for further exploration.

However, it is important to consider the broader context. Our findings, along with most existing literature on coordinated routing, suggest positive effects. For this reason, we believe that coordinated routing has the potential to be strategically exploited to improve overall traffic efficiency. Further research that investigates the specific conditions under which coordinated routing leads to unintended consequences will be crucial for its optimal utilization.

This work also sheds light on how real-world coordinated fleets, like ride-hailing services, impact traffic efficiency. Interestingly, our findings suggest that the routing strategies adopted by these services, based on the exploitation of real-time traffic data, probably is not a significant factor in the observed increase of traffic congestion in urban areas. This increased pressure likely stems from other factors associated with these services (no passenger traveling, mode shift away from public transportation). To improve the model's ability to reflect real-world situations, future research will strive to incorporate these factors into the modeling process.

In conclusion, we emphasize that the analysis conducted in this chapter is entirely based on an instance of mixed behavior NRGs presented in Section 2.2.3, where the traffic state is described exclusively in terms of flows, and the network links are not characterized by any supply and demand mechanism that define capacity constraints. Throughout this thesis, we have highlighted the importance of designing models that account for these constraints, and Chapters 3, 4, and 5 propose models that address this need. Therefore, a key future extension of the model

proposed in this chapter is to extend the analysis of coordinated routing to networks equipped with supply and demand mechanisms, allowing for a more accurate assessment of its impact on traffic efficiency.

Conclusion

This thesis delved information-aware routing and its impact on the efficiency of traffic networks. By extending existing models and theories, we were able to provide a more comprehensive analysis of how modern technologies and services based on information-aware routing, such as navigation apps and ride-hailing services, influence traffic dynamics.

Contributions

The first contribution sheds light on the interplay between selfish routing and the capacity and volume constraints that define a traffic network. By examining non-atomic routing games on traffic networks constrained by supply and demand limits—borrowing from Daganzo's cell transmission model—we demonstrate that individual drivers, each aiming to minimize their own travel time, can inadvertently lead to suboptimal network performance. Far from simply increasing congestion, this selfish behavior causes an unexpected phenomenon: only part of the total traffic flow can enter the network, while the rest is left stranded at its origin. This critical inefficiency, one that extends beyond the commonly discussed Price of Anarchy, shows that self-interested decisions can fundamentally disrupt traffic flow.

Building on this, we turn our attention to the role of navigation apps. In our second contribution, we extend the idea of partial demand transfer to dynamic traffic flows, introducing a model that incorporates real-time routing recommendations. Here, we show that the presence of informed users, guided by these apps to follow optimal routes, can have a profound impact on traffic efficiency—particularly when traffic volume is high. When too many users follow apps' recommendations, the network is overwhelmed, resulting in patterns of partial demand transfer, where large portions of the demand go unserved. Despite the hope that navigation apps could alleviate congestion, our findings suggest that the widespread use of such tools may not always lead to the efficiency improvements we expect, aligning with earlier studies on the topic.

We make a third contribution by exploring the impact of delays in the recommendations provided to informed users, which has not been addressed in the literature before. We show that excessive delays in traffic information can significantly compromise the efficiency of the system. We also demonstrate that the system's sensitivity to these delays crucially depends on key factors such as the exogenous flow, the penetration rate of informed users, and their compliance with the recommendations. Hence, as the presence of informed users increases, the system's tolerance to delays diminishes.

Taken together, these first three contributions shine a light on possible shortcomings of information-aware routing. In fact, our research suggests that the broader adoption of navigation apps might not bring about the traffic improvements many expect. Instead, it may simply exacerbate existing inefficiencies, as users blindly follow their app's guidance without considering the collective impact on the network.

Building on these insights, we propose few recommendations to mitigate these unintended consequences. First, app developers should take into account the implications of traffic inefficiencies when designing routing algorithms, ensuring that recommendations avoid fostering suboptimal traffic patterns. Second, to minimize the impact of informational delays, navigation apps should prioritize frequent updates to route suggestions, better reflecting real-time conditions. Finally, regulators can play a pivotal role, encouraging network usage that is consistent with the original design principles of transportation planning. This could involve promoting the use of high-capacity routes for medium- to long-distance travel, while restricting smaller routes to local traffic accessing urban areas.

In stark contrast to the inefficiencies observed in the uncoordinated, selfish routing setting, our fourth and final contribution focuses on the benefits of coordinated fleets of vehicles, where information-aware routing can significantly enhance network performance. We analyze the impact of these fleets, which use real-time traffic data to act jointly and minimize their collective travel time, on overall traffic network efficiency. Our results show that when the fleet is small with respect to the total traffic volume, its effect on network efficiency is minimal. However, in parallel networks, when the fleet represents a substantial portion of traffic, the coordinated behavior of these vehicles can lead to significant improvements in overall network efficiency.

These findings highlight the potential of coordinated routing as a valuable tool in developing traffic control strategies to mitigate congestion. Moreover, they suggest that the increased presence of ride-hailing services, which often rely on such coordination, is unlikely to be a primary cause of urban congestion. Instead, other factors, such as empty trips and a shift away from public transportation, are more likely to be at the root of the problem.

Future perspectives

There are several promising directions for expanding and refining the work presented in this thesis. One primary extension is the generalization of the models to arbitrary network topologies. The models developed in this work focused on simplified network structures to maintain analytical tractability and clarity. However, real-world traffic networks are far more complex, featuring a diverse set of origin-destination pairs and a large number of intersecting routes. By generalizing the models to more complex network topologies, we can achieve a more complete and accurate characterization of the phenomena uncovered here, leading to broader applicability of the results.

In terms of non-atomic routing games with supply and demand, generalizing the model to arbitrary topologies will require formalizing it as an optimization prob-

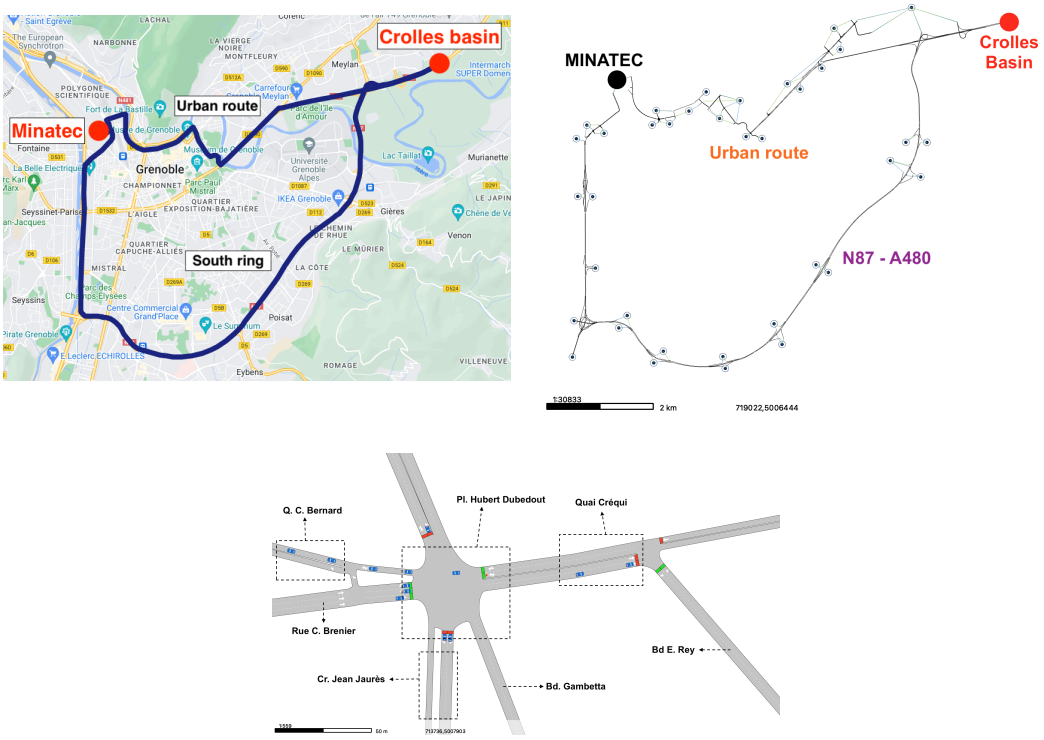

Figure 6.5: Top-left: Two possible routes connecting the Crolles basin to the Minatec area in Grenoble, France. The violet route represents Grenoble's ring road, while the orange route is a common urban itinerary. Top-right: Aimsun model corresponding to the two itineraries in the top-left map. Bottom: A frame from an Aimsun microsimulation on the model shown in the top-right picture.

lem, akin to traditional non-atomic routing games. This will involve adapting the framework to account for supply and demand constraints, taking into consideration the capacity limits of links and intersections. For dynamic models with supply and demand, the challenges are even more pronounced. As real-time information and routing recommendations are incorporated, the dynamics of traffic flow become increasingly complex. Existing conventional techniques for stability analysis—such as those used in static traffic models—are not always applicable in this dynamic setting. Therefore, novel approaches are needed to investigate the stability of these dynamic systems.

Once these more general and dynamic models are developed, a crucial avenue of investigation will be understanding how the introduction of capacity constraints alters our understanding of network routing problems. Traditional models, which often overlook or simplify capacity limitations, will benefit from incorporating more realistic constraints derived from supply and demand dynamics. These constraints can fundamentally shift how we approach traffic management, especially in systems with self-interested users. Various mechanisms, such as tolling, incentives, and information design, have been explored in the literature to address inefficiencies in such networks. It is essential to evaluate whether these existing tools remain effective when applied to networks constrained by supply and demand, or whether

they need to be adapted to the more complex reality of modern traffic systems.

Furthermore, validation of the dynamic models introduced in Chapters 4 and 5 through microscopic traffic simulations is crucial to ensure the accuracy and applicability of the theoretical results. These simulations, conducted in micro-simulation platforms like *Aimsun* [91], will incorporate realistic network structures, traffic dynamics, and user behavior models. By experimenting with diverse network configurations and traffic scenarios, we can refine and corroborate the theoretical findings, identifying potential gaps and improving the models accordingly. Though validation efforts are still underway, this process is vital for enhancing confidence in the models and ensuring they provide meaningful insights for real-world traffic systems.

Finally, the study of coordinated routing offers significant potential for further research, especially in the context of fleet-based transportation systems. Coordinated routing could be instrumental in designing traffic control strategies that optimize network efficiency, particularly in systems involving fleets of vehicles such as ride-hailing services or delivery networks. Understanding the impact of coordinated routing in general network topologies and across multiple fleets will be essential for evaluating its potential as a tool for congestion management and traffic optimization. In the longer term, these studies could lead to novel approaches to traffic control.

Appendix A

Proof of Theorems

A.1 Proof of Theorem 4.1

Theorem 4.1 is proved by exploiting the following result, based on monotonicity. Although it was originally stated in [92, Lemma 3] in terms of the *semiflow* notion, we are going to state it for differential equation systems.

Lemma A.1 (Monotonicity and stability)**.** *Consider a globally Lipschitz system* $\dot{y} = h(y)$ *with* $h : D \to \mathbb{R}^d$ *and* $D \subset \mathbb{R}^d$*. Suppose that:*

- *the system is monotone on* D*;*
- *the system admits a unique equilibrium* y^* *in* D*;*
- *every trajectory of the system has a compact closure.*
- *every neighborhood* N *of every point* $x \in D$ *contains a compact and* order convex neighborhood *of* x*,* $\exists p, q \in N$ *such that* $\{u \in D : p_i \leq u_i \leq q_i, \forall i\} \subseteq N$*.*

Then, the equilibrium y^* *is globally asymptotically stable.*

In order to apply Lemma A.1 to (4.4), we need to verify the four conditions above. The third point is straightforward, after observing that Ω is compact. In the remainder of this section, we shall prove the three remaining points.

A.1.1 Monotonicity

We consider the extension of system (4.4) to the whole positive orthant $\mathbb{R}^2_{\geq 0}$ and we shall prove the stronger property that the extension of (4.4) is a monotone system. To this end, we first show that it satisfies the so called *K-condition*. System (4.4) is said to satisfy the K-condition if, given $a, b \in \mathbb{R}^2_{\geq 0}$ such that $a \leq b$, where the inequality is meant component-wise, and $a_i = b_i$, then $\Sigma_i(b) \geq \Sigma_i(a)$. To verify the K-condition, notice that also for the extended systems we can identify the same system modes $\Sigma^{\text{M}_1\text{-M}_2}$ according to (4.4), with the only difference that the corresponding regions $\overline{\Omega}^{\text{M}_1\text{-M}_2}$ can be unbounded and their union covers $\mathbb{R}^2_{\geq 0}$

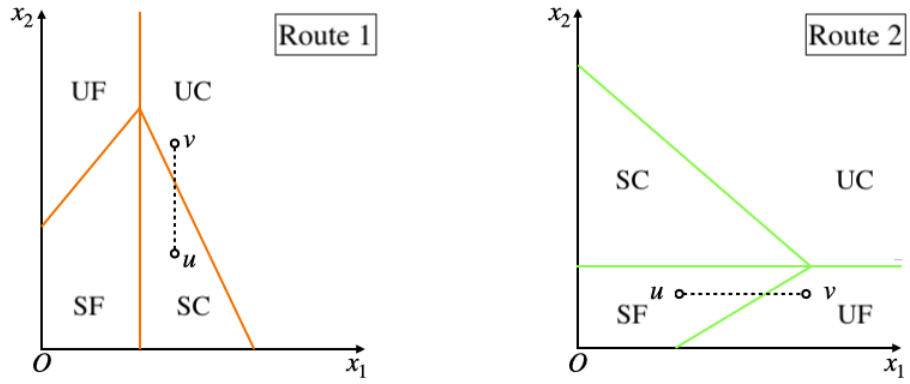

Figure A.1: The K-condition can be easily verified, given a pair of points u and v, considering only one of the two routes.

(the corresponding regions for the original system being their restrictions to Ω.) In particular, Assumption 4.2 guarantess that the jacobian matrices $J^{\text{M}_1\text{-M}_2}$ are Metzler. Then, K-condition holds inside every region $\overline{\Omega}^{\text{M}_1\text{-M}_2}$. The K-condition holds across different regions. To see this, consider $p, q \in \mathbb{R}^2_{\geq 0}$ such that $p_1 = q_1,\ p_2 < q_2$ and p and q belong to different mode regions. We need to prove that $\Sigma_1(p) \leq \Sigma_1(q)$. From Figure A.1, one can see that we only need to compare points belonging to regions such that $p \in \overline{\Omega}^{\text{SF-M}_2}$, $q \in \overline{\Omega}^{\text{UF-M}_2}$, or $p \in \overline{\Omega}^{\text{SC-M}_2}$, $q \in \overline{\Omega}^{\text{UC-M}_2}$.

- $p \in \overline{R}^{\text{SF-M}_2}$, $q \in \overline{\Omega}^{\text{UF-M}_2}$:

$$\Sigma_1(p) = \frac{\Phi R_1(\tau) - v_1 p_1}{L_1} \leq \frac{\overline{f}_1 - v_1 q_1}{L_1} = \Sigma_1(q);$$

- $p \in \overline{\Omega}^{\text{SC-M}_2}$, $q \in \overline{\Omega}^{\text{UC-M}_2}$:

$$\Sigma_1(p) = \frac{\Phi R_1(\tau) - \overline{f}_1}{L_1} \leq \frac{w_1(\overline{x}_1 - p_1) - \overline{f}_1}{L_1} = \Sigma_1(q);$$

The inequalities above follow from the route mode's definition and by symmetry they also apply to Route 2, i.e., to the case in which $p, q \in \mathbb{R}^2_{\geq 0}$ such that $p_1 < q_1,\ p_2 = q_2$, then we get that (4.4) satisfies to the K-condition.

We now prove that system (4.4) is monotone. Let us take $u, w \in \mathbb{R}^2_{\geq 0}$ such that $u \leq w$. Let us take $\delta > 0$ and define $\Sigma_\delta(x) := \Sigma(x) + \delta p$, where $p > \mathbf{0}$. Let $y_\delta(t)$ be the solution to the Cauchy problem

$$\begin{cases} \dot{y}_\delta(t) = \Sigma_\delta(y_\delta(t)) \\ y_\delta(0) = w + \delta p, \end{cases}$$

and let $x(t, u)$ be the solution to (4.4) with initial condition $x(0) = u$. Notice that since $\Sigma_\delta(x)$ is Lipschitz, existence and uniqueness of $y_\delta(t)$ are guaranteed. Clearly, $x(0) < y_\delta(0)$. Our claim is that $x(t, u) < y_\delta(t),\ \forall t \geq 0$. By contradiction, suppose that there exist $\tau > 0$ and i such that

$$x(\tau, u) \leq y_\delta(\tau), \quad x_i(\tau, u) = (y_\delta(\tau))_i,$$

for some l, and $x(s,u) < y_\delta(s)$, $\forall s < \tau$. Then, it must hold that $\dot{x}_l(\tau,u) \geq (\dot{y}_\delta(\tau))_l$, which is equivalent to

$$\Sigma_l(x(\tau,u)) \geq (\Sigma_\delta)_l(y_\delta(\tau)) = \Sigma_l(y_\delta(\tau)) + \delta p_l.$$

This is absurd, since (4.4) satisfies to the K-condition, which implies that

$$\Sigma_l(x(\tau,u)) \leq (\Sigma_\delta)_l(y_\delta(\tau)) = \Sigma_l(y_\delta(\tau)) + \delta p_l.$$

Hence, $x(t,u) < y_\delta(t), \forall t \geq 0$, is proved. Finally, observe that, since Σ and Σ_δ are Lipschitz continuous, by letting $\delta \to 0$, the continuous dependence from initial conditions ensures that $x(t,u) \leq x(t,w)$, $\forall t \geq 0$. We conclude by noting that the monotonicity property that we proved also holds true also for (4.4) restricted to Ω, since Ω is an invariant set.

A.1.2 Uniqueness of equilibrium

Since P is globally attractive, no equilibrium is contained in Q. This implies that the equilibria of (4.4) are contained in P. Hence, we can limit ourselves to only consider the system modes contained in P:

$$\Sigma^{\text{SF-SF}} = \begin{cases} \dot{x}_1 = \Phi R_1(\tau(x)) - v_1 x_1 \\ \dot{x}_2 = \Phi R_2(\tau(x)) - v_2 x_2 \end{cases} \tag{A.1}$$

$$\Sigma^{\text{UF-SF}} = \begin{cases} \dot{x}_1 = \overline{f}_1 - v_1 x_1 \\ \dot{x}_2 = \Phi R_2(\tau(x)) - v_2 x_2 \end{cases} \tag{A.2}$$

$$\Sigma^{\text{SF-UF}} = \begin{cases} \dot{x}_1 = \Phi R_1(\tau(x)) - v_1 x_1 \\ \dot{x}_2 = \overline{f}_2 - v_2 x_2 \end{cases} \tag{A.3}$$

The equilibria of system (4.4) must coincide with the set of *active* equilibria of the sub-systems above, where by active we mean that, if $x^{\Sigma^{\text{M}_1\text{-M}_2}}$ is an equilibrium of $\Sigma^{\text{M}_1\text{-M}_2}$, $x^{\Sigma^{\text{M}_1\text{-M}_2}} \in \overline{\Omega}^{\text{M}_1\text{-M}_2}$. We prove that the equilibrium of (4.4) is unique by showing that each of the above sub-systems admits a unique equilibrium and only one among them is active.

Consider the following sets:

$$\begin{aligned} \psi_l &= \{x \in \Omega : \Phi R_l(\tau(x)) - v_l x_l = 0\}, \quad l = 1,2, \\ \epsilon_l &= \{x \in \Omega : \Phi R_l(\tau(x)) - \overline{f}_l = 0\}, \quad l = 1,2. \end{aligned} \tag{A.4}$$

Sets ψ_1, ψ_2 are always nonempty. In fact, consider $x \in \Omega$ and notice that $x \in \psi_l$ if and only if $x_l = \Phi R_l(\tau(x)))/v_l$. Since $R_l(\tau(x))$ is strictly decreasing in x_l by Assumption 4.2, one can observe that for each $x_k \in [0, \overline{x}_k]$, there exists a unique $x_l \in [0, \overline{x}_l]$ satisfying to the above equality. The fact that $x_l \in [0, \overline{x}_l]$ is guaranteed by Assumption 4.3. This proves that $\psi_l \neq \emptyset$, $l = 1,2$. As for sets ϵ_1, ϵ_2, from (4.6) we have

$$r_l(x) = \frac{1}{\alpha}\left(\frac{\overline{f}_l}{\Phi} - (1-\alpha) r_l^0\right), \quad l = 1,2.$$

Since $0 \leq r_l(x) \leq 1, \forall x \in \Omega,\ l = 1, 2$, in order to have $\epsilon_l \neq \emptyset$, it must be that

$$(1-\alpha)\Phi r_l^0 \leq \overline{f}_l \leq (1-\alpha)\Phi r_l^0 + \alpha\Phi.$$

Now, since routing ratios are assumed to be C^1 in Ω and strictly monotone by Assumption 4.2, when the sets above are nonempty, the implicit function theorem guarantees the existence of functions between the variables x_1 and x_2 making explicit each of the implicit relations in (A.4). Let the following be the functions associated to relations (A.4):

$$\begin{aligned} x_2 &:= \tilde{\psi}_1(x_1), \quad x_2 := \tilde{\psi}_2(x_1), \\ x_2 &:= \tilde{\epsilon}_1(x_1), \quad x_2 := \tilde{\epsilon}_2(x_1). \end{aligned} \tag{A.5}$$

Then, sets in (A.4) are the graphs of functions (A.5). The implicit function theorem and Assumption 4.2 also ensure that they are strictly increasing in x_1:

$$\tilde{\psi}_1'(x_1) = -\frac{\frac{\partial}{\partial x_1}(\Phi R_1(\tau(x) - v_1 x_1)}{\frac{\partial}{\partial x_2}(\Phi R_1(\tau(x)) - v_1 x_1)} = \frac{v_1 - \frac{\partial R_1(\tau(x))}{\partial x_1}}{\frac{\partial R_1(\tau(x))}{\partial x_2}} > 0,$$

$$\tilde{\psi}_2'(x_1) = -\frac{\frac{\partial}{\partial x_1}(\Phi R_2(\tau(x) - v_2 x_2)}{\frac{\partial}{\partial x_2}(\Phi R_2(\tau(x)) - v_2 x_2)} = \frac{\frac{\partial R_2(\tau(x))}{\partial x_1}}{v_2 - \frac{\partial R_2(\tau(x))}{\partial x_2}} > 0,$$

and

$$\tilde{\epsilon}_l'(x_l) = -\frac{\frac{\partial}{\partial x_1}(\Phi R_l(\tau(x)) - \overline{f}_l)}{\frac{\partial}{\partial x_2}(\Phi R_l(\tau(x)) - \overline{f}_l)} = -\frac{\frac{\partial R_l(\tau(x))}{\partial x_1}}{\frac{\partial R_l(\tau(x))}{\partial x_2}} > 0,$$

for $l = 1, 2$.

We now show that each sub-system in P admits a unique equilibrium point. First of all, observe that the image of $\tilde{\psi}_1(x_1)$ is the whole $[0, \overline{x}_2]$, since, for each $x_2 \in [0, \overline{x}_2]$, the map $x_1 = \Phi R_1(\tau_1(x_1), \tau_2(x_2))/v_1$ always admits a unique fixed point in $[0, \overline{x}_1]$. Analogously, $\tilde{\psi}_2(x_1)$ is defined over the whole interval $[0, \overline{x}_1]$, since, for each $x_1 \in [0, \overline{x}_1]$, the map $x_2 = \Phi R_1(\tau_1(x_1), \tau_2(x_2))/v_2$ always admits a unique fixed point $[0, \overline{x}_2]$. This, combined with the fact that both functions are strictly increasing in x_1, implies that $\tilde{\psi}_1(x_1)$ and $\tilde{\psi}_2(x_1)$ intersect lines $\{x_2 = x_2^c\}$ and$\{x_1 = x_1^c\}$ in one and only one point inside Ω, respectively. These two points correspond to the unique equilibria of systems $\Sigma^{\text{SF-UF}}$, $\Sigma^{\text{UF-SF}}$, namely $x^{\Sigma^{\text{SF-UF}}}$, $x^{\Sigma^{\text{UF-SF}}}$.

Now, from above, it follows that the functions $\tilde{\psi}_1(x_1)$ and $\tilde{\psi}_2(x_1)$ admit two points of the form $(x_1, 0)$, $(0, x_2)$ satisfying to their equations, respectively. Moreover, we have that

$$\begin{aligned} \tilde{\psi}_2'(x_1) - \tilde{\psi}_1'(x_1) &= -\frac{v_1 v_2 + \sum_{l \neq k} v_l \frac{\partial R_k(x)}{\partial x_l}}{\frac{\partial R_1(\tau(x))}{\partial x_2}\left(v_2 + \frac{\partial R_1(\tau(x))}{\partial x_2}\right)} \\ &\leq -\frac{v_1 v_2}{K(v_2 + K)} < 0. \end{aligned}$$

The combination of these two facts implies that there exist a unique point in Ω such that $\tilde{\psi}_1(x_1) = \tilde{\psi}_2(x_1)$, which corresponds to the unique equilibrium point of sub-system $\Sigma^{\text{SF-SF}}$, namely $x^{\Sigma^{\text{SF-SF}}}$.

Finally, we prove one and only one among $x^{\Sigma^{\text{SF-SF}}}$, $x^{\Sigma^{\text{UF-SF}}}$, $x^{\Sigma^{\text{SF-UF}}}$ is active. Before proceeding, let us point out some facts. First of all, observe that when ϵ_1 and ϵ_2 are non empty, the graph of the their functions $\tilde{\epsilon}_1(x_1)$, $\tilde{\epsilon}_2(x_1)$ split the state space into two separate regions:

$$E_l^+ := \{x \in \Omega : \ \Phi R_l(\tau(x)) - \overline{f}_l > 0\},$$
$$E_l^- := \{x \in \Omega : \ \Phi R_l(\tau(x)) - \overline{f}_l < 0\},$$

$l = 1, 2$. Regions E_1^+ and E_2^+ are those where there is unsatisfied demand on route 1 and 2, respectively. One can see that E_1^+ and E_1^- stand above and below the graph of $\tilde{\epsilon}_1(x_1)$, respectively, whereas E_2^+ and E_2^- stand below and above the graph of $\tilde{\epsilon}_2(x_2)$, respectively. Observe that the following identities hold:

$$\Omega^{\text{SF-SF}} = E_1^- \cap E_2^- \cap P, \quad \Omega^{\text{UF-SF}} = E_1^+ \cap E_2^- \cap P,$$

$$\Omega^{\text{SF-UF}} = E_1^- \cap E_2^+ \cap P.$$

Assumptions 4.1 and 4.2 ensure that $\tilde{\epsilon}_1(x_1) > \tilde{\epsilon}_2(x_1), \ \forall x \in \Omega$ when both functions are defined, and $\tilde{\epsilon}_1(x_1) \geq \tilde{\psi}_1(x_1), \ x_1 \leq x_1^c$, $\tilde{\epsilon}_2(x_1) \leq \tilde{\psi}_2(x_1), \ x_2 \leq x_2^c$, where the equality holds if and only if $x_1 = x_1^c$ and $x_2 = x_2^c$, respectively. Also, recall that $\tilde{\psi}_2(x_1) \geq \tilde{\psi}_1(x_1), \ x_1 \leq \left(x^{\text{SF-SF}}\right)_1$, where the equality holds if and only if $x_1 = \left(x^{\text{SF-SF}}\right)_1$, and $\tilde{\psi}_2(x_1) < \tilde{\psi}_1(x_1), \ x_1 > \left(x^{\text{SF-SF}}\right)_1$. We are going to distinguish two different cases.

Case 1: $x^{\Sigma^{\text{SF-SF}}} \in P$. In order for $x^{\Sigma^{\text{SF-SF}}}$ to be active, it must hold that $x^{\Sigma^{\text{SF-SF}}} \in \Omega^{\text{SF-SF}}$. This leads to two distinct sub-cases.

- **Sub-case 1:** ϵ_1, ϵ_2 are both empty.
  From Assumption 4.1, ϵ_1, ϵ_2 are both empty if and only $\overline{f}_l \geq \alpha\Phi + (1-\alpha)\Phi r_l^0$, $l = 1, 2$. Hence, unsatisfied demand cannot arise on either route inside Ω. Thus, $P \equiv \overline{\Omega}^{\text{SF-SF}}$. The latter implies that neither $x^{\Sigma^{\text{UF-SF}}}$ nor $x^{\Sigma^{\text{SF-UF}}}$ can be active, hence $x^* \equiv x^{\Sigma^{\text{SF-SF}}}$.

- **Sub-case 2:** at least one among ϵ_1, ϵ_2 is non empty.
  In this case, region P undergoes a partition. Nevertheless, the fact that $\tilde{\epsilon}_1(x_1) > \tilde{\epsilon}_2(x_1)$, $\tilde{\epsilon}_1(x_1) \geq \tilde{\psi}_1(x_1)$, $\tilde{\epsilon}_2(x_1) \leq \tilde{\psi}_2(x_1)$ ensures that $\Omega^{\text{SF-SF}}$ is non-empty and that $x^{\Sigma^{\text{SF-SF}}} \in \Omega^{\text{SF-SF}}$. Thus, $x^* \equiv x^{\Sigma^{\text{SF-SF}}}$. As for $x^{\Sigma^{\text{UF-SF}}}$, it cannot be active, since the intersection between the graph of $\tilde{\psi}_2(x_1)$ and the line $\{x_1 = x_1^c\}$ occurs below the graph of $\tilde{\epsilon}_1(x_1)$. The latter follows from $\tilde{\psi}_2(x_1) \leq \tilde{\psi}_1(x_1), \ x_1 > \left(x^{\text{SF-SF}}\right)_1$. Analogously, $x^{\Sigma^{\text{SF-UF}}}$ cannot be active.

Case 2: $x^{\Sigma^{\text{SF-SF}}} \notin P$. First of all, notice that if $x^{\Sigma^{\text{SF-SF}}} \notin P$, then it is not active and either $x^{\Sigma^{\text{SF-SF}}} \in \{x_1 > x_1^c, \ x_2 \leq x_2^c\}$ or $x^{\Sigma^{\text{SF-SF}}} \in \{x_1 \leq x_1^c, \ x_2 > x_2^c\}$. Indeed, if $x^{\Sigma^{\text{SF-SF}}} \in \{x_1 > x_1^c, \ x_2 > x_2^c\}$, then $\Phi R_l(\tau(x^{\Sigma^{\text{SF-SF}}})) = v_l(x^{\Sigma^{\text{SF-SF}}})_l > \overline{f}_l, \ l = 1, 2$, contradicting Assumption 4.1. Suppose then that $x^{\Sigma^{\text{SF-SF}}} \in \{x_1 > x_1^c, \ x_2 \leq x_2^c\}$. Then, it must be that $\tilde{\psi}_2(x_1^c) > \tilde{\psi}_1(x_1^c)$ and $\tilde{\psi}_2(x_1^c) \leq x_2^c$. Moreover, $\tilde{\epsilon}_1(x_1^c) = \tilde{\psi}_1(x_1^c)$. Thus, $x^{\Sigma^{\text{UF-SF}}} \in \Omega^{\text{UF-SF}}$. Finally, observe that, again, since $\tilde{\psi}_2(x_1^c) > \tilde{\psi}_1(x_1^c)$ and $\tilde{\psi}_2(x_1^c) \leq x_2^c$, $\tilde{\psi}_1(x_1)$ cannot intersect the line $\{x_2 = x_2^c\}$ inside P, i.e, $x^{\Sigma^{\text{SF-UF}}}$ is

not active. One can repeat the same process in the case in which $x^{\Sigma^{\text{SF-SF}}} \in \{x_1 \leq x_1^c,\ x_2 > x_2^c\}$. In this case, $x^{\Sigma^{\text{SF-UF}}}$ is the only active equilibrium point.

To conclude, it might be that $x^{\Sigma^{\text{SF-SF}}}$ coincides with $x^{\Sigma^{\text{UF-SF}}}$ or $x^{\Sigma^{\text{SF-UF}}}$. One can verify that this only happens when $x^{\Sigma^{\text{SF-SF}}} \in \epsilon_1$ and $x^{\Sigma^{\text{SF-SF}}} \in \epsilon_2$, respectively. In this case, the two coinciding equilibria represent the unique equilibrium $\overline{x}$ of the system .

A.1.3 Order convex neighborhoods

Take $N = \{x \in \Omega : |x - \overline{x}| < r\} = \{x \in \mathbb{R}^2 : |x - \overline{x}| < r\} \cap \Omega$, $r > 0$. By taking $p, q \in N$ such that $x \in [p, q]$, it is straightforward that $[p, q] \subset N$.

A.2 Proof of Lemma 5.1

First of all, observe that (5.16) ensures that δ_{CF} and δ_{FC} are well-defined. Then, one can verify that the first condition of Assumption 5.2 is equivalent to $\delta_{FC} < 0 < \delta_{CF}$. Once Φ is fixed, one can define a family of functions $\{x^{\alpha_j,\eta_k}\}, \alpha_j > \max\{\underline{\alpha}_1, \underline{\alpha}_2\},\ \eta_k > 0$ and observe that all functions in this family attain the same value ψ at $\delta = 0$, where

$$\psi := \frac{\Phi}{L}\left(\frac{a_2}{\overline{x}_2} - \left(\frac{a_1}{\overline{x}_1} + \frac{a_2}{\overline{x}_2}\right) r_1^0\right).$$

It holds that $\psi > 0$ when the first inequality in (5.20) holds, whereas $\psi < 0$ when the second does. As pointed out in the proof of Proposition 5.2, the equilibrium point of (5.8) satisfies to $\delta^* = Lx(\delta^*)/v$, i.e., it is the value of δ at which $Lx(\delta^*)/v$ and the identity line intersect. Define $\delta^*_{\alpha_j,\eta_k}$ as the equilibrium point associated to the function x^{α_j,η_k}. Suppose that (5.20) holds, so that $\psi > 0$ and, as one can verify, $\beta_{CF} := x(\delta_{CF}) < 0$. Since neither β_{CF} nor ψ depend on α and η, it follows that $0 < \delta^*_{\alpha_j,\eta_k} < \delta_{CF}$, for all $\alpha_j \in (\max\{\underline{\alpha}_1, \underline{\alpha}_2\}, 1],\ \eta_k > 0$. The proof is complete after observing that we can apply the same process when (5.20) holds, so that $\psi < 0$ and $\beta_{FC} := x(\delta_{CF}) > 0$.

A.3 Proof of Lemma 5.2

The second inequality is trivial, since K is the Lipschitz constant of $g(\delta)$. For the first inequality, if we define $U_1(\delta) := R_1(\delta) - (1 - \alpha)r_1^0$, then

$$g'(\delta) = -\frac{\alpha\Phi}{\eta L}\left(\frac{a_1}{\overline{x}_1} + \frac{a_2}{\overline{x}_2}\right) U_1(1 - U_1), \qquad g''(\delta) = \frac{x'(1 - 2U_1)}{\eta}.$$

We see that $|g'(\delta)|$ increases for $U_1(\delta) \in (0, 1/2)$ and decreases for $U_1(\delta) \in (1/2, 1)$, i.e., $|g'(\delta)|$ increases for $\delta < U_1^{-1}(1/2) = \eta \log(r_2^0/r_1^0)$ and decreases for $\delta > U_1^{-1}(1/2)$, where U_1^{-1} is the inverse of $U_1(\delta)$. From (5.19), $|g'(\delta^*)|$ is greater than at least one between $|g'(\delta_{CF})|$ and $|g'(\delta_{FC})|$.

Bibliography

[1] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*. New York, NY, USA: Cambridge University Press, 2007.

[2] S. C. Dafermos, "The traffic assignment problem for multiclass-user transportation networks," *Transportation Science*, vol. 6, no. 1, pp. 73–87, 1972.

[3] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[4] ——, "The cell transmission model, part II: Network traffic," *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.

[5] M. Carey and M. Bowers, "A review of properties of flow–density functions," *Transport Reviews*, vol. 32, no. 1, pp. 49–73, 2012.

[6] T. C. P. Cabannes, "The impact of information-aware routing on road traffic. from case studies to game-theoretical analysis and simulations," Ph.D. dissertation, UC Berkeley, 2022.

[7] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Informational Braess' paradox: The effect of information on traffic congestion," *Operations Research*, vol. 66, no. 4, pp. 893–917, 2018.

[8] P. T. Harker, "Multiple equilibrium behaviors on networks," *Transportation Science*, vol. 22, no. 1, pp. 39–46, 1988.

[9] F. Facchinei and J.-s. Pang, *Finite-dimensional variational inequalities and complementarity problems*. New York, NY, USA: Springer New York, 2003.

[10] G. Scutari, D. P. Palomar, F. Facchinei, and J.-s. Pang, "Convex optimization, game theory, and variational inequality theory," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 35–49, 2010.

[11] Statista, "Most popular navigation apps in the u.s. 2023, by downloads," last accessed 15 April 2024. [Online]. Available: https://www.statista.com/statistics/865413/most-popular-us-mapping-apps-ranked-by-audience/

[12] ——, "Navigation - worldwide," last accessed 15 April 2024. [Online]. Available: https://www.statista.com/outlook/amo/app/navigation/worldwide

[13] ——, "Ride hailing - worldwide," last accessed 15 April 2024. [Online]. Available: https://www.statista.com/outlook/mmo/shared-mobility/ride-hailing/worldwide

[14] J. Macfarlane, "When apps rule the road: The proliferation of navigation apps is causing traffic chaos. it's time to restore order," *IEEE Spectrum*, vol. 56, no. 10, pp. 22–27, 2019.

[15] J. Rogers, "La traffic is getting worse and people are blaming the shortcut app Waze," *Business Insider*, December 2014.

[16] S. Henry, "Comment Waze a transformé leur village en enfer:"on a compté jusqu'à 14000 véhicule par jour!"," *Le monde*, April 2024.

[17] S. Raphelson, "New jersey town restricts streets from commuters to stop Waze traffic nightmare," *NRP*, May 2018.

[18] "Lombard study: Managing access to the "crooked street"," 2017.

[19] F. F. P. Assemblée Nationale, "Loi n. 2022-1119 du 3 août 2022 relatif aux services numériques d'assistance aux déplacements."

[20] S. Agarwal, D. Mani, and R. Telang, "The impact of ride-hailing services on congestion: Evidence from indian cities," *Manufacturing & Service Operations Management*, vol. 25, no. 3, pp. 862–883, 2023.

[21] M. Behroozi, *Understanding the Impact of Ridesharing Services on Traffic Congestion*. Cambridge University Press, 2023, p. 119–145.

[22] "TNCs & Congestion," 2018.

[23] R. R. Clewlow, "Disruptive transportation: The adoption, utilization, and impacts of ride hailing in the united states," *Transfer Magazine*, 2019.

[24] J. G. Wardrop, "Some theoretical aspects of road traffic research," *Proceedings of the Institute of Civil Engineers, Part II*, vol. 1, no. 3, pp. 325–378, 1952.

[25] 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)*, 2016, pp. 595–601.

[26] M. Wu, S. Amin, and A. E. Ozdaglar, "Value of information in bayesian routing games," *Operations Research*, vol. 69, no. 1, pp. 148–163, 2021.

[27] W. H. Sandholm, *Population games and evolutionary dynamics*. MIT Press, 2010.

[28] L. Cianfanelli, G. Como, and T. Toso, "Stability and bifurcations in transportation networks with heterogeneous users," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 6371–6376.

[29] G. Como, "On resilient control of dynamical flow networks," *Annual Reviews in Control*, vol. 43, pp. 80–90, 2017.

[30] J. A. Jacquez and C. P. Simon, "Qualitative theory of compartmental systems," *SIAM Review*, vol. 35, no. 1, pp. 43–79, 1993.

[31] G. G. Walter and M. Contreras, *Compartmental modeling with networks*. Springer Science & Business Media, 1999.

[32] G. Como and R. Maggistro, "Distributed dynamic pricing of multiscale transportation networks," *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 1625–1638, 2022.

[33] G. Bianchin and F. Pasqualetti, "Navigation systems may deteriorate stability in traffic networks," *IEEE Open Journal of Control Systems*, pp. 1–15, 2024.

[34] R. M. Colombo and H. Holden, "On the Braess paradox with nonlinear dynamics and control theory," *Journal of Optimization Theory and Applications*, vol. 168, no. 1, pp. 216–230, 2016.

[35] A. Festa and P. Goatin, "Modeling the impact of on-line navigation devices in traffic flows," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 323–328.

[36] A. Festa, P. Goatin, and F. Vicini, "Navigation system-based routing strategies in traffic flows on networks," *Journal of Optimization Theory and Applications*, vol. 198, pp. 930–957, 2023.

[37] E. Cristiani and F. S. Priuli, "A destination-preserving model for simulating wardrop equilibria in traffic flow on networks," *Networks and Heterogeneous Media*, vol. 10, no. 4, pp. 857–876, 2015.

[38] T. Boulogne, E. Altman, H. Kameda, and O. Pourtallier, "Mixed equilibrium (ME) for multiclass routing games," *IEEE Transactions on Automatic Control*, vol. 47, no. 6, pp. 903–916, 2002.

[39] H. Yang, X. Zhang, and Q. Meng, "Stackelberg games and multiple equilibrium behaviors on networks," *Transportation Research Part B: Methodological*, vol. 41, no. 8, pp. 841–861, 2007.

[40] A. Houshmand, S. Wollenstein-Betech, and C. G. Cassandras, "The penetration rate effect of connected and automated vehicles in mixed traffic routing," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 1755–1760.

[41] M. Battifarano and S. Qian, "The impact of optimized fleets in transportation networks," *Transportation Science*, vol. 57, no. 4, pp. 1047–1068, 2023.

[42] G. D. Erhardt, S. Roy, D. Cooper, B. Sana, M. Chen, and J. Castiglione, "Do transportation network companies decrease or increase congestion?" *Science advances*, vol. 5, no. 5, 2019.

[43] J. W. Ward, J. J. Michalek, I. L. Azevedo, C. Samaras, and P. Ferreira, "Effects of on-demand ridesourcing on vehicle ownership, fuel consumption, vehicle miles traveled, and emissions per capita in u.s. states," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 289–301, 2019.

[44] S. Wollenstein-Betech, A. Houshmand, M. Salazar, M. Pavone, C. G. Cassandras, and I. C. Paschalidis, "Congestion-aware routing and rebalancing of autonomous mobility-on-demand systems in mixed traffic," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–7.

[45] S. Wollenstein-Betech, M. Salazar, A. Houshmand, M. Pavone, I. C. Paschalidis, and C. G. Cassandras, "Routing and rebalancing intermodal autonomous mobility-on-demand systems in mixed traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 12 263–12 275, 2022.

[46] M. J. Beckmann, C. B. McGuire, and C. Winsten, *Studies in the economics of transportation*. Yale University Press, 1956.

[47] A. Ozdaglar and I. Menache, *Network games: Theory, Models and Dynamics*. Williston, VT: Morgan & Claypool, 2011.

[48] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *STACS 99*, C. Meinel and S. Tison, Eds. Springer Berlin Heidelberg, 1999, pp. 404–413.

[49] S. C. Dafermos, "The traffic assignment problem for multiclass-user transportation networks," *Transportation Science*, vol. 14, no. 1, pp. 42–54, 1972.

[50] R. L. Tobin and T. L. Friesz, "Sensitivity analysis for equilibrium network flow," *Transportation Science*, vol. 22, no. 4, pp. 242–250, 1988.

[51] F. Farokhi, W. Krichene, A. M. Bayen, and K. H. Johansson, "A heterogeneous routing game," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2013, pp. 448–455.

[52] G. Nilsson, P. Grover, and U. Kalabić, "Assignment and control of two-tiered vehicle traffic," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 1023–1028.

[53] D. Bergemann and S. Morris, "Information design: A unified perspective," *Journal of Economic Literature*, vol. 57, no. 1, p. 44–95, 2019.

[54] T. L. Friesz, D. Bernstein, T. E. Smith, R. L. Tobin, and B. W. Wie, "A variational inequality formulation of the dynamic network user equilibrium problem," *Operations Research*, vol. 41, no. 1, pp. 179–191, 1993.

[55] B. Ran and D. E. Boyce, *Modeling Dynamic Transportation Networks*. Heidelberg, Germany: Springer Berlin, 1996.

[56] M. Garavello and B. Piccoli, *Traffic flow on networks*. American Institute of Mathematical Sciences, 2006.

[57] P. Kachroo and K. M. Özbay, *Feedback Control Theory for Dynamic Traffic Assignment*. Springer International Publishing, 2018.

[58] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.

[59] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.

[60] J. Lebacque, "The godunov scheme and what it means for first order traffic flow models," *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, p. 647–677, 1996.

[61] A. Aw and M. Rascle, "Resurrection of "second order" models of traffic flow," *SIAM Journal on Applied Mathematics*, vol. 60, no. 3, pp. 916–938, 2000.

[62] W. Krichene, J. D. Reilly, S. Amin, and A. M. Bayen, "Stackelberg routing on parallel networks with horizontal queues," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 714–727, 2014.

[63] P. Elias, A. Feinstein, and C. Shannon, "A note on the maximum flow through a network," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 117–119, 1956.

[64] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, p. 399–404, 1956.

[65] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, "Robust distributed routing in dynamical networks—part I: Locally responsive policies and weak resilience," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 317–332, 2013.

[66] ——, "Robust distributed routing in dynamical networks – Part II: Strong resilience, equilibrium selection and cascaded failures," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 333–348, 2013.

[67] K. Savla, G. Como, and M. A. Dahleh, "Robust network routing under cascading failures," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 1, pp. 53–66, 2014.

[68] G. Como, E. Lovisari, and K. Savla, "Throughput optimality and overload behavior of dynamical flow networks under monotone distributed routing," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 1, pp. 57–67, 2015.

[69] S. Coogan and M. Arcak, "A compartmental model for traffic networks and its dynamical behavior," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2698–2703, 2015.

[70] A. Y. Yazıcıoğlu, M. Roozbehani, and M. A. Dahleh, "Resilient control of transportation networks by using variable speed limits," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 4, pp. 2011–2022, 2018.

[71] E. Lovisari, G. Como, and K. Savla, "Stability of monotone dynamical flow networks," in *2014 53rd IEEE Conference on Decision and Control (CDC)*, 2014, pp. 2384–2389.

[72] T. Roughgarden and E. Tardos, "How bad is selfish routing?" *J. ACM*, vol. 49, no. 2, p. 236–259, 2002.

[73] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, "Stability analysis of transportation networks with multiscale driver decisions," *SIAM Journal on Control and Optimization*, vol. 51, no. 1, pp. 230–252, 2013.

[74] M. Hirsch and H. Smith, "Monotone dynamical systems," *Handbook of Differential Equations: Ordinary Differential Equations*, vol. 2, 01 2004.

[75] F. Bullo, *Contraction Theory for Dynamical Systems*, 1.1 ed. Kindle Direct Publishing, 2023. [Online]. Available: https://fbullo.github.io/ctds

[76] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transportation Science*, vol. 11, no. 3, pp. 253–274, 1977.

[77] M. Ben-Akiva and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA, USA: The MIT Press, 1985.

[78] C. Canudas de Wit, F. Morbidi, L. L. Ojeda, A. Y. Kibangou, I. Bellicot, and P. Bellemain, "Grenoble Traffic Lab: An experimental platform for advanced traffic monitoring and forecasting," *IEEE Control Systems Magazine*, vol. 35, no. 3, pp. 23–39, 2015.

[79] M. Rodriguez-Vega, L. Senique, and C. Canudas de Wit, "GTL-VILLE: Experimental platform for urban traffic state estimation," in *2022 - 6èmes journées des Démonstrateurs en Automatique*, Angers, France, 2022, pp. 1–9. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03694936

[80] P. Kachroo and S. Sastry, "Traffic assignment using a density-based travel-time function for intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1438–1447, 2016.

[81] A. M. Bayen, A. Keimer, E. Porter, and M. Spinola, "Time-continuous instantaneous and past memory routing on traffic networks: A mathematical analysis on the basis of the link-delay model," *SIAM Journal on Applied Dynamical Systems*, vol. 18, no. 4, pp. 2143–2180, 2019.

[82] Y. Hino and T. Nagatani, "Effect of bottleneck on route choice in two-route traffic system with real-time information," *Physica A: Statistical Mechanics and its Applications*, vol. 395, pp. 425–433, 2014.

[83] J. Wahle, A. L. C. Bazzan, F. Klügl, and M. Schreckenberg, "Decision dynamics in traffic scenario," *Physica A: Statistical Mechanics and its Applications*, vol. 287, no. 3, pp. 669–681, 2000.

[84] E. Friedman, *Introduction to time-delay systems: Analysis and control*. Basel, Switzerland: Birkhauser, 2014.

[85] J. Wei, "Bifurcation analysis in a scalar delay differential equation," *Nonlinearity*, vol. 20, no. 11, pp. 2483–2498, 2007.

[86] Mapbox, "Traffic data," last accessed 05 June 2024. [Online]. Available: https://www.mapbox.com/traffic-data

[87] Google, "Get traffic data with the Waze data feed," last accessed 05 June 2024. [Online]. Available: https://support.google.com/waze/partners/answer/10618035?hl=en

[88] G. Sharon, M. Albert, T. Rambha, S. Boyles, and P. Stone, "Traffic optimization for a mixture of self-interested and compliant agents," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[89] Z. Chen, X. Lin, Y. Yin, and M. Li, "Path controlling of automated vehicles for system optimum on transportation networks with heterogeneous traffic stream," *Transportation Research Part C: Emerging Technologies*, vol. 110, pp. 312–329, 2020.

[90] K. Zhang and Y. M. Nie, "Mitigating the impact of selfish routing: An optimal-ratio control scheme (ORCS) inspired by autonomous driving," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 75–90, 2018.

[91] Aimsun SLU, "Aimsun." [Online]. Available: https://www.aimsun.com

[92] D. Angeli and E. Sontag, "A tutorial on monotone systems - with an application to chemical reaction networks," http://www.sontaglab.org/FTPDIR/04mtns-leenheer-angeli-monotone.pdf, 2004.

Résumé — Les systèmes de transport jouent un rôle central dans la vie urbaine, structurant les déplacements de personnes et de marchandises, stimulant la croissance économique et influençant la durabilité et l'équité sociale. Les innovations récentes, telles que les applications de navigation, les services de covoiturage et les systèmes de gestion intelligente du trafic, ont profondément modifié le comportement des usagers et les dynamiques de circulation. Ces technologies promettent une optimisation en temps réel, mais introduisent également des complexités et des inefficacités. Cette thèse analyse leurs effets sur l'efficacité du trafic à l'aide de modèles de réseaux de trafic et de la théorie des jeux, mettant en lumière à la fois leurs avantages et potentiels inconvénients. La première partie de cette thèse examine l'impact des usagers des applications de navigation qui suivent des recommandations visant à minimiser leur temps de trajet. Nous commençons par définir et analyser un jeu de routage non atomique intégrant des contraintes de capacité routière. Ce cadre amélioré montre que le routage basé sur la minimisation des temps de trajet peut conduire à des configurations de trafic sous-optimales, caractérisées par le phénomène de transfert partiel de la demande, où seule une partie de la demande totale de trafic peut être efficacement prise en charge par le réseau. En nous appuyant sur ce modèle statique, nous étendons l'analyse à un modèle dynamique de flux de réseau, prenant explicitement en compte à la fois les usagers bénéficiant des recommandations en temps réel des applications de navigation et ceux qui n'en bénéficient pas. L'analyse de stabilité de ce modèle dynamique établit un lien direct entre la proportion d'usagers informés, la demande totale de trafic et l'émergence du transfert partiel de la demande. Plus précisément, les résultats montrent que l'augmentation du taux de pénétration des usagers informés et de la demande globale accroît la probabilité de ce phénomène. Dans ce cadre dynamique, nous explorons également l'impact des délais dans les recommandations de routage fournies par les applications de navigation. En examinant l'interaction entre les délais, la demande de trafic et le taux de pénétration des usagers informés, nous montrons comment ces facteurs influencent collectivement la stabilité et l'efficacité du système de trafic. Les résultats révèlent que, dans des conditions de forte demande et d'utilisation généralisée des applications, des délais suffisamment importants peuvent aggraver les inefficacités, entraînant un transfert partiel de la demande et une dégradation des performances globales du réseau. La deuxième partie se concentre sur le rôle des flottes de véhicules coordonnées dans les services de mobilité moderne, tels que les services de covoiturage. À l'aide de jeux de routage à comportements mixtes, cette partie analyse comment des flottes coordonnant leurs trajets pour minimiser leur temps de trajet moyen interagissent avec des usagers individuels optimisant leurs propres trajets. Les résultats montrent que les petites flottes ont une influence minimale sur la distribution globale du trafic, tandis que les grandes flottes peuvent améliorer l'efficacité du réseau, ce qui suggère que le routage coordonné constitue une piste prometteuse pour les stratégies de gestion du trafic.

Mots clés : Réseaux de transport, systèmes de transport intelligents, théorie des jeux, flux dynamiques sur réseaux.

Abstract — Transportation systems are central to urban life, shaping the movement of people and goods, driving economic growth, and influencing sustainability and social equity. Recent innovations, including navigation apps, ride-hailing services, and smart traffic management systems, have profoundly impacted user behavior and traffic patterns. These technologies promise real-time optimization but also introduce complexities and inefficiencies. This thesis analyzes their effects on traffic efficiency using traffic network models and game theory, highlighting both potential benefits and drawbacks. The first part of this thesis investigates the impact of navigation app users who follow recommendations aimed at minimizing travel time. We begin by defining and analyzing a nonatomic routing game that incorporates road capacity constraints. This refined framework reveals that routing based on travel time minimization can result in suboptimal traffic configurations, marked by the phenomenon of partial demand transfer, where only a portion of the total traffic demand can be accommodated by the network. Building on this static model, we extend the analysis to a dynamic network flow model, explicitly accounting for both users who benefit from real-time navigation app recommendations and those who do not. The stability analysis of this dynamic model establishes a direct link between the proportion of informed users, the overall traffic demand, and the emergence of partial demand transfer. Specifically, the results demonstrate that as the penetration rate of informed users and the total traffic demand increase, the likelihood of this phenomenon also rises. Within this dynamical framework, we further explore the impact of delays in the routing recommendations provided by navigation apps. By examining the interplay between delays, traffic demand, and the penetration rate of informed users, we reveal how these factors collectively affect the stability and efficiency of the traffic system. The findings show that under conditions of high demand and widespread app usage, sufficiently large delays can exacerbate inefficiencies, leading to partial demand transfer and reduced overall network performance. The second part focuses on the role of coordinated vehicle fleets in modern mobility services like ride-hailing. Using mixed-behavior routing games, it analyzes how fleets coordinating to minimize their average travel time interact with individual users optimizing their own routes. The findings show that small fleets have minimal influence on overall traffic, but larger fleets can enhance network efficiency, suggesting that coordinated routing holds promise for traffic control strategies.

Keywords: Transportation networks, intelligent transportation systems, game theory, dynamical network flows.

Grenoble Images Parole Signal Automatique (GIPSA-Lab),
11 rue des Mathématiques, 38400 Saint-Martin-d'Hères